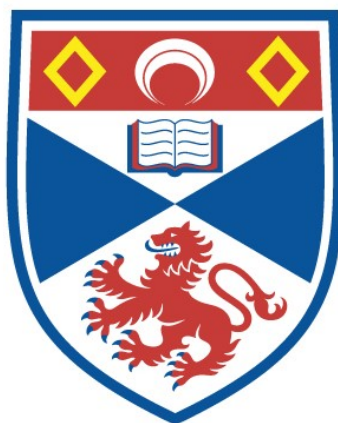


ARTIFICIAL METALLOENZYMES: MODIFIED PROTEINS AS  
TUNEABLE TRANSITION METAL CATALYSTS AND THEIR  
APPLICATION IN OXIDATIVE LIGNIN DEGRADATION

Megan V. Doble

A Thesis Submitted for the Degree of PhD  
at the  
University of St Andrews



2019

Full metadata for this item is available in  
St Andrews Research Repository  
at:

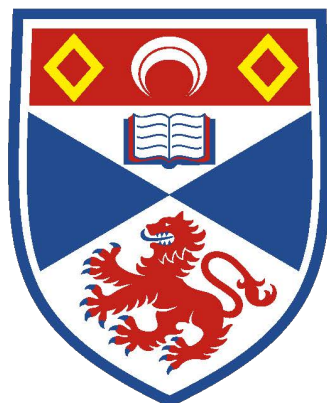
<http://research-repository.st-andrews.ac.uk/>

Please use this identifier to cite or link to this item:

<http://hdl.handle.net/10023/19242>

This item is protected by original copyright

*Artificial Metalloenzymes; Modified Proteins as  
Tuneable Transition Metal Catalysts and their  
Application in Oxidative Lignin Degradation.*



University of  
St Andrews

School of Chemistry  
Fife, Scotland

**Megan V. Doble**

July 2018

*Thesis submitted to the University of St Andrews in application for the degree  
of Doctor of Philosophy*

Supervisor: Professor Paul C. J. Kamer

### **Candidate's declaration**

I, Megan Doble, do hereby certify that this thesis, submitted for the degree of PhD, which is approximately 25,909 words in length, has been written by me, and that it is the record of work carried out by me, or principally by myself in collaboration with others as acknowledged, and that it has not been submitted in any previous application for any degree.

I was admitted as a research student at the University of St Andrews in September 2013.

I received funding from an organisation or institution and have acknowledged the funder(s) in the full text of my thesis.

Date 29/08/2019

Signature of candidate

### **Supervisor's declaration**

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of PhD in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date 10/09/2019

Signature of supervisor

### **Permission for publication**

In submitting this thesis to the University of St Andrews we understand that we are giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. We also understand, unless exempt by an award of an embargo as requested below, that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that this thesis will be electronically accessible for personal or research use and that the library has the right to migrate this thesis into new electronic forms as required to ensure continued access to the thesis.

I, Megan Doble, confirm that my thesis does not contain any third-party material that requires copyright clearance.

The following is an agreed request by candidate and supervisor regarding the publication of this thesis:

**Printed copy**

No embargo on print copy.

**Electronic copy**

No embargo on electronic copy.

Date 29/08/2019

Signature of candidate

Date 10/09/2019

Signature of supervisor



## **Underpinning Research Data or Digital Outputs**

### **Candidate's declaration**

I, Megan Doble, understand that by declaring that I have original research data or digital outputs, I should make every effort in meeting the University's and research funders' requirements on the deposit and sharing of research data or research digital outputs.

Date 29/08/2019

Signature of candidate

### **Permission for publication of underpinning research data or digital outputs**

We understand that for any original research data or digital outputs which are deposited, we are giving permission for them to be made available for use in accordance with the requirements of the University and research funders, for the time being in force.

We also understand that the title and the description will be published, and that the underpinning research data or digital outputs will be electronically accessible for use in accordance with the license specified at the point of deposit, unless exempt by award of an embargo as requested below.

The following is an agreed request by candidate and supervisor regarding the publication of underpinning research data or digital outputs:

No embargo on underpinning research data or digital outputs.

Date 29/08/2019

Signature of candidate

Date 10/09/2019

Signature of supervisor

## Table of Contents

Declarations	II
Table of Contents	V
List of Abbreviations	IX
Abstract	XI

### **Chapter 1: Introduction to Artificial Metalloenzymes and their application to Oxidation**

<b>Catalysis</b>	<b>1</b>
1.1 Natures enzymes for oxidation reactions	1
1.1.2 Heme Containing Oxidative Enzymes	
1.1.2 Non-heme Iron Oxidative Enzymes	
1.1.3 Copper Containing Oxidative Enzymes	
1.2 Advantages in the design of Artificial metalloenzymes	7
1.3 Strategies for creating artificial metalloenzymes	9
1.4 Application of artificial metalloenzymes hybrid catalysts in oxidation reactions	11
1.4.1 Alcohol oxidation	
1.4.2 Sulfoxidation	
1.4.3 C=C and C-H bond oxygenation	
1.5 Outlook on this thesis	14
<b>2 Chapter 2 – Design and Characterisation of Artificial Metalloenzymes</b>	<b>19</b>
2.1 Artificial Metalloenzyme Design	19
2.2 Cofactor Design and Synthesis	21
2.3.1 Nitrogen containing ligand library	
2.3.2 Synthetic Route to Cofactor E	
2.4 Protein Expression and Purification	23
2.5 Protein Scaffold Modification	25
2.5.1 General Methodology	
2.5.2 LCMS Analysis	

2.5.3 CD Analysis	
2.6 Determining Concentration of Artificial Metalloenzymes	29
2.6.1 Concentration calculation using Bradford reagents	
2.6.1 QAA Analysis	
2.7 Metal binding studies of Artificial Metalloenzymes	33
2.8 Conclusions and Further Work	36
2.9 Experimental	37
2.9.1 General Remarks	
2.9.2 Synthesis of Cofactor E	
2.9.3 Protein Expression and Purification of SCP-2L A100C & V83C	
2.9.4 Bioconjugation of cofactor to unique cysteine mutants of SCP-2L	
2.9.5 LCMS Analysis of Modified Proteins	
2.7.5 Determination of Modified Protein Concentration	
<b>Chapter 3 – Engineering thermostability in SCP-2L to increase catalytic activity for hydroformylation</b>	<b>48</b>
3.1 Introduction to protein engineering	48
3.2 Design of thermostable mutants	50
3.3 Synthesis of mutant library	53
3.4 Thermostability testing of mutant library	53
3.4.1 Thermoflour analysis of proteins	
3.4.2 Using CD spectroscopy to determine thermostability of protein scaffolds	
3.5 Design and synthesis of Double mutants	56
3.6 CD spectroscopy analysis of double mutants	56
3.7 Hydroformylation catalysis using double mutants	58
3.7.1 Protein modification	
3.7.2 Hydroformylation of 1-octene catalyzed by rhodium-hydroformylase	
3.8 Conclusions and Further Work	62
3.9 Experimental	63
3.9.1 General Remarks	
3.9. 2 Protein Crystal Structures	

- 3.9.3. Site Directed Mutagenesis
- 3.9.4 Protein Expression and Purification
- 3.9.5 CD Studies to determine thermostability
- 3.9.6 Protein Modification
- 3.9.7 LCMS Analysis of Modified Proteins
- 3.9.8 Hydroformylation

## **Chapter 4 – Artificial Metalloenzymes as oxidation catalysts for lignin degradation 79**

- 4.1 Introduction to lignin 79
  - 4.1.1 Structure and Functionality of Lignin
  - 4.1.2 Naturally occurring lignin degradation
  - 4.1.3 Transition Metal Catalysts for Lignin Oxidation
  - 4.1.4 Biomimetic metalloporphyrins utilised in lignin oxidation
  - 4.1.5 Artificial Metalloenzymes for lignin oxidation hypothesis
- 4.2 Results 87
  - 4.2.1 Lignin Model Compounds Synthesis
  - 4.2.2 Oxidation catalysis of lignin model compounds
  - 4.2.3 Artificial metalloenzymes as oxidation catalysis of lignin model compounds
  - 4.2.4 Kinetic Study of Lignin Model Oxidation with ArM SCP-2L A100C-D-Fe<sup>2+</sup>
  - 4.2.5 Varying the Substrate Scope of Lignin Model Compounds
  - 4.2.6 Investigating ArM reactivity with Organic Solvent Tolerance
  - 4.2.7 Mutants to increase cofactor stability
    - 4.2.7.1 Design of mutants in which protein secondary coordination sphere would stabilize cofactor
    - 4.2.7.2 Lignin model oxidation with mutant in optimum conditions
- 4.3 Experimental 103
  - 4.3.1 General Remarks
  - 4.3.2 Protein Expression and purification
  - 4.3.3 LCMS analysis of SCP-2L A100C mutants F94E/H
  - 4.3.4 Lignin Model Compound Synthesis
  - 4.3.5 Lignin Model Compound Oxidation

4.3.6 Kinetics Study

**Chapter 5 Summary and Conclusion** **117**

**Acknowledgements** **119**

## List of Abbreviations

<b>acac</b>	Acetylacetonato
<b>ArM</b>	Artificial Metalloenzyme
<b>BSA</b>	Bovine serum albumin
<b>BVMO</b>	Baeyer-Villiger monooxygenases
<b>CD</b>	Circular dichroism
<b>DCM</b>	Dichloromethane
<b>DMSO</b>	Dimethylsulfoxide
<b>DNA</b>	Deoxyribonucleic acid
<b>DNase</b>	Deoxyribonuclease (DNase I from bovine pancreas)
<b>DTT</b>	Dithiothreitol
<b>E. coli</b>	Escheria coli
<b>EDTA</b>	Ethylendiaminetetraacetic acid
<b>ESI</b>	Electrospray ionisation
<b>eq</b>	Equivalent
<b>GC</b>	Gas chromatography
<b>h</b>	Hour(s)
<b>HPLC</b>	High pressure liquid chromatography
<b>ICP</b>	Inductively coupled plasma
<b>IPTG</b>	Isopropyl $\beta$ -D-1-thiogalactopyranoside
<b>Kan</b>	Kanamycin
<b>LB</b>	Lysogeny broth
<b>LCMS</b>	Liquid chromatography-mass spectrometry
<b>MALDI</b>	Matrix assisted laser desorbition/ionisation
<b>MeCN</b>	Acetonitrile
<b>MES</b>	2-( <i>N</i> -morpholino)ethanesulfonic acid
<b>MFE</b>	Multifunctional enzyme
<b>MWCO</b>	Molecular weight cut off
<b>NMR</b>	Nuclear magnetic resonance
<b>PB</b>	Production broth

<b>rpm</b>	Revolutions per minute
<b>RT</b>	Room temperaure
<b>SCP</b>	Sterol carrier protein
<b>SDS</b>	Sodium dodecyl sulfate
<b>Syn gas</b>	Synthesis gas, a 1:1 mixture of H <sub>2</sub> and CO
<b>TEV</b>	Tobacco etch virus
<b>THF</b>	Tetrahydrofuran
<b>TLC</b>	Thin layer chromatography
<b>TON</b>	Turnover number
<b>TPA</b>	Tris(2-picolyl)amine
<b>TPP</b>	Triphenylphosphine
<b>TPPTS</b>	Sodium triphenylphosphine trisulfonate
<b>Tris</b>	Tris(hydroxymethyl)aminoethane
<b>UV Vis</b>	Ultraviolet-visible



## **Abstract: Artificial Metalloenzymes; Modified Proteins as Tuneable Transition Metal Catalysts and their Application in Oxidative Lignin Degradation.**

The selective oxidation of organic molecules is fundamentally important to life and immensely useful in industry. Metalloenzyme catalysed oxidations often display exquisite substrate specificity as well as regio and/or stereoselectivity. Huge strides have occurred in the field of biocatalysis in recent years. Work has developed by taking inspiration from nature's enzymes, to use directed evolution and engineering methods to create tailor made catalysts. Artificial Metalloenzymes (ArMs) provide the possibility to expand this repertoire further by combining the advantageous features of enzymes with the versatile reaction scope of transition metals.

The initial chapter in this thesis takes a look into recent literature about artificial metalloenzymes and their application in oxidation catalysis.

Chapter two describes the design rationale and synthesis of protein templates and synthetic cofactors for the development of artificial metalloenzymes. Successful modification was achieved for a wide library of nitrogen donor ligands, creating an array of artificial metalloenzymes that can be tested in catalytic reactions. In the absence of a crystal structure of the modified protein, UV and CD analysis were carried out to gather characteristic information about the artificial metalloenzymes and their metal binding properties. An investigation was also carried out to determine the most accurate method to calculate protein concentration once it has been modified with a cofactor.

The third chapter describes the application of protein engineering to increase the thermostability of the target protein. Variants of an artificial metalloenzyme were created by rational design using structural and bioinformatic information. The variants were tested to identify mutations that enhanced the stability of the protein scaffold. Significant increases in melting temperature were observed in a number of the modified metalloenzymes. Their ability to withstand higher reaction temperatures resulted in increased activity in the hydroformylation of 1-octene, with >5-fold improvements in turnover numbers (TON).

The fourth chapter reports the use of artificial metalloenzymes in oxidation catalysis. In particular their application to the degradation of lignin is investigated. Using a model compound that mimics the most abundant linkage within lignin as a substrate, a wide array of artificial metalloenzymes were tested to study if any oxidation or cleavage occurs. Investigations were carried out to find the optimum conditions varying catalyst loading and buffer/solvent composition. Complete selective conversion to ketone product is observed using SCP-2L A100C modified with a tris(2-pyridylmethyl) amine based cofactor, coordinated to  $\text{Fe}(\text{OTf})_2 \cdot 2\text{MeCN}$ . Engineering the protein scaffold to incorporate glutamic acid was found to improve the ArM activity, showing that rational design of the protein environment using metal binding amino acids can be a method to improve the overall activity of an artificial metalloenzyme.

## Chapter 1 - Introduction to Artificial Metalloenzymes and their application to Oxidation Catalysis

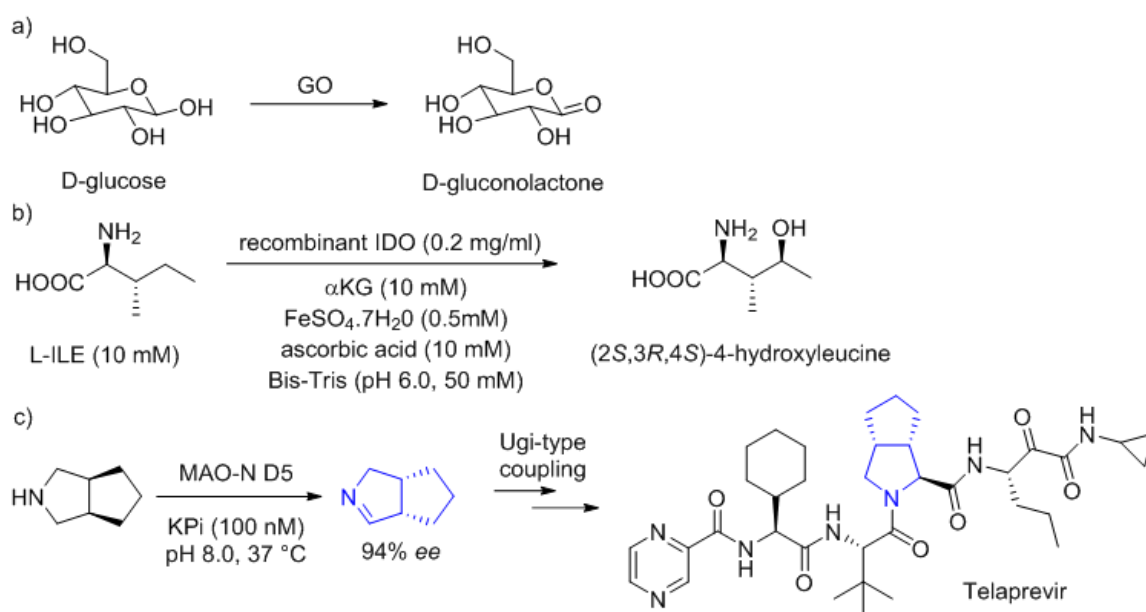
The selective oxidation of organic molecules is fundamentally important to life and immensely useful in industry.<sup>1, 2</sup> Metalloenzyme catalysed oxidations often display exquisite substrate specificity as well as regio- and/or stereoselectivity.<sup>3</sup> Huge strides have occurred in the field of biocatalysis in recent years. Work has developed by taking inspiration from nature's enzymes, to using directed evolution and engineering methods to create tailor made catalysts. Artificial Metalloenzymes (ArMs) provide the possibility to expand this repertoire further by combining the advantageous features of enzymes with the versatile reaction scope of transition metals.

### 1.1 Nature's enzymes for oxidation reactions

Nature has evolved over billions of years to utilise enzymes for extremely efficient and selective multi-step reactions. Metalloenzymes have been used increasingly within oxidation catalysis and they often outperform the low molecular weight metal complexes utilised by chemists. Metalloenzyme catalyzed oxidations often demonstrate high substrate specificity alongside high levels of enantio- and regio-selectivity.<sup>4</sup> Naturally occurring enzymes have been shown to perform oxidative reactions including hydroxylation, dihydroxylation of aromatic and aliphatic C-H bonds, epoxidation, Baeyer-Villiger oxidation of ketones to lactones, and halohydrin formation from alkenes.<sup>5</sup> Enzymatic oxidation reactions have been introduced on an industrial scale, and interesting examples can be seen in Scheme 1. Within the food and wine industry glucose oxidase (GO) is being used for the oxidation of D-glucose to D-gluconolactone (Scheme 1a).<sup>6</sup> A well-known industrial process that utilises enzymatic oxidation is the two-step process for the synthesis of ethyl (*R*)-4-cyano-3-hydroxybutyrate, an important intermediate in atorvastatin production.<sup>7</sup>

Another example is the production of hydroxy amino acids; their antidiabetes activity make them a useful target. An example of their synthesis can be seen in Scheme 1b using L-isoleucine dioxygenase (IDO).<sup>8</sup> The GO and IDO examples are interesting because these utilise the wild type enzymes whereas a significant number of current examples of industrial applications use designed mutants. More recently, examples have appeared in medicinal chemistry, such as the use of a mutant monoamine oxidase in the production of cyclic imine

intermediates for the synthesis of telaprevir, a pharmaceutical drug used for the treatment of hepatitis C (Scheme 1c).<sup>9</sup> In order for enzymes to be used in industrial applications, they need to be of high stability and productivity. Therefore, in many circumstances protein engineering is used to enhance the natural ability of the enzyme or adapt an enzyme to specific needs.



*Scheme 1: Industrial Examples utilising natural enzymes.*

*a) Oxidation of D-glucose, b) production of hydroxy amino acids c) Synthesis of telaprevir. ( $\alpha\text{KG}$  =  $\alpha$ -ketoglutarate,  $\text{KPi}$  = potassium phosphate buffer).<sup>10</sup>*

There are a wide range of different oxidative enzyme types, but they predominately use iron or copper metal centres.<sup>11</sup> In Figure 1 the different active sites found within nature's oxidative enzymes are shown. Iron and copper ions are the metal ions of choice for many biological oxidative systems due to their natural abundance, electronic properties and accessible redox potentials. I will briefly discuss key examples of oxidation catalysis from each different class of oxidation enzymes.

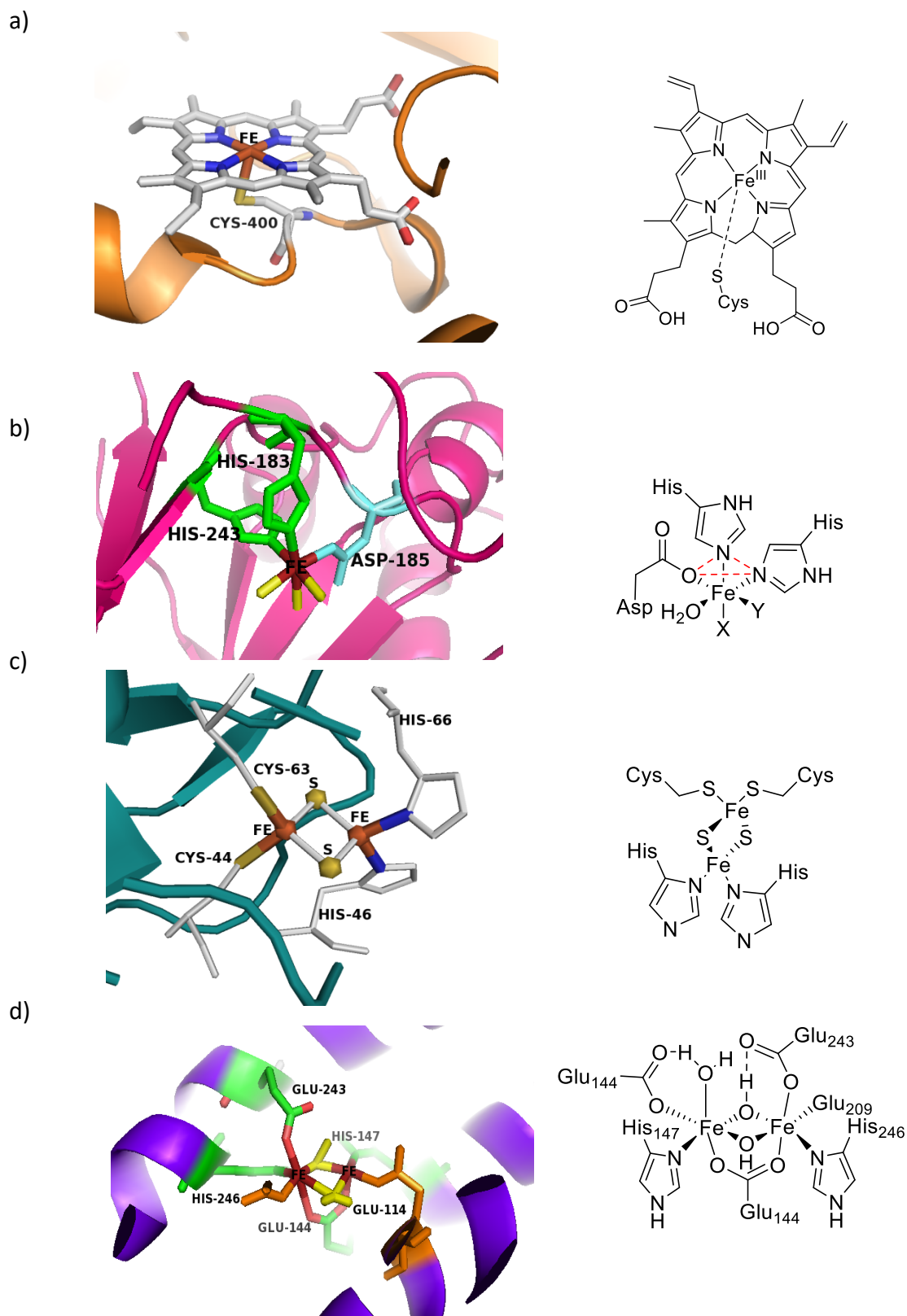
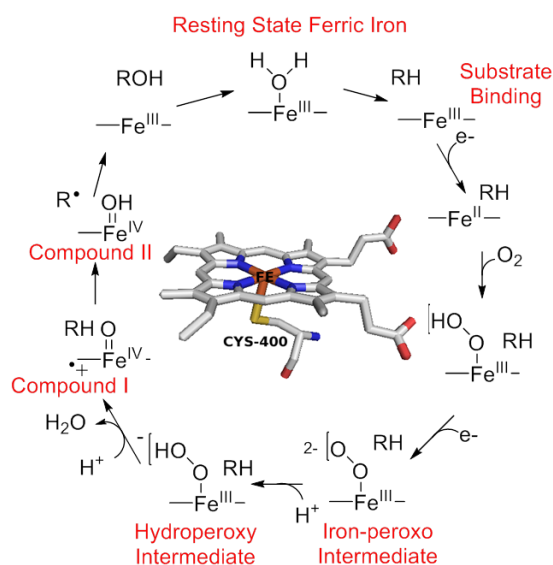


Figure 1: Diagrams to represent the different Iron containing active sites found within nature  
a) P450 Heme Active Site (PDB 1JPZ) b) 2-His-1-carboxylate facial triad motif, His (Green) Asp (Blue) Water (Yellow) (PDB 1RXF) c) [2Fe-2S] Cluster (PDB 1UUW) d) Di-iron centre seen in monooxygenases Glu (Orange) His (Green) Water (Yellow) (PDB 1MTY).<sup>10</sup>

### 1.1.2 Heme Containing Oxidative Enzymes

The heme containing enzymes are among the most studied classes of oxidative enzymes, and include peroxidases, catalases and P450s. These redox enzymes can be used to activate hydrogen peroxide or molecular oxygen, and transfer oxygen to the substrate *via* high-valent iron-oxo intermediates (Scheme 2).<sup>12</sup> The active site comprises of a porphyrin ring anchored to the protein backbone through a histidine or cysteine residue and also *via* ionic, van der Waals and hydrogen bonding interactions (Figure 1a).<sup>13</sup> In nature, different protoporphyrin rings are found, and consequently different types of heme can exist. Generally, iron-protoporphyrin IX complex, known as “heme b”, is referred to as heme.

An important intermediate of the heme enzyme’s catalytic cycle is Compound I, an oxoiron(IV) porphyrin  $\pi$  cation radical, which has the ability to oxidize a variety of substrates, a factor often exploited in biomimetic approaches to oxidative catalysis.<sup>14,15</sup>

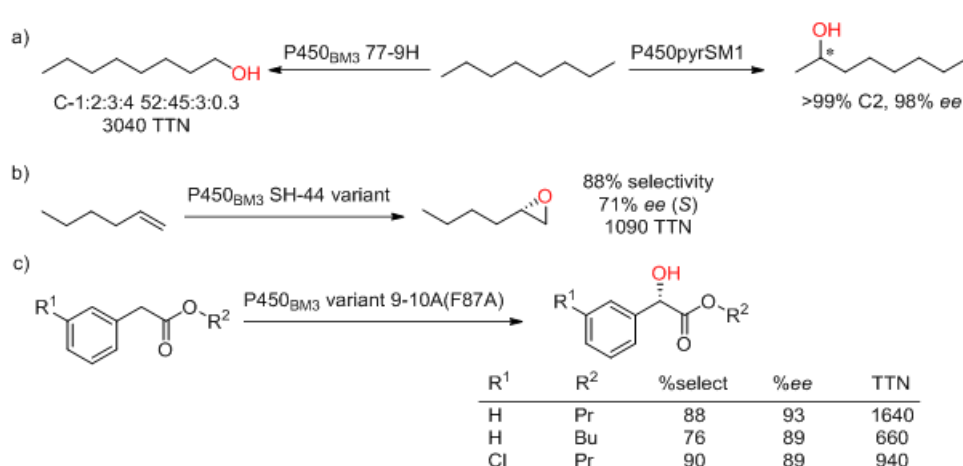


*Scheme 2: The P450 catalytic cycle highlighting important intermediates. Centre of cycle indicates the active site with a conserved threonine and cysteine.<sup>10</sup>*

The most extensively studied type of heme containing enzymes is the cytochrome P450 family.<sup>16,17,18</sup> Some of the most significant reactions by P450s are monooxygenation of aliphatic and aryl C-H bonds, alkene epoxidation and sulfoxidation. Although naturally occurring P450 enzymes possess a narrow substrate scope, their high catalytic activity and

solubility provide an attractive starting point for a bioengineering approach to adapt to other substrates.

Bioengineering of P450s has resulted in impressive regioselectivities in alkane oxidation. For the selective oxidation of *n*-octane, the P450<sub>BM3</sub> mutant 77-9H obtained *via* directed evolution gave rise to 52% selectivity for the terminal position (Scheme 3a). Generally, <10% of the products show 1-octanol for P450<sub>BM3</sub> enzymes. The regioselectivity was shown to be octane specific as the results were not that good for other hydrocarbon chain lengths.<sup>19</sup> The same authors have also targeted the hydroxylation of more complex targets<sup>20, 21</sup> and explored epoxidation<sup>22</sup> utilizing the chirality in the protein scaffold to give rise to enantioselective transformations (Scheme 3b and c).



*Scheme 3: Example applications of bioengineered P450 enzymes*

*a) Hydroxylation of Octane using P450<sub>BM3</sub> 77-9H, b) Enantioselective Epoxidation of Terminal Alkenes Scheme, c) Oxidation of  $\alpha$ -aryl-acetic acid esters. (TTN = total turnover number).<sup>10</sup>*

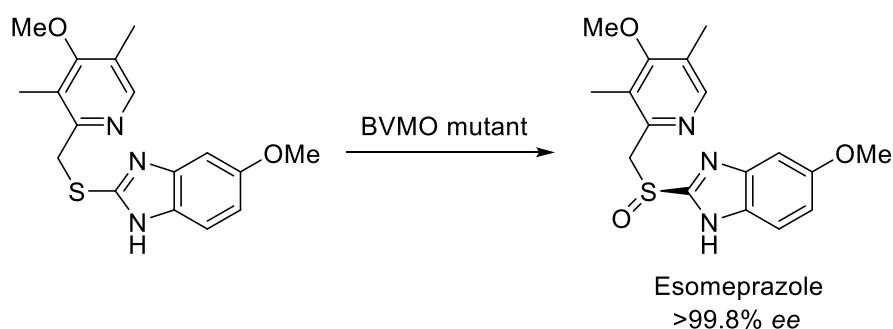
### 1.1.2 Non-heme Iron Oxidative Enzymes

Monooxygenases produce the selective oxidation of a variety of substrates depending on the subtype of enzymes and in all cases they use molecular oxygen as an oxidant.<sup>23</sup>

A class of monooxygenases that has been utilised within the pharmaceutical industry are the Baeyer-Villiger monooxygenases (BVMO), which are capable of oxidation of both linear and cyclic ketones to esters and lactones using molecular oxygen. The BVMO enzymes have been applied within microbes, as enzyme extracts and as recombinant enzymes in organic



synthesis.<sup>24</sup> Several reviews have appeared in recent years describing the use of BVMOs in organic reactions such as steroid transformations, degradation of ketones, heteroatom oxidation, aldehyde oxidation and epoxidation.<sup>25,26</sup> The reactions, however, are complicated due to the need for regeneration and retention of a cofactor. Significant progress has been made in co-factor recycling with solutions such as immobilization, continuous substrate feeding, resins and using biphasic reactions. Regeneration is often achieved using a parallel enzymatic or non-enzymatic redox reaction. A recently developed example involved protein engineering on a BVMO to invert the enantioselectivity and improve the activity, stability and chemo-selectivity of the enzyme. This novel BVMO catalyses a sulfoxidation reaction to give the active pharmaceutical ingredient Esomeprazole (Scheme 4). Nevertheless, more progress is needed on improving the stability and efficiency of BVMO enzymes for their effective use within organic synthesis.

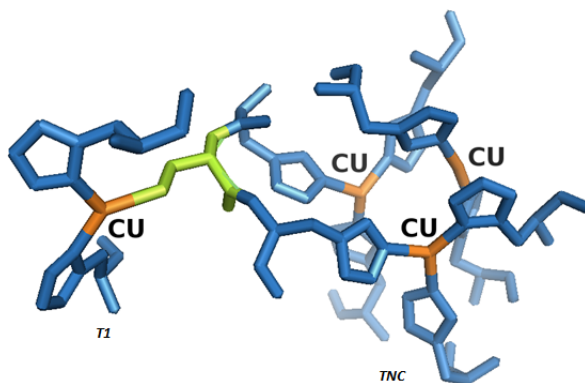


Scheme 4: BVMO sulfoxidation to synthesise Esomeprazole.<sup>10</sup>

### 1.1.3 Copper Containing Oxidative Enzymes

Copper containing oxidases and oxygenases comprise a large class of enzymes that use intriguing mechanisms to bind and activate oxygen to oxidize organic substrates.<sup>27</sup> The active sites within these copper-containing enzymes contain a variable number of copper ions. Laccases are the largest subgroup of blue multicopper oxidases (MCO); they have the capability to catalyze the oxidation of various aromatic substrates including *ortho*- and *para*-diphenols, aminophenols, polyphenols, as well as polyamines and aryl diamines, and inorganic ions.<sup>28</sup> They contain four copper atoms in the active site, which are organised into three different copper centres (Figure 2).<sup>29</sup> Most substrates used by laccase enzymes are

phenols with redox potentials similar to that of the laccase itself; hence their ability to reduce the T1 centre.<sup>30</sup>



*Figure 2: Laccase Active Site (PDB 25HU) illustrating the trinuclear cluster (TNC) of copper atoms and the single copper atom known as a T1 site (His in blue, Cys in Green).<sup>10</sup>*

Laccase proteins are able to maintain stability at 60 °C, which is beneficial for developing large-scale processes that require high temperatures. However, they have low substrate specificity and the enzyme is difficult to recycle. The redox potential of laccases prevents the enzyme from oxidizing primary alcohols but a biphasic system of laccase with a chemical mediator, such as TEMPO, has been successfully developed.<sup>31</sup>

## **1.2 Advantages in the design of Artificial metalloenzymes**

The line between simply mutating an existing metalloenzyme and engineering a novel hybrid catalyst is quite ambiguous. The idea behind ArMs is to combine the advantages of both enzymes and transition metal catalysts to produce an ideal system. Expanding upon the vast scope of enzymes created through evolution by introducing artificial metals creates opportunities for exploring a wider range of enzyme-catalyzed reactions.

Catalysts which have been designed by using scaffolds provided by nature offer many possibilities for tackling such selectivity problems, as they are widely studied, and a wide array of engineering tools are available. Nature's enzymes are the product of evolution and generations of mutations. They can continuously update and optimize their enzymatic repertoires. Directed evolution is a protein engineering tool that can mimic evolution by artificial selection.<sup>32</sup> This method has been widely used to broaden the substrate scope of enzymes or improve their selectivity and/or stability in a targeted manner. A directed

evolution approach can also be applied to ArMs, as shown by random mutagenesis carried out by error-prone PCR to generate dirhodium ArM variants with improved enantioselectivity for a model styrene cyclopropanation reaction.<sup>33</sup>

Introducing non-native metal ions into a protein can expand the repertoire of protein functionalities and its applications. The initial work was published in 1978 by Wilson and Whitesides where they reported a metalloenzyme<sup>34</sup> that catalyzes asymmetric olefin hydrogenation as a homogenous catalyst. The limitations in structural characterization and analysis slowed down the progression of this area at that time. The development of rational design of proteins and metal complexes has increased in the last decade dramatically due to development of bioanalytical techniques and also a larger crossover between biosciences and inorganic chemistry.

The artificial enzyme is unique compared to other catalysts and enzymes, in that two levels of optimization can be envisioned. The first level is the first coordination sphere containing the transition metal / ligand environment, which is open to optimization using synthetic techniques usually within the ligand moiety. The second coordination sphere is the surrounding protein scaffold; this can be fine-tuned using protein engineering.<sup>35,36,37</sup> The concept behind the generation of ArMs is illustrated in Fig 3.

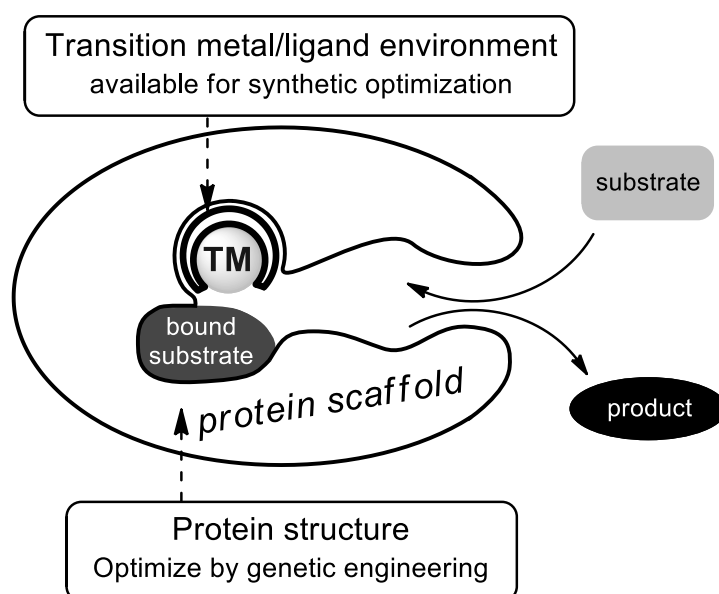


Figure 3: Concept of artificial metalloenzymes and opportunities for catalyst optimisation.<sup>10</sup>

There are no current examples of ArM's being used in industry. However, recent examples have looked at the possible application of ArM's in the pharmaceutical industry. This is an interesting example that has shown the transformation of wild type surface receptors of living cells into ArMs.<sup>38</sup> Work using the Streptavidin-biotin technology has also been applied *in vivo* to catalyze olefin metathesis.<sup>39</sup> While there have been remarkable developments in the activity of ArM's in recent years, the activities remain much lower than those seen from nature's enzymes. One example of a reconstituted artificial metalloenzyme made from a P450 enzyme containing an iridium porphyrin exhibits kinetic parameters similar to those of natural enzymes. This system catalyzed the insertion of carbenes into C-H bonds with up to 98% *ee*, 35000 TON and 2500 h<sup>-1</sup> TOF.<sup>40</sup>

### 1.3 Strategies for creating artificial metalloenzymes

The several strategies for creating these ArMs can be divided into two main categories: non-covalent anchoring and covalent modification.<sup>41, 42, 43</sup> The non-covalent approach can be subdivided into dative anchoring and supramolecular anchoring. Dative anchoring is utilizing the affinity of a protein for a specific transition metal.

The first report to use the dative anchoring approach was Kaiser *et al* in 1976. Carboxypeptidase A (CPA) was used, which in its native form is a zinc-containing enzyme that demonstrates peptidase and esterase activity. The Kaiser group replaced the zinc atom with a copper atom, creating a catalyst capable of oxidizing ascorbic acid. The metal atom exchange was a tremendous alteration, and the active site geometry changed from square planar towards a tetrahedral shape, resulting in different enzyme activity.<sup>44</sup> However, this can be problematic as proteins can have multiple sites for metal binding. Reetz reported the creation of an artificial copper binding site in the synthase subunit of imidazole glycerol phosphate synthase from *Thermotoga maritima* (tHisF).<sup>45</sup> The introduced motif contained two histidine and one aspartic acid residues located at a strategic position to bind a copper atom. The system was then applied to copper catalyzed Diels-Alder reaction between azachalcone

and cyclopentadiene. An improved activity compared to the native protein was seen and 35% *ee* was obtained.

Supramolecular anchoring is based on a high affinity protein-substrate interaction such as biotin-(strept)avidin (dissociation constant in the order of  $10^{-14}$  M).<sup>46</sup> The use of avidin-biotin as a technique to introduce a metal complex into a protein scaffold has been explored very successfully by the group of Ward. The initial work from Ward came from the rhodium-diphosphine catalyzed hydrogenation of N-protected dehydroaminoacids producing an *ee* of up to 95%. This was carried out using streptavidin instead of avidin, which improved the catalytic system due to a deeper binding pocket.<sup>47</sup> This approach combines the advantages of simple catalyst preparation, with resultant fast generation and screening of mutant libraries.<sup>48</sup> However, the anchoring may not always be specific to one region of the scaffold. A whole range of reactions were performed with ArM's based on this strategy (e.g. olefin metathesis<sup>49</sup> and asymmetric transfer hydrogenations of imines)<sup>50</sup>

The most effective way of achieving specific binding of a co-factor, is to covalently tether it to a protein scaffold. The covalent modification of protein structures is performed by site selective modification of the protein with a transition metal complex, for example by making use a nucleophilic amino acid residue. Cysteine, with its strongly nucleophilic thiol functionality is a popular choice for bioconjugation owing to its relatively low occurrence (1.7%) in wild type proteins. The important advantage of covalent modification over supramolecular anchoring is that one is not limited to a certain protein scaffold. Protein hosts can be selected based on desired characteristics such as substrate binding, stability and shape of the binding site. Site directed mutagenesis to introduce a cysteine residue allows the co-factor to be anchored at any site in the protein with high specificity.

Kaiser et al. were the first to create an ArM via covalent modification of an amino acid by alkylation of C25 in papain with bromoacetyl functionalized flavins.<sup>51</sup> This work was the inspiration for much future work for the creation of ArM's using the same covalent attachment procedure. Reetz and De Vries suggested that papain was limited as a protein scaffold within an ArM system because of its large conformational flexibility. It is suggested that structural rigidity is an important factor in ArM activity. By reducing the number of

accessible structures of the catalytic species conformational dynamics of the protein can be limited and/or the functional interactions between the host and the metal-complex can be increased.<sup>52,53</sup>

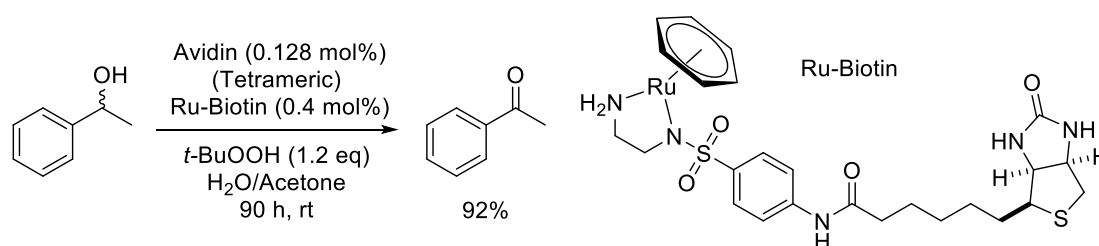
With the development of modelling software<sup>54, 55</sup> *de novo* design has begun to appear, resulting in the design of completely new protein structures. A well-designed example is the novel “Due Ferri” di-iron complex containing protein.<sup>56</sup> DeGrado and co-workers designed a tetra-helical unit, with an iron binding site made up of glutamate and histidine residues.

#### 1.4 Application of artificial metalloenzymes hybrid catalysts in oxidation reactions

Several elegant approaches for the design and application of artificial hybrid oxidation catalysts are highlighted based on the type of oxidation reactions performed.<sup>57,58</sup>

##### 1.4.1 Alcohol oxidation

Oxidation of alcohols is an important reaction in organic chemistry and new mild and green methodology is still in demand. Several elegant bioinspired innovations have been reported using hydrogen transfer pathways.<sup>59, 60, 61</sup> One interesting example of ArM catalysed alcohol oxidation is the oxidation of 1-phenylethanol. This was achieved using a hybrid catalyst developed by supramolecular anchoring of different ruthenium, iridium and rhodium modified biotins (Scheme 5).<sup>62</sup> Benzyl alcohol and cyclohexanol could also be converted into the corresponding aldehyde and ketone respectively (68% and 80% conversion).



*Scheme 5: Benzyl alcohol oxidation using ruthenium amino-sulfamide biotin complexes anchored via supramolecular interaction to streptavidin.*<sup>10</sup>

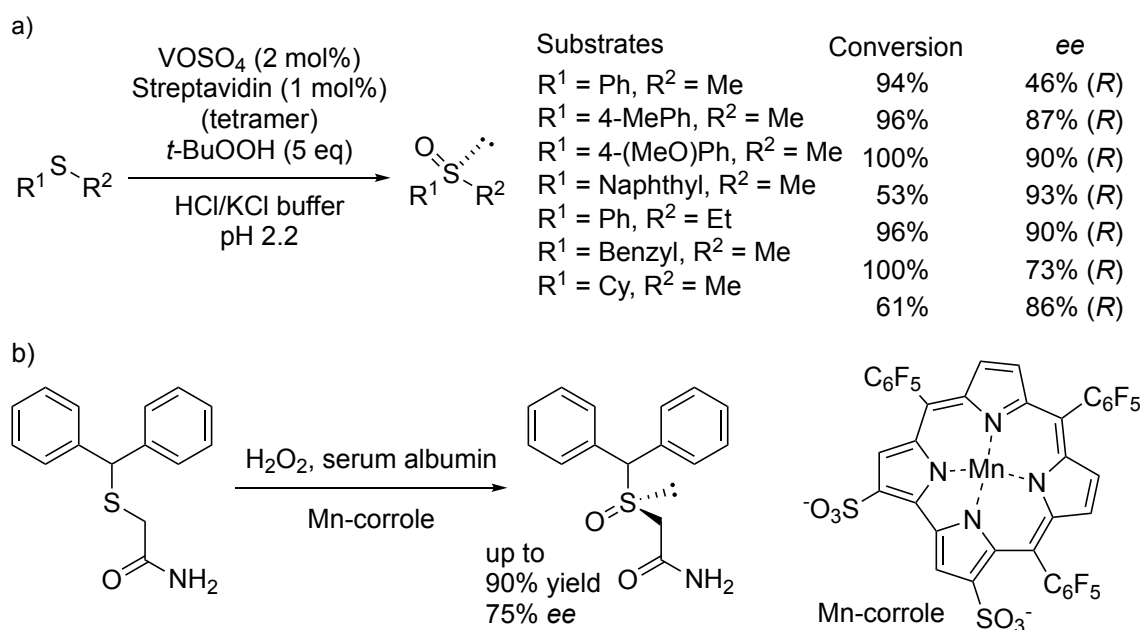
An ArM for the oxidation of catechol derivatives was designed by reprogramming a bacterial metallo- $\beta$ -lactamase into a catechol oxidase.<sup>63</sup> Subsequently, a triple mutant was designed

based on computer-assisted structural analysis of the binding site in order to effectively incorporate a dinuclear copper site.

### 1.4.2 Sulfoxidation

Asymmetric sulfoxidation of methyl phenyl sulfide and similar substrates is a benchmark reaction for assessment of hybrid oxidation catalysts. The aim is obtaining enantioselectivity as it demonstrates the influence of chiral environment of the protein host on the reaction. Ward and co-workers created a highly selective catalyst system for this reaction reaching up to 93% *ee* for naphthalene methyl sulfide using *tert*-butyl hydroperoxide as oxidant. They employed the dative approach incorporating a vanadyl ion in the biotin binding site of streptavidin (Scheme 6).<sup>64</sup> They reasoned that the hydrogen bonds that allow biotin to bind strongly are perfectly suited for coordinating the vanadyl ion while the hydrophobic residues involved in biotin binding are left free for substrate interactions.

Application of the supramolecular interaction between antibodies and metal complexes to obtain enantioselective sulfoxidation catalysts has also been reported (*ee*'s up to 45% *R*, and 43% *S*).<sup>65,66,67</sup> In addition to supramolecular methods, covalent anchoring has also been applied to combine manganese salphen complexes with myoglobin (up to 60% *ee*).<sup>68,69</sup>



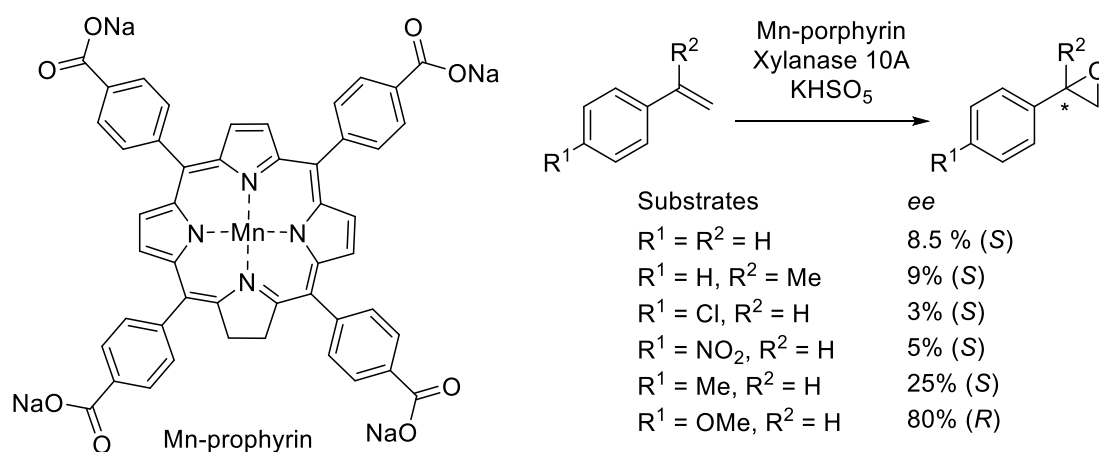


*Scheme 6: (Enantioselective) sulfoxidation by a) complex resulting from the interaction of streptavidin with VOSO<sub>4</sub> b) Supramolecularly anchored manganese corroles into serum albumins*<sup>10</sup>

A simple biomimetic oxidation system was produced by combining metal complexes of amphiphilic corroles with serum albumins. When used in asymmetric sulfoxidation up to 75% *ee* is seen.<sup>70</sup> One drawback of this approach is that synthesis of asymmetric ligands for complexes such as chiral manganese and iron corrole complexes can be extremely time consuming and difficult, compared to achiral derivatives.

### 1.5.3. C=C and C-H bond oxygenation

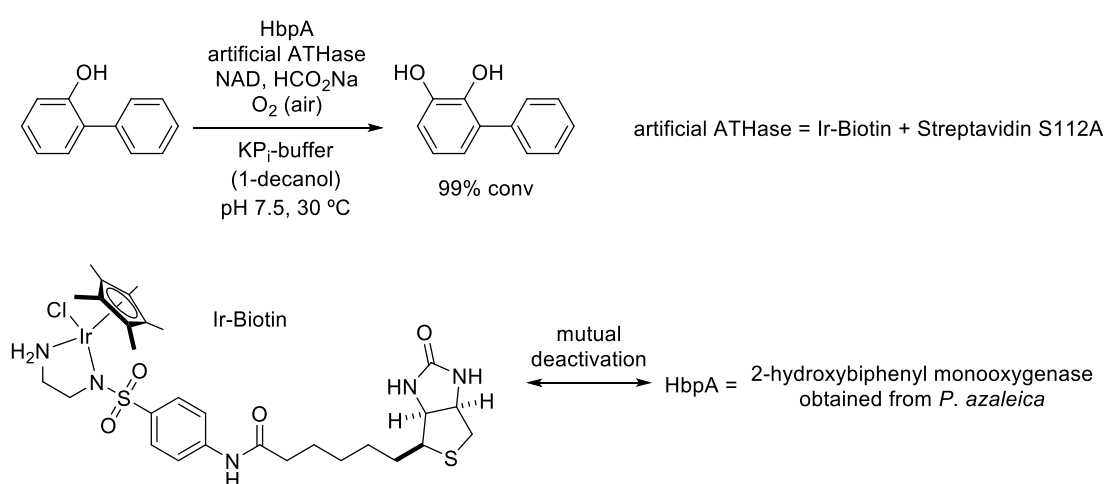
The group of Mahy reported artificial heme containing enzymes which are able to oxidise guaiacol and *o*-dianiside with hydrogen peroxide.<sup>71</sup> These were prepared from Xylanase A, an enzyme that is commercially available. The same group associated manganese porphyrins into Xylanase A for the epoxidation of styrene derivatives using potassium peroxymonosulfate (Scheme 7).<sup>72</sup> With this catalytic system low enantioselectivity was found for the epoxidation of styrene (8.5% *ee*). *p*-Methoxystyrene gave good enantioselectivity (80% *ee*) but conversion was also low (up to 17%).



*Scheme 7: Enantioselective alkene epoxidation using Mn-porphyrins supramolecularly anchored into Xylanase 10A.*<sup>10</sup>

Natural enzymes are able to catalyze complex cascades of chemical reactions in living cells, even in the presence of many other enzyme clusters. This is partly due to efficient compartmentalization of the different active sites protecting them from mutual deactivation. Being able to mimic such tandem reactions would be highly desirable. One-pot combinations

of enzyme reactions with transition metal catalyzed reactions have already been reported<sup>73</sup> but development of such systems is very challenging and only a few examples are reported that utilise ArMs.<sup>74</sup> Despite its attractiveness, the implementation of cascade reactions with an organometallic catalyst and an enzyme is difficult because of the mutual inactivation of both catalysts. The artificial transfer hydrogenase was assembled by combining the high affinity of streptavidin for biotin with a modified iridium-d<sup>6</sup>-piano-stool complex. This resulted in compartmentalization of the hydrogen transfer catalyst and the NADH dependent monooxygenase, allowing regeneration of the NADH (Scheme 8). If the reaction was run without streptavidin the iridium catalyst was quickly deactivated demonstrating the need for compartmentalization.<sup>75</sup>



*Scheme 8: Cascade reaction for the NADH dependent enzymatic hydroxylation of 2-hydroxybiphenyl coupled to a NADH regeneration process based on an artificial transfer hydrogenase (ATHase).<sup>10</sup>*

## 1.5 Outlook of this thesis

In this work the aim is to design and create catalyst systems that can be exploited for a range of different reactions. The envisioned target is to look at how ArM's can be utilised in lignin oxidation reactions. Lignin is a complex aromatic polymer that gives plants their structural integrity by acting as a resin. The pulp and paper industry produce large quantities of extracted lignin and only a small percentage is used commercially, and the rest is simply burnt as low value fuel. Where biorefineries focus mainly on the production of aliphatic platform chemicals from (hemi)cellulose, the degradation of lignin could provide a renewable source of low molecular weight aromatic compounds.<sup>76</sup>

Chapter 2 describes the initial stages of design and synthesis of ArMs and in particular takes a look at how they can be characterised by UV and CD spectroscopy.

Chapter 3 reports how protein engineering can be applied to increase the thermostability of a protein scaffold and how this will affect the catalytic activity when used in an ArM system. This work is applied to previous work by the Kamer group looking at rhodium-protein conjugates in hydroformylation. Protein stability is one of the major limitations for catalytic applications as illustrated by the low melting temperature ( $T_m$ ) and pronounced protein precipitation. Increasing the protein stability would allow for a wider range of possible reaction conditions and could also increase the activity of the catalyst.

Chapter 4 describes the application of created ArM systems in the oxidation of lignin model compounds. The aim was applying an ArM system instead of a small molecular catalyst to improve the selectivity of the reaction in the hope to find a method for the sustainable production of high value-added chemicals.

---

<sup>1</sup> H. Arakawa, M. Aresta, J. N. Armor, M. A. Barteau, E. J. Beckman, A. T. Bell, J. E. Bercaw, C. Creutz, E. Dinjus, D. A. Dixon, K. Domen, D. L. DuBois, J. Eckert, E. Fujita, D. H. Gibson, W. A. Goddard, D. W. Goodman, J. Keller, G. J. Kubas, H. H. Kung, J. E. Lyons, L. E. Manzer, T. J. Marks, K. Morokuma, K. M. Nicholas, R. Periana, L. Que, J. Rostrup-Nielsen, W. M. H. Sachtler, L. D. Schmidt, A. Sen, G. A. Somorjai, P. C. Stair, B. R. Stults, W. Tumas. *Chem. Rev.* **2011**, 101, 953

<sup>2</sup> T. Punniyamurthy, S. Velusamy, J. Iqbal. *Chem. Rev.* **2005**, 105, 2329

<sup>3</sup> L. Que, W. B. Tolman. *Nature*, **2008**, 455, 333

<sup>4</sup> J. C. Lewis, *ACS Catal.* **2013**, 3, 2954.

<sup>5</sup> Turner, N. J. *Chem. Rev.* **2011**, 111, 4073.

<sup>6</sup> S. B Bankar, M. V Bule, R. S Singhal, L. Ananthanarayan. *Biotechnol. Adv.* **2009**, 27, 489.

<sup>7</sup> S. K. Ma, J. Gruber, C. Davis, L. Newman, D. Gray, A. Wany, J. Grate, G. W. Huisman, R. A. Sheldon. *Green Chem.* **2010**, 12, 81

<sup>8</sup> M. Hibi, T. Kawashima, T. Kodera, S. V. Smirnov, P. M. Sokolov, M. Pugiyaama, S. Shimizu, K. Yokozeki, J. Ogawa. *Appl. Environ. Microb.* **2011**, 77, 6926

<sup>9</sup> K. R. Osorio-Lozada, *Catal. Sci. Technol.* **2012**, 2, 1524.

<sup>10</sup> M. V. Doble, A. C. C. Ward, P. J. Deuss, A. G. Jarvis. *Biororganic & Medicinal Chemistry*, **2014**, 22, 5567

<sup>11</sup> B. M. Nestl, S. C. Hammer, B. A. Nebel, B. Hauer, *Angew. Chem., Int. Ed.* **2014**, 53, 3070.

<sup>12</sup> B. Meunier, S. P. de Visser, S. Shaik, S. *Chem. Rev.* **2004**, 104, 3947.

<sup>13</sup> O. Pylypenko, I. Schlichting, I. *Annu. Rev. Biochem.* **2004**, 73, 991.

<sup>14</sup> J. Rittle, M. T. Green, *Science* **2010**, 330, 933.

- 
- <sup>15</sup> J. T. Groves, *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 3569.
- <sup>16</sup> A. W. Munro, H. M. Girvan, A. E. Mason, A. J. Dunford, *Trends. Biochem. Sci.* **2013**, *38*, 140.
- <sup>17</sup> I. G. Denisov, T. M. Makris, S. G. Sligar, I. Schlichtine, *Chem. Rev.* **2005**, *105*, 2253.
- <sup>18</sup> M. Sono, M. P. Roach, R. D. Coulter, J. H. Dawson, *Chem. Rev.* **1996**, *96*, 2841.
- <sup>19</sup> P. Meinhold, M. W. Petersm, A. Hartwick, A. R. Hernandez, F. H. Arnold, *Adv. Synth. Catal.* **2006**, *348*, 763.
- <sup>20</sup> M. Landwehr, L. Hochrein, C. R. Otey, A. Kasrayan, J. E. Backvall, F. H. Arnold, *J. Am. Chem. Soc.* **2006**, *128*, 6058
- <sup>21</sup> D. F. Munzer, P. Meinhold, M. W. Peters, S. Feichtenhofer, H. Griengl, F. H. Arnold, A. Glieder, A. de Raadt, *Chem. Commun.* **2005**, 2597.
- <sup>22</sup> T. Kubo, M. W. Peters, P. Meinhold, F. H. Arnold, *Chem. Eur. J.* **2006**, *12*, 1216.
- <sup>23</sup> M. Merckx, D. A. Kopp, M. H. Sazinsky, J. L. Blazyk, J. Muller, S. J. Lippard, *Angew. Chem., Int. Ed.* **2001**, *40*, 2782.
- <sup>24</sup> G. De Gonzola, M. D. Mihovilovic, M. W. Fraaije, *ChemBioChem* **2010**, *11*, 2208.
- <sup>25</sup> H. Leisch, K. Morley, P. C. K. Lau, P. C. K. *Chem. Rev.* **2011**, *111*, 4165.
- <sup>26</sup> K. Balke, M. Kadow, H. Mallin, S. Saß, U. T. Bornscheuer, *Org. Biomol. Chem.* **2012**, *10*, 6249.
- <sup>27</sup> E. I. Solomon, P. Chen, M. Metz, S. K. Lee, A. E. Palmer, *Angew. Chem. Int. Edn Engl.* **2001**, *40*, 4570
- <sup>28</sup> P. Giardina, V. Faraco, C. Pezzella, A. Piscitielli, S. Vanhulle, G. Sannia, *Cell. Mol. Life. Sci.* **2010**, *67*, 369.
- <sup>29</sup> J. L. Cole, P. A. Clark, E. I. Solomon, *J. Am. Chem. Soc.* **1990**, *112*, 9534.
- <sup>30</sup> R. Silaghi-Dumitrescu, A. Mot, *Biochemistry (Moscow)* **2012**, *77*, 1395.
- <sup>31</sup> C. Chirivi, G. Fontana, D. Monti, G. Ottolina, S. Riva, B. Danieli, *Chem. Eur. J.* **2012**, *18*, 10355.
- <sup>32</sup> F. H. Arnold. *Angew. Chem. Int. Ed.* **2018**, *57*, 4143
- <sup>33</sup> H. Yang, A. M. Swartz, H. J. Park, P. Srivastava, K. E. Guardiola, D. M. Upp, G. Lee, K. Belsare, Y. Gu, C. Zhang, R. E. Moellering, J. C. Lewis, *Nature Chemistry*, **2018**, *10*, 318
- <sup>34</sup> R. L. Baughn; O. Adalsteinsson; G. M. Whitesides, *J. Am. Chem. Soc.* **1978**, *100*, 306
- <sup>35</sup> P. J. Deuss, R. den Heeten, W. Laan, P. C. J. Kamer, *Chem. Eur. J.* **2011**, *17*, 4680.
- <sup>36</sup> M. Creus, T. R. Ward, *Org. Biomol. Chem.* **2007**, *5*, 1835.
- <sup>37</sup> R. Rosati, G. Roelfes, *ChemCatChem* **2010**, *2*, 916.
- <sup>38</sup> W. Ghattas, V. Dubosclard, A. Wick, A. Bendelac, R. Guillot, R. Ricoux, J. Mahy, *J. Am. Chem. Soc.* **2018**, *140*, 8756
- <sup>39</sup> M. Jeschek, R. Reuter, T. Heinisch, C. Trindler, J. Klehr, S. Panke, T. R. Ward. *Nature*, **2016**, *537*, 661
- <sup>40</sup> P. Dydio, H. M. Key, A. Nazarenko, J. Y. E. Rha, V. Seyedkazemi, D. S. Clark, J. F. Hartwig, *Science*, **2016**, *354*, 102
- <sup>41</sup> P. J. Deuss, R. den Heeten, W. Laan, P. C. J. Kamer, *Chem. Eur. J.* **2011**, *17*, 4680
- <sup>42</sup> F. Rosati, G. Roelfes, *ChemCatChem* **2010**, *2*, 916
- <sup>43</sup> M. Raynal, P. Ballester, A. Vidal-Ferran, P. W. N. M. van Leeuwen, *Chem. Soc. Rev.* **2014**, *43*, 1734
- <sup>44</sup> K. Yamamura; E. Kaiser. *J. Chem. Soc. Chem.* **1976**, 11
- <sup>45</sup> J. Podtetenieff, A. Taglieber, E. Bill, E. J. Reijerse, M. T. Reetz. *Angew. Chem. Int. Ed.* **2010**, *49*, 5151
- <sup>46</sup> A. Loosli. *Inorg. Chem.* **2006**, *45*, 660

- 
- <sup>47</sup> T. R. Ward. *Acc. Chem. Res.* **2011**, *44*, 47
- <sup>48</sup> M. Creus and T.R. Ward, *Org. Biomol. Chem.*, **2007**, *5*, 1835-1844.
- <sup>49</sup> C. Lo, M. R. Ringenberg, D. Gnadtt, Y. Wilson, T. R. Ward. *Chem Commun.* **2011**, *47*, 12065
- <sup>50</sup> M. Durrenberger, T. Heinisch, M. Y. Wilson, T. Rossel, E. Nogueira, L. Knorr, A. Mutschler, K. Kersten, M. Zimbron, J. Pierron, T. Schirmer, T. R. Ward. *Angew. Chemie*, **2011**, *50*, 3026
- <sup>51</sup> H. L. Levine, E. T. Kaiser. *J. Am. Chem. Soc.* **1978**, *100*, 7670
- <sup>52</sup> J. R. Carey, S. K. Ma, T. D. Pfister, D. K. Garner, H.k. Kim, J. A. Abramite, Z. Wang, Z. Gio, Y. Lu, *J. Am Chem. Soc.* **2004**, *126*, 10812
- <sup>53</sup> J.. L. Zhang, D. K. Garner, L. Liang, W. Chen, Y. Lu. *Chem. Commun. (Cambridge, U. K.)* **2008**, 1665
- <sup>54</sup> D. Hilvert, *Annu. Rev. Biochem.* **2013**, *82*, 447
- <sup>55</sup> D. Baker, *Biochem. Soc. Trans.* **2014**, *42*, 225
- <sup>56</sup> M. Faiella, C. Andreozzi, R. T. M. de Rosales, V. Pavone, O. Maglio, F. Nastri, W. F. DeGrado, A. Lombardi, *Nat. Chem. Biol.* **2009**, *5*, 882
- <sup>57</sup> M. Dürrenberger, T. R. Ward, *Curr. Opin. Chem. Biol.* **2014**, *19*, 99.
- <sup>58</sup> J. Bos, G. Roelfes, *Curr. Opin. Chem. Biol.* **2014**, *19*, 135.
- <sup>59</sup> T. Orbegozo, J. G. de Vries, W. Kroutil, W. *Eur. J. Org. Chem.* **2010**, 3445.
- <sup>60</sup> I. Matijosyte, I. W. C. E. Arends, S. de Vries, R. A. Sheldon, *J. Mol Catal. B. Enzym.* **2010**, *62*, 142.
- <sup>61</sup> C. Letondor, N. Humbert, T. R. Ward, *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 4683.
- <sup>62</sup> C. M. Thomas, C. Letondor, N. Humbert, T. R. Ward, *J. Organomet. Chem.* **2005**, *690*, 4488.
- <sup>63</sup> N. Fujieda, A. Hasegawa, K. Ishihama, S. Itoh, *Chem. Asian J.* **2012**, *7*, 1203
- <sup>64</sup> A. Pordea, M. Creus, J. Panek, C. Duboc, D. Mathis, M. Novic, T. R. Ward, *J. Am. Chem. Soc.* **2008**, *130*, 8085.
- <sup>65</sup> S. Nimri, E. Keinan, *J. Am. Chem. Soc.* **1999**, *121*, 8978.
- <sup>66</sup> R. Ricoux, E. Lukowska, F. Pezzotti, J. P Mahy, *Eur. J. Biochem.* **2004**, *271*, 1277
- <sup>67</sup> R. Q. Raffy, R. Ricoux, E. Sansiaume, S. Pethe, J. P Mahy, *J. Mol. Catal. A Chem.* **2010**, *317*, 19.
- <sup>68</sup> J. R. Carey, S. K. Ma, T. D. Pfister, D. K. Garner, H. K. Kim, J. A. Abramite, Z. Wang, Z.. Guo, Y. Lu, *J. Am. Chem. Soc.* **2004**, *126*, 10812.
- <sup>69</sup> J. Zhang, D. K. Garner, L. Liang, Q. Chen, Y. Lu, *Chem. Commun.* **2008**, 1665.
- <sup>70</sup> A. Mahammed, Z. Gross. *J. Am. Chem. Soc.* **2005**, *127*, 2883
- <sup>71</sup> R. Ricoux, R. Dubuc, C. Dupont, J. D. Marechal, A. Martin, M. Sellier, J. P. Mahy, *Bioconjugate Chem.* **2008**, *19*, 899.
- <sup>72</sup> M. Allard, C. Dupont, V. Muñoz Robles, N. Doucet, A. Lledós, J. D. Maréchal, A. Urvoas, J. P Mahy, R. Ricoux, *ChemBioChem* **2012**, *13*, 240.
- <sup>73</sup> C. A. Denard, J. F. Hartwig, H. Zhao, *ACS Catal.* **2013**, *3*, 2856.
- <sup>74</sup> K. Engström, E. V. Johnston, O. Verho, K. P. J. Gustafson, M. Shakeri, C. W. Tai, J. E. Bäckvall, *Angew. Chem., Int. Ed.* **2013**, *52*, 14006.

---

<sup>75</sup> V. Köhler, Y. M. Wilson, M. Dürrenberger, D. Ghislieri, E. Churakova, T. Quinto, L. Knörr, D. Häussinger, F. Hollmann, N. J. Turner, T. R. Ward, *Nat. Chem.* **2013**, *5*, 93.

<sup>76</sup> J. Zakzeski, P.C. A. Bruijninx, A. L Jongerius, B. M. Weckhuysen. *Chem. Rev.* **2010**, *110*, 3552

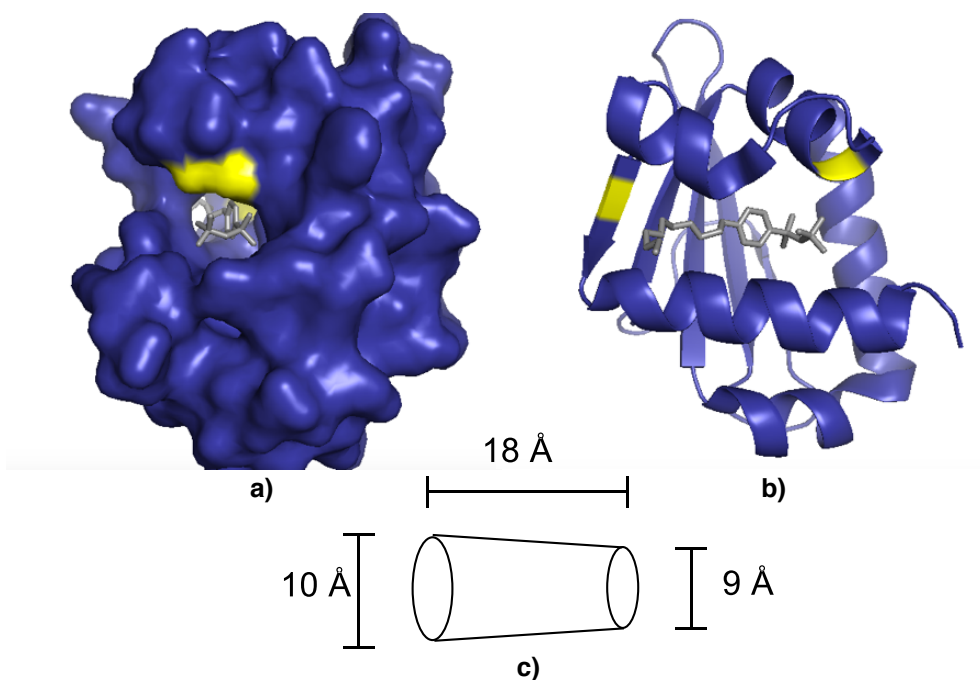
## Chapter 2 – Design and Characterisation of Artificial Metalloenzymes

### 2.1 Artificial Metalloenzyme Design

The first step of artificial metalloenzyme design is the choice of protein to be used as a scaffold. A set of demands is required to be able to select a suitable protein host from the millions of available proteins. Protein hosts are selected based on desired characteristics such as substrate binding, stability and shape of the binding site. One vital component when selecting a protein scaffold is that a large amount of structural information must be known about the protein. Regardless of recent advances in the field of bioengineering, it remains difficult to predict the tertiary structure of a protein, solely based on the amino acid sequence and so de-novo design of an appropriate structure is troublesome.<sup>1</sup> Hence, the protein has to be structurally characterised by either X-ray crystallography or solution NMR. The next important factor is that the protein requires simple production and purification protocols, preferentially resulting in a high yield. Finally, stability of the protein is important. Conditions used in catalysis can vary in temperature; pH, solvents and reagents so using a protein with high stability will be advantageous.

Successful results have been obtained within the Kamer group using human Steroid Carrier Protein SL (SCP-2L),<sup>2,3,4,5</sup> this is one of the three domains of the Human Multifunctional Enzyme Type 2 (MFE-2). The protein MFE-2 is one class of proteins in mammals catalyzing  $\beta$ -oxidation. MFE-2 consists of three domains, the C-terminal domain shows similarities to sterol carrier protein type 2 (SCP-2) and is therefore referred to as a sterol carrier protein 2-like (SCP-2L).<sup>6</sup> The crystal structure of the SCP-2L protein (PDB 1IKT) displays a large apolar tunnel, which in this crystal structure contains a Triton X-100 molecule (Figure 1). SCP-2L is part of the superfamily of SCP-like, this is part of a class of proteins which were found to be involved in the biosynthesis of cholesterol. Their secondary structure is of particular interest as they often contain a hydrophobic binding pocket. The apolar tunnel seen within SCP-2L is of interest within ArM work because it can be used to direct substrates like hydrocarbons into the tunnel and provide more opportunity for substrate coordination to the protein's secondary structure. This would allow the ArM's to tailor or increase selectivity of a reaction.



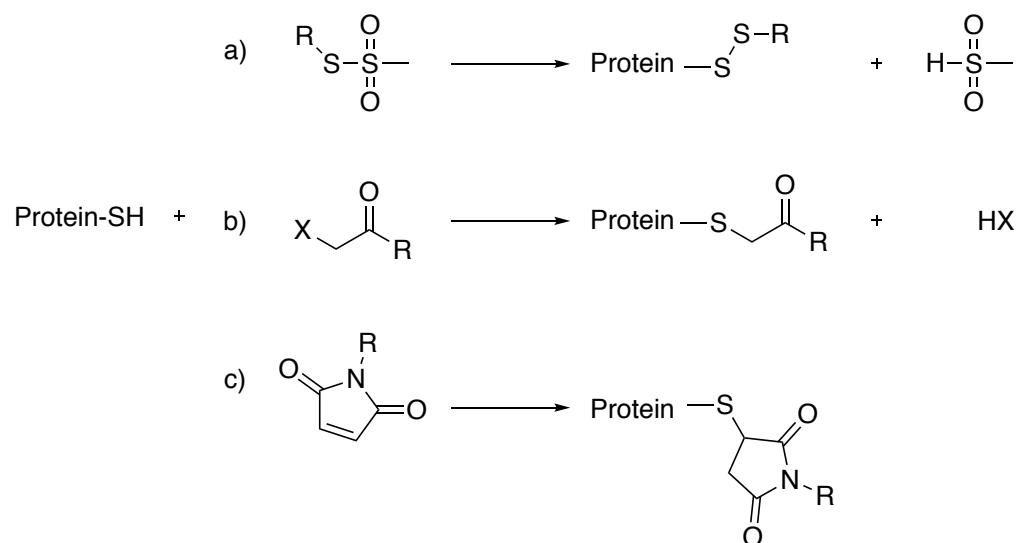


*Figure 1 – Ligand-modified Sterol Carrier Protein Type 2 like Domain (SCP-2L) PDB 1IKT forming a complex with Triton X-100. V83 and A100 are highlighted in yellow to show the placement of Cysteine when mutated. a) Space filled model to represent the apolar tunnel, b) Cartoon representation with Triton X-100 shown in grey c) Tunnel Dimensions*

In order to modify the selected protein template with our transition metal catalyst system, bioconjugation is carried out using a unique cysteine. Cysteine is an amino acid with a highly nucleophilic thiol function that is frequently exploited for site-selective bioconjugation.<sup>7</sup> The presence of a unique cysteine is utilized to incorporate one transition metal binding ligand into the protein, at a unique position to benefit from the selective binding properties of the protein. In figure 1, the two amino acids V83 and A100 are highlighted as this is where unique cysteine residues have been previously introduced into the SCP-2L protein.<sup>3</sup> Each mutant is introduced at either end of the large hydrophobic tunnel. Previous work also showed both mutants can be expressed in high yield and purification is straightforward.<sup>2</sup>

The method of covalently attaching the designed maleimide-containing cofactor is achieved by utilizing the cysteine residue as a target (Scheme 1), this residue is a useful target because it has a high reactivity, due to its thiol group, but it also has a low abundance within a protein. The bioconjugation can be carried out using activated sulfides,  $\alpha$ - halogen

carbonyls and maleimide functionalities. In this work we have concentrated on using the maleimide functionality.



*Scheme 1 – Selective Cysteine Modification using a) activated disulfides b)  $\alpha$ -Halogenocarbonyls c) Maleimide Functionality*

## 2.2 Cofactor Design and Synthesis

### 2.2.1 Nitrogen containing ligand library

The library of nitrogen containing cofactors depicted in Figure 2 was designed based on derivatives of known ligands, which have been successfully used in oxidation reactions with a variety of metals and substrates.<sup>8,9,10,11</sup> This collection of cofactors has an array of useful properties including varied flexibility/ varied chain length from maleimide to metal ion and they also vary between bidentate and tridentate.

Distefano *et al.* demonstrated the successful coupling of a phenanthroline moiety selectively to a unique cysteine residue within a protein using 5-iodoacetamide-1,10-phenanthroline.<sup>12</sup> Many other nitrogen donor ligands have been introduced via a site selective linkage to cysteine in a similar way.<sup>3, 13</sup>

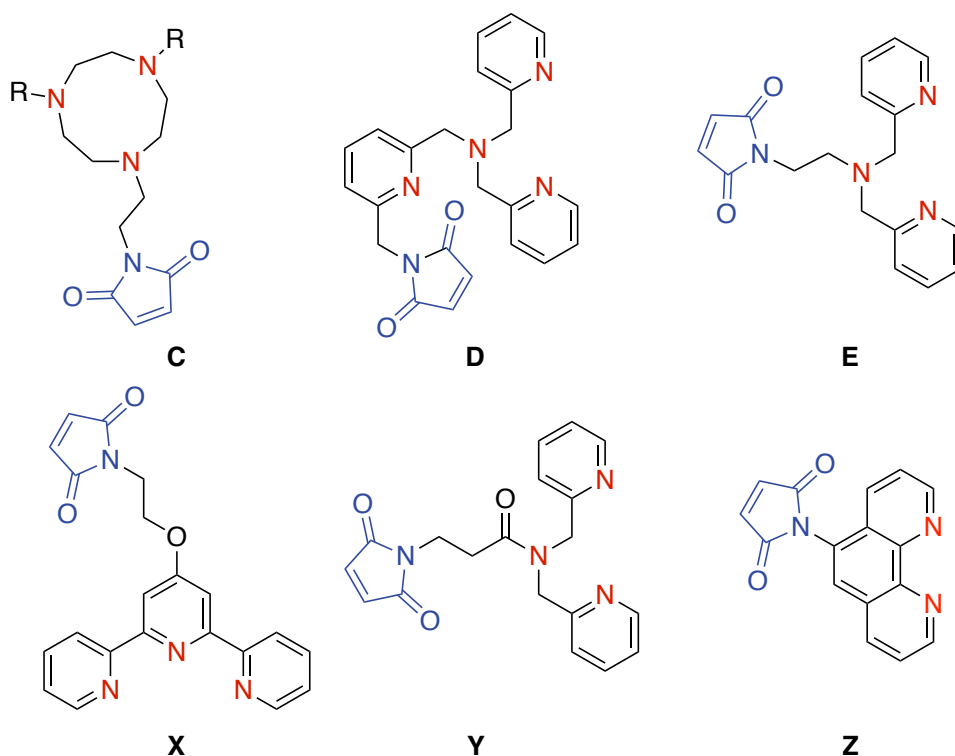
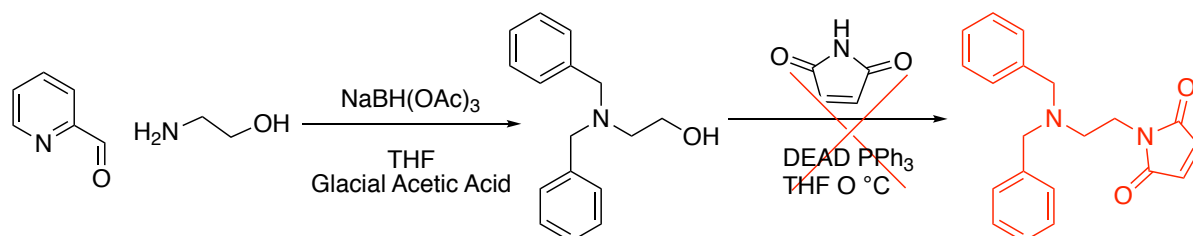


Figure 2 - The library of nitrogen containing cofactors

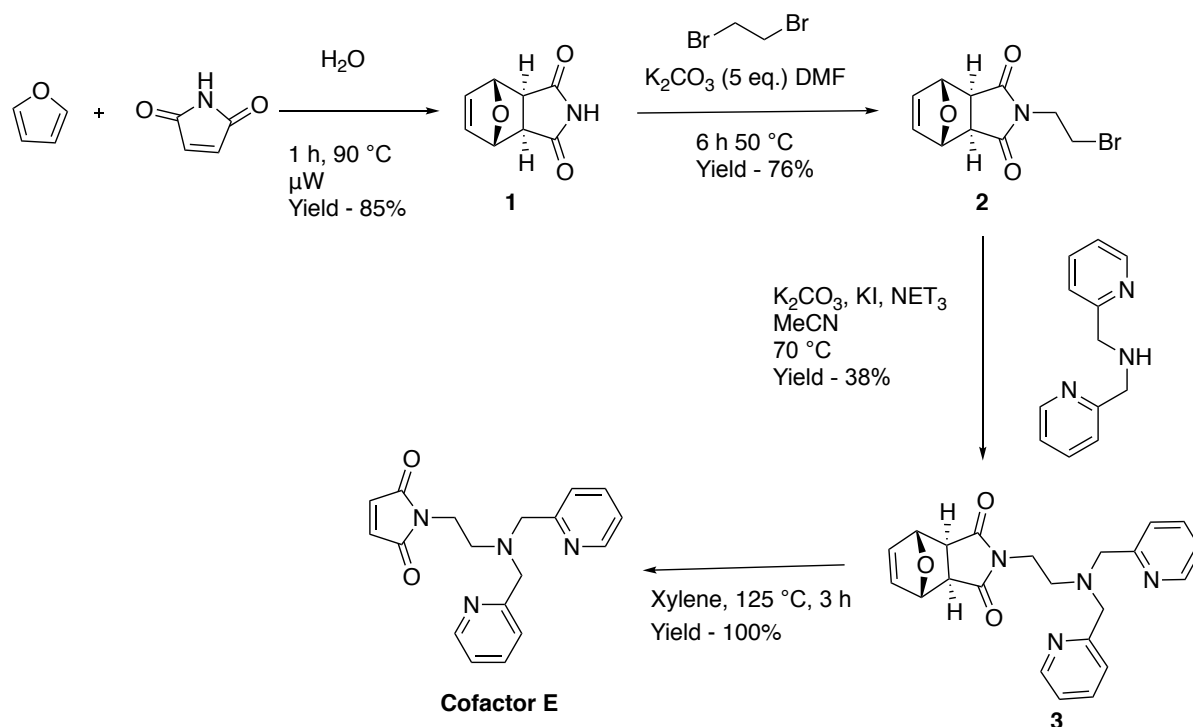
### 2.2.2 Synthetic Route to Cofactor E

The synthesis of this ligand has been previously reported in literature<sup>14</sup> using a Mitsunobu reaction between maleimide and bis(picoly)aminoethanol (Scheme 2). However, even with a large number of attempts varying the conditions, the synthesis could not be reproduced.



Scheme 2 – Synthesis of Cofactor E via the Mitsunobu Reaction

In order to protect the maleimide functionality during reactions, it was reacted with furan under microwave conditions to create a stable Diels-Alder adduct. This Diels-Alder maleimide adduct (**1**) was isolated as the pure *exo* adduct as a protected Michael acceptor. This adduct was then applied to S<sub>N</sub>2 type reactions to yield the protected cofactor (**3**), which was then subjected to a retro Diels Alder to produce the final product (Cofactor E) (Scheme 3). The overall yield is 25%, which is low mainly due to the troublesome step of addition of dipicolylamine.



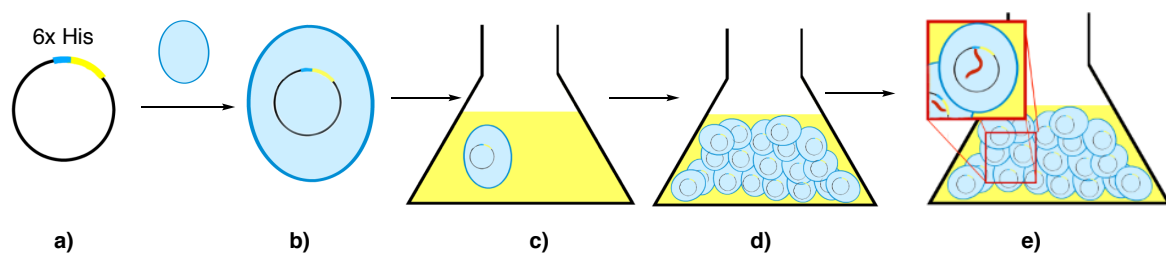
Scheme 3 – Synthetic Route to Cofactor E

The other cofactors within the designed library were made by Dr Amanda Jarvis, Andrew Ward and Dr Peter Deuss from the Kamer Group.

## 2.4 Protein expression and purification

To acquire our desired protein, the gene of interest is cloned into a plasmid containing a cleavable histidine tag using TEV protease, and a specific antibiotic resistance. This plasmid is then transformed into a host cell; in this work *Escherichia coli* (*E. coli*) Rosetta DE3 cells were used. This is the most common bacterial strain used to obtain protein in laboratory

conditions due to its fast growth rate, its relatively simple growth requirements, the simplicity of growing the organism, and the avirulent nature of laboratory strains. A number of cells harbouring the desired plasmid are used to inoculate a large volume of media. This media is a broth that contains sugars to feed the bacteria and the antibiotic the plasmid provides resistance to. Any cells that don't contain the same antibiotic resistance will die as a consequence. Once the cells have grown to an optical density of 0.6 (600nm), the expression of the desired gene encoding the production of the desired protein is triggered. Protein expression is induced with isopropyl  $\beta$ -D-1- thiogalactopyranoside (IPTG); a molecular analogue of allolactose, a lactose metabolite that prompts transcription of the lac operon. IPTG binds to the lac repressor, which is then released from the lac operator in an allosteric manner, thereby allowing the transcription of genes in the lac operon. The lac operon is a working unit of genomic DNA under the control of a single regulatory signal or promoter, which is required for transport and metabolism of lactose in *E. coli*. (Figure 3).



*Figure 3 – Diagram representation of protein expression procedure. a) Plasmid containing desired gene and cleavable His Tag, b) Transformation of vector into bacterial cell, c) Inoculation of bacterial cell into media, d) Bacterial cells grow to a high optical density e) Protein expression induced by addition of IPTG*

Once protein is expressed, the cells were harvested by centrifugation, and the cells are disrupted by sonication. To aid cell lysis, lysozyme, deoxyribonuclease (DNase), and the protease inhibitor benzamidine were added to samples. Once the cells had been lysed, the soluble protein was isolated by centrifugation, as any insoluble cell debris would form a pellet and the supernatant could then be extracted and purified. The advantage of using a protein that contains a histidine tag is that this provides a very simple and effective purification method. The binding affinity of the six tandem histidine residues towards a nickel column can be utilised (Figure 4). The supernatant is applied to a Ni-column, only

proteins with a high nickel affinity will be adsorbed and all other proteins will be washed out. Using an appropriate buffer containing a large excess of imidazole, the desired protein can then be eluted. The histidine tag can then be cleaved using TEV protease; a highly specific protease found in Tobacco Etch Virus (TEV). The residues ENLYFQG, introduced next to the His-tag, form the recognition sequence used and cleavage occurs between the glutamine and glycine residues.

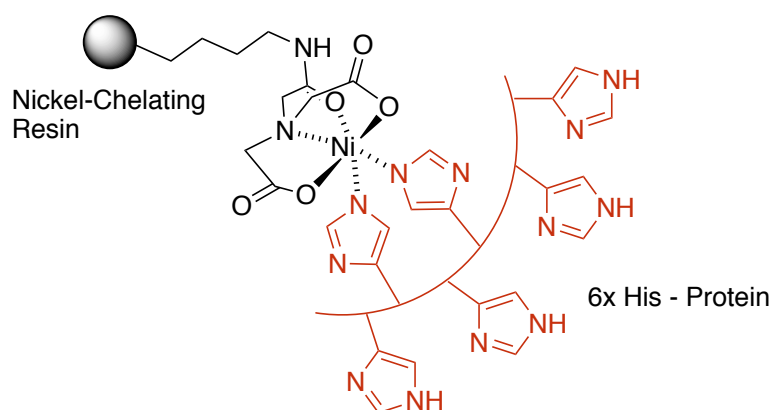


Figure 4 – His tagged Protein binding to a Nickel affinity column

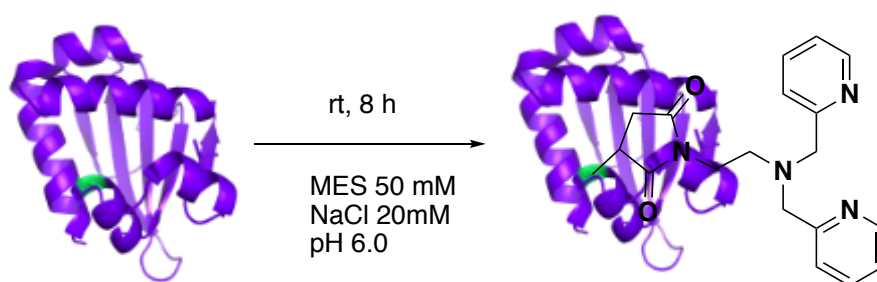
The initial runs of protein expression and purification produced very low yields, approx. 2mg/ L as determined by Bradford assay. This is obviously not suitable to try and apply in catalysis work and thus required further optimization. The starting point was to increase the levels of DNase used, as an efficient lysis procedure is essential for high protein yields. Another step of the procedure that was looked at was the histidine tag cleavage, this reaction occurs by stirring the protein solution with TEV protease. Both SCP-2L mutants are prone to precipitation, so to try and increase the proteins stability in this reaction the protein used is kept in very dilute solution. Also, the reaction is stirred on the lowest possible setting to try not to agitate the protein solution. These alterations proved to be successful as both SCP-2L A100C and V83C were successfully produced and optimised to yield on average 60mg/ml and 40 mg/ml respectively.

## 2.5 Protein Scaffold Modification

### 2.5.1 General Methodology

The scheme for modification of both SCP-2L A100C and V83C is carried out by covalent modification as previously described and is depicted in Scheme 4. The reactive thiol group

of the unique cysteine residue within the protein scaffold, reacts with the maleimide functionality in the cofactor to modify the protein. Most ligands are not soluble in the aqueous conditions of the protein solution so these are added in minimal (50 $\mu$ l) amount of organic solvent. After modification, extensive buffer exchange using centrifugal concentrators is carried out to ensure there is no unbound ligand in the protein solution. One possible disadvantage of this modification technique is that a chiral carbon is formed during the reaction, which gives the possibility of the formation of different stereoisomers. This could be undesirable if the target application is asymmetric catalysis. However, it is possible that the chiral protein environment induces a preference for one of the stereoisomers.

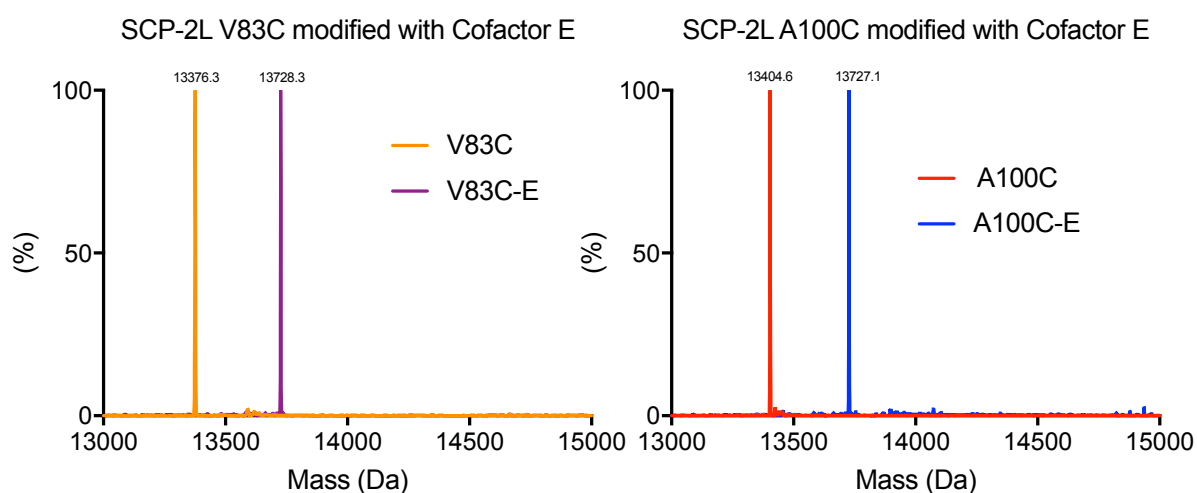


*Scheme 4 –Modification of SCP–2L V83C with Cofactor E*

In order to form the ArM a complex must be formed, this is carried out by the addition of a metal salt to the nitrogen donor ligand modified protein. Once analysed, any excess metal ion must be removed to ensure that any catalysis carried out with the ArM is not being affected by any metal ions in solution.

### 2.5.2 LCMS analysis

For all of the reported nitrogen containing ligands the modification was seen to proceed with 100% conversion to afford the modified protein. To confirm this, LCMS analysis was carried out to see the increase in mass, an example of this is shown in Figure 5 using SCP-2L modified with Cofactor E.



SCP-2L A100C-E:

Calculated mass: 13726.7 Da, Observed: 13727.1 Da.

SCP-2L V83C-E:

Calculated mass: 13698.7 Da. Observed: 13699.8 Da.

*Figure 5 – Deconvoluted mass spectrum showing the increase in signal upon modification for SCP-2L A100C-E*

*The Abundance shown as a (%) on the y axis.*

Analysis of the ArM complex with a metal ion is very difficult by LCMS, as it is unknown if a signal can't be seen of the complex whether this means the complexation was unsuccessful, or the mass spectrometry conditions caused removal of the metal ion. In one circumstance I successfully managed to see SCP-2L A100C- Cofactor Z-  $\text{Cu}^{2+}$  (Figure 6), but analysis by UV studies were carried out to get a better understanding of the ArM complexes.

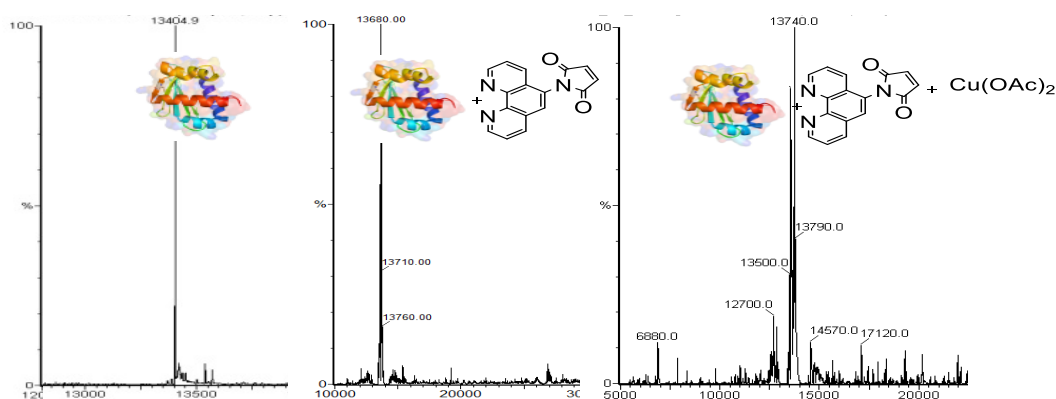




Figure 6 – Mass Spectrum of SCP-2L A100C, modified with Cofactor Z, and then complexed with  $Cu^{2+}$

### 2.5.3 CD analysis

CD spectroscopy measure differences in the absorption of left-handed vs right-handed polarized light, which occurs due to structural asymmetry. This tool is often used for protein structural studies, as it can give an indication how the protein is folded and can provide information about its secondary and tertiary structure.<sup>15,16,17</sup> The CD spectrum in the near-UV region (250-350nm) can provide some information about the proteins tertiary structure. The chromophores that are absorbed at these wavelengths are any aromatic residues and disulfide bonds.<sup>18</sup>

The near UV CD spectra of SCP-2L V83C and A100C are compared to their spectrums when modified with Cofactor D (Figure 7). For both mutants a change in the near UV CD-spectrum is seen between the modified and unmodified proteins. It can be assumed from this that by modifying the protein with the cofactor the protein folding has slightly changed. However, the significant change in signal between 250 and 270nm is due to the introduction of pyridine rings into the protein.

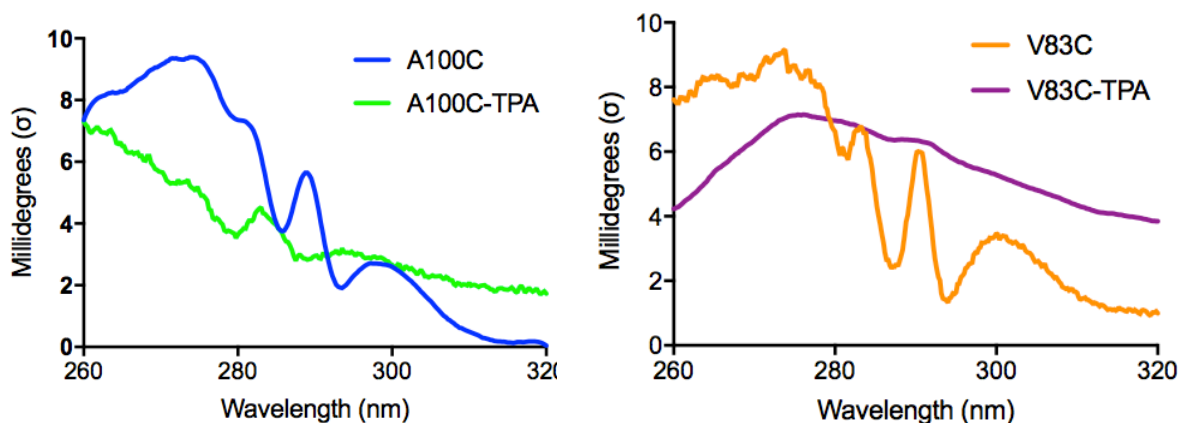


Figure 7 – CD Analysis of SCP-2L A100C and SCP-2L V83C and the modified proteins with Cofactor D

Interestingly, when the near UV CD spectrum of SCP-2L V83C was compared with its hybrid containing cofactor E (Figure 8), it showed very little changes in the folding pattern and the significant changes appear from the introduction of aromatic residues. This is interesting as it suggests that Cofactor D and E are in different chiral environments when bound to SCP-2L mutants.

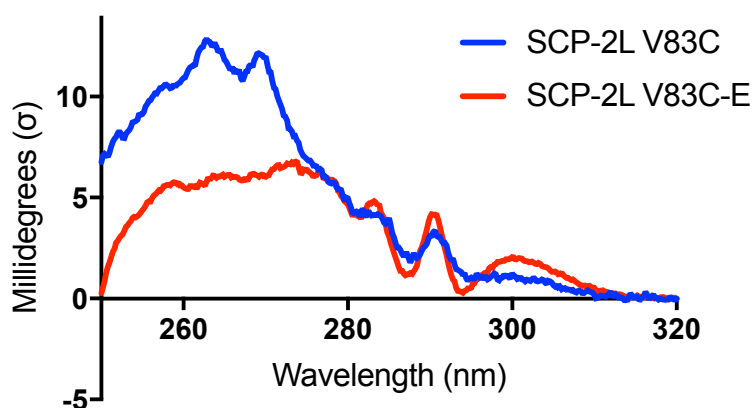


Figure 8 – Near UV spectrum of SCP-2L V83C and the modified hybrid with Cofactor E.

## 2.6 Determining concentration of Artificial metalloenzymes

### 2.6.1 Concentration calculation using Bradford reagents

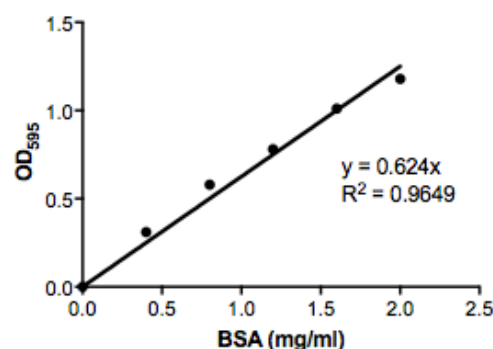
Common methods to determine protein concentration are using the absorbance at A280 from aromatic signals and also using Bradford Reagent.<sup>19</sup> Quantifying protein concentration by directly measuring A280 is a fast and convenient method as no additional reagents or incubation periods are necessary. Using the proteins extinction coefficient and the A280 it is possible to calculate a reliable value for the concentration using Beer's Law.

The alternative method is a Bradford protein assay, which is used to measure the concentration of total protein within a sample. The principle is that the binding of protein molecules to Coomassie dye under acidic conditions results in a colour change from brown to blue. An estimation is performed using a reference protein standard, generally bovine serum albumin (BSA) is the commonly accepted reference. However, if the reference

protein has very different properties to the protein in question then this method becomes unreliable.

The modification of protein introduces new aromaticity into the structure and therefore, analysis was required to determine if normal methods of calculating protein concentration were reliable. In order to obtain reliable catalytic data using ArMs it is critical that the method used to determine the concentration of the ArM is accurate. Experimental analysis was carried out using SCP-2L A100C modified with a TPA cofactor (Cofactor D). The initial task was to compare these two methods, to see if simply measuring concentration at A280 using a nanodrop and using a Bradford protein assay gave comparable results. (Figure 9) Concentration using a nanodrop was calculated using the calculated extinction coefficient for SCP-2L A100C based on the protein sequence using ExPASy ProtParam tool.

	SCP-2L A100C TPA		
	OD <sub>595</sub>	OD <sub>595</sub> -Blank	Calculated Concentration (mg/ml)
Repeat 1	1.12	0.84	1.68
Repeat 2	1.13	0.85	1.71
Repeat 3	1.13	0.84	1.69



Concentration using NanoDrop

Average from 4 repeats = 1.76 mg/ml

*Figure 9 – Comparison of Protein Concentration measurements of SCP-2L A100C TPA using A280 with a nano drop and using a Bradford protein assay*

When using the ArM in catalysis it is crucial that we know the accurate concentration of the protein so further investigation was required to determine which method was giving the more reliable value.

## 2.6.2 QAA analysis

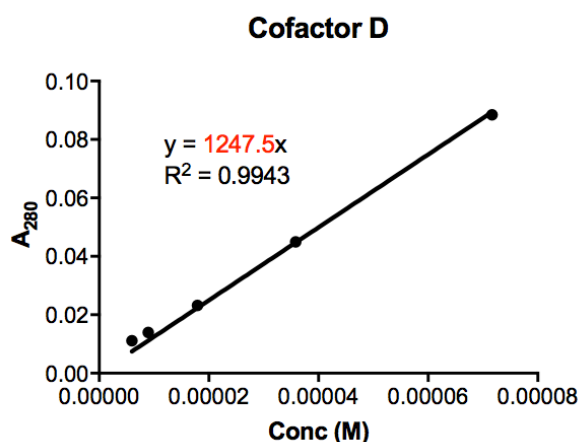
Quantitative Amino Acid Analysis (QAA) is known as the gold standard analysis for accurate quantitation of proteins and peptides.<sup>20</sup> It is a fundamental biochemical technique used for the determination of the amino acid composition of proteins, peptides and samples that contain primary or secondary amino groups within their molecular structure. We submitted the protein sample SCP-2L A100C TPA for QAA analysis at PNAC Facility, University of Cambridge. The results for the same sample previously studied showed an average from two repeats of 1.47mg/ml.

The QAA results showed that using Bradford technique is not a suitable method to determine concentration of modified proteins. We therefore, calculated the extinction coefficient of the sample used for QAA analysis using the obtained concentration and compared this to calculated extinction coefficients for the modified protein. The calculated extinction coefficient was calculated by combining the theoretical extinction coefficient for the protein based on its sequence (Calculated using Protparam), with the extinction coefficients of the synthesised cofactors.

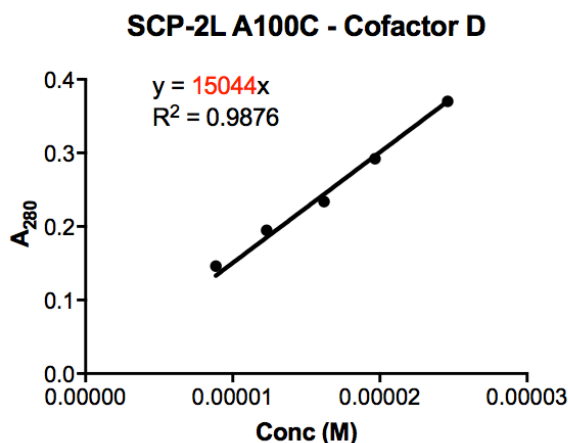
Computed Extinction coefficient at  $A_{280}$  for SCP-2L is  $13980 \text{ M}^{-1} \text{ cm}^{-1}$

Experimentally determined Extinction coefficient at  $A_{280}$  for TPA is  $1247.5 \text{ M}^{-1} \text{ cm}^{-1}$

SCP-2L A100C TPA Extinction coefficient at  $A_{280}$  by addition is  **$15227.5 \text{ M}^{-1} \text{ cm}^{-1}$**



Using the SCP-2L A100C TPA the extinction coefficient at  $A_{280}$  was experimentally calculated as  $15044 \text{ M}^{-1} \text{ cm}^{-1}$



Nanodrop value readjusted with experimental extinction Coefficient ( $15044 \text{ M}^{-1} \text{ cm}^{-1}$ ) is **1.56 mg**

Nanodrop value readjusted with calculated extinction Coefficient ( $15227 \text{ M}^{-1} \text{ cm}^{-1}$ ) is **1.54mg**

The calculated and experimental extinction coefficients are within error, and both of them are within 5% of the concentration obtained by QAA analysis. Therefore, for all the work reported the concentrations were determined by  $A_{280}$  using the calculated extinction coefficients given below. (Table 1) As cofactor C does not absorb significantly at 280 nm, the calculated extinction coefficient of the protein was used. This experiment was the repeated for all of the cofactors moving forwards.

Cofactor	Synthesised Cofactor	Modified Protein
	Extinction Coefficient ( $\text{M}^{-1} \text{ cm}^{-1}$ )	Extinction Coefficient ( $\text{M}^{-1} \text{ cm}^{-1}$ )
X	2769	16749
Y	1043	15023
Z	16744	30724
D	1247.5	15227
E	384	14364

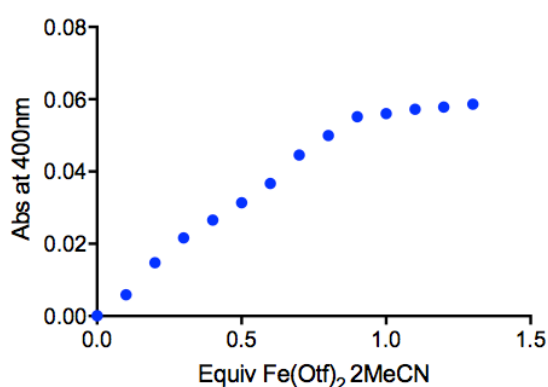
*Table 1 – Extinction Coefficient values of modified proteins*

## 2.7 Metal Binding Studies of Artificial metalloenzymes

UV studies have been widely used for protein analysis and also within artificial metalloenzyme work.<sup>2, 21, 22</sup> To determine if the metal binds to the modified metalloenzyme in a 1:1 ratio as designed, UV analysis was used. Equivalents of the metal salt are added gradually, and the UV spectrum is recorded at each point to determine the saturation point.

The saturation plot for both SCP-2L A100C and V83C hybrids with Cofactor D show that the  $\text{Fe}(\text{Otf})_2 \cdot 2\text{MeCN}$  metal ion appears to bind in a 1:1 ratio with the protein (Figure 10). After 1.0 equivalent of metal ion is added to the ArM both plots level out, showing it has reached its binding capacity.

**SCP-2L A100C -Cofactor D**



**SCP-2L V83C -Cofactor D**

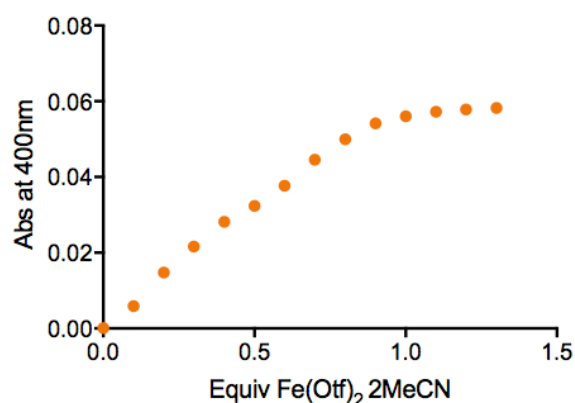
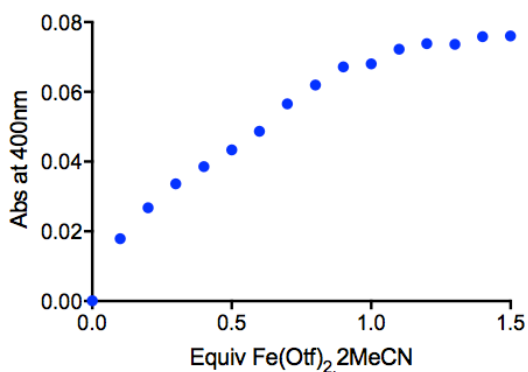


Figure 10– Saturation plot of SCP-2L A100C and V83C modified with cofactor D with  $\text{Fe}(\text{Otf})_2 \cdot 2\text{MeCN}$

The saturation plot for both SCP-2L A100C and V83C hybrids with Cofactor E also indicated a 1:1 binding with  $\text{Fe}(\text{Otf})_2 \cdot 2\text{MeCN}$ . (Figure 11) However, in this instance it appears the absorbance at 400nm continues to increase till 1.1 equivalents of metal were added before it reached full binding capacity. This could also indicate a relatively low binding constant.

### SCP-2L A100C -Cofactor E



### SCP-2L V83C -Cofactor E

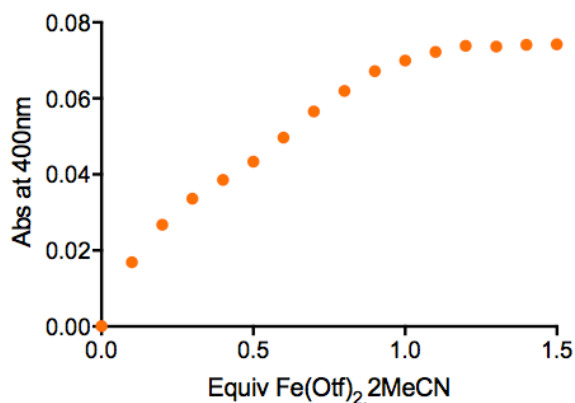
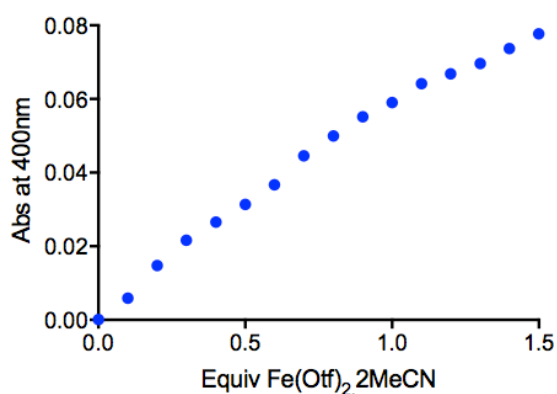


Figure 11 – Saturation plot of SCP-2L A100C and V83C modified with cofactor E with  $Fe(Otf)_2.2MeCN$

The binding capacity of  $Fe(Otf)_2.2MeCN$  with both hydrids of SCP-2L A100C and V83C with Cofactor Z is significantly greater than 1:1, the absorbance continued to increase past 1.5 equivalentents (Figure 12). This was also pushed all the way up 3 equivalentents and still no saturation was seen, this could be due to a very low binding constant.

### SCP-2L A100C -Cofactor Z



### SCP-2L V83C -Cofactor Z

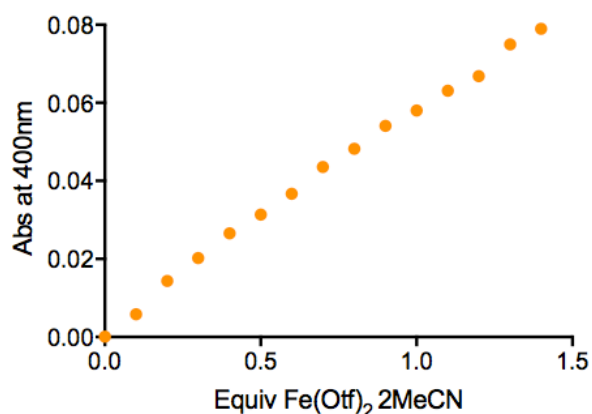
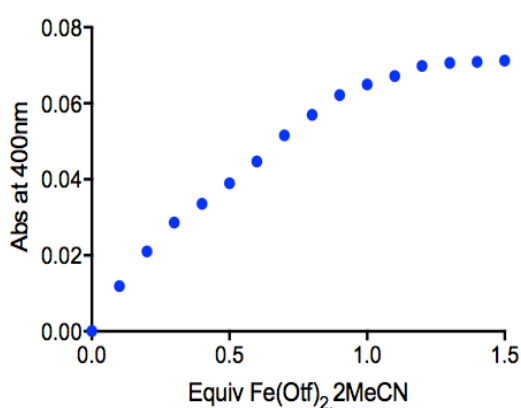


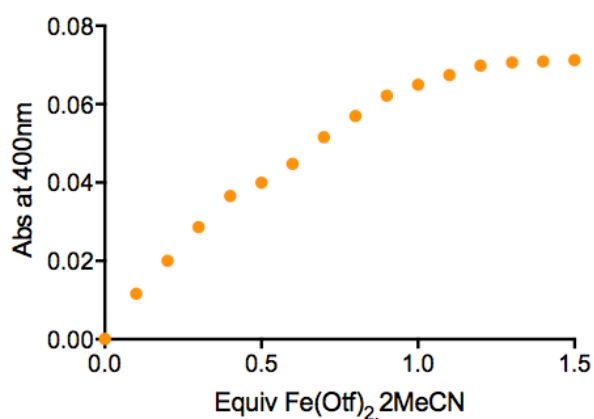
Figure 12 – Saturation plot of SCP-2L A100C and V83C modified with cofactor Z with  $Fe(Otf)_2.2MeCN$

Both Cofactor X and Y hybrids showed similar results when bound with  $\text{Fe}(\text{Otf})_2 \cdot 2\text{MeCN}$  (Figure 13). It can be seen in both circumstances that the ArM does reach binding capacity. For Cofactor X, it appears to take 1.2 equivalents of metal ion until the hybrid protein has reached saturation point. For Cofactor Y, it has increased further and requires 1.4 equivalents of metal ion until it reaches capacity. This could also be a result of a poor binding constant, resulting in the requirement of excess of metal to drive the equilibrium to completion.

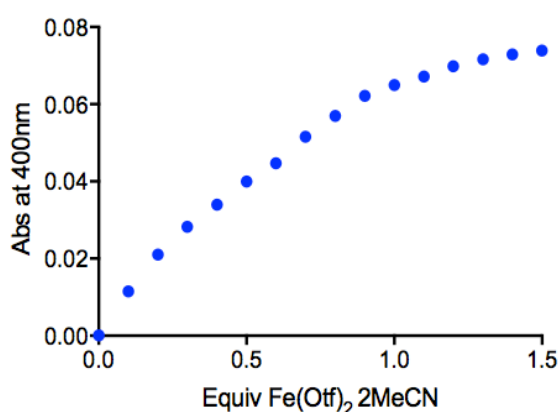
**SCP-2L A100C -Cofactor X**



**SCP-2L V83C -Cofactor X**



**SCP-2L A100C -Cofactor Y**



**SCP-2L V83C -Cofactor Y**

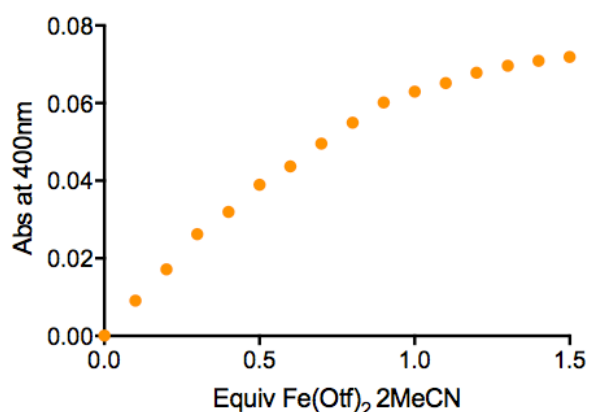


Figure 13 – Saturation plot of SCP-2L A100C and V83C modified with cofactor X and Y with  $\text{Fe}(\text{Otf})_2 \cdot 2\text{MeCN}$



## 2.8 Conclusions and Further Work

In this chapter the successful synthesis of an array of ArMs, based upon SCP-2L protein scaffold and a nitrogen containing ligand library, has been shown. It is of great importance when using the ArM in catalysis to know the exact modified protein concentration. Using QAA analysis and UV studies to experimentally calculate extinction coefficients, an accurate method was successfully determined.

LCMS studies of ArM systems showed most metals are only weakly coordinated to the modified protein and are destroyed during LCMS conditions. UV and CD analysis provided insight into the protein scaffold structure and how it is affected by modification with a synthetic cofactor. By creating a saturation plot using UV, the binding nature of the metal to protein was studied and identified which ArM systems showed a 1:1 binding as designed.

Analysis carried out by LCMS, UV and CD can give us a good understanding of the ArM, but in order to have greater knowledge a crystal structure of the modified protein and/or complex should be obtained. While this was attempted no success was seen, so further optimisation of crystallisation methods should be studied to gain structural information about the ArM. As an alternative the protein scaffold could be isotopically labelled by cultivation in a medium containing  $N^{15}$  as the only source of nitrogen. NMR methods could then be used to gain structural information about the modified protein. If structural characterisation of the modified protein was obtained, then it could be used to model future ArMs. It would provide an insight into the structural properties next to the introduced cofactor and fine-tune this active site, and the coordination site of the metal.

## 2.9 Experimental

### 2.9.1 General Remarks

All reagents and solvents were obtained commercially and used without further purification. *n*BuLi was titrated against 2-3 dimethoxybenzyl alcohol prior to use to determine the correct concentration. Thin layer chromatography was performed on silica plates (TLC silica gel 60 F254 Merck) and visualised with UV light (254 nm). Column chromatography was performed using silica gel 60 particle size 40-63  $\mu\text{m}$  from Merck. NMR spectra were taken at room temperature with Bruker Avance 300, 400 or 500 NMR spectrometers. Chemical shifts ( $\delta$ ) are given (in ppm) and processed using MestReNova software. The residual undeuterated solvent of the deuterated solvent ( $\text{CHCl}_3$   $\delta = 7.26$  and  $77.16$ ,  $^1\text{H}$  and  $^{13}\text{C}$  respectively) was used to reference the chemical shifts. The abbreviations for the multiplicity of the proton, carbon and fluorine signals are as follows: s singlet, d doublet, dd doublet of doublets, ddd doublet of doublet of doublets, t triplet, dt double triplets, q quartet, m multiplet, br s broad singlet.

Circular dichroism spectra were obtained using a Biologic MOS-500 spectrometer. A Xe lamp was used, and the near UV spectra were taken from 190 – 260 nm using a 10 mm pathlength cell.

Protein images were processed by Pymol 0.99rc6.

LC-MS used for analysis of protein was performed on a Waters Alliance HT 2795, equipped with a Micromass LCT-TOF mass spectrometer using positive electrospray ionisation and applying a Waters MASS PREP On-line desalting 2.1 x 10mm cartridge. ESI results were analysed by Mass Lynx V 4.0 and its MaxEnt algorithm.

The following MS-tune settings were used for both LC-MS and direct injection: Capillary = 3000 V, Sample Cone = 50 V, Extraction Cone = 3 V, RF Lens = 350, Desolvation Temperature = 300 °C, Source Temperature = 100 °C, RF DC Offset 1 = 10, RF DC Offset 2 = 8, Aperture = 10, Acceleration = 200, Focus = 0, Steering = 0.0, MCP Detector = 2850, Manual

Pusher = Yes, Max. Flight time = 70, Ion Energy = 32, Tube Lens = 0, Grid 2 = 60, TOF flight tube = 4600, Reflection = 1788.

The gradient used for LCMS protein analysis is shown in Table 2 below.

Time (min)	Solvent A (%)	Solvent B (%)	Flow (ml/min)
0.0	2	98	0.05
0.5	2	98	0.05
2.5	98	2	0.05
4.5	98	2	0.05
4.6	2	98	0.05
12	2	98	0.05

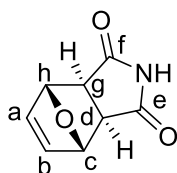
Table 2 – Gradient used for LCMS analysis. Solvent A = 1% formic acid in acetonitrile, Solvent B = 1% formic acid in water.

Buffer exchange for protein was carried out using centrifugal concentrators with a molecular weight cut-off of 10 kDa.

The pH of buffers were adjusted using aqueous solutions of sodium hydroxide or hydrochloric acid.

### 2.9.2 Synthesis of Cofactor E

Step 1: Synthesis of (3aR,4S,7R,7aS)-3a,4,7,7a-tetrahydro-1H-4,7-epoxyisoindole-1,3(2H)-dione (i) was carried out according to literature.<sup>23</sup>

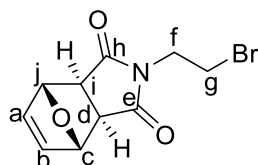


Mp-160-161 °C

A mixture of maleimide (0.855 g, 1 equiv., 8.8 mmol), furan (1.1 mL, 1.7 equiv., 15 mmol) and H<sub>2</sub>O (10 mL) was heated in a 35 mL microwave reaction vessel to 90 °C (100 W) for one hour. The flask was left to cool and a white crystalline solid was precipitated and filtered, then washed with diethyl ether and dried under vacuum to afford the product as white crystalline solid. (4.83 g, 84%) yield.

$^1\text{H}$  NMR (300 MHz,  $\text{CDCl}_3$ )  $\delta$  7.98 (br s, 1H, NH), 6.46 (t,  $J = 2.9$  Hz, 2H,  $\text{H}_{a,b}$ ), 5.25 (t,  $J = 2.9$  Hz, 2H,  $\text{H}_{c,h}$ ), 2.83 (s, 2H,  $\text{H}_{d,g}$ ).

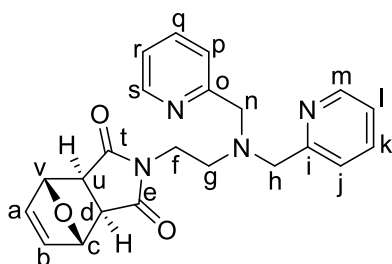
Step 2: (3aR,4S,7R,7aS)-2-(2-bromoethyl)-3a,4,7,7a-tetrahydro-1H-4,7-epoxyisoindole-1,3(2H)-dione (ii) carried out according to literature<sup>24</sup>.



Under an argon atmosphere, a solution of furan protected maleimide (1.5g, 1 equiv, 9.08mmol), 1,2 dibromoethane (3.94ml, 5 equiv, 45.4mmol) and  $\text{K}_2\text{CO}_3$  (6.27g, 5 equiv, 45.4mmol) in DMF (90 mL) was stirred at 50 °C overnight. The reaction was monitored by TLC analysis and upon completion, the solvent was removed under reduced pressure. Ethyl acetate (600 mL) was added to the reaction mixture and the product was extracted using water (4 x 300 mL). The organic fractions were combined and dried over  $\text{MgSO}_4$ . The product was filtered, concentrated and purified by column chromatography (2%MeOH : DCM) to give an off white solid (0.9g, 38% yield).

$^1\text{H}$  NMR (300 MHz,  $\text{CDCl}_3$ )  $\delta$  6.60 – 6.43 (m, 2H,  $\text{H}_{a,b}$ ), 5.37 – 5.18 (m, 2H,  $\text{H}_{c,j}$ ), 3.88 (t,  $J = 6.9$  Hz, 2H,  $\text{H}_f$ ), 3.47 (t,  $J = 6.9$  Hz, 2H,  $\text{H}_g$ ), 2.88 (s, 2H,  $\text{H}_{d,i}$ ).

Step 3: Synthesis of bis(picoly)aminoethylmaleimide-furan protected (iii)

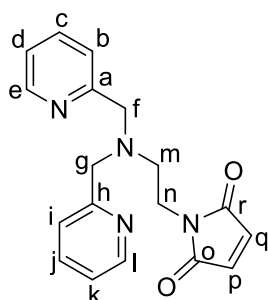


In a 250ml Schlenk flask, dipicolylamine (1.23ml, 1 equiv, 6.83mmol), maleimide tether (ii) (1.85g, 1 equiv, 6.83mmol), triethylamine (0.95ml, 1 equiv, 6.83mmol) and potassium iodide (1.13g, 1 equiv, 6.83mmol) were added to acetonitrile (90ml) under argon. The reaction mixture was stirred under argon at 70 °C for 72 hours. When the reaction had reached completion according to TLC analysis (eluent Ethyl Acetate: Hexane 3:1), water (100ml) was added and then the product was extracted with ethyl acetate (3 x 100ml). The organic fractions were combined, dried over  $\text{MgSO}_4$  and concentrated. The product was purified by

column chromatography (10% Methanol: Dichloromethane) and the product isolated was a light brown oil and was isolated in a 35% yield.

$^1\text{H}$  NMR (500 MHz,  $\text{CDCl}_3$ )  $\delta$  8.52 (ddd,  $J = 5.0, 1.8, 1.0$  Hz, 2H,  $\text{H}_{s,m}$ ), 7.65 (td,  $J = 7.6, 1.8$  Hz, 2H,  $\text{H}_{k,q}$ ), 7.48 (dt,  $J = 7.8, 1.1$  Hz, 1H,  $\text{H}_{j,p}$ ), 7.16 (ddd,  $J = 7.6, 5.0, 1.0$  Hz, 1H,  $\text{H}_{l,r}$ ), 6.53 (t,  $J = 0.9$  Hz, 1H,  $\text{H}_{a,b}$ ), 5.26 (t,  $J = 1.1$  Hz, 4H,  $\text{H}_{c,v}$ ), 3.85 (s, 4H,  $\text{H}_{h,n}$ ), 3.67 (t,  $J = 6.3$  Hz, 1H,  $\text{H}_f$ ), 3.49 (s, 2H,  $\text{H}_{d,u}$ ), 2.80 (t,  $J = 6.3$  Hz, 2H,  $\text{H}_g$ ).

#### Synthesis of bis(picoyl)aminoethylmaleimide (iv)- Cofactor 1



The bis(picoyl)aminoethylmaleimide-furan protected (iii) (0.13g, 0.34mmol) was dissolved in xylene (25ml). The reaction mixture was distilled to remove all traces of furan, upon boiling the thermometer reached a temperature of 49 °C before decreasing. The solvent was removed under reduced pressure. The crude product was recrystallized from ethanol/hexane (50:50). The product isolated was a light brown oily solid and was produced quantitatively.

$^1\text{H}$  NMR (500 MHz,  $\text{CDCl}_3$ )  $\delta$  8.53 (d,  $J = 5.1$  Hz, 2H,  $\text{H}_{e,l}$ ), 7.65 (t,  $J = 7.7$  Hz, 2H,  $\text{H}_{c,j}$ ), 7.42 (d,  $J = 7.9$  Hz, 2H,  $\text{H}_{b,i}$ ), 7.17 (t,  $J = 5.1$  Hz, 2H,  $\text{H}_{d,k}$ ), 6.68 (s, 2H,  $\text{H}_{p,q}$ ), 3.87 (s, 4H,  $\text{H}_{f,g}$ ), 3.70 (t,  $J = 6.1$  Hz, 2H,  $\text{H}_n$ ), 2.80 (t,  $J = 6.1$  Hz, 2H,  $\text{H}_m$ ).

$^{13}\text{C}$  NMR (126 MHz,  $\text{CDCl}_3$ )  $\delta$  170.6 ( $\text{C}_{o,r}$ ), 158.9 ( $\text{C}_{a,h}$ ), 149.3 ( $\text{C}_{e,l}$ ), 137.0 ( $\text{C}_{c,j}$ ), 135.2 ( $\text{C}_{p,q}$ ), 123.2 ( $\text{C}_{b,i}$ ), 122.2, 60.0 ( $\text{C}_{f,g}$ ), 51.5 ( $\text{C}_m$ ), 35.8 ( $\text{C}_n$ ).

Spectral data correspond to those given in literature<sup>14</sup>

HRMS (ESI)  $m/z$   $[\text{MH}]^+$  323.1503 (calculated  $\text{C}_{18} \text{H}_{18} \text{N}_4 \text{O}_2$  H: 323.1503)

### 2.9.3 Protein Production and Purification of SCP-2L A100C & V83C

SCP-2L A100C was prepared by transforming the pEHISTEV::dΔhΔSCP-2L A100C plasmid into competent *E. coli* Rosetta(DE3) cells. A single transformed colony was inoculated into 10 mL PB medium (Production Broth medium; containing 20 g/L tryptone, 10 g/L yeast extract, 5 g/L dextrose, 5 g/L NaCl, 8.7 g/L K<sub>2</sub>HPO<sub>4</sub>, pH 7.0) with 50 µg/mL kanamycin and 34 µg/mL chloramphenicol (in ethanol). This was used to inoculate 100 mL PB media (50 µg/mL kanamycin and 34 µg/mL chloramphenicol) and agitated at 37 °C, 200 rpm overnight (16-18 h). This starter culture (10 mL) was used to inoculate 0.5 litre of PB medium (2 litres in total). The cells were allowed to grow at 37 °C to an OD<sub>600</sub> of 0.6 (about 2 h) after which the temperature was lowered to 16 °C. Isopropyl-1-thio-β-D-galactopyranoside (IPTG) was added after one hour to a final concentration of 0.2 mM from a 0.2 M stock solution to initiate the expression of the recombinant protein. The culture was left overnight at 16 °C (16-18 h). The cells were harvested by centrifugation (20 min, 7500 g), washed with 250 mL of buffer solution (16 mM kPi (potassium phosphate) 120 mM NaCl, pH 7.4) and centrifuged again (20 min 7500 g). The pellet from the 2 litre cell culture was resuspended at 4 °C in 100 ml of buffer solution (50 mM Tris·HCl, 20 mM imidazole, 150 mM NaCl, 0.5 mM benzamidine, pH 8) and frozen at -80 °C.

After defrosting, 20 mg of lysozyme, 1 mg of DNase and 1 mL of 1 M of MgCl<sub>2</sub> was added to the suspension. The suspension was left for 1 hour at 4 °C and was sonicated in portions of 10 mL (Hielscher UP200S ultrasonic processor, 0.5 s pulses 90% power) with 0.5 second pulses for 1 minute. The cell-extract, obtained after centrifugation (50 min, 30000 g) was filtered (0.22 µm PES Millex Filter unit) and applied to a nickel column (5 mL HisTrap FF) equilibrated with 30 mM Tris·HCl, 20 mM imidazole, 150 mM NaCl, pH 8 (wash buffer) at flow rates up to 5 mL/min. The column was washed with 5 column volumes of wash buffer, 5 column volumes of high salt buffer (wash buffer containing 1 M NaCl) and another 5 column volumes of wash buffer. The protein was obtained by eluting with 6 column volumes elution buffer (wash buffer containing 330 mM imidazole) into an equal volume of wash buffer to prevent precipitation of the protein and the resulting solution was dialysed

against 5 litre buffer solution (30 mM Tris·HCl, 10 mM imidazole, 150 mM NaCl, pH 8 at 4°C). If precipitate had formed the resulting mixture was centrifuged (30000 g, 50 min). 0.014 equivalents of TEV-protease<sup>25</sup> and final concentrations of 1 mM DTT and 0.5 mM EDTA were added to resulting His-tagged SCP-2L A100C solution. This mixture was left overnight at room temperature.

If precipitate had formed the mixture was again centrifuged (50 min, 30000 g) and the pellet was discarded. The supernatant was then filtered (0.22 µm PES Millex Filter unit) and used for nickel affinity chromatography on a nickel column (5 mL HisTrap FF) equilibrated with wash buffer. The flow-through containing pure SCP-2L A100C was collected, concentrated and the buffer was exchanged to the storage and coupling buffer (20 mM MES, 50 mM NaCl, pH 6) using a centrifugal concentrator. Three consecutive rounds of buffer exchange were performed to remove all the DTT. Buffer exchange by centrifugal concentrator was found to be more effective than dialysis. Yields up to 75 mg L<sup>-1</sup> culture, with a purity of more than >99% according to SDS-PAGE (Nupage™ 4-12% Bis-Tris gel), were obtained.

#### **2.9.4 Bioconjugation of cofactor to unique cysteine mutants of SCP-2L**

A 25-50 µM solution of protein in buffer (20 mM MES 50 mM NaCl pH 6) was incubated with 10 equivalents cofactor added from a 50 mM stock solution in the corresponding buffer. Typically, no residual free protein was observed. The reaction mixture is shaken gently and left at room temperature overnight. When full conversion was shown the mixture was washed using a centrifugal concentrator with buffer to remove the excess cofactor.

#### **2.9.5 Deconvoluted Mass Spectrums of Modified Proteins.**

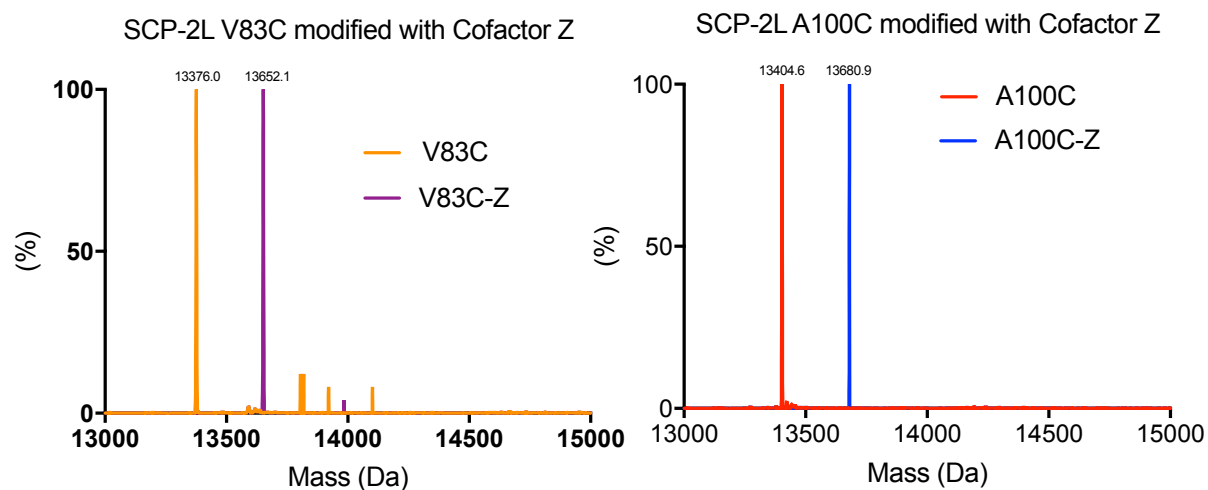
LC-MS(ES+) used for analysis of protein and conjugation-reactions was performed on a Waters Alliance HT 2795 equipped with a Micromass LCT-TOF mass spectrometer, using positive electrospray ionisation and applying a Waters MASSPREP® On-line Desalting 2.1x10 mm cartridge using a gradient of 1% formic acid in H<sub>2</sub>O to 1% formic acid in acetonitrile. ESI-MS results were analysed by MassLynx V. 4.0 and its MaxEnt algorithm. The spectrums shown are deconvoluted and the y axis is given in % abundance.

SCP-2L V83C-Z:

Calculated mass: 13652.6 Da, Observed: 13652.1 Da.

SCP-2L A100C-Z:

Calculated mass: 13679.7 Da, Observed: 13680.9 Da.

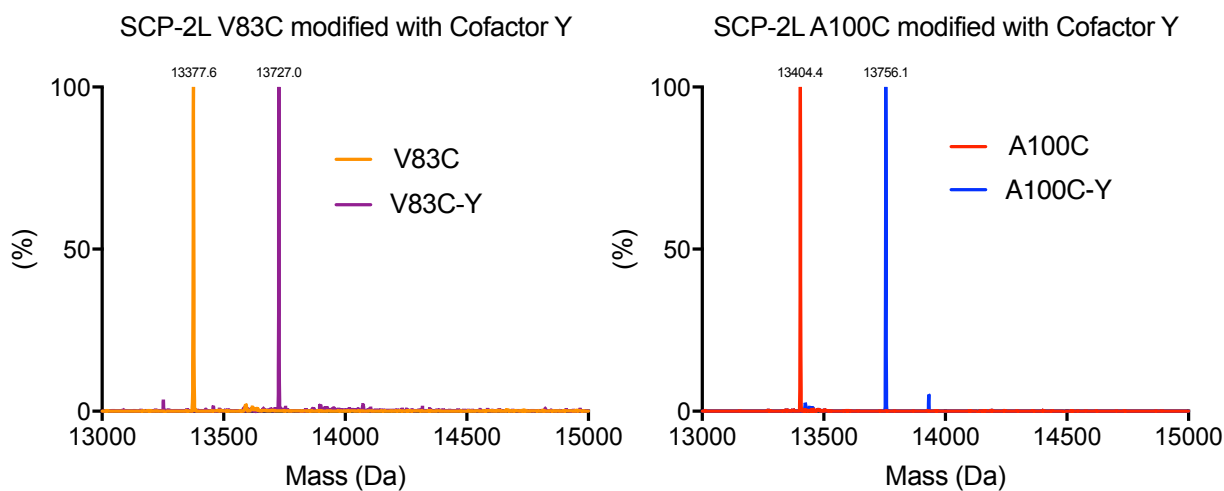


SCP-2L V83C-Y:

Calculated mass: 13727.7 Da, Observed: 13727.0 Da.

SCP-2L A100C-Y:

Calculated mass: 13754.7 Da, Observed: 13756.1 Da.



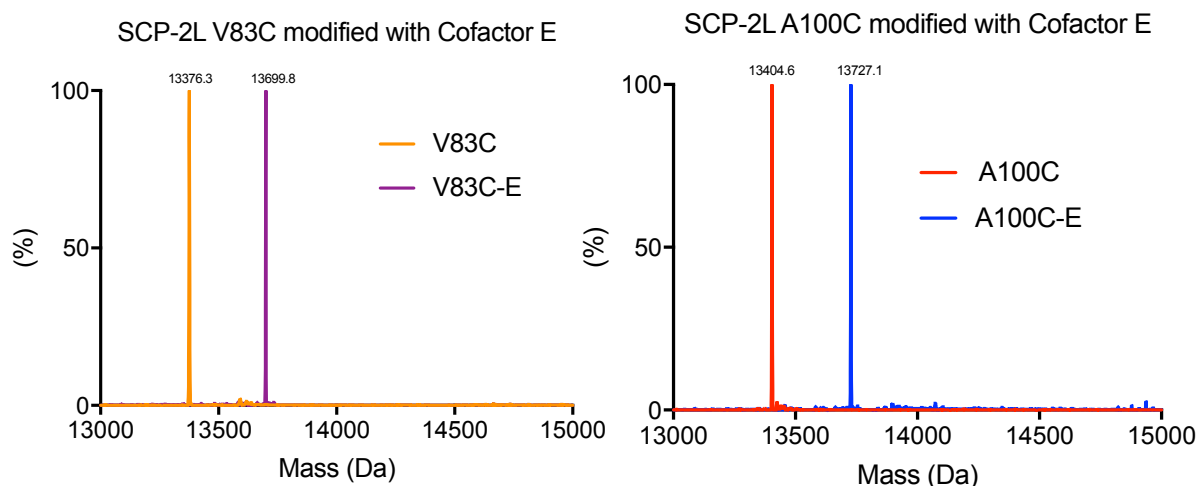


SCP-2L A100C-E:

Calculated mass: 13726.7 Da, Observed: 13727.1 Da.

SCP-2L V83C-E:

Calculated mass: 13698.7 Da. Observed: 13699.8 Da.

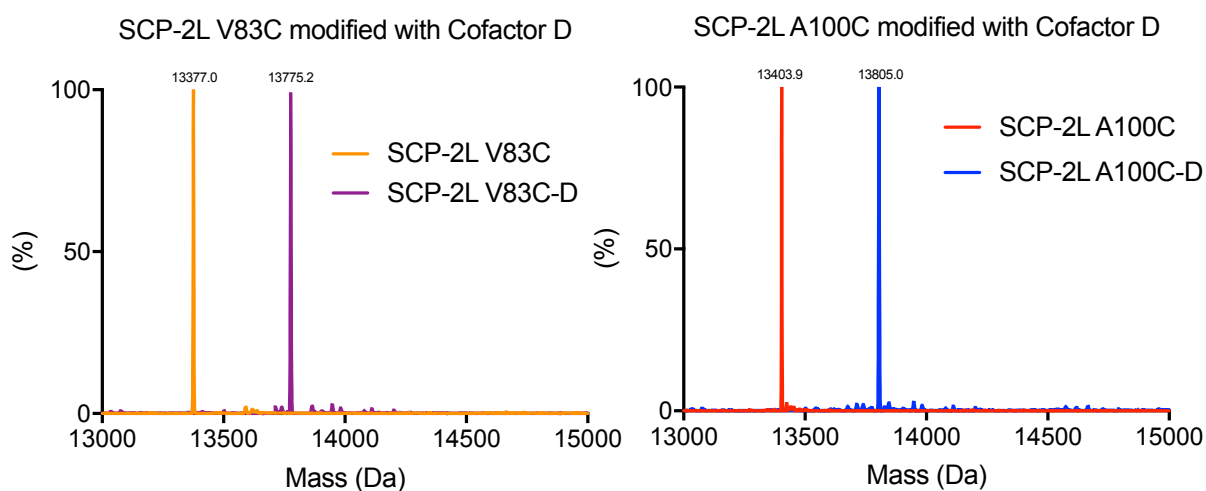


SCP-2L A100C-D:

Expected mass: 13803.8 Da, Observed: 13805.0 Da.

SCP-2L V83C-D:

Calculated mass: 13775.8 Da, observed: 13775.2 Da.

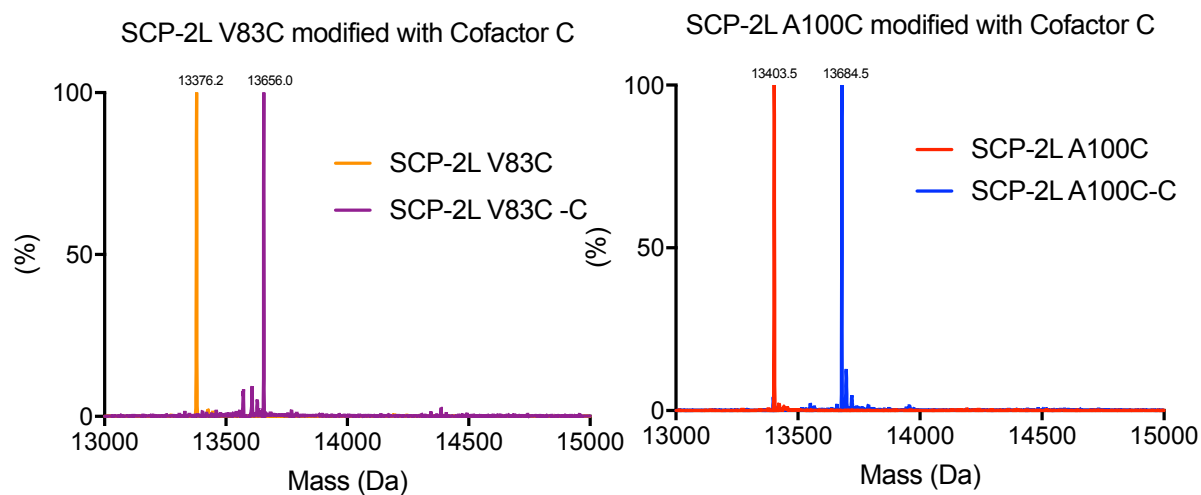


SCP-2L V83C-C:

Calculated mass: 13656.8 Da, Observed: 13656.0 Da.

SCP-2L A100C-C:

Calculated mass: 13684.8 Da, Observed: 13684.5 Da.

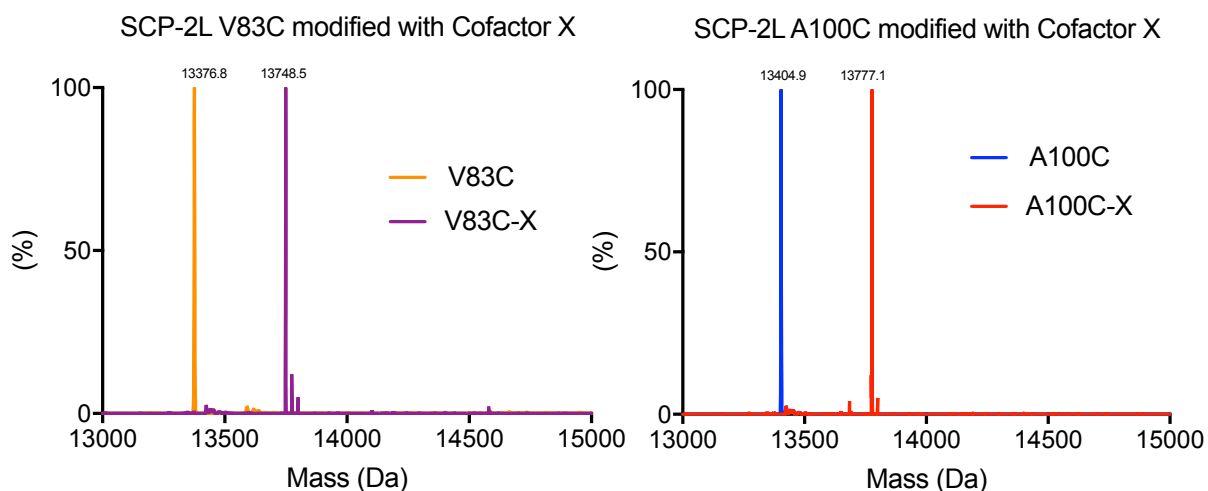


SCP-2L V83C-X:

Calculate mass: 13748.7 Da, Observed: 13748.5 Da.

SCP-2L A100C-X:

Calculate mass: 13776.7 Da, Observed: 13777.1 Da.



## 2.7.5 Determination of Modified Protein Concentration

QAA analysis was carried out PNAC facility, Dept of Biochemistry, University of Cambridge and the full detailed set of results are shown below.

Sample: **AA1**

*Integer fit of measured mole ratios to expected values*

Amino acid	Expected value	Observed value	Closeness of fit to expected value
Cys	1	not determined	-
Asp	10	10.22	better than 5%
Thr	5	5.06	better than 5%
Ser	6	5.81	better than 5%
Glu	12	11.83	better than 5%
Gly	15	14.27	better than 5%
Ala	8	8.13	better than 5%
Val	8	7.60	better than 5%
Met	4	3.94	better than 5%
Ile	9	8.55	better than 5%
Leu	11	11.13	better than 5%
<b>Norleu std</b>			
Tyr	2	2.11	within 5-10%
Phe	6	6.57	within 5-10%
His	1	1.09	within 5-10%
Lys	15	15.38	better than 5%
Arg	4	4.29	within 5-10%
Pro	0	excluded	-
Trp	2		(not determined)
<b>Total (used)</b>	<b>116</b>	<b>residues</b>	

**Average AA1,2**

Total sample	2.647	nmoles
	36.55	ug
Concentration	52.95	nmoles/ml
	730.94	ug/ml

**2.665**  
**36.79**  
**53.30**  
**735.78**

Sample: **AA2**

*Integer fit of measured mole ratios to expected values*

Amino acid	Expected value	Observed value	Closeness of fit to expected value
Cys	1	not determined	-
Asp	10	10.20	better than 5%
Thr	5	5.11	better than 5%
Ser	6	5.91	better than 5%
Glu	12	11.74	better than 5%
Gly	15	14.33	better than 5%
Ala	8	8.00	better than 5%
Val	8	7.61	better than 5%
Met	4	4.02	better than 5%
Ile	9	8.55	within 5-10%
Leu	11	11.09	better than 5%
<b>Norleu std</b>			
Tyr	2	2.08	better than 5%
Phe	6	6.56	within 5-10%
His	1	1.09	within 5-10%
Lys	15	15.37	better than 5%
Arg	4	4.34	within 5-10%
Pro	0	excluded	-
Trp	2		(not determined)
<b>Total (used)</b>	<b>116</b>	<b>residues</b>	

Total sample	2.682	nmoles
	37.03	ug
Concentration	53.65	nmoles/ml
	740.61	ug/ml

- 
- <sup>1</sup> K. J. Jensen, *Peptide and Protein Design for Biopharmaceutical Applications*, **2009**, 207
- <sup>2</sup> P. J. Deuss, G. Popa, A. M. Z. Slawin, W. Laan, P. C. J. Kamer. *ChemCatChem*, **2013**, 5, 1184
- <sup>3</sup> P. J. Deuss, G. Popa, C. H. Botting, W. Laan, P. C. J. Kamer. *Angew. Chem. Int. Ed. Engl.* **2010**, 49, 5315
- <sup>4</sup> P. J. Deuss, R. den Heeten, W. Laan, P. C. J. Kamer. *Chem. Eur. J.* **2011**, 17, 4680
- <sup>5</sup> A. G. Jarvis, L. Obrecht, P. J. Deuss, W. Laan, E. K. Gibson, P. P. Wells, P. C. J. Kamer. *Angew. Chem. Int. Ed. Engl.* **2017**, 56, 13596
- <sup>6</sup> A. Haapalainen, D. van Aalten, G. Merilainen, J. Jalonen, P. Pirila, R. Wierenga, J. Hiltunen, T. Glumoff, *J. Mol. Biol.* **2001**, 313, 1127
- <sup>7</sup> G. T. Hermanson, *Bioconjugate Techniques*, Second ed., Academic Press, San Diego, **2008**
- <sup>8</sup> I. W. C. E. Arends, G. J. ten Brink, R. A. Sheldon. *J. Mol. Catal. A Chem.* **2006**, 251, 246
- <sup>9</sup> S. M. Islam, M. Mobarok, P. Mondal, A. S. Roy, S. Mondal, D. J. Hossain. *Appl. Polym. Sci.* **2012**, 123, 3789
- <sup>10</sup> M. Klopstra, R. Hage, R. M. Kellogg, B. L. Feringa. *Tetrahedron Lett.* **2003**, 44, 4581
- <sup>11</sup> J. Kim, R. G. Harrison, C. Kim, L. Que. *J. Am. Chem. Soc.* **1996**, 118, 4273
- <sup>12</sup> R. R. Davies, M. D. Distefano, *J. Am. Chem. Soc.* **1997**, 119, 11643
- <sup>13</sup> M. T. Reetz, M. Rentzsch, A. Pletsch, A. Taglieber, F. Hollmann, R. J. G. Mondiere, N. Dickmann, B. Hoecker, S. Cerrone, M. C. Haeger, R. Sterner, *ChemBioChem*, **2008**, 9, 552
- <sup>14</sup> S. R. Banerjee, P. Schaffer, W. Babich, J. F. Valliant, J. Zubieta. *Dalton Trans*, **2005**, 3886
- <sup>15</sup> F. Xie, D. E. K. Sutherland, M. J. Stillman, M. Y. Ogawa, *J. Inorg. Biochem.* **2010**, 104, 261
- <sup>16</sup> M. Allard, C. Dupont, V. Munoz Robles, N. Doucet, A. Lledos, D. D. Marechal, A. Urvoas, J. P. Mahy, R. Ricoux, *ChemBioChem*, **2012**, 13, 240
- <sup>17</sup> C. Bertucci, C. Botteghi, D. Giunta, M. Marchetti, S. Paganelli, *Adv. Synth. Catal.* **2002**, 344, 556
- <sup>18</sup> N. J. Greenfield, *Nat Protoc*, **2006**, 1, 2876
- <sup>19</sup> M. M. Bradford. *Analytical Biochemistry*, **1976**, 72, 248
- <sup>20</sup> S. M. Rutherford, B. M. Dunn. *Curr Protoc Protein Sci*, **2011**, Chapter 3: Unit 3.2
- <sup>21</sup> T. Matsuo, C. Imai, T. Yoshida, T. Saito, T. Hayashi, S. Hirota, *Chem. Commun. (Cambridge, U. K.)* **2012**, 48, 1662
- <sup>22</sup> A. N. Zaykov, B. V. Popp, Z. T. Ball, *Chemistry*, **2010**, 16, 6641
- <sup>23</sup> O. K. Farha, R. L. Julius. M. F. Hawthorne. *Tetrahedron Letters*, **2006**, 47, 2619
- <sup>24</sup> D. Fujita, K. Suzuki, S. Sato, M. Yagi-Utsumi, E. Kurimoto, Y. Yamaguchi, K. Kato, M. Fujita. *Chem. Lett.* **2012**, 41, 313

## Chapter 3 – Engineering thermo stability in SCP-2L to increase catalytic activity for hydroformylation

### 3.1 Introduction to protein engineering

Over the last decades, the power of biocatalysis in chemical synthesis, particularly within an industrial setting, has become undeniable. Enzymes such as transaminases and ketoreductases are increasingly used in the synthesis of chiral intermediates in pharmaceutical processes<sup>1,2,3</sup> showing that biocatalytic technology may replace chemical steps as part of existing chemical pathways by its simple nature and attractive sustainability metrics. The recent work towards artificial metalloenzymes (ArMs) has added to the potential of biocatalysis, and provides an elegant approach to combining enzymatic and chemical methods, with the potential to create novel catalysts.<sup>4,5,6,7</sup> ArMs combine the molecular recognition properties of enzymes, which are responsible for the high selectivities observed in many enzymatic processes, with synthetic transition metal cofactors allowing unnatural reactions to be introduced into the biocatalysis arena. The most successful method to create ArMs has been non-covalent binding utilizing protein scaffolds with supramolecular recognition motifs e.g. (strept)avidin.<sup>8</sup> The inclusion of a cofactor within the protein scaffold can also be achieved by covalent modification using a free thiol group on a unique cysteine residue. Using these methods, a whole range of ArMs have been developed for unnatural reactions including transfer hydrogenation<sup>9</sup>, hydroformylation,<sup>10</sup> cross coupling reactions and other carbon-carbon bond forming reactions such as the Diels-Alder reactions.<sup>11</sup>

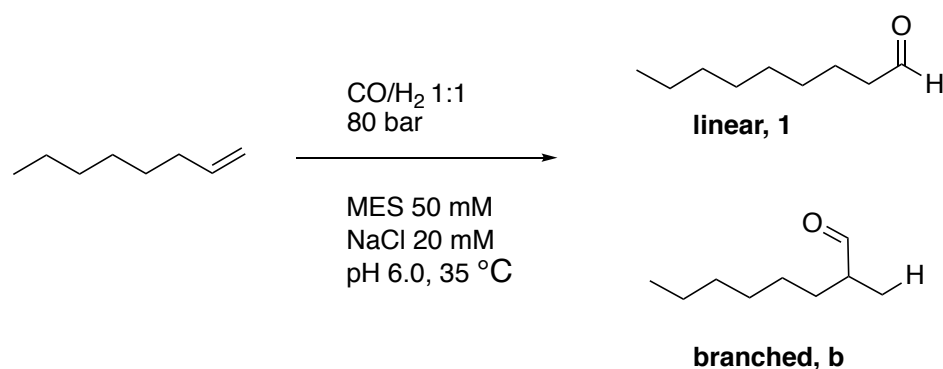
Whilst the use of proteins and enzymes in industrial processes has become increasingly common, the combination with organometallic cofactors can be difficult as the conditions they require for efficient catalysis can include high temperatures and the presence of organic solvents. Enzymes have evolved over time to work in a cellular environment; therefore, harsh conditions often lead to denaturation of proteins and deactivation of the catalyst.<sup>12</sup> This is a particular problem in transferring chemical reactions into biocatalytic reactions as ArMs are designed to do. Several successful examples showed that by targeting a protein using computational methods, the stability of the protein could be improved, which in turn can demonstrate higher reactivity.<sup>13, 14</sup> Based on these documented examples we hypothesized

that if we could increase the thermostability of the protein scaffold, we could increase the activity of our hybrid catalysts and their tolerance to organic solvents.<sup>15</sup> This would expand the repertoire of possible reactions we could conduct and increase the experimental design space we can operate in.

The strategy used to engineer thermostability into a protein is the rational design of mutations based on structural information. Such mutations can be focused on flexible regions in the protein with the aim to reduce flexibility.<sup>16</sup> Introducing disulphide bridges can stabilize early unfolding regions or destabilize the unfolded state by reducing its entropy.<sup>17</sup> However within ArM work introducing disulphides is not applicable as it could result in multiple coordination sites for protein modification. Another strategy to enhance stability is the introduction of surface-located charged groups that can form salt bridges.<sup>18</sup> The electrostatic interaction contributes to the stability of the protein, so it would be beneficial to introduce this into flexible regions within the proteins scaffold.<sup>19</sup> Several studies show that modifying electrostatic interactions, such as removal of isolated surface charges<sup>20, 21</sup> contribute to the stability of the protein. Analysis of the superfamily, that the protein in question belongs to, can also be utilised to identify any consensus sequences that could be introduced or adjusted to benefit the protein. The consensus approach is based on the hypothesis that the more highly conserved amino acids of a sequence alignment generally add more to the protein than the non-conserved amino acids.<sup>22</sup>

The Kamer group has previously reported covalent approaches for synthesizing ArMs containing metal binding ligands via a maleimide linker to the protein scaffold human steroid carrier protein SL (SCP-2L).<sup>23</sup> The SCP-2L protein is a domain from the multifunctional enzyme type 2 (MFE-2) and was specifically selected for its ability to bind linear substrates in its hydrophobic tunnel.<sup>24</sup> This hybrid catalyst showed high activities and selectivities for the production of long chain linear aldehydes in rhodium-catalyzed hydroformylation of alkenes, under benign aqueous conditions (Scheme 1).<sup>25</sup> This reaction is of industrial interest due to the high added-value of the aldehyde products in comparison to the alkene substrates. Within industry, rhodium-catalyzed hydroformylation is used on a large scale under biphasic conditions, enabling the recovery of the expensive rhodium catalyst. However, this reaction is challenging for longer alkenes (>5 carbon atoms) due to their low solubility in aqueous

solution.<sup>26</sup> ArMs provide a catalyst for higher alkene substrates that can work at low temperatures compared to industrial conditions (typically 110-120 °C), with high selectivity and turnover numbers, overall providing a much greener reaction. We set out to see if we could use sequence-based / rational protein engineering techniques to increase the stability of our artificial hydroformylase and thus potentially increase the reaction temperature and, therefore, rate of hydroformylation.

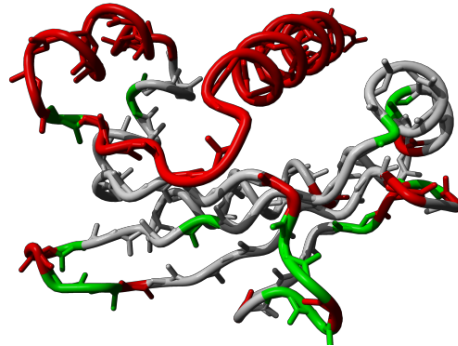


*Scheme 1. Hydroformylation of alkenes to aldehyde products. Conditions from previous work within the Kamer group.<sup>10</sup>*

### 3.3 Design of thermostable mutants

The design of mutations that could increase stability of SCP-2L was carried out both using rational design and sequence based mutant selection with SCP-like family 3DM database.<sup>27</sup> The 3DM databases consist of structure-based alignments of homologous sequences and assigns each residue a position number that allows comparison of positions that are conserved in the protein sequence. Novel crystal structures of the SCP-2L variants with a cysteine residue introduced for covalent bonding (SCP-2L variants A100C and V83C) were added to the SCP-like family database. Using the 3DM database and structural analysis of the protein, the positions in the protein where mutations towards consensus could likely be introduced without loss of structural integrity were determined (Figure 1). The consensus sequences are the most likely residues found at a specific position within the protein super family. Therefore if SCP-2L was missing any of the consensus residues they would be an interesting target to be introduced. The hydrophobic tunnel, which could be involved in ligand binding, was excluded from mutagenesis, as were the four helices that are involved in

substrate binding, i.e. long-chain fatty acyl-CoAs<sup>28</sup> and cholesterol derivatives.<sup>29</sup> Mutations were especially sought in three loops that appear more flexible based on the B factor of the structure.



*Figure 1. SCP-2L (1IKT) Mutation points to consider and avoid, highlighted in red are areas to avoid and in green are the areas that could be of interest.*

To stabilize the protein, I evaluated a number of rational mutation strategies, including stabilizing the flexible loop regions by salt bridge formation, strengthening alpha helix dipoles and to mutations introducing conserved amino acids in the target protein. The first step of the design process was to identify residues that would lead to the formation of salt bridges to stabilize the flexible loops and regions on the proteins surface. I44 is positioned within a loop between beta sheets in a flexible region so creating a salt bridge here was anticipated to help increase stability. I44 is in close proximity to a positively charged Lys residue and by introducing a Glu or Asp an electrostatic interaction can be introduced (Figure 2). Residues S56, K58, K65 and E81 were also identified as sites for salt bridge formation, which can be achieved by reversing / or introducing a charge to create new ionic interactions.



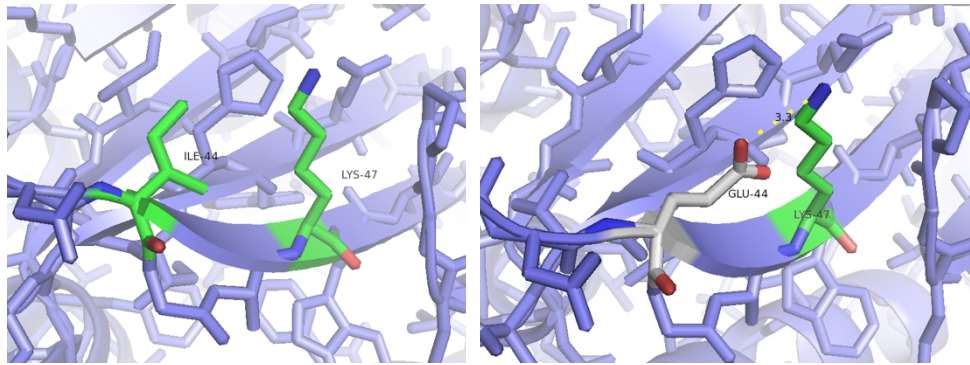


Figure 2. Introduction of a possible Glu44-Lys47 salt bridge by mutation I44E.

The alignment of dipoles of the polypeptide backbone parallel to the axis of an alpha helix causes a net dipole moment with its positive pole at the amino terminus and negative pole at the carboxy terminus. Residue N31 is found on the terminus of an alpha helix and to strengthen the helix dipole a charged residue was introduced by mutation N31D.

Identifying amino acids that are highly conserved within the protein superfamily using the 3DM database enables introduction of more conserved residues within SCP-2L. Residue V26 is positioned within an alpha helix and is 15.5% conserved within the SCP-like 3DM database. However, F26 is the more abundant residue within the whole protein family (27.5% conserved). Furthermore, introduction of an aromatic residue at this position may provide potential  $\pi$ -stacking interactions (Figure 3). A residue that follows the protein super family's consensus can also be introduced by mutations G41X, S54X, D116X and P68X. A full list of the designed mutants is shown in Table 1.

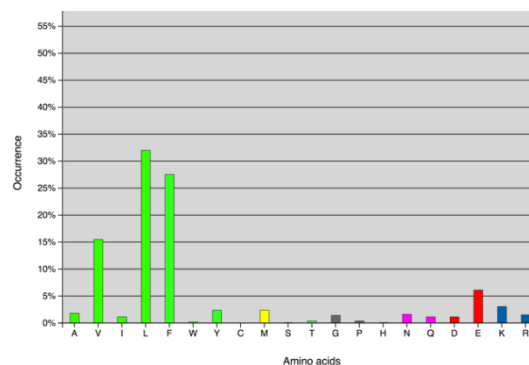


Figure 3. Amino acid distribution for residue V26 within the SCP-like family in 3DM database

### 3.4 Synthesis of mutant library

Site directed mutagenesis<sup>30</sup> was performed both for SCP-2L A100C and V83C to separately obtain plasmids encoding mutations G41D, S56D, K58E, K58D, I44E, I44D, E81K, S54N, S56E, D116K, I44D, K40N, A68P, K65E, N31D, or V26F. Successful mutagenesis was seen for all of the designed mutants and this was determined by DNA sequencing.

### 3.5 Thermostability testing of mutant library

#### 3.5.1 Thermofluor analysis of proteins

One of the most common and convenient methods to measure the thermostability of a protein is to use a thermofluor assay. The thermofluor assay is a quick, temperature-based assay to assess the stability of the protein. The assay uses a Sypro Orange Dye (Figure 4) and is mixed with the target protein. The dye binds to the hydrophobic patches/ denatured protein and fluoresces. With increased temperature the protein denatures and unfolds, and it is therefore possible to monitor the increase in fluorescence signal and determine a melting temperature. This method was attempted on the SCP-2L mutants but proved to be unsuccessful. Due to the large hydrophobic tunnel within the protein, the dye binds immediately to protein and from time point zero a large signal was seen. Due to the initial large fluorescent signal, differences from protein unfolding were difficult to determine, therefore an alternative method was used.

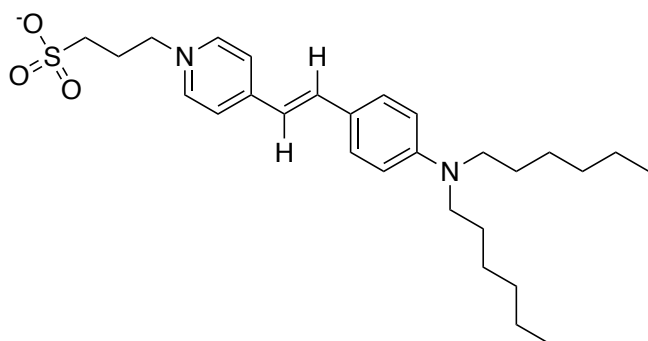


Figure 4 – Sypro orange Dye

### 3.5.2 Using CD spectroscopy to determine thermostability of protein scaffolds

Using circular dichroism (CD) spectroscopy an approximate  $T_{m,app}$ <sup>31</sup> was measured and compared to SCP-2L A100C and V83C encoded by the original plasmid. The CD method used a fast scan approach only measuring 6 different temperature points; therefore the  $T_{m,app}$  values are approximate and serve the purpose of indicating how the thermostability is affected by the new amino acid introduced. If the folding of the protein would be reversible then true thermodynamic parameters could be calculated. In this circumstance some precipitate is formed at higher temperatures, so the value given is  $T_{m,app}$ . However, this still gives information about the relative stability.

The results showed some significant improvements in thermostability over the original SCP-2L A100C and V83C proteins with increases of up to 12 and 10 °C observed for mutants constructed in an A100C and V83C background, respectively. In general, the A100C protein template was more responsive to the designed mutations. The introduction of N31D gave an increase of 12 °C for A100C, but only a 3 °C increase for V83C (Table 1, entry 1). The introduction of salt bridges was successful in raising the  $T_{m,app}$  of both mutants, with both E81K and S56D giving large increases in  $T_{m,app}$ . The  $T_{m,app}$  of mutant A100C increased by 9 and 11 °C for E81K and S56D, respectively, and with V83C increases of 6 and 10 °C were observed (Table 1, entries 4, 10). Mutation S56D improved the  $T_{m,app}$  in both mutants whereas S56E led to a decrease in  $T_{m,app}$  for both mutants. Introducing more conserved amino acids based on the results found using the 3DM database led to more modest increases in  $T_{m,app}$ . An exception was the G41D mutation in A100C which increased the  $T_{m,app}$  by 7 °C. In comparison, this mutation in V83C led to a decrease in  $T_{m,app}$  of 6 °C (Table 1, entry 14).

Reaction Entry	Design Method		Mutation	Template	
				A100C $\Delta T$ ( °C )	V83C $\Delta T$ ( °C )
1	Stabilizing dipole	$\alpha$ -helices	N31D	12	3
2			I44D	2	-1
3			I44E	4	1
4			S56D	11	10
5	Introducing a salt bridge	a salt	S56E	-2	-7
6			K58D	5	n.p.
7			K58E	u	1
8			K65D	n.p.	n.p.
9			K65E	7	1
10			E81K	9	6
11			V26F	u	3
12			K40N	4	3
13	Amino Conservation	Acid	G41D	7	-6
14			S54N	n.p.	n.p.
15			A68P	3	4
16			D116K	2	0

*Table 1. The  $\Delta T$  (°C) of the new designed thermostable mutants compared to SCP-2L V83C and A100C ( n.p, no data due to lack of protein; u, unreliable  $T_{m,app}$  obtained) This fast scan method is used serve the purpose of indicating how the thermostability is affected by the new amino acid introduced the data is approximate and the  $\Delta T$  (°C) error is within 3 °C.*

### 3.6 Design and synthesis of Double mutants

Within the literature, it is reported that combining mutations that increase thermostability can give an enhanced effect.<sup>32</sup> Therefore, the three best results were taken, S56D, E81K, N31D and six double mutants were successfully produced according to the same protocol as the single mutants.

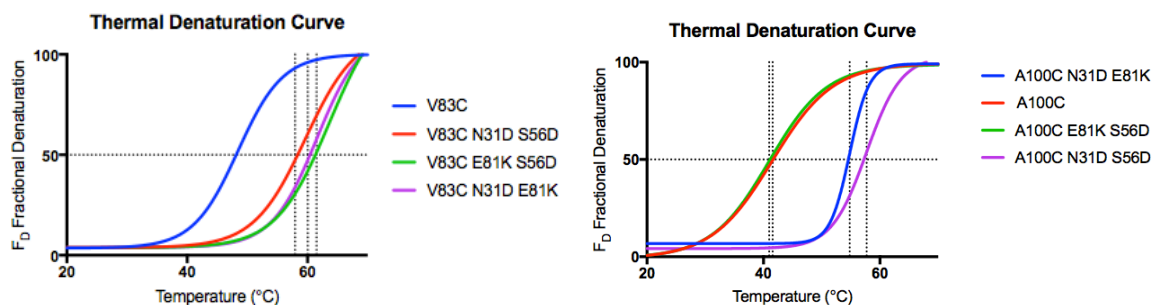
### 3.7 CD spectroscopy analysis of double mutants

Using CD spectroscopy, the melting temperature ( $T_{m,app}$ ) of V83C and A100C and the thermostable double mutants were measured using data points taken at every degree point at 222 nm (Table 2). This CD method provides a more accurate value than the fast scan approach used in the previous screen as a larger number of data points are utilised. A Thermal denaturation curve is produced and from this the  $T_m$  app can be extrapolated.

Variant	$T_{m,app}$ (°C)	$\Delta T$ °C
V83C template	48	
A100C template	42	
A100C N31D S56D	58	+16
A100C N31D E81K	55	+13
A100C E81K S56D	41	-1
V83C N31D S56D	60	+12
V83C N31D E81K	62	+14
V83C E81K S56D	64	+16

Table 2. Apparent melting temperatures ( $T_{m,app}$ ) of SCP-2L mutants determined using CD, measuring the absorbance at 222 nm at 1°C steps. The data is accurate to the nearest degree.

The new thermostable mutants showed impressive improvements in thermostability, increasing the  $T_{m,app}$  by 12-16 °C in all but one case. However, variant A100C/E81K/S56D, showed a negative effect of combining E81K and S56D mutations and the  $T_{m,app}$  decreased by 1 °C. This protein was obtained in very low yields (2mg/L) so the protein was not used further in this work.



*Fig. 5 – Thermal Denaturation Curve of Double SCP-2L Mutants*

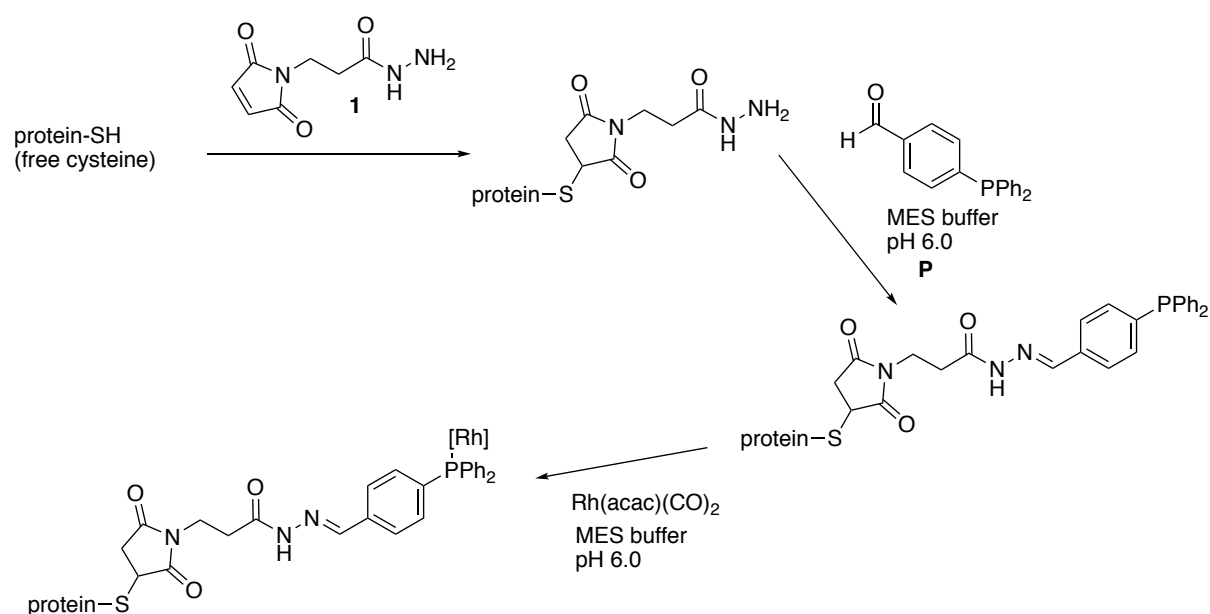
The rational design method proved successful and an increase of protein thermostability was obtained. Therefore, the next step was to test the hypothesis to see if this would result in a more stable catalyst. Hydroformylation was chosen as a benchmark reaction because it is the most studied reaction using SCP-2L as a protein scaffold for ArMs. Previous results reported by the Kamer group showed that SCP-2L A100C and V83C rhodium proteins showed significant hydroformylation activity of up to 400 and 100 TONs respectively over 48 h. These activities were surprisingly high considering that the reactions were carried out at 35 °C, which is considerably lower than typically used in industrial conditions. With a more stable protein scaffold, the activity could increase further as less catalyst degradation will occur over time allowing for higher temperatures.

## 3.8 Hydroformylation catalysis using double mutants

### 3.8.1 Protein modification

The SCP-2L mutants were successfully modified with an aldehyde phosphine (**P**) through a two-step cysteine modification (Scheme 2). The rhodium proteins (SCP-2L A100C/V83C-1-P-Rh) were obtained by the addition of Rh(acac)(CO)<sub>2</sub> to the phosphine modified proteins. In order to ensure all rhodium is bound an excess (2 eq.) of protein was used. If 'free Rh' is present it will have hydroformylation activity leading to low selectivity for the linear aldehyde (ca. 55%) and high TON's (TONs >500) (Table 3 Reaction Entry 15).

Previous work in the Kamer group investigated the effect of the structure of the phosphine cofactor used. The activity was found to increase as the phosphine moved from the *ortho* to *meta*, and *ortho* to *para* position by 30-fold and 70-fold respectively for SCP-2L V83C. Therefore, in this work only the phosphine cofactor in the *para* position was utilized.<sup>10</sup>



Scheme 2 – Synthesis of the artificial metalloproteins

### 3.8.2 Hydroformylation of 1-octene catalyzed by rhodium-hydroformylase

The hydroformylation (Scheme 1) activities of each of the five new double mutants were investigated using 1-octene as the substrate and compared to the original SCP-2L A100C and V83C. The reactions were carried out under the same conditions as previously published<sup>10</sup>, except for the reaction time, which was shortened to 16 h (Table 3).

Reaction Entry	Mutant	Temp °C	Selectivity	
			Linear Aldehyde (%) (average)	TON (average)
1	<b>V83C</b>	35	79.2 (0.15)	42.8 (0.34)
2		45	77.8 (0.09)	11.5 (1.27)
3	<b>V83C E81K S56D</b>	35	78.3 (0.08)	92.1 (0.97)
4		45	79.5 (0.12)	190.3 (1.11)
5	<b>V83C N31D E81K</b>	35	80.1 (0.10)	103.7 (0.89)
6		45	78.4 (0.32)	201.4 (0.93)
7	<b>V83C N31D S56D</b>	35	74.5 (0.07)	98.7 (1.02)
8		45	76.8 (0.87)	180.4 (1.23)
9	<b>A100C</b>	35	79.9 (0.06)	210.8 (0.65)
10		45	75.4 (0.15)	98.9 (2.90)
11	<b>A100C E81K S56D</b>	35	79.3 (0.65)	201.5 (1.32)
12		45	78.2 (1.10)	321.6 (3.14)
13	<b>A100C N31D E81K</b>	35	71.9 (0.21)	213.4 (0.44)
14		45	73.4 (0.54)	345.7 (2.12)
15 <sup>[*]</sup>	<b>Rh(acac)(CO)<sub>2</sub></b>	35	55.3 (0.67)	529.7 (53.30)

Table 3. Hydroformylation of 1-octene catalyzed by rhodium-hydroformylase Conditions: 80 bar syngas (1:1), stirring 625 rpm, 16 h in degassed MES buffer, 0.5 mL of catalyst solution and 0.5 mL of alkene containing 9% (v/v) n-heptane and 1% (v/v) diphenyl ether. Results are the average of three repeated reactions, standard deviation between brackets. Yields determined by GC. [\*] Reaction entry 15 is a free Rh(acac)(CO)<sub>2</sub> published control: 41.25 h, total volume: 0.4 mL 1-octene, no water, 150 nmol Rh.<sup>10</sup>



All three of the new thermostable SCP-2L V83C mutants (V83C E81K S56D/ V83C N31D E81K/ V83C N31D S56D) showed an improvement in TONs compared to SCP-2L V83C at 35 °C. SCP-2L V83C E81K S56D gave a TON of 92.1 which is double that seen of SCP-2L V83C under the same conditions (Table 3, entry 3 vs. 1). This suggests a possibility that when using V83C as a rhodium protein catalyst, there is significant protein degradation and increasing the stability of the protein enables the catalyst to be active throughout the 16-h reaction. To test this further, the reaction temperature was increased to 45 °C, the hypothesis being that rates would increase at higher temperatures if the protein continues to be stable. For SCP-2L V83C (Table 3 entry 2) a significantly reduced TON was observed, most likely due to rapid denaturation of the protein. However, the new thermostable rhodium proteins remained active at this higher temperature and produced high TONs of up to 200 (Table 3 entry 6).

The two new thermostable A100C mutants (A100C N31D E81K/ A100C N31D S56D) showed no improvement but consistent results with A100C. However, when the reaction temperature was increased to 45 °C these rhodium proteins were stable unlike the parent A100C and showed an increase in TONs up to 345 (Table 3 entry 14). This increase in TON is what would be expected by increasing the reaction temperature by 10 °C and indicates that the new designed A100C mutant is stable and can retain its structure at 45 °C.

The selectivity of the reaction varied between 75 and 80% for the linear product (nonanal), matching typical selectivities for double phosphine ligated rhodium catalyzed hydroformylation.<sup>33</sup> The rhodium used in this work is mono-phosphine ligated rhodium which normally would give lower selectivities.<sup>34</sup> This suggests that the protein scaffold counter balances the lack of phosphine ligands.

To test the stability of the artificial metalloenzymes further the reaction time was increased to 48 h to see if the TON could be increased (Table 4). The new thermostable V83C mutants showed an increase of TON from the original V83C of 74 up to 220 (Table 4 entry 1 and 3 respectively) and when the temperature was pushed to 45 °C the TON was pushed up to 406 (Table 4, entry 8). This suggests that the new thermostable mutant is capable of retaining its structure at this higher temperature and can maintain this for at least 48 h. The stabilized

A100C thermostable mutants showed a less impressive stability at 45 °C for 48 h, suggesting that after 16 hr the protein scaffold begins to denature, which is corroborated by a large buildup of white precipitate.

Reaction Entry	Mutant	Temp °C	Selectivity	
			Linear Aldehyde (%) (average)	TON (average)
1	<b>V83C</b>	35	78.3 (1.21)	75.7 (1.32)
2		45	75.8 (1.03)	27.3 (0.98)
3	<b>V83C E81K S56D</b>	35	79.8 (0.67)	220.8 (1.84)
4		45	79.3 (0.93)	385.2 (3.87)
5	<b>V83C N31D E81K</b>	35	79.6 (1.10)	198.9 (1.98)
6		45	78.4 (0.65)	404.3 (4.19)
7	<b>V83C N31D S56D</b>	35	77.4 (1.9)	203.7 (2.81)
8		45	79.6 (0.83)	401.8 (4.35)
9	<b>A100C</b>	35	78.8 (0.76)	400.9 (3.87)
10		45	70.0 (1.59)	170.9 (2.69)
11	<b>A100C E81K S56D</b>	35	77.3 (1.24)	406.1 (5.21)
12		45	72.5 (0.87)	414.8 (3.86)
13	<b>A100C N31D E81K</b>	35	71.9 (1.31)	403.4 (3.54)
14		45	72.9 (1.26)	325.5 (2.85)

*Table 4. Hydroformylation of 1-octene catalyzed by rhodium-hydroformylase. The conditions used were 80 bar syngas (1:1), stirring 625 rpm, 48 h in degassed MES buffer, 0.5 mL of catalyst solution and 0.5 mL of alkene containing 9% (v/v) n-heptane and 1% (v/v) diphenyl ether. In all examples, the reaction substrate is 1-octene (Scheme 1), the average taken in all examples is of three repeated reactions, standard deviation between brackets. Yields calculated by GC.*

### 3.9 Conclusions and Further Work

In conclusion, it has been shown that the stability of the protein scaffold within an ArM is very important to the catalytic activity. By improving the thermostability of the SCP-2L V83C scaffold by up to 16 °C, the reactivity of the ArM was increased at higher temperatures. With the stabilized hybrid enzymes, the TON improved over 5-fold reaching values of up to over 400 (Table 4). Thus, rational mutant design for optimization of the ArM stability yielded a biocatalyst that enables an increased range of reaction conditions and higher activities. The rhodium phosphine modified ArMs have been shown to be effective linear selective catalysts in hydroformylation reactions and showed a rate enhancement of at least  $10^3$  compared to traditional Rh/TPPTS system in the biphasic hydroformylation of 1-octene and 1-decene.<sup>10</sup>

Moving forwards this approach has the potential to be used for a range of reactions, which traditionally use phosphines as ligands and convert these into biocatalytic processes. Combining mutation points, which both showed an increased stability, was shown to have an almost additive effect to the protein scaffold. In order to continue to increase the proteins stability, the mutant library should be expanded, and a larger number of mutation points combined. By designing an assay that could be used to screen for ArM activity using a large mutant library would be more convenient. This could even be expanded to utilizing bioengineering tools such as directed evolution.

### 3.10 Experimental

#### 3.10.1 General Remarks

Mass Spec measurements were taken on a Waters Alliance HT 2795 equipped with a Micromass LCT-TOF mass spectrometer, using positive electrospray ionisation. UV-visible measurements were recorded on a Molecular Devices M2 Spectra Max spectrophotometer.

#### 3.10.2 Protein Crystal Structures – Protein Crystallization carried out by Lorenz Obrecht and Refinement Carried out by Megan Doble & Emma Branigan

A100C data set from Jan 2011 at 2.48 Angstrom resolution:

A100C was concentrated to ~2.1 mg/ml in 20 mM Pipes pH 7.5, 1 mM EDTA, 1 mM sodium azide, 150 mM NaCl, 0.15 mM Triton X100.

V83C data set from August 2012 at 2.09 Angstrom resolution:

V83C was concentrated to ~2.5 mg/ml in 20 mM Pipes pH 7.5, 1 mM EDTA, 1 mM sodium azide, 150 mM NaCl, 0.15 mM Triton X100. Crystals were grown at 4 °C by hanging drop vapour diffusion in 2.2 M ammonium sulfate, 0.2 M NaCl, 0.1 M citric acid pH 5.4 and cryo-protected with 25 % sucrose.

CCP4 suite and COOT were used for refinement.

<b>Data Collection</b>	<b>A100C</b>	<b>V83C</b>
<b>Space group</b>	P212121	P212121
<b>Cell dimensions</b>		
<b>a, b, c (Å)</b>	35.38, 50.75, 63.10	63.74, 35.69, 48.85
<b><math>\alpha, \beta, \gamma</math> (°)</b>	90.00, 90.00, 90.00	90.00, 90.00, 90.00
<b>Resolution (Å)</b>	50.00 – 2.48 (2.52 – 2.48) <sup>a</sup>	28.82 – 2.09 (2.15-2.09)
<b>Rsym or Rmerge</b>	7.0 (51.5)	5.9 (18.1)
<b>I / sigma</b>	29.4 (2.9)	23.4 (10.7)
<b>Completeness (%)</b>	85.8 (100)	99.1 (96.5)
<b>Redundancy</b>	6.4 (4.5)	6.8 (5.9)
<b>Refinement</b>		
<b>Resolution (Å)</b>	29.04 – 2.48	28.82 – 2.09
<b>No. reflections</b>	3469	6538

<b>Rwork / Rfree</b>	20.7 / 28.3	21.2 / 25.6
<b>No. atoms</b>	929	979
<b>B factors</b>	53	20
<b>r.m.s. deviations</b>		
<b>Bond lengths (Å)</b>	0.017	0.011
<b>Bond angles (°)</b>	1.6	2.1

<sup>a</sup>Values in parentheses are for highest-resolution shell.

### 3.10.3. Site Directed Mutagenesis

Mutations were introduced using the Quik-Change Site-Directed Mutagenesis Kit (Stratagene) with pEHISTEV::dΔhΔSCP-2L as the template. The primers used for the mutagenesis are shown below. The constructs were verified by sequencing by GATC.

#### Primers for SCP-2L A100C & V83C Templates

<b>Mutation</b>	<b>Sequence (5' - 3')</b>
<b>acquired</b>	
<b>I44E</b>	CCAAAGGCGGAAATGAAGGGGCTAAGTGGAC GTCCACTTAGCCCCTTCATTTCCGCCTTTGG
<b>I44D</b>	CCAAAGGCGGAAATGATGGGGCTAAGTGGAC GTCCACTTAGCCCCATCATTTCGCCTTTGG
<b>S56E</b>	GACCTGAAAAGTGGTGAGGGAAAAGTGTACCA TGGTACACTTTTCCCTCACCCTTTTCAGGTC
<b>S56D</b>	GACCTGAAAAGTGGTGACGGAAAAGTGTACCA TGGTACACTTTTCCGTCACCCTTTTCAGGTC
<b>K58E</b>	CTGAAAAGTGGTTCTGGAGAGGTGTACCAAGGCCCTGC GCAGGGCCTTGGTACACCTCTCCAGAACCACTTTTCAG
<b>K58D</b>	CTGAAAAGTGGTTCTGGAGACGTGTACCAAGGCCCTGC GCAGGGCCTTGGTACACGTCTCCAGAACCACTTTTCAG
<b>K65E</b>	GTACCAAGGCCCTGCAGAGGGTGCTGCTGATACAAC GTTGTATCAGCAGCACCTCTGCAGGGCCTTGGTAC
<b>K65D</b>	GTACCAAGGCCCTGCAGACGGTGCTGCTGATACAAC GTTGTATCAGCAGCACCTCTGCAGGGCCTTGGTAC

**N31D** CTGAGGTGGTGAAGAAAGTAGACGCTGTATTTGAGTGG  
 CCACTCAAATACAGCGTCTACTTTCTTCACCACCTCAG

**V26F** GGATATTGGGCCTGAGTTCGTGAAGAAAGTAAA  
 TTACTTTCTTCACGAACTCAGGCCCAATATCC

**G41D** GGCATATAACCAAAGACGGAAATATTGGGGC  
 GCCCAATATTTCCGTCTTTGGTTATATGCC

**S54N** GTGGACTATTGACCTGAAAAATGGTTCTGGAAAAGTGTAC  
 GTACACTTTTCCAGAACCATTTTTCAGGTCAATAGTCCAC

**D116K** CTCAGAGCTTGGCATATTTTTTAAGAATCATCTGAAGTTTCTGGC  
 GCCAGAACTTCAGATGATTCTTAAAAAATATGCCAAGCTCTGAG

**K40N** GTGGCATATAACCAACGGCGGAAATATTGGG  
 CCAATATTTCCGCCGTTGGTTATATGCCAC

**A68P** GCCCTGCAAAGGTGCTCCTGATACAACAATCATAC  
 GTATGATTGTTGTATCAGGAGCACCTTTTGCAGGGC

<b>Mutation acquired</b>	<b>Protein Scaffold</b>	<b>Sequence (5' - 3')</b>
<b>E81K</b>	A100C	CAGATGAAGATTTTCATGAAGGTGGTCTGGGCAAG CTTGCCCAGGACCCTTCATGAAATCTTCATCTG
	V83C	GATGAAGATTTTCATGAAGGTGTGCCTGGGC GCCCAGACGCACCTTCATGAAATCTTCATC

The constructs were prepared as follows: SCP-2L A100C and V83C were used as the template DNA where appropriate. Site directed mutagenesis was carried out. The reaction was set up as follows: Primer fw: 0.01 μM (0.25 μM, 0.5 μL), primer rv: 0.01 μM (0.25 μM, 0.5 μL), SCP-2L template DNA (0.5 μL), PfuUltra Master Mix (10 μL), Sterile MilliQ Water

After mixing, the PCR tubes were spun briefly (4 sec, table top centrifuge, RT). The tubes were transferred to a PCR machine and the following PCR cycle ran.

1. Initial Denaturation 95 °C, 5 min
2. Denaturation 30 sec 95 °C
3. Annealing 1min 55 °C
4. Extension 1min/1kp 72 °C
5. Repeat Steps 2-4 for 16 cycles

#### 6. Final extension 20 min 72 °C

Digestion of the wild type template strand was then performed as follows: PCR product (8.5  $\mu$ L), 10x Fast digest buffer (1  $\mu$ L) and Dpn1 (0.5  $\mu$ L) were incubated at 37 °C for 10 min, followed by inactivation at 72 °C for 5 min. The digested products were transformed into commercial DH5 $\alpha$  cells and grown on kanamycin (50  $\mu$ g/mL) agar plates at 37 °C overnight.

### 3.10.4 Protein Expression and Purification

#### General Expression and Purification Protocol:

Protein expression and purification was performed as described in literature<sup>6</sup>

SCP-2L A100C was prepared by transforming the pEHISTEV::d $\Delta$ h $\Delta$ SCP-2L A100C plasmid into competent *E. coli* Rosetta(DE3) cells. A single transformed colony was inoculated into 10 mL PB medium (Production Broth medium; containing 20 g/L tryptone, 10 g/L yeast extract, 5 g/L dextrose, 5 g/L NaCl, 8.7 g/L K<sub>2</sub>HPO<sub>4</sub>, pH 7.0) with 50  $\mu$ g/mL kanamycin and 34  $\mu$ g/mL chloramphenicol (in ethanol). This was used to inoculate 100 mL PB media (50  $\mu$ g/mL kanamycin and 34  $\mu$ g/mL chloramphenicol) and agitated at 37 °C, 200 rpm overnight (16-18 h). This starter culture (10 mL) was used to inoculate 0.5 litre of PB medium (2 litres in total). The cells were allowed to grow at 37 °C to an OD<sub>600</sub> of 0.6 (about 2 h) after which the temperature was lowered to 16 °C. Isopropyl-1-thio- $\beta$ -D-galactopyranoside (IPTG) was added after one hour to a final concentration of 0.2 mM from a 0.2 M stock solution to initiate the expression of the recombinant protein. The culture was left overnight at 16 °C (16-18 h). The cells were harvested by centrifugation (20 min, 7500 g), washed with 250 mL of buffer solution (16 mM kPi (potassium phosphate) 120 mM NaCl, pH 7.4) and centrifuged again (20 min 7500 g). The pellet from the 2 litre cell culture was resuspended at 4 °C in 100 ml of buffer solution (50 mM Tris·HCl, 20 mM imidazole, 150 mM NaCl, 0.5 mM benzamidine, pH 8) and frozen at -80 °C.

After defrosting, 20 mg of lysozyme, 1 mg of DNase and 1 mL of 1 M of MgCl<sub>2</sub> was added to the suspension. The suspension was left for 1 hour at 4 °C and was sonicated in portions of 10 mL (Hielscher UP200S ultrasonic processor, 0.5 s pulses 90% power) with 0.5 second pulses for 1 minute. The cell-extract, obtained after centrifugation (50 min, 30000 g) was filtered

(0.22  $\mu\text{m}$  PES Millex Filter unit) and applied to a nickel column (5 mL HisTrap FF) equilibrated with 30 mM Tris·HCl, 20 mM imidazole, 150 mM NaCl, pH 8 (wash buffer) at flow rates up to 5 mL/min. The column was washed with 5 column volumes of wash buffer, 5 column volumes of high salt buffer (wash buffer containing 1 M NaCl) and another 5 column volumes of wash buffer. The protein was obtained by eluting with 6 column volumes elution buffer (wash buffer containing 330 mM imidazole) into an equal volume of wash buffer to prevent precipitation of the protein and the resulting solution was dialysed against 5 litre buffer solution (30 mM Tris·HCl, 10 mM imidazole, 150 mM NaCl, pH 8 at 4°C). If precipitate had formed the resulting mixture was centrifuged (30000 g, 50 min). 0.014 equivalents of TEV-protease and final concentrations of 1 mM DTT and 0.5 mM EDTA were added to the resulting His-tagged SCP-2L A100C solution. This mixture was left overnight at room temperature.

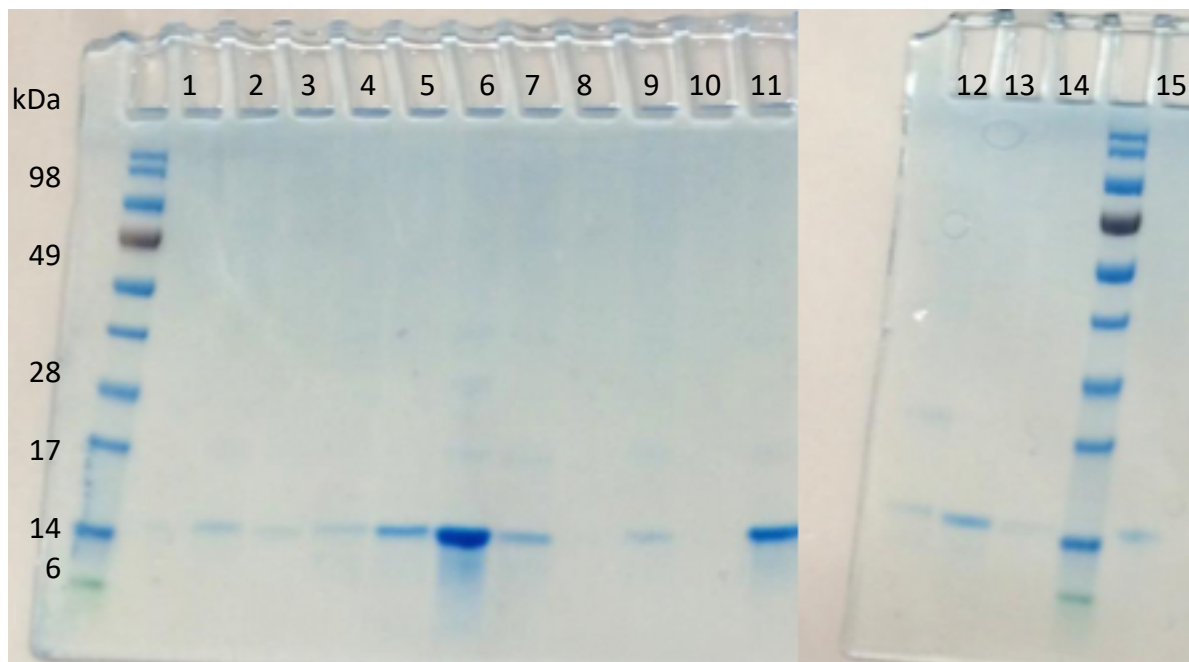
If precipitate had formed the mixture was again centrifuged (50 min, 30000 g) and the pellet was discarded. The supernatant was then filtered (0.22  $\mu\text{m}$  PES Millex Filter unit) and used for nickel affinity chromatography on a nickel column (5 mL HisTrap FF) equilibrated with wash buffer. The flow-through containing pure SCP-2L A100C was collected, concentrated and the buffer was exchanged to the storage and coupling buffer (20 mM MES, 50 mM NaCl, pH 6) using a centrifugal concentrator. Three consecutive rounds of buffer exchange were performed to remove all the DTT. Buffer exchange by centrifugal concentrator was found to be more effective than dialysis. Yields up to 75 mg L<sup>-1</sup> culture, with a purity of more than >99% according to SDS-PAGE (Nupage™ 4-12% Bis-Tris gel), were obtained.

When expressing and purifying protein on a small scale for CD analysis the method was adjusted onto a smaller scale using 500 ml culture for each protein. The only difference was that instead of using a HiTrap pre packed column, the protein supernatant was incubated with Ni-NTA resin for 1 hour and then eluted using a spin column. The Ni-NTA resin used had a binding affinity of 50 mg/ml. Expression was successful for all mutants with an average yield of 20 mg L<sup>-1</sup> culture. This slight decrease in yield is likely due to the accelerated purification protocol.



### SDS Gel Characterisation:

SDS page gel (Nupage™ 4-12% Bis-Tris gel):



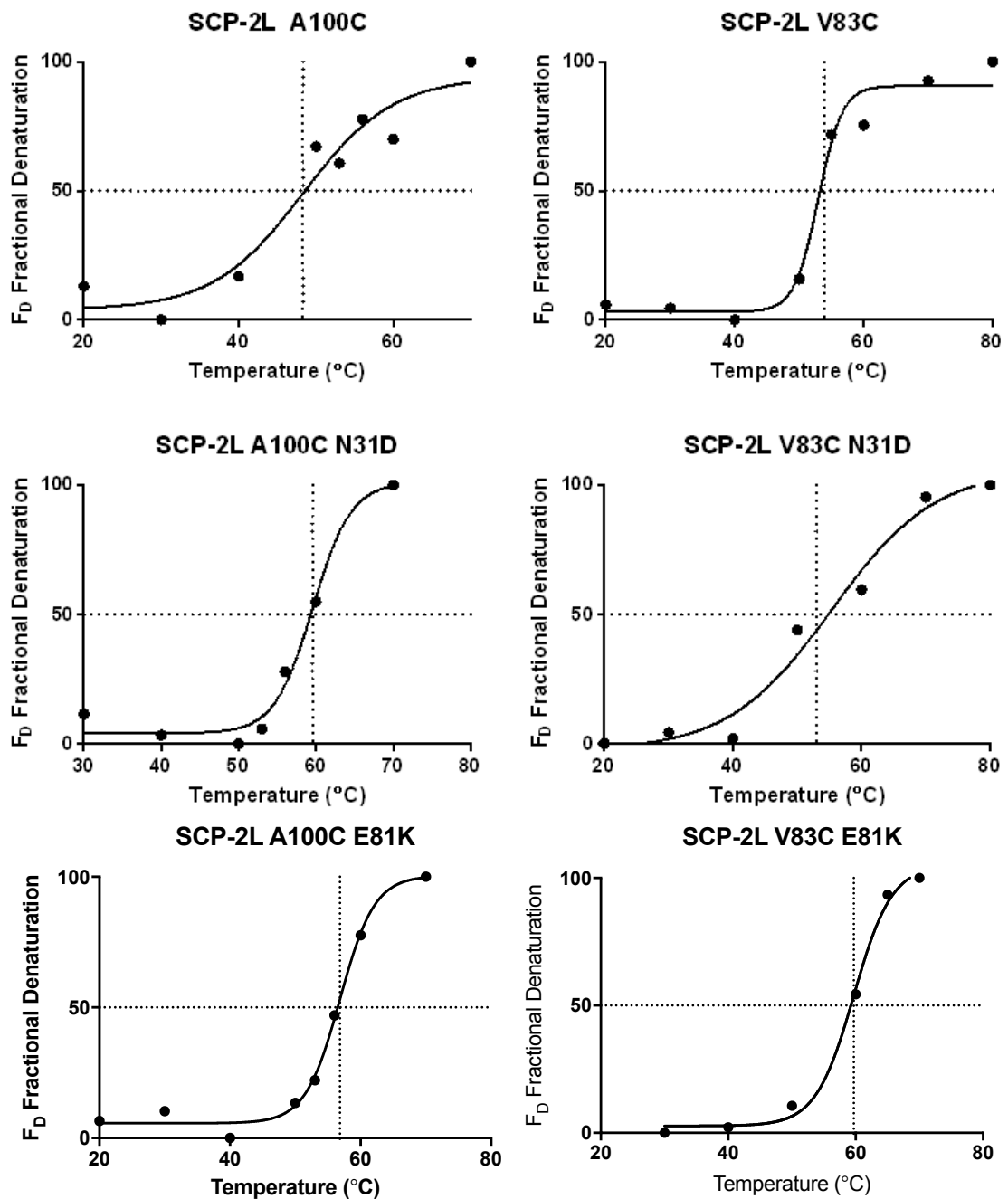
L = SeeBlue Plus 2 Prestained Protein Standard ladder

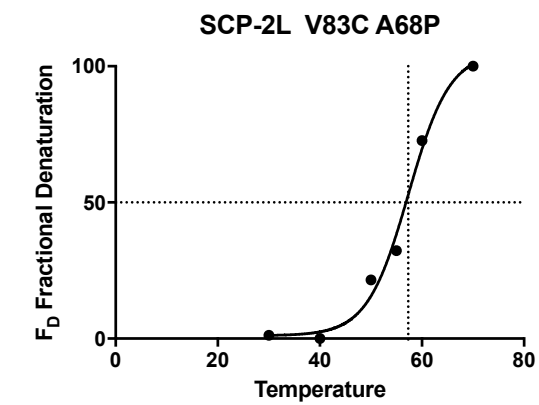
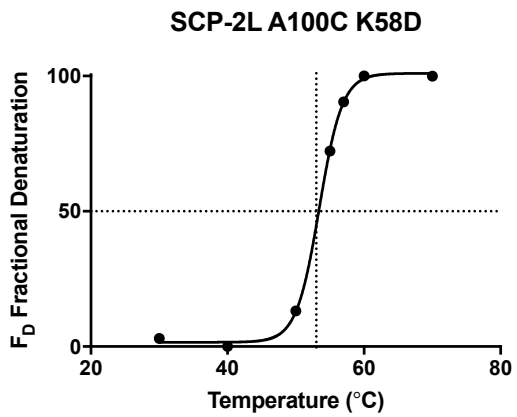
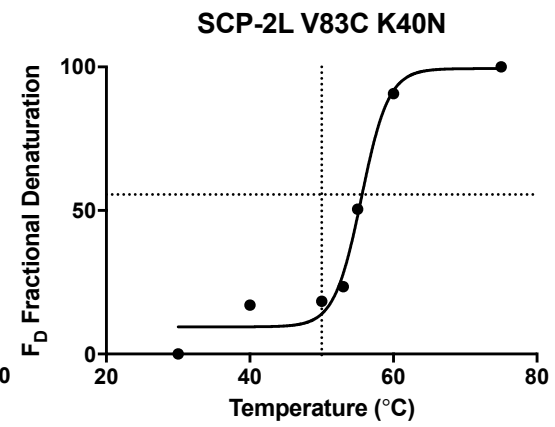
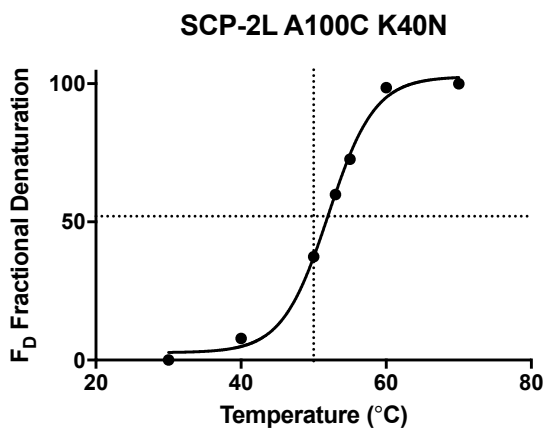
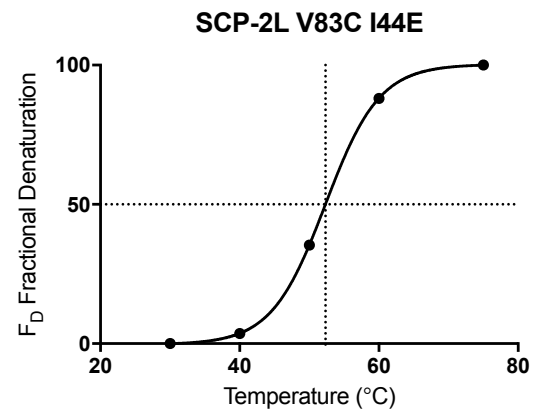
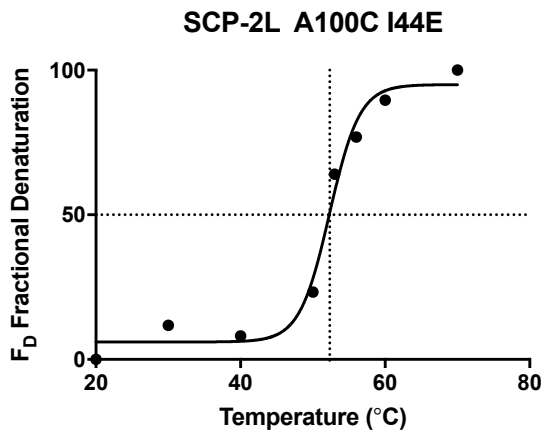
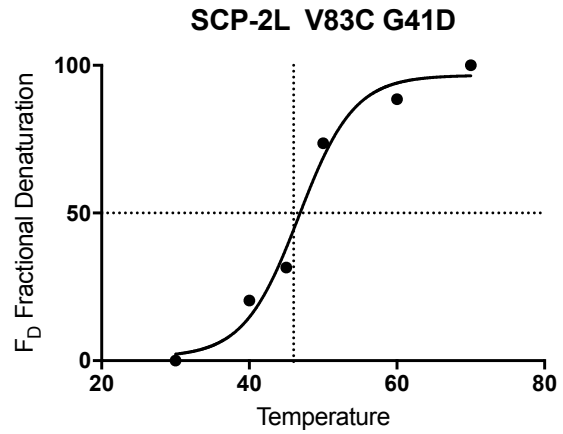
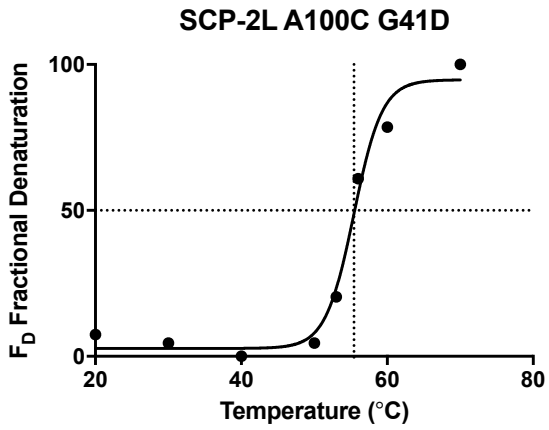
- |               |                |
|---------------|----------------|
| 1. A100C      | 9. A100C K58E  |
| 2. V83C       | 10. A100C I44D |
| 3. A10C N31D  | 11. A100C K58D |
| 4. A100C S56E | 12. A100C K65E |
| 5. A100C E81K | 13. A100C S56D |
| 6. A100C S54N | 14. A100C K40N |
| 7. A100C G41D | 15. A100C V26F |
| 8. A100C K65D |                |

#### 3.10.5 CD Studies to determine thermostability

Circular dichroism spectra were obtained using a Biologic MOS-500 spectrometer. A Xe lamp was used, and the near UV spectra taken from 190 – 260 nm using a 10 mm pathlength cell. The other parameters were as follows: 0.25 nm step, 0.5 s acq period, 3 repeats,  $\pm$ MD 30 and slit size 2 nm. Thermal denaturation curves were obtained by plotting the CD values at 222 nm of each spectrum against temperature. The thermal melt curves were fitted using the Boltzman sigmoidal fit function of Origin 6.0.

Using circular dichroism (CD) spectroscopy an approximate  $T_{mapp}$  was measured and compared SCP-2L A100C and V83C encoded by the original plasmid. The CD method used a fast scan approach only measuring 6 different temperature points; therefore, the  $T_m$  values are approximate and serve the purpose of indicating how the thermostability is affected by the new amino acid introduced (Figure 6).





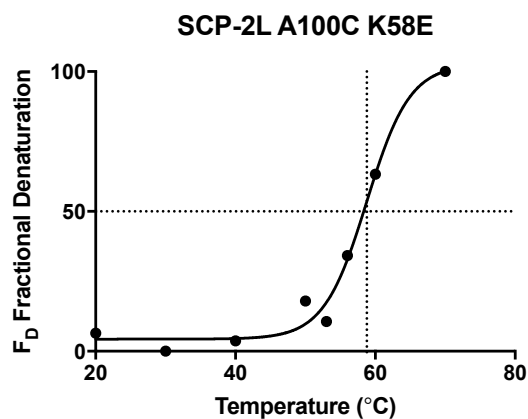


Figure 6 - Using CD spectroscopy, the accurate melting temperature ( $T_m$ ) of V83C and A100C and the thermostable double mutants were measured using data points taken at every degree point at 222 nm

### 3.10.6 Protein Modification

All reactions involving phosphines were performed under an argon atmosphere using degassed solvents and standard Schlenk techniques. Buffer solutions were degassed by bubbling argon through the solution. Protein solutions were degassed by washing and concentrating three times with degassed buffer using a filter unit set up as shown in Figure 5. Protein concentrations were determined by Bradford's reagent.

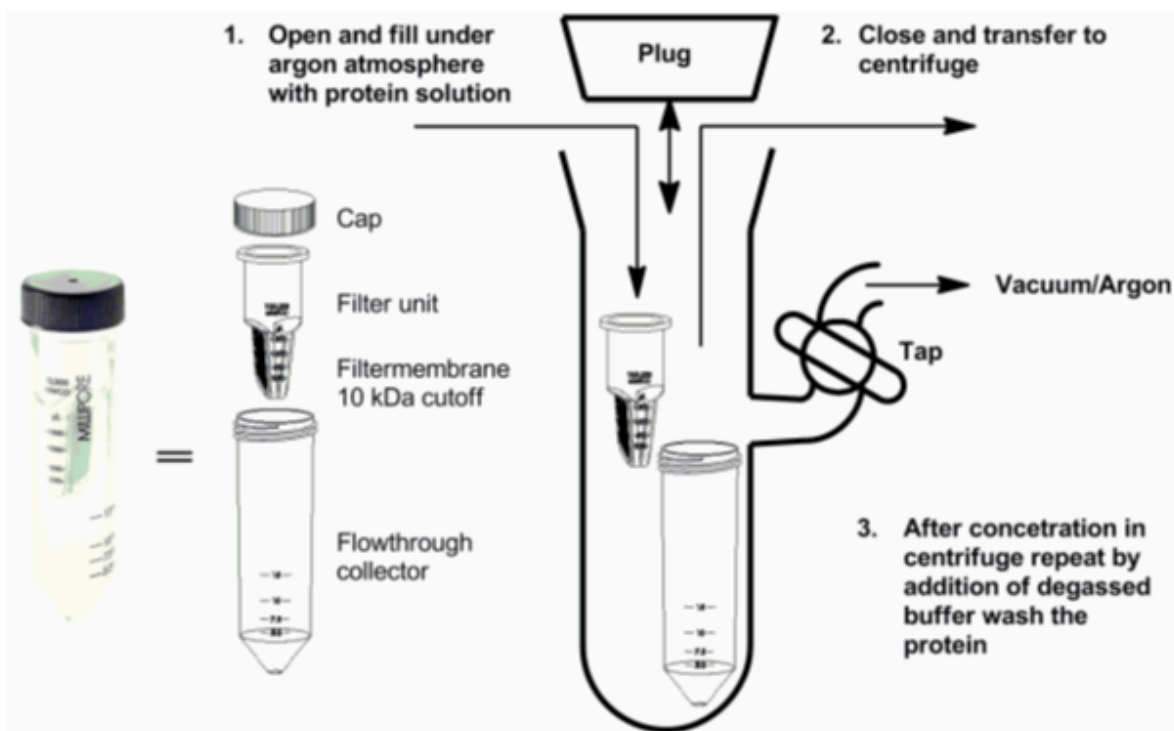


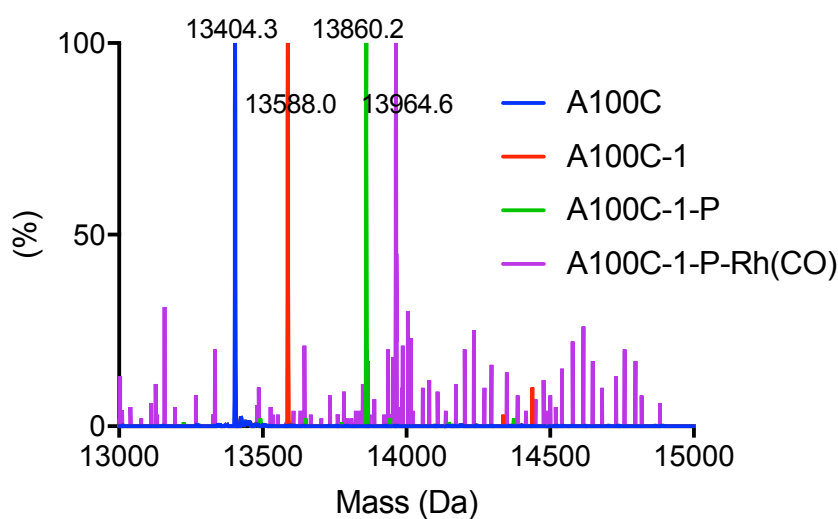
Figure 5: Application of Amicon® Ultra-15 Centrifugal Filter Unit under inert atmosphere using a large custom made Schlenk flask.<sup>10</sup>

Modification of the protein scaffold was carried out according to literature procedures. For the rhodium complex formation, the Rh:P ratio aimed for depended on application. For characterisation a 1:1 ratio was used. For catalysis an excess of phosphine, approximately 2 equiv. per Rh was used to ensure that no 'free Rh' is present. Free Rh-catalysed hydroformylation is characterised by high TON's and low linear to branched selectivities. Slight inconsistencies in protein concentration, through loss in handling or oxidation of the phosphine have the potential to lead to excess Rh over phosphine if a 1:1 ratio is used. Rh(acac)(CO)<sub>2</sub> was used as the rhodium source, and added as a freshly made solution in DMF.

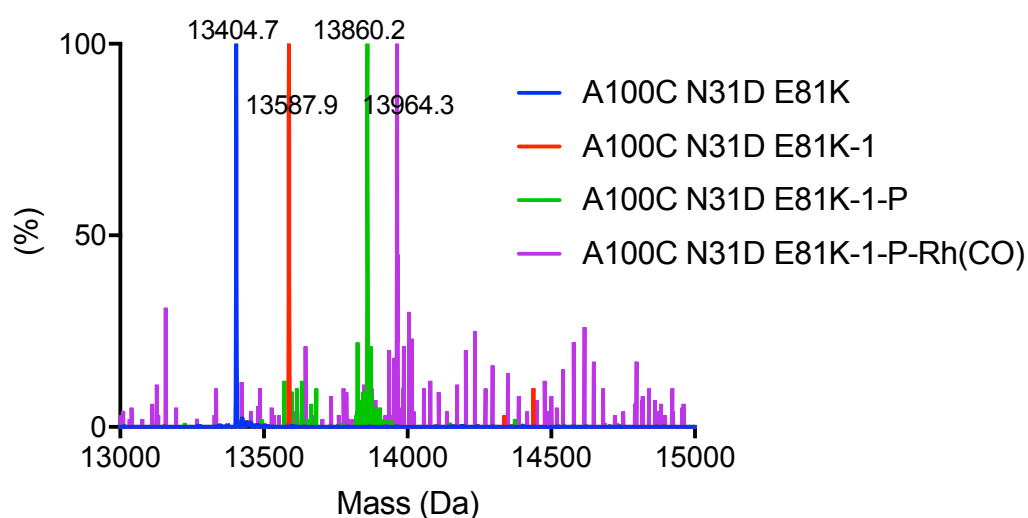
### 3.10.7 Mass Spectrometry of the modified proteins

LC-MS(ES<sup>+</sup>) used for analysis of protein and modification reactions was carried out on a Waters Alliance HT 2795 equipped with a Micromass LCT-TOF mass spectrometer, using positive electrospray ionisation and applying a Waters MASSPREP® On-line Desalting 2.1x10 mm cartridge using a gradient of 1% formic acid in H<sub>2</sub>O to 1% formic acid in acetonitrile. ESI-MS results were analysed by MassLynx V. 4.0 and its MaxEnt algorithm. The samples for

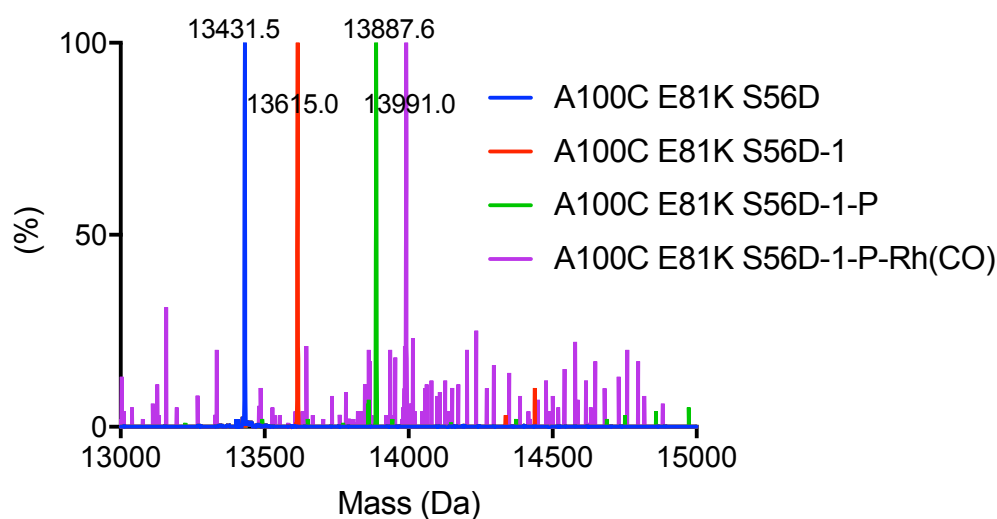
phosphine modified proteins were prepared immediately before injection to reduce the oxidation of the sample.



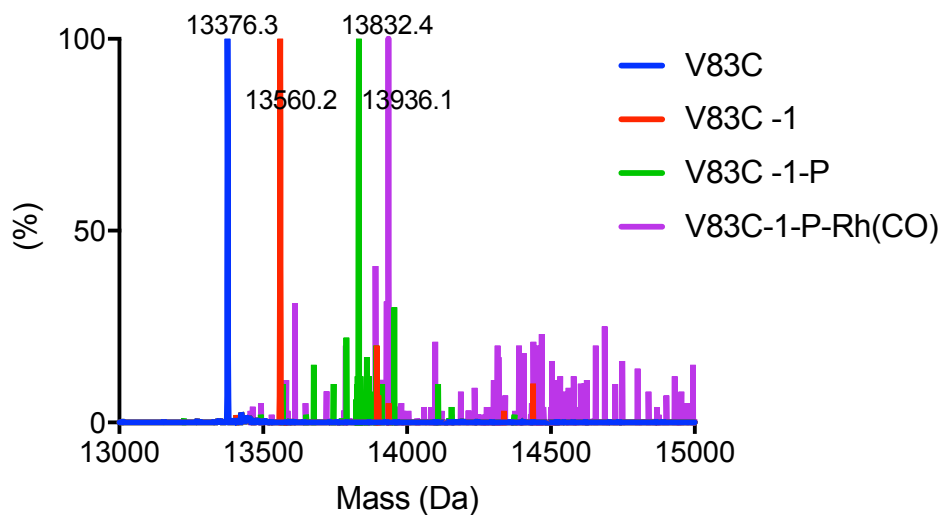
Deconvoluted mass-spectrum of modified proteins Blue A100C 13404.3Da (Calc.13404.6 Da): Red A100C-1 13588.0 Da (Calc. 13587.9 Da): Green A100C-1-P 13860.2 (Calc. 13860.2 Da): Purple A100C-1-P-Rh(CO) 1396462 Da (Calc. 13964.1 Da) obtained by LC-MS (ESI<sup>+</sup>)



Deconvoluted mass-spectrum of modified proteins Blue A100C N31D E81K 13404.7 Da (Calc.13404.7 Da): Red A100C N31D E81K-1 13587.9 Da (Calc. 13587.9 Da): Green A100C N31D E81K-1-P 13860.1 (Calc. 13860.2 Da): Purple A100C N31D E81K-1-P-Rh(CO) 13964.3 Da (Calc. 13964.1 Da) obtained by LC-MS (ESI<sup>+</sup>)



Deconvoluted mass-spectrum of modified proteins Blue A100C E81K S56D 13431.5 Da (Calc.13431.7 Da): Red A100C E81K S56D -1 13615.0 Da (Calc. 13614.9 Da): Green A100C E81K S56D-1-P 13887.6 (Calc. 13887.2 Da): Purple A100C E81K S56D-1-P-Rh(CO) 13991.0 Da (Calc. 13991.1 Da) obtained by LC-MS (ESI<sup>+</sup>)



Deconvoluted mass-spectrum of modified proteins Blue V83C 13376.3 Da (Calc.13376.6 Da): Red V83C-1 13560.2 Da (Calc. 13559.8 Da): Green V83C-1-P 13832.4 (Calc. 13832.1 Da): Purple V83C-1-P-Rh(CO) 13936.1 Da (Calc. 1336.01 Da) obtained by LC-MS (ESI<sup>+</sup>)

### 3.10.8 Hydroformylation

#### General Hydroformylation Procedure:

Hydroformylation reactions were carried out in stainless steel autoclaves (up to two simultaneously) containing up to eight glass reaction vials (volume of approximately 5 mL) each. The reaction vials were equipped with stirring bars as well as a septum cap pierced with a needle in order to allow contact with the reaction gases. The prepared vials were placed in an autoclave which was flushed three times with >20 bar argon before adding the chemicals. The vials were charged under an argon blanket. The aqueous Rh-enzyme solution was added to each vial (0.5 mL, 20 mM MES, 50 mM NaCl, pH = 6). A protein-P : Rh ratio of 2 : 1 was aimed for and an aliquot of this solution was saved for ICP-MS analysis to obtain the actual Rh concentration. An organic solution consisting of 1-alkene, 9% (v/v) heptane and 1% (v/v) diphenyl ether as internal standards, freshly filtered over activated alumina (dried under vacuum at 150 °C for at least 5 hours) to remove peroxides, was added (0.5 mL). The autoclave was then flushed 3 times with about 20 bar syngas (H<sub>2</sub> : CO 1 : 1) and subsequently charged to the desired pressure (80 bar). The autoclave was then placed into an oil bath that was preheated to the desired temperature (35 °C) and pre-set to the desired stirring speed (625 rpm). After the reaction time (48 h) the reaction was stopped by putting the autoclave on ice for at least 15 mins and slowly releasing the pressure. The organic phase was analysed by GC, and ICP-MS for metal leaching. All experiments were run in triplicate.

#### Gas Chromatography:

An Agilent 7820A GC system with autosampler was used for GC analysis. This GC was installed with an Agilent HP-5 column 30m x 0.32 mm x 0.25 µm. Conditions: injector temperature 250 °C; FID detector temperature 300 °C; 6.5 mL/min constant flow; 75:1 split ratio; 1 µL injection; oven method: 25 °C isotherm for 6 min, 10 °C/min to 60 °C, 20 °C/min to 300 °C. Retention times: heptane 2.05 min, 1-octene 4.45 min, 2-methyloctanal 11.34 min, *n*-nonanal 11.83 min, diphenyl ether 14.19 min. Due to the large excess of 1-alkene, resulting in the broad tailing alkene peak, it was not possible to detect small amounts of isomerization products.



This hydroformylation work is an extension of previously published work by Dr Amanda Jarvis, so all controls and hydroformylation using Rh/TPPTS carried out by her can be seen in literature.<sup>10</sup>

- 
- <sup>1</sup> M. Fuchs, J. E. Farnberger, W. Kroutil. *Eur. J. Org. Chem.* **2015**, 6965-6982
- <sup>2</sup> R. N. Patel, *Biomolecules.* **2013**, 3, 741
- <sup>3</sup> J. C. Moore, D. J. Pollard, B. Kosjek, P. N. Devine, *Acc. Chem. Res.*, **2007**, 40, 1412
- <sup>4</sup> M. V. Doble, A. C. Ward, P. J. Deuss, A. G. Jarvis, P. C. J. Kamer. *Bioorg. Med. Chem.*, **2014**, 22, 5657
- <sup>5</sup> F. Rosati, G. Roelfes. *ChemCatChem*, **2010**, 2, 916
- <sup>6</sup> P. J. Deuss, R. den Heeten, W. Laan, P. C. J. Kamer, *Chem. Eur. J.*, **2011**, 17, 4680
- <sup>7</sup> T. Heinisch, T. R. Ward, *Curr. Opin. Chem. Biol.*, **2010**, 14, 184
- <sup>8</sup> T. R. Ward, *Acc. Chem. Res.*, **2011**, 44, 47
- <sup>9</sup> M. Durrenberger, T. Heinisch, Y. M. Wilson, T. Rossel, E. Nogueira, L. Knorr, A. Mutschler, K. Kersten, M. J. Zimbron, J. Pierron. *Angew Chem Int Ed*, **2011**, 50, 3026-3030
- <sup>10</sup> A. G. Jarvis, L. Obrecht, P. J. Deuss, W. Laan, E. K Gibson, P. P. We;;s, P. C. J. Kamer, *Angew. Chem. Int. Ed.* **2017**, 56, 13784
- <sup>11</sup> P. J. Deuss, G. Popa, A. M. Z. Slawin, W. Laan, P. C. J. Kamer, *ChemCatChem*, **2013**, 5, 1184
- <sup>12</sup> H. J. Wijma, R.J. Floor, D. B. Janssen. *Curr. Opin. Struct. Biol*, **2013**, 23, 588
- <sup>13</sup> H. Arabnejad, M. Dal Lago, P. A. Jekel, R. J. Floor, A. W. H. Thunnissen, A. C. Terwisscha van Scheltinga, H. J. Wijma, D. B. Janssen. *Protein Eng Des Sel*, **2017**, 30, 173
- <sup>14</sup> A. Cerdobbel, K. De Winter, D. Aerts, R. Kuipers, H. Joosten, W. Soetaert, T. Desmet. *Protein Engineering, Design and Selection*, **2011**, 24, 829
- <sup>15</sup> F. H. Arnold, *Trends in Biotech*, **1990**, 8, 244
- <sup>16</sup> H. Yu; H. Huang, *Biotechnology Advances*, **2014**, 32, 308-315
- <sup>17</sup> V. G. Eijsink, A. Bjork, S. Gaseidnes, R. Sirevag, B. van den Burg. *J Biotechnol*, **2004**, 113, 105
- <sup>18</sup> M. F. Perutz, *Science.* **1978**, 201, 1187
- <sup>19</sup> J. Chen, H. Yu, C. Liu, J. Liu, Z. Shen, *J Biotechnol.* **2012**, 164, 354
- <sup>20</sup> A. Zarrine-Afsar, Z. Zhang, K. L. Schweiker, G. I. Makhatadze, A. R. Davidson, H. S. Chan. *Proteins*, **2012**, 80, 858
- <sup>21</sup> C. F. Lee, G. I. Makhatadze, K. B. Wong. *Biochemistry.* **2005**, 44, 16817

- 
- <sup>22</sup> K. Steiner, H. Schwab, *Computational and Struct Biotech Journal*, **2012**, 2, 1
- <sup>23</sup> P. J. Deuss, G. Popa, C. H. Botting, W. Laan, P. C. J. Kamer, *Angew. Chem. Int. Ed.* **2010**, 49, 5315
- <sup>24</sup> A. M. Haapalainen, D. M. van Aalten, G. Meriläinen, J. E. Jalonen, R. K. Wierenga, J. K. Hiltunen, T. Glumoff, *J. Mol. Biol.*, **2001**, 313, 1127
- <sup>26</sup> H. Bahrmann, S. Bogdanovic, P. W. N. M. van Leeuwen, *Aqueous- Phase Organometallic Catalysis* **2004**, 391
- <sup>27</sup> H. J. Joosten, NL Patent, 35970, 2008
- <sup>28</sup> M. F. Lensink, A. M. Haapalainen, J. K. Hiltunen, T. Glumoff, A. H. Juffer. *J. Mol. Biol*, **2002**, 323, 99-113
- <sup>29</sup> F. V. Filipp, M. Sattler, *Biochemistry*, **2007**, 46, 7980-7991
- <sup>30</sup> P. Carter, *Biochem. J.*, **1986**, 237, 1
- <sup>31</sup> N. J. Greenfield, *Nat. Protoc.*, **2006**, 1 (6), 2527
- <sup>32</sup> H. J. Wijma, R. J. Floor, P. A. Jekel, D. Baker, S. J. Marrink, D. B. Janssen, *Protein Eng. Des. Se.I*, **2013**, 27 (2), 49
- <sup>33</sup> *Rhodium Catalysed Hydroformylation* (P. W.N.M. eds. Van Leeuwen, C. Claver) (Kluwer Academic Publishers, New York 2002).

## Artificial Metalloenzymes as Catalysts for Oxidative Lignin Degradation

### Chapter 4 – Artificial Metalloenzymes as oxidation catalysts for lignin degradation<sup>1</sup>

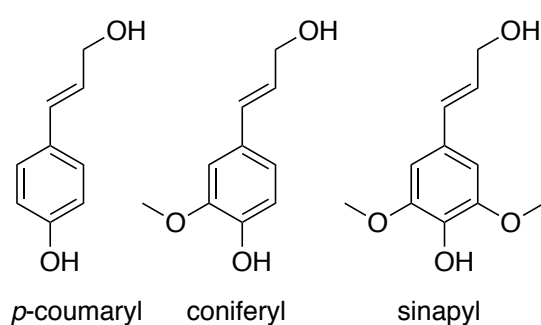
#### 4.1 Introduction to lignin

The rapid depletion of fossil fuels as a resource for energy and chemicals production necessitates a future change to an increase in the use of biomass, ideally through the development of biorefineries that provide a wide range of products. The most abundant renewable resource in this context is lignocellulosic biomass, which is much more plentiful than alternative feedstocks such as sugars, starch, oils and fats.<sup>2</sup> Another advantage is that it does not compete with food sources for land. Lignocellulose, consisting of 40-50% cellulose, 16-33% hemicellulose and 15-30% lignin,<sup>3</sup> is therefore viewed as an abundant future source of fuel and chemicals.<sup>4,5</sup> Whilst the technologies for the exploitation of cellulose and hemicellulose are well established, lignin is currently underutilised with only one commercial process in operation – the preparation of vanillin.<sup>6</sup> The major hurdle for the complete exploitation of lignocellulosic biomass, including lignin, is the efficient depolymerisation of lignin.<sup>7,8</sup> In the biorefinery, cellulose and hemicellulose can provide sugars, alcohols and alkane based products, whilst lignin has the potential to be the major natural source of aromatic compounds to replace oil. Therefore, novel methods to overcome the depolymerisation limitation are required.

##### 4.1.1 Structure and Functionality of Lignin

Lignin is a naturally occurring aromatic non-ordered polymer (Figure 1), which fills the spaces in the cell wall between cellulose and hemicellulose components. Plants obtain their structural integrity by utilising lignin as a resin. The biological design of lignin protects plants from attack by microorganisms, and this helps to prevent hydrolysis of cellulose. This complex three-dimensional polymer is covalently linked to hemicellulose and gives mechanical strength to the cell wall. It is largely hydrophobic in comparison to cellulose, which is hydrophilic, so is intertwined by hemicellulose, which contains both hydrophobic and hydrophilic sections. The distribution of this polymer within the cell wall is not uniform, and the amount of lignin present within a plant also varies dependent on species.

The polymer we refer to as lignin is regarded as the structure after extraction, and protolignin is lignin when still associated with cells. The exact structure of protolignin is still unknown as is its biosynthesis, however it is thought to involve the polymerization of the monomeric units; *p*-coumaryl, coniferyl and sinapyl alcohols (Figure 1).<sup>9</sup> Softwood and hardwood lignin have different compositions based upon the abundance of each of the monolignols. The coniferyl alcohols make up 90% of softwood lignin, whereas comparatively the amount of coniferyl alcohols is similar to the sinapyl alcohols in hardwood lignin.<sup>2</sup> This polymerization is thought to occur through an oxidative coupling reaction via radicals, in nature these would be generated by a peroxidase using H<sub>2</sub>O<sub>2</sub> as an oxidant.<sup>10</sup>



*Figure 1 – The three monolignols, the building blocks of lignin*

Structural studies are complex and limited because different extraction techniques to isolate lignin are known to alter protolignin in different manners. Additionally, the determination of the structure of protolignin itself has so far not been successful. What is known about this structure is, it is comprised of a number of varying linkages that connect together different structural units (Figure 2), these include  $\beta$ -O-4,  $\beta$ -5,  $\beta$ - $\beta$ , 5-5, 4-O-5 and  $\beta$ -1 couplings.<sup>11</sup> The  $\beta$ -O-4, arylglycerol- $\beta$ -aryl ether, is the most abundant linkage found in lignin with approximately 50% in softwood and 80% in hardwood. This is thought to be because the phenolic and  $\beta$ -hydrogen are shared between all three monolignols, so will be involved in a large number of the linkages.

The ability to structurally characterise lignin is a key area of development. This will become crucial in order to benefit from lignin's natural aromatic structure and produce aromatic chemicals.

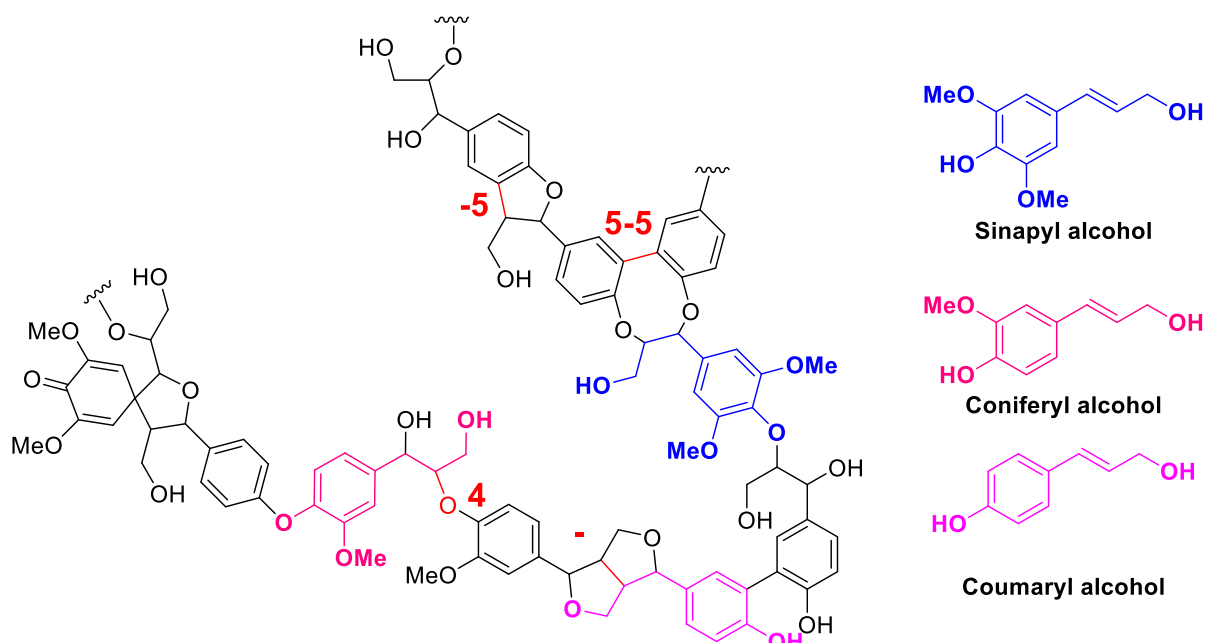


Figure 2 – Representation of lignin type structure highlighting the linkages and monolignols

#### 4.1.2 Naturally occurring lignin degradation

Lignin is predominately resistant to most microorganisms' attack, and any degradation that is seen is very slow. The white rot fungus *Phanerochaete chrysosporium*, is a naturally occurring organism that can cleave lignin, and is the only known organism that can completely degrade lignin to carbon dioxide and water.<sup>12</sup> White rot fungi are named as such due to the specific oxidative bleaching of wood. These fungal organisms utilise four main enzymatic groups for the degradation of lignin, three of these are peroxidases and one is a laccase.

The most extensively studied lignin degrading microorganisms are *Phellinus sp.*, *Daedalea sp.*, *Trametes versicolor* and *Pycnoporus coccineus* as these species possess high ligninolytic activity. In general, the lignin breakdown for fungal species was approx. 20-30% over a 60-day period.<sup>13</sup> This is a very slow rate to apply the biocatalysts on an industrial scale, so the use of these species is unachievable. However, the mechanism of action of white rot fungus can be used as inspiration for a solution.

While a large amount of the present research focuses on fungal lignin degradation, various bacterial species have also been shown to metabolize lignin.<sup>14, 15, 16, 17</sup> However, bacterial systems are often less powerful as oxidation catalysts in comparison to fungal systems.<sup>18</sup> At present research is being carried out looking at microbial systems that can be used to

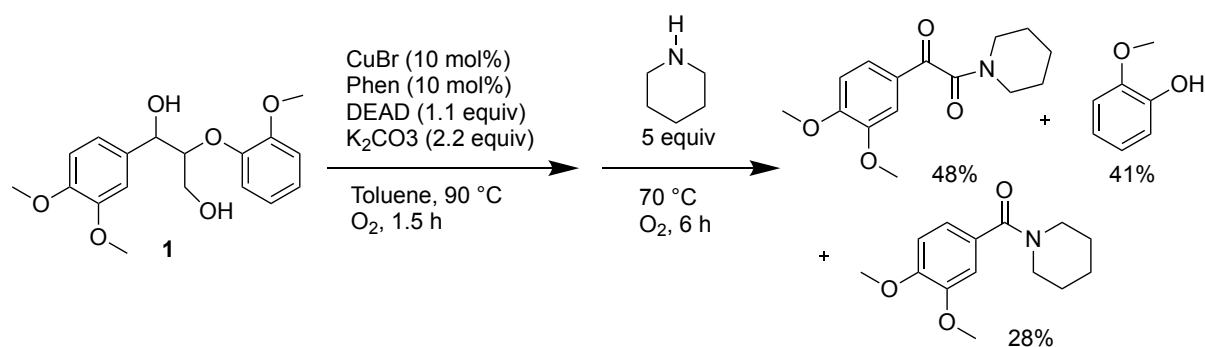
breakdown lignin and could possibly provide molecular information on lignin depolymerisation. Developments in physiological and structural studies of these organisms could lead to identify chemical transformations on lignin.<sup>18</sup>

#### 4.1.3 Transition Metal Catalysts for Lignin Oxidation

At present there are many transition metal catalyzed oxidation methods known for lignin and lignin model compound valorization.<sup>2,19,20</sup>

Currently there is research utilising Methyltrioxorhenium (MTO) as a catalyst for lignin oxidation. MTO is an organometallic compound that contains a Rhenium atom with a tetrahedral geometry with one methyl and three oxo ligands. MTO can activate oxidizing species such as hydrogen peroxide,<sup>21</sup> and MTO-activated hydrogen peroxide is able to oxidize challenging substrates. This MTO-activated peroxide system was shown to have the ability for oxidative cleavage of lignin and a large array of lignin model compounds. MTO was able to oxidize both phenolic and non-phenolic lignin model compounds through side chain oxidations and ring cleavage reactions.<sup>22</sup> Work was also carried out on an MTO system to increase the lifetime by using a polymeric support. This method used polystyrene or poly(4-vinylpyridine) beads showing slightly improved results<sup>23</sup>, however yields and selectivity's were still rather low. In 2014 the group of Herrmann continued this work further looking at methyldioxorhenium (MDO). The MDO is generated *in situ* by reduction of MTO and showed to remain stable with no loss of activity over five cycles. The methyldioxorhenium catalysts were capable of C-O bond cleavage on a range of lignin model compounds.<sup>24</sup> This stability is impressive and shows great potential, but research is needed to see if it will hold the same reactivity when used on lignin biomass.

Methodology for the conversion of lignin  $\beta$ -O-4 linkage models is also being developed using Cu-catalysed aerobic amide bond formation (Scheme 1). This approach utilizes selective oxidation of the secondary alcohol of the  $\beta$ -O-4 link and then amide bond formation using a copper catalyst.<sup>25</sup> Further investigation is required to see whether this system can be applied to natural lignin.

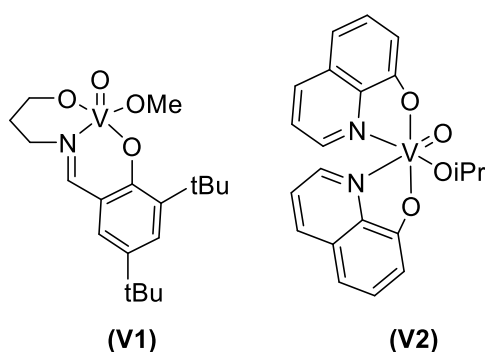


*Scheme 1- One-pot conversion of lignin  $\beta$ -O-4 models through copper catalyzed aerobic amide bond formation<sup>25</sup>*

Another catalyst system being explored for use in lignin oxidation are salen complexes. These complexes are composed of transition metals and are often used with hydrogen peroxide to oxidize a wide variety of substrates.<sup>26</sup> Salen ligands are Schiff bases, synthesized through the condensation of a salicylaldehyde with an amine. The oxidative mechanism of salen complexes was determined through EPR studies. This proceeds via initial formation of a phenoxy radical, which reacts with molecular oxygen to form oxidized lignin model compounds.<sup>27</sup> [Co(salen)] showed a high conversion in a variety of lignin model compounds, although the substrates used are predominately very simple models. An example using salen = [N,N<sup>1</sup>-bis(salicylidene)ethane-1,2-diaminato] and a slightly more realistic  $\beta$ -O-4 model compound was also reported.<sup>33</sup> Investigation using EPR showed that two phenoxy cobalt radicals are utilized in the mechanism of this oxidation reaction.<sup>33</sup> The Co(salen) compounds are often cheaper than alternatives such as biomimetic metalloporphyrin complexes, because they are much simpler to synthesize and often more stable. Modifications of the salen ligand can be used to alter properties of the catalyst for example, such as the addition of sulfonato groups to the salen ligand to increase the solubility.<sup>28</sup>

Catalytic aerobic oxidation is of great interest because the only side product would be water, and could be a cheap alternative to lignin oxidation. The use of vanadium catalysts in aerobic conditions has been studied.<sup>34</sup> Interestingly, depending on the vanadium catalyst (Fig 3), the oxidation proceeded either via C-C cleavage or C-O cleavage. Toste's catalyst (V1) breaks the C-O bond via cleavage of the benzylic C-H bond, whereas the quinolinate complex (V2) breaks the C-C bond. These vanadium complexes showed promising reactivity in oxidation of lignin model compounds, but it is thought that the mechanistic complexity and the number of steps

make them difficult to optimize. Recent work also showed that the vanadium catalysts could be applied to lignin samples and shows  $\beta$ -O-4 cleavage via nonoxidative depolymerisation.<sup>29</sup>



*Fig 3-Vanadium catalysts for the oxidation of lignin models*

Alongside research into organocatalysts for lignin oxidation, the group of Stahl also looked at metal catalysts for lignin oxidation.<sup>30</sup> Chemo selectivity was seen in a few instances for oxidation of the secondary benzylic alcohol, but oxidation of the primary aliphatic alcohol was only observed alongside the secondary alcohol. Often the initial oxidation product of the primary aliphatic alcohol is also not seen because the reaction is followed by retro-aldol bond cleavage.<sup>31,35</sup> An increase in conversion was found upon the addition of oxidants and additives such as TEMPO [(2,2,6,6-Tetramethylpiperidin-1-yl)oxy]. The transition metal catalysis done in this work was complicated by the incorrect mass balance. This could be interpreted in two manners, either an additional aqueous work up was needed, resulting in product loss. Alternatively the discrepancy in mass balance could be due to incomplete recovery of all products, which could be especially relevant with a low molecular weight product. Interestingly, AcNH-TEMPO/ HNO<sub>3</sub>/ HCl metal-free catalytic conditions were applied to Aspen lignin (hardwood) and it was determined that the  $\beta$ -O-4 linkages within lignin follow similar selectivity to that of the model compounds.<sup>35</sup>

A lot of current research is about selectivity in oxidation of lignin model compounds, predominately within oxidation of primary or secondary alcohols. This is important because oxidation of both would lead to a mixture of products that could require difficult separation. However, it is important to be able to achieve both oxidations, because chemoselective oxidation of either alcohol could provide a scaffold for the production of aromatic chemicals from lignin.



It is possible that the different reagents and therefore different reaction mechanisms lead to this variety of products. Studies showed that the metal salt used to form the catalyst complex played a role in the selectivity and reactivity. For example, Cu(OTf)<sub>2</sub> was too active and using CuBr<sub>2</sub> instead while decreasing the activity, proved to be more selective for primary benzylic alcohol oxidation over the secondary position.<sup>32</sup>

The exact mechanisms of many of these pathways often are still not known, as for example for the CuCl/bipy/TEMPO/KO<sup>t</sup>Bu system used to catalyze the aerobic oxidation of primary alcohols.<sup>33</sup> Contradicting studies suggested both coordination of the copper to TEMPO followed by a hydrogen transfer abstraction, and alternatively, an initial single electron transfer mechanism.<sup>34</sup> The CuCl/TEMPO systems catalyze the oxidation of lignin model compounds under mild conditions via C-C bond cleavage. In comparison, a dipicolinate vanadium catalyst reacts to break the C-H bond generating different products.<sup>34</sup> These different reactivity patterns for aerobic oxidation of lignin model compounds suggest that homogenous catalysts offer a possible option for the control of selectivity in lignin reactions. By controlling the catalyst, solvent and reaction conditions a wide array of possible aromatic products could be envisioned.

Recent work of Wang and Rinaldi<sup>35</sup> showed that not enough focus was put on the solvent effects within these catalytic reactions. They showed that Lewis basicity of the solvent controlled the reactions activity and selectivity.

#### **4.1.4 Biomimetic metalloporphyrins utilised in lignin oxidation**

Biomimetic approaches have been used to exploit structural features of heme groups within peroxidase enzymes to produce oxidative catalysts. One of the key differences between using a free metalloporphyrin in comparison to a heme-containing enzyme is that with enzymes they conceal the heme group within the protein scaffold, protecting it from bleaching by the oxidation agent. However, a free metalloporphyrin is less restricted so will have a wider substrate scope.

Highly functionalized porphyrins with aryl substituents in the meso position of the ring are used as a catalytic system that is stable to oxidants. By varying the substituents, the change in electronic properties can control the solubility and redox potential of these catalytic

systems.<sup>36</sup> Within the literature there is a large volume of work using biomimetic catalysts for oxidation of lignin and lignin model compounds, a review was written recently in 2013 by Zucca et al covering some of the recent work.<sup>37</sup>

#### **4.1.5 Artificial Metalloenzymes for lignin oxidation hypothesis**

A number of the homogenous catalytic methods outlined above have been successful at obtaining Lignin<sup>OX</sup> a promising potential commercial intermediate. However, there is still room for improvement in these methods (i.e. the use of greener oxidants, lower catalyst loading) and thus to obtain the greatest benefit from lignin as a possible source of fine chemicals, it is advantageous pursue many avenues. We were interested in pursuing an artificial metalloenzymes approach, as this may identify different selectivity's or enable the improvement in activities of metal complexes particularly in oxidative processes by protecting the complex from oxidative degradation.

The hypothesis is that surrounding the metal centre by a protein scaffold to provide control of the secondary coordination sphere could impact the selectivity of the reactions. The secondary coordination sphere provides additional possibilities to optimise the catalyst via protein engineering methods. An interesting example showed that hydrogen bonding provided by the protein scaffold in an ArM increased the stability of a key intermediate in biological and synthetic oxidations.<sup>38</sup>

The aim of this work is to test a wide array of artificial metalloenzymes in the oxidation of lignin model compounds to determine if either selective oxidation or cleavage can be obtained. Using both chemical methods and protein engineering methods, investigation to find the optimum conditions will be carried out to find the ideal catalyst.

## 4.2 Results

Taking inspiration from the ligands used in homogenous oxidative catalysis, a library of maleimide modified nitrogen cofactors (Figure 4) was prepared as previously discussed in Chapter 2. This was used as a starting point because previous work by Stahl had shown that classic metal complexes such as Cu/bpy and Fe/phen are able to oxidise and in some cases break the  $\beta$ -O-4 linkage.<sup>39</sup> However, in most cases both oxidation of the alcohols and cleavage of the lignin was observed with relatively low yields, and ArM's provide an opportunity to improve upon this.

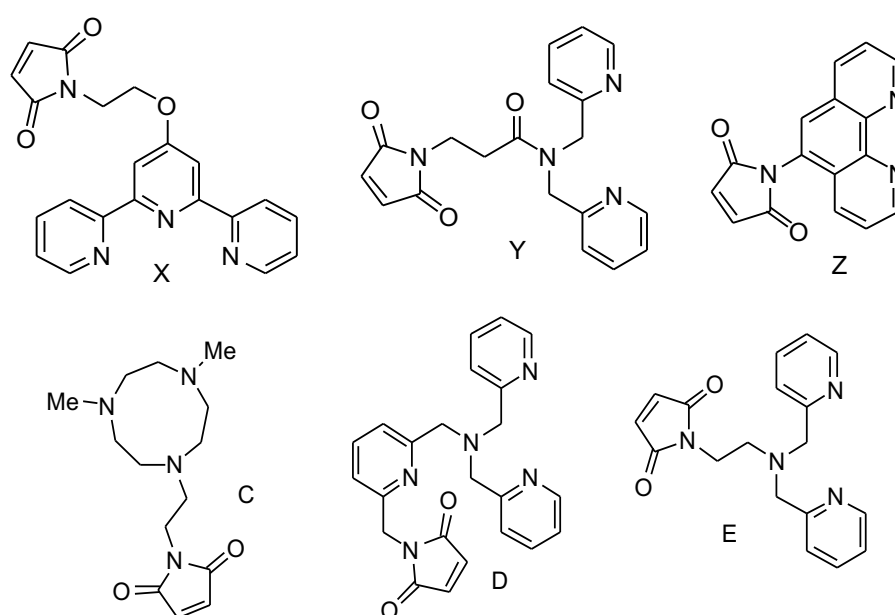


Figure 4 - Library of maleimide modified nitrogen cofactors

### 4.2.1 Lignin Model Compounds

Lignin itself is very complex, therefore initial work on lignin oxidation and depolymerisation generally uses model compounds that simplify the main components. Several model compounds have been used in the literature from basic aryl ethers to synthetic polymers. Some models used for studying lignin, such as aryl ethers and guaiacol (Figure 5 Compound 2 & 3 respectively) can be criticized as being too simple and not a reliable mimic of real lignin. For this work compound 1 was chosen as the initial substrate, as it contains the main components of the  $\beta$ -O-4 linkage including the  $\gamma$ -hydroxyl moiety and is a better representative of how lignin would react.

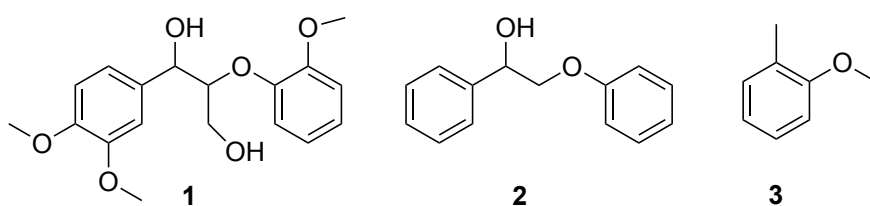
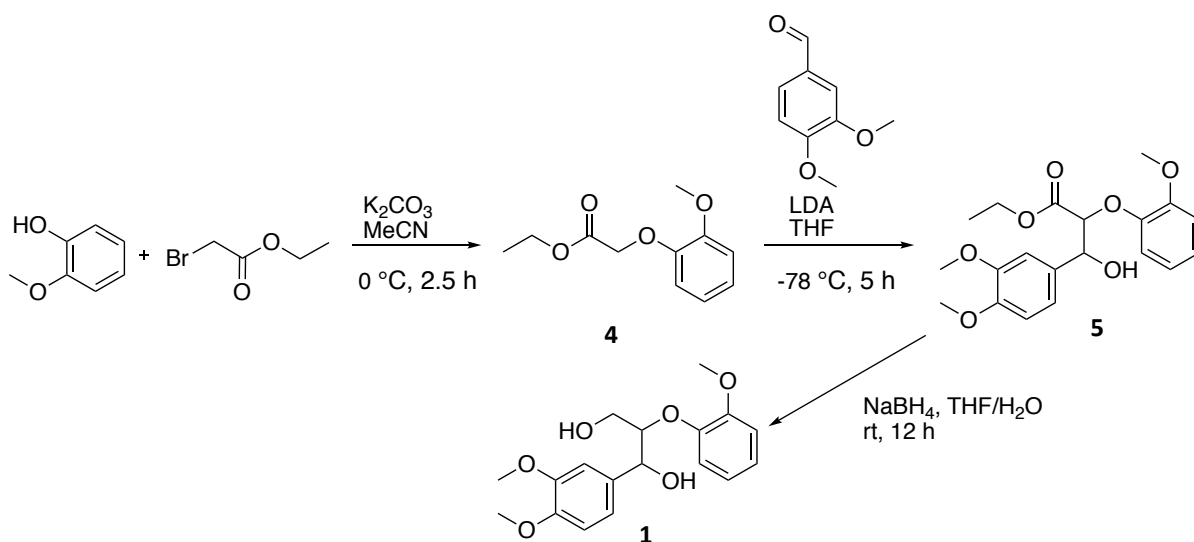


Figure 5 – Range of Lignin Model Compounds studied varying in complexity.

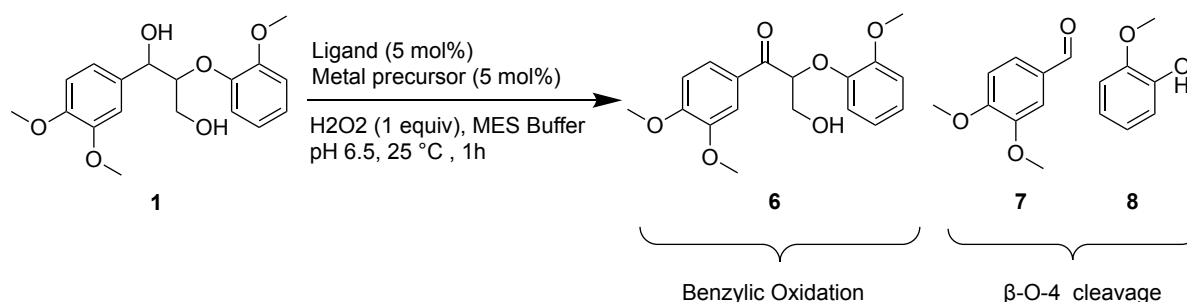
The lignin model compound was synthesised according to literature procedure<sup>40</sup> shown in Scheme 2. While the synthesis of this model compound is well documented, several problems were encountered. The main issue occurred with the lithiation of 2-methoxy(phenoxy)acetate (4) and the trapping of the resulting enolate with veratraldehyde, to produce the  $\beta$ -hydroxy ester (5). Even with several attempts this reaction would not reach completion so required purification. Column chromatography to achieve good separation proved difficult. Good results were achieved using recrystallization conditions with 3:1 ethyl acetate and diethylether.



Scheme 2 – Pathway to synthesise lignin model compound (1)

#### 4.2.2 Oxidation catalysis of lignin model compounds

Initially the first set of reactions with the lignin model substrate were to screen the cofactors without the protein to see if any benzylic oxidation occurs and also if any cleavage occurs. (Scheme 3) The initial screen predominately used iron triflate complexes of the ArM library in the oxidation of lignin model 1.



*Scheme 3 - the observed products from either oxidation of the model compound or from cleavage of the  $\beta$ -O-4 linkage*

In acetonitrile, the TPA based cofactor (Cofactor D) + Fe(OTf)<sub>2</sub> showed moderate conversion and a mixture of oxidation and cleavage products (Table 1, entry 2). However, repeating the reaction in buffer shut off all reactivity (Table 1, entry 1). The phenanthroline based cofactor (Cofactor Z) also showed conversion, however only a small percentage of benzylic oxidation was yielded. In buffer repeating this reaction it only managed to maintain a very small percentage of the activity seen in solvent. The reactions with Cofactor Z showed an unknown product in small yield that could not be identified, so therefore its yield was calculated using the R<sub>f</sub> of the substrate.

Entry	Scaffold/Ligand	Metal	Oxidant	Conversion (%)	Yield (%)			Unknown
					A	B	C	
	5 mol%	5 mol%	1 equiv		A	B	C	Unknown
1	Cofactor D (in Buffer)	Fe(OTf) <sub>2</sub> .2MeCN	H <sub>2</sub> O <sub>2</sub>	0	-	-	-	
2	Cofactor D (in MeCN)	Fe(OTf) <sub>2</sub> .2MeCN	H <sub>2</sub> O <sub>2</sub>	45	8	20	10	
3	Cofactor E (in MeCN)	Fe(OTf) <sub>2</sub> .2MeCN	H <sub>2</sub> O <sub>2</sub>	-	-	-	-	
4	Cofactor E (in Buffer)	Fe(OTf) <sub>2</sub> .2MeCN	H <sub>2</sub> O <sub>2</sub>	-	-	-	-	
5	Cofactor Z (in MeCN)	Fe(OTf) <sub>2</sub> .2MeCN	H <sub>2</sub> O <sub>2</sub>	30	5	-	-	8
6	Cofactor Z (in Buffer)	Fe(OTf) <sub>2</sub> .2MeCN	H <sub>2</sub> O <sub>2</sub>	5	0.5	-	-	3
7	Cofactor Z (in MeCN)	CuBr	H <sub>2</sub> O <sub>2</sub>	8	2	-	-	4
8	Cofactor Z (in Buffer)	CuBr	H <sub>2</sub> O <sub>2</sub>	4	1	0.2	-	3

*Table 1 – Unmodified Transition Metal Catalysis of Lignin Model Compound Oxidation:*

*Conditions: Catalyst formed in situ by addition of the metal precursor (5 mol%, in MeCN stock) Reactions carried out at 25 °C with **1** (1.23 μmol, added from a MeCN stock), H<sub>2</sub>O<sub>2</sub> (1 eq.) in Mes buffer (50 mM, 20 mM NaCl, pH 6.5) or solvent where appropriate. Conversions and yields were determined by HPLC with internal standard (Naphthalene), in triplicate. Unknown Product is a single peak yield calculated with RF for substrate*

#### 4.3.3 Artificial metalloenzymes as oxidation catalysis of lignin model compounds

The next step was to take the cofactors tested and utilise them within an ArM to see how the protein scaffold affected the reaction (Table 2). The pyridyl based Cofactors, D, E and X all showed oxidative activity to yield the benzylic oxidation product between 12 – 85%. The tris(2-pyridylmethyl)amine (TPA) based cofactor (Cofactor D) proved to be the most promising as it yielded the highest 85% of the benzylic oxidation product (Table 2, Entry 1). In the work using the cofactors by themselves, activity was predominately only seen in solvent, and buffer conditions stopped the activity. However, ArM activity is retained despite the aqueous environment and an increased selectivity was observed, with (**6**) being

the sole product observed. This is attributed to a discrete metal catalytic site being formed, as opposed to random iron coordination to the protein. To ensure this was the case and that the protein scaffold did not have any metal binding sites several controls were carried out (Table 3). No conversion was observed if  $\text{Fe}(\text{OTf})_2$  is added directly to the WT SCP-2L protein (Table3, Entry 5). In order to ensure that oxidation was occurring from the ArM system and not any atmospheric oxygen, the reaction was carried out under an argon atmosphere both with and without the ArM (Table3, Entry 3 & 4).

Additional evidence that a discrete metal catalytic site is formed is provided by the change in the extent of conversion when changing the site of modification within the protein scaffold. Both SCP-2L A100C and V83C are located at either end of the hydrophobic tunnel found in SCP-2L, but the surrounding protein environment for each clearly has an effect on the catalysts activity. SCP-2L V83C-D gave lower conversion and yield, suggesting the protein scaffold plays a role either in controlling substrate positioning/access or potentially through interactions that stabilise the catalytic metal centre. The electronic properties of the cofactors in the ArM's also play a crucial role in the activity. When the central nitrogen atom in E is varied by linking through an amide bond, cofactor Y, it proved to be detrimental to the catalytic activity as no reaction was observed. The cofactors based on the phenanthroline moiety (Cofactor Z) or the methyl TACN (Cofactor C) showed no oxidation activity at all. This is an interesting observation because in Chapter 2 the binding of metal to the modified protein was discussed and both Cofactor Z and C did not show the desired 1:1 binding.

Entry	Scaffold/Ligand	Metal	Conversion (%)	Yield (%)		
				A	B	C
1	SCP-2L A100C D	Fe(OTf) <sub>2</sub> .2MeCN	100	85	-	-
2	SCP-2L A100C D	FeCl <sub>3</sub>	0	-	-	-
3	SCP-2L A100C D	Mn(OAc) <sub>2</sub>	0	-	-	-
4	SCP-2L V83C D	Fe(OTf) <sub>2</sub> .2MeCN	85	70	-	-
5	SCP-2L A100C E	Fe(OTf) <sub>2</sub> .2MeCN	35	28	-	-
6	SCP-2L V83C E	Fe(OTf) <sub>2</sub> .2MeCN	0	-	-	-
7	SCP-2L A100C Z	Fe(OTf) <sub>2</sub> .2MeCN	0	-	-	-
8	SCP-2L A100C X	Fe(OTf) <sub>2</sub> .2MeCN	15	12	-	-
9	SCP-2L A100C Y	Fe(OTf) <sub>2</sub> .2MeCN	0	-	-	-
10	SCP-2L A100C C	Fe(OTf) <sub>2</sub> .2MeCN	0	-	-	-

*Table 2 – ArM catalysed Lignin Model Compound Oxidation. Conditions : Reactions carried out at 25 °C with 1 (1.23 μmol, added from a MeCN stock), H<sub>2</sub>O<sub>2</sub> (1 eq.) with rotation at 60 rpm, Buffer conditions 50 mM MES buffer 20mM NaCl pH 6.0, Yields determined by HPLC with internal standard (naphthalene)*

Entry	Scaffold/Ligand 5 mol%	Metal 5 mol%	Conversion (%)	Yield (%)		
				A	B	C
1	-	Fe(OTf) <sub>2</sub> .2MeCN	0	-	-	-
2	-	-	0	-	-	-
3	- *	-	0	-	-	-
4	SCP-2L A100C D*	Fe(OTf) <sub>2</sub> .2MeCN	100	81	-	-
5	WT SCP-2L	Fe(OTf) <sub>2</sub> .2MeCN	0	-	-	-
6	WT SCP-2L + Cofactor D	Fe(OTf) <sub>2</sub> .2MeCN	0	-	-	-

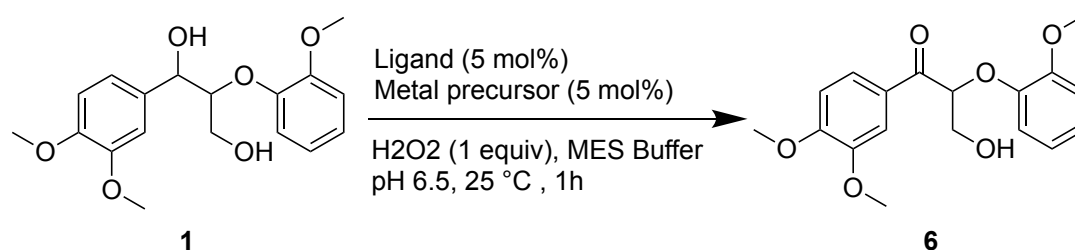
*Table 3 –Controls for ArM catalysed Lignin Model Compound Oxidation: Conditions : Reactions carried out at 25 °C with 1 (1.23 μmol, added from a MeCN stock), H<sub>2</sub>O<sub>2</sub> (1 eq.) with rotation at 60 rpm, Buffer conditions 50 mM MES buffer 20mM NaCl pH 6.0, Yields determined by HPLC with internal standard (naphthalene) \*Reaction carried out under argon.*



Looking at the current research of lignin oxidation, a wide array of metals have been used as previously discussed; therefore one interest was to explore if other metal precursors could be utilised in this ArM system. Mn-tpa complexes are also known oxidation catalysts,<sup>41</sup> however the addition of Mn(OAc)<sub>2</sub> in situ to SCP-2L A100C-D did not result in an active catalyst. Fe(III) precursors were also inactive.

#### 4.2.4 Kinetic Study of Lignin Model Oxidation with ArM SCP-2L A100C-D-Fe<sup>2+</sup>

The next step was to try and gain a greater understanding of this reaction and obtain kinetic data. Initially attempts were made to see if it is possible to lower the catalyst loading. However, the results showed to be unsuccessful as the conversion was found to drop off sharply with decreased catalyst loading, particularly with SCP-2L V83C-D where a 7-fold decrease in activity is observed when halving the catalyst concentration from 5 mol% to 2.5 mol% (Table 4). This was an indication that the reaction may follow non-first order kinetics, possibly the concentration is too close to the metal binding constant so the actual amount of catalyst present in the reaction may be lower than we are expecting. If any metal has dissociated from the ligand we would not expect this to cause any activity as we know that the free metal does not catalyse the reaction (Table 3, entry 1).



Entry	Scaffold/Ligand	Metal loading (mol%)	Conversion (%)	Yield of A (%)
1		5	100	85
2	SCP-2L A100C-D	2.5	35	31
3		1.25	0	0
4		5	85	70
5	SCP-2L V83C-D	2.5	12	7
6		1.25	0	0

*Table 4: Metal loading optimisation Conditions: Reactions carried out at 25 °C with 1 (1.23 μmol, added from a MeCN stock), H<sub>2</sub>O<sub>2</sub> (1 eq.) with rotation at 60 rpm, Buffer conditions 50 mM MES buffer 20mM NaCl pH 6.0, Yields determined by HPLC with internal standard (Naphthalene) \*Reaction carried out under argon.*

The formation of product 6 was then studied overtime as a function of the concentration of the substrate (0.6 – 2.8 mM). For each used substrate concentration, the concentration of the product was plotted against time and the initial rate of the reaction was determined from the linear part of the curve (Figure 6b).

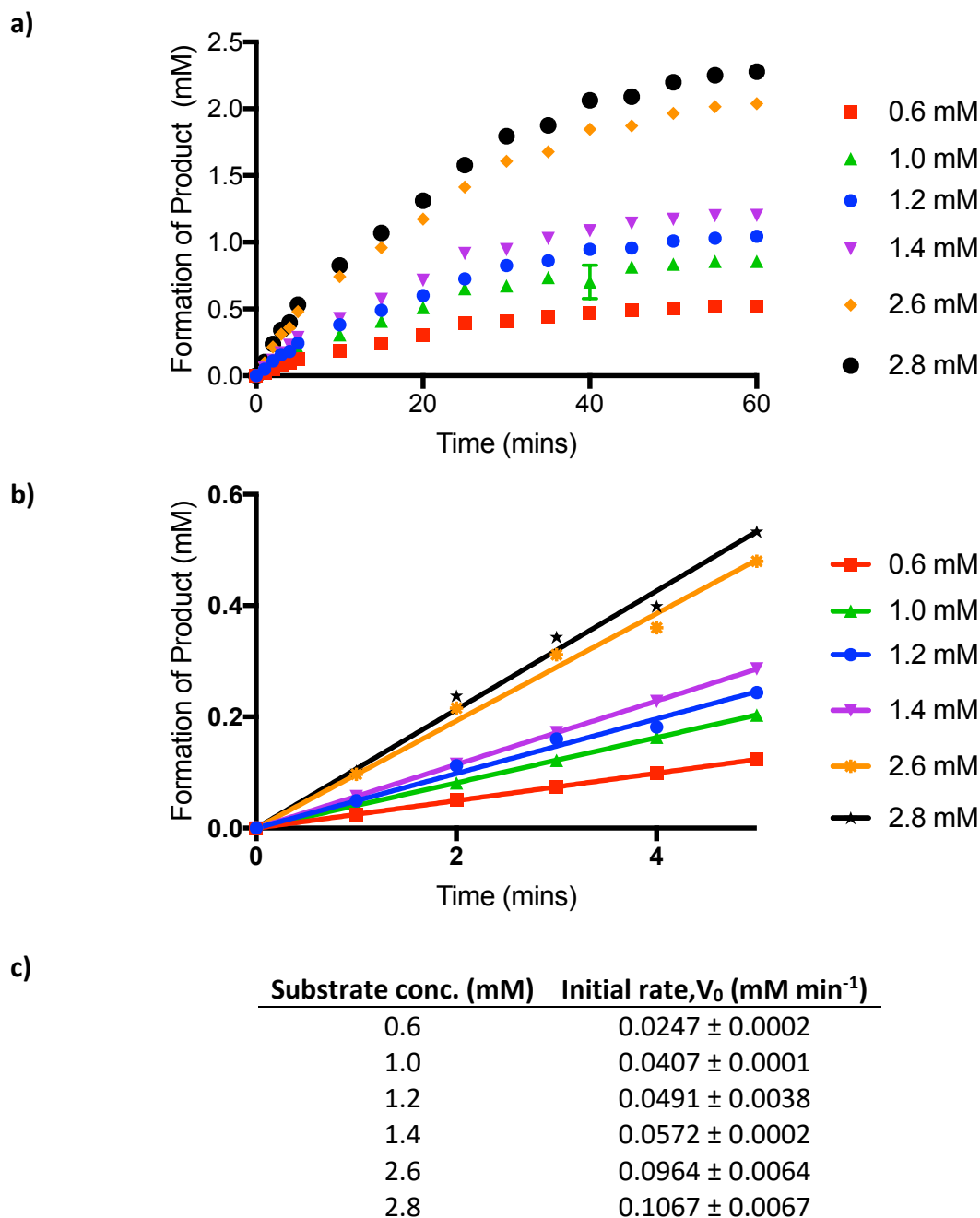


Figure 6 – a) Kinetic Study of the Formation of Product A with varying substrate concentration, b) Initial 5 mins used to calculate the velocity of the reaction. c) Calculated Initial Rate Error quoted for the 95% confidence level.

The next step was to plot the initial velocity vs the varying substrate concentration in order to calculate the kinetic parameters. However, the results showed a linear trend and didn't follow traditional Michaelis-Menten kinetics that you would expect from an enzyme, suggesting the ArM does not reach saturation kinetics. The assumption we can take from this is that the ArM acts more like a small molecular catalyst than an enzyme. By plotting a Lineweaver Burke plot the kinetic parameters can be calculated to give the  $k_{cat}/K_M =$  (Figure 7). Due to the fact that our ArM system does not reach saturation, the  $V_{max}$  value is not reliable as the experimental is far away from that region. The calculated parameters obtained gives a value for  $k_{cat}/K_M = 0.697 \text{ min}^{-1}$ . The gradient =  $K_M/V_{max} = K_M/k_{cat}[E] = 23.65 \text{ min}$ .

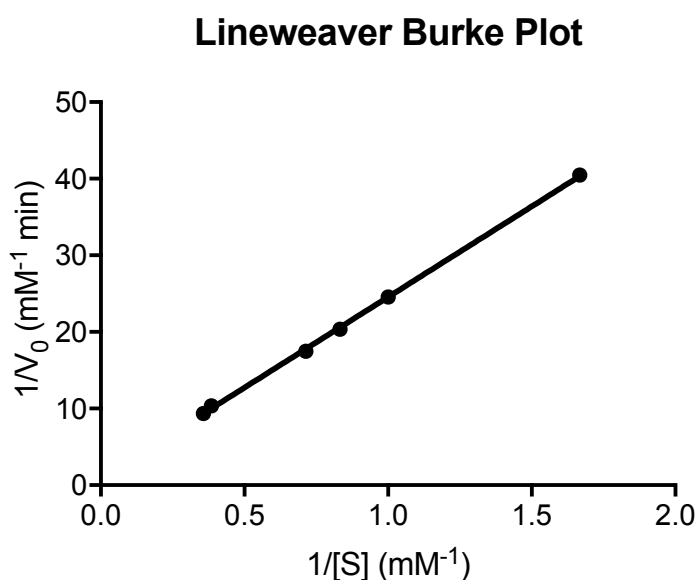


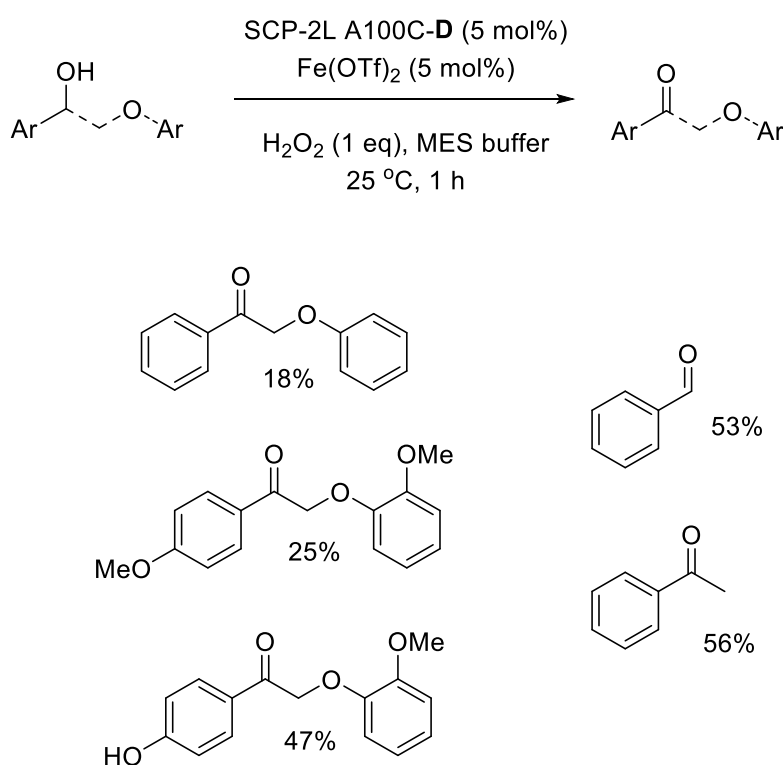
Figure 7 – Lineweaver Burke plot

#### 4.2.5 Varying the Substrate Scope of Lignin Model Compounds

In order to explore how this reaction works, the substrate scope was expanded to look at a range of model compounds that mimic the  $\beta$ -O-4 linkage of lignin. As previously discussed a wide array of models has been studied in literature to mimic lignin (Figure 5). Therefore, a selection of simple models were tested that could be used to explore how the different substituents in lignin may impact on the reactivity of the  $\beta$ -O-4 linkage. A variety of lignin model compounds were subjected to the optimum conditions (Scheme 4). Model compounds that did not contain the  $\gamma$ -hydroxyl showed much lower conversion than

previously studied substrate as they were not fully oxidised under the standard conditions. Introducing a phenol moiety increased conversion over the unsubstituted or methoxy substituted model compounds, albeit still with lower yields than observed with substrate **1**.

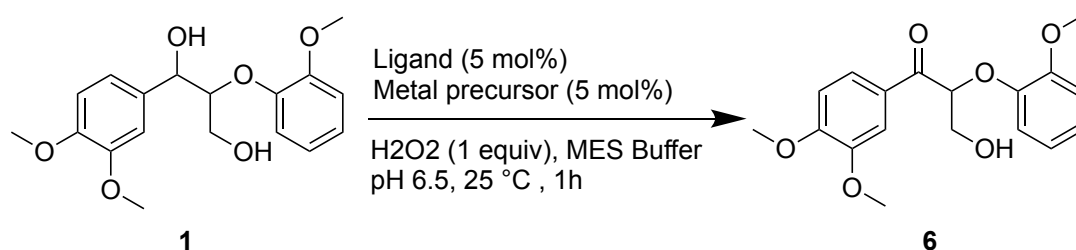
The initial results obtained with SCP-2L A100C-D-Fe<sup>2+</sup> suggested that it could be a good catalyst for simple benzylic oxidation. Therefore, SCP-2L A100C-D was tested in the oxidation of benzyl alcohol and 1-phenylethanol. Both primary and secondary benzylic alcohols are tolerated in the reaction giving moderate yields of the corresponding aldehyde or ketone. Potentially the reduction in yield of the simple model compounds could be due to reduced water solubility of the lignin model. It is also possible that the  $\gamma$ -hydroxyl is an important factor for selective conversion.



*Scheme 4 – Expanding the Substrate Scope for lignin model oxidation. Conditions: Reactions carried out at 25 °C with **1** (1.23  $\mu$ mol, added from a MeCN stock), H<sub>2</sub>O<sub>2</sub> (1 eq.) with rotation at 60 rpm, Buffer conditions 50 mM MES buffer 20mM NaCl pH 6.0, Yields determined by HPLC with internal standard (Naphthalene) \*Reaction carried out under argon.*

#### 4.2.6 Investigating ArM reactivity with Organic Solvent Tolerance

A key drawback in most methodology developed on lignin model compounds is that they cannot be reproduced with real lignin. This is largely due to insolubility of real lignin in most solvents systems. DMSO and dioxane are often solvents of choice for reactions with lignin. One limitation with ArM systems is that protein scaffolds tend to have very low tolerance to organic solvents, and denaturation is often seen. In order to test out this system the reactions were carried out in a buffer solvent mix to see how the activity is affected. Both DMSO and dioxane were also shown to be tolerated as co-solvents under the standard reaction conditions, although substantial decreases in yields were observed (Table 5 Entry 1 vs entries 4-7). The addition of MeCN as a co-solvent was tolerated up to 30% MeCN, albeit with reduced yields of **6** (Table 5). Increasing the co-solvent ratio to 50% led to very low levels of activity. In order to improve upon this protein engineering methods could be used to increase the proteins stability with respect to organic solvents.<sup>42</sup>



Entry	Solvent (Ratio)	Conversion (%)	Yield of A (%)
<b>1</b>	MES	100	85
<b>2</b>	MES : MeCN (7:3)	54	41
<b>3</b>	MES : MeCN (1:1)	12	5
<b>4</b>	MES : Dioxane (7:3)	38	35
<b>5</b>	MES : Dioxane (1:1)	0	0
<b>6</b>	MES : DMSO (7:3)	57	53
<b>7</b>	MES : DMSO (1:1)	9	8

*Table 5: Co-solvent tolerance Conditions : Reactions carried out at 25 °C with **1** (1.23 μmol, added from a MeCN stock), H<sub>2</sub>O<sub>2</sub> (1 eq.) with rotation at 60 rpm, Buffer conditions 50 mM MES buffer 20mM NaCl pH 6.0, Yields determined by HPLC with internal standard (Naphthalene) \*Reaction carried out under argon.*

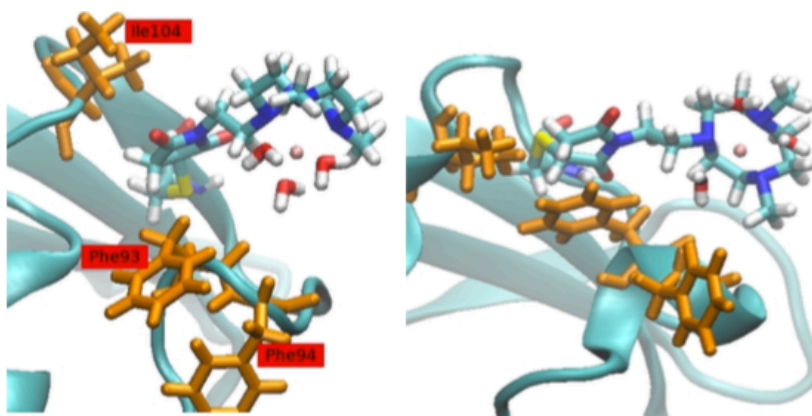
#### **4.2.7 Mutants to increase cofactor stability**

In order to learn more about the ArM activity further characterisation was required, because we had no evidence where the metal ion sits within the protein and which surrounding amino acids it can interact with. Attempts to characterise the obtained ArM with MS were unsuccessful and we were unable to obtain crystal structures of any of the modified proteins SCP-2L A100C/V83C-X, or their metal complexes. Therefore, in parallel to this work, MD simulations were carried out to attempt to probe the position of the cofactors in the protein and any interactions that could be observed.

##### **4.2.7.1 Design of mutants in which protein secondary coordination sphere would stabilize cofactor**

MD simulations were carried out by Michael Buehl, Jan Goetze and Jonathan Colburn. The system used to model was SCP-2L A100C modified with Cofactor C.

One complication with the MD simulation is that modifying the unique cysteine residue has the ability to produce two diastereoisomers and it is not possible to determine which is seen. MD simulations were therefore carried out with each diastereoisomer, and as expected slightly different distances and interactions were observed for each diastereoisomer. The MD simulations showed close interactions of the target construct mainly with the Phe93/Phe94 cluster of a nearby alpha helix. Phe94 was identified as a potential mutation target to introduce coordinating amino acids. (Figure 8) Phe94 was chosen as a preference over Phe93 because it is believed Phe93 is involved in maintaining the structure of the hydrophobic tunnel.



*Figure 8 – Closing arrangement of the “head” of the hydrophobic channel in the two diastereoisomeric arrangements of the bound ligand. Left: S-stereoisomer with open channel. Right: R-stereoisomer with closed channel due to stacked Phe93/ligand rings and Phe93/Ile interactions. MD simulations were carried out by Michael Buehl, Jan Goetze and Jonathan Colburn.<sup>1</sup>*

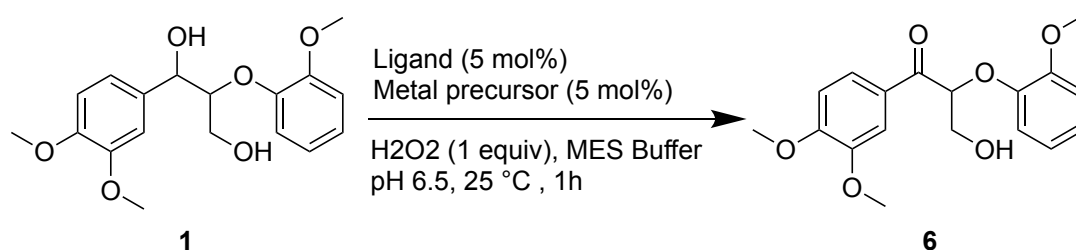
Amino acids that can coordinate to metals such as His or Glu are known to stabilise Fe-O<sub>2</sub> and Fe-OOH intermediates either directly or through hydrogen bonding in a number of enzymes.<sup>43, 44</sup> A similar concept is shown by adding acetic acid to oxidation reactions catalysed by iron complexes, and this was shown to improve yields, possibly by stabilisation of iron intermediates. The mutants SCP-2L A100C F94E and F94H were designed, expressed and purified.

#### **4.2.7.2 Lignin model oxidation with mutant in optimum conditions**

The same optimum conditions were used to test the new mutants to see how what affect they would have on the reactivity. As the Cofactor D showed full conversion of the substrate at a catalytic loading of 5 mol%, the loading was reduced to 2.5 mol% in order to see if any changes occur by introduction of a His or Glu amino acid in the protein scaffold.

The introduction of His at position 94 had no influence on the reaction; however, the introduction of Glu had a positive influence on the reactivity increasing the observed yield. The increase was most remarkable with cofactor SCP-2L A100C-E, where the yield doubled on the introduction of Glu. Acetic Acid was added to the reaction in order to see if it would mimic

the same changes as introducing Glu to the protein scaffold. Smaller increases were observed when acetic acid was added as an additive (Table 6).

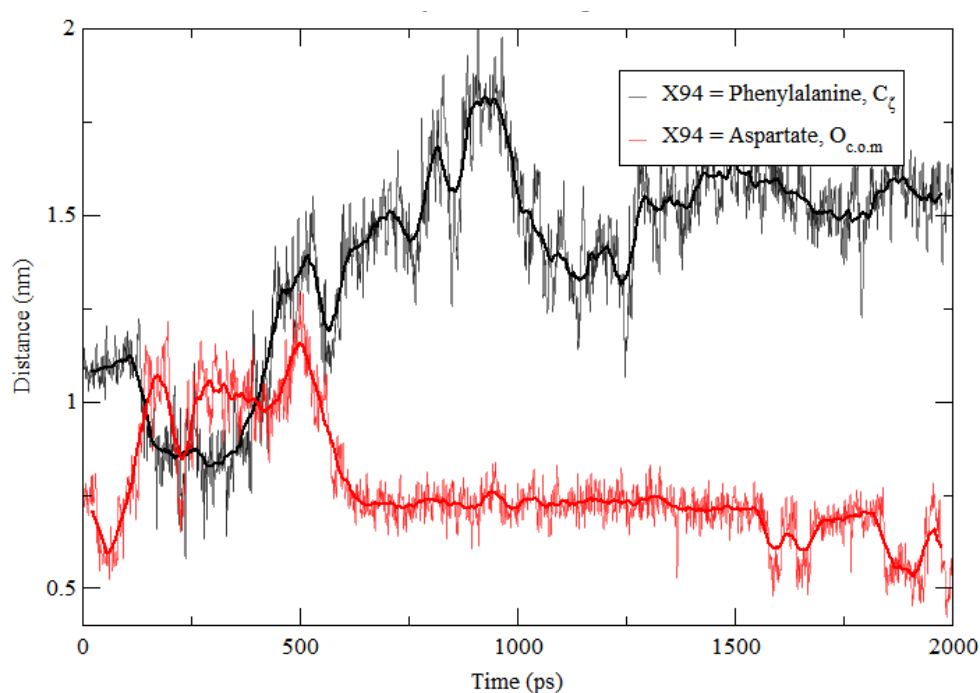


Scaffold/Ligand	Metal	Acetic Acid	Conversion (%)	Yield (%)		
				A	B	C
<b>2.5 mol%</b>	<b>2.5 mol%</b>			<b>A</b>	<b>B</b>	<b>C</b>
SCP-2L A100C D	Fe(OTf) <sub>2</sub> .2MeCN	-	35	31	-	-
SCP-2L A100C F94E D	Fe(OTf) <sub>2</sub> .2MeCN	-	47	43	-	-
SCP-2L A100C F94H D	Fe(OTf) <sub>2</sub> .2MeCN	-	35	29	-	-
SCP-2L A100C D	Fe(OTf) <sub>2</sub> .2MeCN	1 equiv	40	36	-	-
SCP-2L A100C D	Fe(OTf) <sub>2</sub> .2MeCN	5 equiv	53	48	-	-
<b>5 mol%</b>	<b>5 mol%</b>			<b>A</b>	<b>B</b>	<b>C</b>
SCP-2L A100C E	Fe(OTf) <sub>2</sub> .2MeCN	-	35	28	-	-
SCP-2L A100C F94E E	Fe(OTf) <sub>2</sub> .2MeCN	-	59	54	-	-
SCP-2L A100C F94H E	Fe(OTf) <sub>2</sub> .2MeCN	-	35	30	-	-
SCP-2L A100C E	Fe(OTf) <sub>2</sub> .2MeCN	1 equiv	45	41	-	-

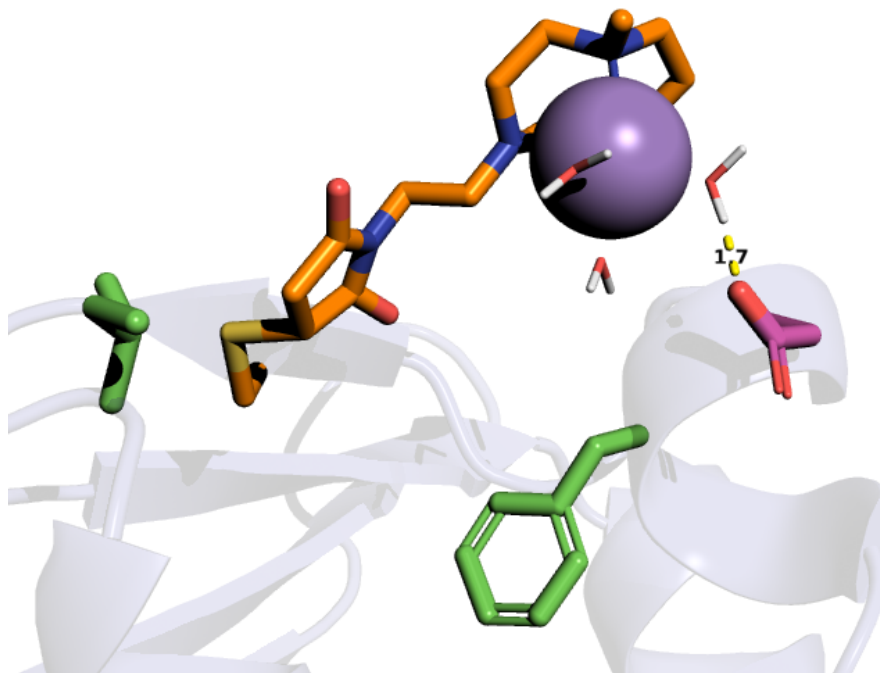
*Table 6 – Lignin Model Compound Oxidation with SCP-2L A100C F94E/H Conditions : Reactions carried out at 25 °C with 1 (1.23 μmol, added from a MeCN stock), H<sub>2</sub>O<sub>2</sub> (1 eq.) with rotation at 60 rpm, Buffer conditions 50 mM MES buffer 20mM NaCl pH 6.0, Yields determined by HPLC with internal standard (Naphthalene)*



Further MD simulations were carried out looking at the Glu mutant to see if closer contact between the carboxylate side chain and the metal is seen. Within the simulations an Asp was used in the place of Glu because of the shorter side chain there are less conformational degrees of freedom to explore. The initial MD simulations were carried out with TACN based cofactor, so the same was utilised in these studies. The simulations showed that on average the carboxylate side chain was closer to the metal the original protein. For a short period the carboxylate side chain came within hydrogen bonded distance to a water ligand coordinated to the metal ( between 1.7 – 2ns in Figure 9, see Figure 10 for a snapshot from this region).



*Figure 9: Fluctuations of the distances between the metal and selected points of the side chain of residue 94 during 2 ns of MD of the protein-ligand complex (results shown for the S-isomer); in black: WT (distance to C<sub>β</sub> of Phe94), in red: F94D mutant (distance to the centre of mass of the two carboxylate O atoms, which, when Phe and Asp residues are overlaid, is very close to the position of C<sub>β</sub>). MD simulations were carried out by Michael Buehl, Jan Goetze and Jonathan Colburn.<sup>1</sup>*



*Figure 10: Snapshot from the simulation of the F94D mutant showing a close contact between the carboxylate side chain and the metal fragment (distance in Å). MD simulations were carried out by Michael Buehl, Jan Goetze and Jonathan Colburn.<sup>1</sup>*

Overall, it has been demonstrated that ArMs can be used to change the selectivity of chemical reactions compared to the parent small molecule catalysts, producing selective catalysts. Complete selective conversion to the ketone product from the lignin model compound was seen using SCP-2L A100C modified with a tris(2-pyridylmethyl) amine based cofactor, complexed with  $\text{Fe}(\text{OTf})_2 \cdot 2\text{MeCN}$ . Engineering the protein scaffold to incorporate glutamic acid was found to improve the ArM activity. This demonstrated that rational design of the surrounding protein environment using metal binding amino acids can be used to improve the overall activity of an artificial metalloenzyme.

Previous use of protein engineering in increasing the activity of a hydroformylase, shown in chapter 3, suggests that further engineering of the ArMs could produce a more stable protein that could potentially be tolerable to organic solvents. Further optimisation of the ArM system could provide a pathway to the ideal catalyst for selective lignin oxidation and also degradation.

## 4.3 Experimental

### 4.3.1 General Remarks

All reagents and solvents were obtained commercially and used without further purification. The *n*BuLi was titrated against 2-3 dimethoxybenzyl alcohol prior to use to determine the correct concentration. Thin layer chromatography was performed on silica plates (TLC silica gel 60 F254 Merck) and visualised with UV light (254nm). Column chromatography was performed using silica gel 60 particle size 40-63 $\mu$ m from Merck. NMR spectra were taken at room temperature with Bruker Avance 300, 400 or 500 NMR spectrometers. Chemical shifts ( $\delta$ ) are given (in ppm) and processed using MestReNova software. The residual undeuterated solvent of the deuterated solvent used ( $\text{CHCl}_3$   $\delta$  = 7.26 and 77.16,  $^1\text{H}$  and  $^{13}\text{C}$  respectively) was used to reference the chemical shifts. The abbreviations for the multiplicity of the proton, carbon and fluorine signals are as follows: s singlet, d doublet, dd doublet of doublets, ddd doublet of doublet of doublets, t triplet, dt double triplets, q quartet, m multiplet, br s broad singlet.

UV-visible measurements were recorded on a Molecular Devices M2 Spectra Max spectrophotometer.

### 4.3.2 Protein Expression and purification:

SCP-2L A100C and V83C were expressed and purified as described before.<sup>45</sup> The same method was used for SCP-2L A100C F94H and F94E giving yields of 20  $\text{mgL}^{-1}$  and 25  $\text{mgL}^{-1}$  respectively. Mutations were introduced using the Quick-Change Site-Directed Mutagenesis Kit (Stratagene) with pEHISTEV:: $\Delta\text{h}\Delta\text{SCP-2L}$  as the template. The primers used for the mutagenesis are shown below. The constructs were verified by sequencing by GATC.

SCP2L A100C F94E

5' CCT CAG AAG GCA TTC **GAA** AGT GGC AGG CTG AAG 3'

5' CTT CAG CCT GCC ACT **TTC** GAA TGC CTT CTG AGG 3'

SCP2L A100C F94H

5' CCT CAG AAG GCA TTC **CAT** AGT GGC AGG CTG AAG 3'

5' CTT CAG CCT GCC ACT **ATG** GAA TGC CTT CTG AGG 3'

The constructs were prepared as follows: SCP-2L A100C was used as the template DNA. Site directed mutagenesis was carried out. The reaction was set up as follows: Primer fw: 0.01  $\mu\text{M}$  (0.25  $\mu\text{M}$ , 0.5  $\mu\text{L}$ ), primer rv: 0.01  $\mu\text{M}$  (0.25  $\mu\text{M}$ , 0.5  $\mu\text{L}$ ), SCP-2L template DNA (0.5  $\mu\text{L}$ ), PfuUltra Master Mix (10  $\mu\text{L}$ ), Sterile MilliQ Water.

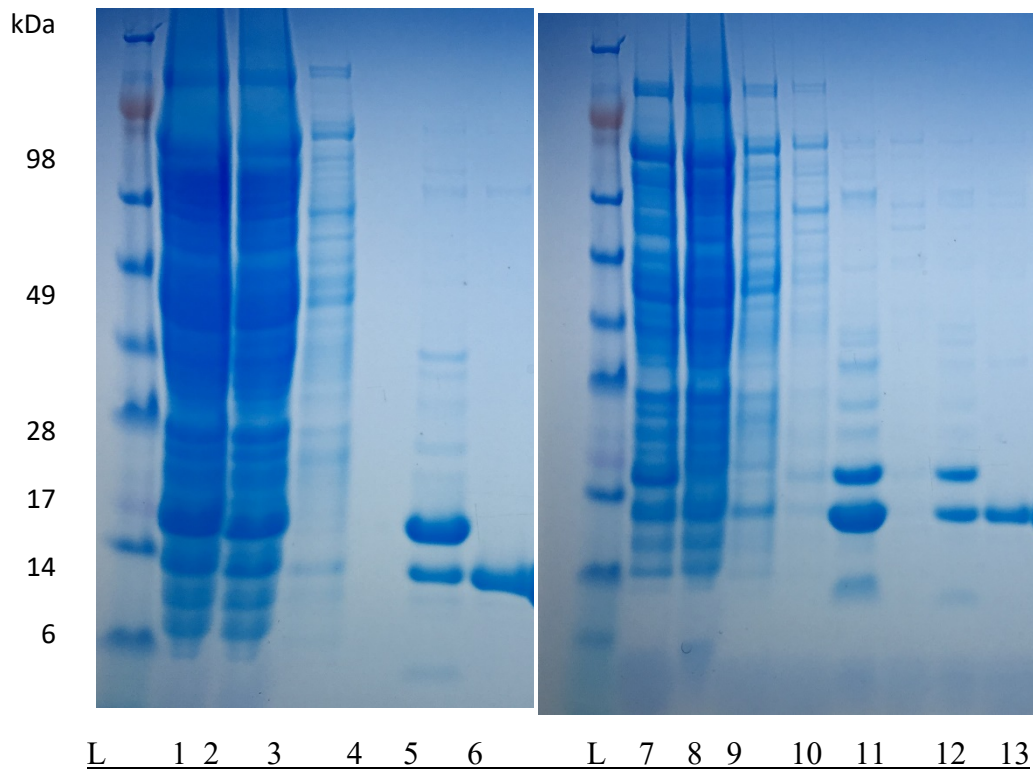
After mixing, the PCR tubes are shortly spun (4 sec, table top centrifuge, RT). The tubes are transferred to a PCR machine and the following PCR cycle is run.

1. Initial Denaturation 95 °C, 5 min
2. Denaturation 30 sec 95 °C
3. Annealing 1min 55 °C
4. Extension 1min/1kp 72 °C
5. Repeat Steps 2-4 for 16 cycles
6. Final extension 20 min 72 °C

Digestion of the wild type template strand was then performed as follows: PCR product (8.5  $\mu\text{L}$ ), 10x Fast digest buffer (1  $\mu\text{L}$ ) and Dpn1 (0.5  $\mu\text{L}$ ) were incubated at 37 °C for 10 min, followed by inactivation at 72 °C for 5 min. The digested products were transformed into commercial DH5 $\alpha$  cells and grown on kanamycin (50  $\mu\text{g}/\text{mL}$ ) agar plates at 37 °C overnight.

One to two colonies were observed for each mutant. The DNA was obtained via a miniprep kit and sent for sequencing at GATC Biotech Co., UK.

SDS Gel to show purification of SCP-2L A100C F94H (left) and SCP-2L A100C F94E (right)



L. SeeBlue Plus 2 Prestained Protein Standard

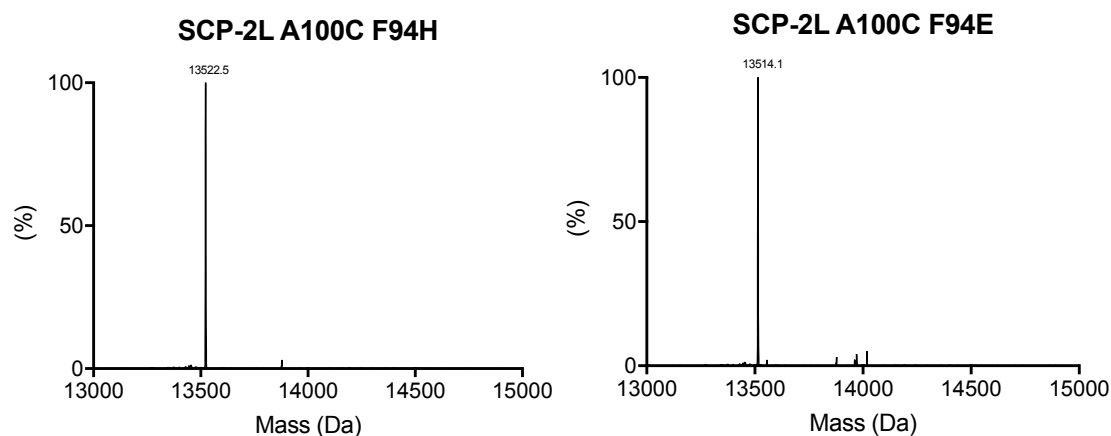
1. SCP-2L A100C F94H Supernatant
2. SCP-2L A100C F94H Supernatant
3. Wash through from Nickel Column
4. High Salt Wash from Nickel column
5. Elution His - Tagged Protein
6. Pure SCP-2L A100C F94H
7. SCP-2L A100C F94E Supernatant
8. SCP-2L A100C F94E Supernatant
9. Wash through from Nickel Column
10. High Salt Wash from Nickel column
11. Elution His - Tagged Protein
12. His Tagged Cleaved Protein
13. Pure SCP-2L A100C F94E

### 4.3.3 LCMS analysis of new SCP-2L A100C mutants F94E/H and the relative modified ArM's

LC-MS(ES+) used for analysis of protein and conjugation-reactions was performed on a Waters Alliance HT 2795 equipped with a Micromass LCT-TOF mass spectrometer, using positive electrospray ionisation and applying a Waters MASSPREP® On-line Desalting 2.1x10 mm cartridge using a gradient of 1% formic acid in H<sub>2</sub>O to 1% formic acid in acetonitrile. ESI-MS results were analysed by MassLynx V. 4.0 and its MaxEnt algorithm.

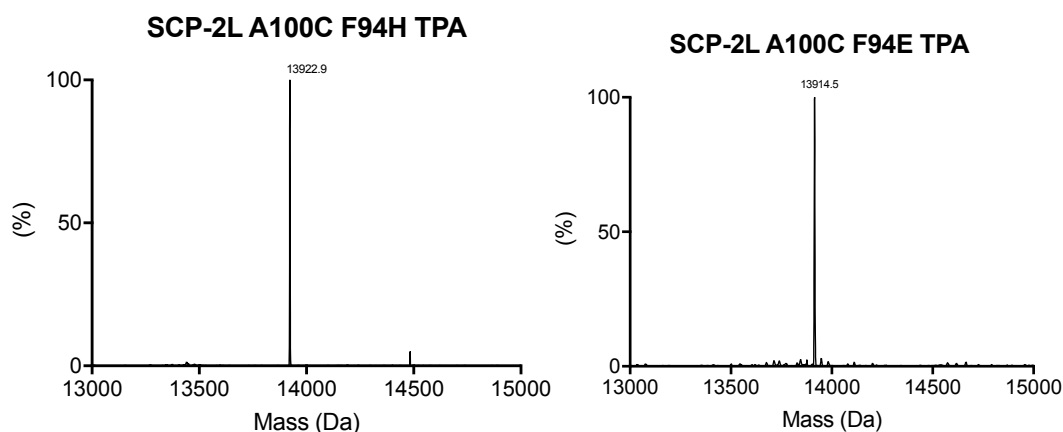
SCP 2L A100C F94H: Calculated mass: 13522.8 Da Observed: 13522.5Da.

SCP 2L A100C F93E: Calculated mass: 13514.8 Da Observed: 13514.1 Da.



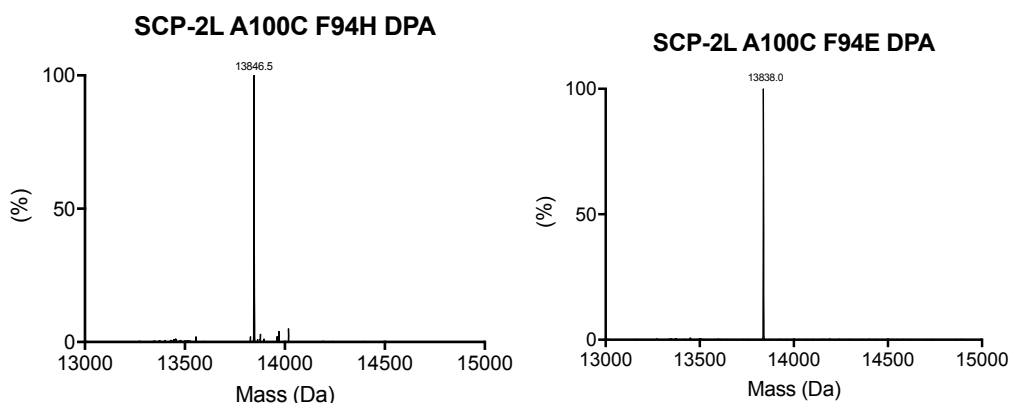
SCP 2L A100C F94H-TPA: Calculated mass: 13922.8 Da Observed: 13922.9Da.

SCP 2L A100C F93E-TPA: Calculated mass: 13914.8 Da Observed: 13914.1 Da.



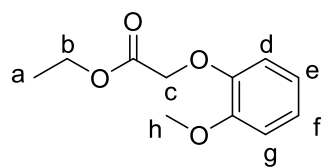
SCP 2L A100C F94H-TPA: Calculated mass: 13846.8 Da Observed: 13846.5 Da.

SCP 2L A100C F93E-TPA: Calculated mass: 13814.8 Da Observed: 13838.0 Da.



#### 4.3.4 Lignin Model Compound Synthesis- Carried out according to published literature<sup>38</sup>

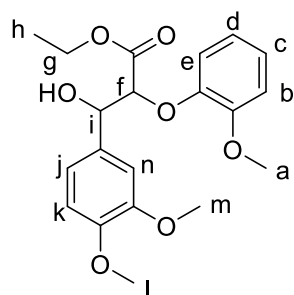
##### Step 1: Synthesis of Ethyl(2-methoxyphenoxy)acetate (4)



A solution containing guaiacol (80 g, 1 equiv., 644 mmol) in acetonitrile (600 mL) and  $K_2CO_3$  (115.6 g, 1.3 equiv., 838 mmol) was stirred for one hour at room temperature. The reaction mixture was then cooled to 0 °C and ethyl bromoacetate (120 g, 1.12 equiv., 718 mmol) was added dropwise over a period of 30 mins and then stirred for an hour at 0 °C. The reaction mixture was filtered, and the light blue filtrate was collected. The reaction mixture was dissolved in water (200 mL) and the product was extracted using ethyl acetate (400 mL). The organic layer was dried over  $MgSO_4$ , filtered and concentrated. The product was washed with 1M NaOH (300 mL) and then dried under vacuum to give the product as dark yellow oil. (41.2 g, 33% yield).

$^1H$  NMR (500MHz,  $CDCl_3$ , TMS):  $\delta$  6.88–6.70 (m, 4H,  $H_{d-g}$ ), 4.57 (s, 2H,  $H_c$ ), 4.24 (q,  $J = 7.1$ Hz, 2H,  $H_b$ ), 3.73 (s, 3H,  $H_h$ ), 1.27 (t,  $J = 7.2$ Hz, 3H,  $H_a$ )

**Step 2: erythro-ethyl 3-(3,4-dimethoxyphenyl)-3-hydroxy-2-(2-methoxyphenoxy)propanoate (5)**



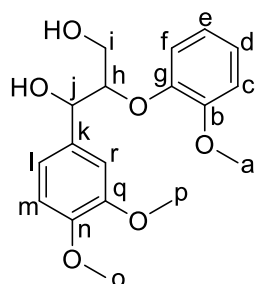
Diisopropylamine (5.1 mL, 36.1 mmol, 1.2 equiv.) in dry THF (40 mL) is stirred under argon at  $-78^{\circ}\text{C}$ . 2.3M *n*BuLi (14.4ml, 36.1mmol, 1.2 equiv) was added dropwise to the reaction over a period of 30mins and then stirred for a further 30 mins at  $-78^{\circ}\text{C}$ . Ethyl(2-methoxyphenoxy)acetate (7.6g, 36.1mmol, 1.2 equiv) is added dropwise over a period of 30 mins and then stirred for a further 60 mins at  $-78^{\circ}\text{C}$  and the reaction turned from colourless to yellow. Veratrylaldehyde (5g, 30.1mmol, 1 equiv) was added to the reaction mixture and then was stirred for a further 3 hours at  $78^{\circ}\text{C}$  during which time the reaction mixture turned cloudy. The reaction mixture was raised from the ice bath and allowed to warm up to room temperature, then the reaction mixture was diluted with water. The product was then extracted with ethyl acetate (4 x 50 ml) and the organic fraction was dried over  $\text{MgSO}_4$ , filtered and concentrated. The solid product was recrystallized from ethanol/ hexane (3:1) and the product washed with diethyl ether to give a white solid. (2.4g, 21% yield).

M.p.  $104\text{-}105^{\circ}\text{C}$

$^1\text{H}$  NMR (500MHz,  $\text{CDCl}_3$ ):  $\delta$  7.07 (d,  $J = 2.0$  Hz, 1H, Ar-*H*), 7.05-6.99 (ddd,  $J = 8.2, 5.2, 1.6$  Hz, 2H, Ar-*H*), 6.88-6.84(m, 2H, Ar-*H*), 6.84-6.82 (m, 2H, Ar-*H*), 5.14 (br t,  $J = 3.8$  Hz, 1H, OH), 4.74 (d,  $J = 5.0$  Hz, 1H,  $\text{H}_i$ ), 4.16 (q,  $J = 5.0$  Hz, 2H,  $\text{H}_g$ ), 3.90 (s, 3H,  $\text{H}_a$ ), 3.89 (s, 3H,  $\text{H}_l$ ), 3.87 (s, 3H,  $\text{H}_m$ ), 3.66 (d,  $J = 5.8$  Hz, 1H), 1.15 (t,  $J = 7.1$  Hz, 3H,  $\text{H}_h$ )



**Step 3: Synthesis of 1-(3,4-dimethoxyphenyl)-2-(2-methoxyphenoxy)propane-1,3-diol – Model Compound (1)**



$\beta$ -Hydroxy ester (**5**) (0.88 g, 1 equiv., 2.34 mmol) in 3:1 THF/H<sub>2</sub>O (13 mL) was stirred at room temperature and NaBH<sub>4</sub> (0.44g, 5 equiv, 11.7mmol) was added to the reaction mixture, which was stirred overnight. Water was added to quench the reaction mixture, and then the product was extracted using ethyl acetate. The organic fraction was dried over MgSO<sub>4</sub>, filtered and concentrated to produce a viscous yellow oily solid. (0.67g, 85% yield).

Mp 104-105 °C

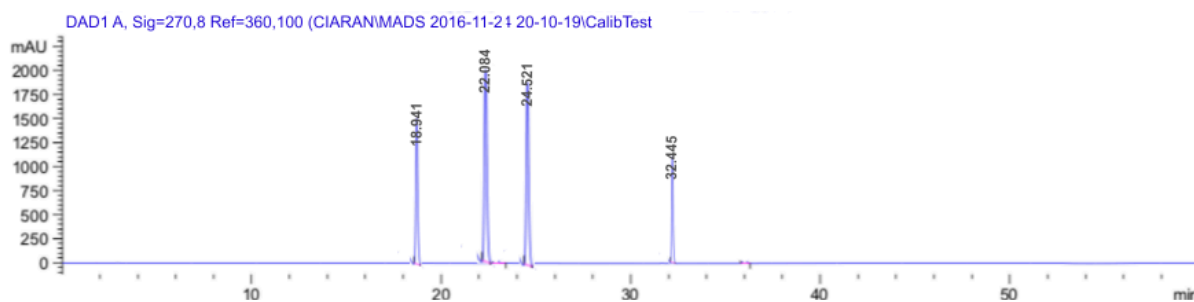
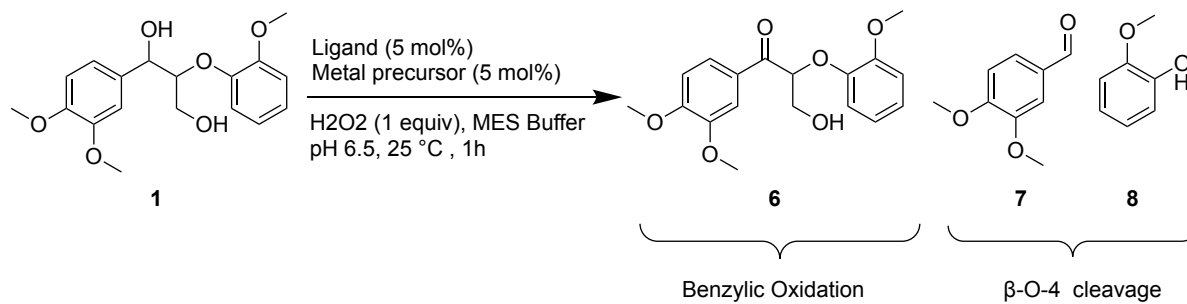
<sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>)  $\delta$  7.10 (ddd, *J* = 8.1, 7.3, 1.8 Hz, 1H, Ar-*H*), 7.02-6.90 (m, 5H, Ar-*H*), 6.86 (d, *J* = 8.2 Hz, 1H, Ar-*H*), 5.01 (dt, *J* = 4.6, 1.8 Hz, 1H, H<sub>j</sub>), 4.19 (ddd, *J* = 6.1, 4.7, 3.4 Hz, 1H, H<sub>h</sub>), 3.95 (m, 1H, H<sub>i</sub>), 3.92 (s, 3H, H<sub>a</sub>), 3.91 (s, 3H, H<sub>o</sub>), 3.90 (s, 3H, H<sub>p</sub>), 3.68 (ddd, *J* = 12.4, 7.3, 3.4 Hz, 1H, H<sub>i</sub>), 3.49 (br s, 1H, OH), 2.73 (br s, 1H, OH)

<sup>13</sup>C NMR (126 MHz, CDCl<sub>3</sub>)  $\delta$  151.7 (Ar-C), 149.0 (Ar-C), 148.5 (Ar-C), 146.8 (Ar-C), 132.3 (Ar-C), 124.4 (Ar-C), 121.7 (Ar-C), 121.2 (Ar-C), 118.4 (Ar-C), 112.2 (Ar-C), 111.0 (Ar-C), 109.1 (Ar-C), 87.6 (C<sub>h</sub>), 72.7 (C<sub>j</sub>), 60.7 (C<sub>i</sub>), 55.9 (C<sub>a,o,p</sub>)

Spectral data corresponds to those given in literature<sup>38</sup>

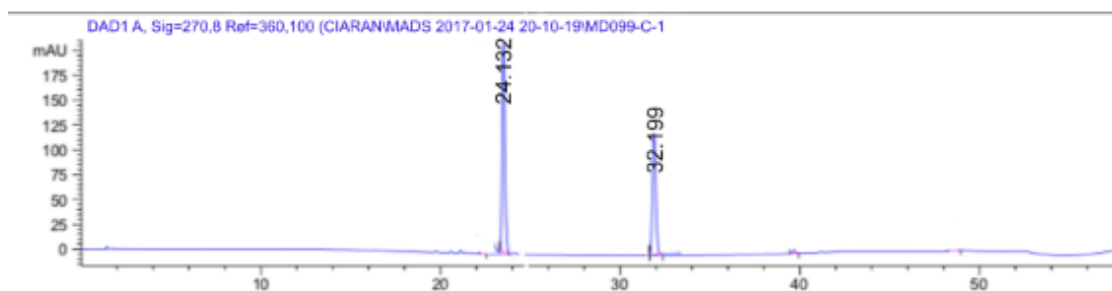
HRMS (ESI) *m/z* [MNH<sub>4</sub>]<sup>+</sup> 352.1756 (calculated C<sub>18</sub>H<sub>22</sub>O<sub>6</sub>NH<sub>4</sub> 352.1755)

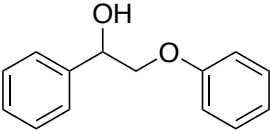
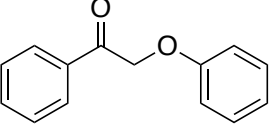
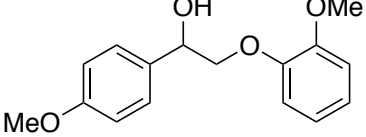
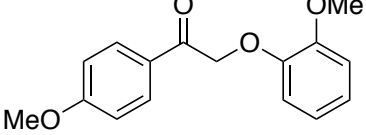
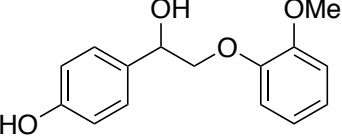
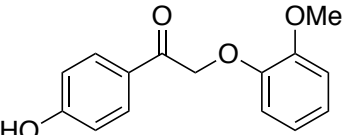
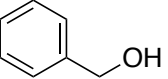
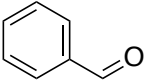
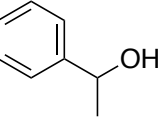
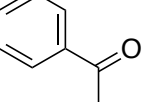




	Retention Time
Dilignol Model Substrate (1)	22.08
Secondary Ketone Dilignol Model (6)	24.52
Veratraldehyde	20.96
Guaiacol	18.94
Napthalene	32.44

Example HPLC trace of SCP-2L A100C -D-Fe(Otf)<sub>2</sub>.2MeCN oxidation with Lignin Model Compound 1



Substrate	Retention Time (mins)
	20.14
	21.84
	21.66
	22.54
	20.94
	22.13
	17.55
	19.23
	19.39
	21.01

**Kinetics:** The kinetics for the oxidation of **1** catalysed by ArM SCP-2L A100C-TPA-Fe<sup>2+</sup> was determined. All measurements were performed at 25 °C in MES Buffer 50 mM MES, 20 mM NaCl, pH 6.0. Analysis was performed by HPLC (conditions given previously) monitoring the formation of product **6** overtime as a function of the concentration of the substrate. (0.6 – 2.8 mM).

The catalytic solution was prepared in the same manner as the standard reactions, by combining the modified ArM (61.5 μM, 5 mol%) with Fe(OTf)<sub>2</sub>·2MeCN (61.5 μM 5 mol%) allowing the complex to form for 20 mins by shaking at 60 rpm. Substrate **5** (0.6 – 2.8 mM) and 1 equivalent of H<sub>2</sub>O<sub>2</sub> (1.23 mM) is added to begin the reaction in a final volume of 1000 μl. To stop the reaction 500 μl DCM was added at the appropriate time point, the reaction mixture was then centrifuged for 2 mins 13000 rpm to remove the catalyst and the product and any remaining starting material were extracted as the organic layer. Analysis was then carried out using HPLC.

For each used substrate concentration, the concentration of the product was plotted against time and the initial rate of the reaction was determined from the linear part of the curve, using linear regression in Prism Graphpad 7.0 software.

---

<sup>1</sup> M. V. Doble; A. G. Jarvis; A. C. C. Ward; J. D. Colburn; J. P. Gotze; M. Buhl; P. C. J. Kamer. *ACS Sustainable Chem. Eng.* 2018, 6, 15100

<sup>2</sup> S. R. Collinson; W. Thielemans. *Coord. Chem. Rev.* **2010**, 254, 1854

<sup>3</sup> J. Zakzeski; P. C. Bruijninx; A. L. Jongerius; B. M. Weckhuysen. *Chem. Rev.* **2010**, 110, 3552

<sup>4</sup> F. S. Chakar; A. J. Ragauskas. *Ind. Crops. Prod.* **2004**, 20, 131

<sup>5</sup> S. R. Collinson; W. Thielemans. *Coord. Chem. Rev.* **2010**, 254, 1854

<sup>6</sup> M. Fache; B. Boutevin; S. Caillol. *ACS Sustainable Chem. Eng.* **2016**, 4, 35

<sup>7</sup> P. J. Deuss; K. Barta; J. G. de Vries. *Catal Sci. Technol.* **2014**, 4, 1174

<sup>8</sup> J. Zakzeski; P. C. A. Bruijninx; A. L. Jongerius; B. M. Weckhuysen. *Chem. Rev.* **2010**, 110, 3552

<sup>9</sup> F. S. Chakar; A. J. Ragauskas. *Ind. Crops. Prod.* **2004**, 20, 131

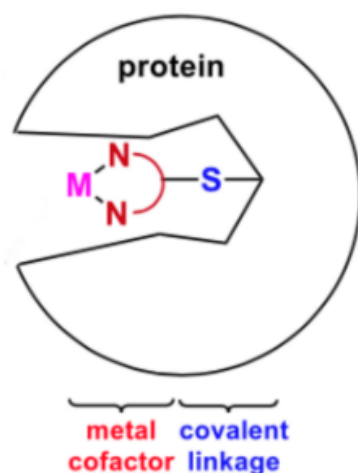
<sup>10</sup> J. Ralph. *Phytochem. Rev.* **2004**, 3, 29

- 
- <sup>11</sup> D. W. S. Wong. *Appl. Biochem. Biotechnol.* **2009**, 157, 174
- <sup>12</sup> R. Ten-Have; P. J. Teunissen; *Chem. Rev.* **2001**, 101, 3397
- <sup>13</sup> C. Y. Liew; A. Husaini; H. Hussain; S. Muid; K. C. Liew; H. A. Roslan. *World J Microbiol Biotechnol.* **2011**, 27, 1457
- <sup>14</sup> H. Sorensen, *J. Gen. Microbiol.* **1962**, 27, 21
- <sup>15</sup> D. L. Crawford. *Appl Environ Microbiol.* **1978**, 35, 1041
- <sup>16</sup> Y. Chen; L. Chai; C. Tang; Z. Yang; Y. Zheng; Y. Shi; H. Zhang. *Bioresour Technol.* **2012**, 123, 682
- <sup>17</sup> T. D. H. Bugg, M. Ahmad, E. M. Hardiman, R. Rahmanpour. *Nat. Prod. Rep.*, **2011**, 28, 1883
- <sup>18</sup> M. E. Brown, M. CY. Chang. *Curr. Opin. Chem. Biol.* **2014**, 19, 1
- <sup>19</sup> C. Crestini, M. Crucianelli, M. Orlandi, R. Saladino. *Catal. Today*, **2010**, 156, 8
- <sup>20</sup> H. Lange; S. Decina; C. Crestini. *Eur. Polym. J.* **2013**, 49, 1151
- <sup>21</sup> W.A. Herrmann; R. W. Fischer. *J. Am. Chem. Soc.* **1995**, 117, 3323
- <sup>22</sup> C. Crestini, . Pro; V. Neri, R. Saladino. *Bioorg. Med. Chem.* **2005**, 13, 2569
- <sup>23</sup> C. Crestini, M. C. Caponi. *Bioorg. Med. Chem.* **2006**, 14, 5292
- <sup>24</sup> R. G. Harms; I. I. E. Markovits; M. Drees. H. C. Mult; W. A. Herrmann; M. Cokoja; F. E. Kuhn. *ChemSusChem.* **2014**, 7, 429
- <sup>25</sup> J. Zhang; Y. Liu; S. Chiba; T. Loh. *Chem. Commun.* **2013**, 49, 11439
- <sup>26</sup> A. Haikaraiene; J. Sipilä; P. Pietikäinen; A. Pajunen; I. Mutikainen. *J Chem Soc Dalton Trans.* **2001**, 7, 991
- <sup>27</sup> L. Zoia; C. Canevali; M. Orlandi; E. L. Tolppa; J. M. F. Sipilä. *J Biol Res.* **2008**, 3, 21
- <sup>28</sup> J. Zakzeski; P. C. Bruijninx; A. L. Jongerius; B. M. Weckhuysen. *Chem. Rev.* **2010**, 110, 3552
- <sup>29</sup> J. M. W. Chan; S. Bauer; H. Sorek; S. Sreekumar; K. Wang; F. D Toste. *ACS Catal.* **2013**, 3, 1369
- <sup>30</sup> A. Rahimi, A. Azarpira, H. Kim, J. Salph, S. Stahl, *J. Am. Chem. Soc.* **2013**, 135, 6415
- <sup>31</sup> V. E Tarabanko; N. A. Fomova; B. N. Kuznetsov; N. M. Ivanchenko; A. V. Kudryashev. *React. Kinet. Catal. Lett.* **1995**, 55, 161
- <sup>32</sup> J. M. Hoover; S. S. Stahl. *J. Am. Chem. Soc.* **2011**, 133, 16901
- <sup>33</sup> P. Gamez; I. W. C. E. Arends; R. A. Sheldon; J. Reedijk. *Adv. Synth. Catal.* **2004**, 346, 805
- <sup>34</sup> B. Sedai; C. Diaz-Urrutia; T. Baker; R. Wu; L. A. Skills; S. Hanson. *ACS Catal.* **2011**, 1, 794

- 
- <sup>35</sup> X. Wang; R. Rinaldi. *ChemSusChem*. **2012**, 5, 1455
- <sup>36</sup> C. Crestini; M. Crucianelli; M. Orlandi; R. Saladino. *Catal. Today*, **2010**, 156, 8
- <sup>37</sup> P. Zucca; A. Rescigno; A. C. Rinaldi; E. Sanjust. *J. Mol. Catal. A: Chem.* **2013**, 388, 2
- <sup>38</sup> S. I. Mann, T. Heinisch, T. R. Ward, A. S. Borovik, *J Am Chem Soc* **2017**, 139, 17289
- <sup>39</sup> A. Rahimi; A. Azarpira; H. Kim; J. Ralph; S. S. Stahl. *J. Am. Chem. Soc.* **2013**, 135, 6415
- <sup>40</sup> Buendia, J; Mottweiler, J; Bolm, C. *Chem. Eur. J.* **2011**, 17, 13877
- 41
- <sup>42</sup> M. T. Reetz, R. Soni, L. Fernandez, Y. Gumulya, J. D. Carballeira. *Chem. Commun.* **2010**, 46, 8657
- <sup>43</sup> J. S. Olsen, A. J. Mathews, R. J. Rohlf, B. A. Springer, K. D. Egeberg, S. G. Sligar, J. Tame, K. Nagai, *Nature*, **1988**, 336, 265
- <sup>44</sup> E. A. Span, D. L. M. Suess, M. C. Deller, R. D. Britt, M. A. Marletta, *ACS Chem Biol*, **2017**, 12, 1095
- <sup>45</sup> P. J. Deuss, G. Popa, C. H. Botting, W. Laan, P. C. J. Kamer, *Angew. Chem. Eur. J.* **2010**, 49, 5315

## Chapter 5 - Summary and Conclusion

This thesis has studied the design, synthesis, application and optimisation of ArM systems. The artificial metalloenzyme is based on a concept which is depicted in figure 1. The catalytic performance of these hybrid catalysts has been optimized by using orthogonal structural-diversity generating procedures: molecular biology for the optimization of the protein structure and synthetic chemistry to tune the structure of the ligand.



*Figure 1- Schematic representation of artificial metalloenzymes obtained by covalent modification*

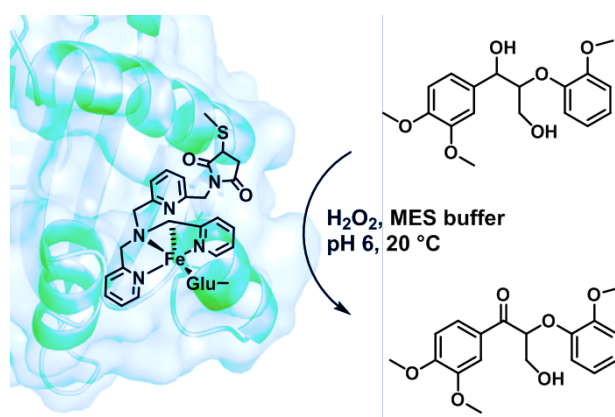
In Chapter 2 a library of ArMs was successfully made using a designed cofactor library and two variants of SCP-2L as the protein scaffold. When using ArM systems in catalysis, it is important to have an accurate method for characterising the catalyst as the synthetic cofactor can cause typical protein protocols to become unreliable. When studying modified SCP-2L A100C in chapter 2, it was shown that normal methods for determining protein concentration such as a Bradford assay could not be relied upon. Instead an accurate extinction coefficient was determined experimentally and varied greatly depending on which cofactor was present.

In chapter 3, the synthesis and application of an ArM for transition metal catalysed reactions not performed by natural enzymes was described. In hydroformylation reactions, an engineered thermostable ArM improved the TON over 5-fold. By improving the stability



of the protein scaffold in the ArM, it provides a much wider selection of reaction conditions that can be used and therefore increases the chances of finding a more active catalyst.

ArM systems were utilised in looking at lignin model compound oxidation in chapter 4. In this scenario the incorporation of the TPA ligand within a protein scaffold gave rise to selective benzylic alcohol oxidation (Figure 2) The observation that the conversion can be increased by mutations in the surrounding protein environment, suggests that future genetic optimisation of the protein will lead to more active enzymes that can be used for larger scale processes and ultimately within the biorefinery.



*Figure 2- selective benzylic alcohol oxidation of lignin model compound*

Overall, this study demonstrates that judiciously chosen protein binding scaffolds can be adapted to obtain enzymes that provide the reactivity of the introduced metal centre combined with specifically intended product selectivity. Such approaches should inspire the development of new catalytic systems that utilize the powerful substrate binding capabilities of proteins. In the long term, as chemogenetic optimisation is used to improve activity, this could open the door to a new era of biocatalytic chemical production.

## **Acknowledgements**

First of all, I would like to thank my supervisor Prof Paul Kamer for the opportunity to work on this interesting project. I am very grateful for everything I have learnt from Paul throughout my PhD studies.

I am incredibly grateful for all the advice from Dr Amanda Jarvis, I have learnt so much from being around her during my time in St Andrews.

I would also like to thank the whole Kamer group for their help, support and creating a nice work environment.

Thank you also to Jane Potter who was always so happy to help with any problems I encountered.

I am also thankful that I was able to go work with Prof D Janssen at Groningen University and Henk-Jan Joosten at BioProduct to learn about bioengineering which was invaluable to my PhD research.

Finally, I am grateful to EastBio for providing the funds for me to be able to carry out my PhD studies.