Spring 5-1-2018

Design of mononuclear nonheme enzyme for oxygen activation

Joel John Rempillo

Design of mononuclear nonheme enzyme for oxygen activation

A Capstone Project Submitted in Partial Fulfillment of the
Requirements of the Renée Crown University Honors Program at
Syracuse University

Joel John Rempillo

Candidate for Bachelor of Science
and Renée Crown University Honors
Spring 2018

Honors Capstone Project in Biochemistry

Capstone Project Advisor: _____
Olga V. Makhlynets, PhD
Assistant Professor

Capstone Project Reader: _____
Ivan V. Korendovych, PhD
Associate Professor

Honors Director: _____
Chris Johnson
Interim Director

Abstract

Natural enzymes are complex systems with intricate folds and binding pockets, which recognize specific substrates and position them ideally for catalysis. Large size and complexity of enzymes prevent their study and use as industrial catalysts. On the other hand, *de novo* designed models of enzymes often use simple and stable protein scaffold, which streamlines mechanistic studies and allows industrial applications. In this work, we use previously designed four-helix bundle as a protein framework to incorporate 2-His-1-carboxylate motif often found in non-heme enzymes. Such coordination environment provides ligands to bind iron(II) and leaves three additional coordination sites to accommodate substrate and oxidant. The resulting protein named MFsc is folded, thermally stable and binds a single Fe(II) ion. MFsc protein loaded with Fe(II) catalyzes oxidation of phenols by dioxygen, thus mimicking function of catechol dioxygenases, natural enzymes that have 2-His-1-carboxylate motif. One of the intermediates in the catechol dioxygenase mechanistic pathway is Fe(II)-semiquinone radical (SQ•) adduct. This intermediate has been observed in the Fe(II)- MFsc model system and characterized by UV-vis spectroscopy and EPR. We recently discovered that even apo protein can stabilize SQ• radical for weeks. A combination of NMR studies and docking showed that the substrate sits in the hydrophobic pocket with hydroxyl groups facing the negatively charged interior of the protein.

Executive Summary

Proteins carry out all the functions needed for cells to survive. They are essentially a long chain of amino acids that fold into a distinct form. There are 20 natural amino acids in living organisms that make up proteins. All amino acids have the same backbone where they are linked, but differ in its side chain (Figure ES1). These side chains are characterized as hydrophobic, polar, or charged. An average protein has a chain of about 300 amino acids (or residues), where it initially forms three distinct secondary structures (i.e. α-helix, β-strands, and random coils) and further folds into its tertiary structure. Hydrophobic amino acids usually form the core and pockets in its three-dimensional shape, while the polar and charged side chains are critical in recognizing its target molecule and performing its function.

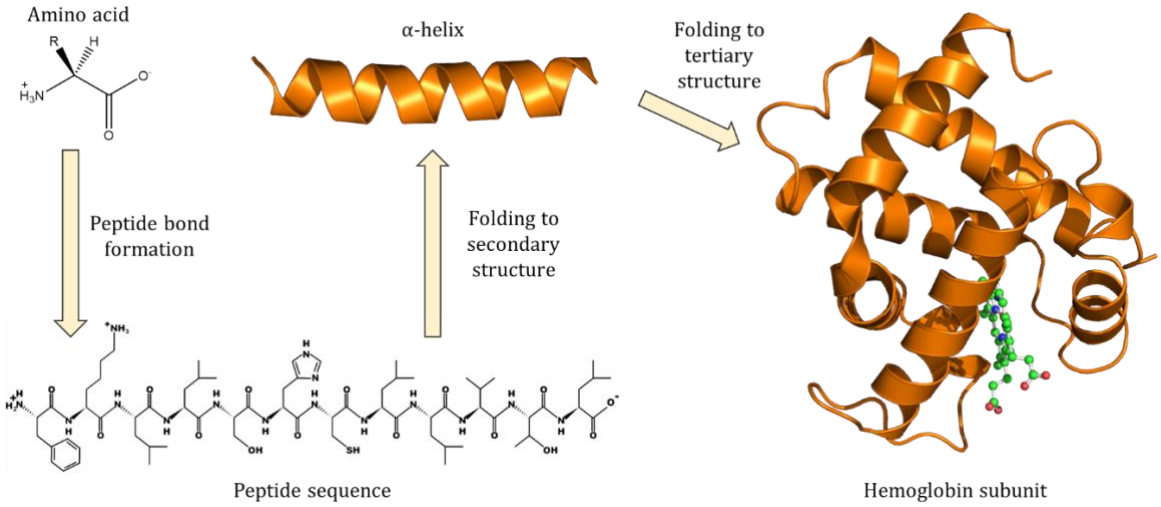


Figure ES1. Amino acids have common amine (NH_3^+) and carboxylate COO^- domains and varying R groups**.** A peptide is formed when several amino acids covalently bond. The peptide folds into a secondary structure such as α-helix (shown), β-sheets and random coils. These secondary structures further folds into a tertiary structure such as the hemoglobin subunit shown above (PDB ID: 1A3N).

The pockets formed are complementary to the shape of the molecule it binds, allowing high specificity even at a single bond. Proteins that catalyze biochemical reaction are called enzymes. They perform a wide range of functions from DNA replication to cell degradation, earning the term "biological machines."

Due to its high efficiency in recognizing target molecules (or substrates) and catalyzing reactions, designing proteins for a desired function has been the "Holy Grail" of biochemistry. Understanding how a protein attains its final three-dimensional structure based solely on its sequence is a difficult task to accomplish. Computational methods may provide insights on how protein folding is achieved by applying the concepts in classical physics. However, positioning the amino acids required for catalysis complicates the process of the *de novo* design of enzymes. Replicating the efficiency of enzymes, both in molecule recognition and catalysis, allows us to design proteins capable of catalyzing target reactions that can be used in the industry and even in medicine, an emerging field called *protein engineering and design*. However, mimicking enzyme functions has only shown moderate success.

One particular product from this field is the Due Ferri single chain (DFsc) protein, a *de novo* designed enzyme that can transform phenols into quinones[1]. It is a small 115-amino acid protein developed from a cluster of designed peptides facilitated by metal ions, which enables oxidation reaction. Given this protein's size, it can serve as a scaffold for designing an enzyme that is small enough to be studied with considerable detail.

Moreover, comprehensive review of a reaction mechanism that occurs in natural enzymes offers clues on how enzymes can be designed. Metal ions such as calcium, magnesium, and transition metals have been used extensively by most enzymes to carry

out their function. In this work, we aimed to design an enzyme that binds a single metal ion to oxidize target molecules. Recent studies have found several enzymes that perform oxidation reactions involving an iron(II) ion. These enzymes bind iron(II) in a common pattern, where two histidine and one carboxylate residues are positioned closely. The binding of iron(II) by these residues facilitates the activation of O_2, which ultimately oxidizes the bound substrate. Using this natural pattern, we incorporated it to the DFsc protein by introducing specific mutations on the residues that bind the two metal ions. We also widened the substrate channel to increase catalytic efficiency by substituting certain alanine residues into glycine. With this approach, we can design an enzyme using a *de novo* designed protein, and can also further our understanding on how protein structure can influence substrate selectivity and function.

Unfortunately, but interestingly, the results presented in this work dictate that the protein did not catalyze the oxidative cleavage of catechols. Rather, a reaction intermediate species was detected. Analytical experiments confirmed that this species is a highly reactive radical trapped and stabilized by the designed protein, which is a phenomenon not commonly observed in nature. Our mononuclear iron(II) protein called **MFsc-4G** exhibited this ability with bound metal. Additionally, the model protein DFsc also stabilized this radical intermediate but in the absence of metal ions. These findings drove this project to study how such proteins can quench these reactive radical species.

Studying this phenomenon would allow us to understand what residues are responsible for stabilizing radicals. Thus, gaining a basis on how we can redesign MFsc-4G to complete the target reaction. Furthermore, outcomes from this study open avenues to such protein designs that can stabilize harmful radical species. *De novo* design of proteins

with the desired function not only serves as a tool to study protein structure-function relationship but may also be used to degrade harmful compounds and replace harsh methods used to transform chemicals in the industry.

Table of Contents

Chapter 1
Introduction

Proteins exist in all biological systems and perform a myriad of crucial functions. Structural proteins such as tubulins polymerize into rigid but dynamic microtubules which provide mechanical support in cells. It also serves as a highway for a family of motor proteins – kinesin and dynein – that transport cellular components. Antibodies are large proteins that recognize and bind antigens, biomarkers for the immune system, with high specificity. Similarly, receptor proteins span the cell membrane to detect extracellular signaling molecules and initiate a metabolic pathway. While these classes of proteins are essential in living systems, studying enzymes attracts the most attention as they are the key in all biological processes.

1.1. Enzymes and Metalloenzymes

Enzymes are biological machines with intricate structures that use precise binding interactions to catalyze chemical transformation of the substrate. These macromolecules are not expended during the reaction, but rather allow an unfavorable chemical reaction to proceed at an incredibly low energy cost. Natural enzymes are composed of amino acid chains that fold into a three-dimensional structure, forming elaborate pockets that are complementary to their target substrate's shape. Additionally, the residues that sculpt the pocket optimize its intermolecular interactions with the substrate and stage it for catalysis, which contributes to its high selectivity. About a third of known enzymes are classified as metalloenzymes, which contain bound metals as prosthetic groups that either participate in catalysis or act as regulators[1].

For example, calmodulin is a well-studied multifunctional metalloenzyme in eukaryotic cells that requires allosteric binding of calcium ions (Ca^{2+}) to form the binding site[2]. Zinc(II) ion facilitates the conversion of carbon dioxide (CO_2) to carbonic acid (H_2CO_3) by the enzyme carbonic anhydrase when we consume carbonated beverages. Biochemists have poured time and effort in investigating the mechanisms of natural enzymes in great detail as we know today. However, their large size and incredibly complex folding remains as the challenging component in understanding its structure-function relationship.

1.2. Protein Engineering

The emerging field of *protein engineering* seeks to design biomimetic proteins, which exploits amino acid properties to manipulate protein folding or its interactions with a specific substrate for a desired reaction. Replicating such functions with high efficiency in

de novo designed proteins is deemed the Holy Grail of biochemistry. Recent studies have only shown modest success in designing functional enzymes by incorporating an active site into a natural protein[3]. Using this method minimizes the problem of generating other possible or unwanted conformations, and only requires a few alterations. Another approach that is gaining more popularity is the *de novo* protein design. In this approach, the exact sequence and side chain positions are unknown, hence the backbone and side chains are designed simultaneously[3]. Designing proteins applies the physical concepts in protein folding using computational methods. For instance, a library of side chain rotamers are sampled around the desired active or binding site until the algorithms determine a stable model[4]. Natural proteins are typically composed of more than 100 units from a selection of 20 canonical amino acids and a variety of post-translational modifications, yielding a virtually infinite number of possible conformations. Thus, designing *de novo* proteins is limited to small sizes or fragments so it can be managed computationally and experimentally[5–7]. This new phase in protein studies provides a novel method in understanding the nature of protein folding and function and serves as a tool to build synthetic enzymes.

1.3. Review of Related Literature

1.3.1. Due Ferri single chain protein: a *de novo* designed enzyme

DeGrado *et al.* designed a catalytic metalloprotein, using one of the Due Ferri (DF) proteins as a starting scaffold, that can oxidize phenols into benzoquinones[3,8]. Peptide sequences were computationally designed to self-assemble into a heterotetrameric α-helical bundle by chelating two zinc(II) ions. These peptide tetramers have shown enzyme-

like activity that follows the Michaelis-Menten kinetics model with a catalytic efficiency (k_{cat}/k_M) of 1540 $M^{-1}\cdot min^{-1}$ (Figure 1.1, *right*). Moreover, the peptide combination with the highest activity was re-designed into a 115-amino acid single strand named Due Ferri single chain or DFsc protein (Figure 1.1, *left*) which can be expressed easily using DNA recombinant technology[3,9].

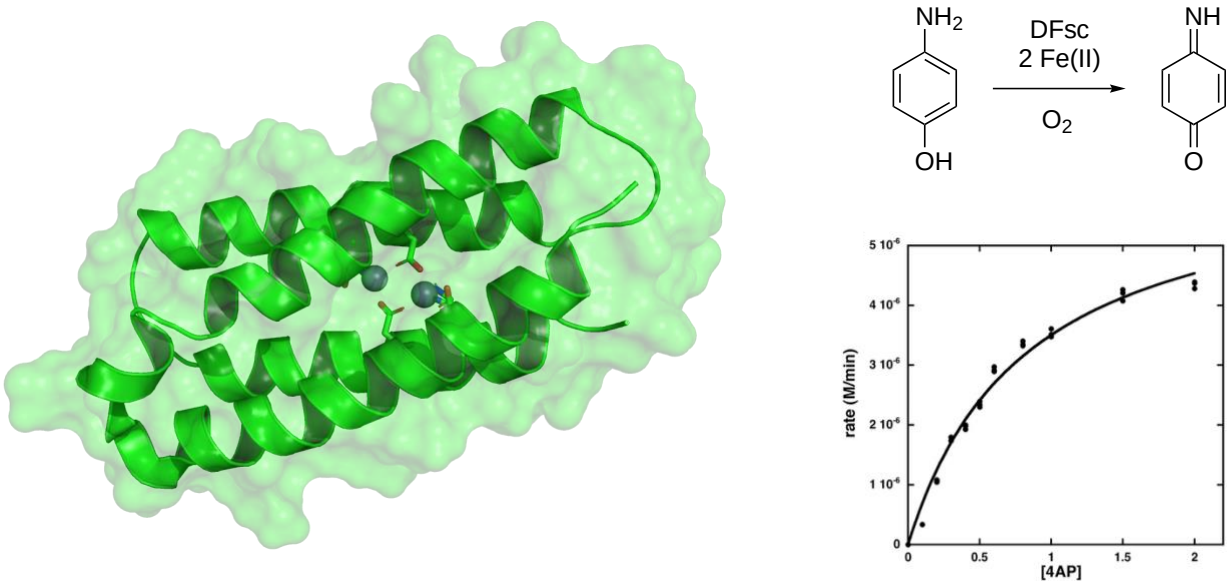


Figure 1.1. **Due Ferri single chain protein** (DFsc, PDB ID: 2HZ8). Cartoon and surface representation of DFsc (***left***), a small and stable *de novo* designed protein that catalyzes oxidation of 4-aminophenol into benzoquinone monoimine effectively (***right***)[3].

DFsc is folded and stable in the absence of metals, but gains function when two equivalents of metal ions such as iron(II) or zinc(II) are bound. The two metals are proximal to each other (4.2 Å apart) in the active site (Figure 1.2, *left*), and facilitate the two-electron transfer in the reaction which prevents the formation of radicals[3]. The metal ions are coordinated by six residues: Glu11, Glu44, His77, His107, and two critical carboxylate bridges Glu74 and Glu104. The active site is buried in the core of the protein

and can be accessed through a narrow hydrophobic channel forming favorable interactions with the aromatic ring of the substrate (Figure 1.2, *right*).

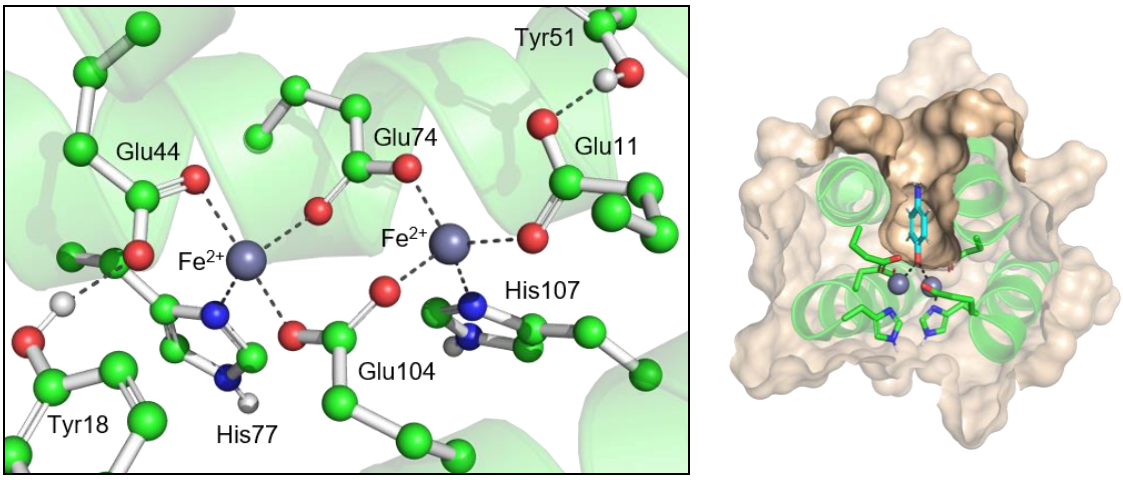


Figure 1.2. DFsc active site. (*left*) The symmetric active site anchors two Fe(II) ions by four glutamates and two histidine residues. Tyr18 and Tyr51 form the secondary shell interactions that stabilize Glu44 and Glu11, respectively. (***right***) A 4-aminophenol molecule (cyan) coordinates with the two metal ions buried in the center of the protein.

Further alterations on the DFsc active site, termed 2A3H-DFsc (PDB ID: 2LFD), revealed its ability to stabilize an unstable semiquinone radical (SQ•) anion in water[10]. Radicals are a very reactive species but are common in enzyme reactions as intermediates. Since enzymes trap the substrate during catalysis, radical intermediates are constrained in the active site to prevent unwanted reactions. This mechanistic feature enforces a specific reaction pathway and the radical is quenched at the end of the catalytic cycle[11]. The radical intermediate was observed spectroscopically when the protein was reacted with 3,5-di-*tert*-butylcatechol. 2A3H-DFsc forms the [2A3H-DFsc-Zn(II)$_2$-SQ•] complex and is stable for more than 120 minutes compared to the nanoseconds half-life of radical intermediates[10–12].

Although 2A3H-DFsc showed this amazing potential, it requires two zinc(II) ions to fold and produce the radical intermediate.

1.3.2. The 2-His-1-carboxylate facial triad motif

Natural metalloenzymes often use only one metal prosthetic group on their active site. In dioxygenases, transition metals such as Fe(II) and Zn(II) are often used for the electron transfer chemistry since these metal ions can adopt multiple oxidation states. The metal ion is situated close the active site and bound by the coordinating ligands of the protein. The unsaturated ligand coordination fine-tunes the metal ion rendering it more reactive. A common structural motif of this coordination environment in nature is referred to as the *2-His-1-carboxylate facial triad*, which has been studied extensively from a number of mononuclear non-heme iron(II) dioxygenases (Figure 1.3)[13].

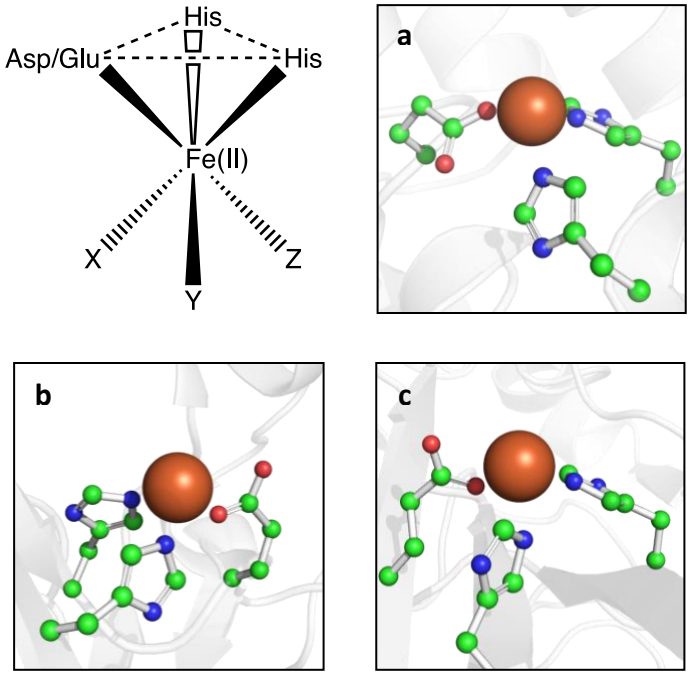


Figure 1.3. 2-His-1-carboxylate facial triad. Fe(II) is anchored by two histidine and one carboxylate residues, occupying a face of the metal ion's octahedron (***top left***). Three labile ligands (X, Y, and Z) on Fe(II) coordination sphere are available sites for cofactor, substrate, or oxidant. (**a**) Active site of phenylalanine hydroxylase binds Fe(II) with this motif to oxidize phenylalanine to tyrosine (PDB ID: 1PAH). (**b**) Clavaminate synthase uses Fe(II) for clavulanic acid biosynthesis (PDB ID: 1DS1). (**c**) Metapyrocatechase exploits Fe(II) for the ring cleavage of catechols (PDB ID: 1MPY).

Several orthologous dioxygenases were characterized with this motif and although they differ significantly in the amino acid sequences, these enzymes follow a common electron transfer mechanism to oxidize their substrate. Generally, two histidine and a side chain carboxylate (Glu or Asp) residues occupy a triangular face of the octahedral coordination sphere of iron(II), vacating three sites for a substrate, cofactor and dioxygen (Figure 1.3, *top left*). The three remaining coordination sites of the bound Fe(II) are saturated with water molecules in aqueous media (Figure 1.4A).

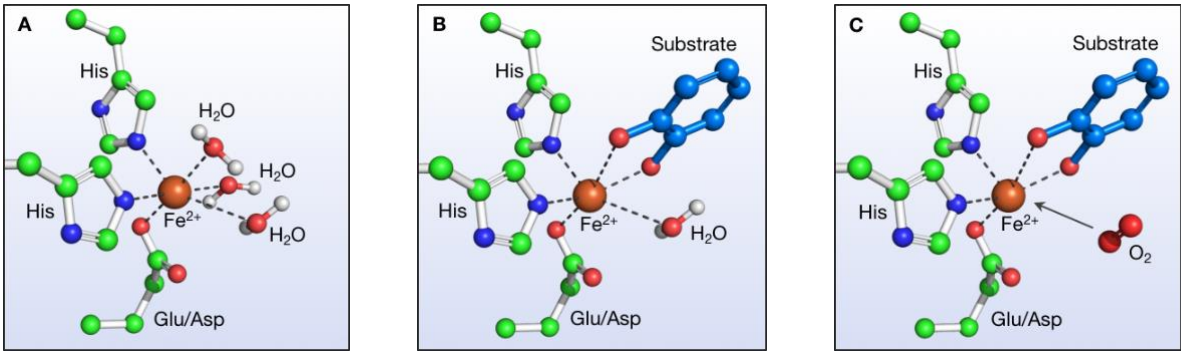


Figure 1.4. 2-His-1-carboxylate facial triad motif consists of Fe(II) coordinated by two histidines and one carboxylate (Glu, Asp) and the remainder of coordination sphere is occupied by labile ligands (H_2O) (**A**). A substrate or cofactor replaces two water molecules (**B**) leaving one coordination site for dioxygen attack (**C**).

Bidentate substrates such as catechols and pyruvate-derived compounds are positioned within the enzyme's pocket and replaces labile water molecules (Figure 1.4B). For catechol-cleaving dioxygenases (CDOs), the five-coordinated Fe(II) ion is primed for dioxygen attack (Figure 1.4C), completely saturating the metal ion to form a superoxo-Fe(III) species (Fe(III)-O_2•)[13–16]. The formation of this intermediate complex is committed to oxidize the adjacent substrate confined in the active site. A class of enzymes called α-ketoglutarate-dependent dioxygenase such as taurine dioxygenase (TauD) uses αKG as the

bidentate co-substrate to activate oxygen, since not all substrates are bidentate that can directly bind Fe(II)[17–19]. In phenylalanine hydroxylase (PheH, Figure 1.3a, PDB ID: 1PAH), the cofactor tetrahydrobiopterin forms a peroxide bridge with PheH-Fe(II) in the presence of O_2. It is followed by the oxidation of the cofactor and the enzyme complex into a Fe(IV)O species, which ultimately hydroxylates the substrate[14,20,21]. Evidently, these enzymes follow different mechanisms for dioxygen activation, but use the same ligands and coordination environment for Fe(II). It is an interesting structural feature particularly on the diverse oxidation reactions it can catalyze.

1.4. Purpose and Objectives

The aim of this study is to design an unnatural mononuclear non-heme iron(II) protein capable of activating dioxygen to oxidize organic molecules. We linked two approaches in protein engineering to design our artificial catalyst. Here, we introduced a single mutation (E104H) to transform DFsc from dinuclear protein into mononuclear protein named Mono Ferri single chain (MFsc) which supports a single metal ion by 2-His-1-carboxylate facial triad motif. Furthermore, we expanded the binding pocket by substituting selected Ala residues with Gly to increase the substrate channel, allowing substrate to easier access the metal center. The resulting protein binds Fe(II) ion and maintained its α-helix structure, thus we renamed the former Mono Ferri single chain protein (MFsc) to MFsc-4G. Our recent spectroscopic experiments with 3,5-di-*tert*-butylcatechol displayed a stabilized SQ• similar to the event observed with 2A3H-DFsc-$Zn(II)_2$ complex. We did not expect to detect SQ• with diiron-type proteins such as DFsc-$Fe(II)_2$ complex, nor the apo form of MFsc-4G, suggesting that only one metal is required.

When tested for metal dependence using DFsc, and found lower absorbance of SQ• with 1:1 protein-to-metal ratio. Interestingly, DFsc also exhibited this phenomenon in the absence of metal ions. This is an example of a stabilized radical by an apo-protein and protein-Fe(II) complex in aqueous solution. Mechanistic studies defining the catechol cleavage in the 2-His-1-carboxylate facial triad motif generally show a radical intermediate that is too short-lived to be detected using UV-vis spectroscopy. Additionally, characterizing the protein-ligand interactions *in silico* using Rosetta combined with NMR studies provide insight on the structure of DFsc-SQ• complex. The results herein describe the potentials of a mononuclear non-heme iron(II) protein and a *de novo* designed protein in using its binding energy to trap highly reactive intermediates. Although this work diverged into two interesting projects, our approach to build such a versatile enzyme offer perspectives on how residues manipulate the coordination sphere of transition metals to activate O_2. Ultimately, understanding the dynamics of enzymes through protein engineering and *de novo* design would allow novel methods in developing synthetic enzymes for applications in industry, medicine, biomaterials, and nanotechnology.

Chapter 2
Design of mononuclear iron(II) protein

2.1. DFsc protein as a modelling scaffold

The DFsc protein is a suitable scaffold to incorporate the 2-His-1-carboxylate motif. Its simplistic structure provides a stable framework and would require minimal alterations to accommodate a mononuclear iron(II) center. We propose that altering Glu104 bridge to His on DFsc would support a mononuclear iron(II) center giving rise to the MFsc protein. Moreover, the resulting deletion of a metal-binding site would offer additional space for optimal positioning of the substrate. Exposing the active site by replacing protein residues into smaller analogs within the substrate channel allows the protein to accommodate larger substrates, which in turn facilitates catalysis. This chapter describes how the active site of DFsc was altered to develop MFsc-4G and determine experimentally which residues are critical to binding the metal ions.

2.2. Results and Discussion

2.2.1. Site-directed Mutagenesis

The DNA sequence that codes for the native DFsc protein was cloned into the pMCSG49 vector, which fuses the protein of interest with a His_6 tagged TEV domain on the N-terminus (see Experimental). The mutation E104H was introduced into the DFsc gene, which translates into a protein named the **MFsc**. For increasing the binding pocket volume, four alanine residues were mapped along the substrate channel, and Ala-to-Gly mutations were introduced in two parts to ensure the stability of the protein. Initially, site-directed mutagenesis (SDM) was performed on the MFsc template with A32G and A36G primers for A10G and A14G mutations, since the gene is fused with a 22-amino acid N-terminus His_6-taged TEV domain. The resulting mutations in the MFsc-2G gene were verified using DNA gel electrophoresis and Sanger sequencing (commercially done by GenScript) (Figure 2.1a). The last part of the substrate channel mutations was introduced on the MFsc-2G gene using A65G A69G primers. The final A43G and A47G mutations were confirmed by agarose gel electrophoresis and Sanger sequencing (Figure 2.1b).

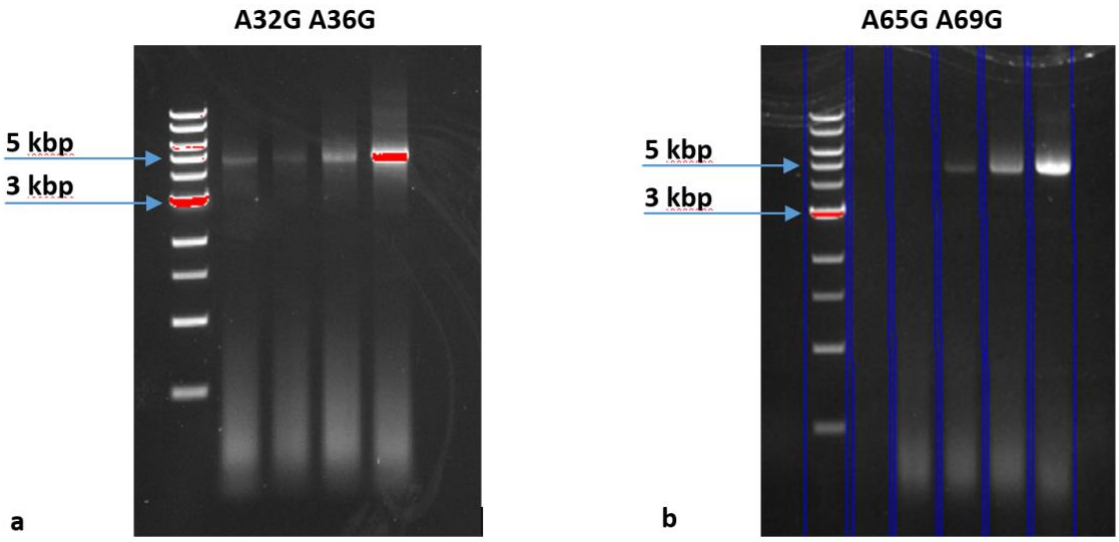


Figure 2.1. DNA gel electrophoresis of A10G A14G **(a)** and A43G A47G **(b)** site-directed mutagenesis on ethidium bromide-stained 1.0% agarose gel. Pronounced bands in line with the 5 kbp reference indicate that the plasmid was extended. Samples with intense bands were digested with DpnI, replicated in *E. coli*, and sent the pure plasmids for sequencing.

Plasmids with the four Ala-to-Gly mutations were purified via MiniPrep (see Experimental) and were submitted to GenScript to validate the sequence. The mutation was confirmed by realigning the sequencing reports with the MFsc template, showing four Ala-to-Gly substitutions on the desired positions (i.e. 10, 14, 43, and 47). The resulting gene was named MFsc-4G gene, and all other mutations were introduced on this template for subsequent studies.

Several mutations were introduced within the active site to examine the influences of the neighboring polar residues towards the binding and stability of the Fe(II) center. In DFsc protein, the carboxylate bridges on positions 74 and 104 anchor the two metal ions in place (Figure 1.2). Replacing bridging E104 with a histidine would allow to eliminate a ligand for one of the metal ions and at the same time increase number of histidine ligands

in the coordination environment of another metal ion. Glu74 is only 2.0 Å from Fe(II), which can interact as a ligand to the metal ion. With two histidines and two glutamates binding Fe(II), this leaves the metal ion with two available coordination sites. Since Glu is a charged polar amino acid, substituting it to its uncharged polar analog, Gln, could provide information on its role in Fe(II) binding. Thus, MFsc-4G E74Q mutant was also made using protocols outlined above.

Additionally, we aim to analyze the positioning of the metal ion in the active site. The metal-binding sites of MFsc-4G protein is symmetrical, so it can be projected that the metal ion can occupy either side. Further mutations were introduced: Gln mutations at positions 11 and 107, replacing the Glu and His respectively (Figure 1.2). Glu11 and His107 mutations aim to investigate how these residues affect metal binding.

Primers for E11Q, E74Q, and H107Q mutagenesis were designed and these mutations were introduced separately into the MFsc-4G template. The resulting mutations were verified with DNA gel electrophoresis (Figure 2.2), and purified samples of these mutations were sent to Genewiz for sequencing to verify the accuracy of mutagenesis (see Experimental). The resulting mutant plasmids were labeled as E11Q, E74Q, and H107Q plasmids.

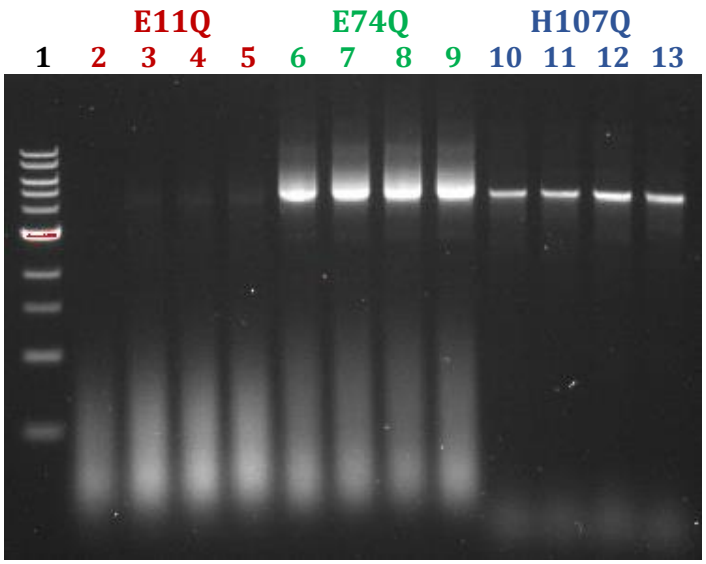
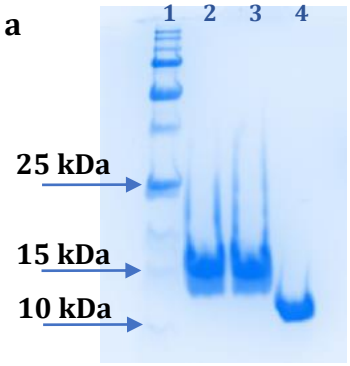


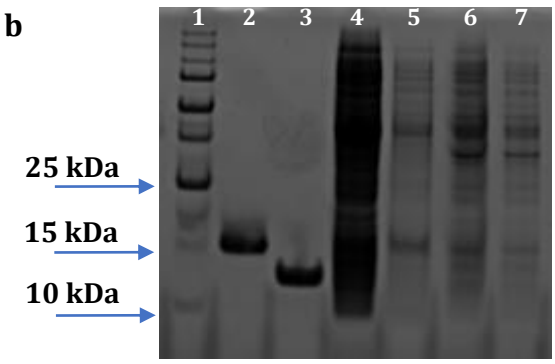
Figure 2.2. SDM PCR samples of E11Q (**Lanes 2-5**), E74Q (**Lanes 6-9**), and H107Q (**Lanes 10-13**) on ethidium bromide-stained 1.0% agarose gel under UV light. E11Q mutation have faint bands compared to the well-pronounced E74Q and H107Q.

2.2.2. Expression and Purification

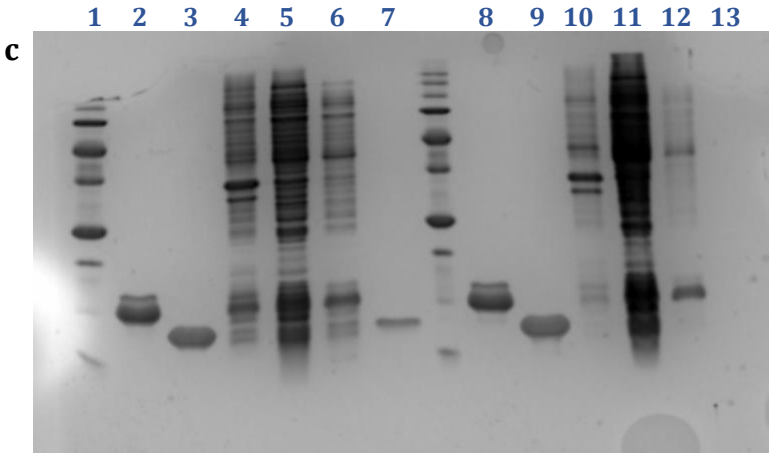
All plasmid variants were transformed in BL21 (DE3) *E. coli* cells (see Experimental). One protocol was used for protein expression and purification for all mutants. Each protein, namely MFsc-4G and its E11Q, E74Q, and H107Q mutants, were purified separately using Ni-NTA affinity chromatography (see Experimental). The His_6-tag was cleaved to isolate the native proteins for each variant (see Experimental). Samples of these proteins were analyzed for purity using SDS-PAGE (Figure 10a, b, c). The theoretical molecular weight of each protein with His_6-tag is 16.0 kDa and the native protein is 13.49 kDa. The bands on the gel indicate high yields of proteins.



MFsc-4G
1. Ladder;
2. Before TEV removal and before addition of EDTA;
3. Before TEV removal and after incubation with EDTA after TEV removal;
4. Purified MFsc-4G after TEV removal.



E11Q
1. Ladder;
2. Pre-TEV;
3. After TEV;
4. Lysate supernatant;
5. Wash;
6. Cell debris;
7. Cell debris 2X diluted.



E74Q
1 – Ladder;
2 – Pre-TEV;
3 – After TEV;
4 – Cell debris;
5 – Flow through;
6 – Wash;
7 – Concentrator flow through.

H107Q
8 – Pre-TEV;
9 – After TEV;
10 – Cell debris;
11 – Flow through;
12 – Wash;
13 – Concentrator flow through.

Figure 2.3. SDS-PAGE analysis of MFsc-4G (**a**), MFsc-4G E11Q (**b**), MFsc-4G E74Q and MFsc-4G H107Q (**c**) proteins for purity at different stages during the purification process.

2.2.3. Metal Binding

We expected that the MFsc-4G proteins would bind metal ions in a 1:1 ratio. However, bound Fe(II) ions are poorly observed with UV-vis spectroscopy, thus, an

alternative and efficient metal ion was used. Co(II) changes its extinction coefficient (ε) significantly when bound to the protein, providing an elaborate and quantifiable signal that can be measured by spectroscopic methods.

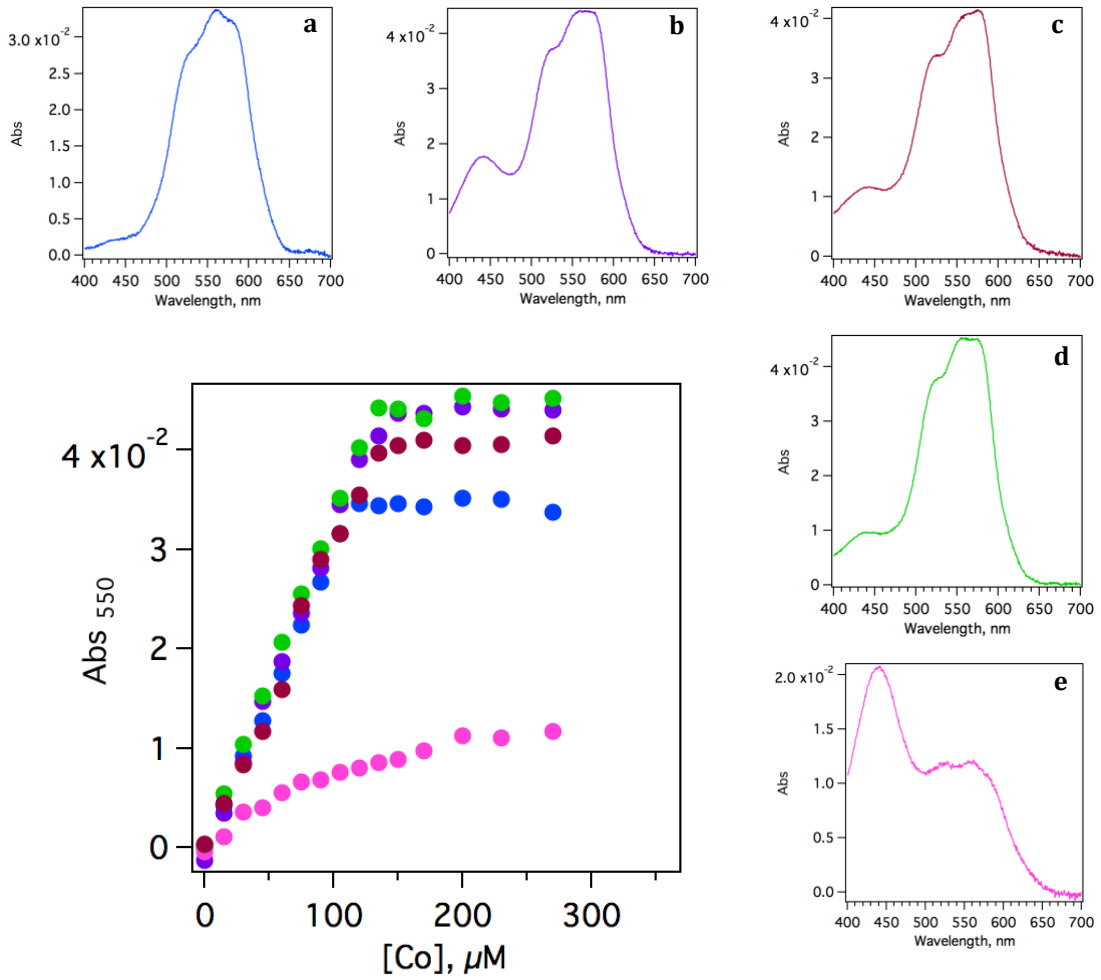


Figure 2.4. UV-vis spectra of cobalt bound to MFsc, MFsc-4G, and its mutants. (**a**) MFsc protein (150 μM) binds a maximum concentration of 120 μM Co(II), giving a metal-to-protein ratio of 0.8:1. (**b**) **MFsc-4G** exhibits a 1:1 ratio, after widening the substrate channel. (**c, d**) E11Q and H107Q share similar features, where both show 0.9:1 metal-to-protein ratio. (**e**) E74Q did not display any metal-binding features with Co(II). An uncharacterized peak was observed at 440 nm for all protein-Co(II) complexes except for MFsc. (**lower left**) Absorbance at 550 nm as a function of cobalt concentration. Data are color coded as in **a-e**.

MFsc, MFsc-4G, and MFsc-4G mutants were incubated with increasing Co(II) concentrations from 0 to 270 μM in buffer at pH 7.6 (see Experimental). The spectra of these protein-Co(II) complexes were recorded between 400 nm to 700 nm. A peak that increases with increasing Co(II) concentration was also observed at 440 nm in MFsc-4G-Co(II) complexes and its mutants. This species is still unknown.

The absorbance at 550 nm for each protein were normalized at 700 nm and plotted as shown on the lower left scatter plot of Figure 2.4. The linear increase of the absorbance on this plot from 0 to ~150 μM is anticipated as a result of increasing concentrations of Co(II) bound to the protein. The scatter plot becomes flat starting at a certain concentration, which implies the maximum Co(II) concentration bound to the protein. The excess unbound Co(II) ions in solution do not contribute to the absorbance of the complex, thus the Co(II) concentration where the absorbance starts to plateau indicates the maximum amount bound to the 150 μM protein. The scatter plot suggests a ~1:1 metal-to-protein binding ratio with the MFsc, MFsc-4G, MFsc-4G E11Q, and MFsc-4G H107Q mutants (Figure 2.4 *color coded*). In contrast, the protein E74Q showed weak metal-binding as shown in Figure 2.4e.

Based on cobalt titration of MFsc-4G E74Q protein, we proposed that the carboxylate bridge in position 74 is critical for the coordination environment of the metal ion. We attempted to increase the distance between Fe(II) and carboxylate by substituting Glu74 with Asp. We anticipate this shorter analog (one -CH_2- short) to weaken its interaction with the metal ion, leaving three available sites on Fe(II). E74D mutation was introduced using MFsc-4G template, expressed in *E. coli*, and purified successfully using the same methods (Figure 2.5).

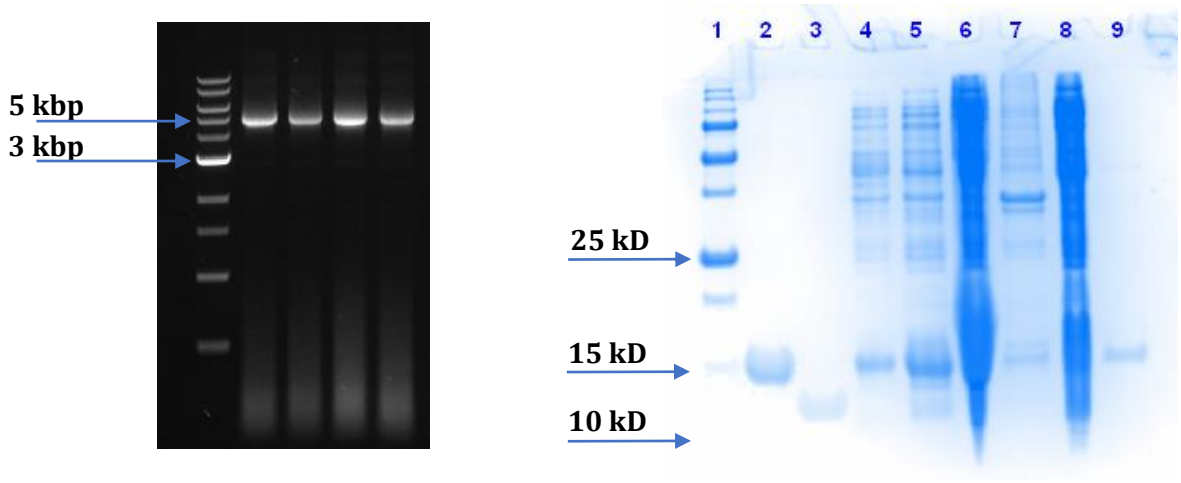


Figure 2.5. E74D mutagenesis, expression, and purification. (***left***) PCR samples of E74D mutagenesis on ethidium bromide-stained 1.0% agarose gel (expected plasmid size 5.5 kbp). (***right***) SDS-PAGE showing E74D mutant protein before (**Lane 2**) and after (**Lane 3**) His_6-tag cleavage. **Lane 4** - control sample with no IPTG; **Lane 5** – induced cells with IPTG; **Lane 6** - lysate supernatant; **Lane 7** - cell debris; **Lane 8** - Ni-NTA flow through (crude lystae); **Lane 9** – wash flow through.

For a tetrahedral conformation, we expected a shift of Co(II)-protein from λ_{max} = 525 ± 50 nm to λ_{max} = 650 ± 50 nm[22]. Spectrum of Co(II) with E74D showed a $\lambda_{max} \approx$ 567 nm and $\varepsilon \approx 199\ M^{-1}cm^{-1}$, indicating a six-coordinate octahedral system[22] (Figure 2.6a). Figure 2.6b suggests weak binding of Co(II) based on the Abs_{550} compared to other MFsc-4G mutants, although Abs_{550} slightly increases for Co(II) concentrations above 150 μM. Based on the scatter plot, metal-to-protein binding ratio cannot be determined accurately.

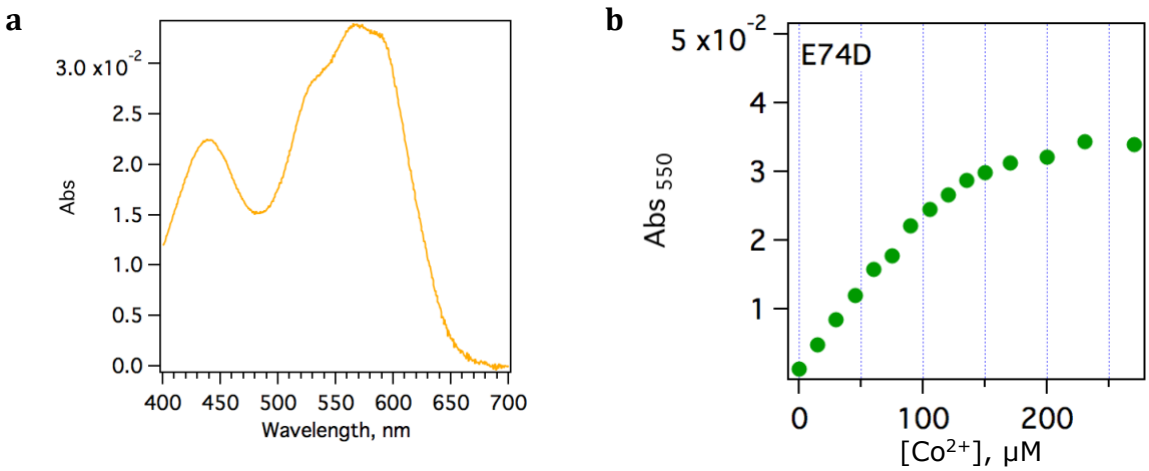


Figure 2.6. UV-vis spectrum of Co(II) bound to E74D (**a**) and titration plot (b) suggest weak binding of Co(II) to the E74D mutant.

We further tested the carboxylate bridge and substituted Glu74 with a His. Using PyMOL and Rosetta to model this mutation (see Experimental), the active site becomes symmetrical with E74H and E104H. Hypothetically, this confers two metal-binding sites, each complementing 2-His-1-carboxylate motif. E74H mutation was introduced in the MFsc-4G template and we named the resulting protein E74H. The same protocol for mutagenesis, expression, and purification were followed and we tested the pure protein for metal binding (Figure 2.7). Absorbance at 550 nm is almost similar for E74H compared to other MFsc-4G mutants, which suggests strong binding of Co(II). Abs_{550} slightly increases with increasing Co(II) concentrations, which indicate that E74H may accommodate more than one ion. However, this rejects our hypothesis that it can bind ions in a 2:1 ratio as in the case of DFsc.

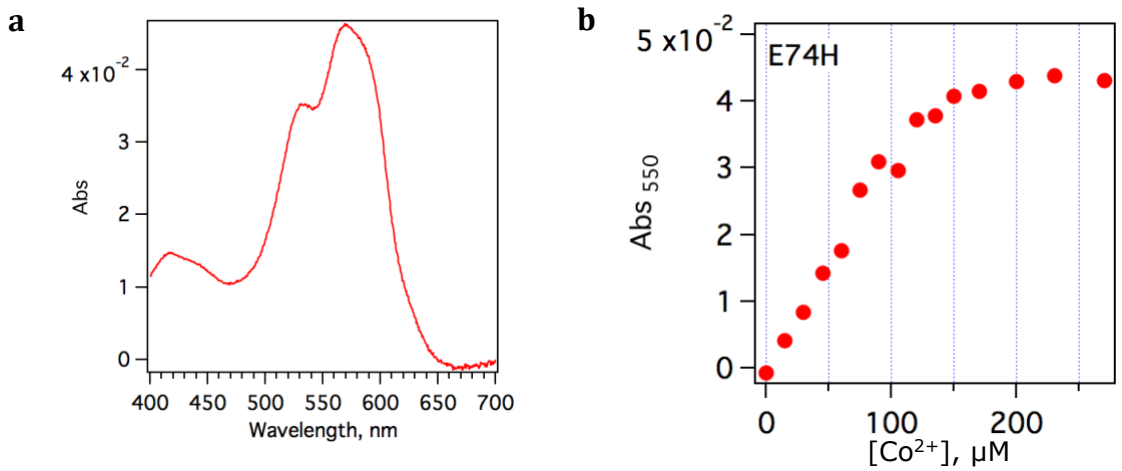


Figure 2.7. UV-vis spectrum of cobalt bound to E74H (**a**) shows absorbance similar to that obtained for MFsc-4G mutants, indicative of strong Co(II) binding. Abs_{550} does not plateau after 150 µM Co(II), suggesting that E74H can potentially bind more than one metal ion (**b**).

Using Beer-Lambert Law, the extinction coefficient of MFsc-4G-Co(II) at Abs_{max} was calculated at 233 M^{-1} cm^{-1}. Recent findings on Co(II) coordination in protein-Co(II) complexes explained that the extinction coefficient of the protein-bound Co(II) signifies a five-coordinate system ($250 > \varepsilon > 50$ $M^{-1}cm^{-1}$)[22]. This is also true for E11Q and H107Q mutants. However, we are uncertain on what ligands are coordinating with the metal ion. Additional experiments are required to determine the coordination environment of the bound metal ion.

2.2.4. Protein Folding and Stability

Introducing multiple mutations may affect amino acid side chain interactions, thus may destabilize or cause improper protein folding[23]. As shown in Figure 1.1, the backbone of MFsc is a bundle of four α-helixes. Circular dichroism (CD) is a spectroscopic method that provides signature signals to determine the secondary structure of the protein. This is

a useful tool as the secondary structure of proteins are sensitive to small perturbations. CD works by generating circularly (left-handed and right-handed) polarized light, where these photons are absorbed by the protein. Since amino acids are chiral (except Gly), each residue absorbs left-handed circularly polarized light (L-CPL) and right-handed circularly polarized light (R-CPL) at different intensities[24]. The difference between the two absorbance values within a range of wavelengths (e.g. 190-260 nm) can be measured and provide a defined spectrum that signifies the sample's secondary structures. We conducted CD analysis on MFsc and MFsc-4G with and without the presence of Zn(II) to determine any changes on their secondary structures. We also tested E74D to see if the mutation caused a change in its secondary structure since metal binding was found to be fragile. As projected, the signals confirmed that there are no changes to its secondary structure, upholding its helical characteristic even with E74D (Figure 2.8).

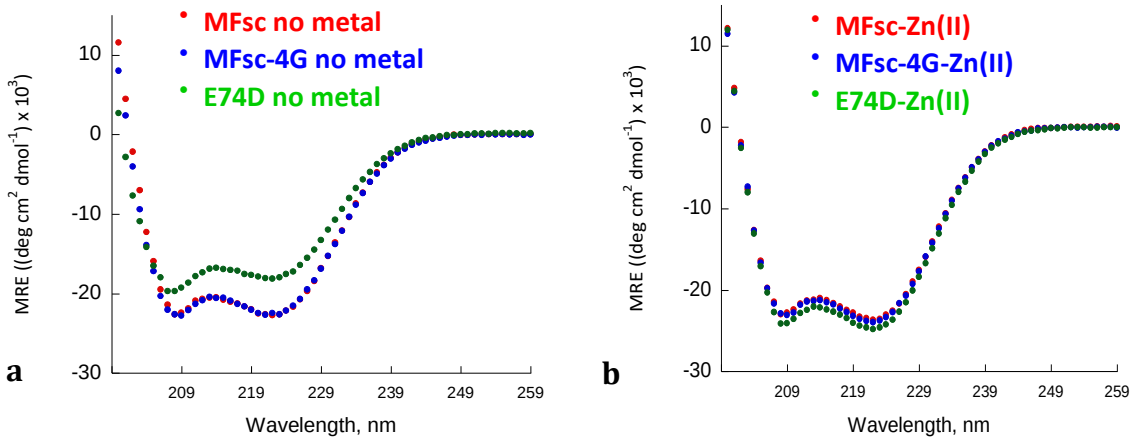


Figure 2.8. CD spectra of (**a**) MFsc (red), MFsc-4G (blue), and E74D mutant (green) with no Zn(II) and (**b**) with 200 μM Zn(II), in 5 mM HEPES pH 7.5 buffer show signature peaks for α-helix, which strongly suggests that proteins are well-folded.

During the purification process, the protein was placed in a 70°C-water bath for ten minutes with 2 mM EDTA to chelate metal ions that may be bound to the protein. No

evidence of precipitation was observed after cooling down the samples to room temperature. Data obtained from CD also imply that the proteins are well-folded.

2.2.5. Computational model of MFsc-4G and crystal structure

Since one significant mutation did not cause a major change in the secondary structure of the MFsc-4G, a computational model can predict the structure of the protein with one metal ion. Using Rosetta fixed backbone (`fixbb`) remodeling application, the structure of MFsc-4G can be predicted from the published NMR solution structure of DFsc (PDB 2HZ8). About 2000 poses were generated and among the assessed top 5%, several models converge into a low-energy structure showing a metal ion coordinated by four residues. The close distance between Glu74 and Fe(II) (1.43 Å) implies that this residue associates with the metal ion.

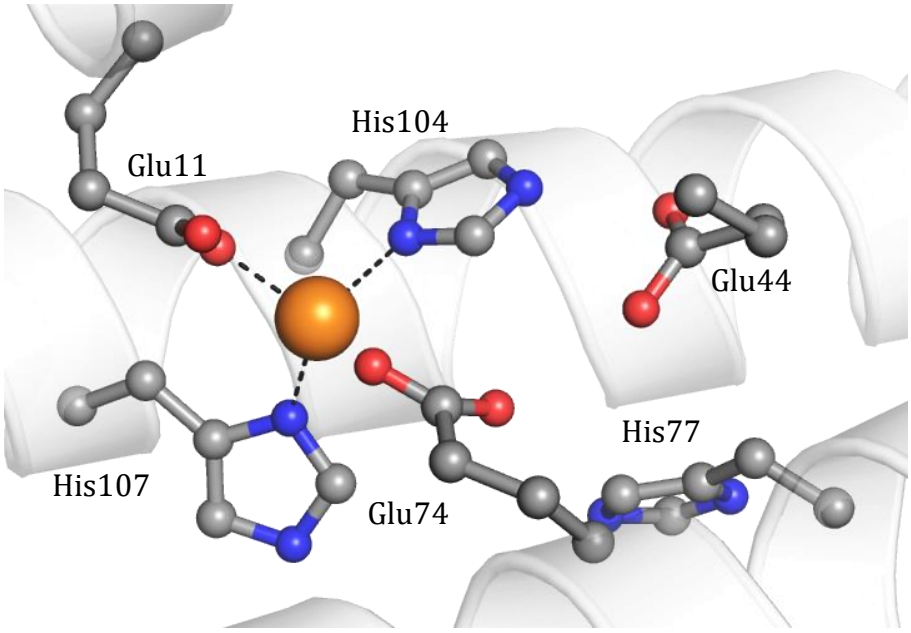


Figure 2.9. Predicted structure of MFsc-4G using Rosetta `fixbb` application. Glu74 may interact with the metal ion and loses this interaction upon substrate binding.

The structure model predicted above does not guarantee the position nor the coordination sphere of the metal ion. Currently Rosetta software suite is not yet optimized to recognize divalent and transition metals, particularly on their oxidation states. However, this does not limit Rosetta for working with metalloproteins. A –`in:auto_setup_metals` flag can be used in Rosetta application to detect metal ions and associated atoms that are covalently bonded. When the algorithms are run, the application creates constraints on the metal ion and the atoms that occupy its coordination sites. In Figure 2.9, the metal was removed when the mutation was introduced using `fixbb` application with the appropriate **resfile**, which contains the codes for replacing Glu104 with a histidine. The metal ion was added back and repacked with the auto setup metals flag.

While computational modeling can be useful to predict structure, our collaborators at Vrije Universiteit Brussel successfully crystallized MFsc in its apo and holo forms (Figure 2.10). With these crystals, the structure of MFsc-4G can be inferred from the crystal structure of MFsc. This also allows us to map the coordination sphere of the metal ion with atomic accuracy. Although, our collaborators reported that the X-ray images captured were of poor quality, particularly on the apo form which formed needle-shaped crystals. Recrystallization of the protein is currently in progress and will be followed by structure determination.

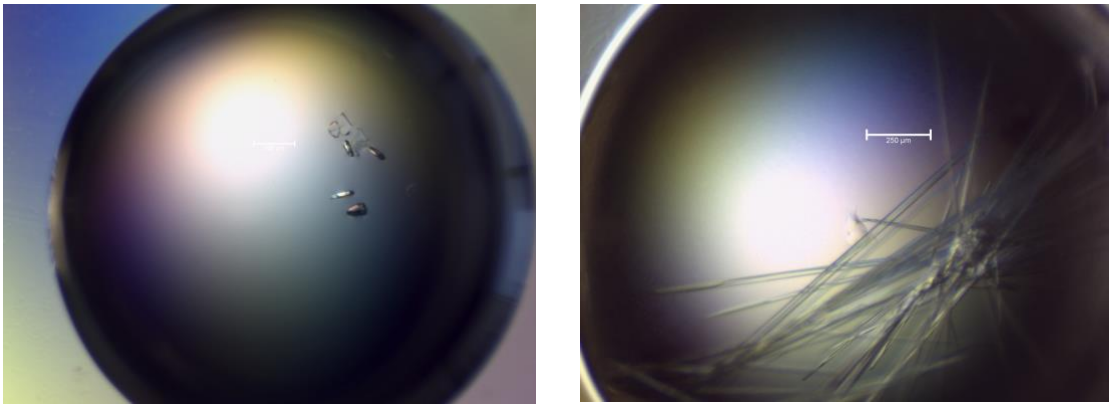


Figure 2.10. Crystals of MFsc with Zn(II) (***left***) and without metal ion (***right***) ready for X-ray crystallography studies.

2.3. Conclusions

In this chapter, we presented our work on designing a mononuclear nonheme Fe(II) protein from a *de novo* designed protein that binds two metal ions. Replacing one of the carboxylate bridges with a His resulted in a protein that binds single metal ion, which we proved using cobalt titration. The substrate channel was modified by mutating four Ala residues into Gly in order to facilitate access for substrates and expose the active site. These mutations did not affect the helical characteristic of the protein as evident on the CD spectra, which also showed that metal ion is not required for proper folding. The resulting protein is named MFsc-4G and mimics the 2-His-1-carboxylate motif active site. Several mutants were also expressed to investigate how the residues influence metal binding and protein folding. Given these results, we expect MFsc-4G to mimic reactions catalyzed by mononuclear nonheme iron(II) enzymes. Chapter 3 will focus on oxidation reactions that occur in the presence of MFsc-4G, metal ions, and dioxygen.

Chapter 3
Analyzing protein activity

Catechol dioxygenases (CDOs) use the same 2-His-1-carboxylate facial triad motif to add two atoms of oxygen on catechols, resulting in cleaving the aromatic molecule[25]. In this work we aim to mimic catechol oxidation reaction to probe its mechanism and to understand the relative factors that contribute to oxygen activation. However, the results described in this chapter show an interesting phenomenon that is rare in nature. MFsc-4G protein was designed to perform dioxygenase activity, particularly on the ring opening of catechols as observed with CDOs. The MFsc-4G-Fe(II) system and its mutants showed evidence of semiquinone radical, which is one of key intermediates in the extradiol cleavage of catechols. However, the radical intermediate was trapped in the protein for several days. Surprisingly, DFsc, the scaffold for MFsc-4G, also exhibited this radical

stabilization even in the absence of metal ions. In this chapter, we present our studies on catechol oxidation and characterization of the radical intermediate.

3.1. Results and Discussion

3.1.1. Oxidation of catechols

Extradiol cleavage mechanism

CDOs that perform extradiol cleavage of catechols utilize Fe(II) while intradiol-cleaving dioxygenases generally use Fe(III)[16]. Since we were focused on catechol oxidation with Fe(II), we followed the ring-opening mechanism of extradiol-cleaving CDOs (Figure 3.1).

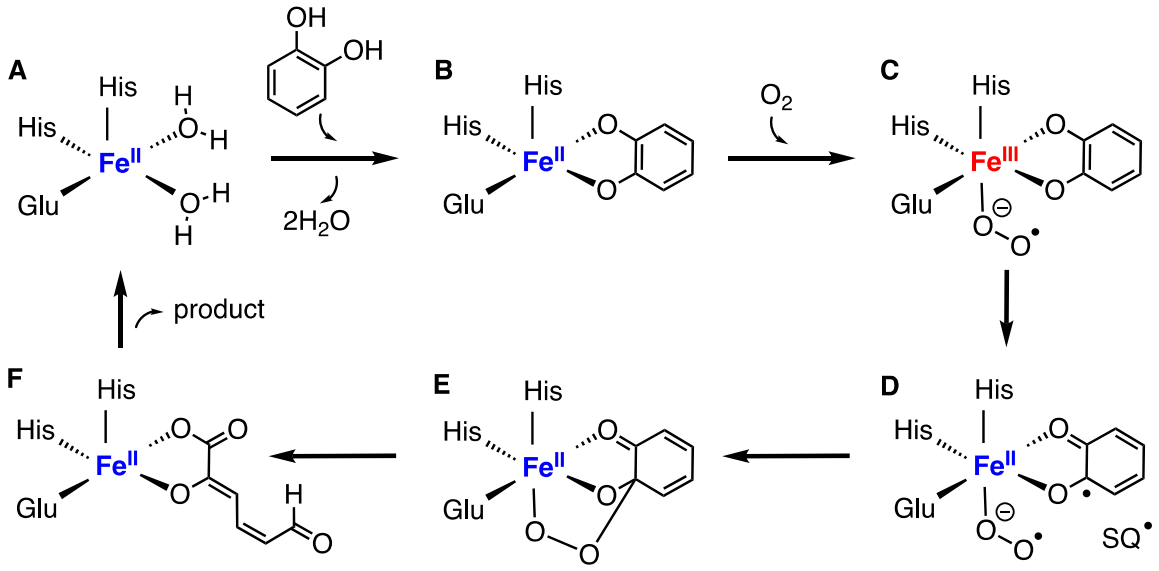


Figure 3.1. General reaction mechanism of extradiol-cleaving dioxygenases.

Once catechol associates to Fe(II) supported by 2-His-1-carboxylate motif (Figure 3.1B), O_2 attacks iron center and forms an Fe(III)-O_2•$^-$ adduct (Figure 3.1C). This species facilitates a one-electron transfer from the catechol to reduce the metal center to its Fe(II)

state (Figure 3.1D), and leaving a semiquinone radical (SQ•)[26]. Metal-bound O_2•$^-$ then reacts with activated SQ• to form a peroxide bridge intermediate (Figure 3.1E). Afterwards, a series of reactions occur including the Criegee rearrangement of substrate with peroxide splitting, which ultimately oxidizes the oxidized catechol to muconate semialdehyde (Figure 3.1F)[12,25]. The reaction is completed when the acyclic product dissociates from the metal center and the enzyme returns to its resting state (Figure 3.1A). Since MFsc and MFsc-4G hold Fe(II) by 2-His-1-carboxylate motif, we tested if the same event would take place in our designed enzyme.

4-*tert*-butylcatechol

We initially selected 4-*tert*-butylcatechol (TBC) to assess its interactions and activity with MFsc and MFsc-4G. The oxidation of TBC was first performed by H_2O_2 in the presence of Fe(II)-loaded MFsc or MFsc-4G protein and the progress of the reaction was monitored by UV-vis spectroscopy. Upon adding the substrate and peroxide, a peak at $\lambda \approx 580$ nm was initially observed for both proteins (Figure 3.2).

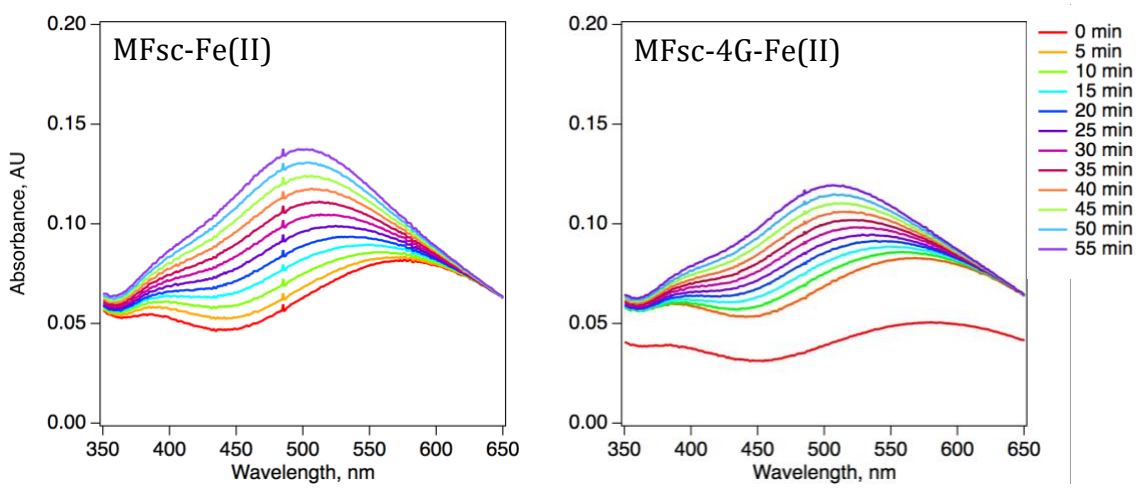


Figure 3.2. Oxidation of TBC by H_2O_2 in the presence of MFsc (***left***) and MFsc-4G (***right***) with Fe(II). Reaction conditions: 20 μM protein, 20 μM Fe(II), 1 mM TBC, and 2 mM H_2O_2 in 50 mM HEPES buffer at pH 8.0.

Having shown that reaction proceeds with formation of colored products we decided to repeat it on a longer timescale. Although MFsc showed relatively higher absorbance, we chose to work on MFsc-4G for the subsequent experiments since it has a more defined 1:1 metal-to-protein ratio, compared to MFsc with 0.8:1 (Figure 2.4). Oxidation of TBC by dioxygen in the presence of Fe(II)-loaded MFsc-4G was monitored for 15 hours (Figure 3.3). The samples were prepared in an anaerobic chamber to prevent oxidation of Fe(II) to Fe(III), and confirm that dioxygen is necessary for the reaction to occur. Clearly, the absorption spectrum indicates that O_2 drives the oxidation of the substrate (Figure 3.3 *red*).

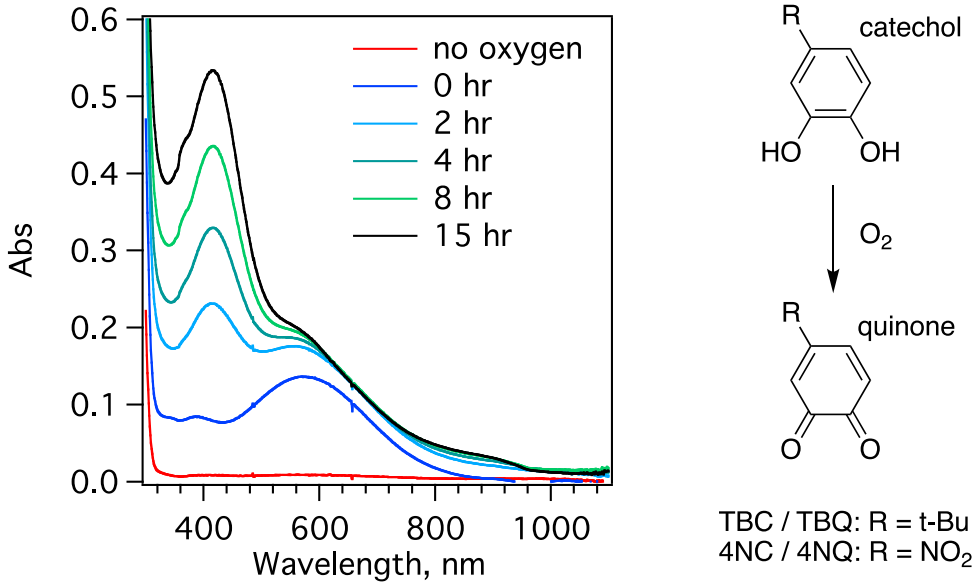


Figure 3.3. Oxidation of TBC by dioxygen in the presence of MFsc-4G-Fe(II). Protein loaded with Fe(II) was prepared in anaerobic glove box to preserve metal ion at Fe(II) state. An intermediate absorbing at 580 nm initially forms once dioxygen become available. The oxidized substrate absorbs at 400 nm and accumulates over time.

After oxygen was introduced we observed formation of species with λ_{max} at 580 nm. Then a large and pronounced peak appeared at 400 nm after 2-hour incubation. We can speculate that the species absorbing at 580 nm is a characteristic of the initial Fe(III)-catecholate intermediate (Figure 3.1C). Spectroscopic studies with 4-nitrocatechol (4NC) by Mbughuni *et al* characterized the intermediate absorbing at 610 nm as the vague catecholate-Fe(III)-$O_2^{\bullet -}$ species, and defined another intermediate at 405 nm as the oxidized product (4-nitro-*o*-benzoquinone, 4NQ) bound to the enzyme[12]. Moreover, free 4-nitro-*o*-benzoquinone product showed maximum absorbance at 380 nm[12]. Since TBC is functionally analogous to 4NC (Figure 3.3 *right*), we can deduce from these data that the peaks absorbing at 400 nm and 580 nm represent 4-*tert*-butyl-*o*-benzoquinone (TBQ) and the catecholate-Fe(III) species respectively.

We gained from this experiment that oxidation of TBC to benzoquinones by dioxygen is promoted by MFsc-4G-Fe(II). However, we still have no evidence of a catechol-cleaving reaction. Next, we tested if the reaction could be extended to other catechol substrates.

3,5-di-*tert*-butylcatechol

A mutant DFsc protein (2A3H-DFsc) remodeled by DeGrado group showcased its ability to confine SQ• in its active site, which is generated from 3,5-di-*tert*-butylcatechol (DTBC) in the presence of two Zn(II) ions[10]. In this work we tested if SQ• radical could be stabilized by MFsc-Fe(II) system as well. In comparison, MFsc-4G and 2A3H-DFsc are structurally homologous based on the sequence alignment of the two proteins since both are derived from DFsc (Figure 3.4).

```
2A3H-DFsc     MDELRELLKAEQQGIKILKEVLKKAKEGDEQELARLNQEIVKAEKQGVKVYKEAAEKARN   60
MFsc-4G       MDELRELLKGEQQGIKIYKEVLKKAKEGDEQELARLIQEIVKGEKQGVKVYKEAAEKARN   60
              *********|*******.******************.*****|*****************

2A3H-DFsc     PEKRQVIDKILEDEEKHIEWHKAASKQGNAEQFASLVQQHLQDEQRHVEEIEKKN        115
MFsc-4G       PEKRQVIDKILEDEEKHIEWLKAASKQGNAEQFASLVQQILQDHQRHVEEIEKKN        115
              ********************.******************.***.***********
```

Figure 3.4. Sequence alignment of MFsc-4G used in this work and 2A3H-DFsc by Ulas *et al*[10]. MFsc-4G and 2A3H-DFsc differ by seven mutations (highlighted). Residues marked with a | represent mutations on the substrate channel.

Oxidation of DTBC by H_2O_2 in the presence of Fe(II)-loaded MFsc was followed for 1 hour (Figure 3.5). The addition of peroxide as a strong oxidant allows the reaction to proceed faster compared to ambient O_2. Peroxide was added last due to its potential to oxidize Fe(II), which can possibly inactivate the protein[27].

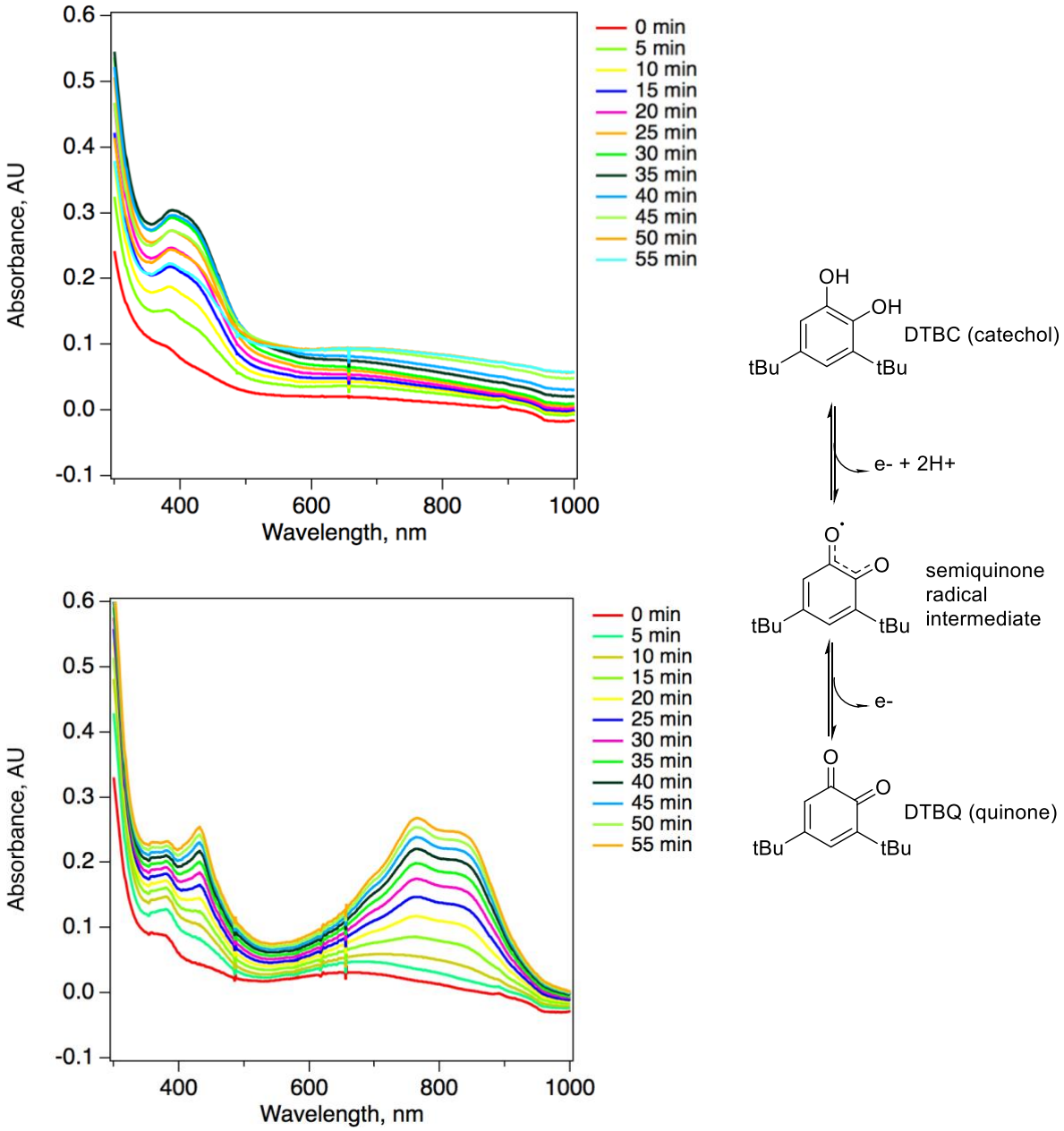


Figure 3.5. Oxidation of DTBC by H_2O_2 in the presence of MFsc-Fe(II). (*top*) DTBC incubated with Fe(II) did not show any significant absorbance changes from 740-850 nm where SQ• absorbs. (*bottom*) Appearance of increasing broad peaks between 740 and 850 nm strongly indicate presence of SQ• supported by MFsc-Fe(II) system. Experimental conditions: 20 μM MFsc (*top*) or MFsc-Fe(II) (*bottom*), 2 mM H_2O_2, 1 mM DTBC, 50 mM TRIS pH 8.0, incubated at room temperature for 1 hour. (*right*) Equilibrium between different oxidation states of DTBC.

Ulas *et al* previously characterized that SQ• absorbs from 740 to 850 nm region and the quinone form (DTBQ) has a λ_{max} of 415 nm[10], which is similar to the observations we recorded with MFsc-Fe(II) (Figure 3.4 *bottom*). We repeated the experiment for a much longer time scale in the presence of MFsc-4G-Fe(II) and dioxygen (Figure 3.6). To ensure that metal ion remains in Fe(II) state, sample was prepared in the anaerobic glove box and obtained UV-vis spectra before adding O_2-saturated buffer.

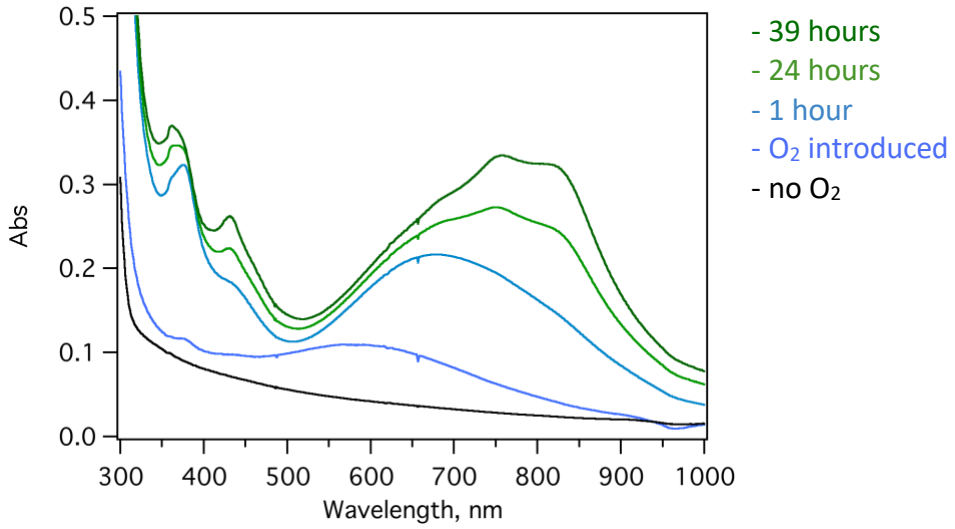


Figure 3.6. Oxidation of DTBC by O_2 in the presence of MFsc-4G-Fe(II). No evidence of oxidation was observed in anaerobic conditions (black). The peak around 600 nm which presumably corresponds to catecholate-Fe(III) appeared once O_2 has been introduced (blue), which began to shift towards the SQ• absorption range after 1 hour (light blue). Absorption spectrum of SQ• becomes apparent after 24 hours (green) and developed with more pronounced peaks in 39 hours (dark green).

Interestingly, the green color of SQ• remained for at least 39 hours (Figure 3.6 *dark green*), which is quite unusual stability for a radical species. SQ•. Given that the SQ• intermediate is trapped by MFsc-4G and 2A3H-DFsc in the presence of metal ions, we invested on studying how the metal ion concentration influence the formation and stabilization of semiquinone radical. For this experiment we needed a protein that could

accommodate variable number of metal ions. Our cobalt titration experiment showed that MFsc-4G can only bind one metal ion. On the other hand DFsc protein can bind one and two metal ions. Therefore, we decided to perform metal dependence tests using DFsc. To assess SQ• yield as a function of metal ion concentration, we UV-vis spectra of DTBC oxidized in the presence of apo DFsc, and with 1 and 2 equivalents of metal bound (Figure 3.7). We assumed that when 1 equivalent of Fe(II) is added to DFsc, the protein exists in mono-iron state. However, it is possible that a fraction of DFsc could be in diiron form and some protein could remain in the apo form.

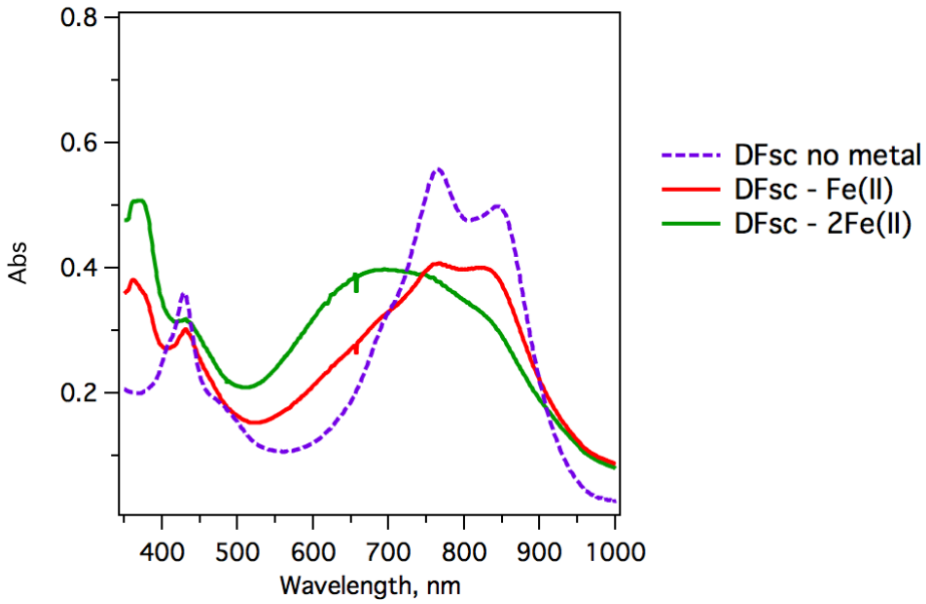


Figure 3.7. SQ• radical stabilization by DFsc protein loaded with 1 (red) and 2 equivalents (green) of Fe(II) and in apo form (purple). SQ• yield was quantified at 740 nm. Samples were prepared in anaerobic glove box then added O_2-saturated buffer. Spectra were taken after 24 hours at room temperature.

Using the reported λ_{max} (≈740 nm) of SQ•, we can quantify the yield of radical at 740 nm and use extinction coefficient (ε) of 7556 M^{-1} cm^{-1}, although these values are based 2A3H-[DFsc-$Zn(II)_2$- SQ•] moiety[10]. In Figure 3.7, the maximum absorbance of DFsc with 1

and 2 equivalents Fe(II) are almost the same, however, their spectra differ in λ_{max} (Figure 3.7). Surprisingly, DFsc showed more pronounced peaks indicative of SQ• and having the highest yield when no metal was present. To accurately measure the amount of radical, we repeated the experiment with DFsc-Zn(II)$_2$ (Figure 3.8).

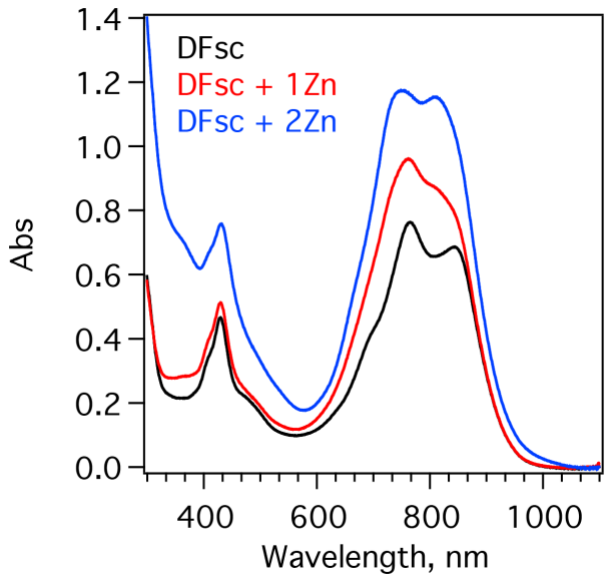


Figure 3.8. SQ• radical stabilization by DFsc protein loaded with 1 and 2 equivalents of Zn(II) (red and blue, respectively) and in apo form (black). Spectra were taken after 14 hours at room temperature. Experimental conditions: 50 μM DFsc, variable Zn(II) (0, 50, 100 μM), 0.5 mM DTBC, 50 mM HEPES pH 8.0

Interestingly, apo DFsc (50 μM) produced large amounts (60-100 μM) of trapped SQ•. To ensure that no metal is bound to the pure apo protein, DFsc analyzed for Fe, Zn, and Mn ions using ICP.

DFsc protein itself is capable of stabilizing SQ•, therefore, we hypothesized that certain residues on the hydrophilic active site contribute to this phenomenon. Using MFsc-4G and its mutants (E11Q, E74Q, H107Q, E74D, and E74H), which were initially analyzed

for metal binding (see Chapter 2.2.3), we can study how these polar residues interact with the dihydroxyl groups of DTBC.

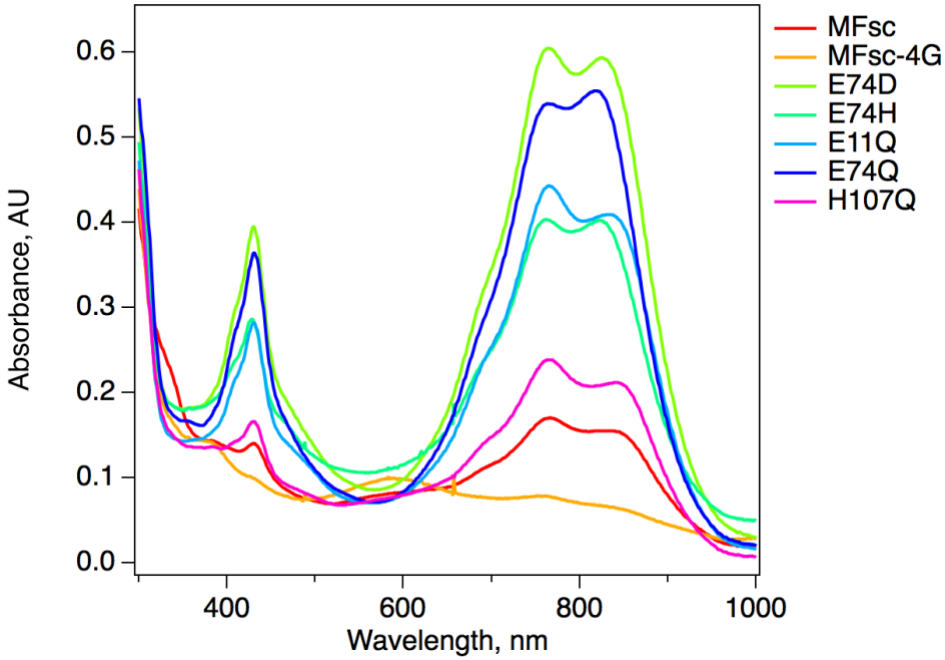


Figure 3.9. UV-vis spectra of SQ• stabilized by apo proteins after a 24-hour incubation. MFsc and H107Q yield low amounts of SQ•, and almost no SQ• was not detected with apo MFsc-4G. Experimental conditions: 50 μM protein, 0.5 mM DTBC, 50 mM HEPES pH 8.0.

From the absorption spectra in Figure 3.9, only MFsc-4G did not produce stabilized radical. MFsc-4G has a His residue at position 104 indicating that Glu104 participates in SQ• formation. Moreover, low yield of radical can be observed with MFsc, which differs by 4 Ala residues that sculpt the substrate channel compared to MFsc-4G. Hence, we hypothesized that substrate dihydroxyl groups face the hydrophilic center and may be interacting with Glu104. Further, Ala residues along the binding cleft form hydrophobic interactions with the bulky *tert*-butyl groups of the DTBC, which confines the molecule in its bound state. These protein-ligand interactions can be studied experimentally using NMR (Chapter 3.1.2) and *in silico* (Chapter 3.1.3).

3.1.2. NMR studies

In Chapter 2.2.5, our collaborators (Virje Universiteit Brussel) presented that MFsc is suitable for X-ray crystallography studies to determine the structure of the protein. Given that the apo proteins used in this work can bind and stabilize SQ•, the proteins can be crystallized with the trapped radical. However, finding the conditions to crystallize these species remain as a challenging task. Fortunately, modern NMR spectroscopy can be used to derive the three-dimensional structure of proteins in solution, which also allows us to study the protein-ligand interactions we have speculated in Chapter 3.1.1. For this technique, we expressed ^{15}N-labeled MFsc and DFsc proteins (see Experimental) to carry out Heteronuclear Single Quantum Correlation (HSQC) studies. Changes in chemical shifts ($\Delta\delta$) of the two-dimensional HSQC spectrum of the protein (showing ^{1}H-^{15}N coupling from protein backbone and side chains) reveals which residues interact with the substrate when added. Since apo DFsc provided the most amount of SQ•, NMR studies were conducted with DFsc.

To understand how SQ• is stabilized in protein environment, we studied the spectrum of apo DFsc in the presence of DTBC. We initially obtained HSQC spectrum of DFsc in HEPES buffer at pH 7.5 since all DTBC experiments presented in the last section (Chapter 3.1.1) were done in HEPES buffer at pH 8.0. However, the spectrum at pH 7.5 showed poor resolution that is not suited for our purpose (Figure 3.10 *red*). We considered taking the spectrum of DFsc at pH 6.0 (MES buffer), although we speculated that the conditions would reduce DFsc function. Rather, protein sample in pH 6.0 showed well-defined signals that are most suitable for NMR studies (Figure 3.10 *blue*).

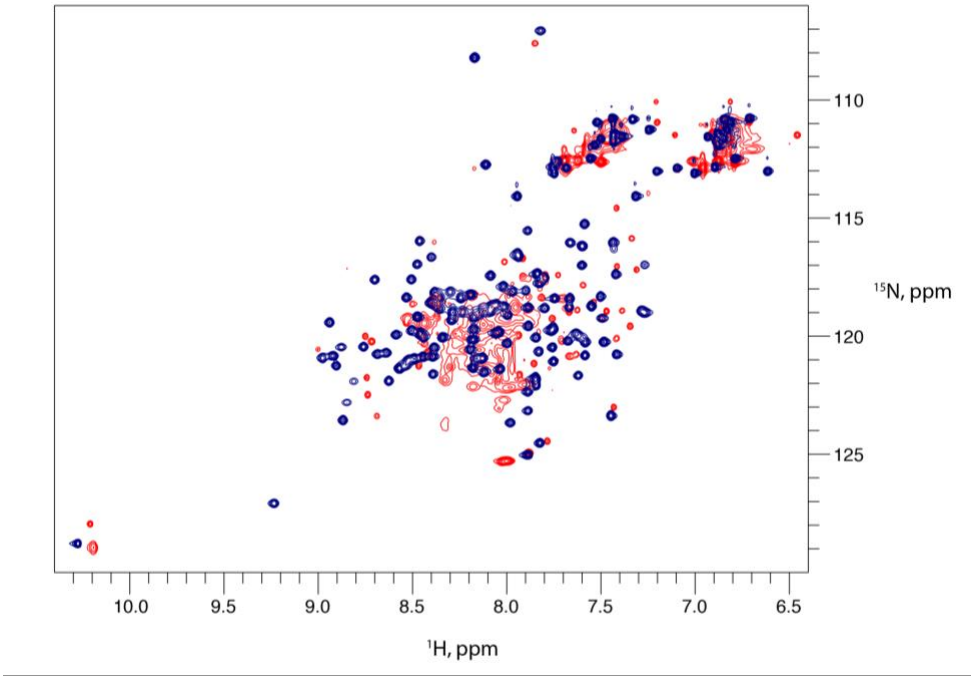


Figure 3.10. HSQC spectra of DFsc protein in 25 mM MES, 100 mM NaCl, pH 6 (blue, 800 µM protein) and 25 mM HEPES, 100 mM NaCl, pH 7.5 (red, 320 µM protein). Samples contained 10% D_2O.

Having a workable experimental condition, we analyzed how much SQ• can be generated at pH 6.0. DTBC was incubated with DFsc for 24 hours (Figure 3.11).

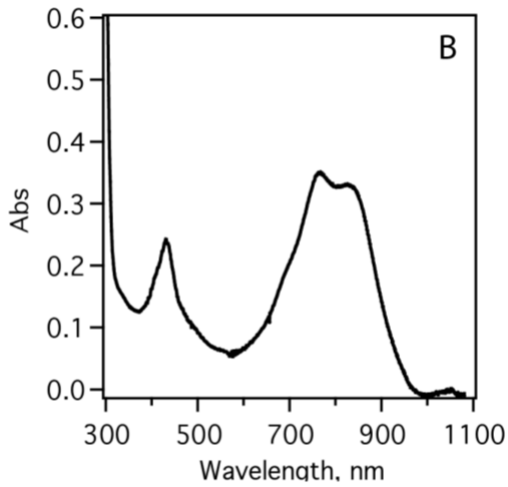


Figure 3.11. UV-vis spectrum of the sample containing DFsc (500 µM) and DTBC (1 mM) in 25 mM MES buffer at pH 6.0.

Based on absorbance the concentration of SQ• is ~46 μM ($\varepsilon_{SQ\bullet}$ = 7556 M^{-1} cm^{-1}), which is only 10 % of total concentration of protein. First, we acquired NMR HSQC spectrum of apo DFsc without DTBC (Figure 3.12 *black*). Next, added DTBC to the same protein sample (Figure 3.12 *maroon*) and acquired another HSQC spectrum. Albeit SQ• yield was low, overlay of HSQC spectra before and after the addition of substrate demonstrated changes in chemical shifts of several signals, which strongly indicates substrate binding.

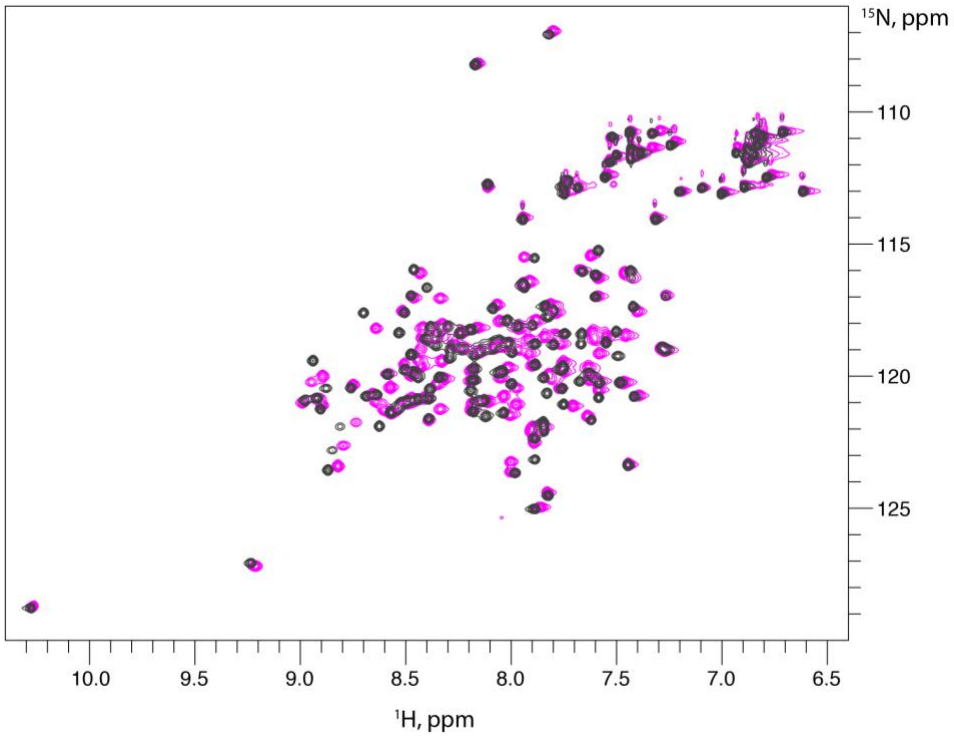


Figure 3.12. DTBC binds to DFsc at pH 6. HSQC spectra of DFsc protein (black, 800 μM protein) and DFsc with 2 equivalents of DTBC at 24 hours (maroon, 500 μM protein). Samples are in 25 mM MES, 100 mM NaCl, pH 6 containing 10% D_2O.

To further this study, our collaborators at Vrije Universiteit Brussel completed a Triple Resonance Backbone Assignment using double-labeled (^{15}N, ^{13}C) DFsc to identify

which residues correspond to the signals in the HSQC spectra. Having these signals assigned to each residue of DFsc, the protein was titrated with DTBC, obtaining HSQC spectrum for each time the substrate was added. Thus, the Δδ for each residue was mapped on the NMR structure of protein (PDB: 2HZ8) to generate a heatmap, where red color signifies the most significant Δδ (Figure 3.13).

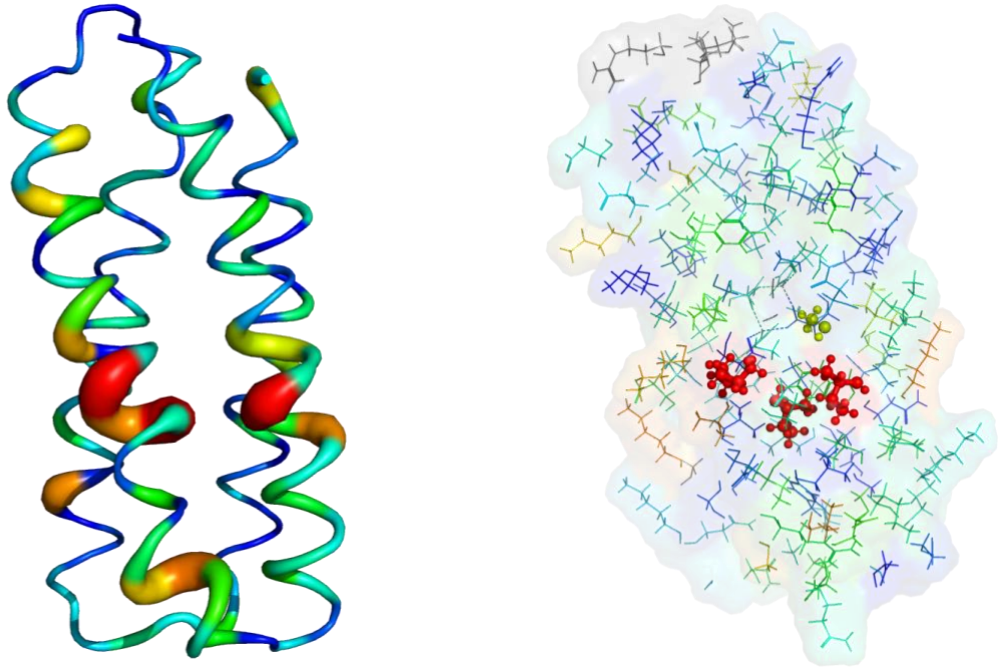


Figure 3.13. B-factor putty cartoon heatmap with showing how much backbone signals shift in the presence of DTBC (*left*). DFsc protein (0.4 mM) was titrated with DTBC up to 2 equiv. Changes in chemical shift (Δδ) were mapped onto DFsc structure (2HZ8): red corresponds to maximum shift and blue minimum shift. Residues I17, I40, and K42 (*right*, ball-and-stick) strongly interact with substrate.

Based on the generated heatmap, the highest changes in chemical shifts were assigned to Ile17, Ile40, and Lys42 (Figure 3.13 *right*) with Δδ > 0.15 ppm when substrate is present. This strongly suggests that these residues interact with the substrate. Relatively strong NOEs were observed between Hα's of residues Ala14, Ile40, and 3-*tert*-butyl

substituent of DTBC. We presumed that these protons are within 6.0 Å from each other, which we used to predict DFsc-DTBC model *in silico*.

3.1.3. Protein-ligand interactions (Rosetta ligand docking)

We utilized a set of applications from the Rosetta Software Suite to generate a model of SQ• binding to apo DFsc. The protocol used for docking the substrate was adopted from Combs *et al* [28]. Using PDB file 2HZ8, we averaged the xyz coordinates of the metal ions before removing them and the solvent molecules. The averaged xyz point was chosen as the center of a user-defined imaginary box with a radius of 8.0 Å. It has a volume spanning over the binding cleft and active site, where Rosetta algorithms positionally translate (i.e. moves and transforms the molecule) random conformers and positions of DTBC within this user-defined box several times until its position has minimum overlaps with any of the atoms in the protein. For this experiment, DTBC conformers were not generated since there are only two insignificant rotatable bonds. The pose is refined by sampling side chain rotamers to minimize the energy of the pose and maximize substrate fitting on the active site. Experimental NOE data suggests that Ala14 and Ile40 Hα's are within 6.0 Å to the *tert*-butyl protons. Since these protons are equivalent and rotate freely, we decided to apply atom distance constraints between the tertiary carbon of 3-*tert*-butyl substituent and the Hα's of Ala14 and Ile40. We initially set the distance constraints for both residues to 7.0 and 8.0 Å to reflect the linear distance between methyl protons and tertiary carbon of *tert*-butyl (2.2 Å). Conversely, results from these parameters do not mirror our NMR data. Thus, we decided to use 6.0 Å constraints with a standard deviation (σ) of 0.5. We assumed that the dihydroxyl groups are buried into the anionic interior of the protein, particularly

towards Glu104 which we identified as an important residue in generating SQ• (see Chapter 3.1.1 and Figure 3.9). However, constraints to Glu104 were not applied to prevent biased results. Scoring function for distance constraints (HARMONIC) was not utilized to avoid significant perturbations on the total score.

We generated 2000 poses and ranked lowest scoring models based on `total_score`. Overlay of these models converged into a pose shown in Figure 3.14. The substrate is positioned reasonably well on the binding pocket, where the hydroxyl groups face the negatively-charged core. Substrate alkyl substituents fit suitably on the surface along the substrate channel lined with hydrophobic residues. Distance between atoms with constraints agree well within the prescribed parameters (distance constraint = 6.0 Å and σ = 0.5). Amine group of Lys42 faces the exterior of the protein and no polar contacts were detected between catechol and residue, indicating that only the alkyl segment contributes to the experimentally observed NMR $\Delta\delta$ with DTBC. Specifically, tertiary carbon of 3-*tert*-butyl substituent is 5.9 Å and 6.5 Å away from Ala14 and Ile40 respectively. Hα's of Ala14 and Ile40 are 3.9, Å and 4.3 Å, respectively, to the closest proton of 3-*tert*-butyl substituent, which complements the NOEs observed and NMR $\Delta\delta$.

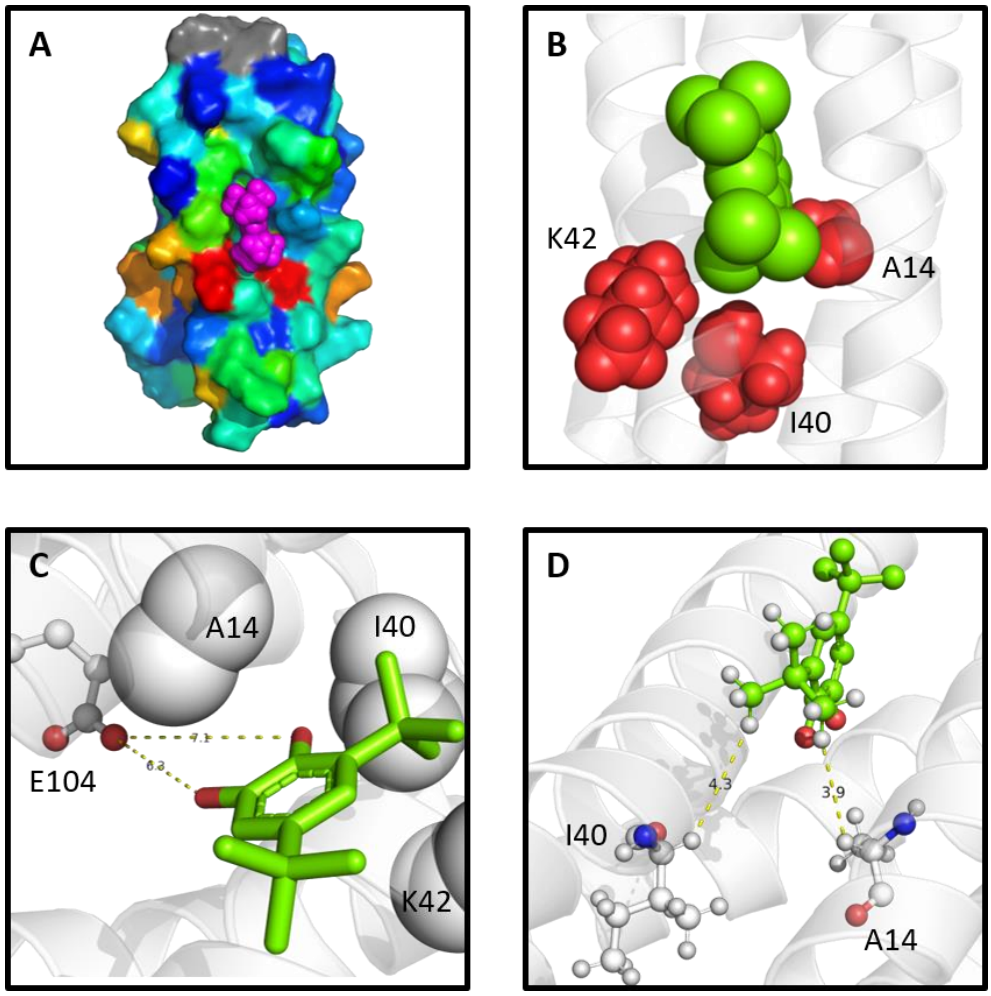


Figure 3.14. Ligand docking of DTBC (green) into apo-DFsc using Rosetta. (**A**) Surface representation of DFsc shows that DTBC (magenta) is positioned inside binding cleft. Red areas indicate residues with strong interactions. (**B**) Ala14, Ile40, and Lys42 (red) in contact with 3-*tert*-butyl substituent via hydrophobic interactions. (**C**) Dihydroxyl groups face anionic interior of DFsc, positioned 6.3-7.1 Å to carboxylate group of Glu104. (**D**) 3-*tert*-butyl substituents of DTBC are ~6.0 Å from Ala14 and Ile40 (5.9 Å and 6.5 Å) with the protons 3.9 Å and 4.3 Å respectively.

Using the predicted model, we can deduce that Ala14 and Ile40 participate in binding DTBC in the substrate channel via hydrophobic intermolecular forces. Additionally, we speculate that DFsc alters the reduction potential of catechol when bound to the substrate. Based on UV-vis experiments with MFsc-4G mutants, distance between

dihydroxyl functional groups and side chain of Glu104 (6.3 – 7.1 Å) may offer insights on how protein promotes oxidation of DTBC into SQ•.

3.2. Conclusion

In this chapter, we examined the reactivity of designed protein with 2-His-1-carboxylate motif on the ring-cleaving oxidation of catechols. We showed that oxidation of 4-*tert*-butylcatechol into quinones is catalyzed by MFsc-4G-Fe(II). We also investigated the oxidation of 3,5-di-*tert*-butylcatechol with MFsc-4G system and showed spectroscopically that protein can stabilize semiquinone radical intermediate in the presence of O_2 and Fe(II). Additionally, we demonstrated that DFsc can produce and stabilize high yields of this highly reactive species even without metal ion. Further studies through a series of site-directed mutagenesis suggest that Glu104 participates in radical formation, while NMR studies revealed specific residues that interact with the substrate. Experimental data were integrated with *in silico* studies to predict a rational model for protein- SQ• moiety, which logically agreed with NMR results. We hypothesized that DFsc binds the catechol substrate and alters its reduction potential, making it more susceptible to oxidation. Having a protein crystallized with and without metal, the proteins used in this work would be suitable for X-ray crystallography studies to understand how metalloprotein and its apo form traps the radical intermediate. Moreover, this would provide insights on developing an ideal mononuclear nonheme iron(II) protein that can catalyze oxidation reactions. The results herein stemmed into two ongoing studies: to study the coordination environment of the active site to develop MFsc-4G into an enzyme and determine how the protein stabilizes a highly reactive species, which we focused on the latter.

Chapter 4
Experimental

The gene encoding the DFsc was cloned into a pMCSG49 expression vector (with Amp^R) using ligase independent cloning (LIC). The DFsc gene is fused to a His_6-tag domain on the N-terminus. All mutations were performed using site-directed mutagenesis and expressed in *E. coli*. Proteins were purified using affinity chromatography followed by TEV digestion. All proteins were stored at -80°C.

4.1. Cloning and Mutagenesis

cDNA sequence that codes for DFsc protein was synthesized by BioBasic. The gene was then cloned into pMCSG49 expression vector (DNASU) using ligase independent cloning (LIC).1 The resulting sequence has a N-terminal His6-tag and tobacco etch virus (TEV) protease recognition site (ENLYFQ/S). Site-directed mutagenesis was performed

using IDT primers and Phusion DNA polymerase (Thermo) following the manufacturer's standard protocol.

4.2. Protein expression and purification

The plasmid containing the gene of interest was transformed into E. coli BL21(DE3) and the cells were grown on agar/LB plates with Amp (100 μg/mL) overnight at 37°C. Next day a single colony was inoculated into 1xLB (Luria Bertani, 100 mL) containing Amp (100 μg/mL). The culture was incubated at 37°C overnight with shaking at 180 rpm. The following day 10 mL of the overnight culture were diluted into 1 L of 1xLB media supplemented with Amp and this culture was incubated at 37°C until OD600 reached 0.6 – 0.7. The culture was induced with IPTG to a final concentration of 0.5 mM and continued to grow at 30°C. After four hours of expression, cells were harvested by centrifugation at 4°C (20 min, 4,000 g). Cell pellet was suspended in a lysis buffer (5 mL for 1 g of cell paste) containing 25 mM TRIS, 20 mM imidazole, 100 mM NaCl (pH 8) with protease inhibitor added (phenylmethylsulfonyl fluoride, 0.5 mM final concentration). Cells were lysed by sonication on ice for 10 min (20 s pulse, 20 s rest). The crude cell lysate was then centrifuged at 20,000 g for 30 min, and the supernatant was passed slowly through Ni-NTA resin (Clontech, 2 mL). The resin was then washed with 50 mL of the lysis buffer. The protein was eluted with a buffer containing 25 mM TRIS, 250 mM imidazole, 100 mM NaCl (pH 8). Fractions containing the protein were identified by UV-Vis spectroscopy, combined and concentrated to 2 mL and then exchanged into a TEV cleavage buffer (50 mM TRIS, 75 mM NaCl, pH 8) using a desalting column (BioRad, Econo-Pac 10 DG). To help proper folding, proteins were incubated in 70°C water bath for 10 min (Figure S13). The same

treatment of wt DFsc protein resulted in precipitation and therefore was avoided. The affinity tag was cleaved with TEV protease in the presence of 1 mM dithiothreitol (DTT) and 2 mM EDTA. TEV protease was added to a final (OD280 protein)/(OD280 TEV protease) ratio of 20/1. The solution was sterilized using a 0.22 μm PES filter (UltraCruz) and incubated overnight at 34°C. After the digest, DTT and EDTA were removed on a desalting column and the obtained solution in the storage buffer (25 mM HEPES, 100 mM NaCl, pH 7.6) was loaded onto Ni-NTA column. The pure digested protein was collected in the flow through fractions. Finally, the protein was concentrated to ~1 mM, aliquoted and flash frozen in liquid nitrogen. The purity (> 95%) of the final samples was checked by SDS-PAGE analysis. Protein concentration was determined by measuring the absorbance at 280 nm using the calculated (ExPASy) extinction coefficient of 8480 M-1 cm-1 for cut protein and 9970 M-1 cm-1 for uncut protein.

4.3. Expression of labeled proteins

The plasmid containing DFsc gene was transformed into *E. coli* BL21 (DE3) and the cells were grown overnight at 37°C. The next day, a single colony was transferred into 2 mL of LB medium supplemented with Amp and grown at 37°C for 4 hours. This culture was diluted with 18 mL of M9 minimal medium solution containing $^{15}NH_4Cl$ (1 g/L) as nitrogen source, D-glucose (6 g/L) or D-glucose-^{13}C (2 g/L) as carbon source and Amp, then incubated at 37°C for 2 hours. The culture was further diluted with 1 L of M9 minimal medium solution and incubated at 37°C until OD_{600} reached 0.6-0.7. Protein expression was induced by addition of IPTG to a final concentration of 0.5 mM and the culture was

incubated at 18°C overnight. Cells were harvested by centrifugation (4000 g for 20 mins). Typical yield of wet cell paste was 2 g/L.

4.4. Cobalt Titration

Determining the protein-to-metal ratio was performed by titrating protein with Co(II). Fifteen tubes were prepared with 150 μM protein, each containing $CoCl_2$ from 0 μM to 270 μM (up to 300 μM with DFsc). These samples were incubated overnight and centrifuged to pellet any precipitates. Spectra of the solutions were obtained from 400-700 nm using Agilent Cary 60 UV-vis spectrophotometer. Each spectrum was normalized at 700 nm and recorded the absorbance at λ_{max}. Optical data were plotted using KaleidaGraph and Igor Pro 8.

4.5. UV-vis spectroscopy for kinetics experiments

Kinetics measurements were done on Agilent 8453 UV-vis spectrophotometer, using a quartz cuvette (1 mL) with 1 cm path length. Protein was degassed using Schlenk line and brought inside anaerobic glove box. For samples containing Fe(II), solution of $(NH_4)_2Fe(SO_4)_2$ in water was added to the protein to deliver 1 or 2 equivalents of metal ion with respect to protein. The samples were incubated at room temperature for 15 minutes inside glove box. Solution of substrate – 3,5-di-*tert*-butyl catechol (DTBC) or 4-*tert*-butyl catechol (TBC) – was freshly prepared in methanol at 20 mM concentration and then added to the protein in water (500 μL total volume). The mixture was then transferred into anaerobic cuvette (Starna cells) and the UV-vis spectrum was acquired outside the glovebox. Oxygen-saturated buffer (100 mM HEPES, pH 8.0) was prepared by bubbling

pure dioxygen gas for 15 minutes through buffer solution kept on ice. Oxygenated buffer (500 μL) was added to the mixture of protein and UV-vis spectrum was taken immediately and then with 1-hour intervals for 16 – 24 hours. The final concentrations of reagents are as follows: 50 μM protein with or without metal ion, 500 μM substrate in 50 mM HEPES, pH 8 and 0.65 mM O_2.

4.6. Inductively Coupled Plasma (ICP) analysis

Protein stock solutions were diluted to a final concentration of 25 μM in 3 mL using storage buffer (25 mM HEPES, 100 mM NaCl, pH 7.6). The concentration of Fe, Zn, and Mn were measured using ICP-MS (Perkin Elmer Optima 3300 DV).

4.7. Product extraction for NMR

The reaction sample was adjusted to about pH 3.0 using concentrated HCl and measured with a pH paper. Dichloromethane was used to form the organic layer in a separatory funnel and mixed vigorously by inverting the sealed funnel and releasing the gas each time the funnel is inverted. The organic layer was collected in a clean vial and the extraction was repeated twice, combining the organic layer into the vial. The organic layer was transferred into a clean separatory funnel and added saturated sodium bicarbonate of equal volume, then mixing vigorously as before. The final organic layer was transferred into a vial and removed the solvent using rotovap. The solid sample was dissolved in 300 μL of D-chloroform and transferred into an NMR tube. The sample was analyzed using the Bruker 400 MHz NMR.

4.8. Computational modeling

For introducing mutations, a **resfile** was generated with the following syntax: `<residue#> <chain> **PIKAA** <one-letter amino acid>`. PDB ID 2HZ8 (PDB code for DFsc) was used, which was cleaned using `clean_pdb.py` to remove solvent and metal ions. Mutation was introduced by running `fixbb` application with `-in:file:extra_res_fa <resfa file>` flag. Only one pose was generated using the flag `-nstruct 1`. Metal ion was added back by concatenating the coordinates from the original PDB file into the new file. The protein was repacked using `fixbb` application with `-repack_only` and `-in:auto_setup_metals` flags.

4.9. XML-based Rosetta ligand docking

4.9.1. Preparing the protein and ligand files

The ligand docking protocol was based on the approach by Lemmon et al[29] using XML scripting to control the parameters. DFsc protein (PDB ID: 2HZ8) was prepared by stripping the water molecules and two zinc ions using the `clean_pdb.py` tool in Rosetta. The side chain positions of the resulting apo protein was optimized without perturbing the backbone using the `fixbb` protocol with `-packing:repack_only -ex1 -ex2` flags. 3D structure of DBC was obtained from PubChem (CID: 66099) in .sdf format. Since DBC has only two rotatable bonds on its *tert*butyl groups, generating conformers presents no significant change. Ligand .params and .pdb files were generated using the `molfile_to_params.py` application. Using the .pdb format of DBC, ligand and protein structures were concatenated into a single file using the following Terminal command:

```
$cat <protein.pdb> <ligand.pdb> > <complex.pdb>
```

4.9.2. Parameters for ligand docking

The scoring parameters used for ligand docking was based on DeLuca et al[28]. The XML script can be used to apply constraints and guide the ligand to the known binding or active site. The algorithm begins by positioning the ligand at a user-specified starting point with the StartFrom mover. The ligand is then sampled within a specified radius via the Transform mover and Monte Carlo Minimization algorithm until clashing between side chains and ligand are minimized. We used the xyz coordinates between the two water molecules originally present in the protein PDB file as the starting point for the ligand with a box radius of 8.0 Å.

Using the data derived from the NOE, atomic distance constraints of 6.0 Å were applied between A14 and I40 Hα atoms and the tertiary carbon of 3' *tert*butyl group. The XML script was run using the following command:

```
$<path_to_Rosetta>/main/source/bin/rosetta_scripts.default.macos
clangrelease -database <path_to_database>
```

With the flags:

```
-s <complex.pdb>
-nstruct 2000
-in:ignore_unrecognized_res
-run
     -ignore_zero_occupancy false
-qsar
     -max_grid_cache_size 1
-mistakes
     -restore_pre_talaris_2013_behavior
-score:analytic_etable_evaluation true
-ex1
-ex2
```

The .params file and XML script were utilized by adding the flag `-extra_res_fa <ligand.params>` and `-parser:protocol <path_to_XML_script.xml>` respectively. Evaluation of the lowest energy models were solely based on the total_score, where the lowest scoring pose most likely represents DBC and DFsc interactions. The XML script is as follows:

```
<ROSETTASCRIPTS>
    <SCOREFXNS>
        <ScoreFunction name="ligand_soft_rep" weights="ligand_soft_rep">
            <Reweight scoretype="fa_elec" weight="0.42"/>
            <Reweight scoretype="hbond_bb_sc" weight="1.3"/>
            <Reweight scoretype="hbond_sc" weight="1.3"/>
            <Reweight scoretype="rama" weight="0.2"/>
        </ScoreFunction>

        <ScoreFunction name="hard_rep" weights="ligand">
            <Reweight scoretype="fa_intra_rep" weight="0.004"/>
            <Reweight scoretype="fa_elec" weight="0.42"/>
            <Reweight scoretype="hbond_bb_sc" weight="1.3"/>
            <Reweight scoretype="hbond_sc" weight="1.3"/>
            <Reweight scoretype="rama" weight="0.2"/>
        </ScoreFunction>
    </SCOREFXNS>
    <LIGAND_AREAS>
        <LigandArea name="docking_sidechain" chain="X" cutoff="6.0"
add_nbr_radius="true" all_atom_mode="true" minimize_ligand="10"/>
        <LigandArea name="final_sidechain" chain="X" cutoff="6.0"
add_nbr_radius="true" all_atom_mode="true"/>
        <LigandArea name="final_backbone" chain="X" cutoff="7.0"
add_nbr_radius="false" all_atom_mode="true" Calpha_restraints="0.3"/>
    </LIGAND_AREAS>

    <INTERFACE_BUILDERS>
        <InterfaceBuilder name="side_chain_for_docking"
ligand_areas="docking_sidechain"/>
        <InterfaceBuilder name="side_chain_for_final"
ligand_areas="final_sidechain"/>
        <InterfaceBuilder name="backbone" ligand_areas="final_backbone"
extension_window="3"/>
    </INTERFACE_BUILDERS>

    <MOVEMAP_BUILDERS>
        <MoveMapBuilder name="docking" sc_interface="side_chain_for_docking"
minimize_water="true"/>
        <MoveMapBuilder name="final" sc_interface="side_chain_for_final"
bb_interface="backbone" minimize_water="true"/>
    </MOVEMAP_BUILDERS>

    <SCORINGGRIDS ligand_chain="X" width="17.0">
        <ClassicGrid grid_name="vdw" weight="1.0"/>
```

```
    </SCORINGGRIDS>

    <RESIDUE_SELECTORS>
      <Index name="116X" resnums="116"/>
      <Index name="A14" resnums="14"/>
      <Index name="I40" resnums="40"/>
    </RESIDUE_SELECTORS>

    <MOVERS>
        <StartFrom name="start_from" chain="X">
            <Coordinates x="3.82999992" y="0.16" z="-1.33899999"/>
        </StartFrom>
      <AddConstraints name="add_csts" >
            <DistanceConstraintGenerator name="gen_my_csts"
residue_selector1="116X" residue_selector2="A14" atom_name1="C1"
atom_name2="HA" function="HARMONIC 6.0 0.5" />
          <DistanceConstraintGenerator name="gen_my_csts"
residue_selector1="116X" residue_selector2="I40" atom_name1="C1"
atom_name2="HA" function="HARMONIC 6.0 0.5" />

      </AddConstraints>
        <Transform name="transform" chain="X" box_size="8.0"
move_distance="0.3" angle="5.0" cycles="500" repeats="1" temperature="5"
initial_perturb="3.0" />

        <ParsedProtocol name="low_res_dock">
            <Add mover_name="start_from"/>
          <Add mover_name="add_csts"/>
            <Add mover_name="transform"/>
        </ParsedProtocol>


        <HighResDocker name="high_res_docker" cycles="6" repack_every_Nth="3"
scorefxn="ligand_soft_rep" movemap_builder="docking"/>
        <FinalMinimizer name="final" scorefxn="hard_rep"
movemap_builder="final"/>
        <InterfaceScoreCalculator name="add_scores" chains="X"
scorefxn="hard_rep" compute_grid_scores="0"/>

        <ParsedProtocol name="high_res_dock">
            <Add mover_name="high_res_docker"/>
            <Add mover_name="final"/>
        </ParsedProtocol>

        <ParsedProtocol name="reporting">
            <Add mover_name="add_scores"/>
        </ParsedProtocol>
    </MOVERS>

    <PROTOCOLS>
        <Add mover_name="low_res_dock"/>
        <Add mover_name="high_res_dock"/>
        <Add mover_name="reporting"/>
    </PROTOCOLS>
    <OUTPUT scorefxn="hard_rep" />
</ROSETTASCRIPTS>
```

The functions used for applying the constraints were not applied to penalize or reward the `total_score`. The ligand was assigned as chain X and residue 116. All scoring functions, with the exception of the HARMONIC function used in DistanceConstraintGenerator, were based on the parameters by DeLuca et al[28], which also used in the Rosetta Online Server that Includes Everyone (ROSIE).

References

1. Waldron, K. J. & Robinson, N. J. How do bacterial cells ensure that metalloproteins get the correct metal? *Nat. Rev. Microbiol.* **7,** 25–35 (2009).
2. Valeyev, N. V., Heslop-Harrison, P., Postlethwaite, I., Kotov, N. V. & Bates, D. G. Multiple calcium binding sites make calmodulin multifunctional. *Mol. BioSyst.* **4,** 66–73 (2008).
3. Kaplan, J. & DeGrado, W. F. De novo design of catalytic proteins. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 11566–11570 (2004).
4. Huang, P.-S., Boyken, S. E. & Baker, D. The coming of age of de novo protein design. *Nature* **537,** 320–327 (2016).
5. Dahiyat, B. I., Sarisky, C. A. & Mayo, S. L. De Novo protein design: towards fully automated sequence selection 1 1Edited by P. E. Wright. *J. Mol. Biol.* **273,** 789–796 (1997).
6. Woolfson, D. N. *et al.* De novo protein design: How do we expand into the universe of possible protein structures? *Curr. Opin. Struct. Biol.* **33,** 16–26 (2015).
7. Huang, P.-S., Boyken, S. E. & Baker, D. The coming of age of de novo protein design. *Nature* **537,** 320–327 (2016).
8. Faiella, M. *et al.* An artificial di-iron oxo-protein with phenol oxidase activity. *Nat. Chem. Biol.* **5,** 882–4 (2009).
9. Reig, A. J. *et al.* Alteration of the oxygen-dependent reactivity of de novo Due Ferri proteins. *Nat. Chem.* **4,** 900–906 (2012).
10. Ulas, G., Lemmin, T., Wu, Y., Gassner, G. T. & DeGrado, W. F. Designed metalloprotein

stabilizes a semiquinone radical. *Nat. Chem.* **8,** 354–359 (2016).

11. Chakrabarty, S., Austin, R. N., Deng, D., Groves, J. T. & Lipscomb, J. D. Radical intermediates in monooxygenase reactions of Rieske dioxygenases. *J. Am. Chem. Soc.* **129,** 3514–3515 (2007).
12. Mbughuni, M. M. *et al.* Trapping and spectroscopic characterization of an FeIII-superoxo intermediate from a nonheme mononuclear iron-containing enzyme. *Proc. Natl. Acad. Sci.* **107,** 16788–16793 (2010).
13. Hegg, E. L. & Que, L. The 2-His-1-carboxylate facial triad--an emerging structural motif in mononuclear non-heme iron(II) enzymes. *Eur. J. Biochem.* **250,** 625–629 (1997).
14. Kal, S. & Que, L. Dioxygen activation by nonheme iron enzymes with the 2 - His - 1 - carboxylate facial triad that generate high - valent oxoiron oxidants. (2017). doi:10.1007/s00775-016-1431-2
15. Liu, A. *et al.* Alternative Reactivity of an r -Ketoglutarate-Dependent Iron (II) Oxygenase : Enzyme Self-Hydroxylation. 8711–8714 (2001).
16. Mehn, M. P., Fujisawa, K., Hegg, E. L. & Que, L. Oxygen Activation by Nonheme Iron (II) Complexes : r-Keto Carboxylate versus Carboxylate. 7828–7842 (2003).
17. Ryle, M. J. *et al.* O2- and alpha-ketoglutarate-dependent tyrosyl radical formation in TauD, an alpha-keto acid-dependent non-heme iron dioxygenase. *Biochemistry* **42,** 1854–1862 (2003).
18. Price, J. C., Barr, E. W., Hoffart, L. M., Krebs, C. & Bollinger, J. M. Kinetic dissection of the catalytic mechanism of taurine:alpha-ketoglutarate dioxygenase (TauD) from Escherichia coli. *Biochemistry* **44,** 8138–8147 (2005).

19. Ye, S. *et al.* Electronic structure analysis of the oxygen-activation mechanism by Fe II- and α-ketoglutarate (αKG)-dependent dioxygenases. *Chem. - A Eur. J.* **18,** 6555–6567 (2012).

20. Roberts, K. M., Pavon, J. A. & Fitzpatrick, P. F. Kinetic mechanism of phenylalanine hydroxylase: Intrinsic binding and rate constants from single-turnover experiments. *Biochemistry* **52,** 1062–1073 (2013).

21. Goodwill, K. E., Sabatier, C. & Stevens, R. C. Crystal structure of tyrosine hydroxylase with bound cofactor analogue and iron at 2.3 ?? resolution: Self-hydroxylation of Phe300 and the pterin- binding site. *Biochemistry* **37,** 13437–13445 (1998).

22. Brophy, M. B., Hayden, J. A. & Nolan, E. M. Calcium ion gradients modulate the zinc affinity and antibacterial activity of human calprotectin. *J. Am. Chem. Soc.* **134,** 18089–18100 (2012).

23. Goldenberg, D. P. MECHANISMS OF FOLDING. (1988).

24. Greenfield, N. J. Using circular dichroism spectra to estimate protein secondary structure. *Nat. Protoc.* **1,** 2876–2890 (2006).

25. Siegbahn, P. E. M. & Haeffner, F. Mechanism for catechol ring-cleavage by non-heme iron extradiol dioxygenases. *J. Am. Chem. Soc.* **126,** 8919–8932 (2004).

26. Emerson, J. P., Kovaleva, E. G., Farquhar, E. R., Lipscomb, J. D. & Que, L. Swapping metals in Fe- and Mn-dependent dioxygenases: evidence for oxygen activation without a change in metal redox state. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 7347–7352 (2008).

27. Hassett, D. J. *et al.* Hydrogen peroxide sensitivity of catechol-2,3-dioxygenase: A cautionary note on use of xylE reporter fusions under aerobic conditions. *Appl.*

Environ. Microbiol. **66,** 4119–4123 (2000).

28. Combs, S. A. *et al.* Small-molecule ligand docking into comparative models with Rosetta. *Nat. Protoc.* **8,** 1277–1298 (2013).

29. DeLuca, S., Khar, K. & Meiler, J. Fully flexible docking of medium sized ligand libraries with rosettaligand. *PLoS One* **10,** 1–19 (2015).