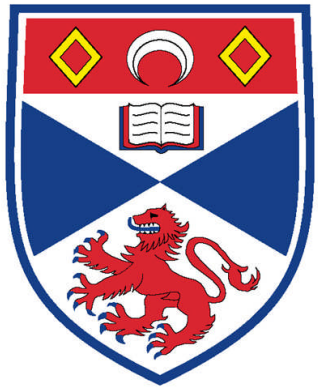
Artificial Metalloenzymes; Modified Proteins as Tuneable Transition Metal Catalysts



University
of
St Andrews

School of Chemistry
Fife, Scotland

Peter J. Deuss

February 2011

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

Supervisor: Professor Paul C. J. Kamer

Declarations

I, Peter Joseph Deuss, hereby certify that this thesis, which is approximately 67000 words in length, has been written by me, that it is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree.

I was admitted as a research student in September, 2006 and as a candidate for the degree of Doctor in Philosophy in September, 2007; the higher study for which this is a record was carried out in the University of St Andrews between 2006 and 2011.

Date …… Signature of candidate ……..…

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

Date …… Signature of supervisor ……..…

In submitting this thesis to the University of St Andrews I understand that I am 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. I also understand 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 my thesis will be electronically accessible for personal or research use unless exempt by award of an embargo as requested below, and that the library has the right to migrate my thesis into new electronic forms as required to ensure continued access to the thesis. I have obtained any third-party copyright permissions that may be required in order to allow such access and migration, or have requested the appropriate embargo below.

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

Embargo on both all and part of the printed copy and electronic copy for the same fixed period of 2 years on the following ground:

- Publication would preclude future publication;

Date …… Signature of candidate……..…Signature of supervisor ……

Abstract

This thesis describes the design, synthesis and application of artificial metalloenzymes for transition metal catalysed reactions not performed by natural enzymes. Unique cysteine containing protein templates were covalently modified with transition metal ligand complexes that generate catalytic activity, which allows for the use of virtually any protein template. SCP-2L was selected as template for the linear hydrophobic tunnel that traverses the protein, which has high affinity for linear aliphatic molecules.

The use of catalysts based on this protein to induce increased activity in the biphasic hydroformylation of linear α-olefins is investigated in this work. For this purpose, unique cysteine containing mutants of SCP-2L were modified with phosphine ligands by application of a novel bioconjugation procedure. Application of rhodium adducts of the phosphine modified protein constructs led to up to a 100 fold increase of the turn over numbers was measured compared to a Rh/TPPTS model system which is used in industry. Furthermore, good selectivity towards the linear product was observed. If it can be confirmed that the found catalytic results truly are the result of substrate encapsulation by the protein scaffold, this system represents the first rationally designed artificial metalloenzyme which exploits the shape selectivity of the protein scaffold to direct the outcome of a catalytic reaction.

In addition, a study was performed for the development of enantioselective artificial metalloenzymes. Nitrogen ligands were covalently introduced in SCP-2L and the obtained conjugates were applied in the copper catalysed Diels-Alder and Michael addition reaction. A promising 25% ee was found for the Diels-Alder reaction between azachalcone and cyclopentadiene using one of the created constructs.

Further development of these catalyst systems with the use of both synthetic (e.g. optimisation of ligand structure) and biomolecular tools (e.g. optimisation of protein environment) for optimisation can lead to very efficient and enantioselective conversions in the future.

Table of Contents

List of abbreviation

Ac = Acetyl

acac = Acetylacetonate

ALBP = Adipocyte lipid binding protein

CD = Circular dichroism

CDI = 1,1'-carbonyldiimidazole

CoA = Coenzyme A

cod = 1,5-cyclooctadiene

DCC = *N*,*N*'-Dicyclohexylcarbodiimide

DCM = Dichloromethane

Dha = Dehydroalanine

DMF = *N*,*N*-Dimethylformamide

DMSO = Dimethylsulfoxide

DNA = Deoxyribonucleic acids

DTNB = 5,5′-dithio-bis(2-nitrobenzoic acid (Ellman's reagent)

DTT = Dithiothreitol

EDC = *N*-(3-dimethylaminopropyl)-*N*'-ethylcarbodiimide

EDTA = Ethylenediaminetetraacetic acid

ee = Enantiomeric excess

EPR = Electron paramagnetic resonance

ES($^{+}$) = Electron spray ionization (positive mode)

HEPES = 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

HOBT = Hydroxybenzotriazole

HPLC = High-performance liquid chromatography

HSA = Human serum albumin

iBCF = Isobutylchloroformate

ICP = Inductively coupled plasma

IPA = Isopropanol

IPTG = Isopropyl-1-thio-β-D-galactopyranoside

KPi = Potassium phosphate (buffer)

LCMS = Liquid chromatography-mass spectrometry

MES = 2-(*N*-morpholino)ethanesulfonic acid

MFE = Multifunctional enzyme

MS = Mass-spectrometry

MSH = *O*-mesitylenesulfonylhydroxylamine

NHS = *N*-hydroxysuccinimide

NMR = Nuclear magnetic resonance

Orn = Ornithine

PB = Production broth (see section 2.5.3 for contents)

pdb = protein data bank (file code)

PYP = Photoactive yellow protein

Pyr-C12 = 12-pyrene dodecanoic acid

RNA = Ribonucleic acids

TCEP = Tris(2-carboxyethyl)phosphine

TEV = Tobacco etch virus

TFA = Trifluoroacetic acid

THF = Tetrahydrofuran

TLC = Thin layer chromatography

TMS = Trimethyl silane

TNB = 5-thio-2-nitrobenzoic acid

TON = Turnover number

TPP = Triphenylphosphine

TPPDS = Triphenylphosphine disulfonate

TPPMS = Triphenylphosphine monosulfate

TPPTS = Triphenylphosphine trisulfonate

Tris = Tris(hydroxymethyl)aminomethane

SCP2 = Sterol carrier protein 2

SCP-2L = Sterol carrier protein 2-like

UV = Ultraviolet

Watergate = Water suppression by gradient tailored excitation

Protein mutant coding:

[Name of the protein] [Original amino acid][Location][Newly introduced amino acid]

For example name SCP-2L V83C indicates that the valine (V) at position 83 in SCP-2L is replaced by a cysteine (C) to obtain the named mutant protein.

Compound codes:

A# = Aldehydes for modification to hydride modified proteins other than phosphines

B# = Branched hydroformylation products

C# = Carboxylic acid containing phosphine ligands

E# = Additional compounds

H# = Rhodium complexes involved in the catalytic hydroformylation

I# = Intermediate products in the synthesis towards ligand for protein modification

K# = "Click" phosphines

L# = Linear hydroformylation products

M# = Maleimide containing compounds for selective bioconjugation to free cysteines

N# = Nitrogen ligands

P# = Phosphino aldehydes

S# = DNA-binding ligands

T# = Tetrazoles

Chapter 1. Introduction to bioinspired catalysis

Bioinspired catalyst design and artificial metalloenzymes

For content, please refer to the publication of this chapter as review article in Chemistry, A - European Journal:

P. J. Deuss, R. Den Heeten, W. Laan, P. C. J. Kamer, *Chem.—Eur. J.* **2011**, *17*, 4680

No rights for online publication of this chapter were obtained.

For content, please refer to the publication of this chapter as review article in Chemistry, A - European Journal:
P. J. Deuss, R. Den Heeten, W. Laan, P. C. J. Kamer, *Chem.—Eur. J.* **2011**, *17*, 4680

No rights for online publication of this chapter were obtained.

For content, please refer to the publication of this chapter as review article in Chemistry, A - European Journal:
P. J. Deuss, R. Den Heeten, W. Laan, P. C. J. Kamer, *Chem.—Eur. J.* **2011**, *17*, 4680

For content, please refer to the publication of this chapter as review article in Chemistry, A - European Journal:
P. J. Deuss, R. Den Heeten, W. Laan, P. C. J. Kamer, *Chem.—Eur. J.* **2011**, *17*, 4680

No rights for online publication of this chapter were obtained.

For content, please refer to the publication of this chapter as review article in Chemistry, A - European Journal:
P. J. Deuss, R. Den Heeten, W. Laan, P. C. J. Kamer, *Chem.—Eur. J.* **2011**, *17*, 4680

No rights for online publication of this chapter were obtained.

For content, please refer to the publication of this chapter as review article in Chemistry, A - European Journal:
P. J. Deuss, R. Den Heeten, W. Laan, P. C. J. Kamer, *Chem.—Eur. J.* **2011**, *17*, 4680

No rights for online publication of this chapter were obtained.

For content, please refer to the publication of this chapter as review article in Chemistry, A - European Journal:
P. J. Deuss, R. Den Heeten, W. Laan, P. C. J. Kamer, *Chem.—Eur. J.* **2011**, *17*, 4680

No rights for online publication of this chapter were obtained.

For content, please refer to the publication of this chapter as review article in Chemistry, A - European Journal:
P. J. Deuss, R. Den Heeten, W. Laan, P. C. J. Kamer, *Chem.—Eur. J.* **2011**, *17*, 4680

No rights for online publication of this chapter were obtained.

For content, please refer to the publication of this chapter as review article in Chemistry, A - European Journal:
P. J. Deuss, R. Den Heeten, W. Laan, P. C. J. Kamer, *Chem.—Eur. J.* **2011**, *17*, 4680

No rights for online publication of this chapter were obtained.

For content, please refer to the publication of this chapter as review article in Chemistry, A - European Journal:
P. J. Deuss, R. Den Heeten, W. Laan, P. C. J. Kamer, *Chem.—Eur. J.* **2011**, *17*, 4680

No rights for online publication of this chapter were obtained.

For content, please refer to the publication of this chapter as review article in Chemistry, A - European Journal:
P. J. Deuss, R. Den Heeten, W. Laan, P. C. J. Kamer, *Chem.—Eur. J.* **2011**, *17*, 4680

No rights for online publication of this chapter were obtained.

For content, please refer to the publication of this chapter as review article in Chemistry, A - European Journal:
P. J. Deuss, R. Den Heeten, W. Laan, P. C. J. Kamer, *Chem.—Eur. J.* **2011**, *17*, 4680

No rights for online publication of this chapter were obtained.

For content, please refer to the publication of this chapter as review article in Chemistry, A - European Journal:
P. J. Deuss, R. Den Heeten, W. Laan, P. C. J. Kamer, *Chem.—Eur. J.* **2011**, *17*, 4680

No rights for online publication of this chapter were obtained.

For content, please refer to the publication of this chapter as review article in Chemistry, A - European Journal:
P. J. Deuss, R. Den Heeten, W. Laan, P. C. J. Kamer, *Chem.—Eur. J.* **2011**, *17*, 4680

No rights for online publication of this chapter were obtained.

For content, please refer to the publication of this chapter as review article in Chemistry, A - European Journal:
P. J. Deuss, R. Den Heeten, W. Laan, P. C. J. Kamer, *Chem.—Eur. J.* **2011**, *17*, 4680

No rights for online publication of this chapter were obtained.

For content, please refer to the publication of this chapter as review article in Chemistry, A - European Journal:
P. J. Deuss, R. Den Heeten, W. Laan, P. C. J. Kamer, *Chem.—Eur. J.* **2011**, *17*, 4680

No rights for online publication of this chapter were obtained.

For content, please refer to the publication of this chapter as review article in Chemistry, A - European Journal:
P. J. Deuss, R. Den Heeten, W. Laan, P. C. J. Kamer, *Chem.—Eur. J.* **2011**, *17*, 4680

No rights for online publication of this chapter were obtained.

For content, please refer to the publication of this chapter as review article in Chemistry, A - European Journal:
P. J. Deuss, R. Den Heeten, W. Laan, P. C. J. Kamer, *Chem.—Eur. J.* **2011**, *17*, 4680

No rights for online publication of this chapter were obtained.

For content, please refer to the publication of this chapter as review article in Chemistry, A - European Journal:
P. J. Deuss, R. Den Heeten, W. Laan, P. C. J. Kamer, *Chem.—Eur. J.* **2011**, *17*, 4680

No rights for online publication of this chapter were obtained.

For content, please refer to the publication of this chapter as review article in Chemistry, A - European Journal:
P. J. Deuss, R. Den Heeten, W. Laan, P. C. J. Kamer, *Chem.—Eur. J.* **2011**, *17*, 4680

No rights for online publication of this chapter were obtained.

For content, please refer to the publication of this chapter as review article in Chemistry, A - European Journal:
P. J. Deuss, R. Den Heeten, W. Laan, P. C. J. Kamer, *Chem.—Eur. J.* **2011**, *17*, 4680

No rights for online publication of this chapter were obtained.

For content, please refer to the publication of this chapter as review article in Chemistry, A - European Journal:
P. J. Deuss, R. Den Heeten, W. Laan, P. C. J. Kamer, *Chem.—Eur. J.* **2011**, *17*, 4680

No rights for online publication of this chapter were obtained.

For content, please refer to the publication of this chapter as review article in Chemistry, A - European Journal:
P. J. Deuss, R. Den Heeten, W. Laan, P. C. J. Kamer, *Chem.—Eur. J.* **2011**, *17*, 4680

No rights for online publication of this chapter were obtained.

For content, please refer to the publication of this chapter as review article in Chemistry, A - European Journal:
P. J. Deuss, R. Den Heeten, W. Laan, P. C. J. Kamer, *Chem.—Eur. J.* **2011**, *17*, 4680

No rights for online publication of this chapter were obtained.

For content, please refer to the publication of this chapter as review article in Chemistry, A - European Journal:
P. J. Deuss, R. Den Heeten, W. Laan, P. C. J. Kamer, *Chem.—Eur. J.* **2011**, *17*, 4680

No rights for online publication of this chapter were obtained.

For content, please refer to the publication of this chapter as review article in Chemistry, A - European Journal:
P. J. Deuss, R. Den Heeten, W. Laan, P. C. J. Kamer, *Chem.—Eur. J.* **2011**, *17*, 4680

No rights for online publication of this chapter were obtained.

For content, please refer to the publication of this chapter as review article in Chemistry, A - European Journal:
P. J. Deuss, R. Den Heeten, W. Laan, P. C. J. Kamer, *Chem.—Eur. J.* **2011**, *17*, 4680

No rights for online publication of this chapter were obtained.

For content, please refer to the publication of this chapter as review article in Chemistry, A - European Journal:
P. J. Deuss, R. Den Heeten, W. Laan, P. C. J. Kamer, *Chem.—Eur. J.* **2011**, *17*, 4680

No rights for online publication of this chapter were obtained.

For content, please refer to the publication of this chapter as review article in Chemistry, A - European Journal:
P. J. Deuss, R. Den Heeten, W. Laan, P. C. J. Kamer, *Chem.—Eur. J.* **2011**, *17*, 4680

No rights for online publication of this chapter were obtained.

For content, please refer to the publication of this chapter as review article in Chemistry, A - European Journal:
P. J. Deuss, R. Den Heeten, W. Laan, P. C. J. Kamer, *Chem.—Eur. J.* **2011**, *17*, 4680

No rights for online publication of this chapter were obtained.

For content, please refer to the publication of this chapter as review article in Chemistry, A - European Journal:
P. J. Deuss, R. Den Heeten, W. Laan, P. C. J. Kamer, *Chem.—Eur. J.* **2011**, *17*, 4680

For content, please refer to the publication of this chapter as review article in Chemistry, A - European Journal:
P. J. Deuss, R. Den Heeten, W. Laan, P. C. J. Kamer, *Chem.—Eur. J.* **2011**, *17*, 4680

No rights for online publication of this chapter were obtained.

For content, please refer to the publication of this chapter as review article in Chemistry, A - European Journal:
P. J. Deuss, R. Den Heeten, W. Laan, P. C. J. Kamer, *Chem.—Eur. J.* **2011**, *17*, 4680

No rights for online publication of this chapter were obtained.

For content, please refer to the publication of this chapter as review article in Chemistry, A - European Journal:
P. J. Deuss, R. Den Heeten, W. Laan, P. C. J. Kamer, *Chem.—Eur. J.* **2011**, *17*, 4680

1.7 Outlook on this thesis

The true power of enzymes, i.e. their tremendous rate acceleration, high substrate selectivity and selective transformation of multifunctional unprotected substrates, is dependent on highly organised protein substrate interactions. Therefore, combining the shape selectivity and powerful substrate recognition of proteins with optimal catalyst performance achieved via rational ligand design requires the use of protein structures that can encapsulate the substrate to enforce optimal orientation with respect to the artificial cofactor (Figure 4).

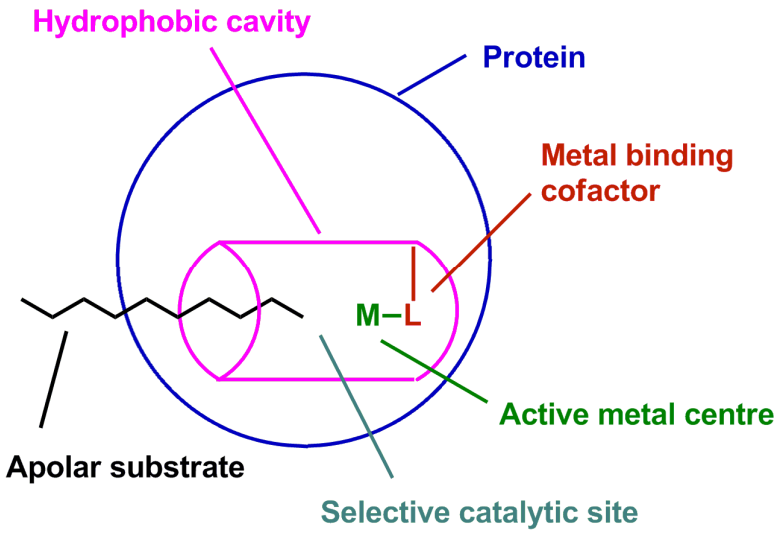


Figure 4. Protein induced shape selective binding of apolar substrates in aqueous environment, obtained by covalent attachment of transition metal catalyst to a selected protein template.

Covalent anchoring allows exploitation of protein structures, which have the desired structural properties for efficient substrate protein interactions, because, unlike for supramolecular assembled artificial metalloenzymes, the binding site is not required for the catalyst assembly. This can lead to the full exploitation of the shape-selectivity of promising protein structures.

In this work we set out to design and create such catalyst systems. The envisioned catalyst system to which the main part of this thesis is dedicated is designed for application as shape selective catalyst in the hydroformylation of higher olefins. This reaction is not performed in nature and thus no natural hydroformylation enzymes are known. The protein part of this catalyst was selected based on its capacity to bind the substrates for this reaction, which will offer unique possibilities to obtain unequalled (chemo)selectivity and substrate specificity. For this purpose, we envisaged the use of a protein with binding affinity towards linear aliphatic substrates in combination with rhodium-phosphine complexes as the catalytically active centre. The approach taken for the development process described in this thesis can be divided into seven stages which are used as directive in the following chapters (Table 1).

Table 1. The seven stages involved in the development of the artificial metalloenzymes described in this thesis.

Stage	Description
1	Protein selection
2	Design of unique cysteine protein templates
3	Production and purification of the unique cysteine templates
4	Site selective conjugation of the ligands to the unique cysteine templates
5	Transition metal coordination to the ligand containing protein conjugates
6	Application of the transition metal conjugates in catalysis
7	Optimization

Chapter 2 describes the first three stages in the development process of a shape selective system. The first stage is the choice of the protein structure based on desired

characteristics like substrate binding, structural characterisation of the binding site, stability etc. The second stage is the conversion of the desired protein into a unique cysteine containing template using site directed mutagenesis. The unique cysteine is required for site selective anchoring of the catalytically active transition metal ligand system. The locations of the cysteines inside the proteins have to be carefully considered as the transition metal should be placed in close proximity to the binding site of the protein to ensure selective catalysis. The third stage is the production and purification of the protein templates. This production by bacterial expression and purification has to be optimized to ensure appropriate scale production to be able to obtain enough material to perform combinatorial screening.

Chapter 3 describes the fourth stage of the development process which is the site selective bioconjugation of phosphine ligands. This key step in the development process has to be achieved with high yield and selectivity as purification of the conjugates is problematic. In addition, the modification procedure needs to be generally applicable to any desired unique cysteine protein template to ensure applicability of protein structure optimisation techniques.

Chapter 4 describes the rhodium complex formation step and application of the rhodium conjugates in hydroformylation. The main application will be the hydroformylation of linear aliphatic 1-alkenes in a biphasic process for which initial catalysis experiments will be reported. Natural enzymes underwent many millennia of the evolutionary process to become the efficient enzymes known to us. It is not likely that the particular system prepared in this thesis will represent the optimal structure for our purposes. Therefore, the catalyst optimisation is the very important last step of the development process and will have to be performed to obtain the envisioned highly selective and active catalyst system. The foundations and opportunities for this final process will be presented throughout this thesis.

The fifth chapter will describe the application of created unique cysteine containing proteins in asymmetric reactions using copper-nitrogen donor ligand systems as catalytically active metal centres. In this approach the chiral environment of the protein is utilized to induced enantioselectivity in catalytic reactions. A model system will be created that can be optimised using all available synthetic and biochemical tools.

Finally, the sixth chapter will give a summary, discussion and final conclusion on the taken approach and methodology applied in this thesis.

No rights for online publication of this chapter were obtained.

For content, please refer to the publication of this chapter as review article in Chemistry, A - European Journal:
P. J. Deuss, R. Den Heeten, W. Laan, P. C. J. Kamer, *Chem.—Eur. J.* **2011**, *17*, 4680

No rights for online publication of this chapter were obtained.

For content, please refer to the publication of this chapter as review article in Chemistry, A - European Journal:
P. J. Deuss, R. Den Heeten, W. Laan, P. C. J. Kamer, *Chem.—Eur. J.* **2011**, *17*, 4680

No rights for online publication of this chapter were obtained.

For content, please refer to the publication of this chapter as review article in Chemistry, A - European Journal:
P. J. Deuss, R. Den Heeten, W. Laan, P. C. J. Kamer, *Chem.—Eur. J.* **2011**, *17*, 4680

No rights for online publication of this chapter were obtained.

For content, please refer to the publication of this chapter as review article in Chemistry, A - European Journal:
P. J. Deuss, R. Den Heeten, W. Laan, P. C. J. Kamer, *Chem.—Eur. J.* **2011**, *17*, 4680

No rights for online publication of this chapter were obtained.

For content, please refer to the publication of this chapter as review article in Chemistry, A - European Journal:
P. J. Deuss, R. Den Heeten, W. Laan, P. C. J. Kamer, *Chem.—Eur. J.* **2011**, *17*, 4680

No rights for online publication of this chapter were obtained.

For content, please refer to the publication of this chapter as review article in Chemistry, A - European Journal:
P. J. Deuss, R. Den Heeten, W. Laan, P. C. J. Kamer, *Chem.—Eur. J.* **2011**, *17*, 4680

No rights for online publication of this chapter were obtained.

For content, please refer to the publication of this chapter as review article in Chemistry, A - European Journal:
P. J. Deuss, R. Den Heeten, W. Laan, P. C. J. Kamer, *Chem.—Eur. J.* **2011**, *17*, 4680

Chapter 2. Unique cysteine containing protein templates

Design and production of protein templates for the synthesis of artificial metalloenzymes

2.1 Abstract

This chapter describes the synthesis of protein templates for the development of artificial metalloenzymes that can bind substrates of interest for catalytic reactions. Various unique cysteine containing mutants of lipid binding proteins mSCP2 and SCP-2L were designed and created. Large scale production of SCP-2L V83C and A100C was successfully achieved, with yields of more than 30 mg/L of bacterial culture. CD-spectroscopy and ^{1}H-NMR showed that the chosen mutations do not influence the tertiary structures and fluorescent binding studies showed retention of substrate binding capability. Variable temperature CD-spectroscopy showed that these proteins have remarkable temperature stability.

2.2 Introduction

2.2.1 Design of protein templates for artificial metalloenzyme synthesis

2.2.1.1 Protein template selection

A key part in the design and synthesis of artificial metalloenzymes is the choice of the protein template.[1] A set of demands was created reflecting desired characteristic for the approach taken in this work, to be able to select a suitable protein host from the millions of available proteins. 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.[2] Hence, the first demand is that the protein has to be structurally characterised by either X-ray crystallography or solution NMR to have a well-defined basic structure on which modifications will be applied. Secondly, the 3D structure has to contain a hydrophobic pocket for binding apolar substrates of our interest. Thirdly, a clearly allocated site for binding must have been identified to be able to rationalize the position for introduction of our transition metal catalyst. Fourthly, it is required that the protein can be produced in considerable amounts to be able to produce sufficient material for screening purposes. Finally stability of the protein is important. The protein must be resistant to the conditions of the catalytic reactions, which would preferably mean sufficient temperature stability and organic solvent tolerance.

2.2.1.2 Protein engineering to create a unique cysteine containing protein template

The next step is to modify the selected protein template with our transition metal catalyst system, which will be done through bioconjugation with a unique cysteine. Cysteine is a thiol containing amino-acid with a particular nucleophilic character that is generally exploited for site-selective bioconjugation.[3] This approach has already been applied to some extent by Distefano *et al.*.[1b, 4] The presence of a unique cysteine is desired to incorporate one transition metal binding ligand into the protein at a unique position to benefit from the selective binding properties of the protein. Hence, when a desirable protein has been identified this protein may have to be modified to position the cysteine at an appropriate location. The unique cysteine will be introduced at a selected position using site directed mutagenesis on a DNA-vector coding for the native protein. In addition, site directed mutagenesis of the same gene

will have to be employed when the protein template also contains other undesired cysteines. Site directed mutagenesis is performed using commercially available kits which describe requirement for primers used and PCR conditions. The modified plasmid containing the DNA coding for a unique cysteine containing mutant of the original protein will than be transferred into an *E. Coli* production strain and the protein will be produced using pET vectors and T7 promoters as this commonly leads to high expression levels.[5]

2.2.1.3 Purification of the protein template

A high yielding and easy purification method is essential for catalyst screening, so the preferred method for purification would be to use affinity chromatography. Well known affinity tags are maltose,[6] glutathione-S-transferase,[7] chitin[8] and histidine tags.[9] The latter will be used in this work in combination with nickel affinity chromatography.

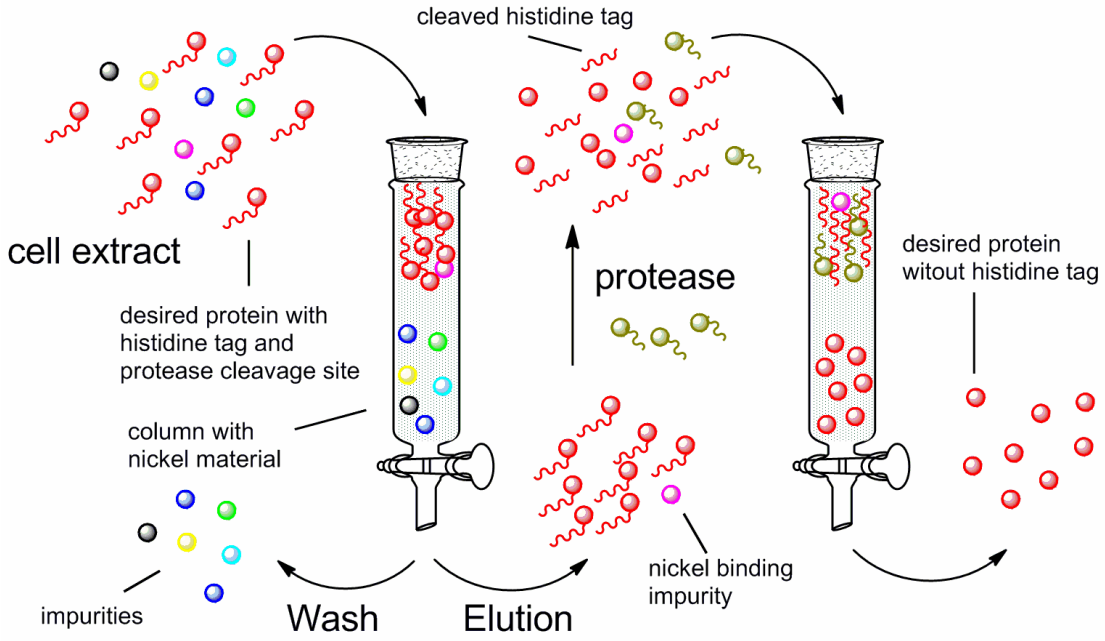


Figure 1. Schematic overview of nickel affinity purification including tag cleavage by protease.

For this purification procedure the desired protein is equipped with a histidine tag consisting of 6 sequential histidines that have high affinity for binding to nickel ions. This histidine tag is usually introduced on the N-terminus with an amino acid linker of choice by an extra DNA sequence on the plasmid. This linker can be designed to

contain a protease cleavage site for removal of the tag.[10] For our approach this removal is essential as the histidine tag also has high affinity for transition metals which can interfere with our site selective transition metal introduction.[11] The overall purification process is displayed in Figure 1. In the first step of nickel affinity chromatography the cell-free extract will be loaded on a column containing immobilized Ni^{2+} ions. After washing away contaminating proteins, the target-protein is eluted by applying a buffer typically containing a high imidazole concentration to compete with the histidine tag for nickel binding. The histidine tag will then be cleaved off by a histidine tag containing protease followed by a second round of nickel affinity chromatography to separate the removed tag, the protease and any other nickel binding protein impurities still present after the first run from the desired protein.

2.2.2 SCP as protein templates

In this project there is a special interest in proteins that can bind substrates of relevance to the catalytic processes of our choice. Linear aliphatic substrates are involved in many catalytic processes for which homogeneous transition metal catalysts are applied. A few examples of such processes are hydroformylation of linear alkenes to linear and branched aldehydes, oxidation of alkanes or alkenes and oligomerisation or polymerisation of alkenes. Selectivity in such reactions is a key issue that we will address by using our protein based catalysts. This will be done by using shape-selectivity, an approach which is often used in heterogeneous catalysis. For example selective α-oxidation of alkanes can be achieved by using porous materials containing appropriate size pores with catalysts embedded at key positions.[12] This concept will be transferred by us to protein based catalysis by choosing proteins with apolar cavities of appropriate size and shapes and subsequent optimization by mutagenesis.

The protein databank contains over 50,000 structurally characterized proteins,[13] providing plenty of choice for the selection of a protein-template according to the demands described in the previous section. A class of proteins that caught our attention is sterol carrier proteins (SCP's). Proteins of this class, which were found to be involved in the biosynthesis of cholesterol, often contain hydrophobic binding pockets. Particular interest was developed for sterol carrier protein 2 (SCP2) and analogues, because for several of these protein linear hydrophobic tunnel shaped

binding pockets were reported (Figure 2).[14] SCP2 proteins from several organisms have been well studied. These proteins are also known as a non-specific lipid transfer protein (nsLTP), because in addition to high binding affinity for cholesterol and other steroids also high affinity for various phospholipids was found.[15] These proteins were also reported to bind long-chain fatty acids, which is of interest for our purposes.[16] It has been shown that these proteins have evolved to transfer substrates to assist enzyme complexes in α-oxidation of branched chain fatty acids and β-oxidation of long chain fatty acids.[17] This was also confirmed by the finding that SCP2 type proteins are often expressed as fusion proteins with enzymes involved in oxidation pathways for fatty acid metabolism.[14c, 18]

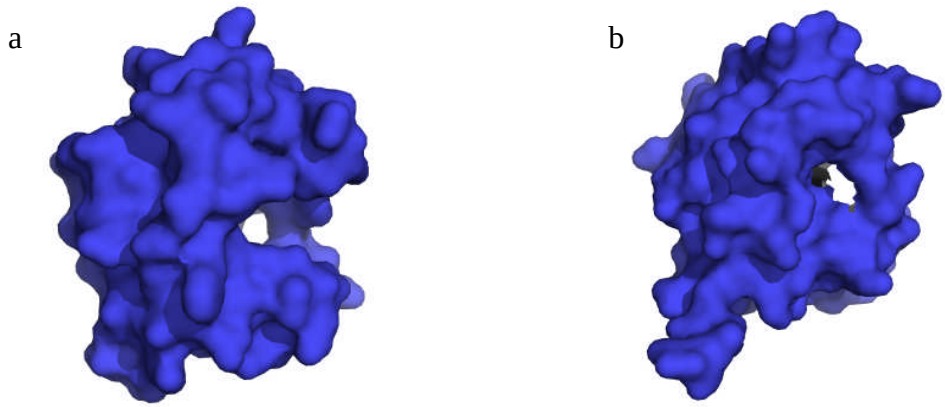


Figure 2. Crystal structures of a) SCP-2L excluding the triton-X molecule[19] b) osSCP2 excluding side chains M109 Q112 K118 P120 that cap the hydrophobic cavity.[20]

Various analogues of SCP2 have been reported since its first isolation from rat liver.[21] Several examples from mammals are osSCP2 from *Oryctolagus cuniculus* (rabbit),[20] hSCP2 and SCP-2L from *Homo sapien*s (human)[19, 22] and other SCP2 analogues from beef liver.[23] Additional analogues are also found in insects like aaSCP2 from *Aedes aegypti* (mosquito),[24] in plants as SCP-2 from *Arabidopsis thaliana* (thale cress)[25] and in bacteria like *Methanococcus jannaschii* (thermophilic methanogen and first archea of which the genome was sequenced).[26] Structural characterization has shown that the general structure of these proteins contains a rigid β-sheet backbone structure and three C-terminal alpha helixes involved in substrate recognition which can open up to bind aliphatic substrates inside the proteins hydrophobic core. This hydrophobic core was well visualised in the structures of SCP-2L and osSCP2 (Figure 2). Small structural variations in the main beta-sheet

backbone and flexible C-terminal alpha helixes are present in all these analogues (Figure 3). Despite their structural similarity, differences have been found in the binding modes of these analogues. SCP-2L co-crystallised with a molecule of Triton-X-100 present in a linear apolar tunnel that traverses the entire protein, while aaSCP2, which crystallised with a molecule of palmitate, shows an apolar cavity oriented perpendicular to the cavity found in SCP-2L.[24] Extensive binding studies of mSCP2 using NMR have shown that especially the linear cavity as depicted for SCP-2L is involved in binding aliphatic substrates.[27] The different binding modi and small structural differences between SCP2 analogues result in small differences in substrate binding affinity and dynamics. Nevertheless research with various SCP2 analogues has shown stronger binding affinity for acyl-CoA derivatives compared to long chain fatty acids typically by an order of 50.[15-16, 28]. The binding affinity was also found to be dependent on the chain length and the branching of the substrate, with carbon chains >24 and high linearity being favoured.[28b, 29] The affinity of these stretched hydrophobic pockets for linear apolar substrates provides a good basis for the design of our artificial metalloenzyme catalyst for reactions involving linear aliphatic substrates. In this chapter the design, DNA-synthesis, production and purification of several unique cysteine containing protein templates for the creation of our artificial metallo enzymes based on SCP2 are described. In addition some structural and stability investigations and some initial substrate binding studies are reported.

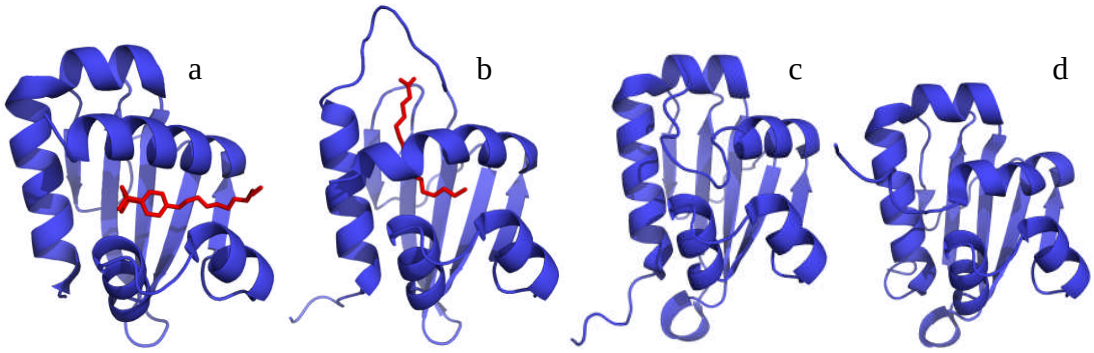


Figure 3. Crystal structures of a) SCP-2L in complex with a Triton X-100 molecule (red) (pdb = 1IKT)[19] b) aaSCP2 in complex with a molecule of palmitic acid (red) (pdb = 1PZ4)[24] c) osSCP2 (pdb = 1C44)[20] d) mSCP2 (in complex with pex5p which has been excluded) (pdb = 2C0L).[30]

2.3 Results and discussion

2.3.1 Design and plasmid synthesis of unique cysteine protein templates

2.3.1.1 mSCP2

Human mSCP2 is the most studied of the SCP2 proteins. It has been crystallized as a complex with the principal import receptor Pex5p[30] and also structurally characterized by NMR.[31] mSCP2 is the product of a post-translational modification of preSCP2, where part of the N-terminus is cleaved (mSCP2 = mature SCP2).[14c] To create artificial metalloenzymes suitable for selective substrate binding mSCP2 has to be altered to introduce a unique cysteine at a position in close proximity to the hydrophobic pocket. The exact binding site of mSCP2 has been studied using crystallography[30] and NMR structural analyses from which an inner cavity was identified (Figure 4).[27, 31] This cavity was found to be located between the C-terminal α-helices, very similar to other SCP2 analogues. NMR studies also indicated that the β-sheet backbone remains rigid during the binding process while the α-helices undergoes conformational changes at fast (picosecond to nanosecond) and slow (microsecond to millisecond) timescales. All these data indicate that the α-helical structure is involved in the access of the substrate into the hydrophobic pocket.

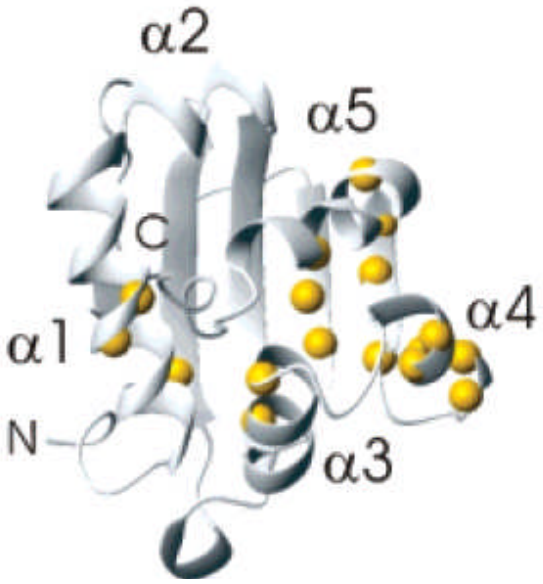


Figure 4. Visualisation of binding site of mSCP for hydrophobic part of long chain fatty acid CoA's as determined by ^{15}N-NMR perturbations of amino acids indicated with yellow dots (Reprinted with permission from numbered reference[27] Copyright 2007 American Chemical Society).

Wild-type mSCP2 contains one native cysteine in position 91, located far away from the known hydrophobic cavity as was evident from the model shown in Figure 5a. To create a suitable artificial metalloenzyme template the native cysteine had to be removed and a new cysteine introduced. A structurally similar serine residue was chosen as replacement for cysteine 91. Positions K104 and M120 were considered for introduction of the unique cysteines. These position are in the proximity of the apolar cavity as was observed in SCP2 analogues and NMR-binding studies of this particular protein.[27] To assess the suitability of these positions for the introduction of our organometallic catalyst systems docking simulations were performed. The program GOLD[32] was used for the docking studies, using the reported NMR-structure as template[31] and applying protein templates generated by the SWISS-MODEL online modelling service.[33] GOLD creates models based on an algorithm describing hydrogen bonding, electrostatic and steric interactions between the protein cavity and the ligand, keeping the ligand fully and the protein partly flexible.[32d] The program then creates a range of structures by optimisation and ranks them based on most favoured ligand protein interactions in the final structure. As covalent bound ligands maleimide-triazacyclononane iron complexes were used for this set of docking experiments as these were a potential catalytic cofactor for alkane oxidation. The highest ranking structures obtained from these docking studies are shown in Figure 5b and c. These structures show that for both these mutants the cofactor is in close proximity of the binding site between the α-helixes.

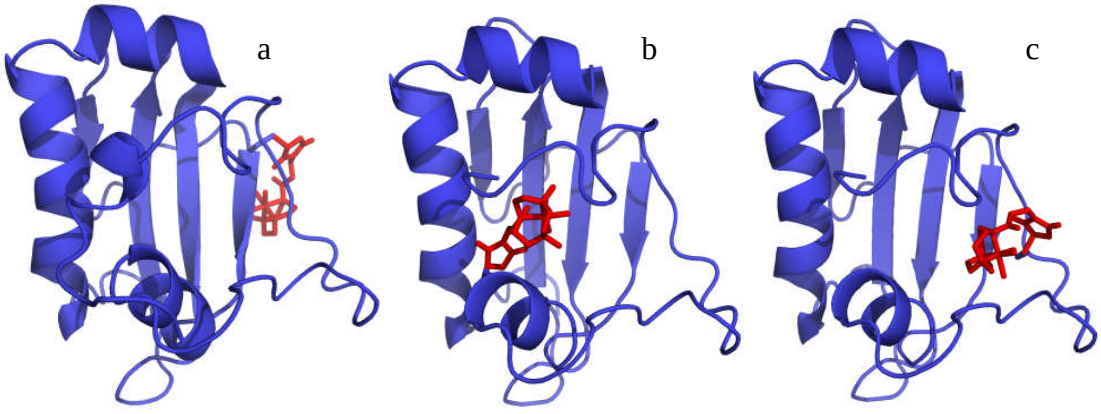


Figure 5. Highest ranking models from GOLD docking studies of a maleimide-triazacyclononane iron complex (red) covalently bound to a) mSCP2 (cysteine at position 91) b) mSCP2 C91S M104C c) mSCP2 C91S K120C.

For initial studies mSCP2 C91S M104C was chosen as unique cysteine bearing target mutant since the models show that the ligand is positioned in closer proximity to the central cavity. The DNA for mSCP2 was acquired as part of the pET24d::dΔhΔhSCP2 plasmid[27] (the pET24d used was modified and called pETM30).[34] This particular pET24d plasmid attaches a histidine and GST tag to the protein expressed in the multiple cloning site. The histidine tag allows for easy nickel affinity chromatography purification. The GST-tag is a long aminoacid sequence which has a high affinity for glutathione, which also makes the whole protein more soluble and so improves the yield from the lysis step. Both the GST and the histidine tag can be removed by Tobacco Etch Virus (TEV) protease.[35]

The removal of the native cysteine and the introduction of the unique cysteine at position 104 could in principle be achieved using two rounds of regular site directed mutagenesis. However nowadays multi-site-directed mutagenesis kits are available to do multiple mutagenesis steps at once. A QuickChange® Multi Site-Directed Mutagenesis Kit was used to introduce the M104C and C91S mutations simultaneously. The isolated plasmids from five different bacterial colonies obtained after polymerase chain reaction (PCR) were sent for sequencing. Three contained only the C91S mutation and only one was found to contain both mutations. The desired pET24d::dΔhΔhSCP2 C91S M104C plasmid was successfully transformed to production strain *E. Coli* BL21 (Rosetta).

2.3.1.2 SCP-2L mutants

SCP-2L originates, like mSCP2, also from *Homo sapien*s and is one of three domains of Human Peroxisomal Multifunctional Enzyme Type 2 (MFE-2) a protein cluster involved in the β-oxidation of long chain fatty acids.[36] SCP-2L was crystallized with Triton X-100 revealing an apolar tunnel (Figure 6).[19] This tunnel has a length of 18 Å and a diameter of 9 Å at the smaller exit of the tunnel (where the “tail” of the Triton X-100 molecule is located). The diameter is more than 10 Å at the wider entrance of the tunnel. The location of the Triton X-100 is presumed to be similar to the location of the binding site for long chain fatty acids as was found in NMR binding studies with the analogue mSCP2.[27] Overall this crystal structure provides good insight into where our transition metal centre has to be introduced to be in close proximity to the binding site.

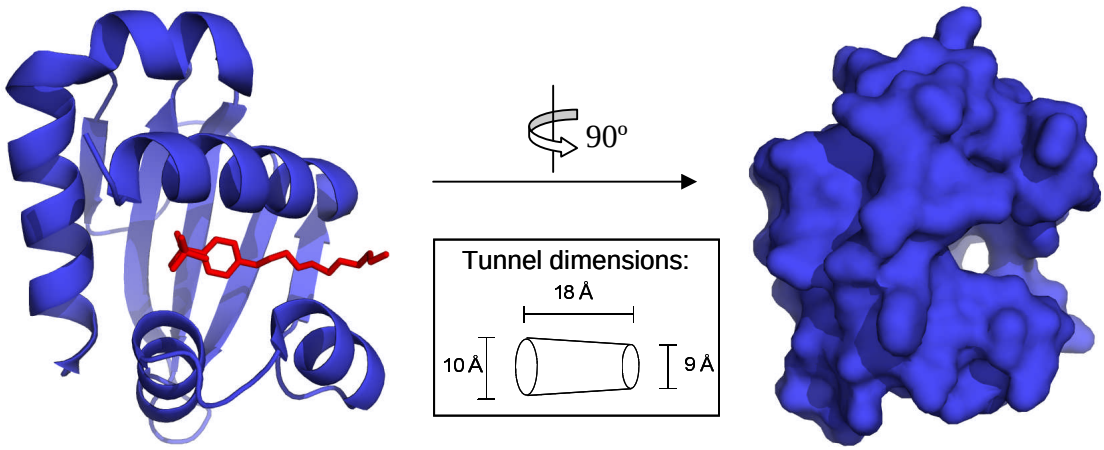


Figure 6. Cartoon model including Triton X-100 (left) and surface model (right) of SCP-2L showing the hydrophobic tunnel.

As SCP-2L does not contain any native cysteines one amino acid mutation suffices. The locations for the introduction of the cysteine were chosen at various positions around the tunnel. To test if these locations indeed bring the metal binding cofactors in close proximity to the tunnel docking studies using GOLD were performed.[32] For this purpose several models of mutants were created using the SWISS-MODEL online service based on the reported crystal structure.[33] These models were then used for docking with maleimide triazacyclononane iron complexes and several hydrazone-phosphane ligands of which a selection is shown in Figure 7 (see appendix A4 for more models). SCP-2L mutants: E14C, W36C, V82C, V83C and A100C were selected as target unique cysteine templates based on these models. The amino acid residues involved in E14C, V82C and V83C are located at the wide entrance of the tunnel, E14C on the rigid β-sheet backbone and V82C and V83C on the flexible α-helix structure. W36C is located in the middle of the tunnel and A100C at the smaller exit of the tunnel, both on the β-sheet backbone.

Initially both V82C and V83C mutants were selected because they are located at the wider opening of the apolar tunnel, which might prevent low coupling efficiency due to steric hindrance. The DNA of SCP-2L was obtained as part of the pET3d::dΔhΔSCP-2L plasmid.[19] The mutants were created by site directed mutagenesis on the original pET3d::dΔhΔSCP-2L plasmid to successfully create the pET3d::dΔhΔSCP-2L V82C and pET3d::dΔhΔSCP-2L V83C plasmids. However expression of the protein from this plasmid provides untagged protein, which complicates the purification. Because of this it was decided to clone the inserts into

the pEHISTEV vector to obtain a protein containing a histidine-tag and a TEV protease cleavage site.[10]

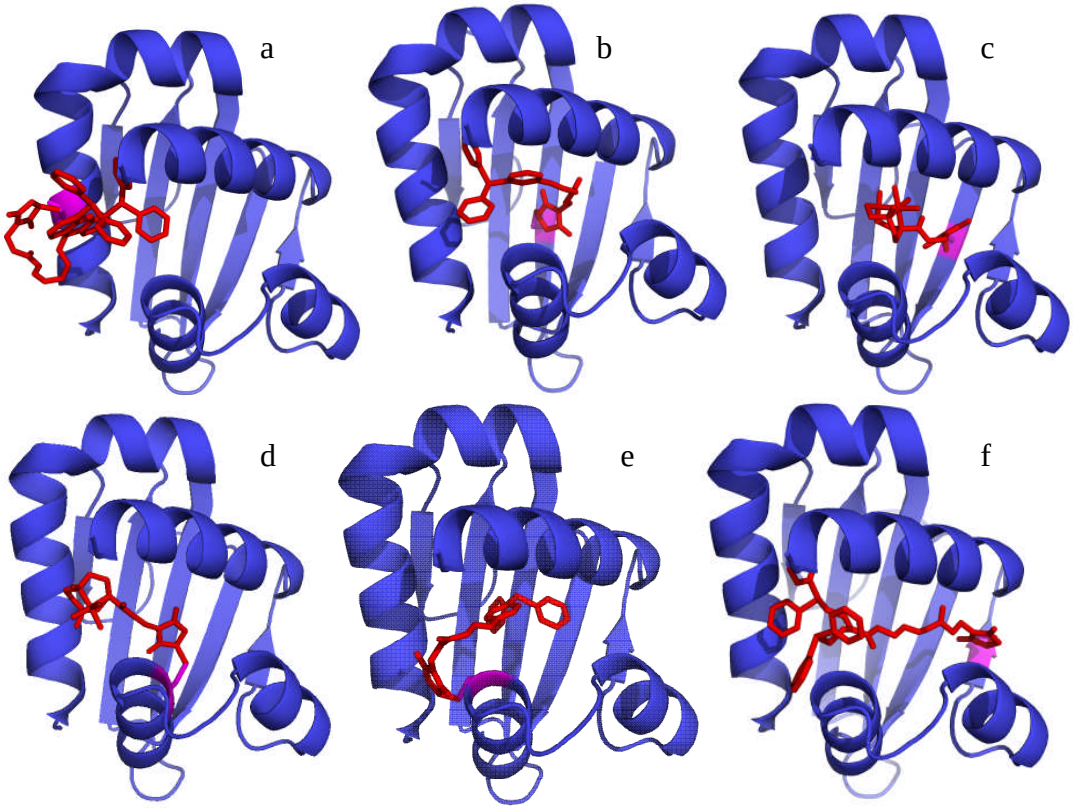
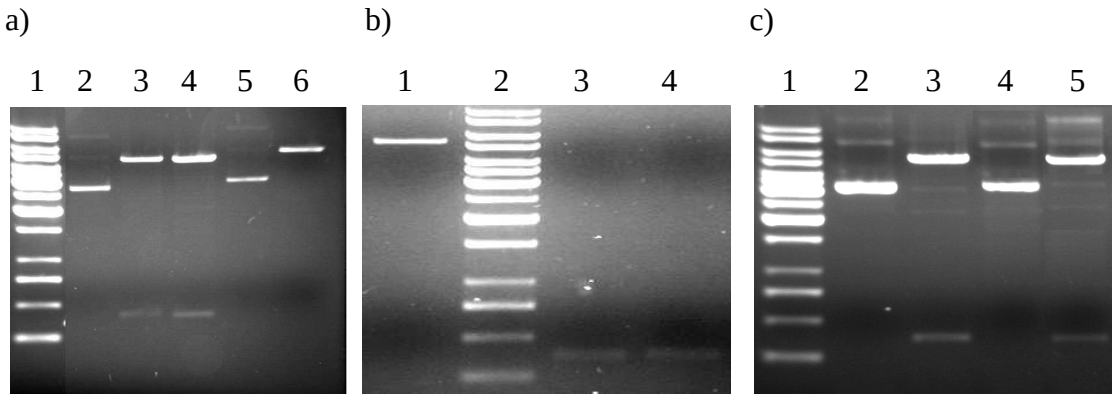


Figure 7. Selected highest ranked models from GOLD docking studies using various covalent bound ligands (red) to cysteine (purple) a) SCP-2L E14C b) SCP-2L W36C c) SCP-2L I72C d) SCP-2L V82C e) SCP-2L V83C f) SCP-2L A100C.

Both dΔhΔSCP-2L and dΔhΔSCP-2L V83C were successfully inserted into pEHISTEV using BamHI and NcoI restriction enzymes to cut the insert from their parent plasmids and sequential ligation with pEHISTEV treated with the same restriction enzymes to obtain pEHISTEV::dΔhΔSCP-2L and pEHISTEV::dΔhΔSCP-2L V83C (Figure 8). For unclear reasons the same method proved unsuccessful for pET3d::dΔhΔSCP-2L V82C. Site directed mutagenesis was performed on pEHISTEV::dΔhΔSCP-2L to obtain plasmids coding for the E14C, W36C and A100C mutants, successfully creating pEHISTEV::dΔhΔSCP-2L E14C and pEHISTEV::dΔhΔSCP-2L A100C. Unfortunately no successful mutagenesis was achieved for the W36C mutation.



a)
1. ladder
2. pET3d::dΔhΔSCP-2L untreated
3. pET3d::dΔhΔSCP-2L cleaved with BamHI and NcoI
4. pET3d::dΔhΔSCP-2L V83C cleaved with BamHI and NcoI
5. pEHISTEV untreated
6. pEHISTEV cleaved with

b)
1. isolated BamHI and NcoI cleaved pEHISTEV
2. ladder
3. isolated dΔhΔSCP-2L
4. isolated dΔhΔSCP-2L V83C

c)
1. ladder
2. pEHISTEV::dΔhΔSCP-2L untreated
3. pEHISTEV::dΔhΔSCP-2L cleaved with BamHI and NcoI
4. pEHISTEV::dΔhΔSCP-2L V83C untreated
5. pEHISTEV::dΔhΔSCP-2L V83C cleaved with BamHI and NcoI

Figure 8. Agarose gel analysis of subcloning dΔhΔSCP -2L and dΔhΔSCP -2L V83C into pEHISTEV, a) pET3d::dΔhΔSCP-2L, pET3d::dΔhΔSCP-2L V83C and pEHISTEV treated with BamHI and NcoI restriction enzymes b) Isolated SCP-2L and SCP-2L V83C inserts and cleaved pEHISTEV c) Control for plasmid isolated after ligation, transformation and isolation treated with BamH and NcoI.

2.3.2 Unique cysteine protein production and purification

2.3.2.1 mSCP2 and mSCP2 C91S M104C

Using the created pET24d plasmids, the mSCP2 protein is produced with a histidine (His) and glutathione S-transferase (GST)-tag attached to the N-terminus. The His-tag consists of 6 histidines which can be used for convenient purification of the protein by nickel affinity chromatography. The GST-tag is a large (~30kDa) protein, which has a high affinity for glutathione, which also makes the whole protein more soluble and so improves the yield from the lysis step. This gives also the opportunity for a purification step which involves glutathione affinity chromatography. Both tags can be cleaved off using His-tag containing TEV-protease, after which final purification can be achieved by another run of nickel affinity chromatography. Transformation of the bacterial expression strain *E. coli* BL21(Rosseta) with the pET24d::dΔhΔhSCP2 and pET24d::dΔhΔhSCP2 C91S M104C plasmids was successful. Both proteins were

expressed after addition of the expression initiator (isopropyl-1-thio-β-D-galactopyranoside) IPTG. Purification of the wild type mSCP2protein was successful following a procedure reported by Sattler *et al.*[37] This procedure included nickel affinity chromatography followed by cleavage of the His/GST-tag by TEV and another run of nickel affinity chromatography. The overall yield was only around 5 mg/L. Further optimisation of the production and purification was desired to improve the yield for application of the protein as artificial metallo enzyme catalyst. The purification of the mSCP2 C91S M104C mutant proved however more troublesome. Although expression was similar to the wild type protein, this mutant proved to be unstable especially during the TEV protease cleavage step. This step is usually performed at room temperature to ensure full conversion to the cleaved protein. However, this resulted in precipitation of the desired protein after cleavage as observed on SDS-PAGE gel, where only a very weak band for the cleaved protein could be observed (Figure 9). The precipitation could not be avoided by using lower temperature.

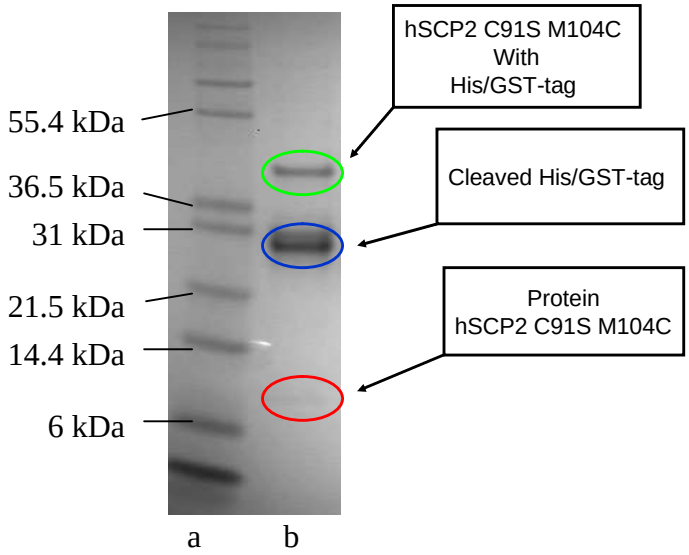


Figure 9. SDS-PAGE analysis of TEV cleavage of mSCP2 C91S M104C, a) protein ladder b) mSCP2 C91S M104C after cutting with TEV.

There have been previous reports concerning mSCP2 mutants[38] and it has been reported that introducing the C91S mutation leads to decreased thermal stability of the protein.[39] Therefore, screening for other amino acids to replace the cysteine at position 91 to help retain stability and substrate binding capacity is required. In

conclusion this particular SCP2 was found to be unsuitable for our purposes of creating stable artificial transition metal catalysts.

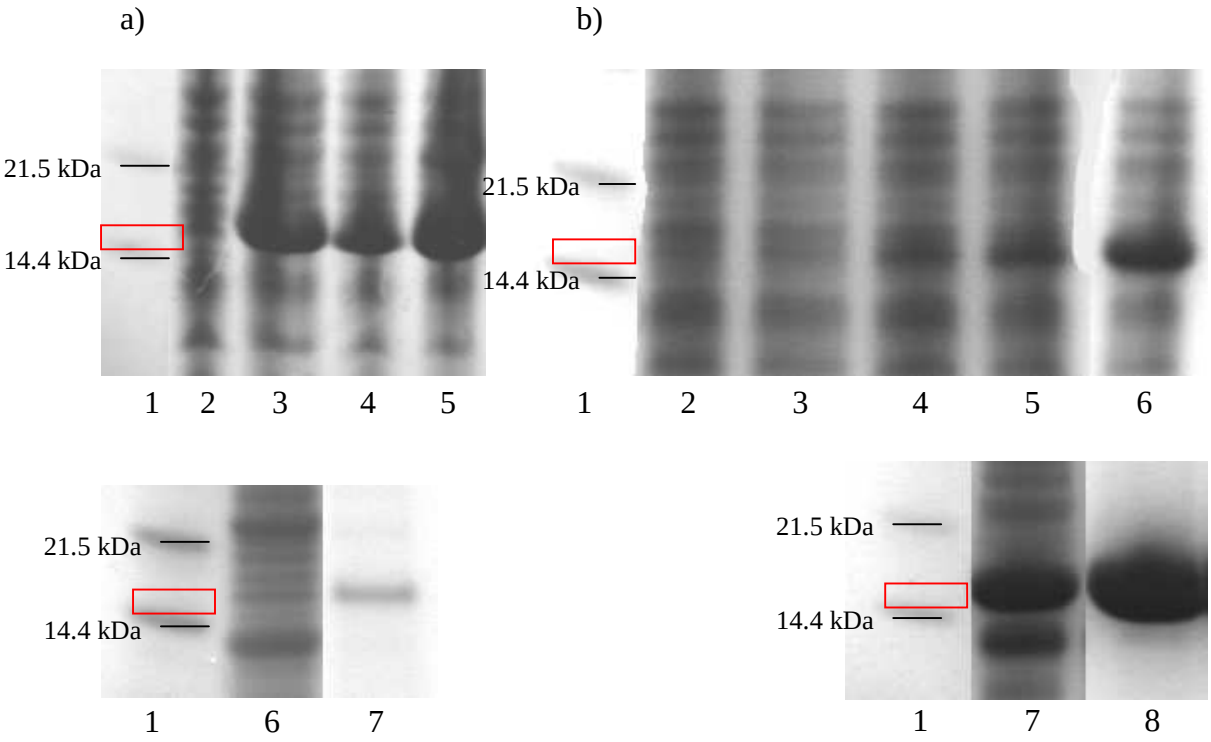
2.3.2.2 SCP-2L unique cysteine mutants

The pET3d:dΔhΔSCP-2L V83C plasmid was transferred to BL21 (DE3) pLysS competent cells. Expression was successful using IPTG as initiator and purification could be achieved using the reported procedure for SCP-2L with an average yield of 20 mg/L.[19] However, because this purification method proved time consuming the plasmids pEHISTEV::dΔhΔSCP-2L V83C and pEHISTEV::dΔhΔSCP-2L were created, which allows for the production of His-tagged protein, which can be purified by nickel affinity chromatography. In this construct the histidine-tag could be cleaved off by the highly specific Tobacco Etch Virus (TEV) protease. This protease cleaves at a specific E-N-L-Y-F-Q-G aminoacid sequence, cleaving between the glutamine (Q) and the glycine (G). This glycine is left on the protein together with an alanine (A) which is used as linker on this particular vector, resulting in a protein with a two aminoacid extension compared to the original SCP-2L.

Optimisation of the expression and purification procedure was done for SCP-2L V83C, and the same conditions were later applied for other mutants. Bacterial expression strain *E. coli* BL21(Rosetta) was successfully transformed with pEHISTEV::dΔhΔSCP-2L V83C. Good over-expression of the His-SCP-2L V83C protein was observed on SDS-PAGE, however after lysis only small amounts of soluble protein were obtained. The problem was successfully circumvented by performing the expression over 16 hours at 16ºC instead of 37ºC, which resulted in slightly lower expression, but significantly more soluble protein after lysis (Figure 10). The conditions for the lysis were also optimised for several parameters like buffer, pH and amount of lysozyme. Freezing the cells to -80ºC before lysis and sonication after lysis also proved very effective for increasing the yield of soluble protein.

The first run of nickel affinity chromatograph already yielded the desired protein with a purity of more than 99% according to SDS-PAGE analysis. However, the efficient removal of the histidine tag by TEV protease proved problematic. The application of standard conditions[40] for this step resulted in large amounts of precipitated protein leaving only small amounts of cleaved SCP-2L V83C. The cleavage was optimised

for buffer, pH and temperature but the most important factor appeared to be the concentration of the protein and TEV (Figure 11a). High concentrations of Hist-SCP-2L V83C resulted in large amounts of precipitation. Lowering the protein concentration to below 1.5 mg/ml and performing the cleavage at room temperature to ensure full cleavage proved to be the optimal conditions for removal of the histidine tag from Hist-SCP-2L using TEV protease (Figure 11b).



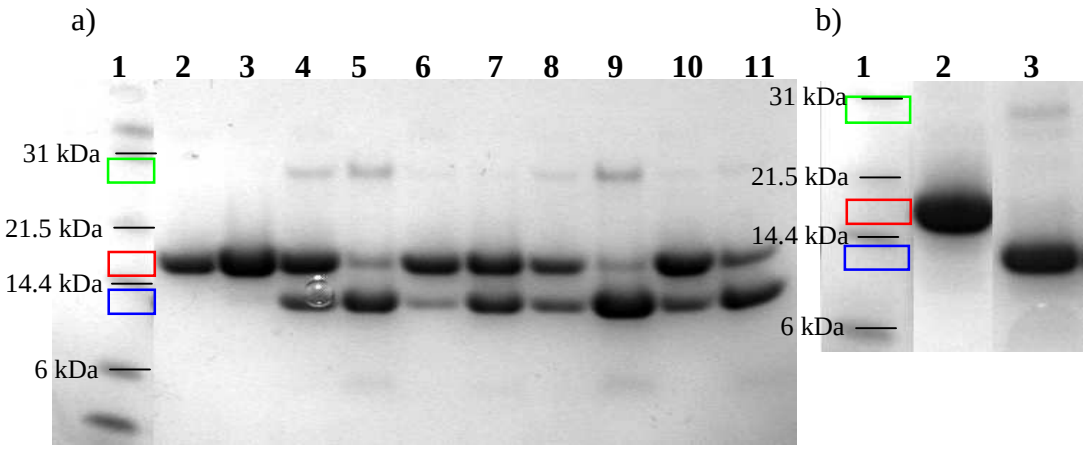
Expression at 37ºC
1. Protein ladder
2. Expression t = 0 h (IPTG addition)
3. Expression t = 1 h
4. Expression t = 3 h
5. Expression t = 16 h
6. Soluble fractions after lysis
7. After nickel affinity chromatography

Expression at 16ºC
1. Protein ladder
2. Expression t = 0 h (IPTG addition)
3. Expression t = 1 h
4. Expression t = 2 h
5. Expression t = 3 h
6. Expression t = 16 h
7. Soluble fractions after lysis
8. After nickel affinity chromatography

Figure 10. Comparison of expression and amount of soluble Hist-SCP-2L V83C at a) 37ºC and b) 16ºC with expected location of protein with molecular weight of 16344.7 Da indicated in red on the ladder.

After a second run of nickel affinity chromatography SCP-2L V83C could be obtained with very high purity (no other protein bands could be detected upon SDS-

PAGE analysis). The yields proved to be high reaching an average of more than 60 mg per litre of bacterial culture.



1. Protein marker
2. 3 mg/ml, Blanc Prec.
3. 3 mg/ml, Blanc Sup.
4. 3 mg/ml, 0.014eq TEV Prec.
5. 3 mg/ml, 0.014eq TEV Sup.
6. 3 mg/ml, 0.003eq TEV Prec.
7. 3 mg/ml, 0.003eq TEV Sup.
8. 1.5 mg/ml, 0.014eq TEV Prec.
9. 1.5 mg/ml, 0.014eq TEV Sup.
10. 1.5 mg/ml, 0.003eq TEV Prec.
11. 1.5 mg/ml, 0.003eq TEV Sup.

1. Protein marker
2. Before TEV cleavage
3. After TEV cleavage

Figure 11. SDS PAGE analysis of TEV protease cleavage experiments of Hist-SCP-2L V83C. Expected bands, Hist-SCP-2L V83C (red) SCP-2L V83C (blue), TEV protease (green) as depicted on the ladder. Conditions; a) 50 mM Tris 300 mM NaCl 20 mM imidazole pH 8 buffer, 0.5 mM EDTA (from 50 mM stock in 50 mM kPi pH 8 buffer), 1 mM DTT (from 100 mM stock in H_2O), 3 mg (180 nmol)/ml or 1.5 mg (90 nmol)/ml Hist-SCP-2L V83C (from 3 mg/ml stock in 50 mM Tris, 150 mM NaCl 10 mM imidazole pH 8 +20% glycerol), 0.07 mg (2.5 nmol)/ml or 0.017 mg (0.06 nmol)/ml TEV-protease (from 1.7 mg/ml stock in 25 mM Tris 150 mM NaCl 50% glycerol) at 4ºC overnight. Spun down next morning 16100 *g* for 30 min and separate precipitate (Prec.) from supernatant (Sup.). b) Diluted to below 1.5 mg/ml 0.014 eq TEV at room temperature overnight).

The wild type SCP-2L, SCP-2L A100C and SCP-2L E14C were successfully purified using the same procedure as developed for SCP-2L V83C. The yield for SCP-2L was the same as for SCP-2L V83C. The yield for SCP-2L A100C was lower with an

average of 30 mg per litre bacterial culture. SCP-2L E14C was found to be unstable forming large amounts of precipitate during the purification procedure resulting in a yield of only 5 mg per litre bacterial culture. This mutant could not be stored in 20 mM MES, 50 mM NaCl pH 6 at 4ºC like the other proteins as this resulted in slow precipitation of the protein over time.

2.3.2.3 ^{15}N-SCP-2L

In general, solution structures of proteins are determined by ^{13}C-NMR and ^{15}N-NMR of isotopically labelled proteins. Although a solution structure of SCP-2L would be very valuable to acquire, obtaining such NMR-structures is a time consuming and costly process, and is not within the scope of this project. The creation of ^{13}C labelled protein is very expensive because of the ^{13}C-labelled glucose that has to be used. The production of ^{15}N-labelled protein is done by growing cells on medium containing ^{15}N-labelled ammonia as nitrogen source, which is affordable compared to ^{13}C glucose. Several structural features of the protein can be determined using ^{15}N-labelled protein, and it can be used for substrate binding characterization. Also for EPR-studies of artificial metalloenzymes containing EPR active transition metals it would be interesting to be able to use ^{15}N-labelled protein as this would help to determine the environment of EPR-active metal centres by comparison with the unlabeled protein.

Production of isotopically labelled protein is performed using a defined medium to make sure the bacteria solely use the nitrogen from the isotopically labelled source for protein synthesis. This has the disadvantage that yields are usually ten fold lower than from growth in rich media such as the production broth (PB) used for large scale production of unlabelled SCP-2L and mutants. By growth at 16ºC overnight good expression of the ^{15}N-labelled SCP-2L could be achieved. The protein was purified using the same protocol used for unlabelled SCP-2L giving a good overall yield of the labelled protein of 6 mg/L culture compared to the 60 mg/L usually obtained for unlabelled SCP-2L. Peak analysis of the mass-spectrum of the protein indicated more than 99% incorporation of ^{15}N (calculated 13533.1 Da; found 13533.9 Da).

A ^{1}H,^{15}N 2D HSQC NMR-spectrum of a 0.1 μM protein solution in H_2O containing 6% D_2O was recorded (Figure 12). The number of peaks found (~120) is about the correct amount of expected signals when overlap of several peaks is considered. Two

distinct signals could be assigned to the two tryptophan side chains present in SCP-2L. The signals for these residues 35 and 47 are found in the bottom left corner of the spectrum (red rectangle). As these two residues are located at the entrance of the hydrophobic tunnel in SCP-2L these may be potential indicator signals for the binding of substrates in titration experiments.

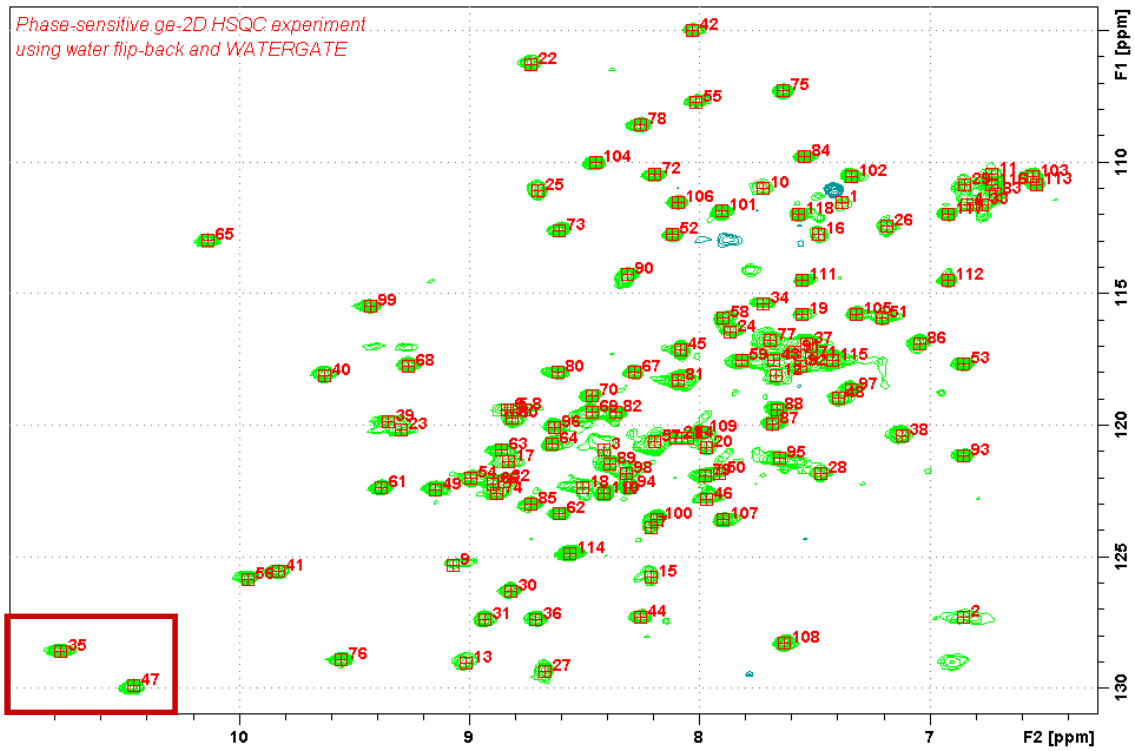


Figure 12. $^{1}H,^{15}N$ 2D HSQC NMR-spectrum of ^{15}N-SCP-2L.

2.3.3 Structure, stability and binding capacity of SCP-2L and mutants

2.3.3.1 Protein folding

When the models of the SCP-2L V83C and SCP-2L A100C proteins generated by the SWISS-MODEL service[33] are compared to the original crystal structure of SCP-2L the fold seems largely unaffected by the mutations (Figure 13). Experimentally the fold of the protein is often investigated using circular dichroism (CD).[41] In addition watergate ^{1}H-NMR, which is often used for obtaining 2D NMR spectra used for structural characterization, can be used as a relatively quick experiment to compare the fold of two proteins by assessment of the perturbations in the signals.

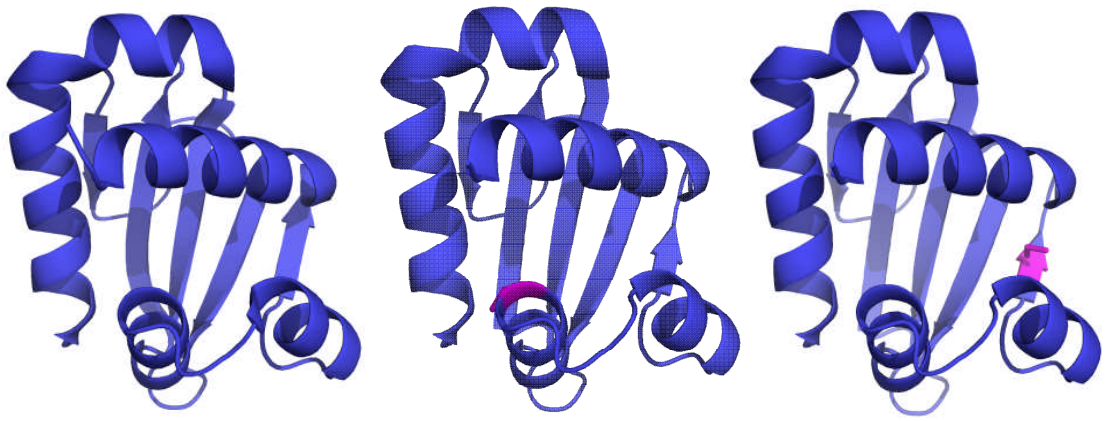


Figure 13. Models created by the SWISS-MODEL service[33] of a) SCP-2L b) SCP-2L V83C c) SCP-2L A100C (cysteines highlighted in purple).

The near UV-CD-spectra of SCP-2L, SCP-2L V83C and SCP-2L A100C proved to be identical (Figure 14). The chosen wavelength region shows the signals for aromatic amino acid residues in the protein.[41] The signal observed shows that the residues are in a chiral environment giving a finger print for this particular protein fold. As the signals for SCP-2L and the two mutants are almost identical it can be concluded that the structure of the protein is unaffected by the two mutations.

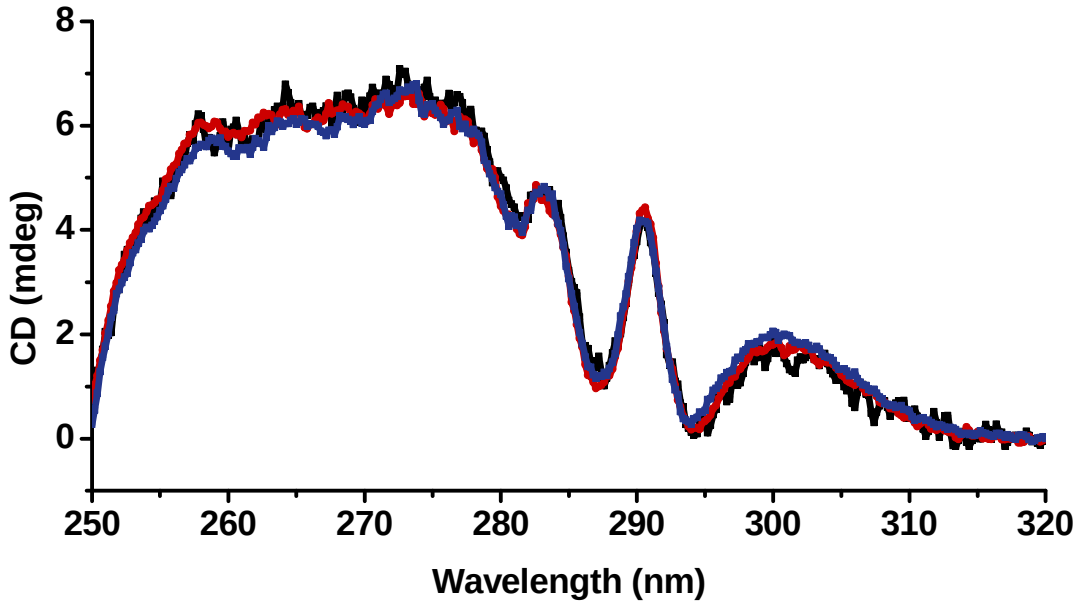


Figure 14. Near UV CD-spectra of SCP-2L (red), SCP-2L V83C (blue) and SCP-2L A100C (black).

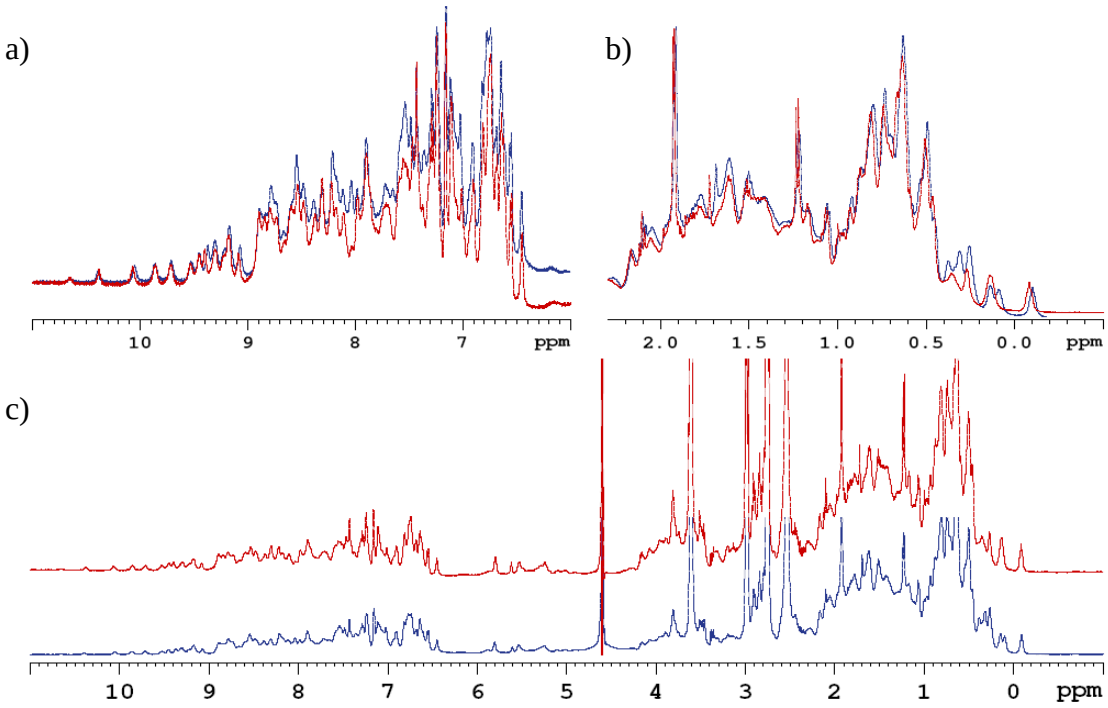


Figure 15. Watergate ^{1}H-NMR spectra of SCP-2L (blue) and SCP-2L V83C (red) in 1.5 mM in 20 mM MES pH 6 buffer a) overlay for the region between 11 and 6 ppm b) overlay for the region between 2.3 and –0.5 ppm c) full spectra, offset for clarity.

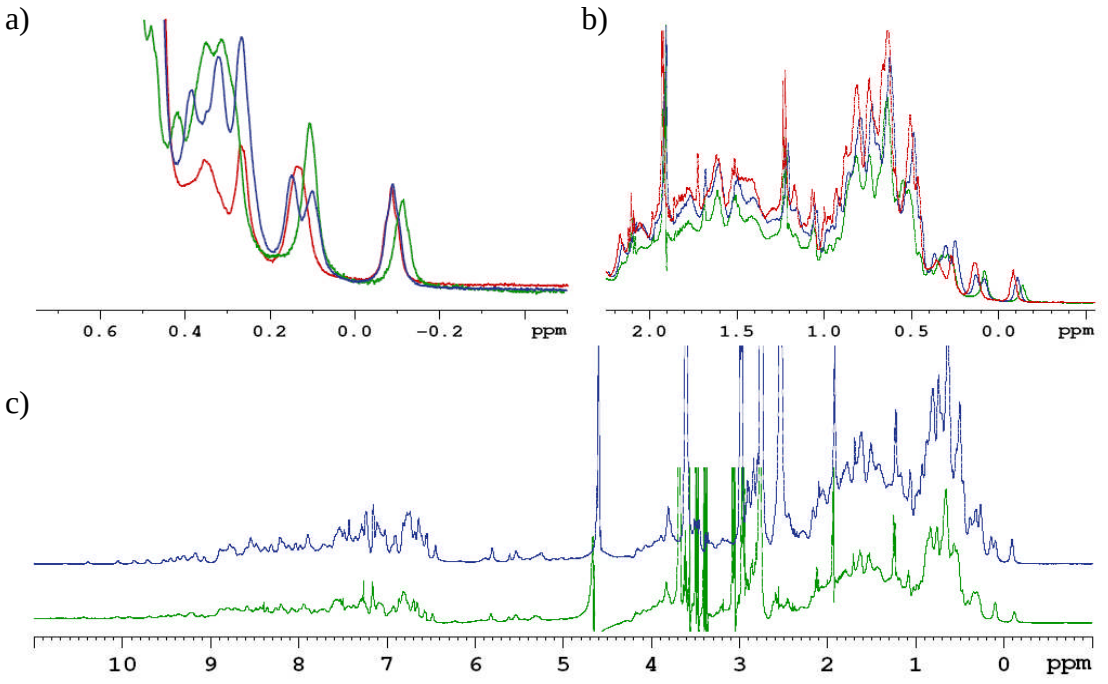


Figure 16. Watergate ^{1}H-NMR spectra of SCP-2L (blue), SCP-2L V83C (red) and SCP-2L A100C (green) in 1.5 mM in 20 mM MES pH 6 buffer a) overlay for the region between 0.75 and -0.5 ppm b) overlay for the region between 2.3 and –0.5 ppm c) full spectra, offset for clarity.

Also the watergate ^{1}H-NMR spectra of SCP-2L and SCP-2L V83C were recorded (Figure 15). In watergate (water suppression by gradient tailored excitation) ^{1}H-NMR experiments[42] the water signal is suppressed like in other solvent suppression ^{1}H-NMR experiments.[43] However, the difference is that in a watergate experiment no irradiation saturation is used, which means that protons that exchange with the solvent like N-H's using water as the solvent can be observed. From the spectra it can be seen that the proteins are both folded as the signals are spread out over the full region of the spectrum. Again the spectra of SCP-2L and SP-2L V83C were found to be very similar indicating a similar tertiary structure. Interestingly a significant difference is observed at round 0.5 ppm (Figure 15b and Figure 16a and b). This region corresponds to aliphatic residues in a hydrophobic surrounding. The V83C mutation is the replacement of a valine containing an isopropyl side chain with a cysteine so the loss of a signal in this particular region in SCP-2L V83C indicates that the signal for the valine at position 83 in SCP-2L is found here. Also the watergate ^{1}H-NMR of SCP-2L A100C mutant was recorded, however due to problems with the phasing the whole spectrum could be not overlapped as well as the WT protein and the V83C mutant. Overall the spectrum was found to be very similar to the WT protein as can be seen in Figure 16. Interestingly, differences can be observed again in the region around 0 ppm. The signal at -0.25 ppm has clearly shifted upfield and now another signal has disappeared at 0 ppm compared to V83C. As this mutant has an alanine replaced by a cysteine it is likely that this missing signal in the spectrum of SCP-2L V83C belongs to the alanine methyl in SCP-2L. The disappearance of signals in this region in both mutants confirms that amino acids inside the hydrophobic core have been replaced in both mutants.

2.3.3.2 Thermal stability

The protein thermal stability is usually expressed in T_m, the temperature at which protein denaturates in solution and can be determined by measuring circular dichroism at different temperatures.[44] The T_m of proteins can range from 20 to over 100ºC and is usually linked to the living environment of the natural host organism. The thermal stability of the proteins used in this study is very important to assess as for most catalytic processes elevated temperatures are desired.[45]

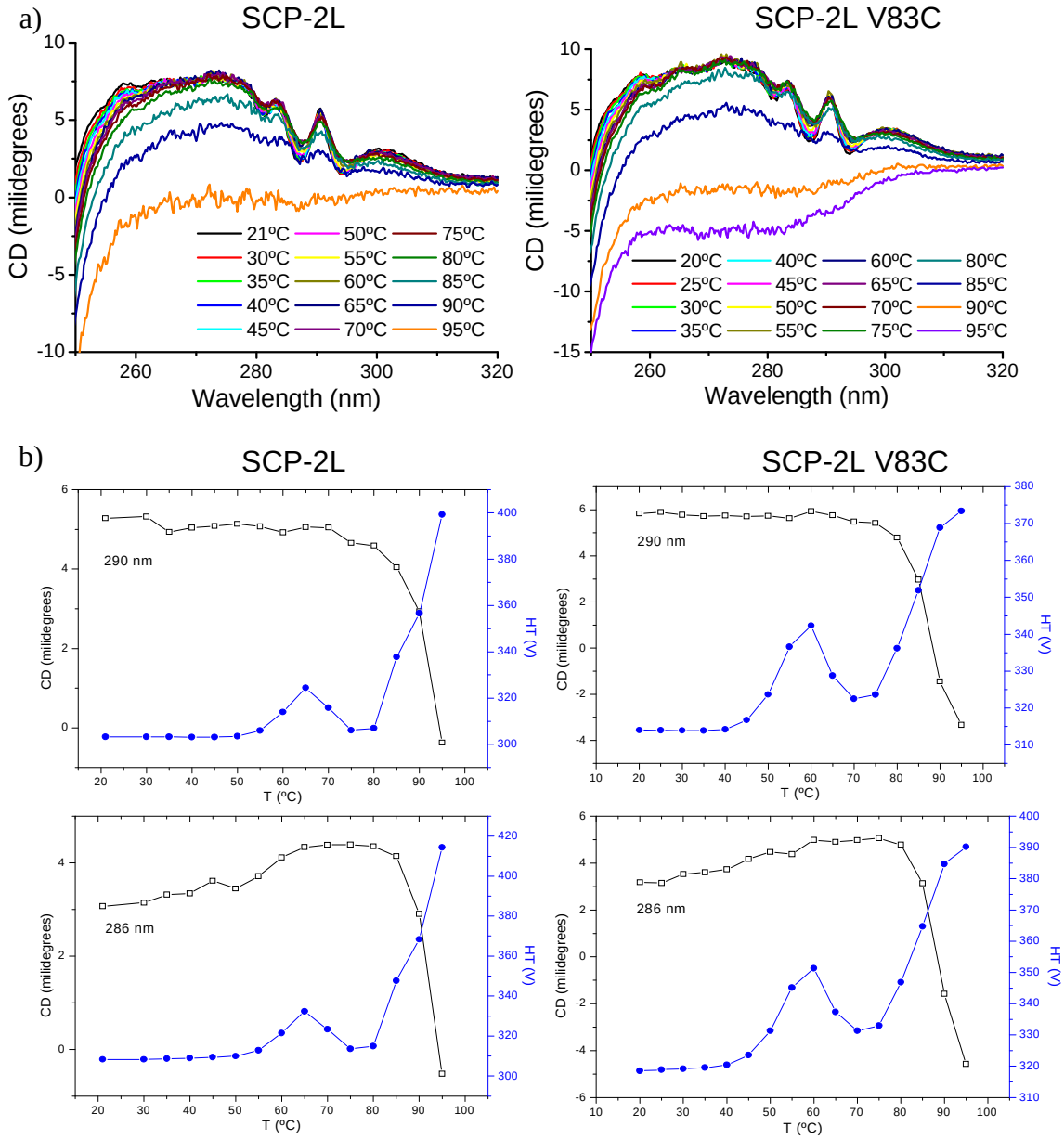


Figure 17. Variable temperature CD-analysis a) spectra of SCP-2L and SCP-2L V83C between 250 and 320 nm b) signals at 290 nm and 286 nm of SCP-2L and SCP-2L V83C, CD-signal (black) and stray light (blue).

Figure 17 shows the results for the variable temperature CD-spectroscopy analysis of wild type SCP-2L and SCP-2L V83C. For SCP-2L the CD signal appears to be constant up to 70 ºC, after which it gradually starts to decrease before rapidly decreasing at temperatures over 80 ºC (Figure 17 first column). At this stage a significant amount of precipitate had formed and when the sample was cooled down again only 1% of the original signal returned and most of the protein seemed to have precipitated. Therefore, no T_m could be determined as this requires reversible

unfolding. Based on these results the T_m can be estimated and must be over 70 ºC. Overall the protein fold appears to be stable at elevated temperatures but at high temperature above 70 ºC it starts to denature irreversibly. High melting temperatures (>70 ºC) have been reported for structurally similar mSCP2 consistent with the results shown above.[39] However in that study it was suggested that unfolding of the protein was reversible.

Similar results were obtained for SCP-2L V83C, with the difference that the signal started to decrease about 10ºC lower in temperature compared to the wild type protein (Figure 17 second column). Irreversible denaturation of the protein was also observed for this mutant. A closer look reveals that in both the spectra of SCP-2L and SCP-2L V83C the signals at the minima at 286 and 294 nm increase with raising the temperature before disappearing upon denaturation. If the temperature is plotted against the CD signal at 286 and 290 nm this effect can be seen more clearly (Figure 17b). The graph for SCP-2L obtained at 286 nm shows that around 55ºC the protein seems to go through a small conformational change. Although less clearly, the same effect can also be seen in the 286 nm graph for SCP-2L V83C, but at a slightly lower temperature (~45ºC). Interestingly at about the same temperature there is a peak in the amount of detected stray light. An increase in the amount of stray light is usually caused by scattering of the light beam by particle or precipitate formation inside the sample.[46] It seems likely that the small conformational change initiates the irreversible protein denaturation. Overall the protein appears remarkable stable to increased temperature, which is very useful for application as template for artificial metalloenzyme catalysts. The observation that the protein seems to precipitate after denaturation can also be used as an advantage as denatured protein can be easily removed by centrifugation.

2.3.3.3 Substrate binding capability of SCP-2L and SCP-2L V83C

Substrate binding to proteins can be measured using various techniques. A frequently used technique to monitor fatty acid binding is titration of the protein with fluorescent fatty acids.[15] Particularly a method using “Förster resonance energy transfer” (FRET) between tryptophan residues in the protein and a fluorescent probe proved useful.[28b] By titration of the proteins with a fluorescent probe K_d-values can be determined as schematically depicted in Figure 18 for the present system using 12-

pyrene dodecanoic acid (Pyr-C12) as probe. During this type of titration the native fluorescence of the tryptophan at 336 nm is quenched (red arrow) by energy transfer to the pyrene of the fatty acid which emits fluorescence at various wavelengths (blue arrows). This energy transfer is only possible if the donor and acceptor are in close proximity (~5 nm), making the amount of observed FRET a measure for binding. Also binding of non-fluorescent substrates can in principle be studied by competitive binding experiments.

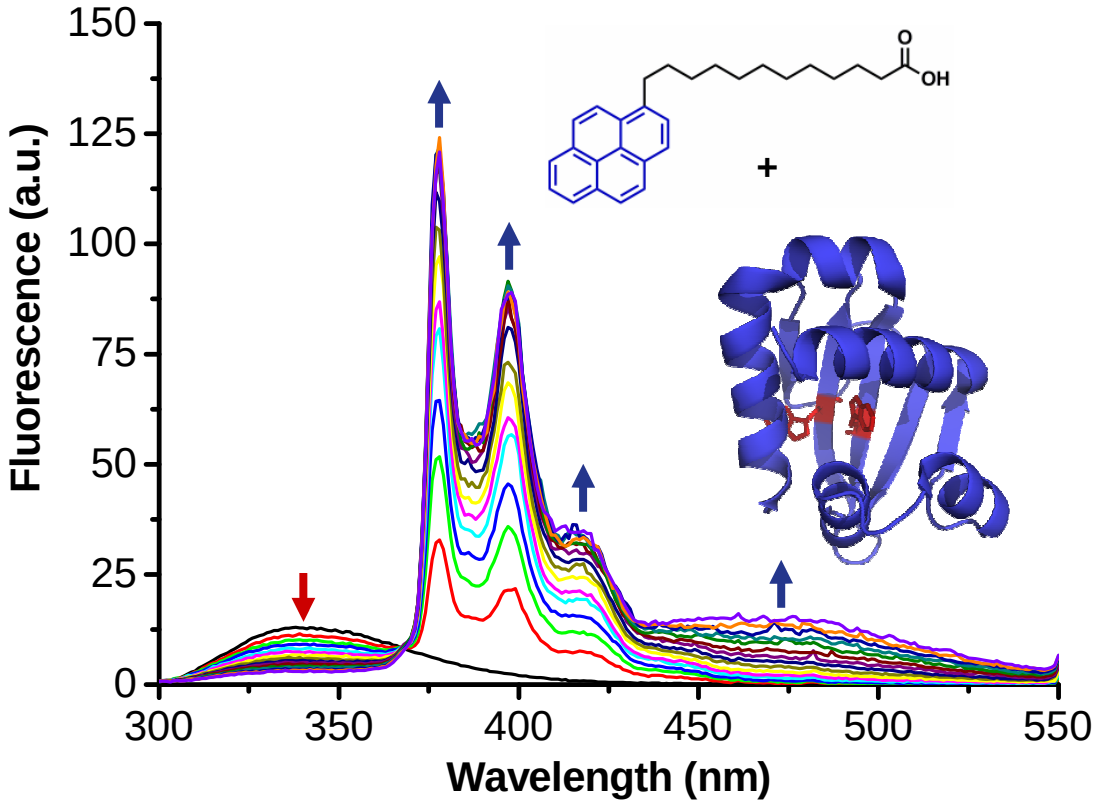


Figure 18. Example of FRET between SCP-2L tryptophan residues (shown in red) and fluorescent Pyr-C12 as observed by measuring the fluorescence upon titration of SCP-2L with Pyr-C12.

Titrations of SCP-2L, SCP-2L V83C were performed and the calculated K_d values are shown in Table 1 (entries 1 and 2). Although the error margins are significant the obtained K_d values show that the mutations do not affect the binding affinity of the protein for the aliphatic substrate used in this titration.

The found K_d values are in same order of magnitude as those of the same and similar fatty acids obtained for mSCP2 (Table 1, entries 3-5). Interestingly, the found K_d values are lower than those for fatty acid CoA binding to SCP-2L. Usually stronger binding is observed for CoA derivatives for this class of proteins. This could indicate

that SCP-2L actually does not have a higher preference for binding fatty acid CoA derivatives. This unusually high affinity for long chain fatty acids is supported by patient studies containing defects in the gene coding for MFE2 (in which SCP-2L is located). The patient typically show high levels of these fatty acids in plasma and fibroblasts are found, showing that the metabolic pathway for these compounds is deregulated.[18, 47] These results indicate that SCP-2L has a high affinity for linear aliphatic substrates which are of interest to this study.

Table 1. K_d values of (un)modified SCP-2L and reported values.

Entry	Protein	K_d (μM) (+/-σ)	Substrate[a]	Method[b]	Reference
1	SCP-2L	0.18 (+/-0.05)	Pyr-C12	FRET	This work
2	SCP-2L V83C	0.18 (+/-0.07)	Pyr-C12	FRET	This work
3	mSCP2	0.24	Pyr-C12	FRET	[28b]
4	mSCP2	0.18 (+/-0.01)	*cis*-Par	Direct	[15]
5	mSCP2	0.56 (+/-0.04)	*trans*-Par	Direct	[15]
6	mSCP2	4.57 (+/-0.75)[c]	*cis*-Par-CoA	Direct	[15]
7	mSCP2	2.76 (+/-0.1)[c]	*trans*-Par-CoA	Direct	[15]
8	SCP-2L	20	Pal-CoA	SPR	[19]
9	SCP-2L	40	*trans*-C16-CoA	SPR	[19]

a) Pyr-C12: pyrene dodecanoic acid, *cis*-Par: *cis*-parinaric acid, *trans*-Par: *trans*-parinaric acid, Pal-CoA: palmitoyl-CoA, *trans*-C16-CoA: *trans*-2-hexadecenoyl-CoA b) FRET: Förster resonance energy transfer, Direct: direct measurement of fluorescence of the probe, which has its emission wavelength altered upon binding, SPR: Surface plasmon resonance measurements c) K_d in nM instead of μM.

2.3.3.4 The free cysteine

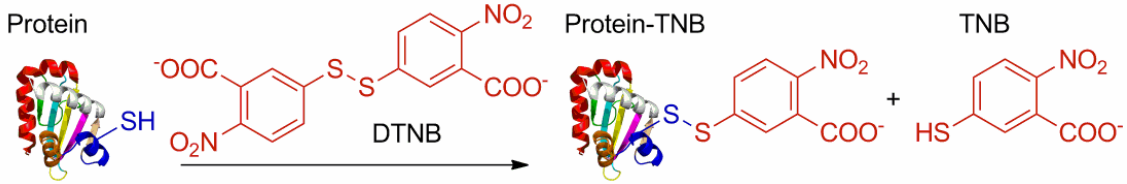
For the creation of artificial metallo enzymes by our methodology it is essential that the unique cysteine is available for selective modification. In order to confirm this, several test experiments were performed. Maleimides conjugate very selectively to the thiol of a cysteine at low pH (<7). Maleimide **M1**[48] was chosen as a test maleimide for modification of the cysteine in SCP-2L V83C and A100C (Scheme 1). After reaction of the proteins with maleimide only one set of peaks was observed by LCMS(ES^+) analysis corresponding to the expected modified proteins.

SCP-2L V83C 13376.6 Da
SCP-2L A100C 13404.6 Da

SCP-2L V83C-**M1** (calc 13619.9 Da)
Found: LCMS (ES^+) 13619.1 Da
SCP-2L A100C-**M1** (calc 13647.9 Da)
Found: LCMS (ES^+) 13646.9 Da

Scheme 1. Reaction between unique cysteine of SCP-2L V83C and A100C with maleimide **M1**.

Selective modification of the cysteine was confirmed by application of Ellman's reagent[49] to the modified and unmodified proteins. Ellman's reagent (5,5′-dithio-bis(2-nitrobenzoic acid) (DTNB) forms disulfide bridges with cysteine releasing an equivalent of the yellow free 5-thio-2-nitrobenzoic acid (TNB) (Scheme 2).[50] The yellow colour of the reaction mixture after 5 minutes clearly showed the formation of TNB upon reaction of DTNB with the unmodified proteins compared to no colour change for the modified proteins. This indicated that the unique cysteine is indeed modified by maleimide **M1**. Further analysis of the reaction mixtures by LCMS(ES^+) confirmed this.[51] While for the reaction of the unmodified SCP-2L mutants full conversion to the SCP-2L A100C-TNB and nearly quantitative conversion to SCP-2L V83C-TNB was found, no new products were observed for the reactions with SCP-2L V83C-**M1** and SCP-2L A100C-**M1**.



SCP-2L V83C 13376.6 Da
SCP-2L A100C 13404.6 Da

SCP-2L V83C-TNB (calc 13573.8 Da)
Found: LCMS (ES^+) 13573.5 Da
SCP-2L A100C-TNB (calc 13601.8 Da)
Found: LCMS (ES^+) 13601.0 Da

Scheme 2. Reaction between unique cysteine of SCP-2L V83C and A100C with Ellman's reagent (DTNB).

The reaction mixtures of the reactions between the SCP-2L mutants and maleimide **M1** were also subjected to a tryptic digest followed by detailed analysis by MALDI-

MS. Trypsin is a protease that digests the protein by cleavage at the carboxylic side of lysines or arginine amino acid residues in the peptide sequence except when the amino acid is modified or followed by a C-terminal proline.[52] By using the protein peptide sequence the results of the MALDI MS analysis after tryptic digest can be compared to the expected most abundant oligopeptides after cleavage (Figure 19 a and b). This way a search can be performed for oligopeptides which have been modified. By applying MS-MS the full amino acid sequence of the oligopeptide can often be obtained providing the exact location of the modification. Analysis of the digests of SCP-2L V83C-**M1** and SCP-2L A100C-**M1** by MALDI MS showed for both proteins modification of the expected oligopeptide only. For SCP-2L V83C-**M1** also a signal of +18 compared to the expected oligopeptide was found which is likely to correspond to maleimide ring opening by water, which is known to occur at basic conditions such as used for the trypsin digest (pH 8). MS-MS analysis of the peak corresponding to the expected oligopeptide sequence is shown in Figure 19c. The found sequence does confirm that the correct oligopeptide is modified. Also the precise location of the modification can be narrowed down to the short peptide CLGK. Because trypsin does not cleave near modified lysines (the only reactive amino acid except for cysteine in this short peptide),[53] this spectrum confirms the successful modification of the cysteine.

Maldi-MS analysis SCP-2L A100C-**M1** was challenging because the signals usually observed for the matrix are found in the same region as the signal for the expected short peptide containing the modified cysteine after trypsin digestion. A signal was found that corresponded to the expected peptide including the N-terminal lysine of the cleavage site (**KC(M1)R)**. Signals corresponding to lysine (K) and arginine (R) which are part of this small peptide could be found upon MS-MS analysis confirming the signal belongs to this short oligopeptide. Although lysines are known for their nucleophilic character modification of this lysine is highly unlikely as trypsin is known to be unable to cleave peptide chains next to modified lysines confirming the modification of the expected cysteine.[53]

All the above experiments confirm that the unique cysteine is present and available for modification using known bioconjugation techniques. This enables the use of these proteins as templates for artificial metalloenzymes by site selective covalent modification with transition metal ligands. The results of these experiments will be described in the following chapters.

a)

MEGGK/LQSTFVFEEIGR/R/LK/DIGPEVVK/K/VNAVFEWHITK/GGNIGAK/WTIDLK/SGSGK/VYQGPAK/**GAADTTIILSDEDFMEVC(M1)LGK**/LDPQK/AFFSGR/LK/AR/GNIMLSQK/LQMILK/DYAK/L

b)

MEGGK/LQSTFVFEEIGR/R/LK/DIGPEVVK/K/VNAVFEWHITK/GGNIGAK/WTIDLK/SGSGK/VYQGPAK/GAADTTIILSDEDFMEVVLGK/LDPQK/AFFSGR/LK/**C(M1)R**/GNIMLSQK/LQMILK/DYAK/L

c)

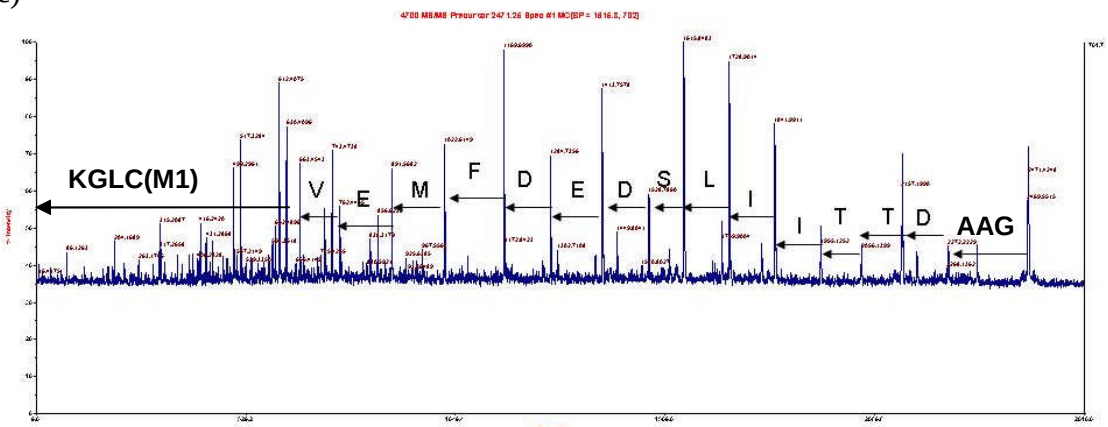


Figure 19. Amino acid sequence of a) SCP-2L V83C and b) SCP-2L A100C showing expected cleavage sites with "**/**", expected modified peptide in bold and modified cysteine in red. This figure only shows the expected most abundant cleave site but cleavage at the other sides of the lysines can also occasionally be observed c) MS-MS analysis of cysteine containing oligopeptide of SCP-2L V83C confirming the expected sequence.

2.4 Conclusions and future prospects

In this chapter the production and purification of unique cysteine containing protein templates for the development of artificial metalloenzymes is described. This was successful for several SCP-2L mutants which according to structural investigations and substrate binding studies maintain the same tertiary structure as the wild-type protein. Mutants of this protein were found suitable for use as artificial metalloenzymes as good thermal stability was found as well as full selective modification using a model maleimide. Further investigation into the structure should be possible using ^{15}N-labelled proteins as shown by the example of the expression and purification of ^{15}N labelled SCP-2L. Crystal structures and NMR-structures of modified proteins would be very useful in the further design of artificial metalloenzymes based on these templates. Some modelling has been performed but

the reliability of the obtained structures still has to be assessed. Although in principle any protein template can be used by this approach to create artificial metalloenzymes, this chapter also shows some limitations. Some of the desired protein templates could not be obtained in sufficient quantities and some were lacking long term stability. The removal of the unique cysteine in mSCP2 leading to an unstable protein and the unstable SCP-2L E14C protein are examples of this instability. Other limitations are related to problems with expression and purification systems used for protein templates.

2.5 Experimental

2.5.1 General equipment;

2.5.1.1 Methods

For processing of the mass-spectra the MaxEnt algorithm of Masslynx V4.0 (Micromass Ltd) was used. Plasmids were sequenced by the Sequencing Service, School of Life Sciences, University of Dundee, Dundee. Protein concentrations were determined using Bradford's reagent.[54] Curve fits to calculate K_d values were performed using OriginPro 8.

2.5.1.2 Equipment

Nickel affinity column protein purifications were done using 5 ml HisTrap FF columns (Amersham Bioscience UK Limited). LC-MS (ES) used for analysis of protein and protein reactions, Waters Alliance HT 2795 equipped with a Micromass LCT-TOF mass spectrometer, using positive electrospray ionisation and using a Waters MASSPREP® On-line Desalting 2.1x 10 mm cartridge and application of a gradient of 1% formic acid in H_2O to 1% formic acid in acetonitrile. Spin-filter concentrators with a molecular weight cut-off of 10kDa (Amicon® Millipore Ultra®, Ultracel 10K, Millipore®, Carrigtwohill, Co. Cork, Ireland) were used to concentrate and/or wash protein solutions. CD-spectra and variable temperature CD-spectra were recorded of 2 mg/ml protein solutions in 20 mM MES pH = 6 on a Jasco J-810 spectropolarimeter equipped with a CDF-426S temperature and stirring controlling unit connected to a Julabo awc100 cooler/heater. Watergate ^{1}H-NMR and ^{1}H,^{15}N 2D HSQC NMR spectra were recorded on a Bruker Avance 500 spectrometer. A Varian

Cary Eclipse was used for fluorescent binding assays equipped with a temperature controller and magnetic stirring unit.

2.5.1.3 Materials

His-TEV(S219V)-Arg[55] obtained from the E. Coli BL21(DE3)-pRK793 strain[35, 56] was used as protease for TEV cleavage of the histidine and GST-tags. The pET24d::dΔhΔhSCP2 plasmid[27] was kindly provided by M. Sattler from the European molecular Laboratory of Heidelberg, Germany (the pET24d used was a modified version, which is also known as pETM30).[34] The pET3d::dΔhΔSCP-2L plasmid,[19] kindly provided by T. Glumoff of the Biocenter Oulu and the Department of Biochemistry of the University of Oulu. All protein sequences used in this thesis can be found in appendix A1. Production Broth medium (PB); containing 20g/L tryptone, 10g/L yeast extract, 5 g/L dextrose, 5 g/L NaCl, 8.7 g/L K_2HPO_4, pH = 7.0 was used for protein expression.

2.5.2 Plasmid construction

2.5.2.1 Construction of pET24d::dΔhΔhSCP2 C91S M104C and pET24d::dΔhΔhSCP2 C91S plasmids

The mSCP2 C91S M104C mutant was constructed from pET24d::dΔhΔhSCP2 plasmid with a QuickChange® Multi Site-Directed Mutagenesis Kit (Stratagene. La Jolla, CA, USA) using the provided pET24d::dΔhΔhSCP2 plasmid as the template and the primers shown in Table 2 for PCR. The construct was verified by sequencing. One of the failed constructs was identified as pET24d::dΔhΔhSCP2 C91S coding for the mSCP2 C91S mutant. pEHISTEV was kindly donated by the group of Prof. Dr. J. H. Naismith (University of St. Andrews).

Table 2. Primers used for mutagenesis to create mSCP2 C91S M104C.

Primer	Mutation acquired	Sequence
Primer 1	C91S TGC/AGC	5'-CAGATAAGAAGGCTGAC**AGC**ACAATCACAATGGCT GACTC-3
Primer 1	M104C ATG/TGC	5'-CCTGGCTTTA**TGC**AC TGGTAAAATGAATCCTCAGTCGGCC-3'

2.5.2.2 Construction of the pET3d::dΔhΔSCP-2L V82C and pET3d::dΔhΔSCP-2L V83C plasmid

Mutations were introduced using the Quick-Change Site-Directed Mutagenesis Kit (Stratagene). pET3d::dΔhΔSCP-2L plasmid was used as templates for the PCR reactions to create the plasmids coding for SCP-2L and SCP-2L mutants V82C and V83C. The primers shown in Table 3 were used for site directed mutagenesis. The constructs were verified by sequencing.

Table 3. Primers used for mutagenesis to create SCP-2L mutants.

Primer	Mutation acquired	Sequence
Primer 1	V82C GTG/TGC	5' GATTTCATGGAG**TGC**GTCCTGGGCAAGCTTGACCC 3'
Primer 2		5' GGGTCAAGCTTGCCCAGGAC**GCA**CTCCATGAAATC 3'
Primer 1	V83C GTC/TGC	5' GATTTCATGGAGGTG**TGC**CTGGGCAAGCTTGAC 3'
Primer 2		5' GTCAAGCTTGCCCAG**GCA**CACCTCCATGAAATC 3'

2.5.2.3 Construction of the pEHISTEV::dΔhΔSCP-2L and pEHISTEV::dΔhΔSCP-2L V83C plasmids

To enhance the production yield of SCP-2L mutants the DNA coding for SCP-2L and mutants V82C and V83C from pET3d::dΔhΔSCP-2L plasmids were subcloned into pEHISTEV.[10]. The pEHISTEV plasmid was digested with BamHI and NcoI in BamHI buffer (Fermentas Life sciences) and isolated from a 0.8% agrose gel. The SCP-2L inserts were similarly isolated from digestion of the pET3d::dΔhΔSCP-2L plasmids. The digested pEHISTEV and corresponding insert were ligated using a Rapid DNA Ligation Kit (Roche Diagnostics GmbH Mannheim, Germany) to create the pEHISTEV::dΔhΔSCP-2L and pEHISTEV::dΔhΔSCP-2L V83C plasmid. The constructs were verified by sequencing. This process was unsuccessful for the SCP-2L V82C. These constructs afford the corresponding protein with an N-terminal hexahistidine tag attached which can be removed *in vitro* by TEV-protease.[35]

2.5.2.4 Construction of pEHISTEV:: dΔhΔSCP-2L A100C, pEHISTEV:: dΔhΔSCP-2L E14C and pEHISTEV:: dΔhΔSCP-2L W36C plasmids

Mutations were introduced applying a Quick-Change Site-Directed Mutagenesis Kit (Stratagene) using pEHISTEV::dΔhΔSCP-2L as template. Table 4 shows the primers

used for mutagenesis to create SCP-2L mutants: SCP-2L E14C, W36C and A100C. The constructs were verified by sequencing.

Table 4. Primers used for mutagenesis to create SCP-2L E14C, W36C and A100C mutants.

Primer	Mutation acquired	Sequence
Primer 1	E14C GAA/TGC	5' ACCTTTGTATTTGAG**TGC**ATAGGACGCCGCCTAAAGGATATTG 3'
Primer 2		5' CAATATCCTTTAGGCGGCGTCCTAT**GCA**CTCAAATACAAAGGT 3'
Primer 1	W36C TGG/TGC	5' AGGTAAATGCTGTATTTGAG**TGC**CATATAACCAAAGGC 3'
Primer 2		5' GCCTTTGGTTATATG**GCA**CTCAAATACAGCATTTACCT 3'
Primer 1	A100C GCC/TGC	5' AGTGGCAGGCTGAAG**TGC**AGAGGGAACATCATG 3'
Primer 2		5' CATGATGTTCCCTCT**GCA**CTTCAGCCTGCCACT 3'

2.5.3 Protein expression and purification

2.5.3.1 mSCP2

PB medium with 50 μg/ml kanamycin and 34 μg/ml chloramphenicol (in ethanol) was used for expression experiments. 10 ml of an overnight culture of E. coli BL21(Rosetta) cells harbouring plasmid pET24d::dΔhΔhSCP2 was used to inoculate 0.5 litre of PB medium using the same antibiotics concentration as during the expression experiments. The cells were allowed to grow under aerobic conditions in 3L baffled conical flasks at 37ºC to an OD_{600} of 0.6 after which the temperature was lowered to 21ºC. IPTG was added after one hour to a final concentration of 0.2 mM, which initiated the expression of the recombinant protein. The culture was than allowed to growth overnight at 21ºC. The cells were harvested by centrifugation (20 min, 3000 *g*, 4ºC). A portion of the pellet (3.5 g of wet weight) was suspended in 30-40 ml 30 mM Tris, 20 mM imidazole, 150 mM NaCl, 2 mM β-mercaptoethanol, pH 8 (Wash) buffer. To this suspension 1 mg/ml lysozyme, 25 μg/ml DNase A and a spatula tip of Pefabloc® (Roche, Applied sciences UK)[57] were added. The resulting mixture was shaken at 4 ºC for 1 hour and sonicated for 15 minutes with pulses of 15 sec with 90% power output. The cell-extract, obtained after centrifugation (50 min, 30000 *g*, 4ºC) was filtered and applied on a 5 ml nickel column equilibrated with wash buffer. The flow through was collected and the protein was obtained by eluting with 30 mM Tris, 330 mM imidazole, 150 mM NaCl, 2 mM β-mercaptoethanol, pH 8 (Elutrion) buffer. The combined fractions containing the mSCP2 protein were dialysed against a 30 mM Tris, 10 mM imidazole, 150 mM NaCl, 5 mM β-

mercaptoethanol, pH 8 (TEV) buffer. To the resulting mixture add 1-2 mg TEV-protease was added per 30-40 mg protein and was left to stand overnight at 4ºC. This mixture was than applied again on a nickel column equilibrated with the (TEV) buffer. The flow-through containing purified mSCP2 was collected and stored at 4ºC. The yield was 4.4 mg (0.3 μmol) per litre culture, with a purity of more than 99% according to SDS-PAGE. Electron spray mass-spectrometry provided a mass of 13334.9 Da (calculated mass 13471.5 Da). The found mass corresponded to the loss of a N-terminal glycine and alanine (calculated mass 13341.3 Da).

2.5.3.2 mSCP2 C91S M104C

The protein mSCP2 C91S M104C was produced by the same procedure as used for the production of mSCP2 using the pET24d::dΔhΔhSCP2 C91S M104 plasmid. The yield was lower than 1 mg per litre culture. Electron spray mass-spectrometry provided a mass of 13427.5 Da (calculated mass 13427.5 Da). A second signal at 13503.5 Da was observed corresponding to the protein containing a disulfide bond to β-mercaptoethanol and one oxidation

2.5.3.3 SCP-2L V83C

PB medium with 100 μg/ml ampicillin was used for expression experiments. Six portions of 10 ml of an overnight culture of *E. coli* BL21(DE3) cells harbouring plasmid pET3d::dΔhΔSCP-2L V83C were used to inoculate six times half a litre of PB medium. The cells were allowed to grow under aerobic conditions at 37ºC to an OD_{600} of 1.5. The addition of IPTG to a final concentration of 0.2 mM initiated the expression of the recombinant protein, followed by further growth overnight. The cells were harvested by centrifugation (20 min, 3000 *g*, 4ºC) and washed wit ice-cold 120 mM NaCl, 16 mM potassium phosphate, pH 7.4 buffer. A portion of the pellet (3.5 g of wet weight) was suspended in 50 ml of 20 mM 3-cyclohexylamino)-1-propanesulfonic acid (CAPS), 0.5 mM benzamidine (BA), pH 11.0. To this suspension 1 mg/ml lysozyme, 25 μg/ml RNase A and 10 mM $MgCl_2$ was added. The resulting mixture was shaken at 4 ºC for 1.5 hours. To the cell-extract, obtained after centrifugation (50 min, 30000 *g*, 4ºC), 0.1 M NaCl was added. The resulting mixture was applied on an anion exchange Hiload Q Sepharose High performance 16/10 column (GE Healthcare, Amersham Biosciences limited, Uppsala, Sweden),

equilibrated with 20 mM CAPS, 0.5 mM BA, 0.35 M NaCl, pH 11. The flow-through was collected by eluting with the equilibration buffer. The combined fractions (250 ml), containing the V83C protein, were dialysed in two steps against a 20 mM 2-(*N*-morpholino)ethanesulfonic acid (MES), 35 mM NaCl, pH 6.0 buffer and applied in portions of 50 ml on a cation exchange Hitrap SP XL 1 ml column (GE Healthcare, Amersham Biosciences AB, Uppsala, Sweden), equilibrated with the dialysis buffer. The bound proteins were eluted with a gradient of 0.35 to 1 M NaCl in the equilibration buffer. The combined fractions containing the V83C protein were combined to give a yield 10-20 mg (0.75-1.5 μmol) / litre culture, with a purity of more than 99% according to SDS-PAGE. Electron spray mass-spectrometry provided a mass of 13245.3 Da, (calculated mass 13248.4 Da).

2.5.3.4 SCP-2L V83C with hexahistadine tag

The reported expression and purification method for SCP-2L[19] resulted in low yields in our hands for both SCP-2L and SCP-2L V83C (<10 mg (0.75 μmol) /L medium,). Therefore the expression and purification procedure was optimized using an attached histidine tag in combination with affinity chromatography:

PB medium with 50 μg/ml kanamycin and 34 μg/ml chloramphenicol (in ethanol) was used for protein production. 10 ml of an overnight culture of *E. coli* BL21(Rosetta) cells harbouring plasmid pEHISTEV::dΔhΔSCP-2L V83C was used to inoculate 0.5 litre of PB medium. The cells were allowed to grow at 37ºC in 3L baffled conical flasks to an OD_{600} of 0.6 after which the temperature was lowered to 16ºC. IPTG was added to a final concentration of 0.2 mM after one hour to initiate the expression of the recombinant protein. The culture was left overnight at 16ºC. The cells were harvested by centrifugation (20 min, 7500 *g*, 4ºC), washed with 16 mM kPi (potassium phosphate) 120 mM NaCl pH 7.4 buffer and centrifuged again (20 min, 3000 *g*, 4ºC). The pellet from 2 litre culture was suspended in 100 ml 50 mM Tris, 20 mM imidazole, 150 mM NaCl, 0.5 mM benzamidine, pH 8 (4ºC) buffer and frozen at -80ºC. After defrosting 20 mg lysozyme, 1 mg DNAse and 1mM $MgCl_2$ was added to the suspension. The suspension was left for 1 hour at 4ºC and was sonicated in portions of 10 ml (Hielscher 4P200S ultrasonic processor, 0.5s pulses 90% power) with 0.5 second pulses for 1 minute. The cell-extract, obtained after centrifugation (50 min, 30000 *g*, 4ºC) was filtered (0.22 μm Millex Filter unit,

Carrigtwohill Co. Cork, Ireland) used for nickel affinity chromatography on a nickel column equilibrated with 30 mM Tris-HCl, 20 mM imidazole, 150 mM NaCl, pH 8 (wash buffer). The column was washed with 5 column volumes of wash buffer, 5 column volumes of high salt buffer (wash buffer containing 1M 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). To the elution containing the Hist-SCP-2L V83C protein, was immediately added an equal volume of wash buffer (to prevent precipitation of the protein) and the resulting solution was dialysed against 5 litre 30 mM Tris, 10 mM imidazole, 150 mM NaCl, pH 8 buffer. If precipitate had formed the resulting mixture was centrifuged (50 min, 30000 *g*, 4ºC). To this resulting Hist-SCP-2L V83C solution was added 0.014 equivalents of TEV-protease and final concentrations of 1 mM DTT and 0.5 mM EDTA from 100x concentrated stocks. This mixture was left overnight at 4ºC. If precipitate had formed the mixture was again centrifuged (50 min, 30000 *g*, 4ºC) and the supernatant used. The resulting mixture was then filtered and used for nickel affinity chromatography on a nickel column equilibrated with wash buffer. The flow-through containing pure SCP-2L V83C was collected and concentrated and buffer exchanged to 20 mM MES pH 6.6 using a centrifugal concentrator. To the resulting concentrated SCP-2L V83C mixture 20% glycerol was added and stored at -20ºC. Average yield is ~60 mg (4.5 μmol) / litre culture, with a purity of more than >99% according to SDS-PAGE. Electron spray mass-spectrometry provided a mass of 13376.0 Da (calculated mass 13376.6 Da).

2.5.3.5 SCP-2L A100C and E14C

Expression and purification of SCP-2L A100C and E14C was performed using the same method as for SCP-2L V83C containing a hexahistidine tag except for a small difference in the preparation of the overnight culture before expression. For the overnight culture 10 ml LB medium (Lysogeny Broth medium; containing 10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl, pH 7.0) was used for plasmid pre-culture which was inoculated with a glycerol stock with *E. coli* BL21(Rosetta) cells containing the pEHISTEV::dΔhΔSCP-2L A100C or E14C plasmid. A small amount (100 μl) of the preculture was than used for the actual overnight culture.

The yield for SCP-2L A100C was ~30 mg (2.23 μmol) / litre culture, with a purity of more than >99% according to SDS-PAGE. Electron spray mass-spectrometry provided a mass of 13405.30 Da (calculated mass 13404.60 Da).
The yield for SCP-2L E14C was ~5 mg (0.37 μmol) / litre culture, with a purity of more than >99% according to SDS-PAGE. Electron spray mass-spectrometry provided a mass of 13346.0 Da (calculated mass 13346.6 Da).

2.5.3.6 ^{15}N labelled SCP-2L

An Erlenmeyer flask containing LB with 50 μg/ml kanamycin stock and a 34 μg/ml chloramphenicol (from stock in ethanol) was inoculated with *E. coli* BL21(Rosetta) cells harbouring plasmid pEHISTEV::dΔhΔSCP-2L from a glycerol stock and left overnight at 37ºC. 10 ml of overnight culture was centrifuged (30 min, 3000 *g*, 4ºC), and the pellet was suspended in 500 ml M9 medium (500 ml M9: 380 ml H_2O, 100 ml M9 salts (39 g/L Na_2HPO_4, 15 g/L KH_2PO_4, 2.5 g/L NaCl), 1 ml 1 M $MgSO_4$, 0.5 ml 0.5 M $CaCl_2$, 10 ml 20% glucose, 7.5 ml 10% NH_4Cl, 0.1 ml 1% thiamine and 0.1 ml trace elements mix (3.3 mg/ml biotin, 160 μM $MnSO_4$, 10 μM $CuSO_4$, 30 μM $CoCl_2$, 0.4 mM boric acid, 10 μM $ZnSO_4$)) containing 50 μg/ml kanamycin and 34 μg/ml chloramphenicol. The cells were allowed to grow under aerobic conditions at 37ºC to an OD_{600} of ~0.55 after which the temperature was lowered to 16ºC. After about 20 minutes IPTG was added to a final concentration of 0.2 mM, to initiate the expression of the recombinant protein. The culture was than allowed to grow overnight at 16ºC after which a final OD_{600} of about 2 was reached. The harvesting and purification of the protein was done according to the same procedure as for the non-isotopic labelled protein described above. Electron spray mass-spectrometry provided a mass of 13528.9 Da (calculated mass 13533.13 Da). The obtained yield was ~6 mg (0.44 μmol) / litre culture

2.5.4 Fluorescent binding assay

A 1 cm cuvette was used equipped with a stirring bar. Samples were only carefully stirred. Pyr-C12 was diluted from a DMSO stock to 75 μM in 17.5 mM β-cyclodextrin in a 20 mM MES 35 mM NaCl pH 6 buffer. This solution was vortexed for 1 min and sonicated for 15 min. The obtained solution was directly used for the titration measurements. The titrations were done by addition of increasing amounts of

the Pyr-C12 stock to a 0.5 μM SCP-2L V83C solution in 20 mM 50 mM NaCl pH 6 buffer. To measure the binding the emission maximum of Pyr-C12 at 378.93 nm was followed after excitation of the tryptophans at 280nm. The emissions were corrected for direct excitation of Pyr-C12 and dilution. The K_d was obtained as by a non-linear curve fit using Equation 1.to the obtained data provided by measurement of the increase in sensitised emission at 378.93 nm versus increasing Pyr-C12 concentration.

Curve fit to the equation

$Y=c*((1+G_0*K_d+H_0* K_d-sqrt(-4*G_0*H_0*K_d\text{^}2+(-1-G_0*K_d-H_0*K)\text{^}2))/(2*a))$

Obtained from

$G = G_0 – H_0$

$H = H_0 – HG$

$K_d * H\ G = HG$

with

Y = measure fluorescence

c = constant

G = Concentration guest

H = Concentration host

H_0 = Host concentration before addition of G

G_0 = Total concentration of added G

Equation 1. Equations used for curve fits to calculate K_d

2.5.5 Experiments to confirm the incorporation of the free cysteine

*2.5.5.1 Bioconjugation of maleimide **M1** 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 maleimide **M1** added from a 50 mM stock solution in DMF. ESI-MS analysis after mixing overnight yielded a set of signals for SCP-2L V83C corresponding to a mass of 13619.1 Da (SCP-2L V83C-**M1** calculated 13619.9 Da) and for SCP-2L A100C corresponding to a mass of 13646.9 Da (SCP-2L A100C-**M1** calculated 13647.9 Da). Typically no residual free protein was observed. The mixture was washed using a centrifugal concentrator with buffer to remove the excess **M1**.

2.5.5.2 General procedure for coupling reactions with Ellman reagent

These experiments are based on a published procedure.[51] A final concentration of 0.6 mM 5,5'dithio-bis(2-nitrobenzoic acid) (Ellman's reagent/DTNB) was added from a 20 mM stock solution in water to a 13 μM solution of protein in 0.1 ml of buffer solution (20 mM MES, 50 mM NaCl, pH 6). The reaction was shaken for an hour at room temperature before analysis by LCMS (ES^+). For the unique cysteine containing proteins conversion to the TNB adducts was found, while for the modified proteins no modification with TNB was observed.

2.5.5.3 Tryptic Digests

The samples (5 μL, 10 pmoles/μL) were dialysed against buffer (50 mM ammonium bicarbonate pH 8.0) using a membrane filter (Millipore, Billerica, MA). Trypsin (0.5 μL, 0.1 μg, Promega, Madison, WI) was added to the resulting solutions. The solutions were incubated at 37°C overnight and the digested solution (0.5 μL). This solution together with an alpha-cyano-4-hydroxycinnamic acid matrix (0.5 μL, 10 mg/mL in 50:50 acetonitrile:0.1% TFA) and 0.1% TFA (0.5 μL) was applied to the MALDI target, and after being allowed to dry, analyzed by MALDI-MS.

MALDI MS was acquired using a 4800 MALDI TOF/TOF Analyser (Applied Biosystems, Foster City, CA) equipped with a Nd:YAG 355 nm laser and calibrated using a mixture of peptides. The samples were initially analysed in positive MS mode between 800 and 4000 m/z, by averaging 1000 laser shots. The most desired peptides were selected for MSMS analysis and acquired to a maximum of 3000 laser shots or until the accumulated spectrum reached a S/N ratio of 35. All MS/MS data were acquired using 1 keV collision energy.

The combined MS and MSMS data were analysed, using GPS Explorer (Applied Biosystems) to interface with the Mascot 2.1 search engine (Matrix Science, London, UK), against the UniProt (Swiss-Prot and TREMBL combined) database (April 2009). No species restriction was applied. The data were searched with tolerances of 100 ppm for the precursor ions and 0.5 Da for the fragment ions, with trypsin as the cleavage enzyme, assuming up to one missed cleavage, carbamidomethyl modification of cysteines as a fixed modification and methionine oxidation selected as a variable modification. The depicted peptides were the only peptides found matching maleimide (including +18) modification.

In the above experiment an active search was performed for other peptides with the same modification but no evidence for those was found. In addition, the samples were analysed by ESI where again no evidence was found for other peptide sequences containing the modification. For ESI analysis the peptides were acidified with TFA and then separated using an UltiMate nanoLC (LC Packings, Amsterdam) equipped with a PepMap C18 trap & column, using a 60 min gradient of increasing acetonitrile concentration, containing 0.1 % formic acid (5-35% acetonitrile in 35 min respectively, 35-50% in a further 20 min, followed by 95% acetonitrile to clean the column). The eluent was sprayed into a Q-Star XL tandem mass spectrometer (Applied Biosystems, Foster City, CA) and analysed in Information Dependent Acquisition (IDA) mode, performing 1 sec of MS followed by 3 sec MSMS analyses of the 2 most intense peaks seen by MS. These masses are then excluded from analysis for the next 60 sec. MS/MS data for doubly and triply charged precursor ions were converted to centroid data, without smoothing, using the Analyst QS1.1 mascot.dll data import filter with default settings. The MS/MS data file generated was analysed using the Mascot 2.1 search engine (Matrix Science, London, UK) against an internal database containing the protein sequence of interest. The data were searched with tolerances of 0.2 Da for the precursor and fragment ions, with trypsin as the cleavage enzyme and allowing for one missed cleavage. Maleimide **M1** modification of cysteine and lysine and the oxidation of methionine were selected as variable modifications.

2.6 References

[1] a) 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; b) R. R. Davies, M. L. Brown, M. D. Distefano, *Protein Eng.* **1997**, *10*, 61; c) T. Heinisch, T. R. Ward, *Curr. Opin. Chem. Biol.* **2010**, *14*, 184.

[2] K. J. Jensen, *Peptide and Protein Design for Biopharmaceutical Applications* **2009**, 207.

[3] G. T. Hermanson, *Bioconjugate Techniques*, Second ed., Acadamic Press, San Diego, **2008**.

[4] a) R. R. Davies, M. D. Distefano, *J. Am. Chem. Soc.* **1997**, *119*, 11643; b) R. R. Davies, H. Kuang, D. Qi, A. Mazhary, E. Mayaan, M. D. Distefano, *Bioorg. Med. Chem. Lett.* **1999**, *9*, 79.

[5] F. W. Studier, *J. Mol. Biol.* **1991**, *219*, 37.

[6] a) C. Di Guan, P. Li, P. D. Riggs, H. Inouye, *Gene* **1988**, *67*, 21; b) C. V. Maina, P. D. Riggs, A. G. Grandea, III, B. E. Slatko, L. S. Moran, J. A. Tagliamonte, L. A. McReynolds, C. Guan, *Gene* **1988**, *74*, 365.
[7] a) K. L. Guan, J. E. Dixon, *Anal. Biochem.* **1991**, *192*, 262; b) D. J. Hakes, J. E. Dixon, *Anal. Biochem.* **1992**, *202*, 293; c) D. B. Smith, K. S. Johnson, *Gene* **1988**, *67*, 31.
[8] a) T. C. Evans, Jr., J. Benner, M.-Q. Xu, *J. Biol. Chem.* **1999**, *274*, 3923; b) T. Watanabe, Y. Ito, T. Yamada, M. Hashimoto, S. Sekine, H. Tanaka, *J. Bacteriol.* **1994**, *176*, 4465.
[9] a) D. J. Kroll, H. A. M. Abdel-Hafiz, T. Marcell, S. Simpson, C. Y. Chen, A. Gutierrez-Hartmann, J. W. Lustbader, J. P. Hoeffler, *DNA Cell Biol.* **1993**, *12*, 441; b) F. W. Studier, B. A. Moffatt, *J. Mol. Biol.* **1986**, *189*, 113.
[10] H. Liu, J. H. Naismith, *Protein Expr. Purif.* **2009**, *63*, 102.
[11] Http://www.gelifesciences.com/aptrix/upp00919.nsf/Content/54F1E49 F7BCDD1E5C1257628001D110D/$file/11000888AF.pdf, *GE Healthcare*, *HisTrap FF 1 and 5 ml Instructions manual*, 17/10, **2010**.
[12] a) R. Raja, J. M. Thomas, M. Xu, K. D. M. Harris, M. Greenhill-Hooper, K. Quill, *Chem. Commun. (Cambridge, U. K.)* **2006**, 448; b) N. Herron, C. A. Tolman, *J. Am. Chem. Soc.* **1987**, *109*, 2837.
[13] a) H. M. Berman, Henrick Kim, Nakamura Haruki, in *Nat. Struct. Biol., Vol. 10*, **2003**, p. 980; b) H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, *Nucleic Acids Res.* **2000**, *28*, 235; c) www.pdb.org, *World Wide Protein Data Base (wwPDB)*, 01-04, **2010**.
[14] a) F. Schroeder, B. P. Atshaves, A. L. McIntosh, A. M. Gallegos, S. M. Storey, R. D. Parr, J. R. Jefferson, J. M. Ball, A. B. Kier, *Biochim. Biophys. Acta, Mol. Cell Biol. Lipids* **2007**, *1771*, 700; b) A. M. Gallegos, B. P. Atshaves, S. M. Storey, O. Starodub, A. D. Petrescu, H. Huang, A. L. McIntosh, G. G. Martin, H. Chao, A. B. Kier, F. Schroeder, *Prog. Lipid Res.* **2001**, *40*, 498; c) U. Seedorf, P. Ellinghaus, J. R. Nofer, *Biochim. Biophys. Acta, Mol. Cell Biol. Lipids* **2000**, *1486*, 45.
[15] A. Frolov, T.-H. Cho, J. T. Billheimer, F. Schroeder, *J. Biol. Chem.* **1996**, *271*, 31878.
[16] F. Schroeder, S. C. Myers-Payne, J. T. Billheimer, W. G. Wood, *Biochemistry* **1995**, *34*, 11919.
[17] R. J. A. Wanders, P. Vreken, S. Ferdinandusse, G. A. Jansen, H. R. Waterham, C. W. T. Van Roermund, E. G. Van Grunsven, *Biochem. Soc. Trans.* **2001**, *29*, 250.
[18] D. W. Russell, *Annu. Rev. Biochem.* **2003**, *72*, 137.
[19] A. M. Haapalainen, D. M. F. van Aalten, G. Merilaeinen, J. E. Jalonen, P. Pirilae, R. K. Wierenga, J. K. Hiltunen, T. Glumoff, *J. Mol. Biol.* **2001**, *313*, 1127.
[20] T. Choinowski, H. Hauser, K. Piontek, *Biochemistry* **2000**, *39*, 1897.
[21] B. J. Noland, R. E. Arebalo, E. Hansbury, T. J. Scallen, *J. Biol. Chem.* **1980**, *255*, 4282.
[22] M. Raabe, U. Seedorf, H. Hameister, P. Ellinghaus, G. Assmann, *Cytogenet. Cell Genet.* **1996**, *73*, 279.
[23] R. C. Crain, D. B. Zilversmit, *Biochemistry* **1980**, *19*, 1433.
[24] D. H. Dyer, S. Lovell, J. B. Thoden, H. M. Holden, I. Rayment, Q. Lan, *J. Biol. Chem.* **2003**, *278*, 39085.

[25] J. Edqvist, E. Roennberg, S. Rosenquist, K. Blomqvist, L. Viitanen, T. A. Salminen, M. Nylund, J. Tuuf, P. Mattjus, *J. Biol. Chem.* **2004**, *279*, 53544.
[26] C. J. Bult, O. White, G. J. Olsen, L. Zhou, R. D. Fleischmann, G. G. Sutton, J. A. Blake, L. M. FitzGerald, R. A. Clayton, et al., *Science (Washington, D. C.)* **1996**, *273*, 1058.
[27] F. V. Filipp, M. Sattler, *Biochemistry* **2007**, *46*, 7980.
[28] a) W. A. Stanley, K. Versluis, C. Schultz, A. J. R. Heck, M. Wilmanns, *Arch. Biochem. Biophys.* **2007**, *461*, 50; b) T. B. Dansen, J. Westerman, F. S. Wouters, R. J. A. Wanders, A. Van Hoek, T. W. J. Gadella, Jr., K. W. A. Wirtz, *Biochem. J.* **1999**, *339*, 193.
[29] A. Frolov, K. Miller, J. T. Billheimer, T.-H. Cho, F. Schroeder, *Lipids* **1997**, *32*, 1201.
[30] W. A. Stanley, F. V. Filipp, P. Kursula, N. Schueller, R. Erdmann, W. Schliebs, M. Sattler, M. Wilmanns, *Mol. Cell* **2006**, *24*, 653.
[31] F. Lopez Garcia, T. Szyperski, J. H. Dyer, T. Choinowski, U. Seedorf, H. Hauser, K. Wuthrich, *J. Mol. Biol.* **2000**, *295*, 595.
[32] a) J. C. Cole, J. W. M. Nissink, R. Taylor, *Drug Discovery Series* **2005**, *1*, 379; b) M. J. Hartshorn, M. L. Verdonk, G. Chessari, S. C. Brewerton, W. T. M. Mooij, P. N. Mortenson, C. W. Murray, *J. Med. Chem.* **2007**, *50*, 726; c) G. Jones, P. Willett, R. C. Glen, *J. Mol. Biol.* **1995**, *245*, 43; d) G. Jones, P. Willett, R. C. Glen, A. R. Leach, R. Taylor, *J. Mol. Biol.* **1997**, *267*, 727; e) J. W. M. Nissink, C. Murray, M. Hartshorn, M. L. Verdonk, J. C. Cole, R. Taylor, *Proteins Struct., Funct., Genet.* **2002**, *49*, 457; f) M. L. Verdonk, G. Chessari, J. C. Cole, M. J. Hartshorn, C. W. Murray, J. W. M. Nissink, R. D. Taylor, R. Taylor, *J. Med. Chem.* **2005**, *48*, 6504; g) M. L. Verdonk, J. C. Cole, M. J. Hartshorn, C. W. Murray, R. D. Taylor, *Proteins Struct., Funct., Genet.* **2003**, *52*, 609.
[33] J. Kopp, T. Schwede, *Nucleic Acids Res.* **2004**, *32*, D230.
[34] G. Stier, *Unpublished*.
[35] R. B. Kapust, J. Toezser, J. D. Fox, D. E. Anderson, S. Cherry, T. D. Copeland, D. S. Waugh, *Protein Eng.* **2001**, *14*, 993.
[36] a) Y. De Launoit, J. Adamski, *J. Mol. Endocrinol.* **1999**, *22*, 227; b) M. Baes, S. Huyghe, P. Carmeliet, P. E. Declercq, D. Collen, G. P. Mannaerts, P. P. Van Veldhoven, *J. Biol. Chem.* **2000**, *275*, 16329.
[37] W. A. Stanley, A. Sokolova, A. Brown, D. T. Clarke, M. Wilmanns, D. I. Svergun, *Journal of Synchrotron Radiation* **2004**, *11*, 490.
[38] U. Seedorf, S. Scheek, T. Engel, C. Steif, H. J. Hinz, G. Assmann, *J. Biol. Chem.* **1994**, *269*, 2613.
[39] C. Jatzke, H.-J. Hinz, U. Seedorf, G. Assmann, *Biochim. Biophys. Acta, Proteins Structure and Molecular Enzymology* **1999**, *1432*, 265.
[40] D. S. Waugh, http://mcl1.ncifcrf.gov/waugh_tech/faq/tev.pdf, *TEV Protease FAQ*, 17/10, **2010**.
[41] http://www.ap-lab.com/circular_dichroism.htm, *Alliance Protein Laboratories Webpage*, 30-10, **2010**.
[42] a) M. Piotto, V. Saudek, V. Sklenar, *J. Biomol. NMR* **1992**, *2*, 661; b) V. Sklenar, M. Piotto, R. Leppik, V. Saudek, *Journal of Magnetic Resonance, Series A* **1993**, *102*, 241.
[43] R. T. McKay, *Annu. Rep. NMR Spectrosc.* **2009**, *66*, 33.
[44] G. D. Fasman, Editor, *Circular Dichroism and the Conformational Analysis of Biomolecules*, **1996**.

[45] J. Podtetenieff, A. Taglieber, E. Bill, E. J. Reijerse, M. T. Reetz, *Angew. Chem., Int. Ed.* **2010**, *49*, 5151.
[46] http://www.jascoinc.com/Products/Spectroscopy/J-815-Circular-Dichroism-Spectrometer/Features-Benefits.aspx, *Jasco Circular Dichroism website*, 30-10, **2010**.
[47] M. Kristian Koski, A. M. Haapalainen, J. K. Hiltunen, T. Glumoff, *J. Mol. Biol.* **2005**, *345*, 1157.
[48] P. Jeschke, A. Harder, N. Mencke, (Bayer A.-G., Germany). DE19538960, **1997**, p. 80 pp.
[49] G. L. Ellman, *Arch. Biochem. Biophys.* **1959**, *82*, 70.
[50] C. K. Riener, G. Kada, H. J. Gruber, *Anal. Bioanal. Chem.* **2002**, *373*, 266.
[51] J. M. Chalker, C. S. C. Wood, B. G. Davis, *J. Am. Chem. Soc.* **2009**, *131*, 16346.
[52] Promega, http://www.promega.com/tbs/9piv511/9piv511.pdf, *Abstract for Sequencing Grade Modified Trypsin*, 31-10, **2010**.
[53] a) S. Bhat, M. G. Sorci-Thomas, E. T. Alexander, M. P. Samuel, M. J. Thomas, *J. Biol. Chem.* **2005**, *280*, 33015; b) C. L. Swaim, D. L. Smith, J. B. Smith, *Protein Sci.* **2004**, *13*, 2832.
[54] M. M. Bradford, *Anal. Biochem.* **1976**, *72*, 248.
[55] D. S. Waugh, S. Cherry and J. E. Tropea, http://mcl1.ncifcrf.gov/waugh_tech/protocols/pur_histev.pdf, *Expression & Purification of His-TEV(S219V)-Arg*, 17/10, **2010**.
[56] D. S. Waugh, http://mcl1.ncifcrf.gov/waugh_prk793.html, *E. coli strain BL21(DE3)-RIL/pRK793*, 17/10, **2010**.
[57] Roche, http://www.roche-applied-science.com/pack-insert/1429868a.pdf, 07-11, **2010**.

Chapter 3. Phosphine-protein conjugates

The development of an efficient modification procedure of unique cysteine SCP-2L templates with phosphine ligands

3.1 Abstract

In this chapter, strategies are explored for cysteine selective modification of the SCP-2L templates described in chapter 2 with phosphine-ligands. A reported modification procedure using carboxylic acid modified phosphines was found to be unselective and two-step procedures using the copper catalysed Huisgen 1,3-dipolar cycloaddition and a light induced 1,3 dipolar cycloaddition were found to be inefficient. A two-step procedure involving introduction of a hydrazide via maleimide linkage followed by hydrazone bond formation with phosphine aldehydes proved highly selective and efficient, and several phosphine moieties were successfully introduced using this methodology. This is the first generally applicable and effective bioconjugation procedure for phosphine-ligands to proteins. The conjugation products were analysed by LC-MS and ^{31}P-NMR to study the modification products. The cysteine selectivity of the reactions was confirmed by analysis of the products of tryptic digestion of the protein and by the application of Ellman reagent. The effects of the modifications on the protein fold were studied using CD-spectroscopy and watergate ^{1}H-NMR. The extent of the distortion to the protein structure upon modification was highly dependent on the ligand used, but using a study with a fluorescent probe it was demonstrated that the substrate binding capability of the protein remains largely unaffected. This work resulted in a set of phosphine modified proteins ready for transformation into phosphine containing artificial metalloenzymes.

The developed phosphine modification procedure has been published as a communication in Angewandte Chemie International Edition:
P. J. Deuss, G. Popa, C. H. Botting, W. Laan, P. C. J. Kamer, *Angew. Chem., Int. Ed.* **2010**, *49*, 5315.

3.2 Introduction

3.2.1 Phosphines as ligands in homogeneous catalysis

In chapter 1 the diversity of the various artificial metalloenzymes developed to date is outlined. A wide variety of transition metal complexes have been introduced in a number of different proteins, and the assemblies have been applied in a variety of catalytic reactions, showing the versatility and promise of these systems. The developed protein templates described in the previous chapter were specifically chosen for the particular shape of their hydrophobic pocket and the possibilities this offers in catalytic reactions involving linear aliphatic substrates.

Since the first report on the use of rhodium phosphine complexes as hydrogenation catalysts,[1] phosphine-transition metal complexes have evolved to one of the most successful types of homogeneous catalysts for industrial reactions. They have been successfully applied in a variety of reactions like hydroformylation, hydrogenation, hydrocyanation, oligomerization and cross-coupling reactions.[2] Phosphine transition metal complexes are renowned for their tunability of their activity and selectivity by applying small changes to the ligand structure.[3] Thus, such complexes are a very attractive class of catalysts for the development of artificial metalloenzymes via introduction into our unique cysteine bearing protein template.

3.2.2 Phosphine containing artificial metalloenzymes

Phosphine transition metal complexes have already been successfully applied in artificial metalloenzymes. Whitesides successfully employed a supramolecular assembly of avidin with a biotin modified rhodium diphosphine complex in the enantioselective hydrogenation of prochiral alkenes (Figure 1a).[4] This system was later optimized in the groups of Ward[5] and others[6], demonstrating that combining chemical and biomolecular techniques is a very effective approach for hybrid catalyst optimization. Ward applied the same approach to create phosphine based artificial metalloenzymes for the palladium catalysed asymmetric allylic alkylation.[7] Another very interesting system is the use of monoclonal antibodies generated against compound **CA** (Figure 1b).[8] A combination of this rhodium diphosphine and one such antibody resulted in a very enantioselective hydrogenation catalyst. Reetz covalently attached the same diphosphine moiety to lipases via a phosphonate

linkage; however the formed phosphonate bond proved hydrolytically unstable (Figure 1c).[9]

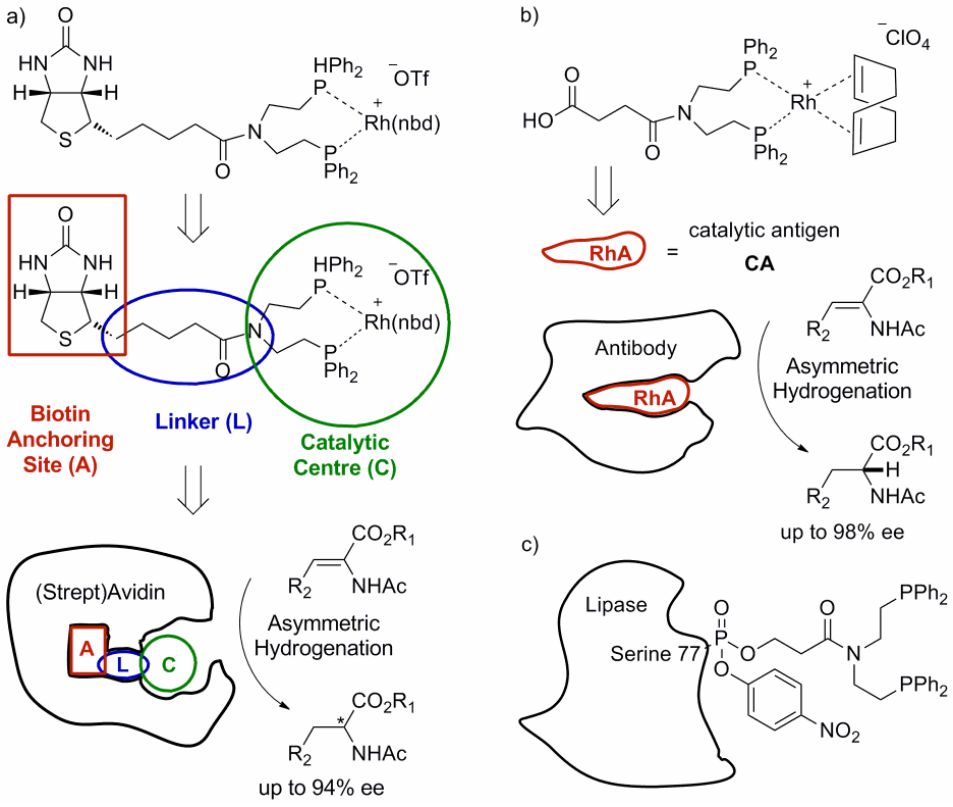


Figure 1, a) Diphosphine systems based on the supramolecular anchoring of modified biotin to avidin b) Catalyst system based on monoclonal antibodies raised against diphosphine complex **CA** c) Diphosphine ligand covalently linked to a lipase by phosphonate linkage to serine 77.

All the assembly approaches outlined above have the disadvantage that only specific proteins can be used. In our approach a free choice of the protein structure is desired in order to be able to choose the protein based on specific characteristics like affinity for a substrate of choice. This is why we opt for the phosphine transition metal complex to be covalently and site selectively introduced. Such a covalent modification approach has already been applied to DNA.[10] Phosphines were successfully introduced into DNA by site selective modification of amine modified oligonucleotides using *N*-(3-dimethylaminopropyl)-*N*'-ethylcarbodiimide (EDC)/ *N*-hydroxysuccinimide (NHS) activated carboxylic acid phosphines. Unfortunately these systems have so far proved difficult to employ in catalysis.[11]

3.2.3 Site selective introduction of phosphine ligands into proteins

We decided to use selective cysteine bioconjugation for the site selective covalent introduction of transition metal binding ligands. A variety of methods for cysteine-selective bioconjugation has been developed,[12] but selective cysteine modification of proteins with phosphines using such techniques has proven anything but straightforward.[13] Unprotected phosphines containing a maleimide or iodo-alkane cannot be synthesized because of the nucleophilic character of the phosphine, leading to the formation of phosphonium salts and phosphorus ylides.[14] Disulfide-bridge formation is also incompatible with free phosphines because free phosphines are known to reduce this bond.[15] Copper catalysed or copper free Huisgen 1,3-dipolar cycloadditions have proven very successful in protein modification,[16] but azides are incompatible with free phosphines as demonstrated by the Staudinger reaction[17], which in fact has been employed for biofunctionalisation.[18] Our group has introduced borane protected phosphine maleimides in different proteins, however subsequent attempts at deprotection using various methods proved unsuccessful or ineffective.[13b] Several other transition metal catalysed reactions have been applied in protein modification,[16b, 19] however the phosphine ligand to be introduced can interfere with the reaction and/or complicate transition metal removal following the coupling reaction. Because a generally applicable method for the cysteine-selective covalent modification of proteins with phosphines was lacking, we have explored several different approaches towards such a method.

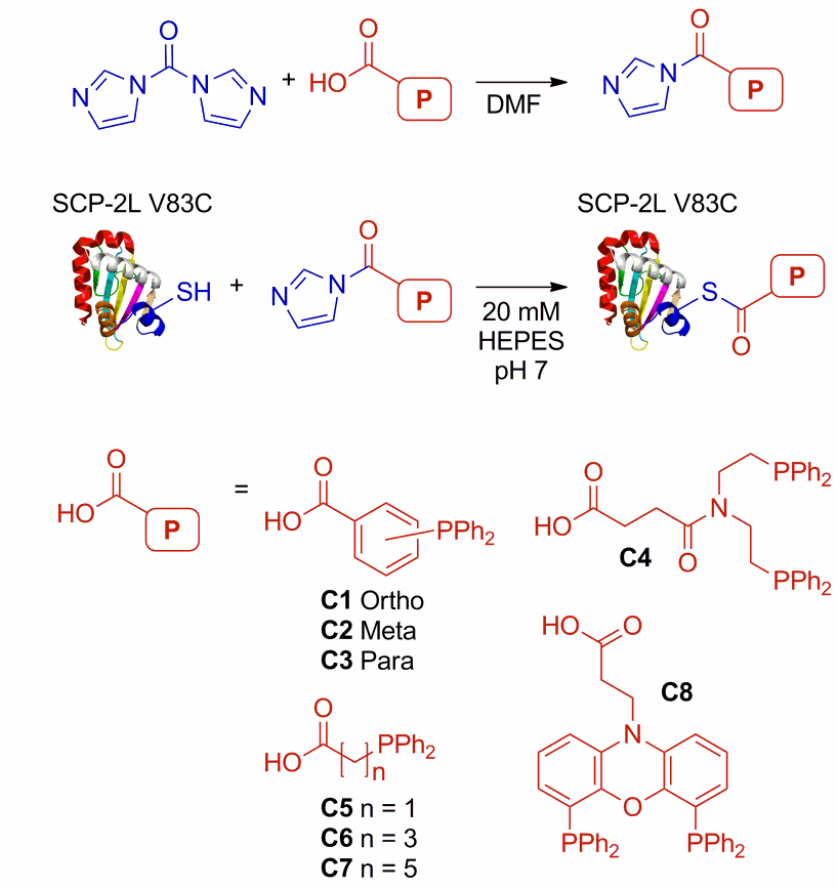
3.3 Results and discussion

3.3.1 Bioconjugation using CDI activated phosphino-carboxylic acids

3.3.1.1 Modification of SCP-2L V83C using CDI activated carboxylic acid containing mono and diphosphines

Recently our group selectively functionalized the cysteine of photoactive yellow protein using 1,1-carbodiimidazole (CDI) activated carboxylic acid bearing phosphines.[20] This is a straightforward procedure as carboxylic acid bearing phosphines are commercially available or can be readily synthesised. The carboxylic acid is activated with CDI, a reaction that can be monitored with ^{31}P-NMR. The obtained (crude) imidazolide solution is added upon completion of the activation reaction to the unique cysteine bearing protein template. This results in highly

selective and efficient thioester bond formation with the cysteine of PYP. We investigated whether we could apply the same method for the modification of the SCP-2L V83C template described in chapter 2. Phosphine carboxylic acids **C1-C8** were used for modification reactions using conditions similar to those established for PYP (Scheme 1, Table 1).[20b] The modification of PYP was successfully performed at pH 8, although the primary amine of lysines present in proteins are expected to act as nucleophiles at this pH. Modification of lysines was found to be a problem when this method was applied to other unique cysteine containing proteins used in our group.[13b] A 20 mM 4-(2-hydroxyethyl)piperazine-1-ethanesulfonic acid (HEPES) pH 7 buffer was used for the modification reactions to decrease the reactivity of the lysines.



Scheme 1. Reaction of unique cysteine of SCP-2L V83C with CDI activated carboxylic acids **C1-C8**.

Conversion to mono modified SCP-2L V83C was found to be low using up to 5 equivalents of CDI activated carboxylic acids (entries 1-5). It was attempted to improve the conversion by increasing the excess of CDI activated carboxylic acids

give, however this gave rise to double, triple and even quadruple modification (entries 6-10 and 12). Experiments were also performed at lower pH by using a 20 mM 2-(*N*-morpholino)ethanesulfonic acid (MES) pH 6.6 buffer, however the same modification patterns were obtained for the modification of SCP-2L V83C. The use of an even lower pH (<6.5) resulted in very little or no modification. For ligands **C6** and **C8** remarkably low conversion was found even with high excesses of ligand (entries 11 and 13).

Table 1. Results of bioconjugation of SCP-2L V83C with CDI activated carboxylic acid phosphines **C1**-**C8**.[a]

#	Ligand	Equiv.	Apo protein (%)	Mono mod. (%)	Multiple coupled ligands		
					2x (%)	3x (%)	4x (%)
1	**C1**	5	69	24	4	3	-
2	**C2**	5	65	30	4	-	-
3	**C3**	5	32	55	8	2	1
4	**C5**	5	97	3	-	-	-
5	**C7**	5	93	7	-	-	-
6	**C2**	10	41	46	13	-	-
7	**C1**	20	72	22	4	2	-
8	**C2**	20	18	51	26	4	-
9	**C3**	20	14	52	22	10	2
10	**C5**	20	61	38	1	-	-
11	**C6**	20	93	7	-	-	-
12	**C7**	20	62	31	7	-	-
13	**C8**	20	88	12	-	-	-
14	**C4**	20	12	88	-	-	-

a) Conditions: 15 µM SCP-2L V83C, 5-20 equivalents of CDI activated carboxylic acid phosphines (from 25 mM stock in DMF) in 20 mM HEPES pH 7 buffer. Product formation (%) determined by signal ratio in mass spectroscopy (ES^+).

It is well known that cysteine containing proteins may form dimers *via* disulfide bridge formation rendering bioconjugation to these cysteines ineffective.[21] However, addition of various amounts of the commonly used disulfide bridge reducing agent

tris(2-carboxyethyl)phosphine (TCEP)[22] (1-100 eq.) to the conjugation reactions did not improve the conversion or selectivity of the reaction. In addition, no proof for the formation of dimers was observed upon LCMS(ES^+) analysis demonstrating that this protein does not readily form dimers. Using EDC as an alternative activator for the carboxylic acids resulted in even poorer selectivity.

These results prove that other amino acid side chains get modified in these reactions using CDI activated carboxylic acid phosphine for cysteine modification. The extent of the formation of multiple-modification products is determined by the excess and nature of the phosphine used. These results show that the chemoselectivity of the conjugation reaction is low, in line with results obtained with other unique cysteine bearing proteins used in our group.[13b] Low cysteine-selectivity of the CDI-methodology can be expected as activated carboxylic acids are generally used for lysine specific protein modification.[8] PYP, whose cysteine is also used *in vivo* for modification with a cofactor via thioester formation, seems exceptionally suited for cysteine-selective bioconjugation with carboxylic acids.

Interestingly, the most intense signal for the modification products found upon LCMS (ES^+) analysis was typically of protein containing non oxidized phosphine(s) even though the conditions under which the mass-spectra are obtained are not oxygen free. In addition, a small signal (5-20%) of +16 mass units compared to the expected modification is usually observed. Although this signal corresponds to the oxidation of the phosphine, a signal with an increase mass of 16 Da is also present for the protein not containing the phosphine and is most likely the result of methionine oxidation, which is a commonly observed phenomenon in mass spectroscopy on proteins.

*3.3.1.2 Bioconjugation of CDI activated **C4***

Interestingly CDI activated carboxylic acid **C4** did give very high mono modification using 20 equivalents ligand at pH 7 (Table 1 entry 14). For this particular ligand better conversion and reproducibility were obtained using a 20 mM MES buffer of pH 6.6. LCMS(ES^+) analysis of mixtures of modified and unmodified protein showed that the unmodified protein ionizes about five times better than the modified, indicating that the conversion to the mono modified protein is in fact higher than 95%. Using ^{31}P-NMR it was demonstrated that the **C4** can be introduced without oxidation of the phosphines, which is very important for application of these phosphine containing

proteins as artificial metalloenzymes (see appendix A3). However, since the other reactions involving CDI activated carboxylic acid ligand presented above, clearly appeared unselective, the question remained if this modification is exclusive to the cysteine. Determining the selectivity of the reaction by analysis of tryptic digest of the protein proved troublesome, as the conditions used for this led to hydrolysis of the thioester bond. However, upon the reaction of wild-type SCP-2L with 5 equivalents of **C4** the same extent of mono modification was found as for SCP-2L V83C (Table 2). This experiment was performed at pH 6.6 (compared to 7 for above experiments) where cysteine modification should be more selective. Thus, this result demonstrates that the modification of SCP-2L V83C with ligand **C4** is clearly unselective for the cysteine. The reason for this apparent selective mono modification might be the result of protein instability after modification with two or more ligands causing it to precipitate and not appear in LCMS(ES^+) analysis of the soluble fraction of the modification reaction. This also explains the low amount of soluble protein (~10%) measured in solution after the coupling reaction and subsequent concentration and washing steps. Nonetheless, several experiments were performed to study the effect of the modification of **C4** on the protein fold (see appendix A3).

Table 2. Bioconjugation of CDI-activated **C4** to SCP-2L (WT) and SCP-2L V83C.[a]

#	SCP-2L mutant	Cofactor	Equivalents	Conversion (%)
1	WT	**C4**	1	17
2	WT	**C4**	5	23
3	V83C	**C4**	1	21
4	V83C	**C4**	5	32

a) Conditions: 15 μM protein solution, 1-5 equivalents of activated CDI activated carboxylic acid phosphine **C4** (from 25 mM stock in DMF) in 20 mM MES pH 6.6 buffer. Product formation (%) determined by signal ratio in mass spectroscopy (ES^+).

In the above experiments it was demonstrated that using CDI activated carboxylic acids to selectively functionalize the unique cysteine in SCP-2L V83C was unsuccessful. Especially the modification of other amino acid residues hinders application of the resulting conjugates as well-defined phosphine modified artificial metalloenzymes. Another method for selective introduction of phosphines that is

applicable to unique cysteine containing SCP-2L mutants and other such protein templates had to be developed.

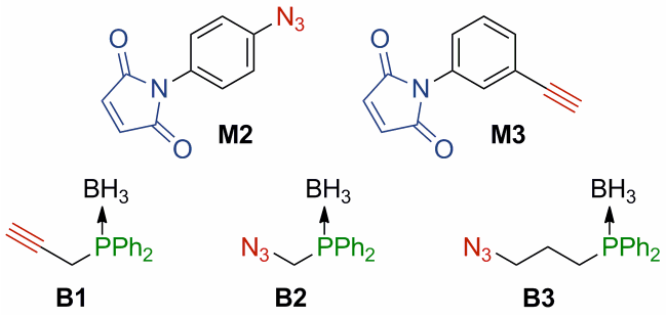
3.3.2. Modification using two-step conjugation procedures

3.3.2.1 Advantages of a two-step modification procedure

Because most common single step modification procedures for bioconjugation to a unique cysteine[12] proved incompatible or ineffective for introduction of phosphines we started looking for possible modification procedures employing two-steps. In a two step procedure a functionality can be introduced that can be specifically modified in a second step with a phosphine moiety. This gives the opportunity to select a method for the second modification step that is compatible with phosphines. For the first step standard cysteine bioconjugation techniques can be employed to introduce a new functional group that will be selectively modified in the subsequent step. It is important that both reaction steps are "click" type reactions where no side products are formed.

3.3.2.2 Huisgen copper catalysed 1,3-dipolar cycloaddition

In our group we did try to employ a two step modification process using modification by the copper catalysed Huisgen 1,3-dipolar cycloadditions. Maleimides **M2**[23] and **M3**[24] (Scheme 2) were easily introduced into SCP-2L V83C using standard conditions for maleimide modification described in section 2.3.3.4.



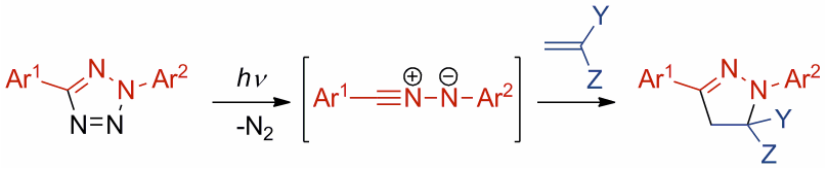
Scheme 2. Maleimide azide and alkyne cross linkers **M2** and **M3** for subsequent copper catalysed Huisgen 1,3-dipolar cycloaddition reactions at the azide or alkyne (red) to azide or alkyne containing phosphines **K1**-**K3**.

However, the subsequent copper catalysed reaction with the protein conjugates using borane protected phosphines **K1**-**K3** was unsuccessful.[13b] This reaction had the

additional disadvantage of the incompatibility of phosphines with azides. Because of this the phosphine has to be introduced as a borane protected phosphine.[25] In our group we found that general procedures for phosphine deprotection on the phosphine were ineffective.[13b] These results demonstrated that a second reaction compatible with free phosphines is essential, thus we refined our search for a suitable modification procedure to such reactions.

3.3.2.3 UV mediated 1,3-dipolar cycloaddition

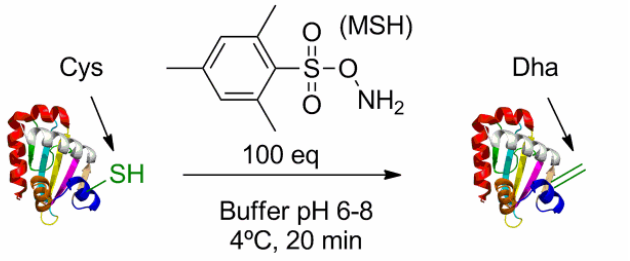
Recently a novel method for the site-selective modification of proteins was reported,[26] which uses a 1,3-dipolar cycloaddition between an alkene and an *in situ* photogenerated nitrile imine (Scheme 3).[27] UV-light activates a tetrazole to form the nitrile imine intermediate after loss of dinitrogen. This intermediate reacts with alkenes to form the pyrazoline product. Lin *et al* reported bioconjugation of a range of compounds to proteins by introduction of either the alkene or the tetrazole into the protein using standard conjugation procedures and a subsequent photo induced 1,3-dipolar cycloaddition. [26, 28] An additional advantage of the reaction is that the formed pyrazoline products are fluorescent, which allows for easy identification on SDS-PAGE when subjected to UV-light. The mild conditions and the large reported functional group compatibility of this reaction prompted us to investigate the possibility to use this reaction for our purpose of introducing phosphines in to proteins. In order to apply this reaction to our protein templates we set out to introduce an alkene functional group into our proteins.



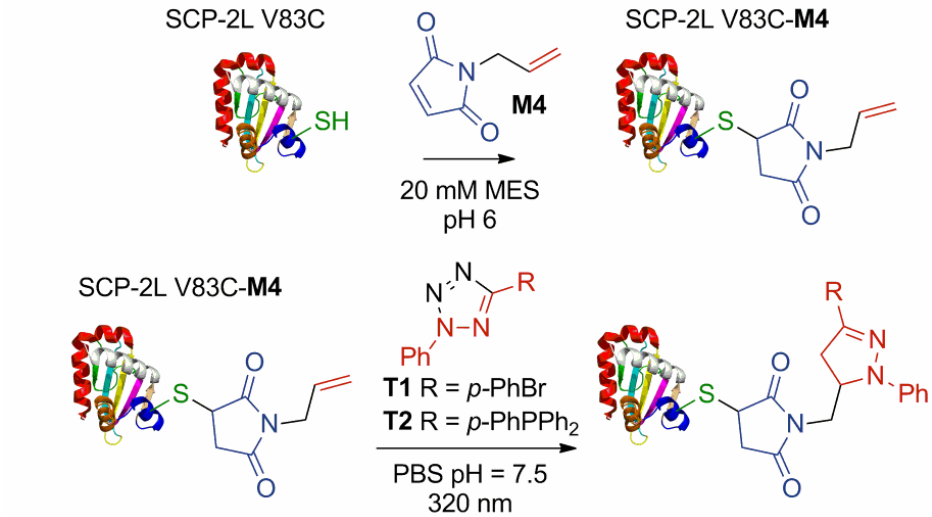
Scheme 3. Overview of the photoinducible 1,3-dipolar cycloaddition between tetrazoles and alkenes.

The first procedure we investigated for introducing an alkene into our unique cysteine containing protein was the procedure using *O*-mesitylenesulfonylhydroxylamine (MSH) to convert cysteine (Cys) into dehydroalanine (Dha) as reported by the group of Davis (Scheme 4).[29] Several oxidative procedures for conversion of cysteine to Dha have been previously reported[30] but this procedure has the advantage that it does

not oxidize methionines, which can cause protein destabilization. In addition, this procedure allows for a very short linker between the protein and the phosphine, which is desired to be able to control the localization of the phosphine inside the protein. However, after the incubation of SCP-2L V83C with 50 to 200 equivalents of MSH no change to the protein was observed upon LCMS(E.S^{+}) even after an overnight reaction at room temperature. Reactions at different pH (7.5 and 8.5) and in MES (pH 6.5) buffer gave no indication of any product formation either. Also for other proteins used in our group this reaction was found to be ineffective. The exposed nature of the cysteine in the protein used by Davis could play an important role in the reaction. This makes this procedure unsuitable as in all proteins used in our studies the cysteines are located in hydrophobic cavities.



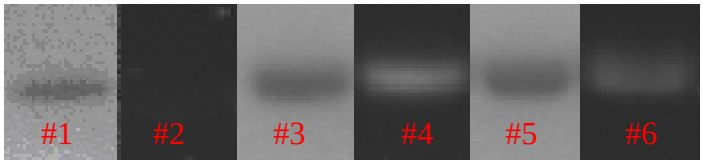
Scheme 4. Selective formation of dehydroalanine (Dha) from cysteine using *O*-mesitylenesulfonylhydroxylamine (MSH).



Scheme 5. Modification of SCP-2L V83C via alkene containing maleimide **M4** followed by photoinduced 1,3-dipolar cycloaddition of tetrazoles **T1** and **T2** with the obtained SCP-2L V83C-**M4**.

An alternative approach for the introduction of an alkene involved the introduction of allyl-maleimide **M4**, synthesised in our group.[31] Using standard conditions for

bioconjugation of maleimides to cysteine full conversion to the expected modification product was observed by mass spectroscopy (found 13513.3 Da, calculated 13513.7 Da) (Scheme 5). As a model reaction for modification of the alkene modified protein the 1,3-dipolar cycloaddition was performed using phenyl bromide substituted tetrazole **T1**,[32] which served as a precursor in the synthesis of **T2**.



Lane #	Contents	Analysis
1	SCP-2L V83C-**M4**	Gel stain
2	SCP-2L V83C-**M4**	UV-light
3	SCP-2L V83C-**M4** + **T1**	Gel stain
4	SCP-2L V83C-**M4** + **T1**	UV-light
5	SCP-2L V83C-**M4** + **T2**	Gel stain
6	SCP-2L V83C-**M4** + **T2**	UV-light

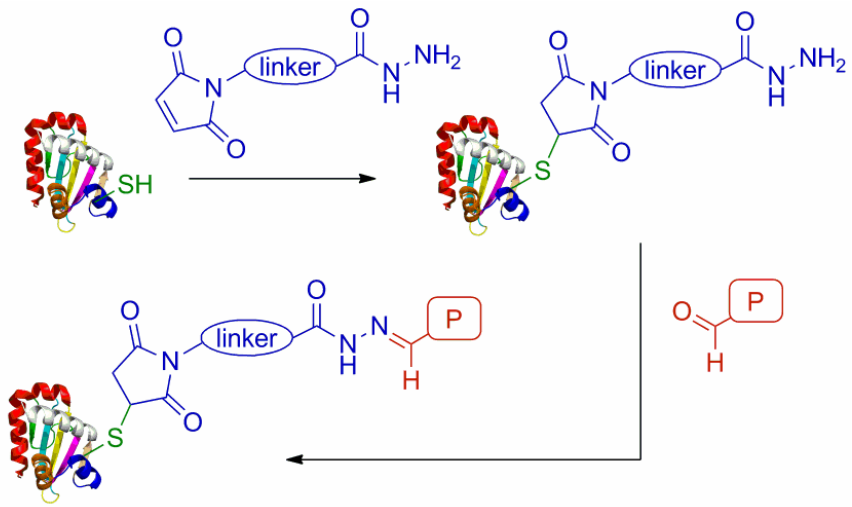
Figure 2. The SDS-PAGE analysis of the reaction between modified SCP-2L V83C-**M4** and tetrazoles **T1** and **T2** using gel staining and UV-light.

An overnight reaction with modified SCP-2L V83C-**M4** and tetrazole **T1** with excitation at 320, gave rise to an emission band at 450 nm. This emission band was assigned to the product using literature data.[26] The intensity of this band increased slowly over 6 hours after which reaction progress seemed to have ceased. Mass-spectroscopy after the reaction did show a signal with the expected mass for the pyrazoline product (calculated *m/z* 13785.6; found *m/z* 13783.2), however, this signal was very weak. SDS-PAGE gel did show a fluorescent band at the position of the protein band when exposed to UV-light using a hand-held UV-lamp, indicating modification of the protein (Figure 2 lanes 1-4). Performing the reaction with the phosphine modified tetrazole **T2** also gave rise to an emission band at 450 nm. Although, analysis of SDS-PAGE confirmed the presence of the pyrazole modified protein, unfortunatly no signal for the modified protein was found by mass-spectroscopy. SDS-PAGE gel showed that the intensity of the fluorescent band for the reaction with **T2** (Figure 2 lane 6) is lower than that obtained for the reaction with **T1** (Figure 2 lane 4).

In summary, it appears that the protein can be modified with the phosphine modified tetrazole, but the conversion was low. Attempts to improve the conversion by addition of more tetrazole only resulted in increased protein precipitation and formation of a product with an emission band at 370 nm, which is attributed to the product of the reaction between the nitrile imine intermediate and water, as reported in literature.[26] It seems that modification of the protein can be achieved using this approach applying a 1,3-dipolar cycloaddition reaction, however the low conversion and formation of side products and protein precipitate render this procedure unsuitable for the formation of phosphine modified proteins.

3.3.3 Introduction of phosphines via hydrazone formation between hydrazide modified proteins and phosphine aldehydes

Another reported efficient method for bioconjugation is hydrazone linkage between a hydrazide and an aldehyde.[33] This reaction is often used for protein cross-linking and also for aldehyde detection.[14, 34] Importantly, this reaction has been reported to be compatible with phosphines.[35] The syntheses of various phosphine aldehydes have been reported and several are even commercially available.



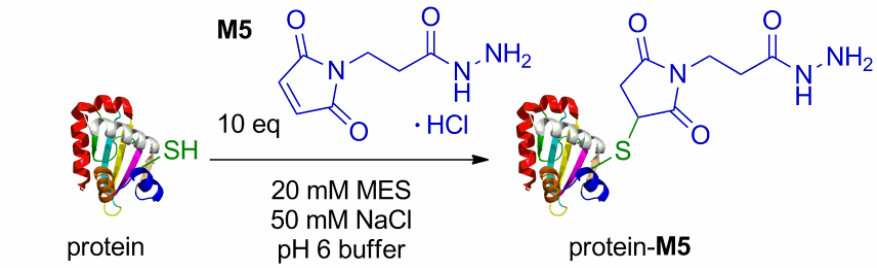
Scheme 6. Modification of SCP-2L through hydrazone formation and subsequent hydrazone formation using phosphine aldehydes.

To exploit this reaction using phosphine aldehydes the hydrazide has to be introduced into the protein. Maleimide-hydrazides are often used as cross-linkers and several are commercially available. This allows for the approach outlined in Scheme 6. In this approach the hydrazide is site selectively introduced by modification of the unique

cysteine by maleimide addition, which was shown before to be very efficient. In a second step the phosphine aldehyde is reacted with the hydrazide to obtain the phosphine modified protein.

*3.3.3.1 Modification of proteins using hydrazide-maleimide cross linker **M5***

The first step of the modification sequence was performed using 3-maleimidopropionic acid hydrazide hydrochloride **M5**, a commercially available cross linker often used for bioconjugation (Scheme 7).[33, 34b] Full conversion was observed upon LCMS(ES^+) analysis of the reaction mixture obtained after overnight mixing of a solution of SCP-2L V83C with 10 equivalents **M5** (see Figure 4 for an a representative mass-spectrum). Complete modification could be achieved with as little as three equivalents of the maleimide **M5** and reaction times of several hours on nmol scale. However it was found that especially at large scale the reaction should be monitored by mass-spectroscopy as reaction times can be long (up to three days at a 5 µmol scale). The reaction rate could be greatly increased by using up to 20 equivalents of **M5**. As for the model maleimide reactions described in chapter 2 section 2.4.3.4, the modification of the cysteine was confirmed using Ellman's reagent followed by analysis with mass spectroscopy (ES^+).[36] The reaction proved to be equally effective for the SCP-2L A100C mutant providing quantitative modification to the hydrazide modified proteins.



Scheme 7. Modification of unique cysteine bearing SCP-2L templates V83C and A100C with 3-maleimidopropionic acid hydrazide hydrochloride **M5**.

Th effect of the modification of SCP-2L A100C and V83C with **M5** on the folding of the protein was investigated by CD spectroscopy (Figure 3). The near UV CD spectra of the A100C mutant before and after modification appear to be nearly identical, showing that the modification does not have a significant influence on the protein fold. The spectrum of the V83C mutant after modification does show a slight increase

of the signals around 287 and 294 nm, meaning that some residues that cause negative signals in that area are affected by the modification. Also a slight rise in the intensity of the signal from 250 to 260 nm can be observed upon modification of SCP-2L V83C with **M5**. This signal increase can be caused by a partial loss of the α-helical structure of which the cysteine at position 83 is a part. The α-helical structures give a strong negative signal in the area of 190-240 nm (explained in more detail in section 2.4.3). Overall the major part of the structure seems to be retained indicating that the modification of the SCP-2L mutants with **M5** does not cause the protein fold to change significantly. This can be explained by the hydrophilicity of the ligand causing it to point outward into the solution and consequently not disturbing the inner structure of the protein.

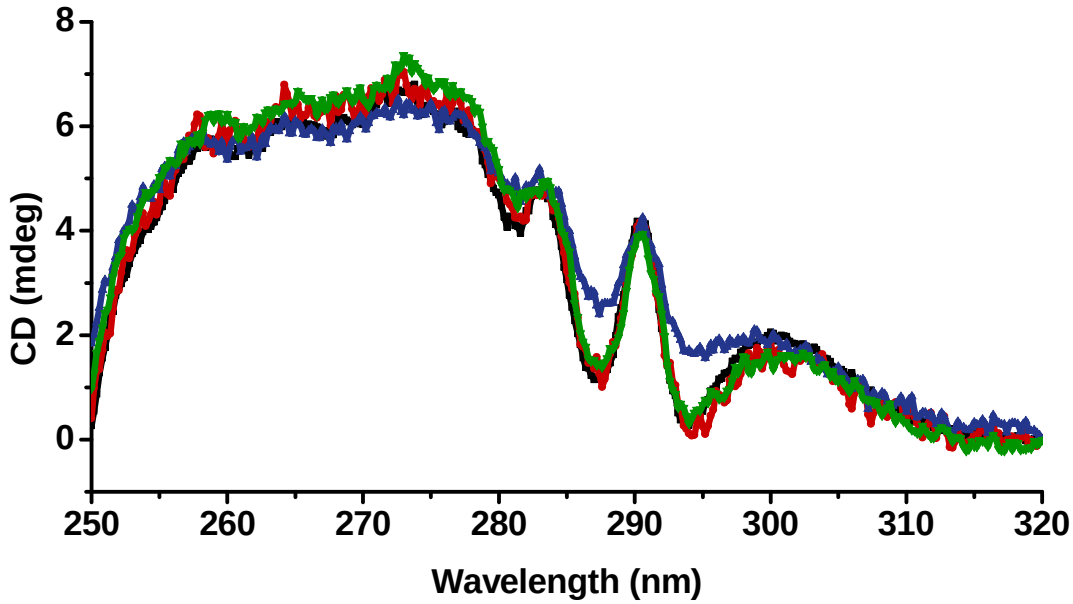
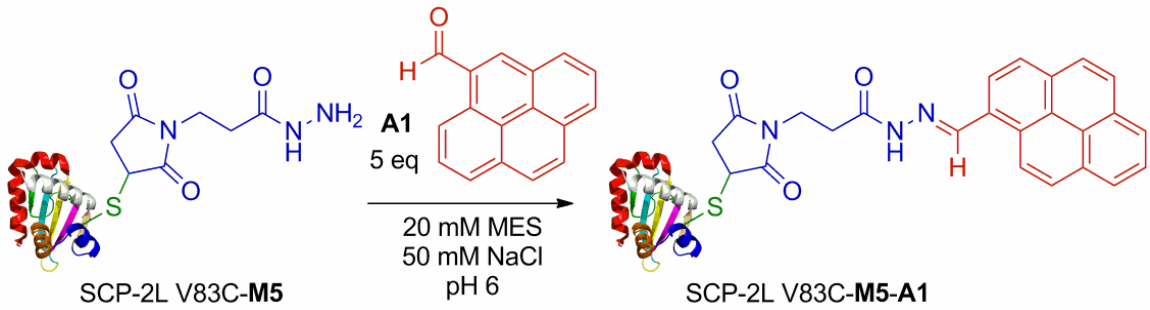


Figure 3. Near UV CD spectra of SCP-2L V83C (black), SCP-2L A100C (red), SCP-2L V83C-**M5** (blue) and SCP-2L A100C-**M5** (green).

Hydrazide modified SCP-2L V83C (SCP-2L-V83C-**M5**) was reacted with fluorescent 1-pyrenecarboxaldehyde **A1** to test the second step of the modification procedure and be able to quantify the product formation using fluorescence (Scheme 8). Although **A1** is insoluble in water, analysis of the soluble fraction, obtained after centrifugation of the modification reaction, by LCMS(ES^+) did show almost full conversion to the hydrazone linked adduct (expected 13771.7 Da, found 13772.1 Da). The signal of the hydrazide modified protein (SCP-2L V83C-**M5**) was also still present. The overall signal to noise ratio was very low which was attributed to the considerable amount of

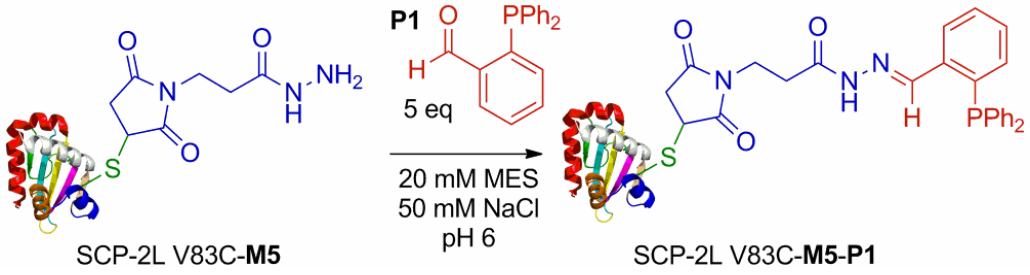
protein lost by precipitation. The formation of the precipitate hindered quantification of the reaction progress using fluorescence. Nevertheless, the coupling reaction was successful which showed promise for the modification with phosphine-aldehydes.



Scheme 8. Modification of SCP-2L V83C-**M5** with 1-pyrenecarboxaldehyde **A1**.

*3.3.3.2 Modification of SCP-2L V83C-**M5** using a commercially available phosphine aldehyde **P1***

A test reaction was also performed with the commercially available 2-diphosphino benzaldehyde **P1**[37] using SCP-2L V83C-**M5** (Scheme 9). This phosphine is not water soluble and was added to the protein solution from a concentrated solution in DMF to obtain a suspension. Full conversion to the phosphine modified protein was achieved after overnight mixing (LCMS (ES^+) analysis, expected 13831.74 Da, found 13832.06 Da, Figure 14).



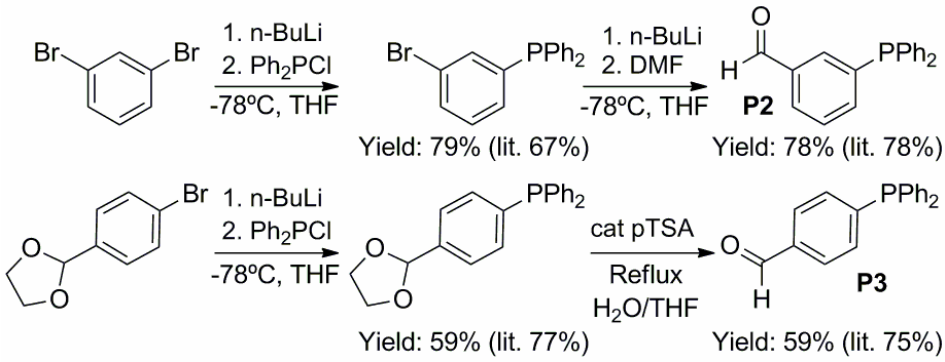
Scheme 9. Conjugation of 2-diphenylphosphino benzaldehyde **P1** to SCP-2L V83C-**M5**.

The obtained signal had a very good signal to noise ratio and only a small amount of protein precipitate was observed. The main signal observed corresponded to the modification with the free phosphine even though no precautions were taken to exclude oxygen. An added advantage of this method was that the excess phosphine could be readily removed by centrifugation due to the insolubility of the phosphine in

water. This result showed the first and only very efficient covalent phosphine modification reaction for the created unique cysteine SCP-2L template so far.

3.3.3.3 Synthesis of phosphine aldehydes

To get further insight into the general applicability of the hydrazone methodology to synthesise phosphine modified proteins for the development of artificial metalloenzymes, we synthesised additional phosphinoaldehydes. The synthesis of 3-diphenylphosphino benzaldehyde **P2**[38] and 4-diphenylphosphino benzaldehyde **P3**[39] was successful using procedures reported in literature outlined in Scheme 10. The meta substituted diphenylphosphino benzaldehyde was synthesised by selective monolithiation of one bromide of 1,3-dibromobenzene and quenching with chlorodiphenylphosphine to introduce the phosphine and reaction with DMF after a second lithiation to introduce the aldehyde.[38] Both steps proceeded with good yields, but unfortunately **P2** was obtained as a yellow oil which proved hard to handle because of sensitivity to oxygen. The para substituted diphenylphosphino benzaldehyde was synthesised through an elegant procedure involving substitution of 2-(4-bromophenyl)-1,3-dioxolane with chlorodiphenylphosphine by lithiation followed by acid catalysed deprotection of the ketal to the aldehyde.[39] The products of the first and second steps could be crystallized and were obtained in reasonable yields. The phosphine **P3** was obtained as a white crystalline solid which could be used and stored under aerobic conditions which make them much easier to handle compared to **P2**.



Scheme 10. Synthetic strategy for the synthesis of phosphine aldehydes **P2**[38]and **P3**.[39]

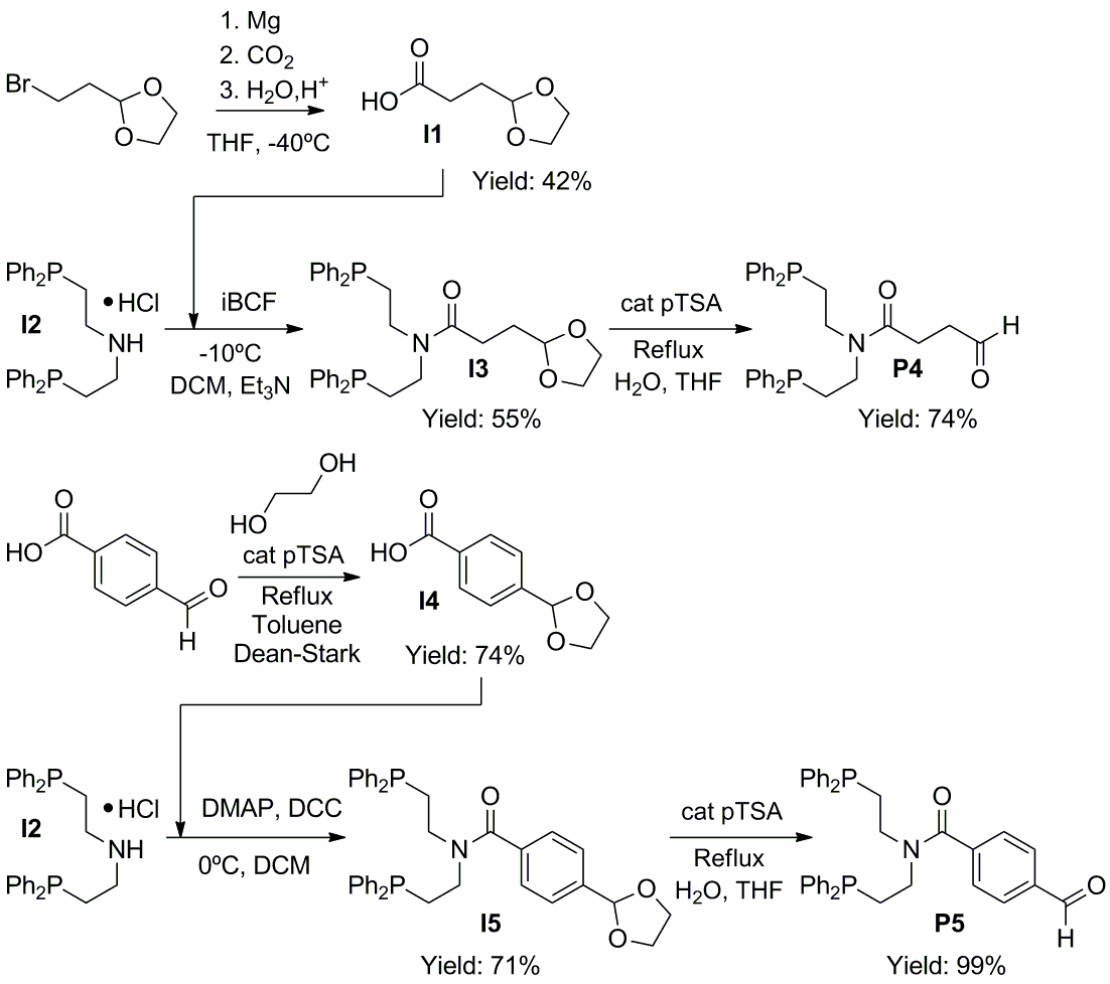
Inspired by the work of Whitesides[40] and others[5a, 6-7] two aldehyde variants **P4** and **P5** containing the bis[2-(diphenylphosphino)ethyl]amine motif were synthesised. This

motif was shown to be very effective as it is very flexible and can easily be influenced by the protein environment to potentially give good enantio selectivity in asymmetric catalysis.[4, 6] The two aldehydes containing diphosphine **P4** and **P5** were synthesised using the strategies as outlined in Scheme 11.

The diphosphine variant containing the aliphatic aldehyde linker **P4** was synthesised by coupling the protected aldehyde 3-(2-dioxolanyl)propionic acid **I1** to the secondary amine of the bis[2-(diphenylphosphino)ethyl]amine **I2** motif which was synthesised through a literature procedure by the reaction of *N*,*N*-bischloroethylamine and lithiodiphenylphosphine.[40-41] 3-(2-dioxolanyl)propionic acid **I1** was also synthesised through a literature procedure involving quenching of the Grignard of 2-(2-bromoethyl)-1,3-dioxolane with crushed dry ice. The product was obtained in very pure form by using a vacuum distillation providing **I1** as crystals with a melting point close to room temperature (see appendix A5.1 for crystal structure). The coupling reaction using isobutylchloroformate (iBCF) to form protected phosphine aldehyde **I3** gave the product in a moderate yield. During column chromatography the main side product was isolated and found to be the bis[2-(diphenylphosphino)ethyl]amine modified at the amine with an isobutylformate group (see appendix A6 for NMR data). It appeared that the low yield is caused either by the similar reactivity of the two carbonyls in the anhydride intermediate or by incomplete conversion of the **I3** to the anhydride, which caused the isobutylchloroformate to react directly with bis[2-(diphenylphosphino)ethyl]amine. The use of other reagents for the coupling reaction like dicyclohexylcarbodiimide (DCC) should circumvent this problem and give higher product yield. The deprotection of **I3** gave the desired phosphine aldehyde **P4** in good yield.

Another diphosphine aldehyde derivative based on bis[2-(diphenylphosphino)ethyl] amine containing a benzaldehyde linker **P5** was synthesised. This compound was also synthesised via coupling of an acid containing a protected aldehyde to the bis[2-(diphenylphosphino)ethyl]amine **I2** motif followed by removal of the ketal to form the aldehyde. The protected aldehyde 4-(1,3-dioxolan-2-yl)benzoic acid **I3** was synthesised from 4-formylbenzoic acid by mixing with ethylene glycol in a Dean-Stark setup according to a literature procedure.[42] A short reaction time and the saponification before the work-up were found to be essential to get the product pure and in high yield. For the amide formation reaction with **I2** DCC was used to get the protected diphosphine aldehyde **I4** in good yield. Deprotection of the ketal was very

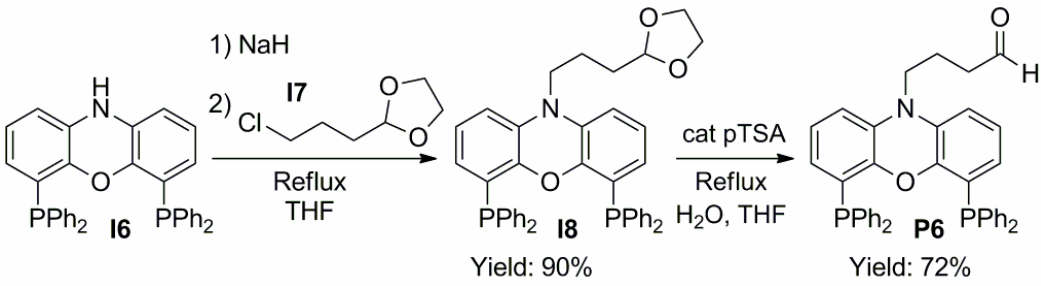
efficient for this compound providing the aldehyde in quantitative yield. Both diphosphine aldehydes **P2** and **P3** were obtained as a white gum which is very oxygen sensitive and needed to be stored under inert atmosphere.



Scheme 11. Synthesis of aldehyde phosphines **P4** and **P5**.

Nixantphos[43] aldehyde derivative **P6** was also synthesised. Wide bite angle xantphos derivatives are well known and successful phosphine ligands for reactions like hydroformylation[43-44], allylic substitution[45] and Buchwald Hartwig amination.[46] This diphosphine ligand motif is very rigid, which means that unlike diphosphine aldehydes containing bis[2-(diphenylphosphino)ethyl]amine **P4** and **P5**, little influence of the protein on the native selectivity of **P6** in catalytic reactions is expected. **P6** was synthesised by introduction of a protected aliphatic aldehyde through modification of the secondary amine of commercially available 4,6-bis(diphenylphosphino)phenoxazine (nixantphos) **I6** with commercially available 2-(3-chloropropyl)-1,3-dioxolane **I7** (Scheme 12). The secondary amine of nixantphos

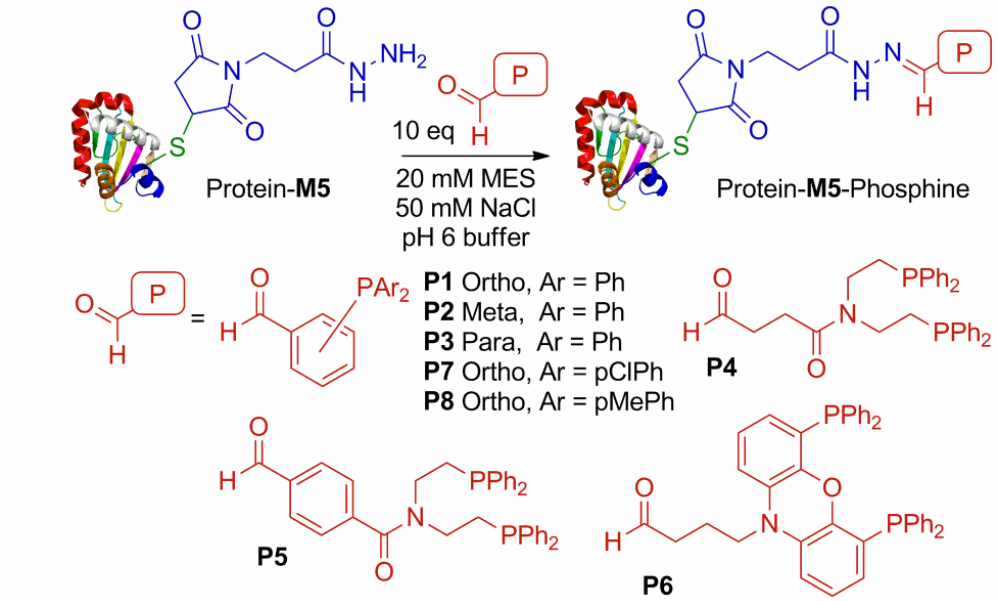
can not easily be functionalised using carboxylic acids due to the weak nucleophilic character. However it has been reported that the nitrogen can be substituted by deprotonation followed by addition to alkylchlorides.[43-44] Deprotonation with sodium hydride and subsequent reaction with 2-(3-chloropropyl)-1,3-dioxolane **I7** was achieved in excellent yield following this procedure. The deprotection of the ketal **I8** (see appendix A5.2 for crystal structure) to the aldehyde diphosphine ligand **P6** was also successful however as for the above ligand containing the aliphatic aldehyde **P4** it proved not very high yielding.



Scheme 12. Synthesis of aldehyde containing nixantphos derivative **P6**.

3.3.3.4 Modification of hydrazide modified proteins using phosphino aldehydes

The successful test reaction between **P1** and SCP-2L V83C-**M5** was extended to SCP-2L A100C-**M5** and various monophosphine benzaldehydes (Figure 4, Table 3, entries 1 - 10).



Scheme 13. Conjugation of mono- and diphosphino aldehydes to **M5** modified protein templates.

Table 3. Results of bioconjugation of hydrazide modified proteins (Protein-**M5**) with aldehydes **P1**-**P8**.[a]

Entry	Protein–**M5**	Phosphine	Calculated mass (Da)	Observed mass (Da)[b]
1	V83C	**P1**	13830.2	13830.4
2	V83C	**P2**	13830.2	13830.8
3	V83C	**P3**	13830.2	13829.9
4	V83C	**P7**	13898.1	13898.1
5	V83C	**P8**	15890.2	13906.0[c]
6	A100C	**P1**	13858.2	13857.4
7	A100C	**P2**	13858.2	13858.0
8	A100C	**P3**	13858.2	13858.1
9	A100C	**P7**	13926.1	13942.1[c]
10	A100C	**P8**	13918.2	13934.6[c]
11	V83C	**P4**	14065.3	14065.9[d]
12	V83C	**P5**	14114.9	14130.1[c]
13	V83C	**P6**	14161.8	14163.5[d]
14	A100C	**P4**	14093.3	14092.8[d]
15	A100C	**P5**	14142.9	14144.5
16	A100C	**P6**	14189.8	-[e]

[a] 10 equivalents phosphine **P1**-**P8** in 20 mM MES 50 mM NaCl buffer pH 6; full conversion of Protein-**M5** was observed unless stated otherwise [b] Main signal for modified protein [c] Main signal corresponds to the calculated mass for the protein containing oxidized phosphine [d] Protein-**M5** observed as main signal and only weak signal for phosphine modified Protein-**M5** was found.[e] no protein signal observed.

The modification reactions were performed using a small excess (up to 10 equivalents) of the phosphino aldehydes. A concentrated solution of the phosphino aldehyde was added to **M5**-modified protein under inert atmosphere to give a suspension. Quantitative conversion to the corresponding phosphine modified proteins was typically observed by LCMS(ES^+) analysis of the soluble fraction after overnight mixing. In addition to the phosphino aldehydes **P1**-**P3**, the more electron rich phosphino aldehyde **P7** and electron poor phosphino aldehyde **P8** were successfully coupled to both SCP-2L mutants (Table 3, entries 4, 5, 9, 10). For all

phosphines the excess of insoluble phosphine aldehyde ligands was easily removed by centrifugation to allow further use of the phosphine modified protein. Generally the protein solution was also washed with degassed buffer in a centrifugal concentrator to remove potential trace amounts of dissolved ligand. The yield of modified protein was typically around 90% according to analysis of the modification reactions by Bradford assays. However when a large excess phosphine (over 20 equivalents) is added or if a protein concentration of over 3 mg/ml is used the yields are typically lower due to protein precipitation.

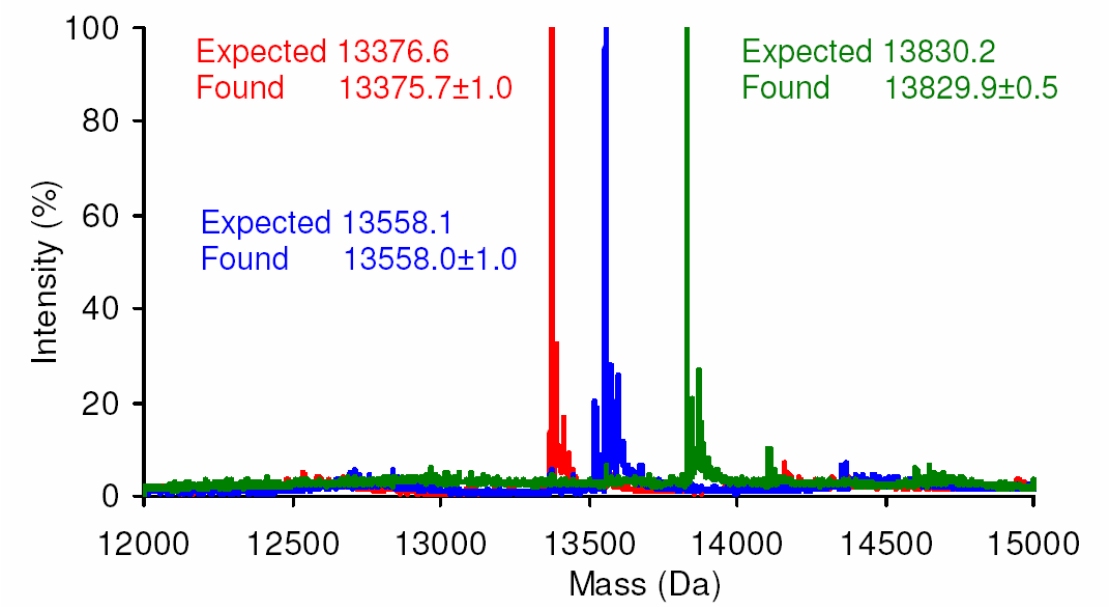


Figure 4. Processed mass-spectra (ES^+) of SCP-2L V83C (Red), SCP-2L V83C-**M5** (Blue) and SCP-2L V83C-**M5-P3** (Green) (Copyright Wiley-VCH Verlag GmbH & Co. KGaA. Reproduced with permission).[47]

An additional signal, which was 16 mass units higher than the modified protein signals shown in Table 3 was also observed in the mass-spectrum of several phosphine conjugates (Figure 4). Typically, this signal has an intensity of 20-40% compared to the signal corresponding to the phosphine modified protein. This signal corresponds to the mono oxidized protein, which indicates possible phosphine oxidation. However, a similar signal of +16 mass units is also observed in the spectra of the unmodified and the hydrazide modified proteins, which indicates that this signal might not be the result of a significant amount of phosphine oxidation, but most likely corresponds to methionine oxidation, a phenomenon often observed in mass spectroscopy of proteins. The phosphine oxide was found to be the most intense signal for reactions with phosphines **P7** and **P8.** These phosphines were most likely

oxidized by prolonged exposure of the sample for LCMS to air or oxidation in the ionization chamber.

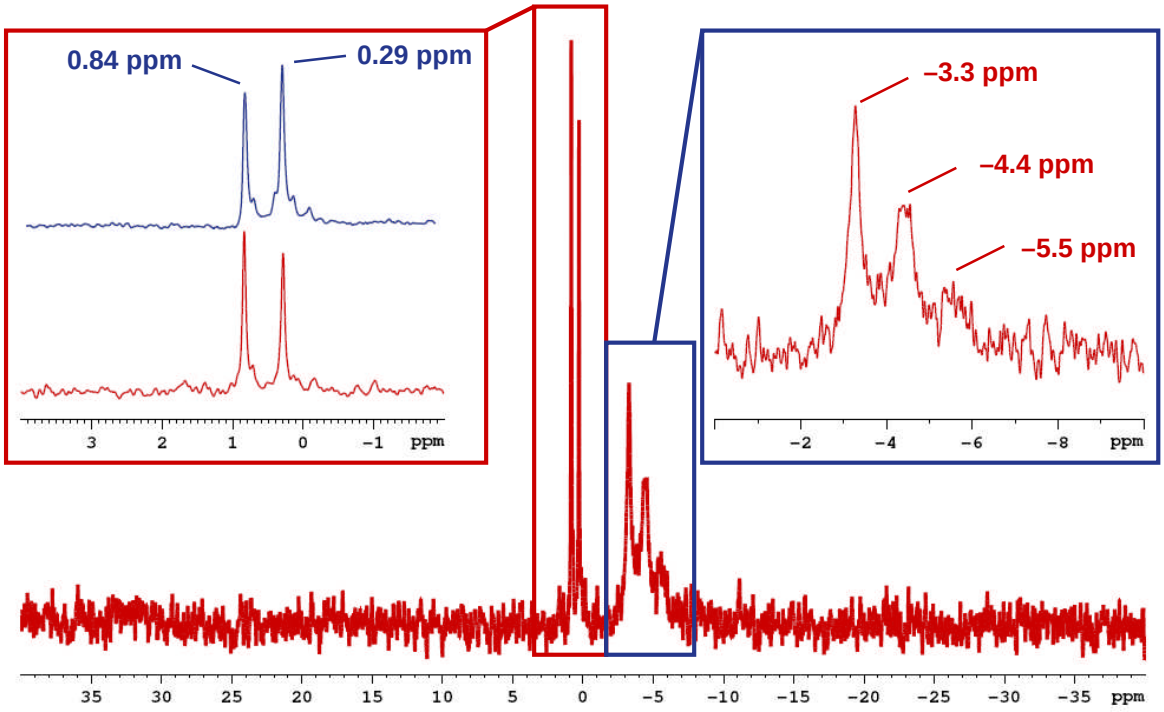
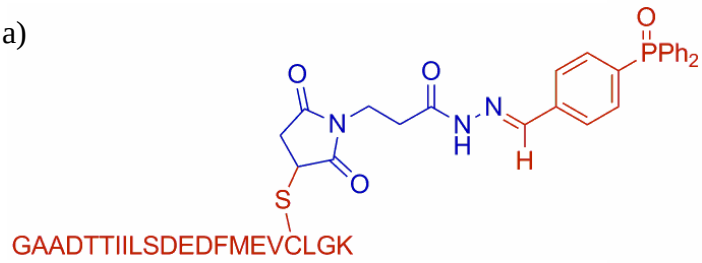


Figure 5. ^{31}P-{^{1}H}-NMR spectrum of SCP-2L V83C-**M5-P3** (red) with the red insert showing a magnification of the area between 4 and –2 ppm with a comparison to the ^{31}P-{^{1}H}-NMR spectrum of SCP-2L V83C (blue) and the blue insert showing a magnification of the area between –1 and –9 ppm.

The ^{31}P-NMR spectrum of SCP-2L V83C modified with **P3** shows broad signals in the range of –2 to –8 ppm (Figure 5). These are typical chemical shifts for a free tri-aryl phosphine, thus confirming that this modified protein contained a phosphine which was not oxidized. The signals appear to consist of at least two overlapping broad signals. This set of signals might result from the presence of different conformations of the conjugates, caused by either the imine bond, the chiral centre formed by the maleimide-sulfide bond or the conformation of the protein. The possibility that the different signals originate from phosphine modification of different amino acid side chains of the protein was ruled out by detailed analysis of trypsin digests, which showed only modification of the expected peptide fragment containing the cysteine (Figure 6a). The found mass of this peptide fragment corresponded to the modification of the peptide with phosphine-oxide analogue of **M5-P3**. MS-MS analysis of the peptide sequence of the found signal confirmed that this signal belonged to the expected peptide (Figure 6b). A search for other modified peptide

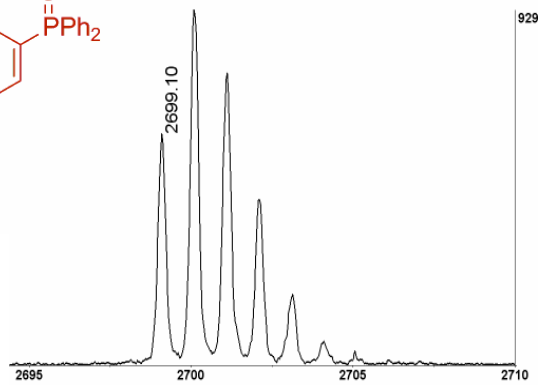
fragments did not provide any indication of any other modified amino acids. In addition, no protein modification was observed upon LCMS(ES^+) analysis of the reaction between unmodified unique cysteine SCP-2L mutants with the phosphine aldehydes.

a)



Expected masses for $C_{121}H_{179}N_{26}O_{38}PS_2$

m/z: 2700.21 (100.0%), 2701.21 (83.6%), 2699.21 (74.4%), 2702.22 (29.5%)



b)

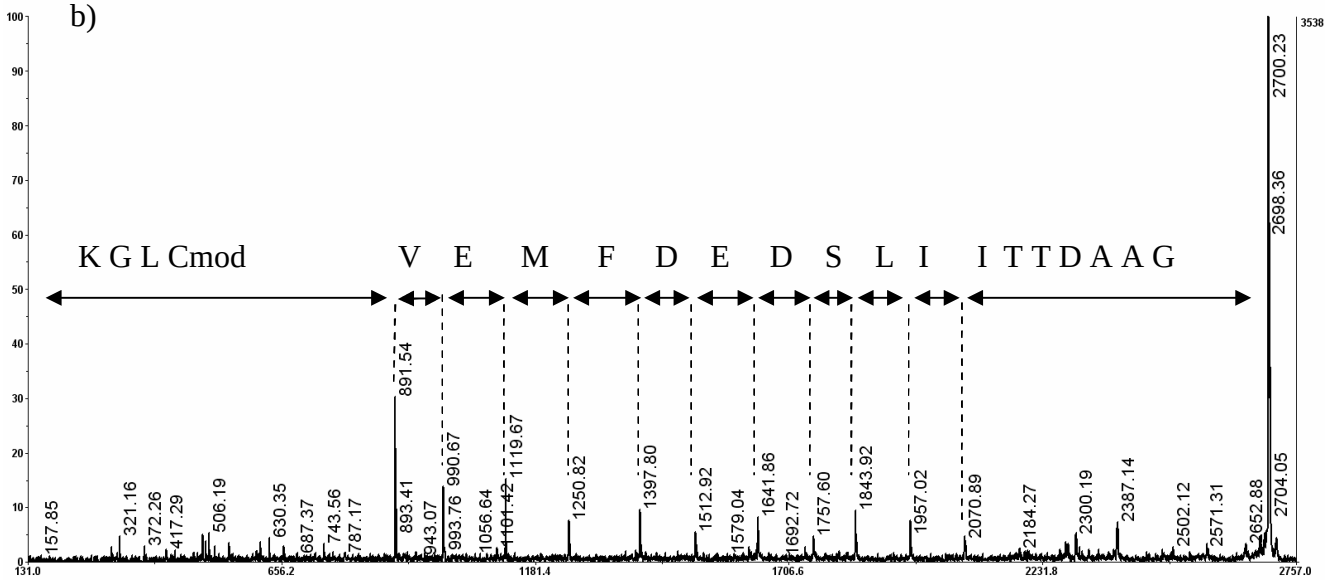


Figure 6. a) MALDI-MS spectrum displaying obtained signal for the expected modified peptide for digested SCP-2L V83C-**M5**-**P3** b) MS-MS of the signal at 2700.23 showing the expected peptide sequence (Cmod stands for cysteine modified with **M5**–**P3**).

The modification procedure shown in Scheme 13 was also performed using diphosphines **P4–P6** (Table 3, entries 11 - 16). Unfortunately, the reactions using diphosphines **P4** and **P6** resulted in substantial protein precipitation which is likely caused by instability of the formed conjugation product. The modification reaction using diphosphine **P5** containing the benzaldehyde did give the desired conjugation product. Although a small signal corresponding to the hydrazide modified protein was

observed the conversions were found to be >95% after correction of the ES^+ data for differences in ionization efficiencies (Table 3, entries 12 and 15). It appeared that the benzaldehyde containing phosphines were more suitable for conjugation than the aliphatic aldehydes as the coupling-products were less prone to precipitation. Analysis of the SCP-2L V83C-**M5-P5** by Maldi MS after trypsin digestion showed a signal corresponding to the expected modified peptide (Figure 7). The found signal corresponded to the peptide modified with an analogue of **M5-P5** containing two oxidized phosphines. MS-MS of this signal confirmed the expected peptide sequence, again demonstrating the site selectivity of the modification procedure (appendix A3).

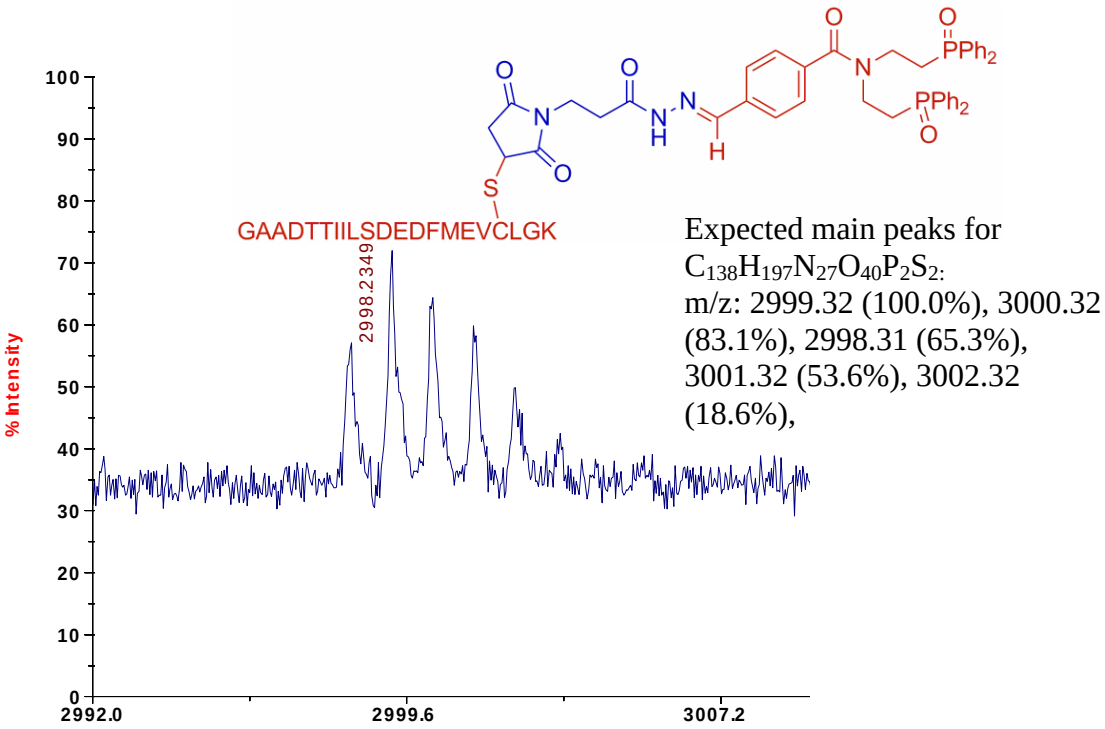


Figure 7. MALDI-MS analysis of trypsin-digested SCP-2L V83C-**M5-P5**, displaying obtained signal for the expected modified peptide for (MS-MS of the signal at 2999.32 showing the peptide sequence is depicted in appendix A3).

It was also demonstrated by our group that the hydrazone methodology described above is applicable to several other unique cysteine containing proteins.[13b, 47] The used proteins were structurally diverse, containing hydrophobic cavities of different sizes and shapes for different applications as artificial metalloenzymes in catalytic reactions. The hydrazone formation procedure proved successful for all these structurally diverse unique cysteine containing proteins encountering no problems with reactivity or selectivity. These results show that this modification procedure is

truly generally applicable and therefore realises the prospect of the use of virtually any unique cysteine bearing protein as phosphine containing artificial metalloenzymes.

3.3.3.5 Structural investigation

The effect of the phosphine modifications using **P3** and **P5** on the folding of the protein templates SCP-2L V83C and SCP-2L A100C was investigated by CD spectroscopy, watergate ^{1}H-NMR and by performing a substrate binding assay. The near-UV CD-spectra of SCP-2L V83C-**M5**-**P3** and SCP-2L V83C-**M5**-**P5** were recorded (Figure 8).

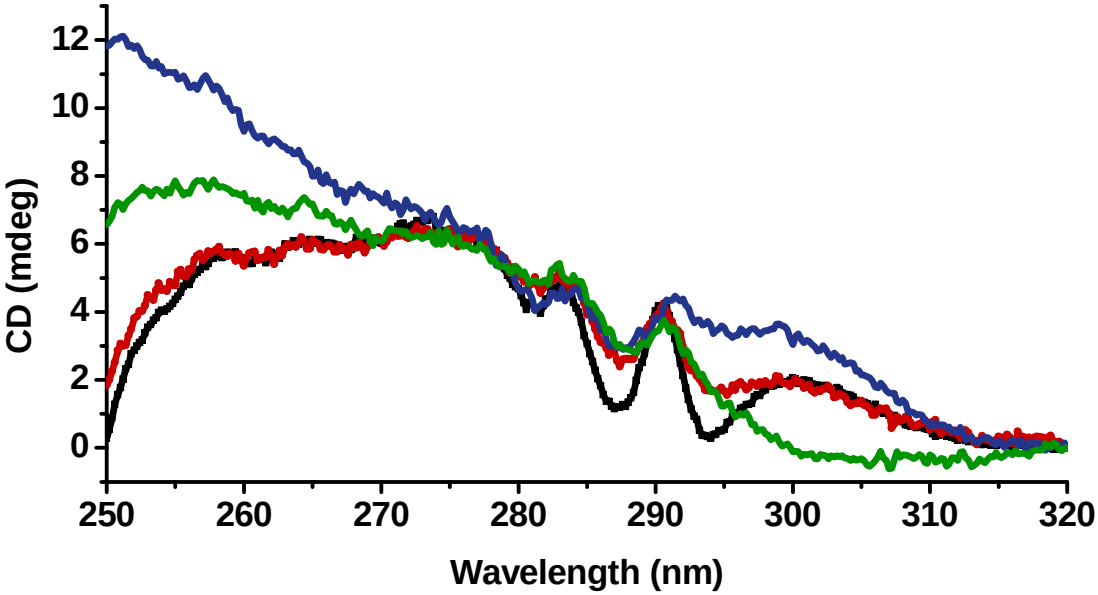


Figure 8. CD spectra of SCP-2L V83C (black), SCP-2L V83C-**M5** (red), SCP-2L V83C-**M5**-**P3** (blue) and SCP-2L V83C-**M5**-**P5** (green).

The CD spectrum of unmodified SCP-2L V83C and SCP-2L V83C-**M5** are also shown for comparison, which allows for the observation of distinct differences between this spectrum and those obtained for SCP-2L V83C-**M5** with **P3** and **P5**. Each modification seems to affect a different area of the CD spectrum. The modification with **P3** leads to an increase in the signal around 300 nm and 250-270 nm, whereas the modification with **P5** leads to a decrease of the signal in the 300 nm area and a similar but less intense increase of the signal around 250-270 nm. Interestingly the signal from 275 to 290 nm seems largely unaffected by the phosphine modification compared to the **M5**-modification indicating that a large part

of the structure is retained. The increase in the region from 250 to 270 nm can be attributed to an increase in the signal due to the introduced aromatic rings being placed in a chiral environment combined with a distortion in the α helical structures present in the protein causing an increase in the strong negative signal between 190-240 nm. We were unable to perform CD spectroscopy under an inert atmosphere making phosphine oxidation a likely event. This means that the observed spectra most likely belong at least partially to the proteins modified with oxidized phosphines, however oxidation of the ligands is expected to have little effect on the protein structure.

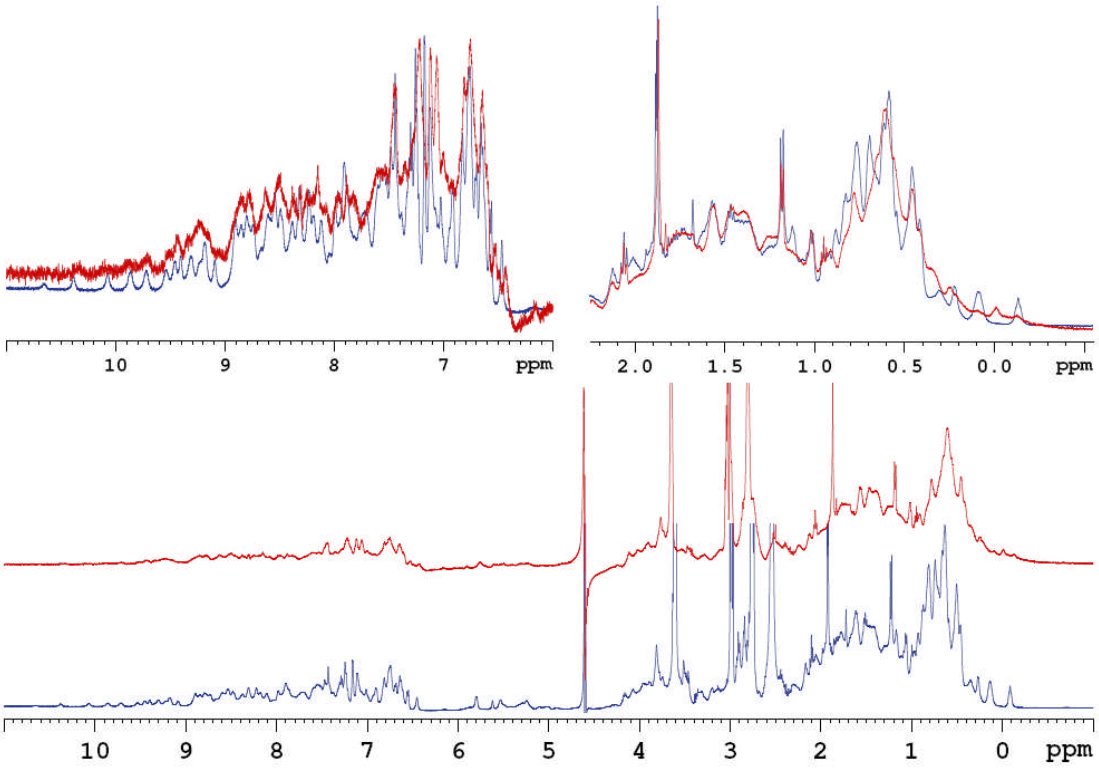


Figure 9. Watergate ^{1}H-NMR spectra of SCP-2L V83C (blue) and SCP-2L V83C-**M5-P3** (red) ~2 mM in 20 mM MES pH = 6 buffer a) overlay for the region between 11 and 6 ppm b) overlay for the region between 2.3 and –0.5 ppm c) full spectra, offset for clarity.

The watergate ^{1}H-NMR spectrum of SCP-2L V83C-**M5-P3** shows that coupling of the phosphine mainly results in broadening of the signals, but also induces several clear changes when compared to the spectrum obtained for unmodified SCP-2L V83C (Figure 9). Several signals around 0 ppm have shifted significantly indicating some

changes to aliphatic residues residing in the hydrophobic core of the protein, which is expected upon the introduction of the hydrophobic ligand.

Both the CD-spectrum and the watergate ^{1}H-NMR spectrum indicate some significant changes to the fold of SCP-2L V83C-**M5** after modification with the phosphines **P3** and **P5**. Changes to the structure of the protein are acceptable as long as the capability of the protein to bind the substrates for which the protein was selected is not significantly affected. The substrate binding capability of SCP-2L V83C-**M5** modified with **P3** and **P5** was determined using the fluorescent binding assay described in section 2.3.3.3. The measured K_d values for binding of Pyr-C12 were 0.24 (±0.06) μM for SCP-2L V83C-**M5**-**P3** and 0.36 (±0.06) μM for the SCP-2L V83C-**M5**-**P5**. The found values are of the same order of magnitude as the K_d of 0.18 (±0.07) μM found for unmodified SCP-2L V83C. This result shows that although the hydrophobic core of the protein is affected by the modification the substrate binding capability of the protein is not significantly affected by the introduction of the phosphine-ligands.

A similar but less thorough study was performed for SCP-2L A100C-**M5** modified with phosphines aldehydes **P3** or **P5**. The near UV CD spectra recorded of the SCP-2L A100C mutant before and after phosphine modification show far more intense changes to the overall signal compared to those recorded for the SCP-2L V83C conjugates (Figure 10). Distinct changes over the whole range from 250 to 300 nm can be observed for both **P3** and **P5** modifications as well as an increase in the signal at 287 nm and in the area of 294 to 320 nm. For both modifications there is a smaller increase at 250 nm compared to the modifications of the SCP-2L V83C mutant, which can be explained by the fact that the cysteine at position 100 is located on the β-sheet structure instead of on the α helix structure (see chapter 2 for more details).This means that a smaller effect on the helical structure of the protein can be expected which causes less distortion to the strong negative signal between 190 and 240 nm. The change in signal around 260 nm can be assigned to the introduced aromatic rings of the ligand being located in a chiral environment; however in this case giving rise to a negative signal as opposed the positive signal observed for the modified V83C mutant. This effect is more pronounced for ligand **P3** than for ligand **P5,** which indicates that **P3** is located in a different environment or that one diphenylphosphine moiety of **P5** might give a positive signal which cancels out the negative signal. The

overall signal seems to be more disturbed by the **P5** modification indicating more significant structural changes of the protein.

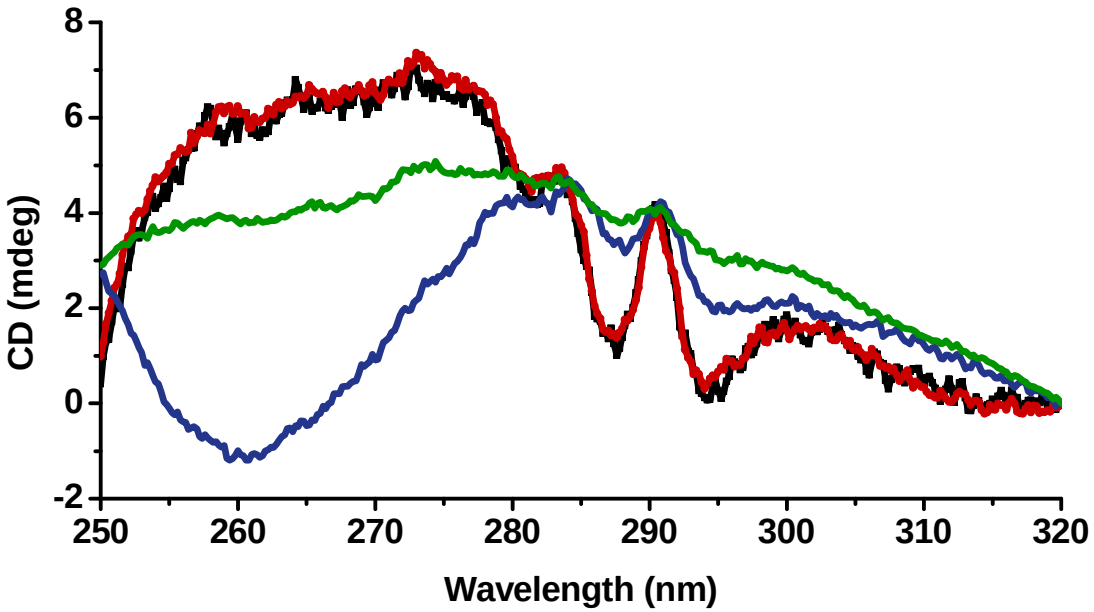


Figure 10. CD spectra of SCP-2L A100C (black), SCP-2L A100C-**M5** (red), SCP-2L-**M5**-**P3** (blue) and SCP-2L A100C-**M5**-**P5** (green)

Overall, the CD, ^{1}H-NMR and fluorescent binding assay data show that the fold of the proteins is definitely altered upon modification with the phosphine aldehydes. This can be expected as the hydrophobic ligand will most likely move inside the protein, changing the structure of the hydrophobic core of the protein. The CD spectra suggest that this effect is greater for the A100C mutant compared to the V83C. How this effects the stability of the protein would be an interesting subject for further research.

3.4 Conclusions and future prospects

This chapter shows a detailed exploration of several procedures for bioconjugation of functionalised phosphines to the unique cysteine containing templates based on SCP-2L. The cysteine-selective introduction of phosphines was found challenging due to the incompatibility of phosphines with standard cysteine selective modification reactions in combination with the inability to apply protected phosphines because of ineffectiveness of deprotection procedures with proteins. The additional selectivity problems of the CDI activated carboxylic acid methodology led to the development of two-step procedures to circumvent these issues. This was successful by application of cysteine-selective introduction of a hydrazide using maleimide introduction and

subsequent modification using phosphine aldehydes. This procedure was shown to be very selective and efficient. This modification strategy can also be applied to other unique cysteine bearing templates, as demonstrated by the modification of two different scaffolds with a variety of different phosphine-ligands. This methodology can now be applied for the modification of virtually any desired unique cysteine bearing protein with phosphines.

Further investigation is required into the formation of isomeric structures for SCP-2L V83C-phosphine conjugates as observed in the form of multiple signals in ^{31}P-NMR. This is important as these isomeric structures can lead to undesired mixtures of different catalytic species. Detailed structural investigations by NMR using isotopically labelled proteins or X-ray crystallography will be required. The results described in this chapter now allows for the exploration of transition metal complex formation using these phosphine containing proteins and their application in catalysis.

3.5 Experimental

3.5.1 General

3.5.1.1 Methods

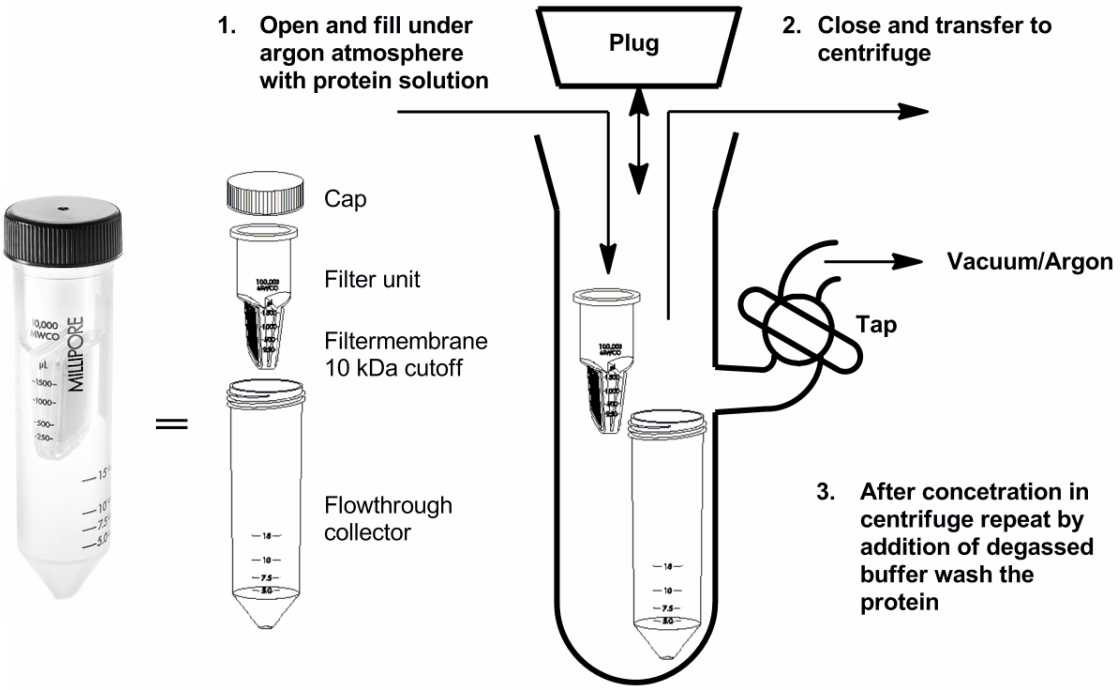


Figure 11. Application of Amicon® Ultra-15 Centrifugal Filter Unit (left) under inert atmosphere using a large Schlenk flask

All reactions involving phosphines and other oxygen sensitive compounds were performed under an argon atmosphere using degassed solvents and standard Schlenk techniques. The centrifugation and washings of protein solutions under an argon atmosphere were performed using a setup involving a large Schlenk flask (Figure 11). Binding constants of the fluorescent probe Pyr-C12 were determined using the method outlined in section 2.3.3.3. For processing of the mass-spectra the MaxEnt algorithm of Masslynx V4.0 (Micromass Ltd) was used. Protein concentrations were determined using Bradford's reagent.[48]

3.5.1.2 Equipment

NMR spectra were taken at room temperature with Bruker Avance 300, 400 or 500 NMR spectrometers. Chemical shifts (δ) are given in ppm for high-frequency shifts relative to a TMS reference (^{1}H and ^{13}C) or 85% H_3PO_4 (^{31}P). Watergate ^{1}H-NMR spectra were recorded on a Bruker Avance 500 spectrometer. 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 2.1x 10 mm cartridge using a gradient of 1% formic acid in H_2O to 1% formic acid in acetonitrile. Standard deviations were calculated manually from the raw mass data. Spin-filter concentrators with a molecular weight cut-off of 10kDa (Amicon® Ultra-15 Centrifugal Filter Units, Millipore®, Carrigtwohill, Co. Cork, Ireland) were used to concentrate and/or wash protein solutions. CD spectra were recorded using a Jasco J-810 spectropolarimeter using a 1 mg/ml solution of the protein. Elemental analysis was performed by Stephen Boyer at the elemental analysis service of the London Metropolitan University or by Sylvia Williamson at the University of St. Andrews. HRMS (TOF MS CI^+) were recorded on the University of St. Andrews by Caroline Horsburgh.

3.5.1.3 Materials

Chemicals were purchased in the highest commercially available purity unless stated otherwise. 3-Maleimidopropionic acid hydrazide hydrochloride **M5** and 2-(diphenylphosphino)benzaldehyde **P1** were obtained from Apollo Scientific Ltd and Aldrich respectively. Dry DMF was purchased from Acros in a sealed bottle

containing 4Å molecular sieves and was used as received. For dry conditions freshly distilled degassed solvents were used. Dichloromethane (DCM) was distilled from CaH_2, triethylamine (Et_3N) from lithium aluminium hydride, tetrahydrofuran (THF) and diethylether (Et_2O) were distilled from sodium/benzophenone and ethanol (EtOH) was distilled from Mg/I_2. Thin-layer chromatography (TLC) was performed on silica (Polygram 0.2 mm silica gel with fluorescent indicator UV_{254}) or alumina (aluminium oxide 60 neutral with fluorescent indicator UV_{254}) plates purchased from Merck. Silica gel (60 particle size 0.063-0.2 mm) and aluminium oxide (90 active neutral) purchased from Merck and Fluka were used for column chromatography. Phosphine carboxylic acids **C1** and **C3** were obtained from commercial sources. Compounds **C5**–**C7** were synthesised by Bianca K. Muñoz and compounds **C2** and **C8** by René den Heeten. Maleimide **M4**[31], **T1** and **T2** were synthesised by Gregorio Guisado-Barrios from our group. The phosphine aldehydes 3-(diphenylphosphino)benzaldehyde **P2**[39], 4-(diphenylphosphino)benzaldehyde **P3**[38], *N*-(4-azidophenyl)maleimide **M2**[23] and *N*-(3-acetylenephenyl)maleimide **M3**[24] were synthesised according to literature procedures. Phosphine aldehydes **P7** and **P8** were kindly donated Piotr Wawrzyniak of the group of Dr. Matt Clark (University of St. Andrews).

3.5.2 Modification using CDI activation of carboxylic acid phosphines

*3.6.2.1 CDI activation of carboxylic acid phosphines (**C1**-**C8**)*

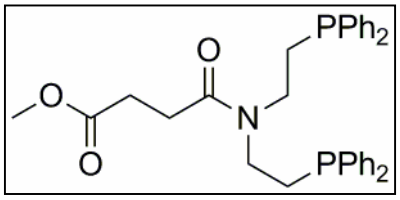
This procedure is based on a previously reported procedure [20b] 5-10 Equivalents of CDI was added as a solid to a 25 mM solution of the corresponding carboxylic acid phosphine in DMF under an argon atmosphere. The reaction was stirred overnight for 16 h, after which the conversion was determined by ^{31}P-NMR using a D_2O locking insert. The reaction-mixture was used without purification for the bioconjugation to SCP-2L V83C.

*3.5.2.2 Bioconjugation of CDI activated carboxylic acid phosphines (**C1**-**C8**) to SCP-2L V83C*

This procedure is based on a procedure reported previously by our group.[20b] A 15 μM solution of SCP-2L V83C in buffer (20 mM HEPES pH 7) was incubated with 5-20 equivalents of the CDI activated carboxylic acid phosphine added from a 25 mM

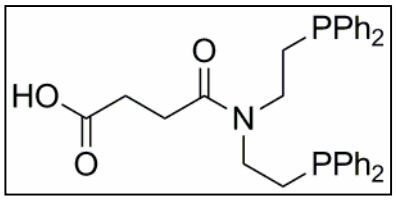
stock solution in DMF. Progress of the reaction was monitored using LCMS (ES^+) analysis of the soluble fraction of the reaction obtained by centrifugation and removal of the precipitate.

*3.5.2.3 Synthesis of methyl 4-(bis(2-(diphenylphosphino)ethyl)amino)-4-oxobutanoate **I9***



This procedure is an optimized synthesis based on a literature procedure.[40] 10.8 g (22.6 mmol) **I2** (Synthesised according to a literature procedure)[40-41] was suspended under argon atmosphere in a flame-dried flask in 240 ml dichloromethane. 14.2 ml (0.1 mol) triethylamine was added to this suspension, which resulted after stirring for several minutes in a clear solution. To this solution was added through a dropping funnel 3.05 ml (25 mmol) methyl 4-chloro-4-oxybutyrate in 30 ml dichloromethane over 30 minutes. The resulting brownish solution was stirred for 16 hours at room temperature. The solution was washe9d two times with a degassed 2 M HCl solution and once with a degassed 0.1 M NaOH solution. The solvent was removed under reduced pressure and the residue purified by column chromatography under argon with degassed silica and a degassed mixture of 2% acetone in dichloromethane. After removal of the solvent 10.06 g (18.12 mmol, 81%) **35** was obtained as an off white gum. ^{1}H-NMR (300 MHz, CDCl3), δ 7.37-7.2 (m, 20H, Ar-*H*), δ 3.60 (s, 3H, CH_3), δ 3.39-3.27 (m, 2H, CH_2), δ 3.24-3.14 (m, 2H, CH_2), δ 2.48 (t, $^3J_{HH}$=6.5 Hz, 2H, CH_2), δ 2.26-2.12 (m, 6H, CH_2), ^{31}P-{^{1}H}-NMR (121.47 MHz, $CDCl_3$), δ -20.31 (1P, -*P*Ph_2), δ -21.77 (1P, -*P*Ph_2). Corresponding to values found before in our group[49].

*3.5.2.4 Synthesis of 4-(Bis(2-(diphenylphosphino)ethyl)amino)-4-oxobutanoic acid **C4***



This procedure is based literature procedure.[40] Under argon atmosphere 10.06 g (18.12 mmol) **I9** was dissolved in tetrahydrofuran. 48 ml (38.05 mmol) of a degassed 0.8 M LiOH solution was added to the resulting solution and the mixture was stirred at room temperature. After 1 Hour 48 ml of water was added and the tetrahydrofuran was removed under reduced pressure.

The free acid **C4** was precipitated from the solution by addition of 25 ml of a degassed 2 M HCl solution. The mixture was filtered and dried to obtain 7.07 g (13.07 mmol, 71%) **C4** as a white powder. ^{1}H-NMR (300 MHz, CDCl3), δ 7.30-7.11 (m, 20H, Ar-*H*), δ 3.29-3.16 (m, 2H, C*H*$_2$), δ 3.08-2.92 (m, 2H, C*H*$_2$), δ 2.30-1.80 (m, 8H, C*H*$_2$), ^{31}P-{^{1}H}-NMR (121.47 MHz, $CDCl_3$), δ -20.41 (1P, -*P*Ph$_2$), δ -21.84 (1P, -*P*Ph$_2$). Corresponding to values found before in our group.[50]

3.5.3 General procedure for bioconjugation maleimides 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 maleimide added from a 50 mM stock solution in the corresponding buffer or DMF for water insoluble maleimides. Progress of the reaction was monitored using LCMS (ES$^+$) analysis. Typically no signal for residual free protein was observed after leaving the reaction for 16 hours at room temperaturet. The mixture was washed using a centrifugal concentrator with buffer to remove the excess maleimide.

3.5.4 Procedure for the coupling of tetrazoles T1 and T2 to SCP-2L V83C-M4

In a quartz cuvette, 200 μM solution tetrazole was added from a 10 mM stock solution in PBS to a 10 μM solution of SCP-2L V83C-**M4** in 1 ml PBS. The resulting mixture was placed in a Varian Cary Eclipse fluorescence spectrometer and irradiated at 320 nm with a filter slot width of 20 nm. The reaction was followed in time by measuring the increase of the emission peak at 450 nm. Only for the reaction with tetrazole **T1** a signal was found in the mass-spectrum belonging to the product (LCMS (ES$^+$), calculated 13785.64 Da; found 13783.2 Da). Product formation was also analysed by SDS-PAGE gel (Invitrogen® NuPAGE® Novex® 4-12% Bis-Tris Gell, ran in MES buffer). After running the gel was analysed under UV-light of ~360 nm to show the formation of the pyrazoline ring. The gel was than stained with Coomassie Blue to see if the fluorescent band was in the same position as the protein band.

3.5.5 General procedure for coupling reactions with Ellman reagent

These experiments are based on a published procedure.[36] A final concentration of 0.6 mM 5,5'dithio-bis(2-nitrobenzoic acid) (Ellman's reagent/DTNB) was added from a 20 mM stock solution in water to a 13 μM solution of protein in 0.1 ml of buffer solution (20 mM MES, 50 mM NaCl, pH 6). The reaction was shaken for an hour at room temperature before analysis by LCMS (ES^+). For the unique cysteine containing proteins conversion to the DTNB adducts was found (Figure 12), while for the modified proteins no modification with DTNB was observed (Figure 13).

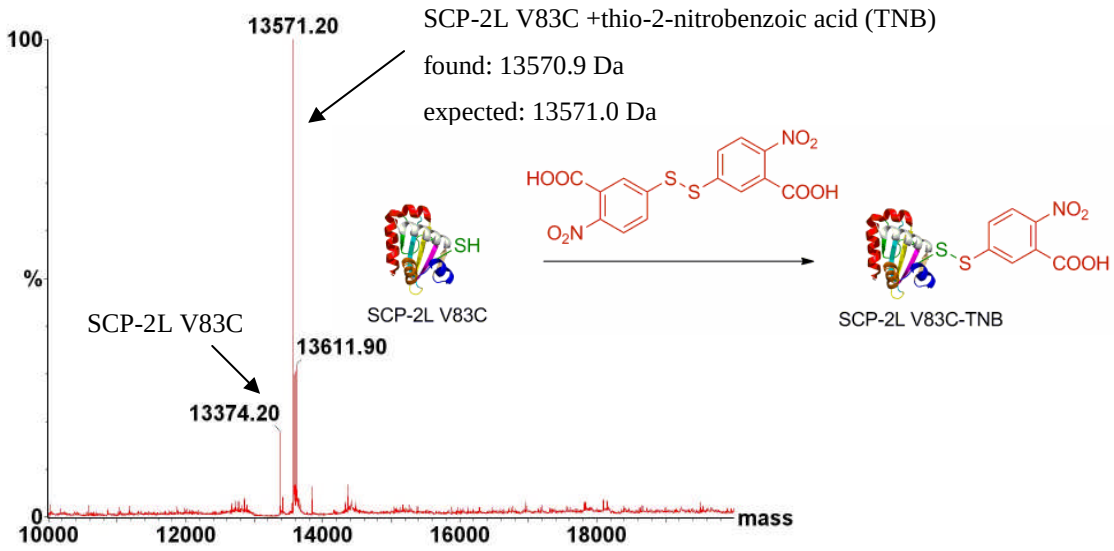


Figure 12. SCP-2L V83C modified upon reaction with Ellman's reagent as observed in the processed mass-spectrum (ES^+).

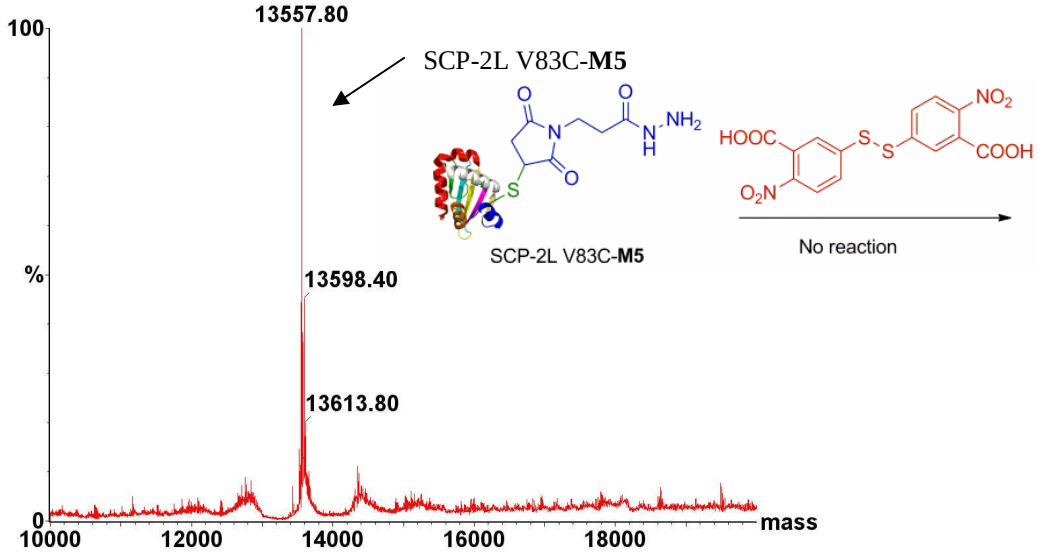
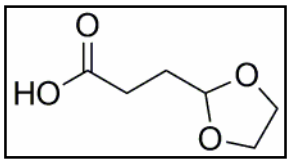


Figure 13. SCP-2L V83C-**M5** showing no modification upon reaction with Ellman's reagent as observed in the processed mass-spectrum (ES+).

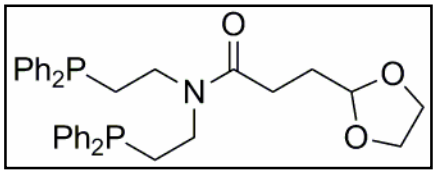
3.5.6 Synthesis of phosphine aldehydes

*3.5.6.1 Synthesis of 3-(1,3-dioxolan-2-yl)propionic acid **I1***



3-(2-dioxolanyl)propionic acid **I1** was synthesised following a literature procedure.[51] The product was purified using fractional distillation (bp 25ºC/1.5 mbar), providing crystalline **I1** in an 42% overall yield. The ^{1}H-NMR spectrum matched the literature data:[51] ^{1}H-NMR (400 MHz, $CDCl_3$), δ 5.00 (t, $^{3}J_{HH}$=4.1 Hz, 1H, -CH_2-C*H*-(O-)$_2$), δ 4.05-3.87 (m, 4H, Ketal-C*H*$_2$'s), δ 2.53 (t, $^{3}J_{HH}$=7.4 Hz, 2H, HOOC-C*H*$_2$-CH_2-), δ 2.07 (td, $^{3}J_{HH}$=7.4 Hz,$^{3}J_{HH}$=4.1 Hz, 2H, -CH_2-C*H*$_2$-CH-); ^{13}C-{^{1}H}-NMR (75.4 MHz, $CDCl_3$), δ 178.3 (C_q, HOO*C*-CH_2), δ 103.4 (CH,-CH_2-*C*H(-O-)$_2$), δ 65.5 (Ketal-*C*H_2's) δ 28.6 (CH_2, -CH_2-*C*H_2-CH-), δ 28.2 (CH_2, HOOC-*C*H_2-CH_2-).

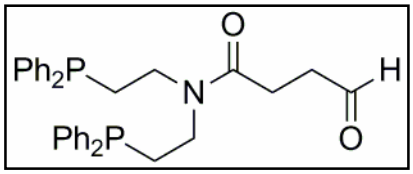
*3.5.6.2 Synthesis of 3-(1,3-dioxolan-2-yl)-N,N-bis(2-(diphenylphosphino) ethyl)propanamide **I3***



This reaction was performed using dry solvents and glassware. 0.139 g (0.95 mmol) of 3-(2-dioxolanyl)propionic acid **I1** was dissolved in 10 ml of DCM and 0.3 ml (2.2 mmol) of Et_3N. The solution was cooled to -10ºC and 136.7 µl (1.05 mmol) of isobutyl chloroformate was added dropwise. The reaction was stirred at -10ºC for 30 minutes after which a cooled solution of 0.5 g (1.05 mmol) of bis(2-(diphenylphosphino)ethyl)amine hydrochloride salt **I2**[40-41] in 4 ml of DCM and 0.3 ml (2.2 mmol) of Et_3N was added dropwise. The reaction mixture was stirred for another hour at -10ºC and for one hour at room temperature after which TLC (Al_2O_3, 7:3 PE(40-60ºC):EtOAC) showed complete conversion of the starting material. The reaction mixture was sequentially washed three times with 10 ml of H_2O (pH 10) and three times with an aqueous HCl solution (pH 3). The organic layer was concentrated and purified on an Al_2O_3 packed column (7:3 PE (40-60ºC):EtOAc), which provided **I3** as white gum (300 mg, 0.53 mmol, 55%). R_f 0.45 (Al_2O_3, 7:3 PE:EtOAc); ^{1}H-NMR (400 MHz, $CDCl_3$), δ 7.3-7.21 (m, 20H, Ar-C*H*), δ 4.77 (t, $^{3}J_{HH}$=4.5 Hz, 1H, -CH_2-C*H*-(O-)$_2$), δ 3.88-3.69 (m, 4H, Ketal-C*H*$_2$'s), δ 3.40-3.15 (m, 4H, Ph_2P-C*H*$_2$-CH_2-), δ 2.26-2.04 (m, 6H, -CH_2-C*H*$_2$-N-, -NOC-C*H*$_2$-CH_2-), δ 1.91-1.82 (m, 2H, -CH_2-C*H*$_2$-CH-); ^{13}C-NMR (101 MHz,

$CDCl_3$), δ 171.6 (C_q, -N-O*C*-CH_2-), δ 137.9 (C_q, $^1J_{CP}$=12.3 Hz, Ar-*C*-P-), δ 137.2 (C_q, $^1J_{CP}$=12.3 Hz, Ar-*C*-P-), δ 132.7 (CH, J_{CP}=19.1 Hz, Ar-*C*H), δ 132.6 (CH, J_{CP}=19.0 Hz, Ar-*C*H), δ 129.0 (CH, Ar-*C*H), δ 128.7 (CH, Ar-*C*H), δ 128.7 (CH, Ar-*C*H), δ 128.7 (CH, Ar-*C*H), δ 128.5 (CH, Ar-*C*H), δ 128.5 (CH, Ar-*C*H), δ 103.5 (CH,-CH_2-*C*H(-O-)$_2$), δ 64.9 (Ketal-*C*H_2's), δ 45.4 (CH_2, J_{CP}=26.3 Hz, -P-*C*H_2-CH_2- or -CH_2-*C*H_2-N-), δ 43.8 (CH_2, J_{CP}=23.9 Hz, -P-*C*H_2-CH_2- or -CH_2-*C*H_2-N-), δ 29.0 (CH_2, -NOC-*C*H_2-CH_2- or -CH_2-*C*H_2-CH-), δ 28.0 (CH_2, J_{CP}=15.3 Hz, -P-*C*H_2-CH_2- or -CH_2-*C*H_2-N-), δ 27.1 (-NOC-*C*H_2-CH_2- or -CH_2-*C*H_2-CH-), δ 26.5 (CH_2, J_{CP}=14.1 Hz, -P-*C*H_2-CH_2- or -CH_2-*C*H_2-N-); ^{31}P-{^{1}H}-NMR (162 MHz, $CDCl_3$), δ -20.4 (s, 1P, -*P*Ph_2), δ -21.8 (s, 1P, -*P*Ph_2). Elemental analysis calculated for $C_{34}H_{37}NO_3P_2$: C 71.69, H 6.55, N 2.46; found: C 71.55, H 6.37, N 2.27.

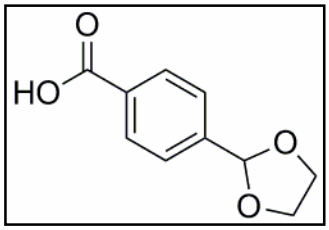
*3.5.6.3 Synthesis of N,N-bis(2-(diphenylphosphino)ethyl)-4-oxobutanamide **P4***



3 g (5.3 mmol) of **I3** was dissolved in a mixture of 150 ml of THF and 50 ml of H_2O. To this mixture was added 46 mg (0.27 mmol) of p-toluenesulfonic acid monohydrate. The mixture was heated at 70ºC for 3 days, after which no further formation of the product was observed on TLC (Al_2O_3, 3:2 PE(40-60ºC: EtOAc). The THF was removed *in vacuo* and 40 ml of DCM was added. The organic layer was separated and the water layer was extracted twice with 40 ml of DCM. The organic layers were combined and the DCM was removed *in vacuo*. Filtration over a small layer of SiO_2 (7:3 EtOAc:PE (40-60ºC)) was performed to remove oxides yielding 2 g (3.8 mmol, 72%) of **P4** as an off-white gum. R_f 0.45 (Al_2O_3, 3:2 PE(40-60ºC):EtOAc); ^{1}H-NMR (400 MHz, $CDCl_3$), δ 9.70 (s, 1H, *H*CO-CH_2-), δ 7.36-7.20 (m, 20H, Ar-C*H*), δ 3.39-3.13 (m, 4H, Ph_2P-C*H*$_2$-CH_2-), δ 2.61 (t, $^3J_{HH}$ = 6.4 Hz, 2H, -CH_2-C*H*$_2$-COH), δ 2.25-2.14 (m, 6H, -CH_2-C*H*$_2$-N-, -NOC-C*H*$_2$-CH_2-); ^{13}C-{^{1}H}-NMR (101 MHz, $CDCl_3$), δ 201.0 (CH, -*C*HO), δ 170.4 (C_q, -N-O*C*-CH_2-), δ 137.9 (C_q, $^1J_{CP}$=12.4 Hz, Ar-*C*-P-), δ 137.2 (C_q, $^1J_{CP}$=12.2 Hz, Ar-*C*-P-), δ 132.7 (CH, J_{CP}=19.1 Hz, Ar-*C*H), δ 129.1 (CH, Ar-*C*H), δ 128.8 (CH, Ar-*C*H), δ 128.7 (CH, Ar-*C*H), δ 128.7 (CH, Ar-*C*H), δ 128.5 (CH, Ar-*C*H), δ 128.5 (CH, Ar-*C*H), δ 45.4 (CH_2, J_{CP}=26.5 Hz, -P-*C*H_2-CH_2- or -CH_2-*C*H_2-N-), δ 43.9 (CH_2, J_{CP}=24.3 Hz, -P-*C*H_2-CH_2- or -CH_2-*C*H_2-N-), δ 38.6 (CH_2, -CH_2-*C*H_2-COH), δ 28.0 (CH_2, J_{CP}=15.4 Hz, -P-*C*H_2-CH_2- or -CH_2-*C*H_2-N), δ 26.5 (CH_2, J_{CP}=14.0 Hz, -P-

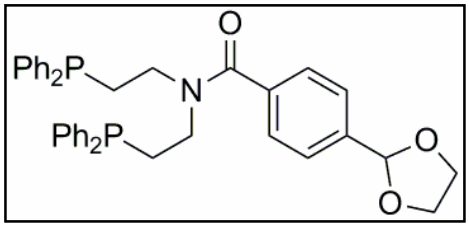
CH_2-CH_2- or -CH_2-*C*H_2-N-) δ 25.6 (-NOC-*C*H_2-CH_2-); ^{31}P-{^{1}H}-NMR (162 MHz, $CDCl_3$), δ -20.2 (s, 1P, -*P*Ph_2), δ -21.6 (s, 1P, -*P*Ph_2); HRMS (TOF MS CI^+), m/z calculated for $C_{32}H_{34}NO_2P_2$ (*M*+H^+) 526.2065, found 526.2067.

*3.5.6.4 Synthesis of 4-(1,3-dioxolan-2-yl)benzoic acid **I4***



4-(1,3-dioxolan-2-yl)benzoic acid **I4** was synthesised following a literature procedure.[42] The product was obtained analytically pure without the need for recrystallization to give 9.4 g of **I4** (48.4 mmol, 74%) as a white solid. The obtained ^{1}H-NMR spectrum was similar to literature data.[52] ^{1}H-NMR (300 MHz, $CDCl_3$), δ 8.07 (d, $^3J_{HH}$=8.3 Hz, 2H, Ar-C*H*), δ 7.53 (d, $^3J_{HH}$=8.3 Hz, 2H, Ar-C*H*), δ 5.82 (s, 1H, Ar-C*H*-(O-)$_2$), δ 4.11-3.96 (m, 4H, Ketal-CH_2's); ^{13}C-{^{1}H}-NMR (75.4 MHz, $CDCl_3$), δ 171.4 (C_q, HOO*C*-Ar), δ 144.2 (C_q Ar-*C*), δ 130.7 (CH, Ar-*C*H), δ 130.3 (C_q Ar-*C*), δ 127.0 (CH, Ar-*C*H), δ 103.3 (CH, Ar-*C*H(-O-)$_2$), δ 65.8 (Ketal-*C*H_2's).

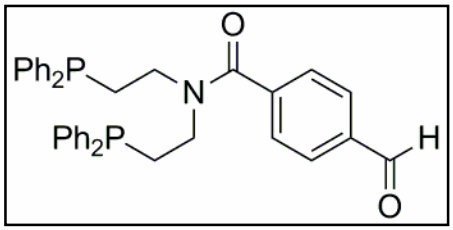
*3.5.6.5 Synthesis of 4-(1,3-dioxolan-2-yl)-N,N-bis(2-(diphenylphosphino) ethyl)-benzamide **I5***



This reaction was performed using dry solvents and glassware. 0.203 g (1.05 mmol) of 4-(1,3-dioxolan-2-yl)benzoic acid **I4**, 0.5 g (1.05 mmol) of bis(2-(diphenylphosphino)ethyl)amine hydrochloride salt **I2**[40-41] and 153 mg (1.26 mmol) of N,N-dimethylaminopyridine were dissolved in 5 ml of DCM. The solution was cooled to 0ºC and a solution of 0.216 g (1.05 mmol) dicyclohexylcarbodiimide in 5 ml of DCM was added dropwise over 30 min. The reaction was stirred at 0ºC for 1 hour and for 16 hours at room temperature after which a white solid precipitated. Thin layer chromatography (SiO_2, 4% acetone in DCM) showed complete conversion. The reaction mixture was filtered over celite and the solvent was removed from the filtrate *en vacuo*. The obtained solid was dissolved in methanol and left at -20ºC overnight after which a white precipitate formed. The white solid was filtered and recrystallized from methanol to provide **I5** as a white crystalline solid (475 mg, 0.74 mmol, 71%). R_f 0.35 (SiO_2, 4% acetone in DCM); ^{1}H-NMR (400 MHz, $CDCl_3$), δ 7.48-7.02 (m,

24H, Ar-C*H*), *δ* 5.76 (s, 1H, Ar-C*H*-(O-)$_2$), *δ* 4.10-3.95 (m, 4H, Ketal-CH_2's), *δ* 3.62-3.40 (m, 2H, Ph_2P-CH_2-CH_2-), *δ* 3.31-3.12 (m, 2H, Ph_2P-CH_2-CH_2-), *δ* 2.45-2.28 (m, 2H, -CH_2-CH_2-N-), *δ* 2.10-1.99 (m, 2H, -CH_2-CH_2-N-); ^{13}C-NMR (101 MHz, $CDCl_3$), *δ* 171.2 (C_q, -N-O*C*-Ar), *δ* 139.2 (C_q Ar-*C*), *δ* 137.8 (C_q, $^1J_{CP}$=13.4 Hz, Ar-*C*-P-), *δ* 137.2 (C_q Ar-*C*), *δ* 136.9 (C_q, $^1J_{CP}$=13.4 Hz, Ar-*C*-P-), *δ* 132.7 (CH, J_{CP}=19.2 Hz, Ar-*C*H), *δ* 132.4 (CH, J_{CP}=19.2 Hz, Ar-*C*H), *δ* 128.8 (CH, Ar-*C*H), *δ* 128.6 (CH, Ar-*C*H), *δ* 128.6 (CH, Ar-*C*H), *δ* 126.7 (CH, Ar-*C*H), *δ* 126.4 (CH, Ar-*C*H), *δ* 103.2 (CH, Ar-*C*H(-O-)$_2$), *δ* 65.3 (Ketal-*C*H_2's), *δ* 46.9 (CH_2, J_{CP}=27.9 Hz, -P-*C*H_2-CH_2- or -CH_2-*C*H_2-N-), *δ* 43.0 (CH_2, J_{CP}=25.1 Hz, -P-*C*H_2-CH_2- or -CH_2-*C*H_2-N-), *δ* 27.8 (CH_2, J_{CP}=16.7 Hz, -P-*C*H_2-CH_2- or -CH_2-*C*H_2-N-), *δ* 26.5 (CH_2, J_{CP}=12.9 Hz, -P-*C*H_2-CH_2- or -CH_2-*C*H_2-N-); ^{31}P-{^{1}H}-NMR (162 MHz, $CDCl_3$), *δ* -20.6 (s, 1P, -*P*Ph_2), *δ* -21.7 (s, 1P, -*P*Ph_2); Elemental analysis calculated for $C_{38}H_{37}NO_3P_2$: C 73.89, H 6.04, N 2.27; found: C 74.04, H 5.97, N 2.18.

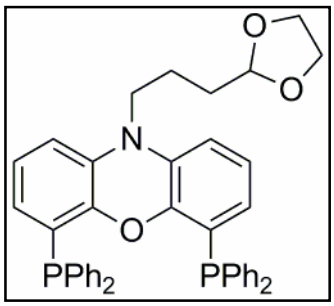
3.5.6.6 Synthesis of N,N-bis(2-(diphenylphosphino)ethyl)-4-formylbenzamide ***P5***



0.4 g (0.65 mmol) of **I5** was dissolved in a mixture of 9 ml of THF and 1 ml of H_2O. To this mixture was added 10 mg (0.05 mmol) of p-toluenesulfonic acid monohydrate. The mixture was refluxed overnight, after which full conversion of the starting material was observed on TLC (SiO_2, 5% acetone in DCM). The THF was removed *in vacuo* and the residue washed with 10 ml of H_2O. The residue was dried *in vacuo* to give **P5** as a white paste (370 mg, 0.65 mmol, 99%). R_f 0.5 (SiO_2, 5% acetone in DCM); ^{1}H-NMR (400 MHz, $CDCl_3$), *δ* 9.94 (s, 1H, *H*CO-CH_2-), *δ* 7.67 (d, $^3J_{HH}$ = 8.3 Hz, 2H, Ar-C*H*), *δ* 7.47-7.02 (m, 22H, Ar-C*H*), *δ* 3.62-3.49 (m, 2H, Ph_2P-CH_2-CH_2-), *δ* 3.25-3.11 (m, 2H, Ph_2P-CH_2-CH_2-), *δ* 2.45-2.34 (m, 2H, -CH_2-CH_2-N-), *δ* 2.02-2.98 (m, 2H, -CH_2-CH_2-N-); ^{13}C-{^{1}H}-NMR (101 MHz, $CDCl_3$), *δ* 191.5 (CH, -*C*HO), *δ* 170.2 (C_q, -N-O*C*-CH_2-), *δ* 142.0 (C_q Ar-*C*), *δ* 137.7 (C_q, $^1J_{CP}$=11.7 Hz, Ar-*C*-P-), *δ* 136.7 (C_q, $^1J_{CP}$=12.7 Hz, Ar-*C*-P-), *δ* 136.5 (C_q Ar-*C*), *δ* 132.7 (CH, J_{CP}=18.5 Hz, Ar-*C*H), *δ* 132.4 (CH, J_{CP}=19.0 Hz, Ar-*C*H), *δ* 129.8 (CH, Ar-*C*H), *δ* 128.9 (CH, Ar-*C*H), *δ* 128.9 (CH, Ar-*C*H), *δ* 128.7 (CH, Ar-*C*H), *δ* 128.6 (CH, Ar-*C*H), *δ* 126.9 (CH, Ar-*C*H), *δ* 46.8 (CH_2, J_{CP}=27.4 Hz, -P-*C*H_2-CH_2- or -

CH_2-*C*H_2-N-), δ 43.0 (CH_2, J_{CP}=23.5 Hz, -P-*C*H_2-CH_2- or -CH_2-*C*H_2-N-), δ 27.9 (CH_2, J_{CP}=15.3 Hz, -P-*C*H_2-CH_2- or -CH_2-*C*H_2-N), δ 25.6 (CH_2, J_{CP}=14.5 Hz, -P-*C*H_2-CH_2- or -CH_2-*C*H_2-N-); ^{31}P-{^{1}H}-NMR (162 MHz, $CDCl_3$), δ -20.9 (s, 1P, -*P*Ph_2), δ -21.8 (s, 1P, -*P*Ph_2); Elemental analysis calculated for $C_{36}H_{33}NO_2P_2$: C 75.38, H 5.80, N 2.44; found: C 75.29, H 5.74, N 2.36.

*3.6.6.7 Synthesis of 10-(3-(1,3-dioxolan-2-yl)propyl)-4,6-bis(diphenyl phosphino)phenoxazine **I8***

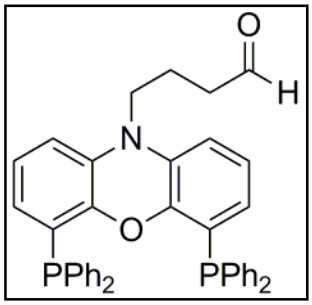


This reaction was performed using dry solvents and glassware. 0.88 g (1.56 mmol) of 4,6-bis(diphenylphosphino)phenoxazine **I6**[43] was dissolved in 30 ml of THF. To this solution 0.08 g (3.2 mmol) of NaH was added and the suspension was refluxed for 1h at 70ºC. The reaction mixture was cooled down to room temperature and 0.49 g (3.13 mmol) of 2-(3-chloropropyl)-1,3-dioxolane **I7** in 10 ml of THF was added. The reaction was then refluxed until TLC (SiO_2, 7:3 PE(40-60ºC):EtOAc) showed complete conversion (60h). The reaction was quenched by adding 20 ml of H_2O and subsequently most of the THF was removed *in vacuo* resulting in a yellow/brown precipitate. The precipitate was filtered off and dried under reduced pressure. Crystallization from DCM/EtOH provided 933 mg crude **I8** (1.4 mmol, 90% yield ~95% purity (^{1}H-NMR)). Crude **I8** was used as such in the following step towards **P6**. Further purification could be achieved by chromatography (SiO_2, 7:3 PE(40-60ºC):EtOAc). R_f 0.27 (SiO_2, 7:3 PE:EtOAc); ^{1}H-NMR (400 MHz, $CDCl_3$), δ 7.25-7.08 (m, 20H, PPh_2(C*H*'s)), δ 6.58 (t, $^3J_{HH}$=7.9 Hz, 2H, Phenoxazin(C*H*)), δ 6.40 (d, $^3J_{HH}$=7.9 Hz, 2H, Phenoxazin(C*H*)), δ 5.92 (d, $^3J_{HH}$=7.9 Hz, 2H, Phenoxazin(C*H*)), δ 4.88 (t, $^3J_{HH}$=3.9 Hz, 1H, -CH_2-C*H*(-O-)$_2$), δ 3.98-3.76 (m, 4H, Ketal-CH_2's), δ 3.48 (t, $^3J_{HH}$=7.6 Hz, 2H, N-CH_2-CH_2-), δ 1.81-1.60 (m, 4H, -CH_2-CH_2-CH_2-CH-); ^{13}C-{^{1}H}-NMR (101 MHz, $CDCl_3$), δ 147.1* (t, J=10.4 Hz, C_q), δ 136.8* (t, J=4.0 Hz, C_q), δ 133.9* (t, J=10.4 Hz, Phenyl *C*H), δ 133.0 (C_q), δ 128.3 (Phenyl *C*H), δ 128.1* (t, J=3.3 Hz, Phenyl *C*H), δ 125.2 (Phenoxazin *C*H), δ 124.4* (t, J=4.4 Hz, C_q), δ 123.8 (Phenoxazin *C*H), δ 111.9 (Phenoxazin CH), δ 103.9 (-CH_2-*C*H(-O-)$_2$), δ 65.0 (Ketal-*C*H_2's), δ 44.5 (N-*C*H_2-CH_2-), δ 30.7 (-CH_2-CH_2-*C*H_2-CH- or -CH_2-*C*H_2-CH_2-CH_2-CH-), δ 18.9 (-CH_2-CH_2-*C*H_2-CH- or -CH_2-*C*H_2-CH_2-

CH-); ^{31}P-{^{1}H}-NMR (162 MHz, $CDCl_3$), δ -18.9 (s); HRMS (TOF MS CI^+), m/z calculated for $C_{42}H_{38}NO_3P_2$ (M+H^+) *m/z* 666.2327, found *m/z* 666.2341.

* These ^{13}C signals appear as a triplet with the above coupling constants, but are actually second order signals resulting from an AA'X spin system.[53]

3.5.6.8 Synthesis of 4-(4,6-bis(diphenylphosphino)-phenoxazin-10-yl)butanal ***P6***



0.93 g (1.4 mmol) of **I8** was dissolved in a mixture of 50 ml of THF and 15 ml of H_2O. To this mixture was added 24 mg (0.14 mmol) of p-toluenesulfonic acid monohydrate. The mixture was heated to 70ºC leading to the formation of a white precipitate. After 2 days no further conversion was observed on TLC (Al_2O_3, 3:2 PE(40-60ºC: EtOAc). THF was removed *in vacuo* and 40 ml of DCM was added. The water layer was removed and the organic layer washed twice with 20 ml of H_2O. The organic layer was then concentrated and 15 ml of H_2O, 50 ml of THF and 24 mg (0.14 mmol) of *p*-toluenesulfonic acid were added and the reaction mixture was heated at 70ºC until full conversion was achieved according to TLC. Filtration of the reaction mixture afforded **P6** as a white powder (630 mg, 0.95 mmol, 72% yield). R_f 0.37 (Al_2O_3, 7:3 PE(40-60ºC):EtOAc); ^{1}H-NMR (300 MHz, $CDCl_3$), δ 9.79 (t, 1H, $^{3}J_{HH}$ = 0.6 Hz, *H*CO-CH_2-), δ 7.20-7.06 (m, 20H, PPh_2(C*H*'s)), δ 6.61 (t, $^{3}J_{HH}$=7.9 Hz, 2H, Phenoxazin(C*H*)), δ 6.48 (dd, $^{4}J_{HH}$=1.3 Hz, $^{3}J_{HH}$=7.9 Hz, 2H, Phenoxazin(C*H*)), δ 5.93 (dtd, $^{4}J_{HH}$=1.5 Hz, $^{5}J_{HH}$=1.9 Hz, 2H, $^{3}J_{HH}$=7.9 Hz, Phenoxazin(C*H*)), δ 3.52-3.40 (m, 2H, N-CH_2-CH_2-), δ 2.57 (t, $^{3}J_{HH}$=0.6 Hz, $^{3}J_{HH}$=6.4 Hz, 2H, -CH_2-CH_2-COH), δ 1.92 (tt, $^{3}J_{HH}$=6.5 Hz, $^{3}J_{HH}$=8.0 Hz, 2H, -CH_2-CH_2-CH_2-); ^{13}C-{^{1}H}-NMR (101 MHz, $CDCl_3$), δ 201.0 (-*C*HO), δ 147.1* (t, J=10.5 Hz, C_q), δ 136.9* (t, J=6.4 Hz, C_q), δ 133.9* (t, J=10.4 Hz, Phenyl *C*H), δ 132.9 (C_q), δ 128.2 (Phenyl *C*H), δ 128.1* (t, J=3.3 Hz, Phenyl *C*H), δ 125.4 (Phenoxazin *C*H), δ 124.9* (t, J = 9.6 Hz, C_q), δ 123.8 (Phenoxazin *C*H), δ 111.8 (Phenoxazin *C*H), δ 43.5 (N-*C*H_2-CH_2-), δ 40.8 (-CH_2-*C*H_2-CHO), δ 17.3 (-CH_2-*C*H_2-CH_2-); ^{31}P-{^{1}H}-NMR (121.5 MHz, $CDCl_3$), δ -19.0 (s); Elemental analysis calculated for $C_{40}H_{33}NO_2P_2$: C 77.28, H 5.35, N 2.25; found: C 77.28, H 5.35, N 2.25, found: C 77.28, H 4.98, N 2.47.

* These ^{13}C signals appear as a triplet with the above coupling constants, but are actually second order signals resulting from an AA'X spin system.[53]

3.5.7 General procedure for phosphine aldehyde modification of hydrazide modified proteins

2-10 equivalents of phosphine aldehyde **P1**-**P8** was added to a 25-50 mM solution of **M5**-modified protein in buffer (20 mM MES 50 mM NaCl pH 6) from a 50 mM stock solution in degassed DMF or MeCN under inert atmosphere. The reaction mixture left overnight after which the excess phosphine was removed by centrifugation. The remaining protein solution was washed three times with the appropriate buffer in a centrifugal concentrator. The mass spectra (LCMS, ES^+) from the different resulting solutions provided the results outlined in Table 3. Examples of the processed mass spectra obtained for the modification using ligands **P1**, **P2**, **P3** and **P4** are shown in Figure 14, Figure 15, Figure 16 and Figure 17 below.

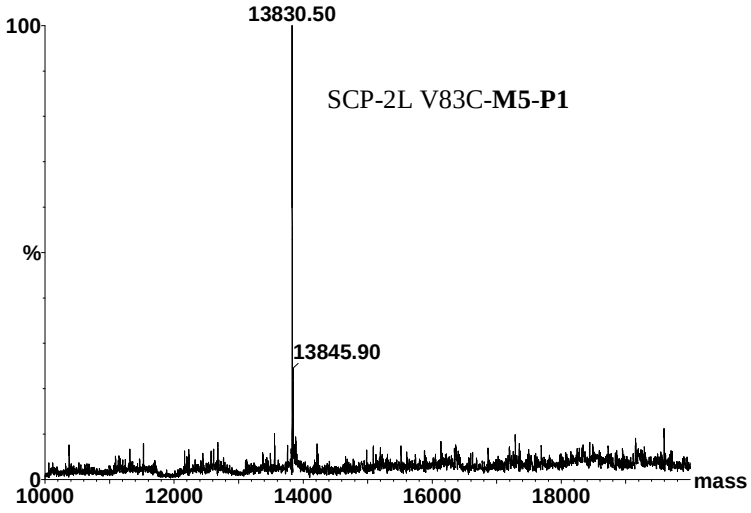


Figure 14. The processed mass-spectrum (ES^+) of SCP-2L V83C-**M5**-**P1**.

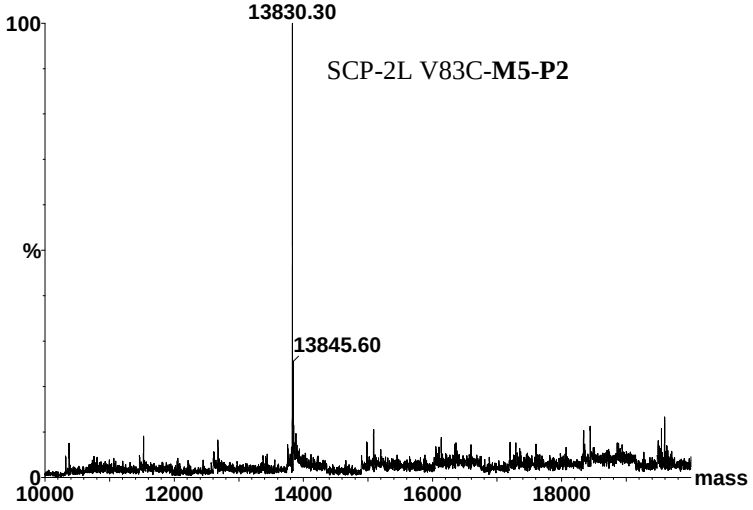


Figure 15. The processed mass-spectrum (ES^+) of SCP-2L V83C-**M5**-**P2**.

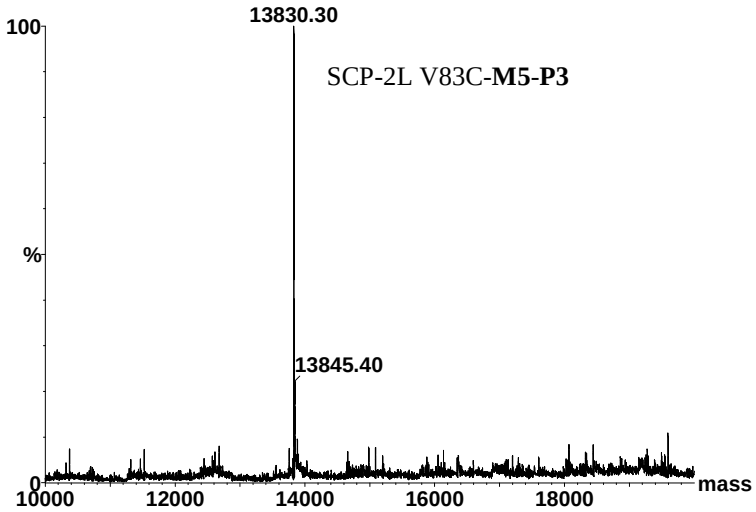


Figure 16. The processed mass-spectrum (ES^+) of SCP-2L V83C-**M5**-**P3**.

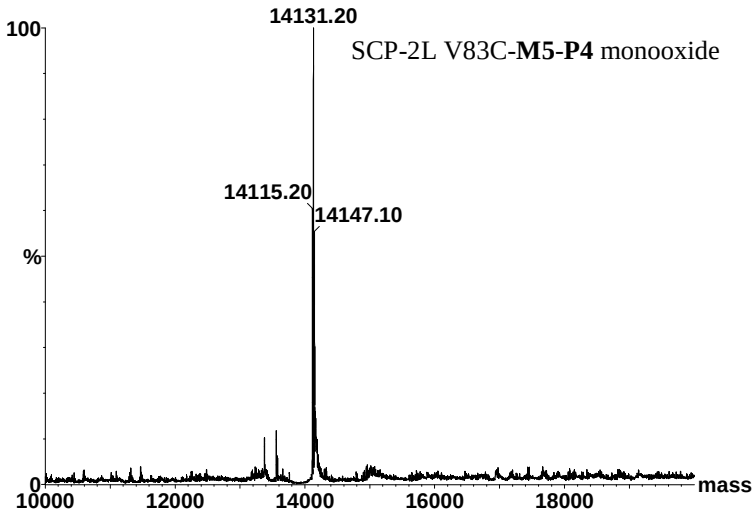


Figure 17. The processed mass-spectrum (ES^+) of SCP-2L V83C-**M5**-**P4**

3.5.8 General procedure for tryptic digests

The samples (5 μL, 10 pmoles/μL) were dialysed against buffer (50 mM ammonium bicarbonate pH 8.0) using a membrane filter (Millipore, Billerica, MA). Trypsin (0.5 μL, 0.1 μg, Promega, Madison, WI) was added to the resulting solutions. The solutions were incubated at 37°C overnight and the digested solution (0.5 μL). This solution together with an α-cyano-4-hydroxycinnamic acid matrix (0.5 μL, 10 mg/mL in 50:50 acetonitrile:0.1% TFA) and 0.1% TFA (0.5 μL) was applied to the MALDI target, and after being allowed to dry, analyzed by MALDI-MS.

MALDI MS was acquired using a 4800 MALDI TOF/TOF Analyser (Applied Biosystems, Foster City, CA) equipped with a Nd:YAG 355 nm laser and calibrated using a mixture of peptides. The samples were initially analysed in positive MS mode between 800 and 4000 m/z, by averaging 1000 laser shots. The most desired peptides were selected for MSMS analysis and acquired to a maximum of 3000 laser shots or until the accumulated spectrum reached a S/N ratio of 35. All MS/MS data were acquired using 1 keV collision energy.

The combined MS and MSMS data were analysed, using GPS Explorer (Applied Biosystems) to interface with the Mascot 2.1 search engine (Matrix Science, London, UK), against the UniProt (Swiss-Prot and TREMBL combined) database (April 2009). No species restriction was applied. The data were searched with tolerances of 100 ppm for the precursor ions and 0.5 Da for the fragment ions, with trypsin as the cleavage enzyme, assuming up to one missed cleavage, carbamidomethyl modification of cysteines as a fixed modification and methionine oxidation selected as a variable modification. Digest results are shown for SCP-2L V83C-**M5-P3** in Figure 6 and for SCP-2L V83C-**M5-P4** in Figure 7 (MS-MS data can be found in appendix A2). The depicted peptides were the only peptide found matching phosphine (oxidized or non oxidized) modification.

In the above experiments an active search was performed for other peptides with the same modification but no evidence for those was found. In addition, the samples were analysed by ESI where again no evidence was found for other peptide sequences containing the modification. For ESI analysis the peptides were acidified with TFA and then separated using an UltiMate nanoLC (LC Packings, Amsterdam) equipped with a PepMap C18 trap & column, using a 60 min gradient of increasing acetonitrile

concentration, containing 0.1 % formic acid (5-35% acetonitrile in 35 min respectively, 35-50% in a further 20 min, followed by 95% acetonitrile to clean the column). The eluent was sprayed into a Q-Star XL tandem mass spectrometer (Applied Biosystems, Foster City, CA) and analysed in Information Dependent Acquisition (IDA) mode, performing 1 sec of MS followed by 3 sec MSMS analyses of the 2 most intense signals observed in MS. These masses are then excluded from analysis for the next 60 sec. MS/MS data for doubly and triply charged precursor ions were converted to centroid data, without smoothing, using the Analyst QS1.1 mascot.dll data import filter with default settings. The MS/MS data file generated was analysed using the Mascot 2.1 search engine (Matrix Science, London, UK) against an internal database containing the protein sequence of interest. The data were searched with tolerances of 0.2 Da for the precursor and fragment ions, with trypsin as the cleavage enzyme and allowing for one missed cleavage. Phosphine (free, oxidized and for **23** double oxidized) modification of cysteine and lysine and methionine oxidation were selected as variable modifications.

3.5.9 Procedure for recording ^{31}P-{^{1}H}-spectra of proteins

The ^{31}P-{^{1}H}-spectra of SCP-2L V83C and SCP-2L V83C-**M5-P3** were recorded on a Bruker Avance 500 spectrometer over 16 hours at room temperature. The protein sample was concentrated to ~2 mM and 10% D_2O was added for locking. Usually, small amounts of precipitate formed during the recording period. The ^{31}P-{^{1}H}-spectrum of SCP-2L V83C before modification is depicted in Figure 5a showing phosphorous signals at 0-1 ppm. These signals were also observed in the ^{31}P-{^{1}H}-spectra of SCP-2L V83C-**M5-P3** (Figure 5b) which have been subjected to extensive washing, which gave rise to the hypothesis that they correspond to covalently or very strongly non covalently bound phosphates.[54]

3.6 References

[1] a) F. H. Jardine, J. A. Osborn, G. Wilkinson, J. F. Young, *Chem. Ind. (London, U. K.)* **1965**, 560; b) D. N. Lawson, J. A. Osborn, G. Wilkinson, *J. Chem. Soc., A* **1966**, 1733.

[2] a) A. Börner, *Phosphorus Ligands in Asymmetric Catalysis; Synthesis and Applications*, Wiley-VCH, Weinheim, **2008**; b) P. W. N. M. van Leeuwen, *Homogeneous Catalysis: Understanding the Art*, Kluwer Academic Publishers, Dordrecht, **2004**.

[3] J. A. Gillespie, D. L. Dodds, P. C. J. Kamer, *Dalton Trans.* **2010**, *39*, 2751.

[4] M. E. Wilson, G. M. Whitesides, *J. Am. Chem. Soc.* **1978**, *100*, 306.
[5] a) G. Klein, N. Humbert, J. Gradinaru, A. Ivanova, F. Gilardoni, U. E. Rusbandi, T. R. Ward, *Angew. Chem., Int. Ed.* **2005**, *44*, 7764; b) J. Collot, J. Gradinaru, N. Humbert, M. Skander, A. Zocchi, T. R. Ward, *J. Am. Chem. Soc.* **2003**, *125*, 9030; c) M. Skander, N. Humbert, J. Collot, J. Gradinaru, G. Klein, A. Loosli, J. Sauser, A. Zocchi, F. Gilardoni, T. R. Ward, *J. Am. Chem. Soc.* **2004**, *126*, 14411; d) U. E. Rusbandi, C. Lo, M. Skander, A. Ivanova, M. Creus, N. Humbert, T. R. Ward, *Adv. Synth. Catal.* **2007**, *349*, 1923; e) U. E. Rusbandi, M. Skander, A. Ivanova, C. Malan, T. R. Ward, *C. R. Chim.* **2007**, *10*, 678.
[6] a) M. T. Reetz, J. J. P. Peyralans, A. Maichele, Y. Fu, M. Maywald, *Chem. Commun. (Cambridge, U. K.)* **2006**, 4318; b) C.-C. Lin, C.-W. Lin, A. S. C. Chan, *Tetrahedron: Asymmetry* **1999**, *10*, 1887.
[7] J. Pierron, C. Malan, M. Creus, J. Gradinaru, I. Hafner, A. Ivanova, A. Sardo, T. R. Ward, *Angew. Chem., Int. Ed.* **2008**, *47*, 701.
[8] H. Yamaguchi, T. Hirano, H. Kiminami, D. Taura, A. Harada, *Org. Biomol. Chem.* **2006**, *4*, 3571.
[9] M. T. Reetz, M. Rentzsch, A. Pletsch, M. Maywald, *Chimia* **2002**, *56*, 721.
[10] a) M. Nuzzolo, A. Grabulosa, A. M. Z. Slawin, N. J. Meeuwenoord, G. A. van der Marel, P. C. J. Kamer, *Eur. J. Org. Chem.* **2010**, 3229; b) M. Caprioara, R. Fiammengo, M. Engeser, A. Jäschke, *Chem.--Eur. J.* **2007**, *13*, 2089.
[11] M. Nuzzolo, Ph.D. Thesis, University of St. Andrews (St. Andrews, UK), **2010**.
[12] G. T. Hermanson, *Bioconjugate Techniques*, Second ed., Acadamic Press, San Diego, **2008**.
[13] a) M. Maywald, Ph.D. Thesis, Universität Bochum (Mülheim an der Ruhr, Germany), **2005**; b) G. Popa, Ph.D. Thesis, University of St. Andrews (St. Andrews, UK), **2010**.
[14] E. Hedaya, S. Theodoropulos, *Tetrahedron* **1968**, *24*, 2241.
[15] M. W. Jones, G. Mantovani, S. M. Ryan, X. Wang, D. J. Brayden, D. M. Haddleton, *Chem. Commun. (Cambridge, U. K.)* **2009**, 5272.
[16] a) K. Nwe, M. W. Brechbiel, *Cancer Biother Radiopharm* **2009**, *24*, 289; b) J. M. Antos, M. B. Francis, *Curr. Opin. Chem. Biol.* **2006**, *10*, 253; c) S. I. van Kasteren, H. B. Kramer, H. H. Jensen, S. J. Campbell, J. Kirkpatrick, N. J. Oldham, D. C. Anthony, B. G. Davis, *Nature (London, U. K.)* **2007**, *446*, 1105.
[17] H. Staudinger, J. Meyer, *Helv. Chim. Acta* **1919**, *2*, 635.
[18] E. Saxon, C. R. Bertozzi, *Science (Washington, D. C.)* **2000**, *287*, 2007.
[19] a) Y. A. Lin, J. M. Chalker, B. G. Davis, *ChemBioChem* **2009**, *10*, 959; b) Y. A. Lin, J. M. Chalker, N. Floyd, G. J. L. Bernardes, B. G. Davis, *J. Am. Chem. Soc.* **2008**, *130*, 9642.
[20] a) R. den Heeten, B. K. Munoz, G. Popa, W. Laan, P. C. J. Kamer, *Dalton Trans.* **2010**, *39*, 8477; b) W. Laan, B. K. Munoz, R. den Heeten, P. C. J. Kamer, *ChemBioChem* **2010**, *11*, 1236.
[21] G. S. Heaton, H. N. Rydon, J. A. Schofield, *Journal of the Chemical Society* **1956**, 3157.
[22] J. A. Burns, J. C. Butler, J. Moran, G. M. Whitesides, *J. Org. Chem.* **1991**, *56*, 2648.

[23] a) R. Rubner, E. Kuehn, H. Ahne, (Siemens A.-G., Fed. Rep. Ger.). Application: DE, **1980**, p. 12 pp; b) W. Trommer, M. Hendrick, *Synthesis* **1973**, 484.
[24] a) P. M. Hergenrother, S. J. Havens, J. W. Connell, *Polym. Prepr. (Am. Chem. Soc., Div. Polym. Chem.)* **1986**, *27*, 408; b) Z. Luo, L. Wei, F. Liu, T. Zhao, *Eur. Polym. J.* **2007**, *43*, 3461.
[25] R. J. Detz, S. Arevalo Heras, R. De Gelder, P. W. N. M. Van Leeuwen, H. Hiemstra, J. N. H. Reek, J. H. Van Maarseveen, *Org. Lett.* **2006**, *8*, 3227.
[26] W. Song, Y. Wang, J. Qu, M. M. Madden, Q. Lin, *Angew. Chem., Int. Ed.* **2008**, *47*, 2832.
[27] a) Y. Wang, C. I. Rivera Vera, Q. Lin, *Org. Lett.* **2007**, *9*, 4155; b) V. Lohse, P. Leihkauf, C. Csongar, G. Tomaschewski, *J. Prakt. Chem.* **1988**, *330*, 406; c) J. S. Clovis, A. Eckell, H. Rolf, R. Sustmann, *Chem. Ber.* **1967**, *100*, 60; d) M. V. George, C. S. Angadiyavar, *J. Org. Chem.* **1971**, *36*, 1589; e) A. Padwa, S. Nahm, E. Sato, *J. Org. Chem.* **1978**, *43*, 1664.
[28] W. Song, Y. Wang, J. Qu, Q. Lin, *J. Am. Chem. Soc.* **2008**, *130*, 9654.
[29] G. J. L. Bernardes, J. M. Chalker, J. C. Errey, B. G. Davis, *J. Am. Chem. Soc.* **2008**, *130*, 5052.
[30] T. J. Holmes, Jr., R. G. Lawton, *J. Am. Chem. Soc.* **1977**, *99*, 1984.
[31] a) R. C. Clevenger, K. D. Turnbull, *Synth. Commun.* **2000**, *30*, 1379; b) T. Kurosaki, T. Matsumoto, K. Sawada, S. Kurishima, T. Numakura, *Journal of Photographic Science* **1988**, *36*, 122.
[32] S. Ito, Y. Tanaka, A. Kakehi, *Bull. Chem. Soc. Jpn.* **1976**, *49*, 762.
[33] T. P. King, S. W. Zhao, T. Lam, *Biochemistry* **1986**, *25*, 5774.
[34] a) I. S. Carrico, B. L. Carlson, C. R. Bertozzi, *Nat. Chem. Biol.* **2007**, *3*, 321; b) A. Dirksen, P. E. Dawson, *Bioconjugate Chem.* **2008**, *19*, 2543.
[35] K. K. Hii, S. D. Perera, B. L. Shaw, M. Thornton-Pett, *J. Chem. Soc., Dalton Trans.* **1992**, 2361.
[36] J. M. Chalker, C. S. C. Wood, B. G. Davis, *J. Am. Chem. Soc.* **2009**, *131*, 16346.
[37] G. P. Schiemenz, H. Kaack, *Justus Liebigs Ann. Chem.* **1973**, *9*, 1480.
[38] D. Rivillo, H. Gulyas, J. Benet-Buchholz, E. C. Escudero-Adan, Z. Freixa, P. W. N. M. van Leeuwen, *Angew. Chem., Int. Ed.* **2007**, *46*, 7247.
[39] Y. S. Hon, C. F. Lee, R. J. Chen, P. H. Szu, *Tetrahedron* **2001**, *57*, 5991.
[40] M. E. Wilson, R. G. Nuzzo, G. M. Whitesides, *J. Am. Chem. Soc.* **1978**, *100*, 2269.
[41] R. G. Nuzzo, S. L. Haynie, M. E. Wilson, G. M. Whitesides, *J. Org. Chem.* **1981**, *46*, 2861.
[42] W. A. Denny, G. J. Atwell, B. C. Baguley, B. F. Cain, *J. Med. Chem.* **1979**, *22*, 134.
[43] L. A. Van der Veen, P. H. Keeven, G. C. Schoemaker, J. N. H. Reek, P. C. J. Kamer, P. W. N. M. Van Leeuwen, M. Lutz, A. L. Spek, *Organometallics* **2000**, *19*, 872.
[44] A. J. Sandee, J. N. H. Reek, P. C. J. Kamer, P. W. N. M. van Leeuwen, *J. Am. Chem. Soc.* **2001**, *123*, 8468.
[45] A. J. Sandee, D. Dimitrijevic, R. J. van Haaren, J. N. H. Reek, P. C. J. Kamer, P. W. N. M. van Leeuwen, *J. Mol. Catal. A: Chem.* **2002**, *182-183*, 309.
[46] B. Schlummer, U. Scholz, *Adv. Synth. Catal.* **2004**, *346*, 1599.
[47] P. J. Deuss, G. Popa, C. H. Botting, W. Laan, P. C. J. Kamer, *Angew. Chem., Int. Ed.* **2010**, *49*, 5315.

[48] M. M. Bradford, *Anal. Biochem.* **1976**, *72*, 248.
[49] R. Den Heeten, P. C. J. Kamer, P. W. N. M. van Leeuwen, **2006**.
[50] R. den Heeten, Ph.D. Thesis, Universiteit van Amsterdam (Amsterdam, The Netherlands), **2009**.
[51] K. J. Shea, E. Wada, *J. Am. Chem. Soc.* **1982**, *104*, 5715.
[52] S. Mahboobi, A. Sellmer, H. Hoecher, C. Garhammer, H. Pongratz, T. Maier, T. Ciossek, T. Beckers, *J. Med. Chem.* **2007**, *50*, 4405.
[53] M. Kranenburg, Y. E. M. van der Burgt, P. C. J. Kamer, P. W. N. M. van Leeuwen, K. Goubitz, J. Fraanje, *Organometallics* **1995**, *14*, 3081.
[54] M. D. Davis, D. E. Edmondson, F. Mueller, *Eur. J. Biochem.* **1984**, *145*, 237.

Chapter 4. Biphasic rhodium catalysed hydroformylation of higher linear alkenes

Rhodium complexation and catalysis using phosphine modified SCP-2L protein ligands

4.1 Abstract

This chapter describes studies of transition metal complexation to phosphine modified proteins described in the previous chapter and initial results of the application of these artificial metalloenzymes in the hydroformylation of alkenes. A phosphine-rhodium conjugate was successfully obtained and characterised. The obtained complexes were applied in the biphasic rhodium catalysed hydroformylation of linear Ŭ-olefins. Increased activity was found compared to the traditional Rh/TPPTS catalyst, along with good linear selectivity between 75 and 85% depending on the substrate and the SCP-2L mutant phosphine rhodium complex used. The results imply a true shape selective catalytic system that involves substrate encapsulation by the protein template. This work demonstrates for the first time such a system obtained through rational design of the appropriate artificial metalloenzyme. In addition, initial results are shown for the hydroformylation of other alkene substrates using the developed catalysts.

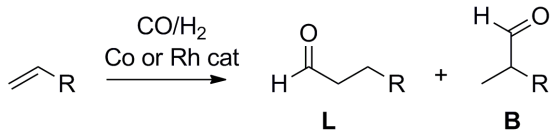
The rhodium complex formation procedure has been published in a communication article in Angewandte Chemie International Edition:

P. J. Deuss, G. Popa, C. H. Botting, W. Laan, P. C. J. Kamer, *Angew. Chem., Int. Ed.* **2010**, *49*, 5315.

4.2 Introduction

4.2.1 Hydroformylation

In the previous chapters the development of phosphine containing artificial metalloenzymes based on SCP-2L was described. SCP-2L was specifically chosen for its ability to bind linear aliphatic substrates. One of the main opportunities for application of this system would be in the rhodium catalysed hydroformylation of linear alkenes. This reaction, which is also known as ñoxo synthesisò, transforms alkenes into aldehyde products using synthesis gas (scheme 1).[1] Catalysts for this reaction usually consist of cobalt or rhodium hydrides containing phosphine ligands.[2] The importance of the reaction lies in the high added value of the aldehyde products compared to the cheap alkene substrates. For example, the products of hydroformylation of higher linear Ŭ-olefins (C6 and higher) are widely used in the detergent and surfactant industry. In the reaction two regioisomers can be formed; the linear (**L**) and branched (**B**) aldehydes. Selectivity towards one of these products is one of the key goals in the catalyst development for this reaction.[3]

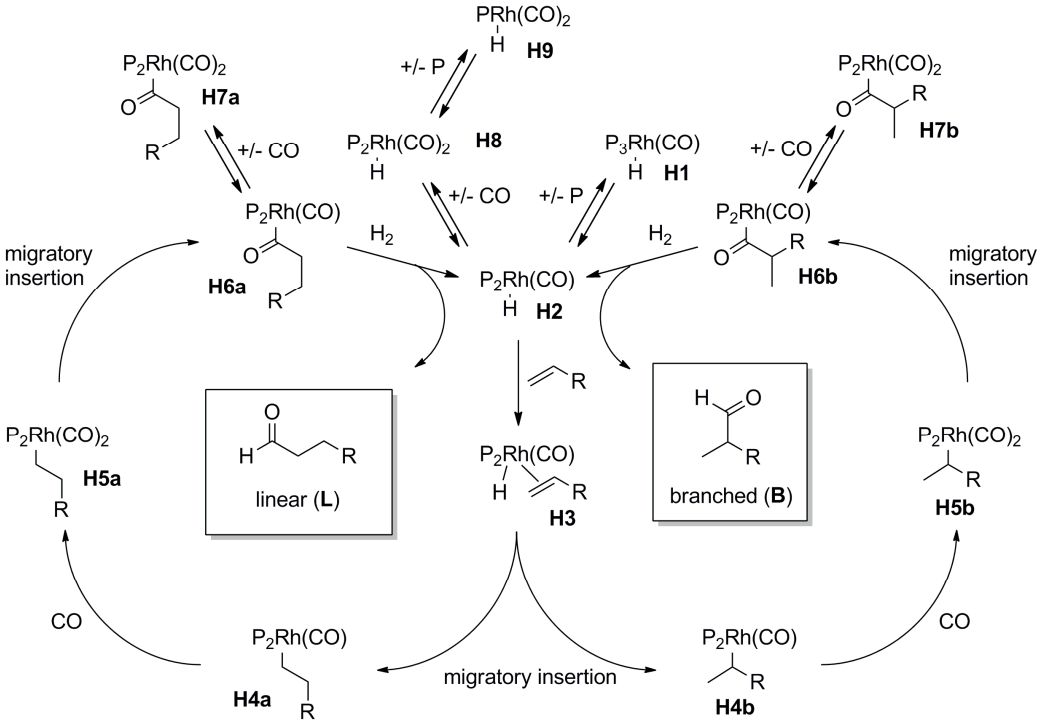


Scheme 1. Hydroformylation of alkenes to aldehyde products.

4.2.2 Rhodium catalysed hydroformylation

The generally accepted catalytic cycle for the rhodium catalysed hydroformylation of a monosubstituted alkene towards the linear and branched aldehyde products is depicted in Scheme 2.[2, 4] The hydride complex **H1**, containing three phosphine ligands, is the typical catalytic precursor, which is obtained by mixing the phosphine with a typical rhodium precursor like [Rh(acac)$(CO)_2$]. Dissociation of a phosphine ligand from the hydride complex **H1** forms **H2**, which is considered to be the catalytically active species. **H2** has so far not been directly observed but indirect evidence suggests phosphine dissociation from complex **H1** is required for initiation of the catalytic cycle.[5] Alkene coordination to **H2** gives **H3**, which undergoes migratory insertion of the olefin into the Rhï H bond. For Ŭ-olefins, this process can give either **H4a**, which yields the linear aldehyde product via **H5a** and **H6a**, or **H4b**, which yields the branched aldehyde product via **H5b** and **H6b**. In general, linear-

selectivity is found for Ŭ-olefins using rhodium phosphine complexes due to the steric bulk of species **H4a**, except in special cases were the branched carbon is activated, such as vinyl acetate or styrene. Species **H2**, **H6a** and **H6b** can undergo complexation of an additional CO molecule to form the resting states **H7a**, **H7b** and **H8** respectively, which have been observed *in situ* by high pressure ^{31}P-NMR and IR studies.[5-6] Dicarbonyl species **H8** can also form inactive dimers, which can be reactivated with hydrogen.[6b, 7]



Scheme 2. Catalytic cycle for the rhodium catalysed hydroformylation.

The electronic parameters of the ligands can play an important role in the activity and selectivity. For example, electron withdrawing phosphines decrease Rh-CO π-backdonation, weakening the Rhï CO bond, thereby positively influencing the equilibrium between the dicarbonyl and monocarbonyl species to provide an increased catalytic activity. There are several commonly observed pathways that lead to undesired side products. Species **H4a** and **H4b** can undergo hydrogenolysis to afford the hydrogenated alkene. When R is an alkyl chain species **H4b** can undergo ß-hydride elimination leading to isomerization of the alkene towards more stable internal olefins, which are far less reactive towards hydroformylation. Dicarbonyl species **H8** can lose a phosphine to form monophosphine complex **H9**, which

typically leads to unselective hydroformylation since the complex has only one phosphine to control the selectivity. Every substrate brings different demands for the catalyst on selectivity and activity, which requires fine tuning of the ligand. Therefore ligand development is a crucial aspect of homogeneous rhodium catalysed hydroformylation. Not surprisingly, numerous groups are actively involved in ligand design and screening using computational and empirical approaches.[3c]

4.2.3. The biphasic rhodium catalysed hydroformylation process

All aspects of the industrial hydroformylation processes have been well developed, like reactor engineering, product separation and catalyst recycling. Especially for lower linear Ŭ-olefins (C3, propene and C4, 1-butene) sophisticated processes exist and the hydroformylation is performed on multi-ton/year scale. The most successful process has been developed by Ruhrchemie/Rhône-Poulenc, which involves a highly effective biphasic rhodium phosphine catalytic system (Figure 1).

No rights for online publication of this figure were obtained.

Please refer to *Homogeneous Catalysis: Understanding the Art* by P. W. N. M. van Leeuwen for this figure.[2]

Figure 1**.** Flow chart showing the hydroformylation process as performed by Ruhrchemie/Rhône-Poulenc

This system uses a water soluble rhodium catalyst by employing triphenylphosphine trisulfonic acid trisodium salt (TPPTS) as the phosphine ligand.[8] Rhodium-TPPTS complexes are dissolved in the aqueous phase while the organic phase consists of the reaction products, mainly linear aldehyde (92%). The substrate and the synthesis gas are fed into the reactor through a gas dispersion system for optimal mixing. The most important part of this process is the phase separation system where the product is separated from the aqueous phase containing the catalyst. After separation the organic layer is led to a distillation unit where the linear product is separated from the branched product.

4.2.4 Application of the biphasic system for higher alkenes

The process described above can be applied to higher alkenes. This is however not carried out on an industrial scale due to a significant reduction in the turnover numbers when the substrate hydrocarbon chain length is increased. A nearly exponential decrease in the reaction rate is observed for substrates of increasing chain length, which arises from the decreasing water solubility of these substrates (Figure 2) Therefore, the industrial hydroformylation of alkenes with a medium chain length (C5-C8) is performed using phosphine rhodium catalysts in a single organic phase. The reaction products obtained from these reactions can be separated from the catalyst solution by distillation.[9] Distillation is less desirable than a phase separation as this causes high thermal stress on the catalyst system leading to significant catalyst loss. For this reason often cheaper cobalt based catalyst systems are used for these processes. Although the use of Rh/TPPTS in a biphasic system is not economically viable for these substrates, several studies have been performed to find the optimal conditions for hydroformylation of medium chain length alkenes in a biphasic system. Selectivity and activity were found to be very sensitive to the equilibrium between **H2** and the less active and selective **H8**, which is dependent on the CO pressure (Scheme 2). It was found that the synthesis gas pressure reaches an optimum between 30 and 80 bar depending on the exact conditions and that selectivity decreases significantly at high pressures (270 bar) due to increased formation of **H8**. Also high ratios of TPPTS to rhodium are required to prevent the formation of **H9**, which also leads to unselective catalysis.[6d] The pH of the reaction is optimal between 5.5 and 6.5. It has been found that over the course of the reaction the pH can drop by one pH unit due to

the formation of CO_2 by the water-gas shift reaction.[10] Variation of the counter ions to replace sodium showed little effect on both rate and selectivity, although increasing the amount of salts leads to a decrease in rate. This ñsalt-effectò is due to a decreased solubility of the alkene in the aqueous phase at increased ionic strength.

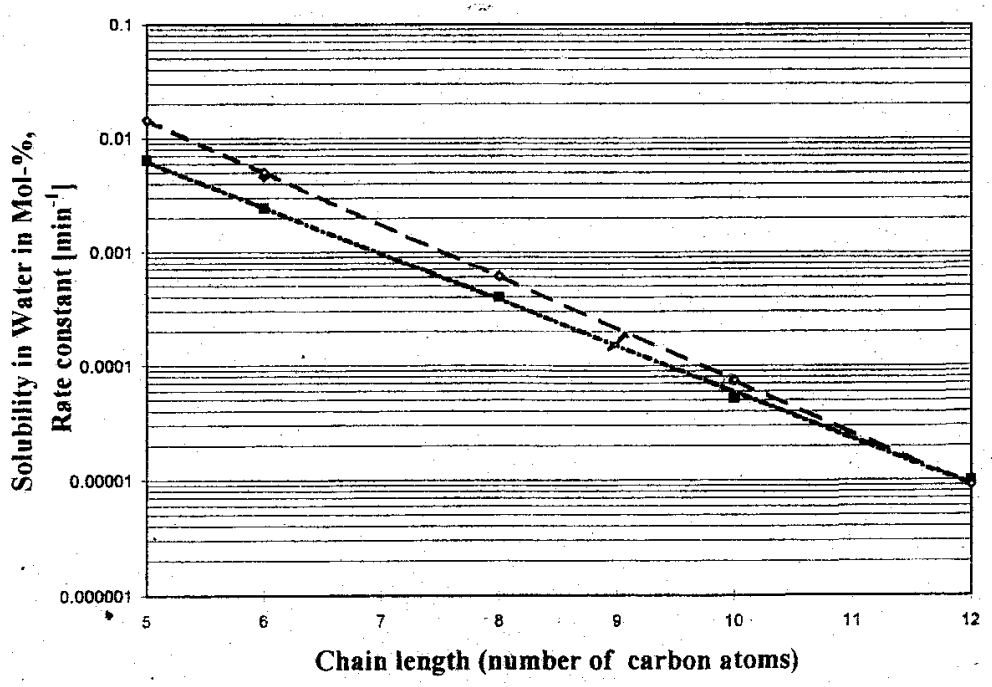


Figure 2. Solubility of alkenes in water (y, solid line) and rate constants for the Ruhrchemie/ Rhône-Poulenc hydroformylation process (◊, dashed line) as a function of substrate chain length (Copyright Wiley-VCH Verlag GmbH & Co. KGaA, reproduced with permission).[9]

4.3.5 Solutions for higher alkenes

For higher alkenes (>C8) it is not possible to separate the aldehyde products by distillation due to thermal sensitivity of the aldehydes, making a biphasic system highly desirable. However, as described above, the low solubility of higher alkenes in water causes very low turnovers in a process similar to that performed by Ruhrchemie/Rhône-Poulenc. Several research groups have studied the possibility to engineer a suitable biphasic system which is effective for the hydroformylation of higher alkenes.[9] To increase mixing, shockwave and ultrasound reactors were engineered leading to a rate enhancement by a factor of 1.5-2. Various monophosphine and diphosphine analogues of TPPTS were found more effective than TPPTS in terms of selectivity and activity.[8-9, 11] The downside is that these phosphine

ligands are much more expensive than TPPTS. Addition of other cheap phosphines like triphenylphosphine monosulfonate (TPPMS) and triphenylphosphine disulphate (TPPDS) or even triphenylphosphine (TPP) to the reaction mixture is very effective at increasing the activity but also significantly increases rhodium leaching because of the increased solubility of these ligands into the organic layer compared to TPPTS.[9, 12] Changing the alkali counter ion from Na^+ to Li^+ or Cs^+ proved to have little influence on the rate and selectivity and a negative effect was found using Al^{3+}.[9, 13] Using additives to enhance phase transfer like cationic detergents,[9, 14] poly(ethylene glycol) PEG derivatives,[15] modified cyclodextrins,[16] crown ethers[9] and calixarenes[17] was often found to be very effective at increasing the rate, but this is accompanied by a decrease in selectivity and complication of the phase separation process. The rate can also be enhanced by addition of polar co-solvents. Alcohols and methanol in particular proved very effective due to enhancement of the solubility of the alkenes into the aqueous layer. The disadvantage is that these solvents also increase the solubility of the catalyst in the organic layer increasing catalyst leaching.[9] These examples show that improvement of the biphasic system using Rh/TPPTS is far from straightforward.

4.2.6 Bioinspired solutions

Various proteins are well known to bind hydrophobic organic compounds, including important substrates for hydroformylation. Artificial metalloenzymes based on such proteins have significant potential as catalysts for the biphasic hydroformylation of poorly water-soluble substrates. This is because the protein can bring substrate molecules into the water layer, while the insolubility of the protein in the organic phase prevents leaching of the catalyst into the organic layer. Moreover, the protein may be used to induce selectivity on the reaction. Some studies have been performed on protein-based rhodium catalyst systems for hydroformylation. Marchetti *et al.* published two papers on using a system involving a dative adduct of human serum albumin (HSA) and $[Rh(acac)(CO)_2]$.[18] HSA is known to be a transporting protein for various hydrophobic compounds and is also known to bind metal ions. High conversion and turnovers were found for several alkene substrates including 1-octene in a biphasic system using pentane as the organic phase. However, selectivity was poor providing a nearly one to one mixture of branched to linear aldehyde. For

substrates which typically afford high branched selectivity the system showed good regioselectivity, for example 90% branched selectivity was obtained for styrene. Their system was most extensively studied using styrene as substrate. Optimal conditions were found to be a reaction temperature of 60 ºC, a high synthesis gas pressure (70-80 atm), near neutral pH of the water layer and 20 molecules of rhodium per HSA. HSA demonstrated high stability under the reaction conditions, showing only minimal amounts of grey precipitate after the reaction. The catalyst solution could be reused for more than 10 cycles as the rhodium leaching was found to be minimal (8 ppm Rh found in the organic layer after the reaction, which corresponds to about 2-3% of total rhodium). When papain and egg albumin were used as alternative proteins, far more leaching was observed and in case of the latter also total precipitation of the protein. Complexation of the rhodium to HSA was analysed by CD spectroscopy giving rise to an intense broad positive signal at 350 nm. This signal was attributed to rhodium-protein adduct. More recent studies showed that the rhodium most likely interacts with the sulfur atoms of the cysteines present in HSA and that cysteine-rhodium complexes are highly active catalysts.[19] The system was also successfully applied in the synthesis of the fragrance Florhydral®.[20]

Another protein based catalyst system was created in the group of Kazlauskas, who replaced the zinc in the active site of carbonic anhydrase by rhodium.[21] This catalyst system was engineered to selectively bind one rhodium metal in a histidine containing active site. This was done by removing other potential binding sites by replacement of various amino acids using site directed mutagenesis. Remarkable selectivity was observed for the rhodium catalysed hydroformylation of styrene, obtaining linear over branched ratios of 8 whereas free rhodium typically gives a ratio of 0.2. This reaction was performed under relative mild conditions using room temperature and 20 bar of synthesis gas. Good yields are reported but no information is given about the amount of substrate used, so the actual activity of this system remains unknown. Unfortunately, also no information is provided about the substrate addition process. However, the catalyst was also applied in rhodium catalysed hydrogenation of *cis* and *trans* stilbene[22] and, assuming a similar experimental setup is used for the hydroformylation and for their hydrogenation, the substrate is partly dissolved in the reaction mixture using acetonitrile. This catalyst system was found less stable

compared to the HSA/Rh complexes described above as the protein was reported to precipitate after one hour.

A rhodium hydroformylation catalyst system will be presented in this chapter, which is based on SCP-2L phosphine conjugates. This catalyst is developed according to the concept shown in Figure 3. A system is envisaged were the protein binds the apolar substrate therefore encapsulating it and presenting the covalently bound rhodium phosphine system with a selectivity inducing environment. Both bioinspired examples outlined above use native dative interactions between amino acid residues of the protein to bind the catalytically active metal centre, making them very different from the system demonstrated in this thesis. But, although, the concept explained here uses phosphine ligands to bind the phosphine metal centre, other amino acid side chains of the proteins will have influence on the catalytic reaction. This is important for the concept as this allows for optimization of the systems by site directed mutagenesis and directed evolution in the future.

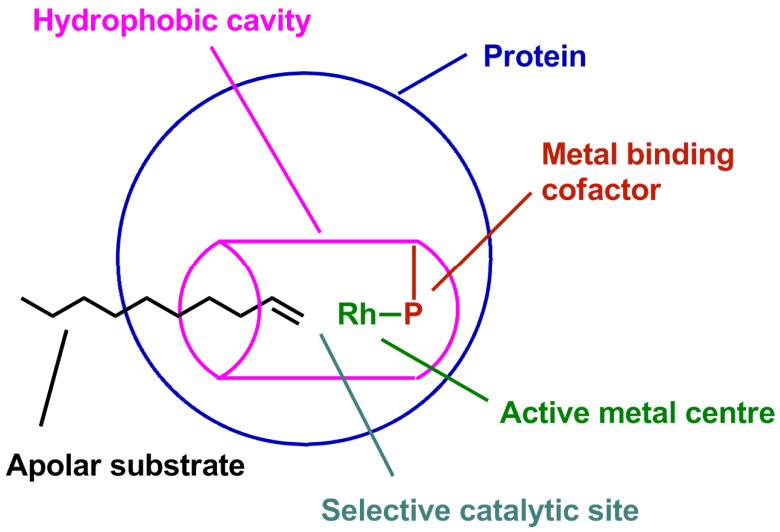


Figure 3. Concept for shape selective hydroformylation catalyst.

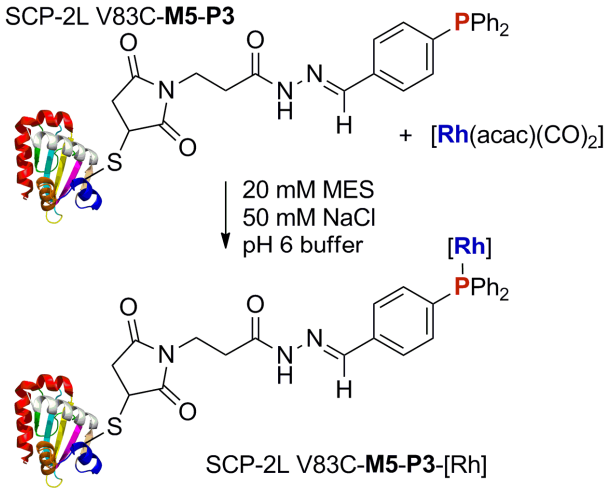
4.3 Results and discussion

4.3.1 Rhodium coordination to phosphine containing SCP-2L

4.3.1.1 Synthesis of rhodium-phosphine proteins adducts

In the previous chapter the synthesis of phosphine modified SCP-2L ligand is described. The formation of rhodium-phosphine complex modified proteins had to be studied for successful application of our approach in hydroformylation. Direct

modification of **M5** modified proteins with preformed rhodium complexes was attempted, which would avoid an additional protein modification step (rhodium coordination) and the extra set of washings this would require. However, this approach was unsuccessful (see appendix A7 for more details). Therefore, we explored the coordination of rhodium to phosphine-modified protein. Mixing SCP-2L V83C-**M5**-**P3** with 1 equivalent of [Rh(acac)$(CO)_2$] for 16 h, resulted in nearly full conversion of the phosphine-modified protein to a mixture of species according to LCMS (ES^+) analysis, with masses ranging from 13933-14100 Da (hereafter collectively referred to as SCP-2L V83C-**M5**-**P3**-[Rh]) (Scheme 3, Figure 4). The strongest signal within the range of species correlates closely to the calculated mass for SCP-2L V83C-**M5**-**P3** modified with a rhodium carbonyl fragment (i.e. SCP-2L V83C-**M5**-**P3**-RhCO) (Table 1 entry 1). Other signals within the mass range belong most likely to rhodium species with other coordinating ligands like water and acetonitrile (the latter from the elution solution of the LCMS).



Scheme 3. Complex formation between SCP-2L-V83C-**M5**-**P3** and [Rh(acac)(CO_2)].

Two other signals with m/z 13572 and 13846 Da are assigned to the oxide signal (methionine oxidation) of some residual SCP-2L V83C-**M5** (calculated 13572.9 Da) and oxidized SCP-2L V83C-**M5**-**P3** (calculated 13845.2 Da). These signals are observed because of a significantly reduced ionization efficiency of SCP-2L V83C-**M5**-**P3**-[Rh] compared to SCP-2L V83C-**M5(-P3)**, which causes to observation of small residual amounts of oxidized SCP-2L V83C-**M5** and SCP-2L V83C-**M5**-**P3**.

The reaction between SCP-2L A100C-**M5**-**P3** and [Rh(acac)$(CO)_2$] resulted in a similar product distribution, showing high conversion to SCP-2L A100C-**M5**-**P3**-

RhCO (Table 1 entry 2). Interestingly, LCMS(ES^+) analysis of a sample taken 15 minutes after the start of the reaction showed a completely different product distribution. Several species were observed corresponding to the coordination of two or three rhodium centres to the SCP-2L A100C-**M5**-**P3**. A sample taken after three hours reaction time showed the product distribution as observed after 16 hours for SCP-2L A100C-**M5**-**P3** upon LCMS(ES^+) analysis, with no trace of the rhodium species observed after 15 minutes. These results suggest that the rhodium centres initially bind to the surface of SCP-2L-A100C-**M5**-**P3** forming statistical mixtures of rhodium modified protein before they coordinate to the phosphorous ligand.

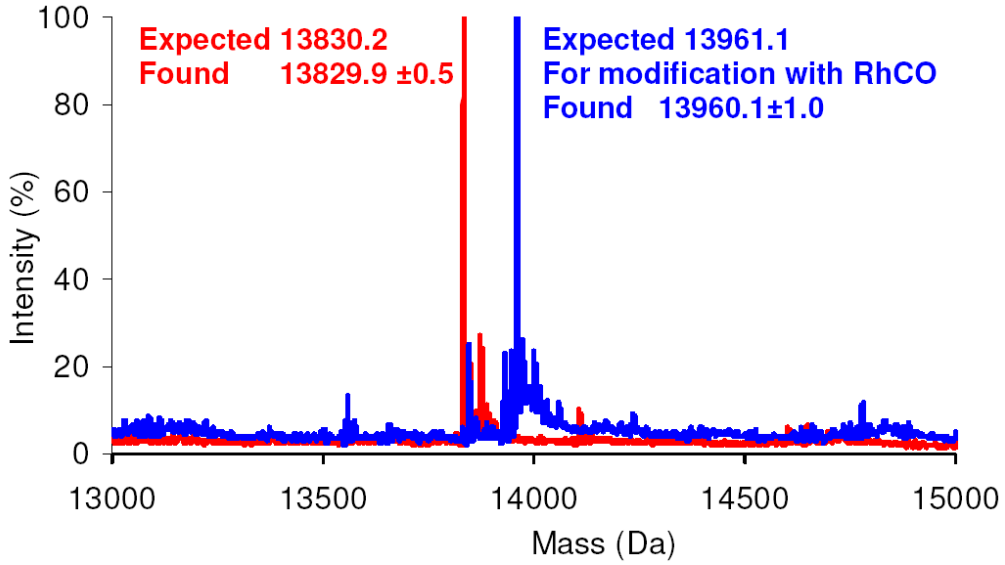


Figure 4. Processed mass spectra of SCP-2L V83C-**M5**-**P3** before (red) and after modification with [Rh(acac)$(CO)_2$] (blue) (Copyright Wiley-VCH Verlag GmbH & Co. KGaA. Reproduced with permission).[23]

Large scale synthesis of rhodium complexes for catalysis experiments allowed for the determination of the yield of the rhodium coordination reaction. These experiments showed that more than 90% of the protein remains in solution after the reaction and subsequent washings and concentration. Moreover, the mass spectrum obtained by LCMS (ES^+) of the complex after 16 hours is essentially identical to that obtained after 3 hours indicating reasonable stability of the formed metalloprotein.

The reaction between SCP-2L V83C-**M5**-**P5** and [Rh(acac)$(CO)_2$] provided a major species corresponding to complexation of RhCO (Table 1 entry 3). The only difference compared to the **P3** conjugates was that a strong signal corresponding to

residual double oxidized SCP-2L V83C-**M5**-**P5** was observed (calculated 14147.5 Da, observed 14147.6 Da) and a medium signal corresponding to oxidized SCP-2L V83C-**M5**-**P5**-RhCO (calculated 14262.4, found 14261.9). Similar products were observed upon reacting SCP-2L A100C-**M5**-**P5** with [Rh(acac)$(CO)_2$] (Table 1 entry 4).

Table 1. Most intense signals observed in the mass-spectra obtained after the reaction between phosphine modified proteins and [Rh(acac)$(CO)_2$].[a]

Entry	Phosphine modified protein	Calculated mass +RhCO	Main signal found[b]
1	SCP-2L V83C-**M5**-**P3**	13961.1	13960.1
2	SCP-2L A100C-**M5**-**P3**	13991.0	13991.1
3	SCP-2L V83C-**M5**-**P5**	14246.4	14247.6
4	SCP-2L A100C-**M5**-**P5**	14276.3	14276.3

a) Reaction conditions: 0.1 mM phosphine modified protein, 1 equivalent [Rh(acac)$(CO)_2$] in degassed 20 mM MES 50 mM NaCl pH 6 buffer, 16 h at room temperature, and analyzed by mass-spectroscopy b) Most intense signal found by analysis of the obtained mass spectrum (ES+).

The reaction between SCP-2L V83C-**M5**-**P3** and [Rh$(MeCN)_2$(cod)]BF_4 also yielded a range of rhodium adducts. Signals for the addition of rhodium and additional corresponding oxidation signals were observed upon LCMS(ES^+) analysis as well as signals belonging to oxidized SCP-2L V83C-**M5**-**P5**. The observation of phosphine oxidation of this ligand was not surprising as the significant oxygen sensitivity of this ligand was already observed in bioconjugation experiments of **P5** to **M5** modified proteins.

4.3.1.1 Spectroscopic studies of the artificial rhodium-phosphine proteins

To confirm the formation of a rhodium phosphine complex and to investigate the effect of this modification on the protein structures several spectroscopic techniques were used. The ^{31}P-NMR spectrum of SCP-2L V83C-**M5**-**P3**-[Rh] was recorded using the same parameters as for the ^{31}P-NMR spectra obtained for the phosphine modified proteins (Figure 5). The spectrum shows signals between 3 and 0 ppm,

which have previously been assigned to phosphorylation of the protein, and a broad signal between 50 and 30 ppm with sharp signals at 44.8 and 43.8 ppm. The chemical shifts of these latter signals are consistent with phosphorus-rhodium complexes. The distance between the sharp signals (~200 Hz) is consistent with a phosphorus-rhodium coupling, however due to the low signal to noise ratio this is hard to confirm. The broad signal suggests multiple phosphorus-rhodium species are present, which is in line with the observation of a range of different rhodium-adducts by mass-spectroscopy. The absence of signals between ï 3 and ï 7 ppm indicates complete conversion of the protein-bound phosphines to rhodium-complexes. CO and the phosphine are the only coordinating ligands identified by mass and NMR spectroscopic analysis of the obtained complex. These will occupy only two of the possible five coordination sites available on rhodium and therefore it is assumed that amino acid side chains contribute to rhodium complexation. As there are many suitable amino acid side chains present, this could be one of the reasons for the observed mixture of rhodium species.

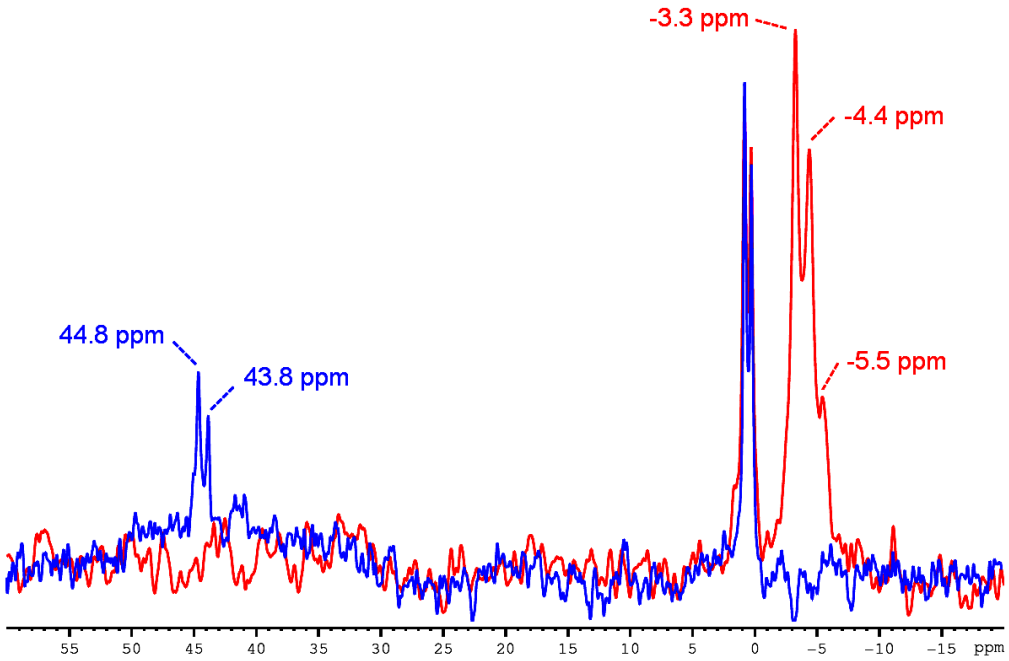


Figure 5. {^{1}H}-^{31}P-NMR spectra of SCP-2L V83C-**M5-P3** (red) and SCP-2L V83C-**M5**-**P3**-[Rh] (blue) (Copyright Wiley-VCH Verlag GmbH & Co. KGaA. Reproduced with permission).[23]

A watergate ^{1}H-NMR spectrum was recorded of SCP-2L V83C-**M5**-**P3**-[Rh] to study the effect of the rhodium coordination on the protein fold (Figure 6). The obtained spectrum is similar to the spectrum of SCP-2L V83C-**M5**-**P3** except for some extra signals which can be identified as residual DMF (7.70, 2.78 and 2.63 ppm, literature: 7.92, 3.01 and 2.85 ppm)[24] used as solvent for the addition of [Rh(acac)$(CO)_2$]. Also, a slight loss in resolution is observed which could be due to the formation of some precipitate during the measurement. Overall the protein structure seems unaffected by the addition of [Rh(acac)$(CO)_2$], which is very promising for application in catalysis.

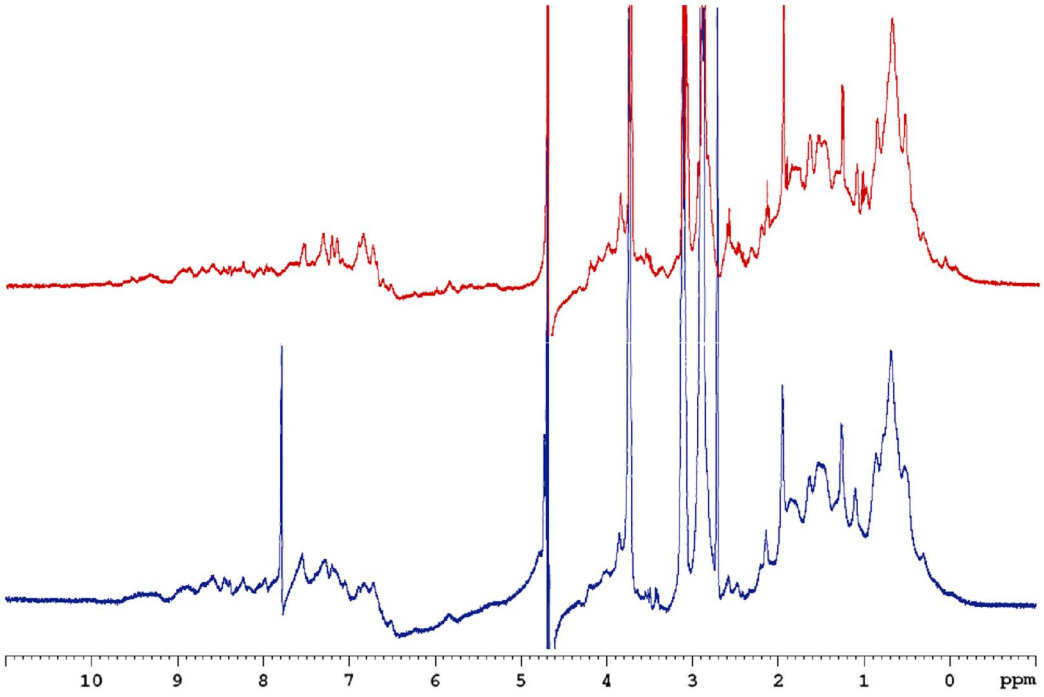


Figure 6. Watergated ^{1}H-NMR spectra of SCP-2L V83C-**M5**-**P3** (red) and SCP-2L V83C-**M5**-**P3**-[Rh] (blue).

Attempts were made to study the structural integrity of rhodium adducts of the phosphine modified proteins by near-UV CD-spectroscopy. However, only an intense broad signal with a maximum at 301 nm was observed in the near UV region of SCP-2L A100C-**M5**-**P3** (Figure 7). This spectrum is very similar to those obtained by Marchetti *et al.* however with a blue-shift of the signal of 50 nm.[18a] Because the watergate ^{1}H-NMR spectrum of the SCP-2L V83C-**M5**-**P3**-[Rh] complex showed the fold of the protein is not significantly changed upon rhodium incorporation, it is assumed that the dramatic change in the CD signal corresponds to rhodium-protein

interactions or phosphine-rhodium interactions which are known to give UV-absorptions in this region.[25]

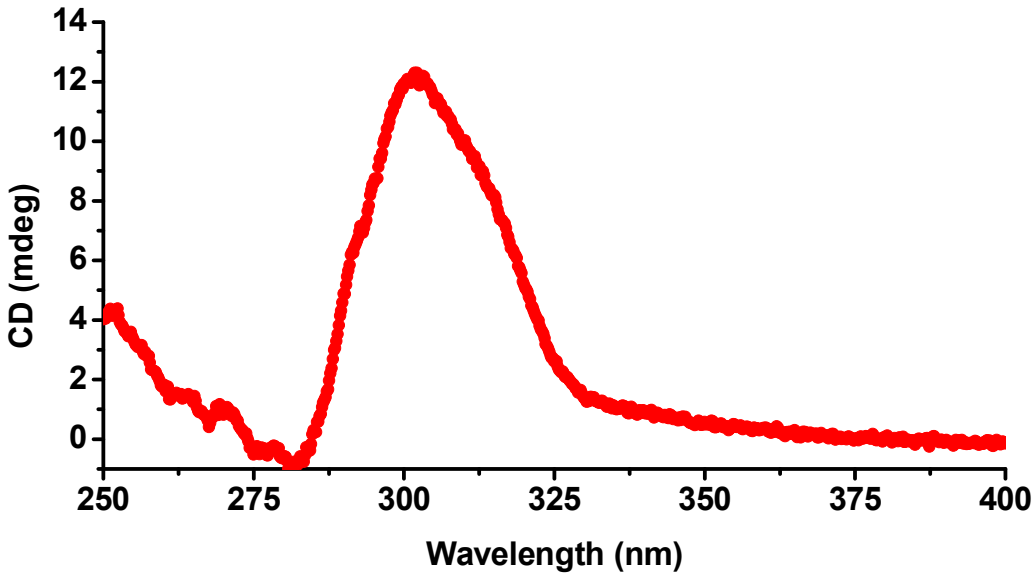
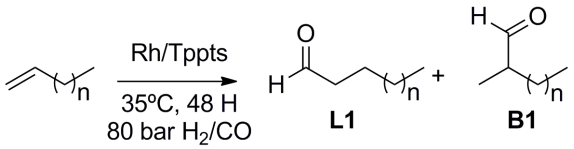


Figure 7. Near-UV CD-spectrum of SCP-2L A100C-**M5**-**P3**-[Rh].

4.3.2 Hydroformylation using Rh/TPPTS

4.3.2.1 Hydroformylation of linear α-olefins

The conditions for the biphasic hydroformylation of linear Ŭ-olefins for using rhodium/TPPTS catalyst systems are well established, as shown in the introduction of this chapter. However, certain conditions and aspects of the process needed to be altered for the application of proteins in these reactions. The rhodium/TPPTS catalyst system was selected as model catalytic system to obtain reference data for the protein based catalysts under the selected conditions (Scheme 4). Initial experiments using the Rh/TPPTS catalyst showed that the complex of TPPTS and [Rh(acac)$(CO)_2$] had to be prepared in a separate reaction vessel. This is done to prevent undissolved rhodium to accumulate on the glass walls, which subsequently dissolves in the organic layer, where it will catalyse unselective hydroformylation. No difference in activity and selectivity was observed using degassed water or 20 mM MES 50 mM NaCl pH 6 buffer. This is important as this indicated that the use of this buffer to stabilize the protein and maintain the pH stable during the reaction will not affect the catalysis results.



Scheme 4. Hydroformylation of linear Ŭ-olefins using Rh/TPPTS.

Table 2. Hydroformylation of linear Ŭ-olefins using Rh/TPPTS.[a]

Entry	Catalyst[b]	Substrate[c]	**L1**/**B1**[d]	TON[d]
1	[Rh(acac)(CO)$_2$]	1-decene	1.27	369
2	Rh:TPPTS 1:1	1-hexene	1.29 (±0.02)	615 ±256
3	Rh:TPPTS 1:1	1-octene	1.26 (±0.01)	427 ±77
4	Rh:TPPTS 1:1	1-decene	1.27 (±0.02)	667 ±170
5	Rh:TPPTS 1:1	1-octadecene	1.27 (±0.01)	223 ±65
6	Rh:TPPTS 1:20	1-hexene	6.31 (±1.71)	16 (±3)
7	Rh:TPPTS 1:20	1-octene	4.55 (±2.00)	~1
8	Rh:TPPTS 1:20	1-decene	1.64 (±0.27)	<1
9	Rh:TPPTS 1:20	1-octadecene	nd	<1

a) 48 hours, 80 bar CO/H_2 = 1:1, 35 ºC, stirred b) Rh (0.3 µmol) as preformed complex with the adequate amount of TPPTS in 0.5 ml 20 mM MES 50 mM NaCl pH 6 c) 0.5 ml substrate added neat containing 9% n-heptane and 1% phenylether as internal standards d) determined by GC, turnover numbers (TON) given in [mol aldehydes/mol cat].

Examples of experiments run using the appropriate conditions with different phosphine to rhodium ratios can be found in Table 2. These results demonstrate the obvious difference in activity and selectivity resulting from the hydroformylation catalysed by leached rhodium species and rhodium-TPPTS complexes dissolved in the water layer. ñFreeò rhodium gives a linear over branched ratio (**L**/**B)** of ~1.27 and high turnover numbers; and similar results are obtained when rhodium and TPPTS are used in a 1:1 ratio (Table 2 entries 1-5). Reproducible results were only obtained if an excess of 20 equivalents of phosphine was used (Table 2 entries 6 and 7). These

experiments also showed the expected selectivity for catalysis by Rh:TPPTS indicating complete prevention of leaching of the rhodium to the organic layer. Reactions with Rh:TPPTS ratios of 1:5 and 1:10 showed variable results either matching the results of Rh:TPPTS 1:1 or Rh:TPPTS 1:20 (data not shown). The rhodium content in the organic layer was determined using ICP to verify the levels of rhodium leaching. The variable activity was found to be directly related to the amount of leached rhodium. The values ranged from 10 to 250 ppm for systems involving Rh:TPPTS 1:10 or lower TPPTS content (here 250 ppm = ~ 0.3 εmol = all rhodium leached to the organic layer). A rhodium content of less than 0.1 ppm was measured for reactions involving rhodium:TPPTS 1:20, indicating that this catalyst system leads to true biphasic catalysis. As expected, the turnover numbers are low for the reactions that are truly catalysed by Rh/TPPTS in the aqueous layer, due to the very low water solubility of the substrates. An effect of the increase in chain length of the alkene substrate can be observed going from 1-hexene to higher alkenes for which nearly no conversion to aldehydes is found (Table 2 entries 5-9). For 1-decene the reaction rate is so low that the traces of rhodium that leach into the organic layer contribute significantly to the total conversion, as shown by the decreased selectivity.

4.3.2.2 High pressure NMR experiments

High pressure NMR experiments were performed to confirm that solubility of the substrates is the cause of the observed low activity of the Rh/TPPPTS system. This experiment should demonstrate if hydride species such as **H1** are formed under the conditions used above (Scheme 2). Especially, the low temperature could affect the formation of such species. A sample containing Rh/TPPTS was continuously scanned under syngas pressure at constant temperature and respective spectra were accumulated for 30 minutes. A low Rh:TPPTS ratio of 1:2 was used to make sure that the signal of the free phosphine would not suppress the signals belonging to the formed complexes.

A selection of these spectra is shown in Figure 8. The initial formed solution shows broad signals at around 32 and 50 ppm, suggesting it contains a mixture of Rh-TPPTS species. The sharp signal at 34.5 ppm is assigned to a small fraction of oxidized TPPTS, which can be observed in the spectrum of the batch of TPPTS used (Figure 8a-b). When synthesis gas pressure was applied new species were observed, which

could be identified using literature reports of similar and related structures.[6, 7b, 7c] After 30 minutes a major species displaying a doublet at 43.3 ppm ($^{1}J_{Rh,P}$ = 155 Hz) was observed, which was identified as [RhHCO(TPPTS)$_3$], the TPPTS analogue of **H1** (Figure 8c).

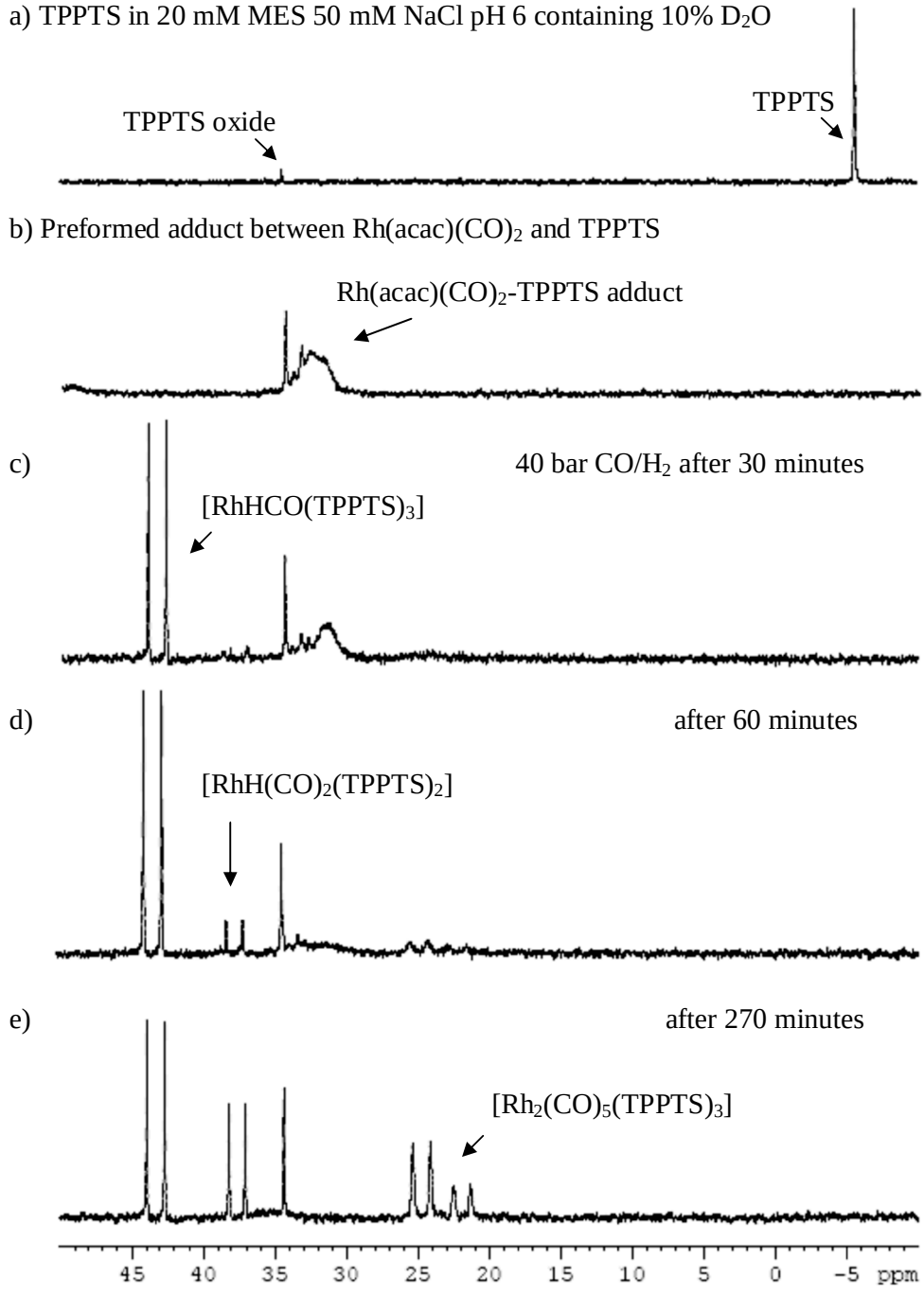


Figure 8. High pressure ^{31}P-NMR experiments using Rh(acac)(CO)$_2$/TPPTS 1:2 under 40 bar CO/H$_2$ at 35 ºC in 20 mM MES 50 mM NaCl pH 6 containing 10% D$_2$O.

After an hour a new species displaying a doublet ($^{1}J_{Rh,P}$ = 136 Hz), at 37.6 ppm was observed (Figure 8d) which was identified as $[RhH(CO)_2(TPPTS)_2]$, whose relative concentration slowly increases in time. In addition, a weak signal consisting of two broad doublets at 24.8 and 21.8 ppm ($^{1}J_{Rh,P}$ = 152 and 156 Hz respectively) with a 2:1 ratio was observed, whose relative intensity increased over time (Figure 8e). According to published data for complexes with the general formula of $[Rh_2(CO)_5(phosphine)_3]$ this species was identified as the TPPTS analogues of these complexes.[6b, 7b, 7c]

In addition, a ^{1}H-NMR experiment (with solvent suppression) was run after three hours (Figure 9) This spectrum clearly showed two sets of hydride multiplet signals between ï 9.5 and ï 9.7 ppm and ï 10.1 and ï 10.4 ppm with a relative intensity similar to the ratio observed by ^{31}P-NMR after this time between $[RhHCO(TPPTS)_3]$ and $[RhH(CO)_2(TPPTS)_2]$.These experiment clearly showed the formation of typical hydride species that are involved in hydroformylation within 30 minutes at 35ºC and 40 bar H_2/CO. These results confirm that the slow reaction rate found for Rh:TPPTS is not due to inefficient formation of catalytically active hydride species.

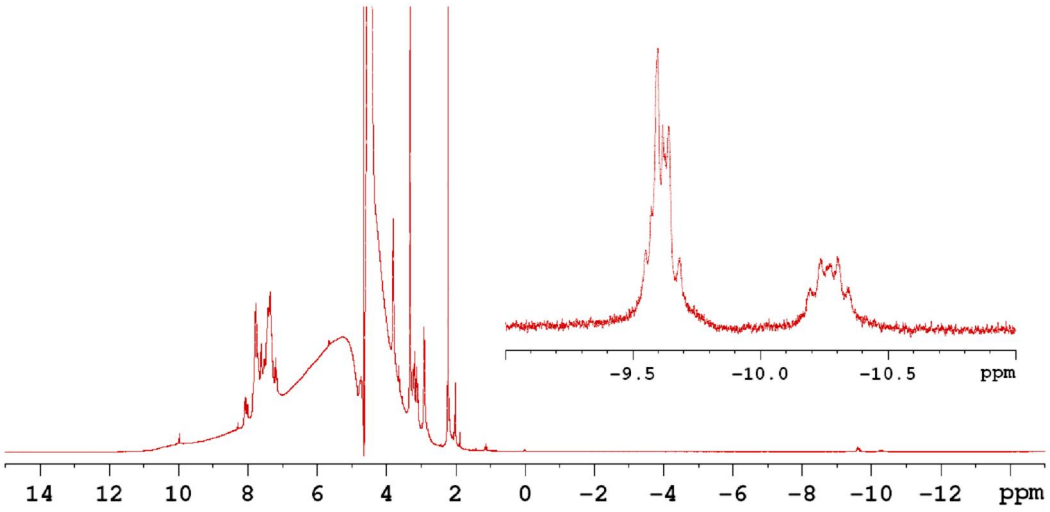


Figure 9. ^{1}H-NMR spectrum (with solvent suppression) of Rh:TPPTS 1:2 after 300 min exposure to 40 bar CO/H_2 at 35 ºC in 20 mM MES 50 mM NaCl pH 6 containing 10% D_2O.

4.3.3 Phosphine rhodium artificial enzyme catalysed reactions

4.3.3.1 Hydroformylation of linear α-olefins

Initial hydroformylation experiments were performed using **P3** modified proteins. These phosphine conjugates were selected because rhodium complexation had been tested for these ligands. Initial experiments were performed using a 1:1 modified protein to rhodium mixture as the catalyst for the hydroformylation of linear Ŭ-olefins, under the conditions described above for the Rh/TPPTS system. This catalytic system showed very high activity for the hydroformylation of linear alkenes, but low **L**/**B** ratios and very poor reproducibility. These results were attributed to different amounts of rhodium leaching in every run. This was confirmed by determination of the rhodium content in the organic layer by ICP, which was performed for several reactions involving 1-decene. Values up to 50 ppm were measured, and it was found once again that the rhodium content showed a direct relationship with the found activity. Differences in rhodium leaching are expected in 1:1 Rh:phosphine mixtures in biphasic catalysis if in one experiment a slight excess rhodium might be present while in other a slight excess of phosphine modified protein is used. To prevent this, we decided to use an excess of ligand. Reproducible catalysis results were obtained using a 2:1 ratio of modified protein and rhodium. Following several hydroformylation reactions of 1-decene with SCP-2L V83C-**M5**-**P3**-[Rh] as catalyst the rhodium content in the organic layer was measured. Values lower than 0.5 ppm were obtained, confirming very little to no influence of the rhodium leaching on the results. An overview of the results of all hydroformylation experiments involving protein based catalysts with 2:1 ratio of phosphine modified protein to rhodium is given in Table 2.

Triton-X-100 was added to the reactions involving the protein constructs to prevent the formation of large amounts of precipitate. This indeed largely prevented the formation of precipitation and thereby simplified the workup of the reaction. The addition of Triton-X-100 had no noticeable effect on the selectivity and activity, but did make the results slightly more reproducible (compare entries 4 and 6 with 5 and 7). SCP-2L A100C-**M5**-**P3**-[Rh] showed very good conversion and selectivity for 1-hexene and 1-octene (entries 1 and 3). The activity and selectivity drop when 1-decene is used as substrate and even more using 1-octadecene (entries 7 and 10). SCP-2L V83C-**M5**-**P3**-[Rh] appears less selective than SCP-2L A100C-**M5**-**P3**-[Rh]

for 1-octene and 1-decene (compare entries 2 and 5 with 3 and 7). In general, the selectivity of SCP-2L V83C-**M5**-**P3**-[Rh] seems to be lower than that found for the A100C construct, however SCP-2L V83C-**M5**-**P3**-[Rh] seems to retain its selectivity better moving to longer alkene substrates (compare entries 9 and 10). Leaching might well contribute in various amounts to the obtained selectivity and activity at the low turnover numbers found for 1-octadecene. The selectivity is however significantly higher than that observed for leached rhodium. Interestingly, also the decrease in activity between 1-octene to 1-decene seems more pronounced for SCP-2L A100C-**M5**-**P3**-[Rh] compared to SCP-2L V83C-**M5**-**P3**-[Rh]. Most importantly, the overall activity was found to be significantly higher when compared to those obtained for Rh/TPPTS (compare Table 2 entries 1-4 with Table 3, visualised in Figure 10).

Table 3. Hydroformylation of linear Ŭ-olefins using Rh-Protein constructs.[a]

Entry	SCP-2L-catalyst[b]	Substrate[c]	**L1**/**B1**[d]	TON[d]
1	A100C-**M5**-**P3**-[Rh]	1-hexene	5.31 (± 1.1)	196 (±71)
2	V83C-**M5**-**P3**-[Rh]	1-octene	3.95 (± 0.20)	57 (±12)
3	A100C-**M5**-**P3**-[Rh]	1-octene	6.72 (± 0.18)	124 (±23)
4[e]	V83C-**M5**-**P3**-[Rh]	1-decene	3.31 (± 0.29)	35 (±11)
5	V83C-**M5**-**P3**-[Rh]	1-decene	3.44 (±0.06)	34 (±11)
6[e]	A100C-**M5**-**P3**-[Rh]	1-decene	3.71 (±0.69)	26 (±9)
7	A100C-**M5**-**P3**-[Rh]	1-decene	4.15 (±0.41)	24 (±8)
8	PYP-**M5**-**P3**-[Rh]	1-decene	1.51 (±0.02)	61 (±10)
9	V83C-**M5**-**P3**-[Rh]	1-octadecene	3.12 (±0.19)	5 (±0.4)
10	A100C-**M5**-**P3**-[Rh]	1-octadecene	2.08 (±0.51)	12 (±8)

a) 48 hours, 80 bar CO:H_2 1:1, 35 ºC stirred b) Protein-P:Rh = 2:1 (0.3:0.15 ε mol) in 0.5 ml 20 mM MES 50 mM NaCl pH 6 containing 0.46 mM Triton-X-100 c) added 0.5 ml substrate containing 9% *n*-heptane and 1% phenylether as internal standards d) determined by GC, turnover number (TON) given in [mol aldehydes/mol Rh] e) no Triton-X-100 was added

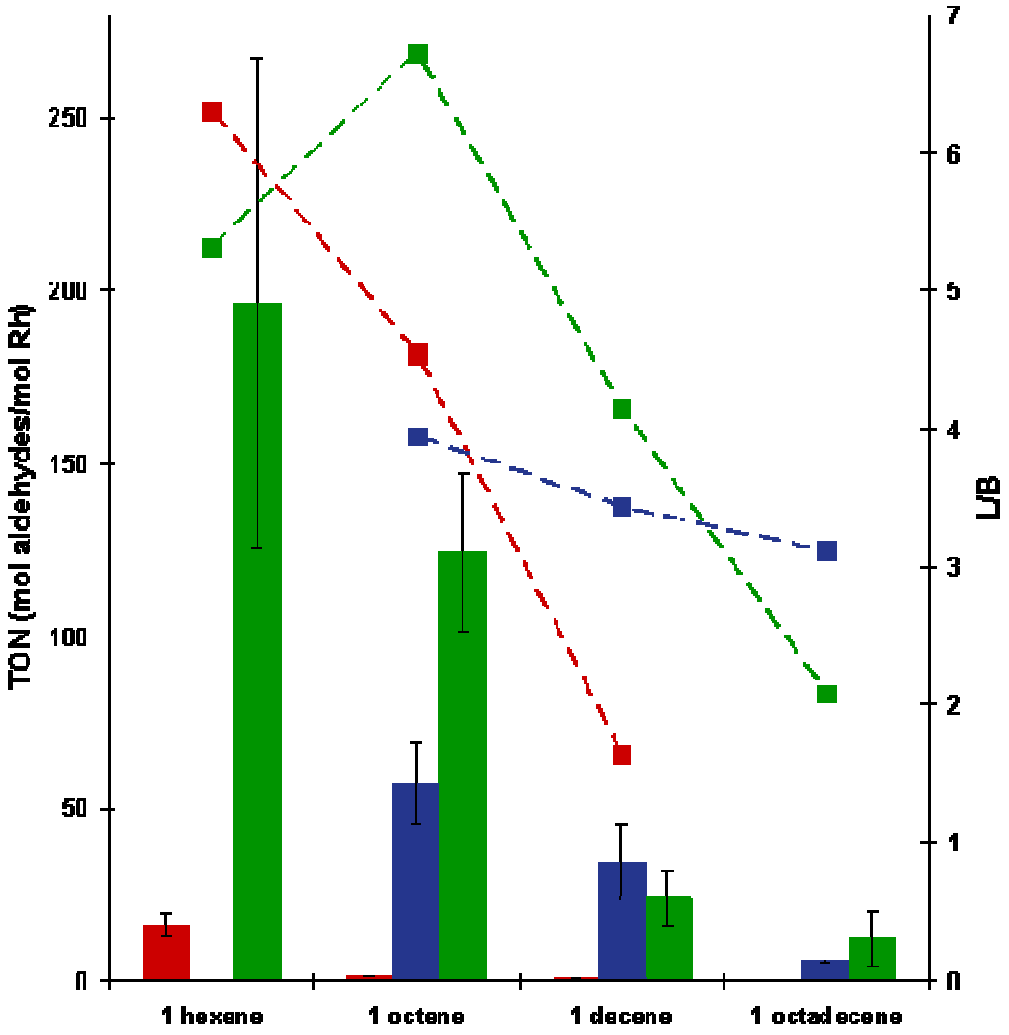


Figure 10. Comparison of the activity (bars) and selectivity (dots and dashed lines) of Rh:TPPTS 1:20 (red), SCP-2L V83C-**M5**-**P3**-[Rh] (blue) and SCP-2L A100C-**M5**-**P3**-[Rh] (green) in the biphasic hydroformylation of alkenes.

The catalyst based on the A100C mutant of SCP-2L showed a more than 100 fold increased activity for all tested substrates compared to Rh/TPPTS. This high activity was accompanied by a high selectivity, which makes these results even more outstanding. That the increased activity for higher alkenes is truly the result of shape selective biphasic catalysis is so far confirmed by the ICP data for 1-decene and the observed selectivity which exclude a significant contribution of hydroformylation catalysed by free rhodium. Although these data are very promising, additional proof is required to corroborate that the high selectivity and activity are truly caused by the

system as we have designed it to work. This means whether or not the substrate is bound by the tunnel exposing the terminal carbon of the alkene followed by selective linear hydroformylation and product release/replacement with a new molecule of substrate should be determined

The unique cysteine of PYP was also modified with **M5**, **P3** and 0.5 equivalents of rhodium using the same methodology as the SCP-2L mutants. This protein only contains a small hydrophobic pocket and does not bind linear aliphatic molecules. The obtained adduct was applied in the hydroformylation of 1-decene to act as a control reaction. When this protein based catalyst was applied also good activity was observed, however selectivity was low. This indicated that either rhodium dissociates relatively easily from this catalyst giving unselective ñfreeò rhodium catalysis or that this catalyst does not induce linear selectivity as SCP-2L does but does still enhances the reaction rate. Unfortunately no ICP data were obtained for this reaction to confirm the first hypothesis. Either way, this experiment demonstrates the importance of the choice of protein template for catalyst development.

The results for PYP are very similar to those obtained for dative systems by Marchetti *et al.* although the differences in reaction setup and conditions do not allow for direct comparison of the results.[18] Various dative Rh-proteins adducts were tested for hydroformylation of 1-octene showing different activities but all with low selectivity (e.g. **L1**/**B1** = 1-1.3). This indicated that other rhodium-amino acid complexes are capable of increasing the activity of 1-octene biphasic hydroformylation catalysis but do not lead to increased selectivity. The activity found using SCP-2L A100C-**M5**-**P3**-[Rh] and SCP-2L V83C-**M5**-**P3**-[Rh] is high similar to the Rh-HSA catalyst developed by Marchetti *et al.* but also high linear selectivity is observed. This observation further underlines the significance of the results obtained using our SCP-2L based catalysts.

4.3.3.2 The fate of the catalyst

Some studies were performed to analyse the effect of the catalytic conditions on the protein catalysts. The amount of remaining soluble catalyst after the reaction was analysed by Bradford assays after separation of the precipitate from the soluble protein by centrifugation. The percentage of protein that remained in solution varied

significantly when no triton-X-100 was added. When triton-X-100 was added the loss of protein could be improved to <50%, but the actual values did still vary.

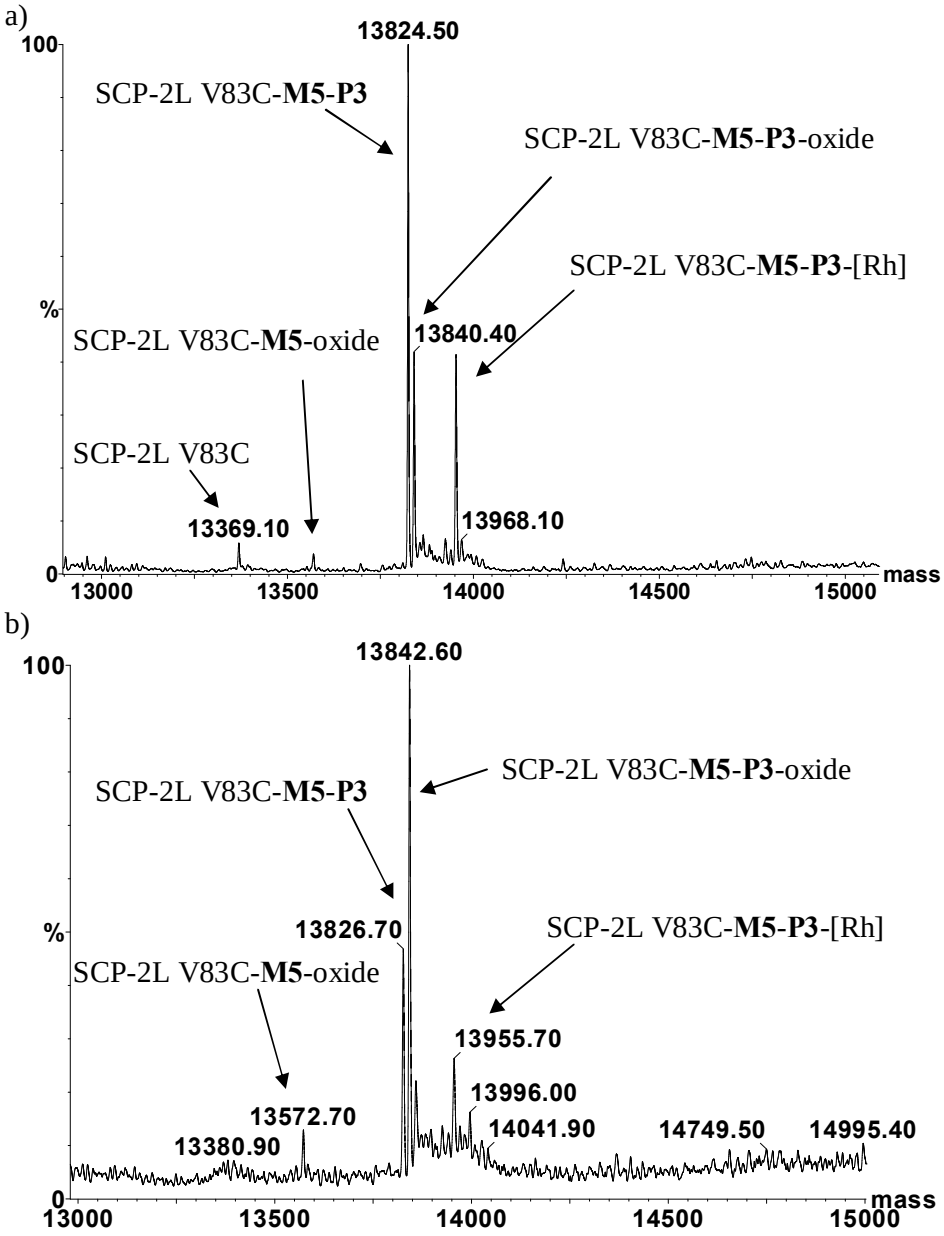


Figure 11. Processed mass-spectra of SCP-2L V83C-**M5**-**P3**-[Rh] (protein-P:Rh 2:1) a) before and b) after catalysis (both spectra are not calibrated, so mass values have shifted slightly)

Interestingly, only a small difference in the reactivity and activity was found between reactions with varying amounts of protein precipitate. In general, slightly lower selectivity and larger variation was found in the activity in reactions were more

precipitate was observed. The precipitate is not expected to be active, which suggests that the precipitate may only be formed near the end of the reaction. However, experiments to confirm this will be required. The water layer containing the soluble protein after the reaction when no triton-X-100 was used was analysed by LCMS(ES^+) (Figure 11). Reactions containing triton-X-100 can not be analysed by LCMS as the triton-X-100 suppresses the protein signal. Although some of the initial rhodium complex was still present, the main product after catalysis was usually found to be the protein modified with an oxidized phosphine. However, as the reaction work-up and LCMS analysis are not performed under inert atmosphere it is unclear if the oxidation occurs during the catalytic runs. The signal to noise ratio of the spectrum after catalysis is clearly lower than of the spectrum before catalysis which is caused by the loss of a significant amount of protein due to precipitation. This explains the slight increase in intensity of the signal belonging to impurities like the trace of oxidized SCP-2L V83C-**M5**. This modified protein appears to be significantly more stable compared to the phosphine modified proteins and thus remains in solution during the reaction. Overall, phosphine oxidation seems the main decomposition pathway for the rhodium-phosphine modified proteins catalysts presented in this work. More research is required to determine whether this happens during the reaction or during the work-up. It remains unclear what happens to the rhodium after phosphine oxidation

4.3.4 Other alkene substrates

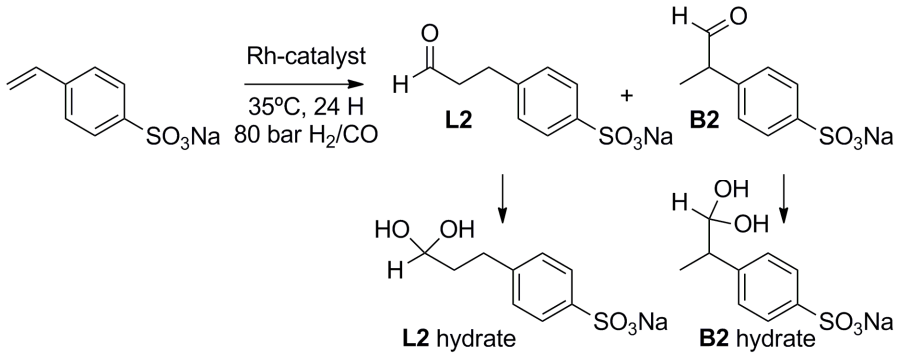
4.3.4.1 Water soluble substrates

4-Styrenesulfonate and *N*-allylacetamide were investigated as alternative substrates for the hydroformylation using our protein based catalysts (see Scheme 5 and Scheme 7). These substrates are water soluble and allow for a true homogeneous reaction. 4-styrenesulfonate was selected because of the high branched selectivity that is usually found, which gives the opportunity of asymmetric hydroformylation. The linear product of the hydroformylation product of *N*-allylacetamide is an important intermediate in the synthesis of the human sleep regulating hormone melatolin.[26] This makes it an interesting substrate for SCP-2L based hydroformylation systems, as enhanced linear selectivity is expected by the use of the protein tunnel to induce shape selectivity. An added advantage is that because the rhodium precursor

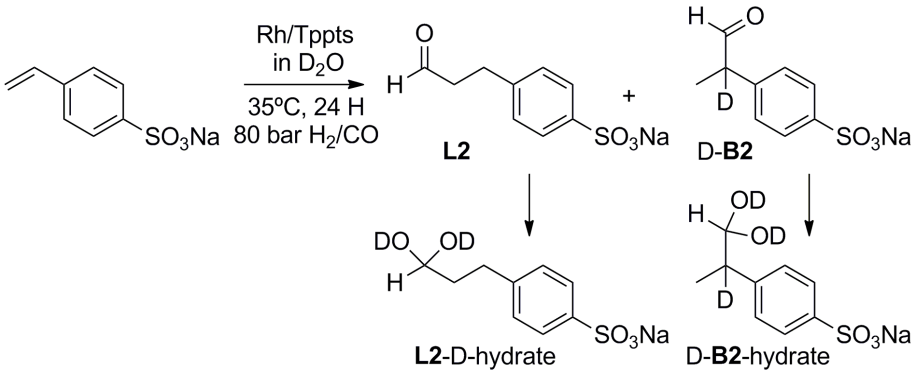
[Rh(acac)$(CO)_2$] is essentially insoluble and thus inactive in water these reactions will experience less influence of catalysis by ñfreeò rhodium species in comparison to the biphasic systems. Rh/TPPTS was again chosen as model catalyst systems for comparison to the protein based catalysts. Lower rhodium to TPPTS ratios can be used for these reactions because no problems caused by the leaching of rhodium are expected. This is an advantage as lower rhodium:TPPTS ratios generally give higher activity due to a positive effect in the equilibrium between **H1** and **H2** (Scheme 2). The reactions were successfully performed using a 5:1 ratio between TPPTS and [Rh(acac)$(CO)_2$]. The conditions were chosen to allow the application of the protein based catalysts, i.e. buffered solutions (preferably 20 mM MES 50 mM NaCl pH 6) and low temperature (35 ºC). The products can be analysed by ^{1}H-NMR using previous published NMR data.[26-27] This was done by using suppression of the solvent signal from which the aldehyde signals are significantly remote to allow accurate peak integration. The aldehyde signal of 2,4-disulphonated benzaldehyde was used as internal standard.

4.3.4.2 Hydroformylation of 4-styrenesulfonate

The hydroformylation of 4-styrenesulfonate using Rh/TPPTS 1:5 in H_2O gave full conversion in 24 hours (TON <1500 [mol product / mol catalyst]) and the **B2**/**L2** ratio was 13 (Scheme 5). The same results were found using 20 mM MES 50 mM NaCl pH 6, which is promising for the application of phosphine modified SCP-2L rhodium catalysts. The selectivity for the aldehyde products was only 31% due the formation of large amounts of hydrated aldehyde products. Low aldehyde selectivity for the hydroformylation reaction with this substrate and the formation of hydrates has been reported before and transformation to the corresponding alcohols during work-up of the reaction using sodium borohydride has been described.[27-28] Interestingly, by using D_2O as solvent product **B2** was not observed (Scheme 6), but **B2** with a deuterium incorporated into the β-carbon was observed instead. No such product was observed for **L2**. These results indicate solvent exchange of the hydrogen at the β-position of **B2** through an enol intermediate. This process is probably catalysed by rhodium and eliminates the prospect of obtaining enantioselectivity as this likely causes racemization of the chiral centre formed.



Scheme 5. Hydroformylation of 4-styrenesulfonate



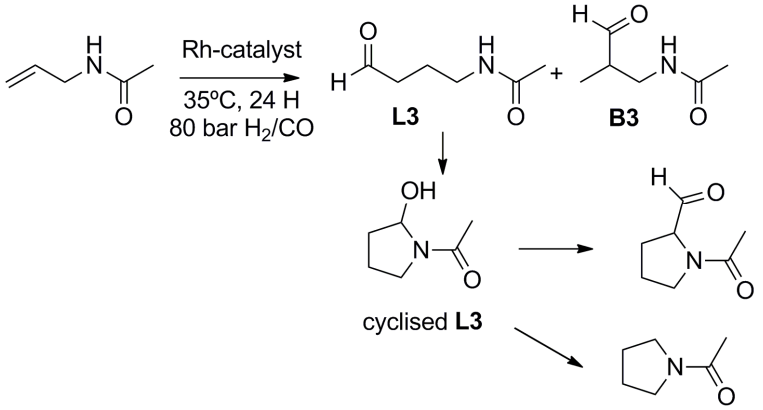
Scheme 6. Hydroformylation of 4-styrenesulfonate in D_2O

The hydroformylation of 4-styrenesulfonate catalysed by SCP-2L A100C-**M5**-**P3**-[Rh] resulted in a low yield of aldehyde (~6%). A **B3**/**L3** ratio of 6 was found for the small amount of aldehyde product observed, indicating a slightly increased linear selectivity compared to the reaction catalysed by Rh/TPPTS. However, the formation of large amounts of polymer during the reaction hampered a more detailed product analysis. This polymer formation led to formation of a viscous water layer and formation of a thick gel. Polymer formation of 4-styrenesulfonate was also observed by Anderson *et al* under hydroformylation conditions using an immobilized rhodium phosphine catalyst.[27, 29] They attributed this phenomenon to chain growth polymerisation on the alkyl-rhodium intermediate **H5** (Scheme 2) of the hydroformylation by continuous alkene insertion competing with CO insertion which would give hydroformylation products. Control experiments using radical scavengers excluded radical polymerization in their system. In our hands the polymerisation was also observed to take place under aerobic conditions after the hydroformylation was terminated by release of the synthesis gas pressure. In fact, a significant amount of

addition polymer was observed after the samples were left overnight. Also, the polymerisation reaction was only initially suppressed upon introduction of a radical scavenger (hydroquinone). No polymer formation was physically observed (the aqueous layer was not viscous) after the reaction was terminated, however the reaction mixture did become a viscous solution overnight. These result showed that the process more likely does involve radical polymerisation of 4-styrenesulfonate instead of the chain growth polymerisation proposed by Anderson *et al.*[29]

4.3.4.3 Hydroformylation of N-allylacetamide

The hydroformylation of *N*-allylacetamide using Rh/TPPTS 1:5 gave 80% conversion after 24 hours (Scheme 7). Similar results were again obtained for H_2O and 20 mM MES 50 mM NaCl pH 6. The selectivity was high (**B3**/**L3** ratio of 7.9), which is significantly more selective than the **B3**/**L3** = 1.3 reported for the same catalyst at higher temperature (70ºC), lower pressure (10 bar) and a Rh:TPPTS ratio of 1:25.[26, 30] The aldehyde selectivity was low (25%) which was due to the formation of cyclization products, which have previously been reported to be formed under hydroformylation conditions.[31]



Scheme 7. Hydroformylation of *N*-allylacetamide

The initial cyclization product is formed when linear aldehyde reacts intramolecularly with the amide to form an alcohol product containing either a four or a five membered ring structure. This cyclized structure has been reported to react in the presence of rhodium to give several cyclic side products. The formation of these cyclization products explains the high **B3**/**L3** ratio found as only the linear product cyclizes. The cyclization process was probably also the cause of the low TON of 358 [mol

aldehyde/mol catalyst] as this number was only based on the aldehyde products found. Correction of this number by including the cyclization products was hampered due to the difficult ^{1}H-NMR analysis of the products because of signal overlap. The hydroformylation of *N*-allylacetamide catalysed by SCP-2L A100C-**M5**-**P3**-[Rh] afforded only trace amounts of aldehyde products. A significant amount of cyclization products were observed by ^{1}H-NMR analysis but again peak overlap did not allow for quantification. A more detailed study of these reaction products using other analytical techniques might provide more insight in the actual activity and selectivity of the rhodium based catalysts in this reaction.

4.4 Conclusion and future prospects

Successful rhodium complex formation using several phosphine modified proteins has been described in this chapter. Mass spectrum and ^{31}P-NMR analysis indicated the formation of a phosphine-rhodium complex with one coordinating molecule of carbon monooxide; however the exact nature of this species and to what extent a mixture of species is obtained is still uncertain.

A 50 to over a 100 fold increase in the TON was found for the hydroformylation of 1-hexene and 1-octene using the created rhodium containing phosphine modified proteins compared to the Rh/TPPTS system. These protein based systems catalysed also the hydroformylation of 1-decene and 1-octadecene for which the Rh/TPPTS system shows very little to no activity. The obtained linear selectivity was remarkable for all systems, with SCP-2L A100C-**M5**-**P3** showing an **L**/**B** >5 in the hydroformylation of 1-hexene and 1-octene as the most promising results. The selectivity is most extraordinary when the results are compared to earlier reported dative rhodium-protein complexes in which also increased activity was found but no significant selectivity. These initial results did indicate that the catalyst systems truly operate trough a mechanism in which the tunnel of SCP-2L facilitates solubility of the substrates and enhances selectivity in a shape selective manner.

Further research will be required to confirm the mechanism of operation of these catalysts. The use of inhibitors that strongly bind to the SCP-2L tunnel is one of the experiments which can be performed to confirm the role of this tunnel. Other interesting experiments will be the use of additional structurally different proteins and detailed kinetic studies. Several parameters that could have a significant effect on the

catalyst performance like the temperature, pressure, and H_2/CO ratio may require optimisation and can provide more mechanistic insight.

The fact that differences in selectivity are observed between the SCP-2L V83C and A100C mutants are an indication of how changes to the protein environment can influence the selectivity of these reactions. This gives these systems plenty of opportunity for optimisation using designed and directed evolution. Opportunities lie in the use of different phosphines and maleimide-hydrazide cross-linkers as synthetic modification. Also biological tools like site directed mutagenesis can be applied to optimise the position of the cysteine inside the protein or the replacement of key amino acids. In addition, other protein environments can be selected since the developed modification strategy is generally applicable. In conclusion, these phosphine based catalytic systems hold great potential for selective hydroformylation of higher linear alkenes under biphasic conditions. In addition, this approach has the opportunity to be extended to other transition metal-phosphine catalysed reactions.

4.5 Experimental

4.5.1 General

4.5.1.1 Methods

The MaxEnt algorithm of Masslynx V4.0 (Micromass Ltd) was used for processing of the mass-spectra. Protein concentrations were determined using Bradfordôs reagent.[32] Protein solutions were handled under argon atmosphere using the experimental setup outlined in section 3.5.1.1.

4.5.1.2 Equipment

^{1}H and ^{31}P NMR spectra were taken at room temperature with Bruker Avance 300, 400 or 500 NMR spectrometers. Chemical shifts (δ) are given (in ppm) for high-frequency shifts relative to a TMS reference (^{1}H and ^{13}C) or 85% H_3PO_4 (^{31}P). Watergate ^{1}H-NMR spectra were recorded on a Bruker Avance 500 spectrometer. High pressure NMR experiments were performed on a Bruker 300 MHz machine using a 10 mm probe and a quartz-NMR tube suitable for high pressure experiments. LC-MS(ES^+) used for analysis of protein and protein 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 2.1x 10 mm cartridge using a gradient of 1% formic acid in H_2O to 1% formic acid in acetonitrile. Spin-filter concentrators with a molecular weight cut-off of 10kDa (Amicon® Ultra-15 Centrifugal Filter Units, Millipore®, Carrigtwohill, Co. Cork, Ireland) were used to concentrate and/or wash protein solutions. ICP-MS was performed using an Agilent 7500 series equipped with a Pt-tip by scanning for Rh(103) using free aspiration. Standards were measured using [Rh(acac)$(CO)_2$] solution in toluene providing a accurate detection range between 50 and 1000 ppb. Each sample was measured 10 times of which an average was taken to give the final value. CD spectra were recorded using a Jasco J-810 spectropolarimeter using a 1 mg/ml solution of the protein. Hydroformylation reactions were carried out in a stainless steel autoclave containing up to 8 glass reaction vials (volume of approximately 5 ml). The vials were equipped with mini-stirring bars and a septum cap, which was pierced with a needle to allow contact with the reaction gases. A CE instruments GC8000 Top equipped with a AS800 autosampler was used for GC analysis.

4.5.1.3 Materials

Triphenylphosphine trisulfonic acid trisodium salt (TPPTS) was purchased from Fluka containing about 5% of the TPPTS-oxide, Dicarbonyl rhodium acetylacetonate [Rh(acac)$(CO)_2$] was purchased from Acros and both were used as received. Degassed acetonitrile (MeCN) was obtained by distillation from CaH_2.

4.5.2 Procedure for the formation of rhodium complexes of phosphine modified proteins

The appropriate amount of [Rh(acac)$(CO)_2$] in a 12.5 mM stock of degassed MeCN was added to a 0.2 mM solution of the phosphine modified protein in 20 mM MES 50 mM NaCl pH 6 under argon. The pink suspension was stirred for at least three hours at room temperature during which it turned into a clear yellow solution. The following procedure was used before application in catalysis. The reaction mixture was centrifuged and washed under argon with 20 mM MES 50 mM NaCl pH 6 using a centrifugal concentrator with the procedure described in chapter 3. Mass spectra were recorded of a sample containing 1 εl of the reaction mixture in 50 εl of degassed

water before and/or after the washings. The procedure described in chapter 3 was used for the preparation of samples for ^{31}P-NMR.

4.5.3 Hydroformylation experiments

4.5.3.1 Preparation of Rh/TPPTS for hydroformylation reactions and high pressure NMR

Dicarbonyl rhodium acetylacetonate [Rh(acac)$(CO)_2$] and triphenylphosphine-3,3',3''-trisulfonic acid trisodium salt, (TPPTS) were dissolved in appropriate molar ratio in degassed methanol (150 εM based on rhodium) under argon atmosphere and stirred for two hours. The methanol was removed in vacuo and the remaining solid redissolved in degassed 20 mM MES 50 mM NaCl pH 6 buffer (0.3 mM based on rhodium). This solution is directly used in hydroformylation. The same procedure was used for the preformation of the complex used for high-pressure NMR using a 2:1 ratio TPPTS:Rh. The quartz NMR tube was loaded with a solution of this complex in degassed 20 mM MES 50 mM NaCl pH 6 buffer (20 mM based on Rh, containing 10% degassed D_2O) and after recording the first spectrum charged with 40 bar CO/H_2. The ^{31}P-NMR spectra were all recorded at a constant temperature of 35ºC.

4.5.3.2 Biphasic hydroformylation of linear α-olefins

Experiments were performed in a stainless steel autoclave containing a stainless steel insert in which glass vials with stirring bars can be placed. The autoclave was loaded with the reactants, pressurized and placed on a stirring plate containing an oil bath. This setup allowed for performing up to 10 small volume (Ò2 ml) batch reactions simultaneously under the same conditions. A pressure of 80 bar of synthesis gas was used similar to the work of Marchetti *et al*.[18b] Lower temperatures are preferred in order to minimise denaturation of the protein. A temperature of 35ºC was selected which could be maintained using an oil bath. Continuous substrate addition or formation of the hydride complex prior to addition of the substrate can not be performed due to the small scale and batch-wise manner at which the reactions are run. To facilitate analysis a biphasic system was used in which the organic layer consisted of substrate (containing GC internal standards) to allow direct analysis of the organic phase for formed aldehyde products by gas chromatography.

Under argon atmosphere the vials were loaded with the preformed Rh/protein or Rh/TPPTS catalyst solutions to a total volume of 500 εL in degassed 20 mM MES 50 mM NaCl pH 6 buffer (0.3 εmol for Rh/TPPTS and 0.15 εmol for the artificial metalloenzymes based on Rh, the artificial metalloenzyme precipitated when more concentrated solutions were used). Triton-X-100 was added from a 23 mM stock in the same buffer. 500 εL of the corresponding alkene substrate solution containing 9% (v/v) heptane and 1% (v/v) diphenyl ether as internal standards was added. This substrate solution was prepared by mixing a solution of 10% (v/v) diphenyl ether in heptane with an appropriate amount of alkene freshly filtered over a plug of neutral alumina. The autoclave was carefully purged three times with 5-10 bar CO/H_2 before being charged to 80 bars and placed on a stirring plate in an oil bath at 35 ºC. After the appropriate time the autoclave was cooled and pressure was very carefully released. The organic layer was separated from the water layer and used directly for GC and ICP-MS analysis. Soluble protein concentration in the water layer was determined by performing a Bradford assay on soluble fraction of the water layer after centrifugation (30 min, 16100 *g*).[32] For LCMS (ES^+) analysis, 1 εL of this soluble fraction was dissolved in 49 H_2O, although this was only possible for reaction not containing triton-X-100 as this compound severely interferes with the protein mass-spectrum. GC analysis conditions: Restek RTX-1 (general purpose) column 30m x 0.25 mm x 0.1 εm; injector temperature 250 ºC; 100 kPa constant pressure; 100:1 splitratio; 1 or 2 εL injection depending on the signal strength; oven method, 45 ºC isotherm for 5 min, 20 ºC/min to 130 ºC, 130 isotherm for 2 min, 20 ºC/min to 250 ºC, 250 ºC isotherm for 10 min and 20 ºC/min to 300 ºC; retention times, 1-hexene 2.1 min, heptane 2.6 min, 1-octene 3.7 min, 2-methylhexanal 4.8 min, *n*-heptanal 5.7 min, 1-decene 7.6 min, 2-methyloctanal 8.3 min, *n*-nonanal 8.7 min, 1-dodecene 9.7 min, 2-methyldecanal 10.3 min, *n*-undecanal 10.7 min, phenylether 11.8 min, 2-methyldodecanal 12.6 min, *n*-tridecanal 13.0 min, 1-octadecene 15.2 min, 2-methyloctadecanal 16.6 min, *n*-nonadecanal 16.6 min.

4.5.3.3 Hydroformylation of water soluble substrates

The reaction vessels were loaded with the appropriate substrate (1 mmol) and 2,4-disulphonated benzaldehyde (0.1 mmol) as internal standard. The autoclave was degassed and under argon atmosphere the vials were loaded with preformed

Rh/protein or Rh/TPPTS (0.15 εmol for the artificial metalloenzymes or 0.3 εmol for Rh/TPPTS based on Rh) catalyst solutions to a total volume of 500 εL in degassed 20mM MES 50 mM NaCl pH 6 buffer. The autoclave was carefully purged three times with 5-10 bar CO/H_2 before being charged to 80 bars and placed on a stirring plate in an oil bath at 35 ºC. After the appropriate time the autoclave was cooled and pressure was very carefully released. The samples were loaded in NMR tubes and ^{1}H-NMR was measures using a 30 second delay time to allow proper quantification. Products were identified using reported ^{1}H-NMR data.[26-27]

4.6 References

[1] For an interesting paper on the history on the discovery of this reaction see: a) O. Roelen, *Chem. Exp. Didakt.* **1977**, *3*, 119; b) B. Cornils, W. A. Herrmann, M. Rasch, *Angew. Chem., Int. Ed.* **1994**, *33*, 2144.

[2] P. W. N. M. van Leeuwen, *Homogeneous Catalysis: Understanding the Art*, Kluwer Academic Publishers, Dordrecht, **2004**.

[3] a) B. Breit, *Top. Curr. Chem.* **2007**, *279*, 139; b) R. V. Chaudhari, *Curr. Opin. Drug Discovery Dev.* **2008**, *11*, 820; c) J. A. Gillespie, D. L. Dodds, P. C. J. Kamer, *Dalton Trans.* **2010**, *39*, 2751; d) P. W. N. M. Van Leeuwen, C. Claver, Editors, *Rhodium Catalyzed Hydroformylation. [In: Catal. Met. Complexes, 2000; 22]*, **2000**.

[4] a) R. F. Heck, *Accounts Chem. Res.* **1969**, *2*, 10; b) J. F. Young, J. A. Osborn, F. H. Jardine, G. Wilkinson, *Chem. Commun.* **1965**, 131; c) D. Evans, J. A. Osborn, G. Wilkinson, *J. Chem. Soc. A* **1968**, 3133; d) D. Evans, G. Yagupsky, G. Wilkinson, *J. Chem. Soc. A* **1968**, 2660.

[5] J. M. Brown, A. G. Kent, *J. Chem. Soc., Perkin Trans. 2* **1987**, 1597.

[6] a) E. Monflier, H. Bricout, F. Hapiot, S. Tilloy, A. Aghmiz, A. M. Masdeu-Bulto, *Adv. Synth. Catal.* **2004**, *346*, 425; b) C. Bianchini, H. M. Lee, A. Meli, F. Vizza, *Organometallics* **2000**, *19*, 849; c) J. M. Brown, L. R. Canning, A. G. Kent, P. J. Sidebottom, *J. Chem. Soc., Chem. Commun.* **1982**, 721; d) I. T. Horvath, R. V. Kastrup, A. A. Oswald, E. J. Mozeleski, *Catal. Lett.* **1989**, *2*, 85.

[7] a) A. S. C. Chan, H. S. Shieh, J. R. Hill, *J. Chem. Soc., Chem. Commun.* **1983**, 688; b) A. Aghmiz, C. Claver, A. M. Masdeu-Bulto, D. Maillard, D. Sinou, *J. Mol. Catal. A: Chem.* **2004**, *208*, 97; c) Z. Freixa, M. M. Pereira, A. A. C. C. Pais, J. C. Bayon, *J. Chem. Soc., Dalton Trans.* **1999**, 3245.

[8] B. Cornils, E. G. Kuntz, *J. Organomet. Chem.* **1995**, *502*, 177.

[9] H. Bahrmann, S. Bogdanovic, P. W. N. M. van Leeuwen, *Aqueous-Phase Organometallic Catalysis (2nd Edition)* **2004**, 391.

[10] P. Kalck, F. Monteil, *Adv. Organomet. Chem.* **1992**, *34*, 219.

[11] F. Joo, A. Katho, *J. Mol. Catal. A: Chem.* **1997**, *116*, 3.

[12] R. V. Chaudhari, B. M. Bhanage, R. M. Deshpande, H. Delmas, *Nature (London, U. K.)* **1995**, *373*, 501.

[13] M. Angeles Garcia del Vado, A. F. Rodriguez Cardona, G. R. Echevarria Gorostidi, M. Carolina Gonzalez Martinez, J. G. Santos Blanco, F. G. Blanco, *J. Mol. Catal. A: Chem.* **1995**, *101*, 137.

[14] a) C. Yang, X. Bi, Z.-S. Mao, *J. Mol. Catal. A: Chem.* **2002**, *187*, 35; b) M. Li, Y. Li, H. Chen, Y.-E. He, X. Li, *J. Mol. Catal. A: Chem.* **2003**, *194*, 13.
[15] E. A. Karakhanov, Y. S. Kardasheva, E. A. Runova, V. A. Semernina, *J. Mol. Catal. A: Chem.* **1999**, *142*, 339.
[16] a) T. Mathivet, C. Meliet, Y. Castanet, A. Mortreux, L. Caron, S. Tilloy, E. Monflier, *J. Mol. Catal. A: Chem.* **2001**, *176*, 105; b) E. Monflier, G. Fremy, Y. Castanet, A. Mortreux, *Angewandte Chemie, International Edition in English* **1995**, *34*, 2269; c) T. Mathivet, C. Meliet, Y. Castanet, A. Mortreux, L. Caron, S. Tilloy, E. Monflier, *J. Mol. Catal. A Chem.* **2001**, *176*, 105; d) M. Dessoudeix, M. Urrutigoity, P. Kalck, *Eur. J. Inorg. Chem.* **2001**, 1797.
[17] S. Shimizu, S. Shirakawa, Y. Sasaki, C. Hirai, *Angew. Chem., Int. Ed.* **2000**, *39*, 1256.
[18] a) C. Bertucci, C. Botteghi, D. Giunta, M. Marchetti, S. Paganelli, *Adv. Synth. Catal.* **2002**, *344*, 556; b) M. Marchetti, G. Mangano, S. Paganelli, C. Botteghi, *Tetrahedron Lett.* **2000**, *41*, 3717.
[19] a) S. Crobu, M. Marchetti, G. Sanna, *J. Inorg. Biochem.* **2006**, *100*, 1514; b) S. Paganelli, M. Marchetti, M. Bianchin, C. Bertucci, *J. Mol. Catal. A: Chem.* **2007**, *269*, 234.
[20] S. Paganelli, A. Ciappa, M. Marchetti, A. Scrivanti, U. Matteoli, *J. Mol. Catal. A: Chem.* **2006**, *247*, 138.
[21] Q. Jing, R. J. Kazlauskas, *ChemCatChem* **2010**, *2*, 953.
[22] Q. Jing, K. Okrasa, R. J. Kazlauskas, *Chem.--Eur. J.* **2009**, *15*, 1370.
[23] P. J. Deuss, G. Popa, C. H. Botting, W. Laan, P. C. J. Kamer, *Angew. Chem., Int. Ed.* **2010**, *49*, 5315.
[24] H. E. Gottlieb, V. Kotlyar, A. Nudelman, *J. Org. Chem.* **1997**, *62*, 7512.
[25] M. M. T. Khan, V. V. S. Reddy, H. C. Bajaj, *Polyhedron* **1987**, *6*, 921.
[26] G. Verspui, G. Elbertse, F. A. Sheldon, M. A. P. J. Hacking, R. A. Sheldon, *Chemical Communications (Cambridge)* **2000**, 1363.
[27] J. R. Anderson, E. M. Campi, W. R. Jackson, Z. P. Yang, *J. Mol. Catal. A Chem.* **1997**, *116*, 109.
[28] M. Schreuder Goedheijt, P. C. J. Kamer, P. W. N. M. van Leeuwen, *J. Mol. Catal. A Chem.* **1998**, *134*, 243.
[29] J. R. Anderson, W. R. Jackson, Z. Yang, E. M. Campi, *Catal. Lett.* **1997**, *45*, 197.
[30] G. Verspui, G. Elbertse, G. Papadogianakis, R. A. Sheldon, *J. Organomet. Chem.* **2001**, *621*, 337.
[31] I. Ojima, Z. Zhang, *J. Org. Chem.* **1988**, *53*, 4422.
[32] M. M. Bradford, *Anal. Biochem.* **1976**, *72*, 248.

Chapter 5. Artificial copper-enzymes for asymmetric reactions

Modification of SCP-2L templates with nitrogen donor ligands and their application in asymmetric copper catalysis

5.1 Abstract

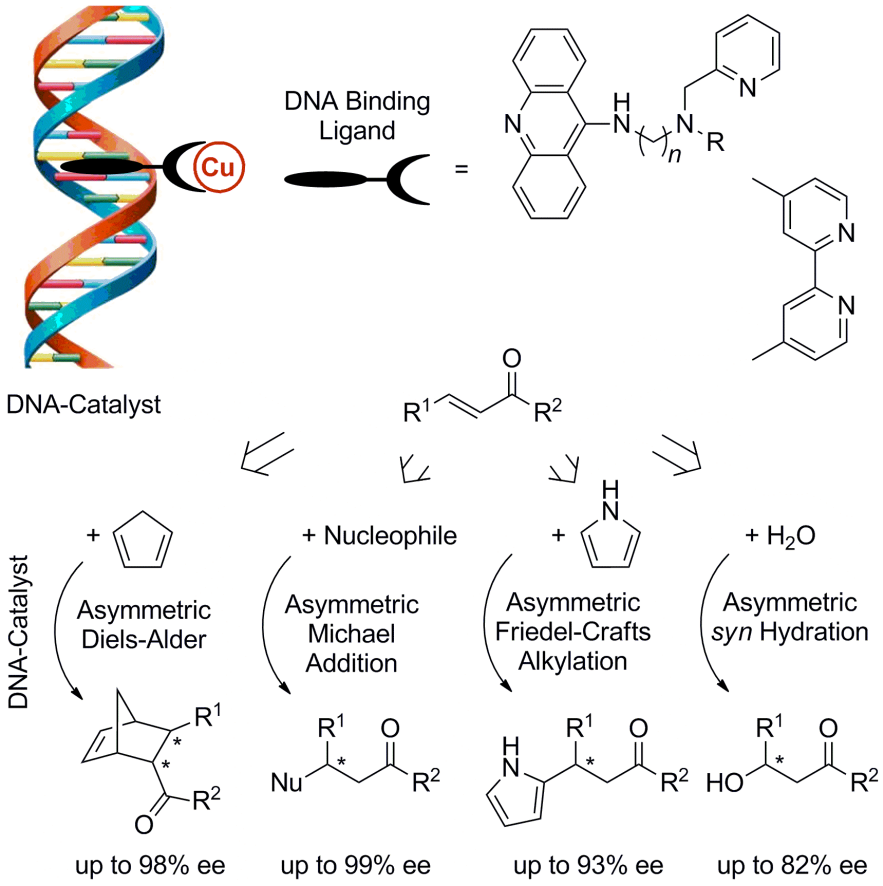
This chapter describes the development of artificial copper enzymes from SCP-2L for the use in asymmetric catalysis. For this purpose, the unique cysteine SCP-2L templates described in chapter two were modified with various nitrogen donor ligands. Maleimide containing ligands were found most suitable for selective cysteine bioconjugation. Fluorescence spectroscopy was used to confirm copper binding to the phenanthroline ligand introduced in SCP-2L. Several copper adducts were applied in asymmetric Diels-Alder and Michael addition reactions. A clear influence of both the protein environment and the introduced ligand was found in the asymmetric Diels-Alder reaction between azachalcone and cyclopentadiene. A promising enantioselectivity of 25% ee was obtained using SCP-2L V83C modified with phenanthroline-maleimide **N3**. This provides a suitable starting point for the implementation of various available techniques to optimize the performance of this system. Application of these catalyst systems in asymmetric Michael addition reactions proved less fruitful.

5.2 Introduction

5.2.1 Bioinspired copper-nitrogen catalyst systems

Many groups have reported the use of proteins as chiral scaffold for the development of artificial metalloenzymes for asymmetric catalysis. Various approaches have led to catalyst systems that induce high enantioselectivity in reactions like, for example, asymmetric hydrogenation,[1] asymmetric sulfoxidation.[2] Particularly copper-nitrogen ligand systems have been widely applied in various bioinspired systems. There are various significant advantages of using nitrogen-ligand based catalytic systems over the phosphine based systems described in the previous two chapters. The main advantage is that nitrogen ligands are not oxygen sensitive, which makes the synthesis of the ligands and the synthesis and manipulation of ligand-modified proteins much more straightforward. This advantage becomes even more pronounced if catalytic systems are going to be used for combinatorial optimisation techniques. Also, unlike phosphines, nitrogen donor ligands are in general compatible with cysteine selective modification procedures,[3] and the introduction of nitrogen donor ligands via covalent modification of biomolecules is well established, particularly because of the widespread application of such systems in DNA-cleavage.[4] In addition, nitrogen donor ligands, like bipyridine can be introduced as unnatural amino acids.[4a]

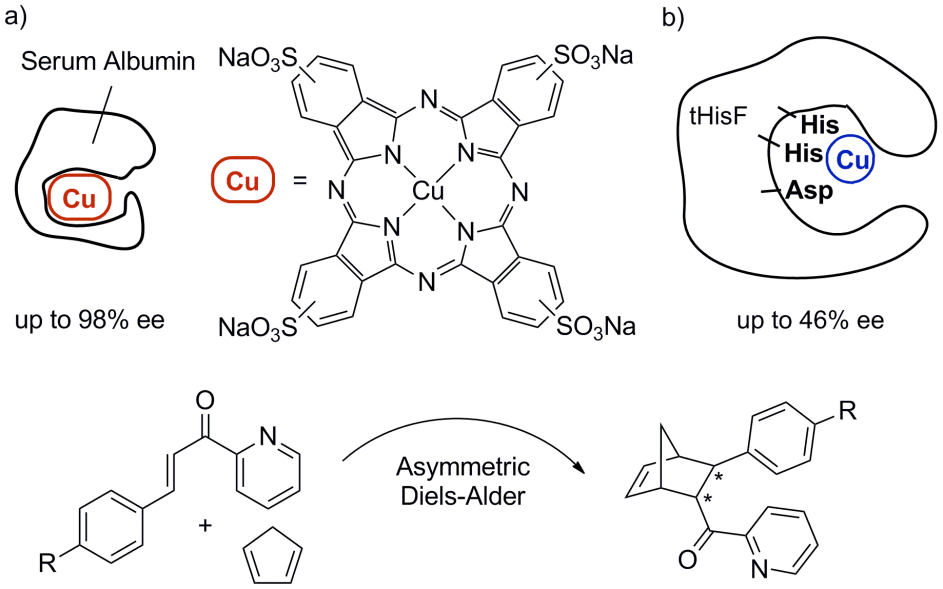
Roelfes *et al* demonstrated how nitrogen copper systems in combination with DNA can be used for a wide range of copper catalysed asymmetric reactions (Scheme 1).[5] The catalyst was assembled by application of nitrogen donor ligands that can intercalate or bind the major groove of DNA. Excellent enantioselectivities were obtained for copper catalysed asymmetric Diels-Alder[6], Michael addition[7] and Friedel-Crafts alkylation[8] reactions using these nitrogen-copper complexes in combination with commercially available salmon testes or calf thymus DNA. Recently, this system was also applied in the enantioselective and diastereospecific syn hydration of α,β-unsaturated ketones.[9] This reaction is not performed by either conventional asymmetric catalysts using transition metals or by DNA/RNAzymes, demonstrating how a biomolecule and a transition metal can be combined to form a catalyst which opens up new reaction pathways.



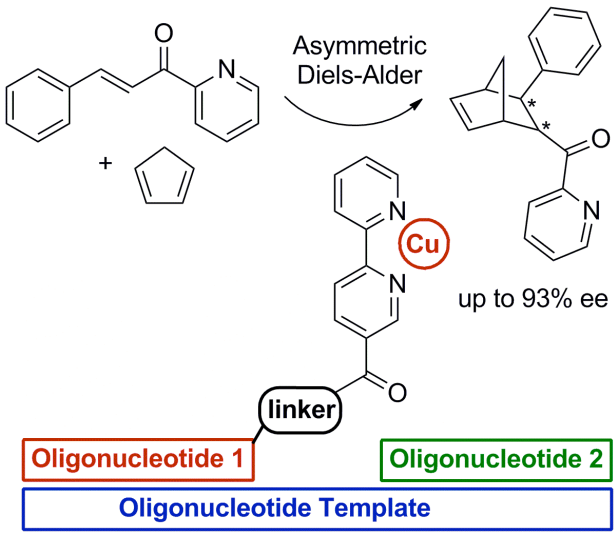
Scheme 1. Nitrogen donor ligands applied in combination with DNA in various copper catalysed asymmetric reactions by Roelfes *et al*.[6-10]

The group of Reetz has developed various protein based copper nitrogen ligand catalysts to obtain enantioselective artificial metalloenzymes.[11] One such system consisted of commercially available copper phthalocyanine in combination with various serum albumins to create artificial metalloenzymes for copper catalysed Diels-Alder reactions (Scheme 2a).[12] These systems proved surprisingly effective showing good conversions, high endo selectivities and high enantioselectivity. Nevertheless, the lack of an efficient *en masse* purification system for serum albumin hampered application of directed evolution for performance optimization. Very recently, Reetz reported the introduction of a copper binding-site in the synthase subunit of imidazole glycerol phosphate synthase from *Thermotoga maritime* (tHisF, Scheme 2b). This protein was specifically selected for its thermo-stability which allows efficient purification via a simple heat treatment to separate this protein directly from the cell extract. The binding site consists of a geometrically appropriately introduced histidine-histidine-asparagine motif, which was inspired by

natural copper binding sites in proteins. These mutations were introduced using step-wise site directed mutagenesis. Copper coordination to this motif was confirmed using EPR spectroscopy. The obtained artificial metalloenzyme showed good conversion and endo-selectivity and more importantly 46% enantioselectivity in the Diels Alder reaction between azachalcone and cyclopentadiene. Reetz *et al.* anticipate that this new system will be suitable for application in directed evolution.

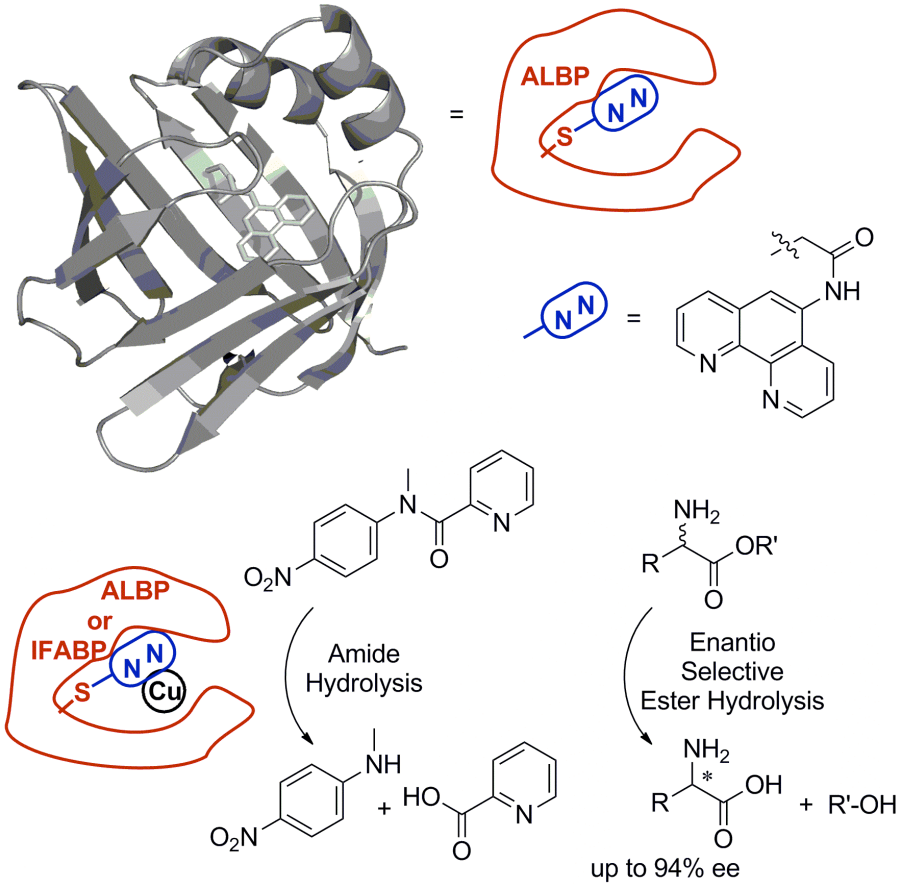


Scheme 2. Copper based artificial metalloenzymes applied by Reetz *et al.* in the asymmetric Diels-Alder reaction a) artificial metalloenzymes assembled by supramolecular interaction between copper phthalocyanine and serum albumins[12] and b) artificial metalloenzymes assembled using the dative interaction between copper and an engineered binding site in tHisF.[13]



Scheme 3. Terminal bipyridine modified oligonucleotide used in copper catalysed Diels-Alder reactions.

Using a different approach Roelfes introduced a bipyridine complex into DNA by covalent modification (Scheme 3).[14] The obtained system was applied in the copper catalysed Diels-Alder reaction in which the enantioselectivity could be improved from 22 to 93% by changing the complementary strand.



Scheme 4. Artificial metalloenzyme developed by Distefano *et al.*[15-16]

Proteins have also been covalently modified with nitrogen donor ligands. The group of Distefano developed the first covalently assembled enantioselective artificial metalloenzyme, based on copper and a cysteine selectively introduced phenanthroline (Scheme 4). The cysteine at position 117 in adipocyte lipid binding protein (ALBP) was site selectively modified using 5-iodoacetamide-1,10-phenanthroline. ALBP was selected for its large hydrophobic cavity (600 $Å^3$).[15] This cavity is large enough to accommodate the covalently introduced phenanthroline and a substrate. Fluorescence spectroscopy was used to characterize the copper coordination to the introduced nitrogen donor ligand. This was possible owing to quenching of the native fluorescence of phenanthroline upon copper coordination. The resulting artificial metalloenzyme was applied in amide and enantioselective ester hydrolysis. High

conversions were found in both amide and ester hydrolysis and up to 86% ee was achieved in the L-specific hydrolysis of the isopropyl ester of alanine. The catalyst system could be optimised by variation of the position of the unique cysteine and thus the location of the phenanthroline-copper inside the cavity of intestinal fatty acid binding protein (IFAP), a structurally related protein.[16] This resulted in an L-specific catalytic system that gave 94% ee in the hydrolysis of the isopropyl ester of alanine.

5.2.2 The potential of SCP-2L based copper-nitrogen ligand systems

SCP-2L and analogous proteins are known to bind a variety of hydrophobic molecules.[17] This makes this class of proteins very suitable for the development of artificial metalloenzymes for asymmetric reactions involving hydrophobic molecules like the typical reactants used in Diels-Alder reactions. The native cavity SCP-2L might not have the optimal structure to induce the desired selectivity in a selected reaction, but the structure can be optimised for this purpose using directed evolution or designed evolution.

In our approach we wish to introduce our transition metal binding moiety covalently. Therefore, an efficient bioconjugation technique is required to make this a viable approach for future applications. The conjugation products in the previous chapters were analysed by mass-spectroscopy and ^{31}P-NMR. No direct substitute is available for the latter to study conjugates containing nitrogen-ligands, however alternative analysis techniques like UV-Vis-, fluorescence- and EPR spectroscopy (on paramagnetic metal adducts) can be applied instead. Especially EPR spectroscopy offers great prospects as this allows for direct analysis of the copper environment.

Several options for bioconjugation of nitrogen donor ligands to SCP-2L will be explored in this chapter along with some initial applications of copper adducts in asymmetric catalysis.

5.3 Results and discussion

5.3.1. Synthesis and bioconjugation of nitrogen donor ligands

5.3.1.1 Synthesis of nitrogen donor ligands

Nitrogen ligands do not suffer from the compatibility problems with standard cysteine selective bioconjugation methods encountered with phosphorous ligands. Therefore single step modification of protein-scaffolds should be straightforward. Distefano *et*

al. already demonstrated the introduction of a phenanthroline moiety site selectively to a unique cysteine containing protein using 5-iodoacetamide-1,10-phenanthroline as described above.[15] However, it has been reported that this reaction can potentially form iodine containing side-products which can influence catalytic results.[18] Nevertheless, many other nitrogen donor ligands have been introduced via site selective linkage to cysteine in a similar way.[19]

The synthesis of a dipicolylamine based nitrogen donor ligand **N4** for this purpose is given below. In addition, nitrogen donor ligands **N1**[20]**, N2**[14] and **N3**[21] were used in this study, (Figure 1). Several other synthetic routes to novel nitrogen donor ligands which were not completed can be found in appendix A9.

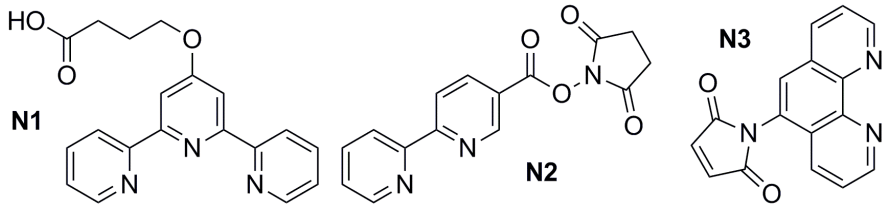
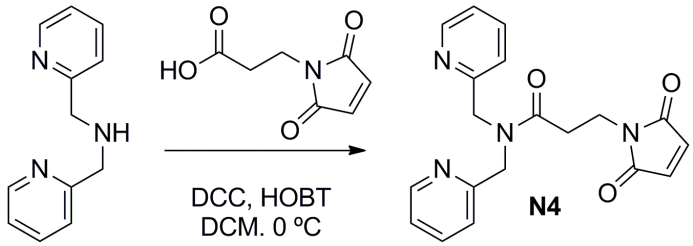


Figure 1. Nitrogen donor ligands used for bioconjugation to unique cysteine SCP-2L templates in addition to dipicolylamine based ligands.

5.3.1.2 Synthesis of dipicolylamine based ligands

Nitrogen donor ligands based on the dipicolylamine moiety have not yet been explored for copper catalysed reactions. They have however been used in radioactive labelling of various macromolecular scaffolds and complex studies involving copper and other transition metals.[22]



Scheme 5. Synthesis of dipicolylamine based ligands **N4**.

Dipicolylamine is commercially available and can be readily modified with functionalities for bioconjugation to cysteine. A single step synthetic procedure did provide functionalized dipicolylamine based ligand **N4** (Scheme 5). Maleimide containing **N4** was synthesised by an amide bond formation procedure between dipicolylamine and 3-maleimidopropionic acid using *N*,*N*'-dicyclohexylcarbodiimide

(DCC) and hydroxybenzotriazole (HOBT), affording **N4** in 95% yield. Other dipicolylamine derivatives were also synthesised and can be found in the appendix A9.

5.3.1.3 Bioconjugation

Ligand **N1** contains a terpyridine motif with a carboxylic acid linker which is suitable for bioconjugation via activation of the carboxylic acid.[23] However, activation using 1,1'-carbodiimidazole (CDI) and subsequent bioconjugation to SCP-2L V83C resulted in unselective modification (Table 1, entries 1-3). These results were similar to those obtained for carboxylic acid containing phosphine ligands described in section 3.3.1.

Table 1. Results of bioconjugation of SCP-2L V83C[a] with nitrogen donor ligands **N1-N4**.[b]

#	Ligand	Equiv.	Apo protein (%)	Mono mod. (%)	Multiple modification		
					2x (%)	3x (%)	4x (%)
1[c]	**N1**	2	63	37	-	-	-
2[c]	**N1**	11	47	50	3	-	-
3[c]	**N1**	20	22	75	3	-	-
4	**N2**	1	90	10	-	-	-
5	**N2**	3	74	24	2	-	-
6	**N2**	5	41	40	16	3	-
7	**N2**	10	38	41	18	3	-
8	**N2**	20	10	31	35	19	5
9	**N3**	10	-	100	-	-	-
10	**N4**	10	-	100	-	-	-

a) SCP-2L V83C obtained by the purification procedure without nickel affinity chromatography was used for entries 1-3 (see chapter 2 section 2.5.3.3 for more details) b) Conditions: 15 μM SCP-2L V83C, 5-20 equivalents of activated CDI activated carboxylic acid (from 25 μM stock in DMF) in 20 mM MES 50 mM NaCl pH 6 buffer. Product formation (%) determined by peak height ratios in mass spectroscopy (ES^+). c) 20 mM HEPES pH 7 buffer.

The *N*-hydroxysuccinimide (NHS) activated dipyridine ligand **N2** showed the same selectivity problems (Table 1, entries 4-6). Therefore, activated carboxylic acid bearing ligands proved unsuitable for selective modification of SCP-2L V83C as already discussed before in chapter 3.

In line with the results described before in chapters 2 and 3, maleimide conjugation of nitrogen-ligands to SCP-2L V83C proved successful. Maleimide-phenanthroline **N3** was readily introduced site selectively to the cysteine of SCP-2L V83C (Table 1, entry 9). Full conversion to the conjugation product was also observed for the modification of SCP-2L A100C with **N3**.

Maleimide containing ligand **N4** was conjugated to SCP-2L V83C giving full conversion to the modification product (Table 1, entry 10). The same results were obtained using SCP-2L A100C as protein template.

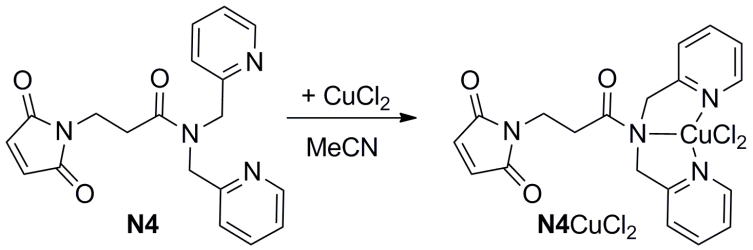
Thus the two maleimide containing nitrogen donor ligands **N3** and **N4** were found to be the most suitable ligands for cysteine-selective bioconjugation. A disadvantage of this modification technique is that a chiral carbon is formed during the reaction which gives the possibility of the formation of different stereoisomers. This is highly undesirable considering the envisaged application in asymmetric catalysis. However, there is a good possibilty that a preference for one of the stereoisomers is induced by the chiral protein environment. The selective formation of stereoisomers is difficult to determine, therefore a different modification procedure might be required for future application. Nevertheless, the modification of SCP-2L V83C and SCP-2L 100C with ligands **N3** and **N4** provided us with a small set of modified proteins, which could be used for initial catalytic studies.

5.3.2 Copper complex formation

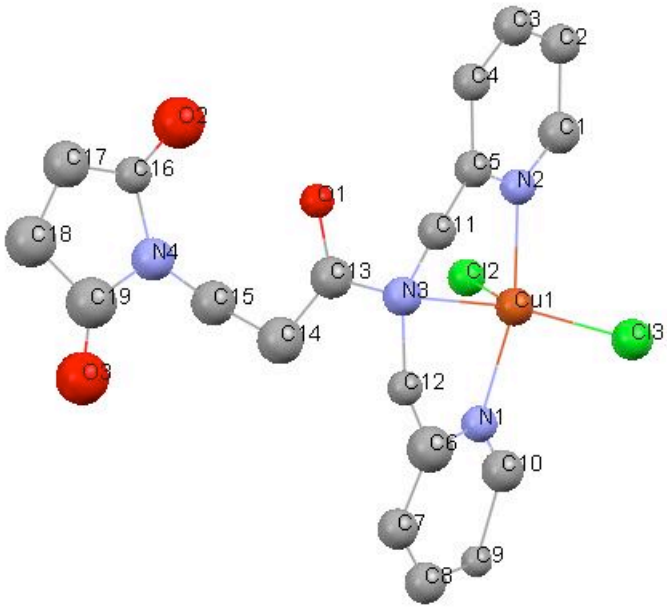
*5.3.2.1 Synthesis of copper complexes of maleimide-nitrogen ligand **N4***

Two approaches can be taken to obtain artificial metalloenzymes of proteins modified with the above nitrogen donor ligands. The first is to couple a pre-formed metal complex to the protein and the second is to add the transition metal after modification. A preformed copper complex was obtained by a reaction of ligand **N4** with copper(II)chloride, which was tested in the bioconjugation to the unique cysteine of the protein (Scheme 6).

Crystals suitable for X-ray crystallography were obtained and the resulting structure confirmed the formation of a **N4**-copperdichloride complex with **N4** acting as a tridentate ligand (Figure 2, see appendix A10 for more information). The structure is found as a trigonal bipyramid with the two chloride atoms and the amide nitrogen atom in the equatorial plane and the pyridine nitrogen atoms in the axial positions. The bond angles show that trigonal bipyramid is distorted with the N1-Cu1-N2 angle being lower than 160º, which is most likely caused by the ligand structure. From the bond distances it can be concluded that the amide nitrogen forms a weaker nitrogen copper bond compared to the two electron rich pyridine nitrogen atoms. The crystal structure also confirmed that the double bond of the maleimide, which is essential for cysteine conjugation, remained unaffected as a bond distance typical for a C=C bond was observed. An EPR spectrum was also recorded of **N4**$CuCl_2$ (see appendix A10).



Scheme 6. Copper complex formation of **N4**.



Selected bond distances:

Cu1-N1	1.989 Å
Cu1-N2	2.014 Å
Cu1-N3	2.381 Å
Cu1-Cl2	2.294 Å
Cu1-Cl3	2.266 Å
C17-C18	1.437 Å

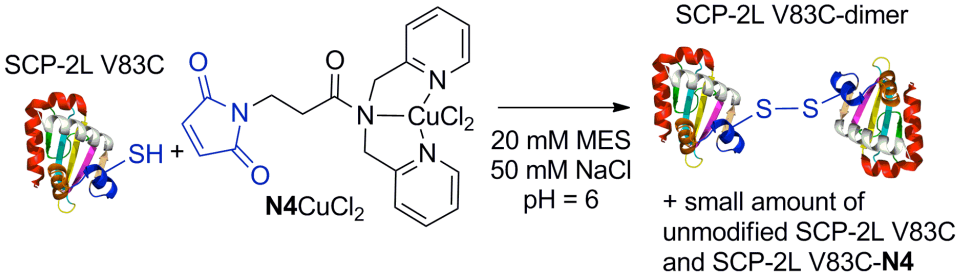
Selected bond angles:

N1-Cu1-N3	76.7º
N2-Cu1-N3	81.1º
N1-Cu1-N2	157.7º
N3-Cu1-Cl2	107.6º
N3-Cu1-Cl3	111.2º

Figure 2. Crystal structure of **N4**$CuCl_2$ and selected bond distances and angles.

5.3.2.2 Modification using preformed copper complexes

Modification of SCP-2L V83C with 3-10 equivalents of **N4**$CuCl_2$ was performed using standard condition for maleimide conjugation to cysteine described in earlier chapters (Scheme 7).



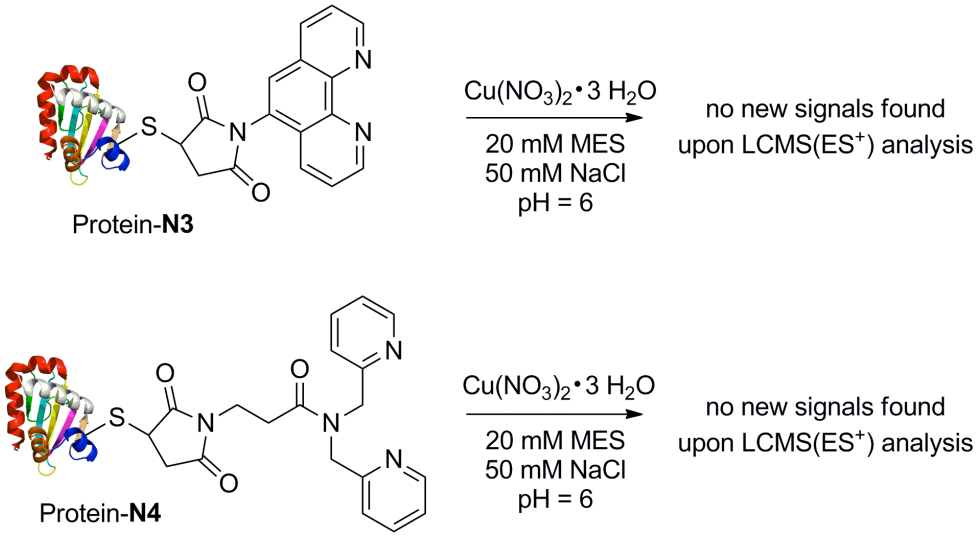
Scheme 7. Bioconjugation of SCP-2L V83C with **N4**$CuCl_2$ resulting in a SCP-2L V83C dimer.

Disappointingly, analysis of the reaction product by LCMS(ES^+) gave only a signal corresponding to the dimer of SCP-2L V83C (expected 26751 Da, found 26750 Da) which is likely the result of copper catalysed disulfide bridge formation. A small fraction of SCP-2L V83C modified with **N4** (<5%) was also found but no peaks corresponding to copper adducts were detected. Notably, the colour of the protein solution remained blue after washing and concentration, similar to the original copper complex. This indicated that copper is still present in the protein. Whether this is the result of copper coordinating to **N4** inside the protein, providing a species which is not observed in mass-spectroscopic analysis or of copper complexes formed with other amino acid side chains remains unclear from these experiments. The dimer formation could be suppressed by addition of the reducing agent tris(2-carboxyethyl)phosphine (TCEP) in an equimolar ratio to **N4**$CuCl_2$. Although, this experiment led to increased formation of SCP-2L V83C-**N4** (75% modified protein, 25% unmodified protein), still no peaks corresponding to any copper adduct were detected.

5.3.2.3 Complex formation by addition of metal salt to nitrogen donor ligand modified proteins

The second approach was to form copper complexes after modification of the unique cysteine with nitrogen donor ligands **N3** and **N4** (Scheme 8). Copper(II)nitrate trihydrate was used for this purpose because of its water solubility. Disappointingly,

no copper adducts were found upon LCMS(ES^+) analysis of the soluble fraction after the reaction. The observed signals in the mass-spectrum were identical to those before the addition of copper.



Scheme 8. Copper complex formation by addition of $Cu(NO_3)_2$ •3 H_2O **N4** and **N5** modified proteins.

The outcome of the above experiments can be explained in two ways. The first is that no copper adducts are formed, which is somewhat unexpected since the introduced nitrogen ligands have a high affinity for copper. The second is that the copper adducts are not suitable for detection by mass-spectroscopic analysis using the setup and conditions used in this research. The second hypothesis is supported by the fact that no mass-spectra of copper adducts have been reported in earlier studies using copper adducts of proteins.[4a, 12-13, 15-16] Therefore, the complexation was also studied using fluorescence spectroscopy. This method was applied by Distefano and co-workers to confirm copper complexation to the ALBP-phenanthroline conjugates developed in his group.[15-16] They showed that the fluorescence emission peak of the protein-bound phenanthroline at 406 nm that occurs upon excitation at 275 nm is quenched upon addition of copper(II)sulphate. It was established that this quenching is the result of copper coordination to phenanthroline by analysis of the fluorescence spectra of the analogous non-conjugated complexes. The fluorescent signal was restored by extraction of the copper from the protein by addition of an excess of a copper chelating agent like the disodium salt of ethylenediaminetetraacetic acid (EDTA).

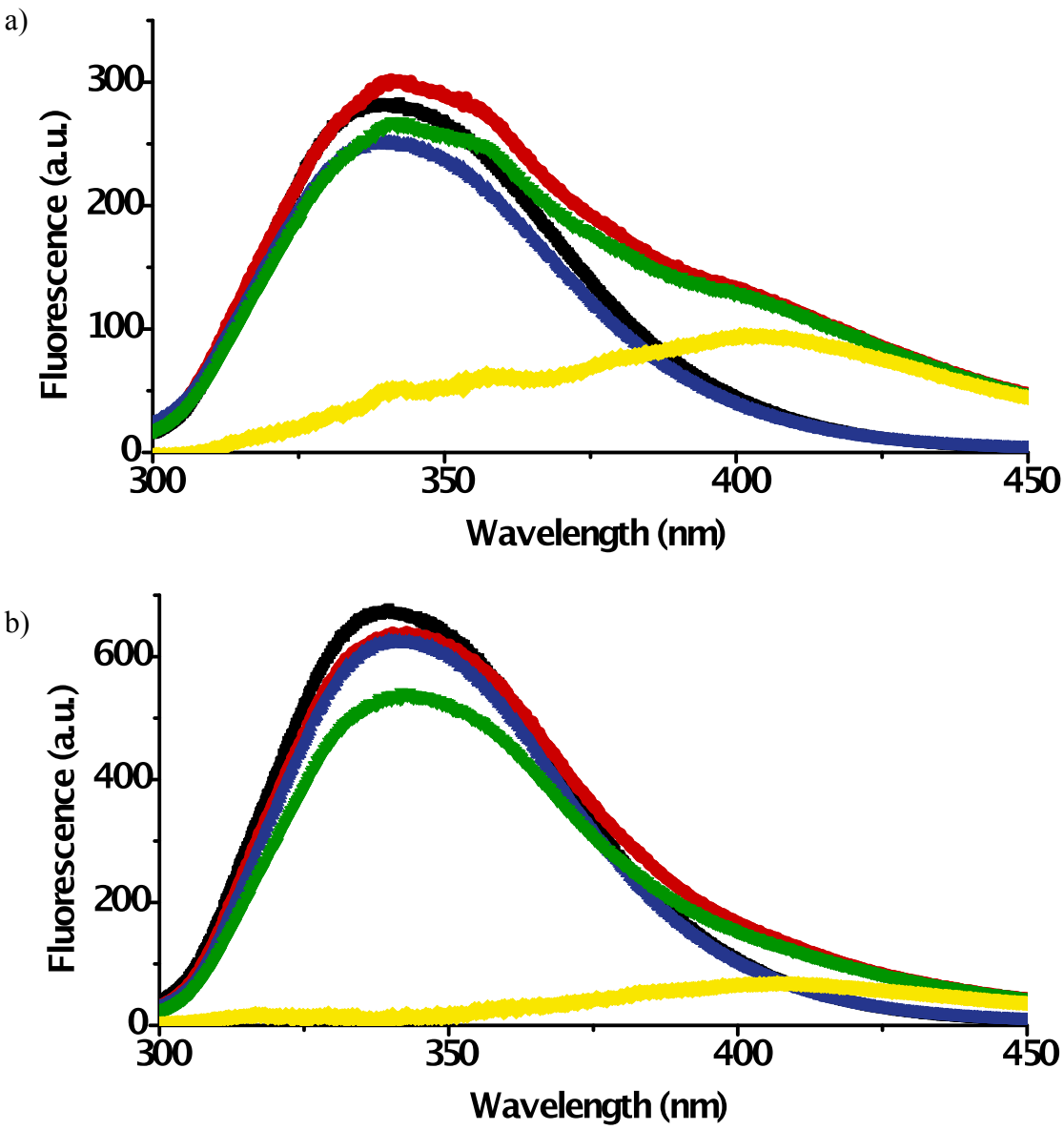


Figure 3. Fluorescence emission spectra obtained upon excitation at 275 nm of a) SCP-2L V83C (black), SCP-2L V83C-**N3** (red), + 1 equivalent of copper(II)nitrate (green), + 100 equivalents of EDTA (blue) and the difference spectrum of SCP-2L SCP-2L V83C-**N3** +1 equivalent of copper(II)nitrate and V83C-**N3** (yellow) b) SCP-2L A100C (black), SCP-2L A100C-**N3** (red), +1 equivalent of copper(II)nitrate (green), + 100 equivalents of EDTA (blue) and the difference spectrum of SCP-2L SCP-2L A100C-**N3** +1 equivalent of copper(II)nitrate and A100C-**N3** (yellow).

The fluorescence emission spectrum of SCP-2L V83C-**N3** upon excitation at 275 nm shows a peak at 336 nm and a shoulder with a maximum at 402 nm (Figure 3a) This shoulder is absent in the spectrum obtained for SCP-2L V83C and is thus assigned to

the conjugated phenanthroline. The shoulder disappears upon addition of one equivalent of copper(II)nitrate, affording a signal that corresponds closely to the emission spectrum of unmodified SCP-2L V83C. The addition of 100 equivalents of EDTA to the solution containing the copper adduct of SCP-2L V83C-**N5** provides a spectrum that is similar to that obtained for SCP-2L V83C-**N3,** indicating removal of the copper.

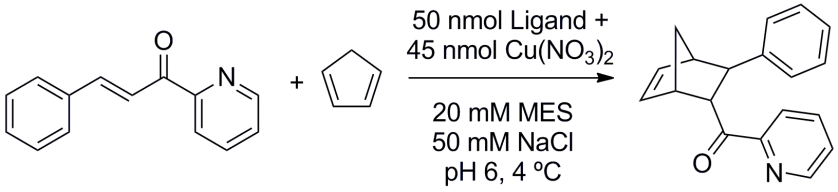
The fluorescence emission spectrum of SCP-2L A100C-**N3** upon excitation at 275 nm also shows a similar shoulder compared to that of SCP-2L V83C-**N3** (Figure 3b). The native fluorescence of SCP-2L A100C is significantly more intense than that of SCP-2L V83C however the intensity of the shoulder remains the same. This shoulder has a maximum of 407 nm. Also this shoulder disappears upon addition of copper and reappears upon the subsequent addition of EDTA.

These results demonstrate the coordination of copper to the phenanthrolines of SCP-2L V83C-**N3** and SCP-2L A100C-**N3**. Copper binding must be very specific for the introduced phenanthrolines as one equivalent of copper already completely quenches their signal. Therefore, it can be concluded that LCMS(ES^+) is an unsuitable technique for analysis of the copper adducts of protein-**N3**. These results suggest that it is likely that copper adducts of protein-**N4** conjugates are also formed. The application of other analysis procedures, for example EPR, might confirm this in the future.

5.3.3 Application of nitrogen donor ligand modified SCP-2L as catalyst in copper catalysed reactions

5.3.3.1 Asymmetric Diels-Alder reaction

We decided to test our nitrogen ligand modified SCP-2L mutants in the Diels-Alder reaction between azachalcone and cyclopentadiene as performed by Roelfes and Reetz using their bioinspired catalyst systems (Scheme 9).[6, 10a, 12-13]



Scheme 9. Copper catalysed Diels-Alder reaction using nitrogen donor ligands.

The results of the reaction using the modified proteins and a series of control reactions are displayed in Table 2. The conversion varied largely depending on the reaction but average to good conversions were observed using the artificial metalloenzymes. Overall, copper adducts with ligand **N3** gave lower conversion than those with **N4** (Table 2, compare entries 1, 3 and 4 with 2, 5 and 6). Higher Endo/Exo ratios were found for copper adduct of **N4** compared to **N3**, which demonstrated that also the Endo/Exo ratio was dependent on the nitrogen ligand used. Interestingly, dipicolylamine ligands have never been used for copper catalysed Diels-Alder reactions, but seem to give good conversion and Endo-Exo selectivity.

Table 2. Results of Diels-Alder reactions using nitrogen ligand modified SCP-2L mutants.[a]

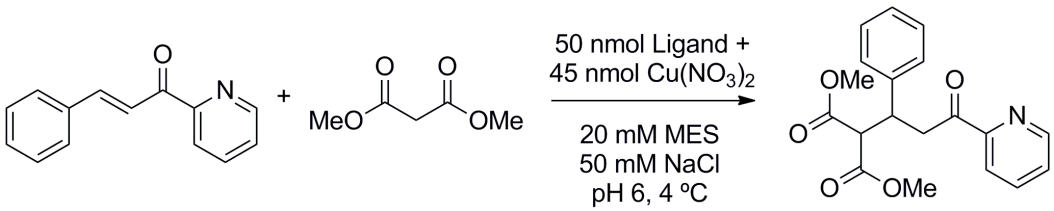
Entry	Ligand	Conversion (%)[b]	Endo/Exo[b]	Ee (%)[b]
1	**N3**	26.9 (±8.5)	5.7 (±0.4)	-
2	**N4**	85.1 (±4.7)	10.6 (±0.1)	-
3	SCP-2L V83C-**N3**	19.6 (±9.2)	7.6 (±2.4)	25.1 (±2.0)
4	SCP-2L A100C-**N3**	43.4 (±3.0)	11.7 (±2.4)	-
5	SCP-2L V83C-**N4**	86.8 (±8.7)	14.3 (±2.9)	4.1 (±2.7)
6	SCP-2L A100C-**N4**	76.3 (±8.9)	17.6 (±1.3)	-5.4 (±3.2)
7	-	90.3 (±7.1)	14.1 (±0.2)	-
8[c]	-	9.8 (±0.5)	6.0 (±0.5)	-
9	SCP-2L V83C	42.7 (±45.7)[d]	17.8 (±0.8)	-5.9 (±0.7)
10	SCP-2L A100C	65.7 (±7.4)	8.8 (±2.4)	8.1 (±1.4)
11[e]	SCP-2L	6.5 (±1.5)	3.4 (±0.6)	-

a) 50 nmol Ligand, 45 nmol $Cu(NO_3)_2$, 0.5 µmol azachalcone and 15 µmol cyclopentadiene in 20 mM MES 50 mM NaCl pH 6, total volume 0.8 ml, shaken for 72 hours at 4ºC b) Analysed by HPLC the ee is given for the endo product and a negative value corresponds to the (-)-enantiomer and a positive to the (+)-enantiomer, standard deviations are given between brackets c) no copper was added d) a large deviation in the conversion was observed with either ~90% or ~10% conversion e) total volume of 1.2 ml.

The reaction catalysed by "free" copper gave the highest conversion and a high Endo/Exo ratio (Table 2, entry 7). The non catalysed reaction also proceeds but with low conversion and a low Endo/Exo ratio (Table 2, entry 8). Product formation was also observed for reactions employing the unique cysteine SCP-2L templates themselves as ligands (Table 2, entries 9 and 10). A significant lower conversion was observed using native SCP-2L not containing any cysteine, which indicates that the free cysteine in SCP-2L V83C and SCP-2L A100C might be involved in copper coordination to give an active catalyst (Table 2, entry 11). On average, very low ee was found but the reaction using SCP-2L V83C-**N3** as catalyst gave a promising ee of 25%. Interestingly, the ee was found to be dependent on the protein template used even when the same nitrogen donor ligand was introduced. SCP-2L A100C-**N3** gave no ee while SCP-2L V83C-**N3** does and SCP-2L V83C-**N4** and SCP-2L A100C-**N4** give low ee but of the opposite enantiomer. The ee found for unmodified SCP-2L V83C and SCP-2L A100C is also interesting as selectivity for the opposite enantiomers was found. In addition, no ee was found for SCP-2L itself once again indicating involvement of the cysteine sulfide on the reaction. Overall, the results suggest that it may be possible to obtain highly enantioselective copper-catalysts by combining SCP-2L and different nitrogen donor ligands. Further optimisation using both changes to the protein template as well as the introduction of different nitrogen donor ligands might lead to more selective catalysts. Also the transition metal can be varied as demonstrated by a ruthenium system developed for Diels-Alder reactions by the group of Salmain.[24]

5.3.3.2 Michael addition reaction

The same catalyst systems as employed for the Diels-Alder reaction were also used for a copper catalysed 1,4- addition reaction (also called Michael addition reaction). This reaction has been successfully applied by the group of Roelfes for several of their copper based catalytic systems involving nitrogen donor ligands.[7] This reaction was performed using the same azachalcone substrate as for the Diels-Alder reactions with dimethylmalonate as nucleophile (Scheme 10).[10a].



Scheme 10. Copper catalysed Michael addition reaction using nitrogen donor ligands.

The reactions using ligands **N3** and **N4** provided only low conversions, with the former performing significantly better than the latter. The **N3** and **N4** modified protein templates performed better as ligands compared to ligands **N3** and **N4** by themselves giving conversions from 36 to 71%. Disappointingly, no ee was found using the protein based catalysts. These results show that either the selected protein environment or the used nitrogen cofactors make the developed catalyst system unsuitable to induce chirality in the reaction above. Further experimentation and optimisation are required to indicate the cause of this and to develop effective catalysts for this reaction.

Table 3. Result for the Michael addition reactions using nitrogen ligand modified SCP-2L mutants.[a]

Entry	Ligand[b]	Conversion (%)[b]
1	**N3**	24.7 (±4.5)
2	**N4**	5.8 (±2.3)
3	SCP-2L V83C-**N3**	36.5 (±7.1)
4	SCP-2L A100C-**N3**	70.7 (±6.1)
5	SCP-2L V83C-**N4**	69.0 (±5.8)
6	SCP-2L A100C-**N4**	61.4 (±8.8)
7	-	31.7 (±10.3)
8[c]	-	-

a) 50 nmol ligand, 45 nmol $Cu(NO_3)_2$, 0.5 µmol azachalcone and 50 µmol dimethylmalonate in 20 mM MES 50 mM NaCl pH 6, total volume 0.8 ml, shaken for 72 hours at 4ºC b) Analysed by HPLC c) no copper was added.

5.3.4. Structural characterization of the modified proteins

SCP-2L V83C-**N4** showed the remarkable results in the copper catalysed Diels-Alders reaction and was therefore studied by CD and watergate ^{1}H-NMR spectroscopy to get some insight in the effect of this modification on the protein fold. The near UV CD-spectra of SCP-2L V83C modified with nitrogen donor ligand **N4** was recorded and compared to that of SCP-2L V83C (Figure 4). The main change is the increased signal between 260 and 275 nm. This increase can be assigned to the introduced pyridine rings of the ligands interacting with the chiral environment of the protein. The increase in the signal between 260 and 275 nm is extended to the region between 250 and 260 nm indicating a change to the helical structure of the protein, which gives strong signals in the far UV between 200 and 250 nm. This was also observed for the modification of SCP-2L V83C with phosphine ligands (described in chapter 3) and can be ascribed to a conformational change in the cysteine-containing helix upon modification of this cysteine. The remainder of the spectrum indicated a loss of structure similar to that observed for the phosphine modifications, which is most clear for the signals at 287, 294 and 300 nm.

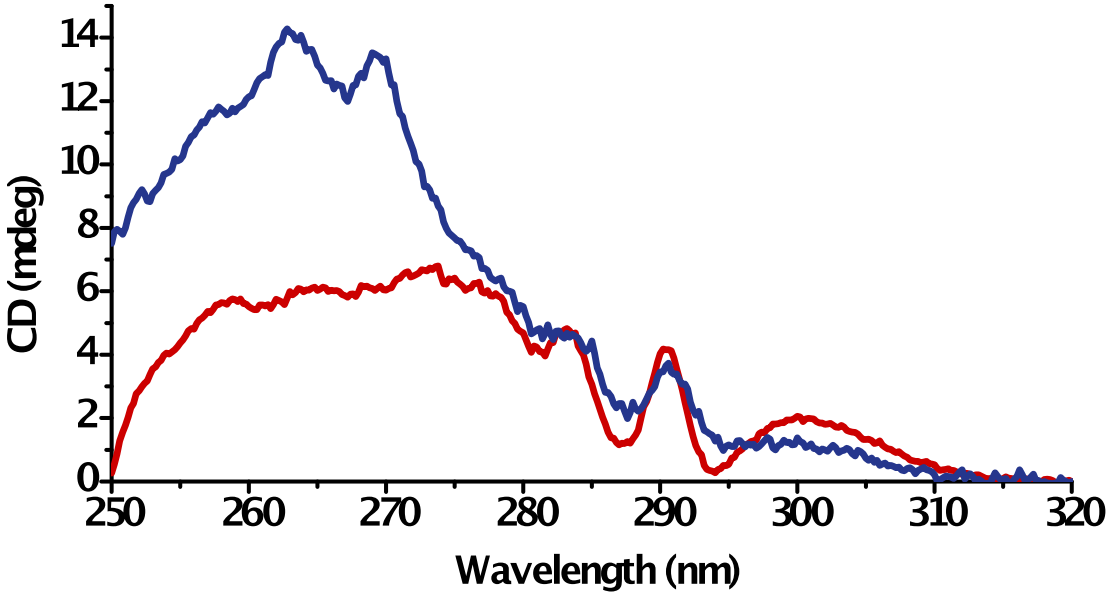


Figure 4. Near UV CD-spectra of SCP-2L V83C (red) and SCP-2L V83C-**N4** (blue).

The watergate ^{1}H-NMR spectra of SCP-2L V83C-**N4** was recorded and compared to that of SCP-2L V83C (Figure 5). This spectrum is recorded in phosphate buffer which does not have proton signals. Therefore no buffer signals are observed like the ones for MES in the spectrum of SCP-2L V83C at 3.61, 2.99 2.76 and 2.54 ppm. The effect

of the different buffers on the fold of the proteins is expected to be relatively small. The spectrum of SCP-2L V83C-**N4** seems only slightly more distorted than that of SCP-2L V83C. Some peaks have shifted slightly but overall the structure seems to be retained These results indicate some minor changes to the structural integrity of the protein.

The influence of the introduction of the ligand on the substrate binding capacity of the protein was assessed by a fluorescent binding study using the procedure described in chapter 2. The combined data provided binding constants (K_d) for the fluorescent probe Pyr-C12 of 0.25 (+/-0.04) µM for SCP-2L V83C-**N4**. This K_d value is in the same order of magnitude as the K_d of 0.18 (±0.07) µM found for SCP-2L V83C itself. This result indicates that the substrate binding capabilities of SCP-2L remain effectively the same after the modification with **N4**. From the results obtained from the different spectroscopic techniques it can be concluded that although the structure of SCP-2L V83C-**N4** is slightly altered compared to that of SCP-2L V83C these changes are only minor and do not alter the binding properties of the protein.

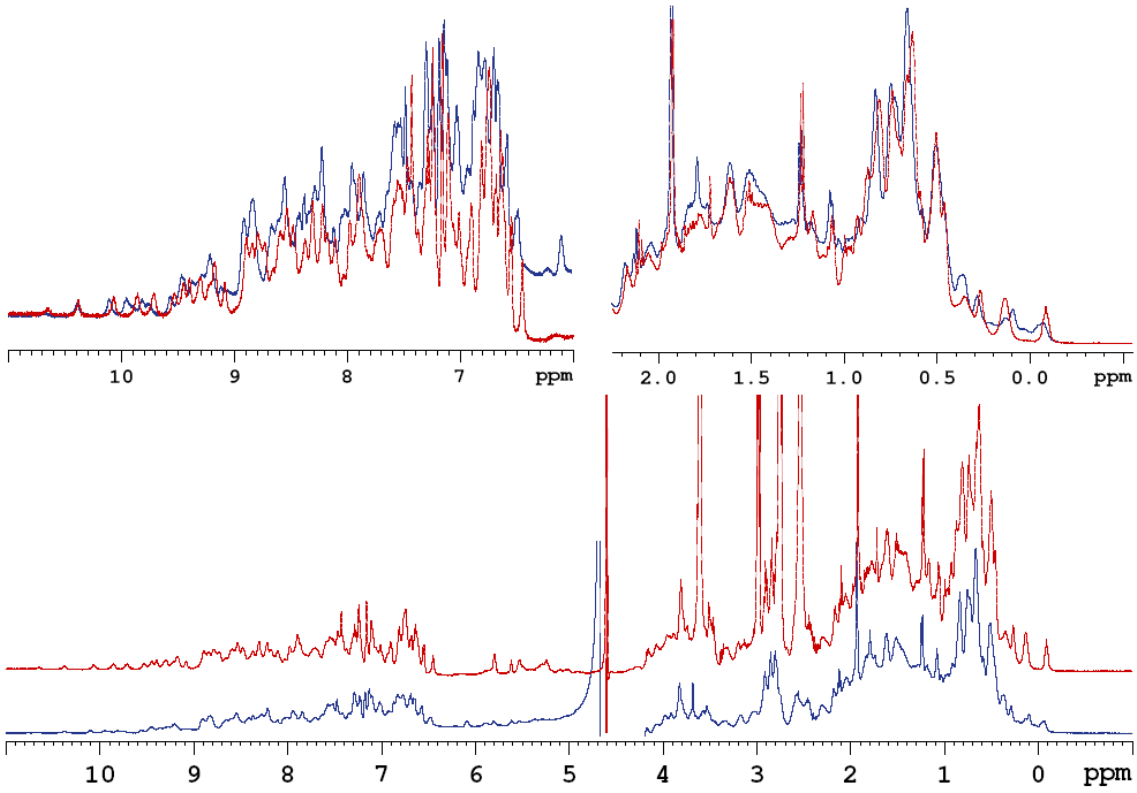


Figure 5. Watergate ^{1}H-NMR spectra of SCP-2L V83C (red) and SCP-2L V83C-**N4** (blue) with overlays of the areas between11 and 6 ppm and 2.3 to -0.5 ppm.

5.4 Conclusions and future prospects

The results in this chapter demonstrate how an enantioselective artificial metalloenzyme can be created from SCP-2L via the covalent introduction of nitrogen-ligands. Maleimide containing nitrogen donor ligands were found most suitable for cysteine selective introduction into SCP-2L. Unfortunately, the analysis of copper conjugates proved troublesome using LCMS(ES^+). Copper coordination was confirmed by fluorescence spectroscopy; however, this approach is limited to phenanthroline and other fluorescent ligands. Other techniques will have to be applied to confirm the formation of coordination products with other ligands. One such a technique is EPR which already has been shown in literature to be useful for characterisation of such systems. Although maleimides have the disadvantage of the formation of diastereomers, initial catalytic results showed promising results. The selectivity and activity of the Diels-Alder reaction between azachalcone and cyclopentadiene proved to be influenced by both the introduced nitrogen donor ligand and the protein environment. This makes the system using SCP-2L V83C-**N3** that gave the highest enantioselectivity of 25% ee a good starting point for further optimisation. The results found for the asymmetric Michael addition of dimethylmalonate to azachalcone were disappointing. However, more suitable catalysts for this reaction might be created by further optimised systems. These catalyst systems can be optimised by the introduction of other nitrogen donor ligands in combination with optimisation of the SCP-2L scaffold.

5.5 Experimental

5.5.1 General

5.5.1.1 Methods

Where required, reactions were performed under an argon atmosphere using degassed solvents and standard Schlenk techniques. For the fluorescent binding study the procedure described in section 2.3.3.3 was used. Protein concentrations were determined using Bradford's reagent.[25]

5.5.1.2 Equipment

NMR spectra were taken at room temperature with Bruker Avance 300, 400 or 500 NMR spectrometers. Chemical shifts (δ) are given (in ppm) for high-frequency shifts

relative to a TMS reference (^{1}H and ^{13}C) or 85% H_3PO_4 (^{31}P). Watergate ^{1}H-NMR spectra were recorded on a Bruker Avance 500 spectrometer. LC-MS(ES^+) used for analysis of protein and protein 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 2.1x 10 mm cartridge using a gradient of 1% formic acid in H_2O to 1% formic acid in acetonitrile. For processing of the mass-spectra the MaxEnt algorithm of Masslynx V4.0 (Micromass Ltd) was used. Spin-filter concentrators with a molecular weight cut-off of 10kDa (Amicon® Ultra-15 Centrifugal Filter Units, Millipore®, Carrigtwohill, Co. Cork, Ireland) were used to concentrate and/or wash protein solutions. CD spectra were recorded using a Jasco J-810 spectropolarimeter using a 1 mg/ml solution of the protein. Elemental analysis was performed by Stephen Boyer at the elemental analysis service of the London Metropolitan University. For the fluorescence measurements a Varian Cary Eclipse apparatus was used with a 1 cm cuvette.

5.5.1.3 Materials

For dry conditions freshly distilled degassed solvents were used: Dichloromethane (DCM) and acetonitrile (MeCN) were distilled from CaH_2. Toluene was distilled from sodium/benzophenone. Dry DMF was purchased from Acros in a sealed bottle containing 4Å molecular sieves and was used as received. Dipicolyl amine and β-maleoyl propionic acid were obtained from Aldrich. Other reagents were obtained from commercial sources in the highest purity. Thin-layer chromatography (TLC) was performed on silica (Polygram 0.2mm silica gel with fluorescent indicator UV_{254}). Silica gel (60 particle size 0.063-0.2 mm) was used for column chromatography.

5.5.2 Synthesis of dipicolylamine based nitrogen donor ligands

*5.5.2.1 Synthesis of N,N-dipicolyl-N'-maleoyl-3-aminopropanamide **N4***

Under an argon atmosphere 138.16 mg (0.75 mmol) dipicolylamine and 138.34 mg (0.82 mmol) *N*-maleoyl-β-analine were dissolved in 10 ml dry DCM. This mixture was cooled to 0ºC and 110.8 (0.82 mmol) 1-hydroxybenzotriazole and 169.2 mg (0.82 mmol) *N,N*-dicyclohexylcarbodiimide were added. The reaction was stirred 1H at 0ºC and than let to heat up to room temperature. Already after 5 minutes white precipitate

had formed and the reaction was followed by TLC. Upon completion after 7h the white powder was filtered off and the DCM was removed *in vacuo* from the filtrate. A small amount of EtOAc was added to the remaining white solid so that all **N4** dissolved and put in the freezer overnight. The white solid that formed was filtered off and the EtOAc was removed from the filtrate by evaporation. The remaining solid was purified by flash column chromatography using 150 ml silica and 5% MeOH in DCM with a few drops of Et_3N. Combination of the fractions and evaporation of the solvent resulted in 250 mg (0.714 mmol, 95%) **N4** as an off-white oily solid, which slowly turned into a hard solid over a few months. ^{1}H-NMR (300 MHz, $CDCl_3$), δ 8.48 (ddd, $^3J_{HH}$=4.8 Hz, J_{HH}=1.7 Hz, J_{HH}=0.9 Hz, 1H, Py-*H*), δ 8.43 (ddd, $^3J_{HH}$=4.9 Hz, J_{HH}=1.7 Hz, J_{HH}=0.9 Hz, 1H, Py-*H*), δ 7.62-7.53 (m, 2H, Py-*H*), δ 7.23 (d, $^3J_{HH}$=7.8, 1H, Py-*H*), δ 7.15-7.06 (m, 3H, Py-*H*), δ 6.60 (s, 2H, Mal-*H*), δ 4.68 (s, 2H, Py-C*H*$_2$-N), δ 5.49 (s, 2H, Py-C*H*$_2$-N), δ 3.89-3.82 (m, 2H, N-C*H*$_2$-CH_2-CO-N), δ 2.81-2.76 (m, 2H, N-CO-C*H*$_2$-CH_2-); ^{13}C-NMR (75.47 MHz, $CDCl_3$), δ 170.9 (C_q, CH_2-*C*O-N), δ 170.5 (C_q, CH-*C*O-N), δ 157.9 (C_q, Py*C*-CH_2), δ 156.2 (C_q, Py*C*-CH_2), δ 150.0 (CH, Py*C*-H), δ 150.0 (CH, Py*C*-H), δ 149.2 (CH, Py*C*-H), δ 136.9 (CH, Py*C*-H), δ 136.8 (CH, Py*C*-H), δ 134.2 (CH, Mal*C*-H), δ 122.7 (CH, Py*C*-H), δ 122.6 (CH, Py*C*-H), δ 122.4 (CH, Py*C*-H), δ 121.0 (CH, Py*C*-H), δ 51.2 (CH_2, Py-*CH*$_2$-N), δ 52.9 (CH_2, Py-*CH*$_2$-N), δ 34.2 (CH_2, N-*C*H_2-CH_2-CO-N), δ 31.4 (CH_2, N-CO-*C*H_2-CH_2-); IR-spectrum (PTFE card), 3066.5 cm^{-1} (w, CH stretch band), 2925.5 cm^{-1} (w, CH stretch band), 1702.9 cm^{-1} (s, amide CO stretch band), 1642.7 cm^{-1} (s, amide CO stretch band). Elemental analysis calculated for $C_{19}H_{318}N_4O_3$: C 65.13, H 5.18, N 15.99; found: C 65.21, H 5.30, N 15.92.

5.5.3 Bioconjugation

*5.5.3.1 CDI activation of carboxylic acid ligand **N1***

This procedure is based on a procedure reported in literature by our group[23b] 5-10 Equivalents of CDI was added as a solid to a 25 mM solution of the corresponding carboxylic acid ligand in dry DMF. The mixture was stirred for 3 or more hours under argon atmosphere. The mixture was used as such for the bioconjugation to SCP-2L V83C.

*3.5.3.2 Bioconjugation of activated carboxylic acid phosphines (**N1-N2**) to SCP-2L V83C*

This procedure is based on a procedure reported in literature by our group.[23b] A 15 μM solution of SCP-2L V83C in buffer (20 mM MES pH 6 or 7) was incubated with 2-11 equivalents of the activated carboxylic acid ligand added from a 25 mM stock solution in DMF. The reaction was stirred for 16 hours and the conversion determined by LCMS(ES^+). Data obtained by mass-spectroscopy: SCP-2L V83C-**N1** (expected 13565.7 Da, found 13565.8 Da) and SCP-2L V83C-**N2** (expected 13558.8 Da, found 13558.6 Da).

3.5.3.3 General procedure for bioconjugation maleimides 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 maleimide added from a 50 mM stock solution in the corresponding buffer or DMF for water insoluble maleimides. Progress of the reaction was monitored using LCMS (ES^+) analysis. Typically no residual free protein was observed after leaving the reaction overnight. The mixture was washed using a centrifugal concentrator with buffer to remove the excess maleimide. Data obtained by mass-spectroscopy: SCP-2L V83C-**N3** (expected 13651.9 Da, found 13651.6 Da), SCP-2L V83C-**N4** (expected 13727.0 Da, found 13727.1 Da), SCP-2L A100C-**N3** (expected 13679.9 Da, found 13679.7 Da) and SCP-2L A100C-**N4** (expected 13755.0 Da, found 13755.1 Da).

3.5.4 Fluorescence spectroscopy

Fluorescence emission spectra were obtained by excitation at 275 nm. A 50 μM solution of the protein or modified protein in 20 mM MES 50 mM NaCl pH 6 was used for the measurements. 50 μM copper(II)nitrate was added from a 2 mM solution in the same buffer. 5 mM EDTA was added from a 1 M stock solution in deionized water. The obtained solutions were mixed well after the addition of each solution. Equipment settings used: 5 nm excitation and emission slits, PMT of 600 V, scan rate 50 nm/min, data interval 0.1667 nm, temperature 25 ºC and detection range 300-450 nm.

3.5.5 Catalysis procedures

3.5.5.1 Diels-Alder reactions

A solution of 50 nmol ligand and 45 nmol copper(II)nitrate in 794 μL 20 mM MES 50 mM NaCl pH 6 buffer was mixed for 1 hour prior to the catalytic reactions.. The ligands **N3** and **N5** were added from 50 mM solutions in DMSO while free proteins and protein conjugates were added from solution in 20 mM MES 50 mM NaCl pH 6 buffer of which the concentration was determined before use after removal of formed precipitate by centrifugation. 5 μL azachalcone from a 100 mM solution in MeCN and 1 μL freshly cracked and distilled cyclopentadiene were added to the reaction mixture. The obtained solution was shaken for 72 hours in a cold room (4ºC). The obtained solutions were extracted three times with about 1 ml diethylether and the combined organic layers dried *en vacuo*. The obtained solid was dissolved in 200 μL 1% IPA in hexanes, which was used for HPLC analysis. HPLC analysis was performed using an OD-H column with a flow of 1 ml/min of 1% IPA in hexanes (retention times: Exo products 13.8 and 15.7 minutes, Endo products 17.5 and 23.1 minutes and azachalcone 26.9 minutes). The chromatogram was obtained by UV-analysis at 215, 230, 254 and 280 nm which all give similar data. The data obtained at 230 nm was used for the results displayed in Table 2 as these results gave the smallest standard deviation for the obtained ee. Conversion was calculated using a calibration using standard samples with different ratios between azachalcone and the product. The enantiomers were identified using literature data.[26]

3.5.5.2 Michael Addition reactions

A solution of 50 nmol ligand and 45 nmol copper(II)nitrate in 794 μL 20 mM MES 50 mM NaCl pH 6 buffer was mixed for 1 hour prior to the reactions. The ligands **N3** and **N5** were added from 50 mM solutions in DMSO while free proteins and protein conjugates were added from solution in 20 mM MES 50 mM NaCl pH 6 buffer of which the concentration was determined before use after removal of formed precipitate by centrifugation. 5 μL azachalcone from a 100 mM solution in MeCN and 0.5 μL dimethylmalonate were added to the reaction mixture. The obtained solution was shaken for 72 hours in a cold room (4 ºC). The obtained solutions were extracted with diethylether and the combined organic layers dried *en vacuo*. The obtained solid was dissolved 200 μL 5% IPA in hexanes, which was used for HPLC analysis. HPLC

analysis was performed using an OD-H column with a flow of 1 ml/min of 5% IPA in hexanes (retention times: azachalcone 16.1 and 19.3 minutes and products 40.9 and 45.3 minutes). The chromatogram was obtained by UV-analysis at 215, 230, 254 and 280 nm. The data obtained at 215 nm was used for the results displayed in Table 3 as the highest peak intensities were obtained. Conversion was calculated using a calibration using standard samples with different ratios between azachalcone and the product.

5.6 References

[1] a) M. E. Wilson, G. M. Whitesides, *J. Am. Chem. Soc.* **1978**, *100*, 306; b) J. Steinreiber, T. R. Ward, *Topics in Organometallic Chemistry* **2009**, *25*, 93 and reference herein, c) M. T. Reetz, J. J. P. Peyralans, A. Maichele, Y. Fu, M. Maywald, *Chem. Commun. (Cambridge, U. K.)* **2006**, 4318; d) H. Yamaguchi, T. Hirano, H. Kiminami, D. Taura, A. Harada, *Org. Biomol. Chem.* **2006**, *4*, 3571.

[2] a) A. Pordea, M. Creus, J. Panek, C. Duboc, D. Mathis, M. Novic, T. R. Ward, *J. Am. Chem. Soc.* **2008**, *130*, 8085; b) A. Pordea, T. R. Ward, *Synlett* **2009**, 3225; c) A. Mahammed, H. B. Gray, J. J. Weaver, K. Sorasaenee, Z. Gross, *Bioconjugate Chem.* **2004**, *15*, 738; d) A. Mahammed, Z. Gross, *J. Am. Chem. Soc.* **2005**, *127*, 2883; e) M. Ohashi, T. Koshiyama, T. Ueno, M. Yanase, H. Fujii, Y. Watanabe, *Angew. Chem., Int. Ed.* **2003**, *42*, 1005; f) T. Ueno, T. Koshiyama, S. Abe, N. Yokoi, M. Ohashi, H. Nakajima, Y. Watanabe, *J. Organomet. Chem.* **2007**, *692*, 142; g) S. Nimri, E. Keinan, *J. Am. Chem. Soc.* **1999**, *121*, 8978; h) J.-P. Mahy, B. Desfosses, S. De Lauzon, R. Quilez, B. Desfosses, L. Lion, D. Mansuy, *Appl. Biochem. Biotechnol.* **1998**, *75*, 103; i) R. Ricoux, R. Dubuc, C. Dupont, J.-D. Marechal, A. Martin, M. Sellier, J.-P. Mahy, *Bioconjugate Chem.* **2008**, *19*, 899; j) R. Ricoux, E. Lukowska, F. Pezzotti, J.-P. Mahy, *Eur. J. Biochem.* **2004**, *271*, 1277; k) Q. Raffy, R. Ricoux, E. Sansiaume, S. Pethe, J.-P. Mahy, *J. Mol. Catal. A: Chem.* **2010**, *317*, 19; l) E. Sansiaume, R. Ricoux, D. Gori, J.-P. Mahy, *Tetrahedron: Asymmetry* **2010**, *21*, 1593; m) 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; n) J.-L. Zhang, D. K. Garner, L. Liang, Q. Chen, Y. Lu, *Chem. Commun. (Cambridge, U. K.)* **2008**, 1665.

[3] G. T. Hermanson, *Bioconjugate Techniques*, Second ed., Acadamic Press, San Diego, **2008**.

[4] For various examples see: a) H. S. Lee, P. G. Schultz, *J. Am. Chem. Soc.* **2008**, *130*, 13194; b) C. Q. Pan, J. A. Feng, S. E. Finkel, R. Landgraf, D. Sigman, R. C. Johnson, *Proc. Natl. Acad. Sci. U. S. A.* **1994**, *91*, 1721; c) R. H. Ebright, Y. W. Ebright, P. S. Pendergrast, A. Gunasekera, *Proc. Natl. Acad. Sci. U. S. A.* **1990**, *87*, 2882; d) J. C. Francois, T. Saison-Behmoaras, M. Chassignol, T. Nguyen Thanh, J. S. Sun, C. Helene, *Biochemistry* **1988**, *27*, 2272; e) D. S. Sigman, C. H. Chen, M. B. Gorin, *Nature (London, U. K.)* **1993**, *363*, 474; f) J. L. Czlapinski, T. L. Sheppard, *Chem. Commun. (Cambridge, U. K.)* **2004**, 2468.

[5] A. J. Boersma, R. P. Megens, B. L. Feringa, G. Roelfes, *Chem. Soc. Rev.* **2010**, *39*, 2083.
[6] a) G. Roelfes, A. J. Boersma, B. L. Feringa, *Chem. Commun. (Cambridge, U. K.)* **2006**, 635; b) G. Roelfes, B. L. Feringa, *Angew. Chem., Int. Ed.* **2005**, *44*, 3230.
[7] D. Coquiere, B. L. Feringa, G. Roelfes, *Angew. Chem., Int. Ed.* **2007**, *46*, 9308.
[8] A. J. Boersma, B. L. Feringa, G. Roelfes, *Angew. Chem., Int. Ed.* **2009**, *48*, 3346.
[9] A. J. Boersma, D. Coquiere, D. Geerdink, F. Rosati, B. L. Feringa, G. Roelfes, *Nat. Chem.* **2010**, *2*, 991.
[10] a) R. P. Megens, G. Roelfes, *Org. Biomol. Chem.* **2010**, *8*, 1387; b) A. J. Boersma, J. E. Klijn, B. L. Feringa, G. Roelfes, *J. Am. Chem. Soc.* **2008**, *130*, 11783.
[11] M. T. Reetz, M. Rentzsch, A. Pletsch, M. Maywald, P. Maiwald, J. J. P. Peyralans, A. Maichele, Y. Fu, N. Jiao, F. Hollmann, R. Mondiere, A. Taglieber, *Tetrahedron* **2007**, *63*, 6404.
[12] M. T. Reetz, N. Jiao, *Angew. Chem., Int. Ed.* **2006**, *45*, 2416.
[13] J. Podtetenieff, A. Taglieber, E. Bill, E. J. Reijerse, M. T. Reetz, *Angew. Chem., Int. Ed.* **2010**, *49*, 5151.
[14] N. S. Oltra, G. Roelfes, *Chem. Commun. (Cambridge, U. K.)* **2008**, 6039.
[15] R. R. Davies, M. D. Distefano, *J. Am. Chem. Soc.* **1997**, *119*, 11643.
[16] R. R. Davies, H. Kuang, D. Qi, A. Mazhary, E. Mayaan, M. D. Distefano, *Bioorg. Med. Chem. Lett.* **1999**, *9*, 79.
[17] a) A. M. Haapalainen, D. M. F. van Aalten, G. Merilaeinen, J. E. Jalonen, P. Pirilae, R. K. Wierenga, J. K. Hiltunen, T. Glumoff, *J. Mol. Biol.* **2001**, *313*, 1127; b) A. M. Gallegos, B. P. Atshaves, S. M. Storey, O. Starodub, A. D. Petrescu, H. Huang, A. L. McIntosh, G. G. Martin, H. Chao, A. B. Kier, F. Schroeder, *Prog. Lipid Res.* **2001**, *40*, 498; c) F. Schroeder, B. P. Atshaves, A. L. McIntosh, A. M. Gallegos, S. M. Storey, R. D. Parr, J. R. Jefferson, J. M. Ball, A. B. Kier, *Biochim. Biophys. Acta, Mol. Cell Biol. Lipids* **2007**, *1771*, 700; d) U. Seedorf, P. Ellinghaus, J. R. Nofer, *Biochim. Biophys. Acta, Mol. Cell Biol. Lipids* **2000**, *1486*, 45.
[18] M. Maywald, Ph.D. Thesis, Universität Bochum (Mülheim an der Ruhr, Germany), **2005**.
[19] 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.
[20] U. Sampath, W. C. Putnam, T. A. Osiek, S. Touami, J. Xie, D. Cohen, A. Cagnolini, P. Droege, D. Klug, C. L. Barnes, A. Modak, J. K. Bashkin, S. S. Jurisson, *J. Chem. Soc., Dalton Trans.* **1999**, 2049.
[21] a) S. A. Trammell, H. M. Goldston, Jr., P. T. Tran, L. M. Tender, D. W. Conrad, D. E. Benson, H. W. Hellinga, *Bioconjugate Chem.* **2001**, *12*, 643; b) A. Taglieber, Ph.D. Thesis, Universität Bochum (Mülheim an der Ruhr, Germany), **2007**.
[22] For examples see: a) N. Lazarova, J. Babich, J. Valliant, P. Schaffer, S. James, J. Zubieta, *Inorg. Chem.* **2005**, *44*, 6763; b) S. R. Banerjee, P. Schaffer, J. W. Babich, J. F. Valliant, J. Zubieta, *Dalton Trans.* **2005**, 3886; c) R. J. Butcher, A. W. Addison, *Inorg. Chim. Acta* **1989**, *158*, 211; d) S. Bhattacharya, K. Snehalatha, V. P. Kumar, *J. Org. Chem.* **2003**, *68*, 2741; e) C. Incarvito, A. L.

Rheingold, A. L. Gavrilova, C. J. Qin, B. Bosnich, *Inorg. Chem.* **2001**, *40*, 4101; f) N. Niklas, F. W. Heinemann, F. Hampel, T. Clark, R. Alsfasser, *Inorg. Chem.* **2004**, *43*, 4663; g) J. Mukherjee, V. Balamurugan, R. Gupta, R. Mukherjee, *Dalton Trans.* **2003**, 3686; h) K. Visvaganesan, R. Mayilmurugan, E. Suresh, M. Palaniandavar, *Inorg. Chem. (Washington, DC, U. S.)* **2007**, *46*, 10294.

[23] a) R. den Heeten, B. K. Munoz, G. Popa, W. Laan, P. C. J. Kamer, *Dalton Trans.* **2010**, *39*, 8477; b) W. Laan, B. K. Munoz, R. den Heeten, P. C. J. Kamer, *ChemBioChem* **2010**, *11*, 1236.

[24] B. Talbi, P. Haquette, A. Martel, F. de Montigny, C. Fosse, S. Cordier, T. Roisnel, G. Jaouen, M. Salmain, *Dalton Trans.* **2010**, *39*, 5605.

[25] M. M. Bradford, *Anal. Biochem.* **1976**, *72*, 248.

[26] D. Coquiere, J. Bos, J. Beld, G. Roelfes, *Angew. Chem., Int. Ed.* **2009**, *48*, 5159.

Chapter 6. Summary, conclusion and outlook

This thesis describes the design, synthesis and application of artificial metalloenzymes for transition metal catalysed reactions not performed by natural enzymes (Figure 1). The artificial metalloenzymes were designed to contain a hydrophobic cavity for binding selected substrates. The proteins were then modified covalently with a transition metal ligand complex which generates catalytic activity. The creation of a shape selective catalyst is envisioned by the combination of the selective substrate binding of the protein and the strategic position of the transition metal centre. The key difference between this approach and several other reported successful applications of artificial metalloenzymes lies in the virtual unlimited choice of proteins for binding of specific substrates. This freedom of choice is facilitated by the application of the covalent modification strategy which is a key factor for the successful application of this approach. Additionally, the formed catalysts are compatible with the use of both synthetic (e.g. optimisation of ligand structure) and biomolecular tools (e.g. optimisation of protein environment) for optimisation, which has previously been proven very powerful for the development of artificial metalloenzymes.

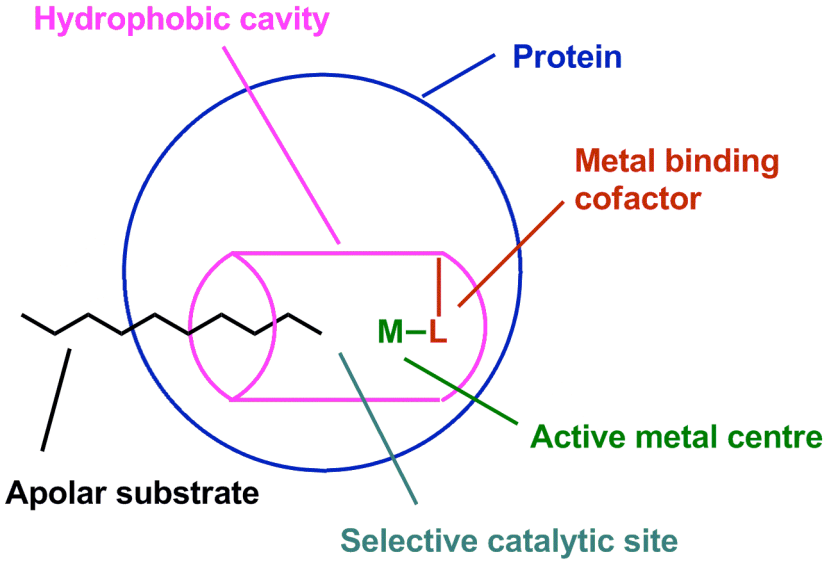


Figure 1. Concept for the design and creation of shape selective artificial metalloenzymes.

Chapter 2 describes the design and synthesis of unique cysteine containing protein templates. Sterol carrier proteins were selected as the class of proteins of choice for the use as protein templates because of their affinity for linear aliphatic substrates. One such protein, the protein sterol carrier protein 2 like (SCP-2L) domain of multifunctional enzyme (MFE) has the advantage that its reported crystal structure clearly reveals the hydrophobic binding site (Figure 2 a and b). Unique cysteines were introduced at strategic locations around the binding site in SCP-2L by site directed mutagenesis to serve as covalent anchoring sites for the transition metal complex. The SCP-2L mutants V83C and A100C were successfully produced at large scale (~60 and ~30 mg/L bacterial culture respectively) and conveniently purified using nickel affinity chromatography (Figure 2c and d). The protein structures of the mutations were studied using CD and watergate ^{1}H-NMR spectroscopy and were found nearly unchanged compared to wild-type SCP-2L.

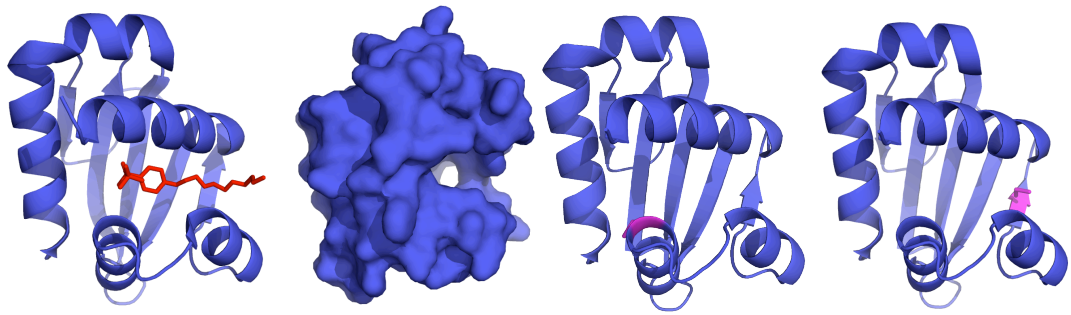
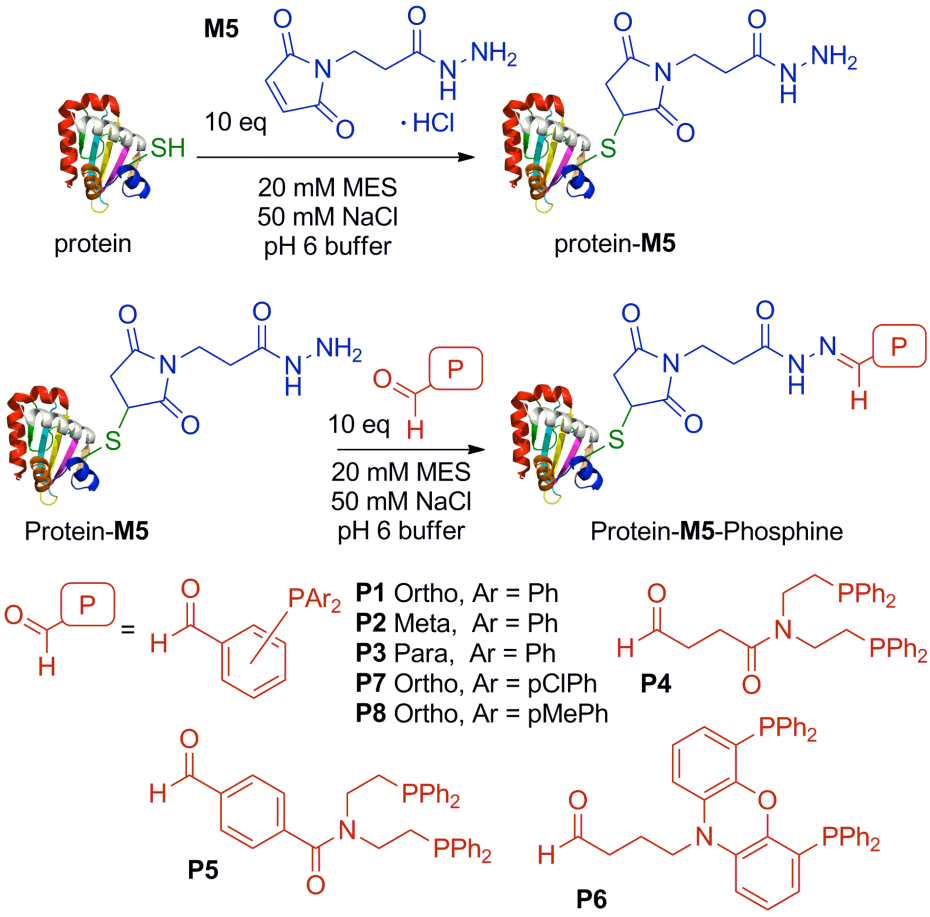


Figure 2. Cartoon and surface models of SCP-2L, a) Crystal structure showing part of a triton-X-100 molecule (red) in the binding site b) side-view of the crystal structure showing, upon removal of the substrate, the hydrophobic tunnel crossing through the protein c) unique cysteine containing template SCP-2L V83C d) unique cysteine containing template SCP-2L A100C (introduced cysteines for covalent modification shown in purple).

Chapter 3 describes the bioconjugation of phosphine ligands to the created unique cysteine protein templates. The covalent introduction of phosphine ligands proved problematic due to the similarity in nucleophilicity of the phosphines compared to the thiol of the cysteine. This led to the development of a two step modification procedure (Scheme 1). The unique cysteine is selectively modified in the first step with a maleimide-hydrazide to introduce an hydrazide functionality into the protein. This hydrazide is modified in the second step using phosphino-aldehydes to give phosphine modified proteins. The procedure proved most successful for benzaldehyde

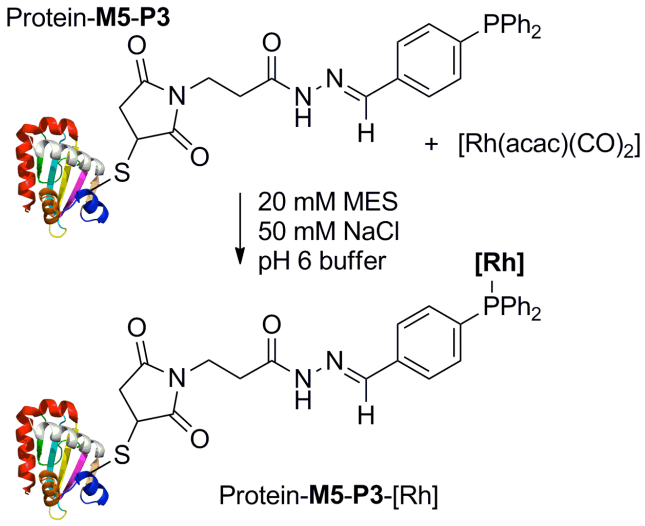
containing phosphines and was also used in our group to successfully modify a variety of other unique cysteine containing proteins. This is the first and so far only general applicable covalent modification method developed for the modification of proteins with phosphines. This key discovery opens up new possibilities for the application of phosphine modified artificial metalloenzymes in addition to overcoming a vital problem in this stage of development of the artificial metalloenzymes presented in this thesis. Although, it was shown by CD and watergate ^{1}H-NMR spectroscopy that the structure of SCP-2L was significantly altered by the phosphine modifications the constructs did still possess good substrate binding capability for a fatty acid labelled with a fluorescent probe.



Scheme 1. Two step modification procedure for the introduction of phosphine ligands to unique cysteine containing protein templates.

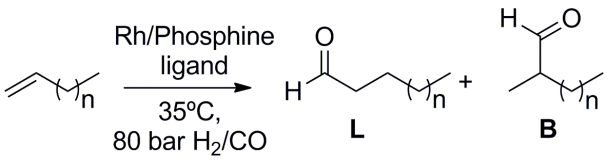
Chapter 4 describes the formation of rhodium adducts of the phosphine modified proteins and application of these systems in the rhodium catalysed hydroformylation of α-olefins. Rhodium complexes of several phosphine modified proteins were obtained by addition of dicarbonyl rhodium acetylacetonate ([Rh(acac)(CO)$_2$],

Scheme 2). Although, rhodium addition to the phosphine of the protein constructs was confirmed by ^{31}P-NMR and a peak corresponding to the addition of RhCO to the phosphine modified proteins was observed upon LCMS(ES^{+}) analysis, the exact nature and composition of the rhodium adducts remains unclear. These studies showed that it is likely that a mixture of different rhodium adducts is present.



Scheme 2. Rhodium complexation to phosphine-protein constructs.

The rhodium-phosphine containing artificial metalloenzymes were applied in the biphasic hydroformylation of linear alkenes (Scheme 3). The aqueous biphasic hydroformylation of linear alkenes (n > 3) using water soluble rhodium catalysts and typical hydroformylation conditions suffers from severe activity problems due to the limited solubility of these substrates in water. This was reflected by the low activity found for this system using triphenylphosphine trisulfonate (TPPTS) as water soluble phosphine ligand (Figure 3). This system only showed activity for 1-hexene but for higher alkenes (n > 5) the found TON's were close to 0. Much higher activity was found using SCP-2L V83C-**M5**-**P3**-[Rh] and SCP-2L A100C-**M5**-**P3**-[Rh] were used as catalysts and also good selectivities were obtained. Activity was still even observed for the hydroformylation of 1-octadecene using these catalysts. The difference in activity and selectivity between the two applied phosphine-rhodium based artificial metalloenzymes suggests that further optimisation of the location of the cysteine might lead to further improved catalysts. The observed activity and selectivity for the linear product point towards a true shape selective catalyst system in which the protein encapsulates the substrate, thereby solubilizing it in the aqueous phase, and creating a preference for the formation of the linear product.



Scheme 3. Hydroformylation of linear alkenes using rhodium-phosphine complexes.

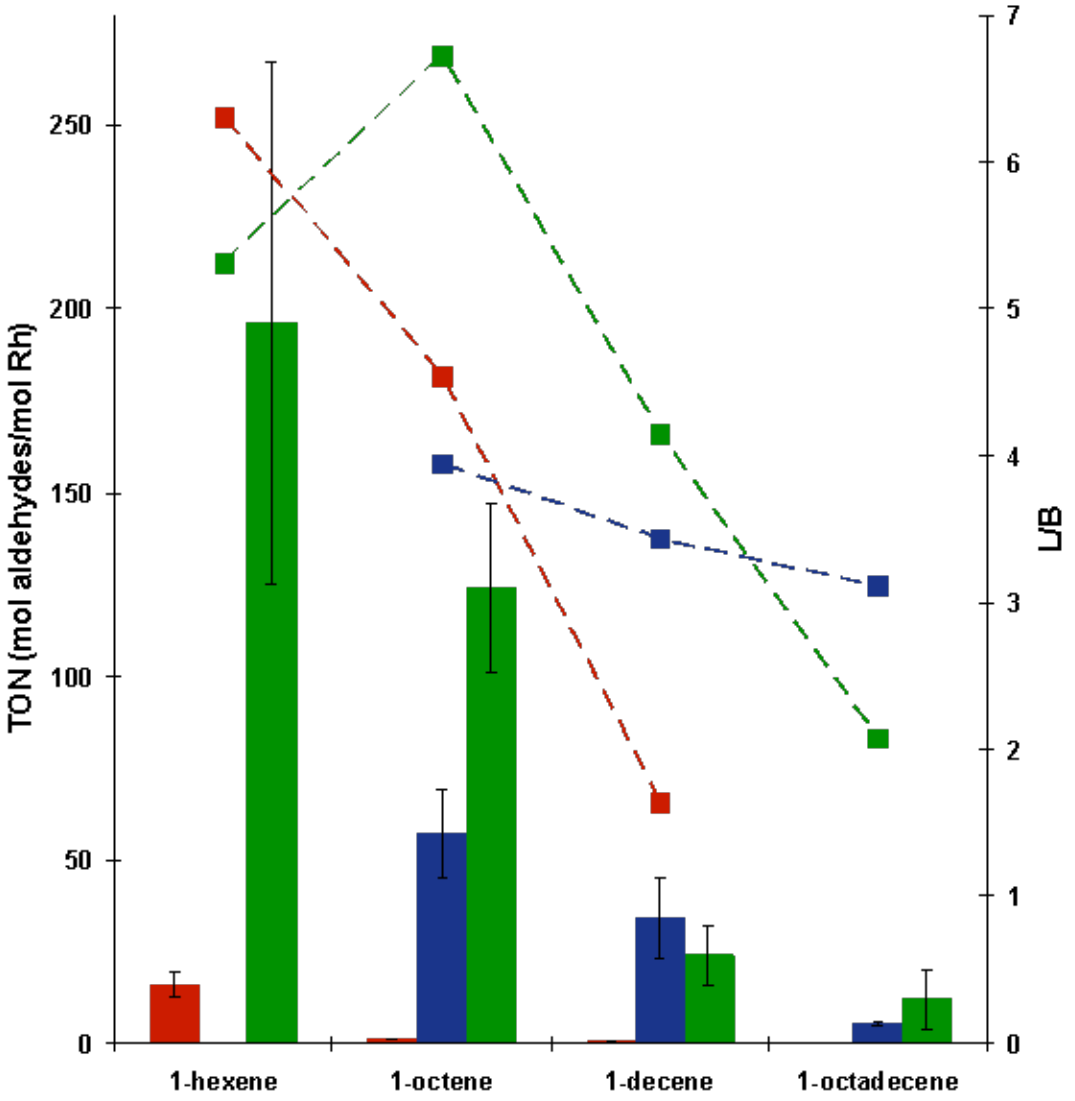
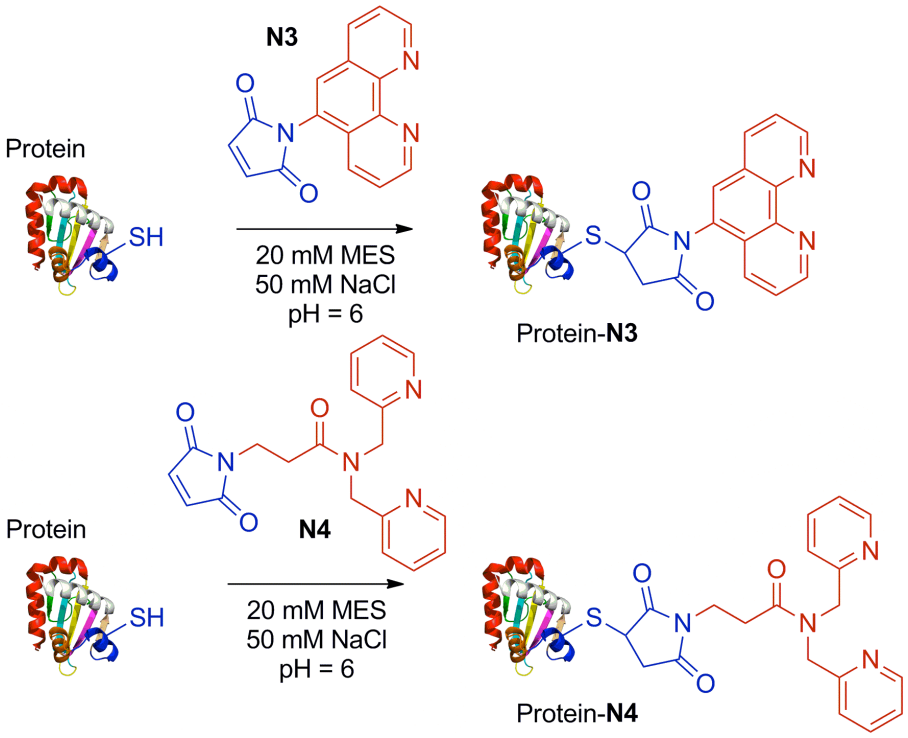


Figure 3. Comparison of the activity (bars) and selectivity (dots and dashed lines) of Rh:TPPTS 1:20 (red), SCP-2L V83C-**M5**-**P3**-[Rh] (blue) and SCP-2L A100C-**M5**-**P3**-[Rh] (green) in the biphasic hydroformylation of alkenes.

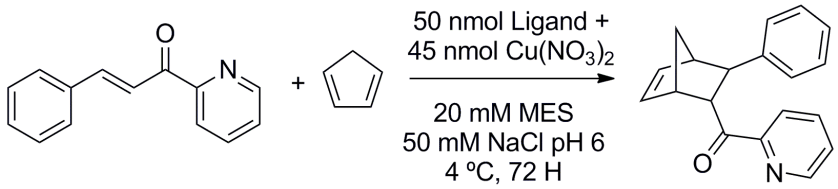
The chapters 2-4 demonstrate a unique approach for the design and creation of a remarkable shape selective catalyst. If it can be confirmed that the found catalytic results truly are the result of substrate encapsulation by the protein scaffold, this system represents the first rationally designed artificial metalloenzyme which exploits

the shape selectivity of the protein scaffold to direct the outcome of a catalytic reaction. This catalytic system has numerous possibilities for optimisation. Several more phosphine aldehydes are available as well as different maleimide-hydrazines to create different linker structures between the protein and the phosphine. Also, the protein scaffold can be further optimised by moving the cysteine to different locations. Another possibility is the application of computational modelling in combination with catalytic results to identify and replace key amino acids. In addition, the protein scaffold itself can be completely substituted for another aliphatic substrate binding protein, which has been made possible by the development of the generally applicable phosphine modification strategy presented in this work.
This catalyst system demonstrates numerous new possibilities in artificial metalloenzymes. New catalytic systems can be designed based on a specific reaction and substrate by selection of the protein host that has binding affinity for the substrates of interest. Combined with the general applicable conjugation method for the introduction of phosphine ligands this opens many new possibilities for the creation of artificial metalloenzymes.

Chapter 5 describes an alternative approach in which SCP-2L is used as chiral template for application in asymmetric copper catalysed reactions. Two maleimide containing nitrogen donor ligands **N3** and **N4** were site selectively introduced into SCP-2L V83C and SCP-2L A100C (Scheme 4). Fluorescent spectroscopy was used to confirm copper complexation to the **N3** ligand after modification of the protein template. The nitrogen ligand modified proteins were applied in the copper catalysed asymmetric Diels-Alder reaction between azachalcone and cyclopentadiene (Scheme 5). A clear difference was found in the performance between ligands **N3** and **N4**, with the latter providing higher endo/exo selectivity and activity. This trend was observed for the **N3** and **N4** modified proteins indicating a clear effect of the nitrogen donor ligands on the catalytic performance of the conjugates. Also differences in activity and selectivity were observed between the different protein templates. The highlight was an enantioselectivity of 25% ee for the Diels-Alder reaction catalysed by SCP-2L V83C-**N3**.



Scheme 4. Modification of SCP-2L unique cysteine templates with nitrogen ligands



Scheme 5. Copper catalysed Diels-Alder reaction between azachalcone and cyclopentadiene

The alternative approach of the use of protein templates as chiral ligands in catalysis presented in chapter 5 has already been used by other research groups. However, no covalent modified proteins have so far been applied in copper catalysed Diels-Alder reactions. The system still has many opportunities for further optimisation using different nitrogen donor ligands and by optimisation of the protein structure. The covalent introduction of the nitrogen donor ligands also allows for the choice of different protein templates that might be more suitable for the induction of selectivity in copper catalysed reactions. An example of this would be a protein that can ensure proper orientation of the substrate when bound to the catalytic centre. In this way it can activate a reaction pathway that leads to the selective formation of one particular reaction product. The proper protein structure can be obtained by searching for a

protein with a specific active site, shape and dimensions or can designed using computational methods starting from a suitable protein host.

Overall, this thesis demonstrates the development of new artificial metalloenzymes based on covalent modification strategies. The freedom of the choice of the protein template has been demonstrated to be useful to rationally select a protein that has the desired binding properties for a type of substrates of choice. The created shape selective hydroformylation catalyst is the first of its kind and might be a milestone for the future development of many more interesting artificial metalloenzymes. New artificial metalloenzyme catalyst can be designed by utilising the binding properties of the protein host for a desired substrate. This provides new opportunities for chemo- and stereo- selectivity and substrate specificity for reactions not performed by natural enzymes unequalled by traditional synthetic homogeneous catalysts.

Appendixes

A1 Peptide sequences

A1.1 SCP-2L sequences

SCP-2L

MEGGKLQSTFVFEEIGRRLKDIGPEVVKKVNAVFEWHITKGGNIGAKWTIDL
KSGSGKVYQGPAKGAADTTIILSDEDFMEVVLGKLDPQKAFFSGRLKARGNI
MLSQKLQMILKDYAKL

-Amino acids marked in yellow are not visible in the X-ray structure (pdb = 1IKT)[1]

SCP-2L V82C

MEGGKLQSTFVFEEIGRRLKDIGPEVVKKVNAVFEWHITKGGNIGAKWTIDL
KSGSGKVYQGPAKGAADTTIILSDEDFMECVLGKLDPQKAFFSGRLKARGNI
MLSQKLQMILKDYAKL

- The introduced cysteine is highlighted in purple

SCP-2L V83C

MEGGKLQSTFVFEEIGRRLKDIGPEVVKKVNAVFEWHITKGGNIGAKWTIDL
KSGSGKVYQGPAKGAADTTIILSDEDFMEVCLGKLDPQKAFFSGRLKARGNI
MLSQKLQMILKDYAKL

- The introduced cysteine is highlighted in purple

Hexahistidine-tagged SCP-2L

MSYYHHHHHHDYDIPTTENLYFQ/GAMEGGKLQSTFVFEEIGRRLKDIGPEVV
KKVNAVFEWHITKGGNIGAKWTIDLKSGSGKVYQGPAKGAADTTIILSDEDF
MEVVLGKLDPQKAFFSGRLKARGNIMLSQKLQMILKDYAKL

- Red amino acids show the target sequences of TEV with the "/" indicating the cleavage site[2]

SCP-2L after TEV cleavage

GAMEGGKLQSTFVFEEIGRRLKDIGPEVVKKVNAVFEWHITKGGNIGAKWTI
DLKSGSGKVYQGPAKGAADTTIILSDEDFMEVVLGKLDPQKAFFSGRLKARG
NIMLSQKLQMILKDYAKL

- Amino acids marked in red are added to the original sequence of SCP-2L

Hexahistidine-tagged SCP-2L E14C

MSYYHHHHHHDYDIPTTENLYFQ/GAMEGGKLQSTFVFECIGRRLKDIGPEV
VKKVNAVFEWHITKGGNIGAKWTIDLKSGSGKVYQGPAKGAADTTIILSDED
FMEVVLGKLDPQKAFFSGRLKARGNIMLSQKLQMILKDYAKL

- Red amino acids show the target sequences of TEV with the "/" indicating the cleavage site[2]
- The introduced cysteine is highlighted in purple

SCP-2L E14C after TEV cleavage

GAMEGGKLQSTFVFECIGRRLKDIGPEVVKKVNAVFEWHITKGGNIGAKWTI
DLKSGSGKVYQGPAKGAADTTIILSDEDFMEVVLGKLDPQKAFFSGRLKARG
NIMLSQKLQMILKDYAKL

- Amino acids marked in red are added to the original sequence of SCP-2L
- The introduced cysteine is highlighted in purple

Hexahistidine-tagged SCP-2L W36C

MSYYHHHHHHDYDIPTTENLYFQ/GAMEGGKLQSTFVFEEIGRRLKDIGPEVV
KKVNAVFECHITKGGNIGAKWTIDLKSGSGKVYQGPAKGAADTTIILSDEDF
MEVVLGKLDPQKAFFSGRLKARGNIMLSQKLQMILKDYAKL

- Red amino acids show the target sequence of TEV with the "/" indicating the cleavage site[2]
- The introduced cysteine is highlighted in purple

Hexahistidine-tagged SCP-2L V83C

MSYYHHHHHHDYDIPTTENLYFQ/GAMEGGKLQSTFVFEEIGRRLKDIGPEVV
KKVNAVFEWHITKGGNIGAKWTIDLKSGSGKVYQGPAKGAADTTIILSDEDF
MEVVLGKLDPQKAFFSGRLKCRGNIMLSQKLQMILKDYAKL

- Red amino acids show the target sequence of TEV with the "/" indicating the cleavage site[2]
- The introduced cysteine is highlighted in purple

SCP-2L V83C after TEV cleavage

GAMEGGKLQSTFVFEEIGRRLKDIGPEVVKKVNAVFEWHITKGGNIGAKWTI
DLKSGSGKVYQGPAKGAADTTIILSDEDFMEVVLGKLDPQKAFFSGRLKCRG
NIMLSQKLQMILKDYAKL

- Amino acids marked in red are added to the original sequence of SCP-2L
- The introduced cysteine is highlighted in purple

Hexahistidine-tagged SCP-2L A100C

MSYYHHHHHHDYDIPTTENLYFQ/GAMEGGKLQSTFVFEEIGRRLKDIGPEVV
KKVNAVFEWHITKGGNIGAKWTIDLKSGSGKVYQGPAKGAADTTIILSDEDF
MEVVLGKLDPQKAFFSGRLKCRGNIMLSQKLQMILKDYAKL

- Red amino acids show the target sequences of TEV with the "/" indicating the cleavage site[2]
- The introduced cysteine is highlighted in purple

SCP-2L A100C after TEV cleavage

GAMEGGKLQSTFVFEEIGRRLKDIGPEVVKKVNAVFEWHITKGGNIGAKWTI
DLKSGSGKVYQGPAKGAADTTIILSDEDFMEVVLGKLDPQKAFFSGRLKCRG
NIMLSQKLQMILKDYAKL

- Amino acids marked in red are added to the original sequence of SCP-2L
- The introduced cysteine is highlighted in purple

A1.2 mSCP2 sequences

GST- and hexahistide-tagged mSCP

MKHHHHHHNTSSNSMSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEG
DKWRNKKFELGLEFPNLPYYIDGDVKLTQSMAIIRYIADKHNMLGGCPKERA
EISMLEGAVLDIRYGVSRIAYSKDFETLKVDFLSKLPEMLKMFEDRLCHKTYL
NGDHVTHPDFMLYDALDVVLYMDPMCLDAFPKLVCFKKRIEAIPQIDKYLKS
SKYIAWPLQGWQATFGGGDHPPTSGSGGGGGSMSENLYFQ/GAMGSASDGF
KANLVFKEIEKKLEEEGEQFVKKIGGIFAFKVKDGPGGKEATWVVDVKNGK
GSVLPNSDKKADCTITMADSDFLALMTGKMNPQSAFFQGKLKITGNMGLAM
KLQNLQLQPGNAKL

- Red amino acids show the target sequences of TEV with the "/" indicating the cleavage site[2]

- Blue amino acids (including green) shows original mSCP2 sequence[3]
- Green indicating the native cysteine

mSCP2 after TEV cleavage

GAMGSASDGFKANLVFKEIEKKLEEEGEQFVKKIGGIFAFKVKDGPGGKEAT
WVVDVKNGKGSVLPNSDKKADCTITMADSDFLALMTGKMNPQSAFFQGKL
KITGNMGLAMKLQNLQLQPGNAKL

- Amino acids marked in red are added to the original sequence of mSCP2
- Blue amino acids (including green) shows original mSCP2 sequence[3]
- The native cysteine is highlighted in green

GST- and hexahistidine-tagged mSCP C91S M104C

MKHHHHHHNTSSNSMSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEG
DKWRNKKFELGLEFPNLPYYIDGDVKLTQSMAIIRYIADKHNMLGGCPKERA
EISMLEGAVLDIRYGVSRIAYSKDFETLKVDFLSKLPEMLKMFEDRLCHKTYL
NGDHVTHPDFMLYDALDVVLYMDPMCLDAFPKLVCFKKRIEAIPQIDKYLKS
SKYIAWPLQGWQATFGGGDHPPTSGSGGGGGSMSENLYFQ/GAMGSASDGF
KANLVFKEIEKKLEEEGEQFVKKIGGIFAFKVKDGPGGKEATWVVDVKNGK
GSVLPNSDKKADSTITMADSDFLALCTGKMNPQSAFFQGKLKITGNMGLAM
KQNLQLQPGNAKL

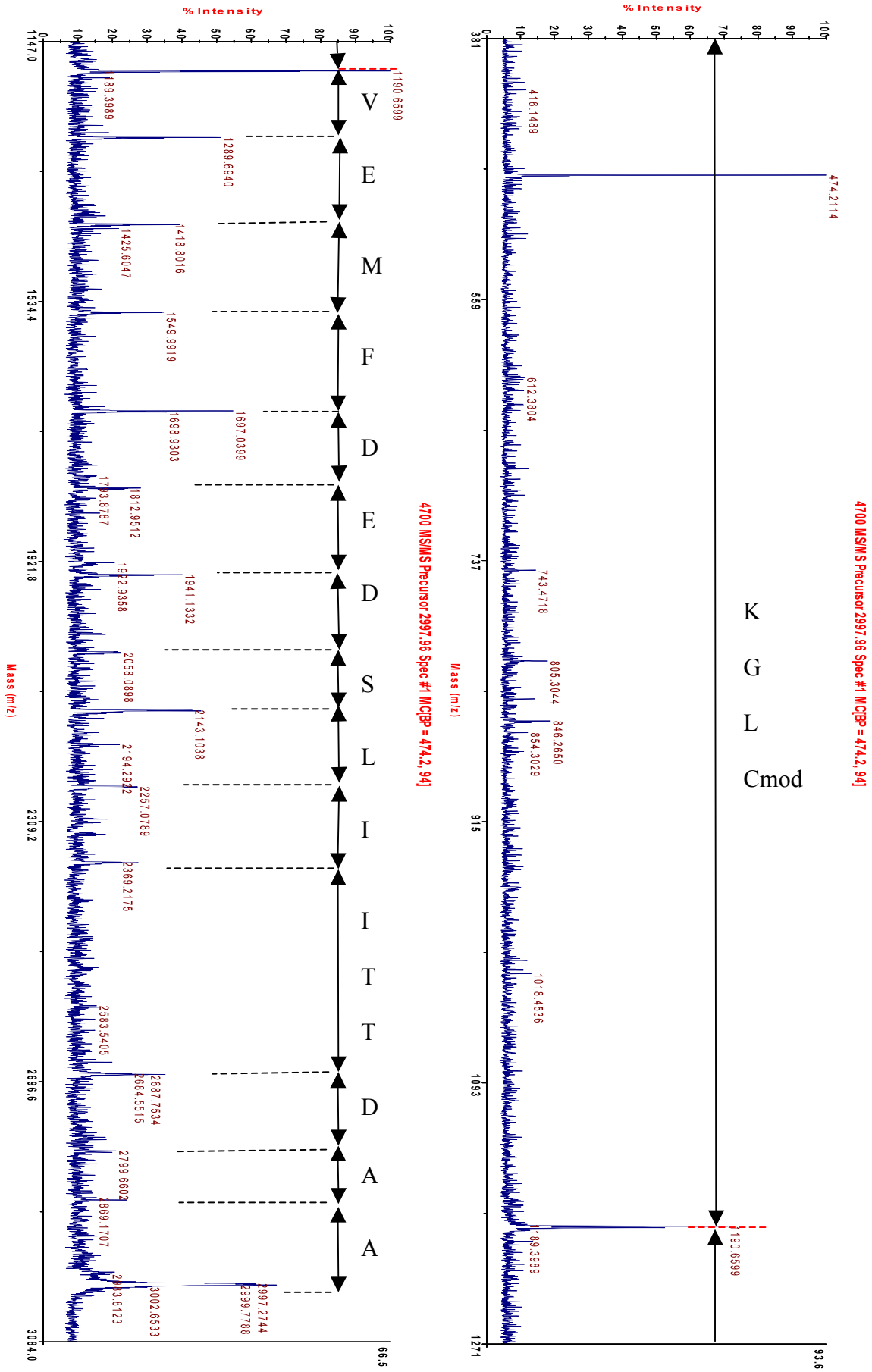
- Red amino acids show the target sequences of TEV with the "/" indicating the cleavage site[2]
- The introduced cysteine is highlighted in purple
- The location of the replacement of the native cysteine for serine is highlighted in orange

mSCP2 C91S M104C after TEV cleavage

GAMGSASDGFKANLVFKEIEKKLEEEGEQFVKKIGGIFAFKVKDGPGGKEAT
WVVDVKNGKGSVLPNSDKKADSTITMADSDFLALCTGKMNPQSAFFQGKLK
ITGNMGLAMKLQNLQLQPGNAKL

- Amino acids marked in red are added to the original sequence of mSCP2
- The introduced cysteine is highlighted in purple
- The location of the replacement of the native cysteine for serine is highlighted in orange

A2 MS-MS analysis of SCP-2L V83C-M5-P5 digested with trypsin



A3 Structural characterization of SCP-2L V83C-P4

Our group has reported several successful ^{31}P-NMR studies of phosphine modified proteins.[4] By application of this method it is possible to confirm whether the introduced phosphine is indeed obtained as a free phosphine. An attempt was made to perform the modification reaction of SCP-2L V83C with CDI activated **C4** at a scale that would allow conformation of the presence of the free phosphine by ^{31}P-NMR spectroscopy.

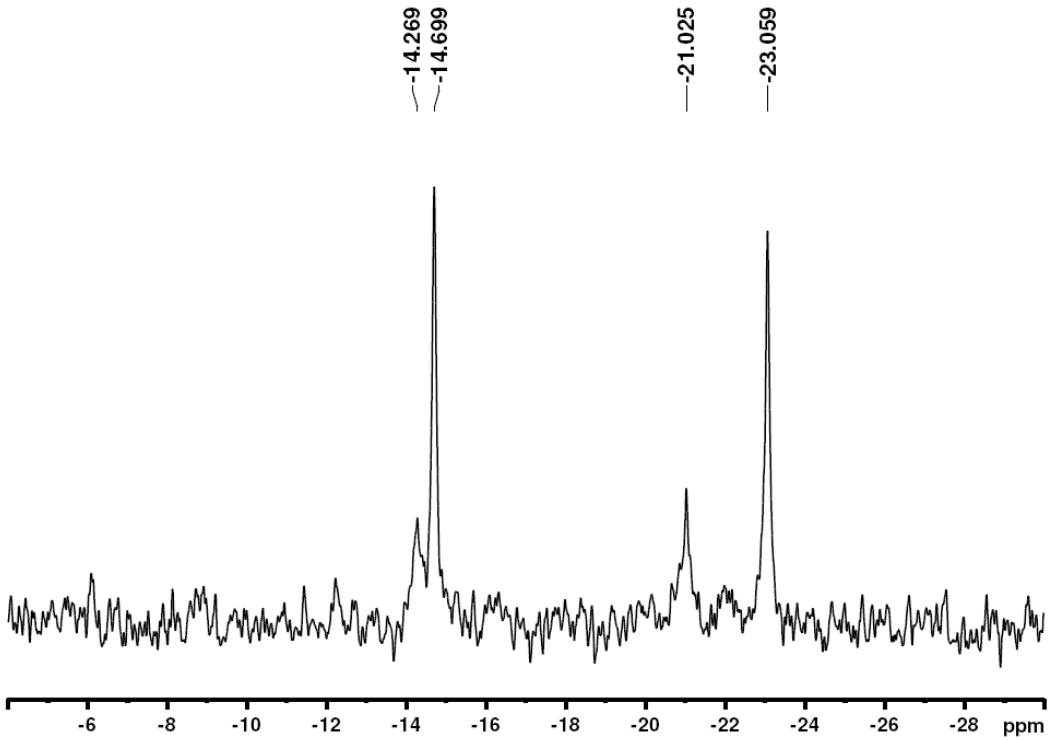


Figure 1. ^{31}P{^{1}H} NMR spectrum of SCP-2L V83C-**C4**.

The ^{31}P{^{1}H} NMR spectrum of SCP-2L V83C-**C4** was recorded at 202 MHz using 30000 scans (Figure 1). The obtained spectrum showed 4 peaks which based on the peak intensities appear to be two sets of two relatively broad peaks (set 1: –14.3 ppm, –21.0 ppm; set 2: –14.7 ppm and –23.1 ppm). The presence of two sets of signals could be the result of modification of the protein in two different positions or the ligand being in two different conformations. The large difference between the two phosphorus peaks within each apparent set (7-8 ppm compared to 1 ppm in the ^{31}P{^{1}H} NMR spectrum of the free ligand) are likely to be the result of two entirely different environments of the two phosphines.

The CD spectrum (Figure 2) and watergate ^{1}H- NMR (Figure 3) spectrum of SCP-2L V83C modified with **C4** were recorded to investigate the structural integrity of the protein upon modification. Although the CD measurement could not be performed under oxygen free conditions, the results still yield relevant structural information as the structures of the phosphine and phosphine oxide are not expected to differ significantly. Watergate ^{1}H-NMR can be performed under oxygen free atmosphere as the sample preparation and the measurement are fully compatible with this.

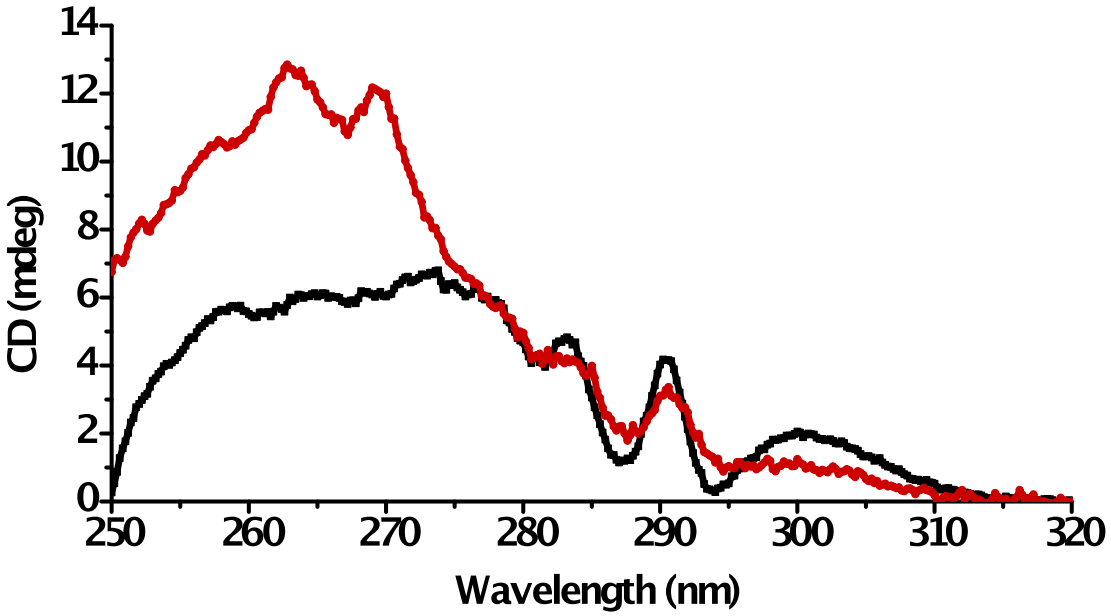


Figure 2. Near UV-CD spectra of SCP-2L V83C (black) and SCP-2L V83C-**C4** (red).

The CD spectrum shows significant changes to the signals at 287, 290, 294 and 300 nm and an increase in the signal at 250 - 275 nm. The latter increase can be assigned to the introduced phenyl substituents attached to the phosphines. This result indicates that the phosphines are affected by the chiral environment of the protein which is promising for application of the conjugate in asymmetric catalytic reactions Another effect that can play a role in the increased signal at around 250 nm can be due to partial unfolding of α-helices, as intact helices give a strong negative signal in the range of 190-240 nm. This is to be expected as **C4** is likely partly introduced at cysteine at position 83 which is part of the front α helical structure (see section 2.4.3 of chapter 2). Thus, it seems the structure of the protein is affected by the introduction of the ligand. However, the global structure seems retained as the general shape of the spectrum between 275 and 320 is retained.

Although the phase correction was troublesome the watergate ^{1}H-NMR spectrum provides some interesting information. Significant changes can be observed when compared to the spectrum obtained from the unmodified protein discussed in chapter 2 section 2.4.3. In the region above 5 ppm (Figure 3a) all peaks have shifted and are broader indicating a more flexible fold. Also in the region around 0 ppm (Figure 3b), where signals corresponding to aliphatic residues inside the hydrophobic core of the protein are located, significant changes are observed. This can indicate significant changes to the cavity structure caused by the incorporation of the hydrophobic ligand. The region between 0.5 and 2.5 ppm (Figure 3b) also shows some perturbations but also clear similarity indicating that part of the structure is retained.

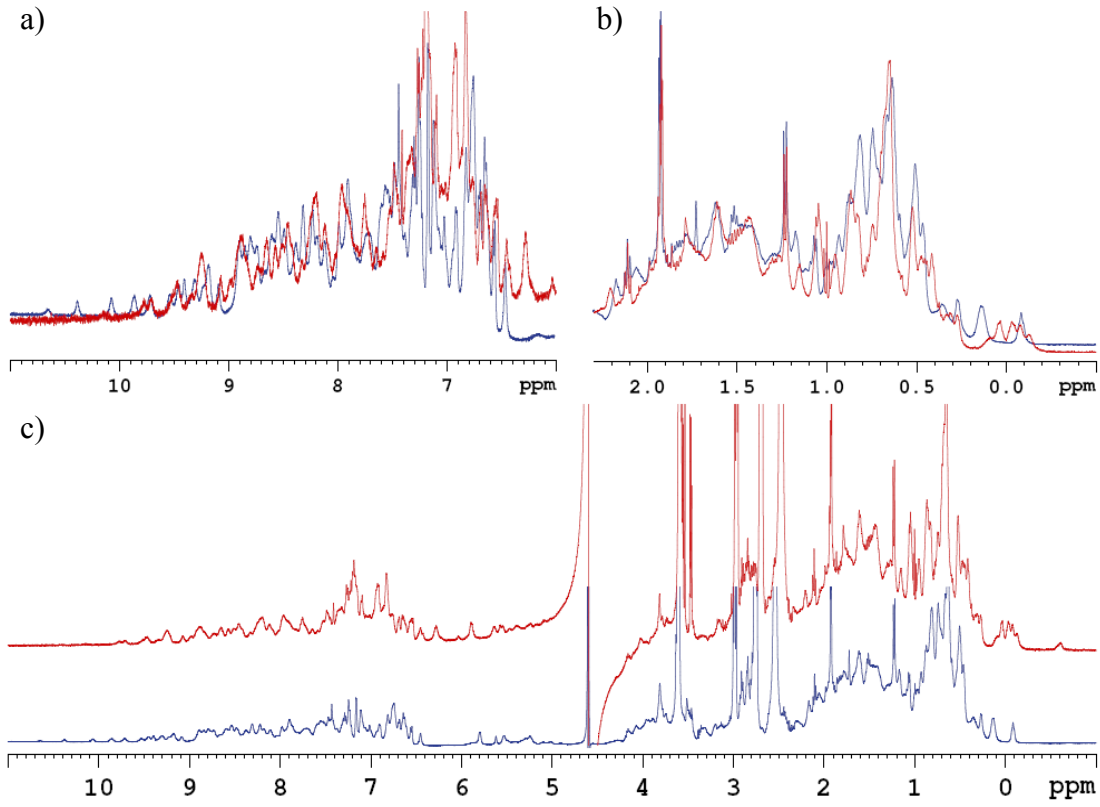


Figure 3. Watergate ^{1}H-NMR spectra of SCP-2L V83C modified with **C4** (red) and SCP-2L V83C (blue) 2 mM in 20 mM MES pH = 6 buffer a) overlaid in the region between 11 and 6 ppm b) overlaid in the region between 2.3 and -0.5 ppm c) full spectra offset for clarity.

A4 Ligand-protein models created using GOLD

Several models were created of SCP-2L with phosphine modifications to study the orientation of phosphine ligands with respect to the hydrophobic cavity. These models

were created with GOLD[5] using models of the unique cysteine templates created by the SWISS-MODEL online service based on the reported SCP-2L crystal structure (1IKT)[6] and structures of the ligands optimised using semi-empirical structure optimisations of Spartan Wavefunction (Figure 4).

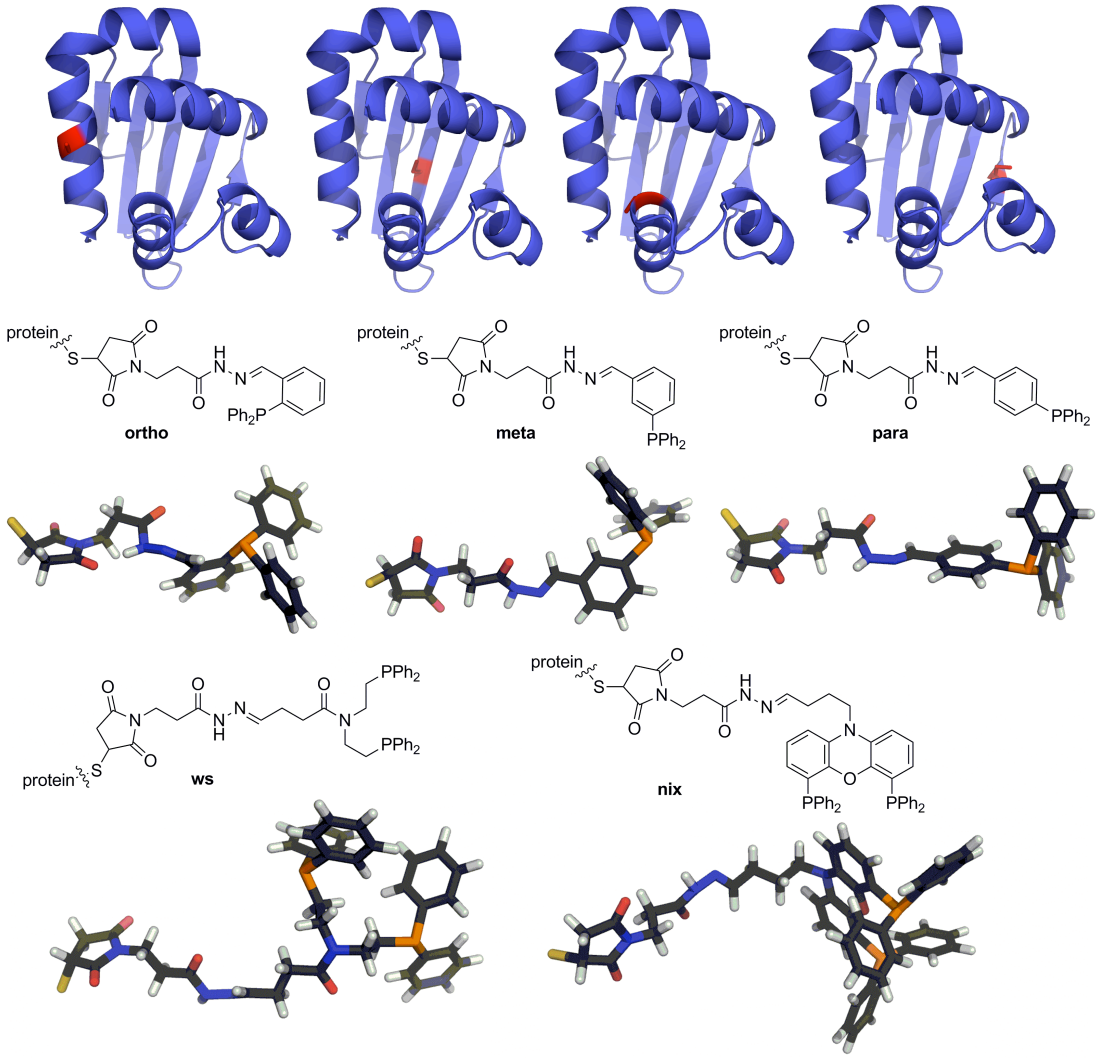
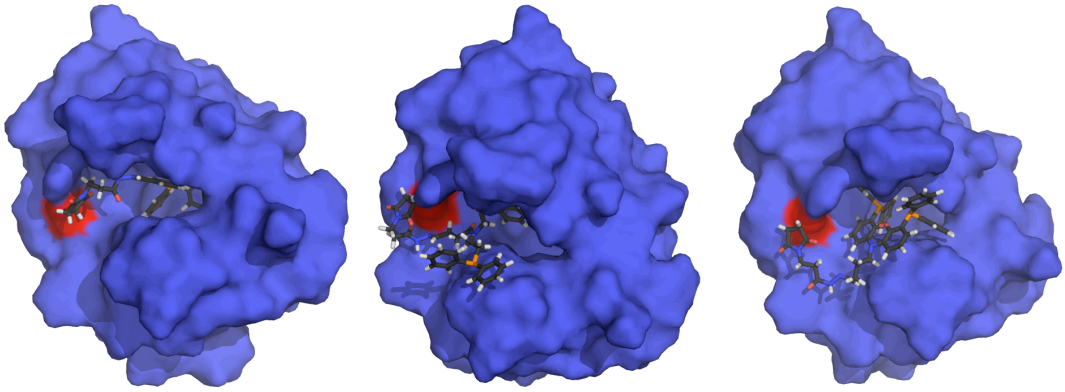


Figure 4. Models of SCP-2L mutant templates of E14C, W36C, V83C and A100C with the introduced cysteine shown in red created by SWISS-MODEL online and phosphine-hydrazone ligands with Chemdraw and Spartan Wavelength optimized structures.

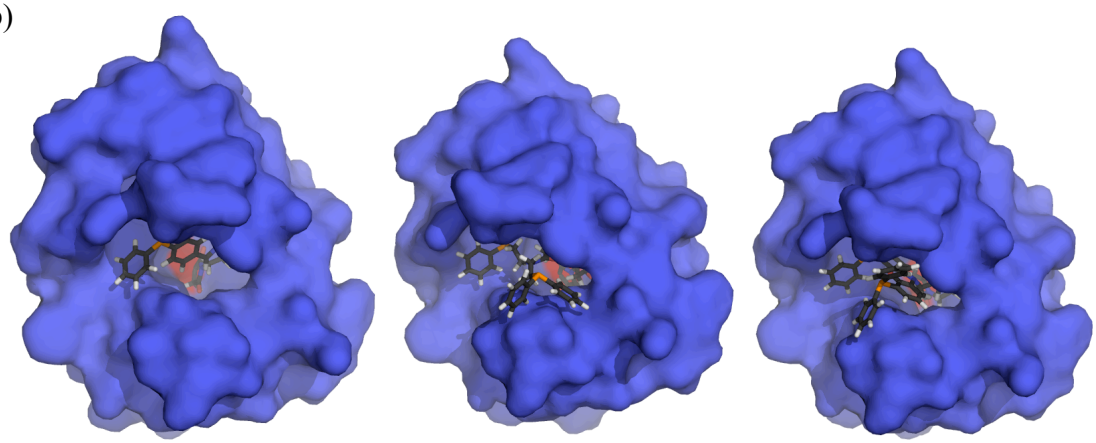
The ligand was covalently docked in the protein structure by overlapping the sulfur atoms in both the ligand and the protein structure. Both the S or R conformation of the chiral centre created by sulfide-maleimide bond were created and used in the optimisation process. Selections of the top ranked structures obtained from 20 optimisations are shown in Figure 5. Little difference was found between the

structures obtained for the maleimide-hydrazone-diphenylphosphine benzaldehyde ligands **P1-P3** (**ortho**, **meta** and **para)** for each separate protein template (that is why only one of each is shown for each template). All the obtained structures for these three ligands show that the triphenylphosphine motif has the tendency to be located in the wide entrance of the tunnel. This causes the ligand to cap the tunnel for the structures obtained for SCP-2L E14C and SCP-2L V83C while the ligand takes up almost the entire space of the tunnel for SCP-2L W36C and SCP-2L A100C. The structures obtained for the Whitesides based ligand **P4** (**ws**) show that tunnel is only large enough to accommodate one of the triphenylphosphine moieties while the other triphenylphosphine moiety is located outside of the tunnel. The structures obtained for the nixantphos based ligand **P6** (**nix)** demonstrate that this ligand is too large to be accommodated by the protein tunnel and therefore is only party located herein. Additional models can be found on the CD accompanying this thesis (The program Pymol[7] is required to view these structures).

a)



b)



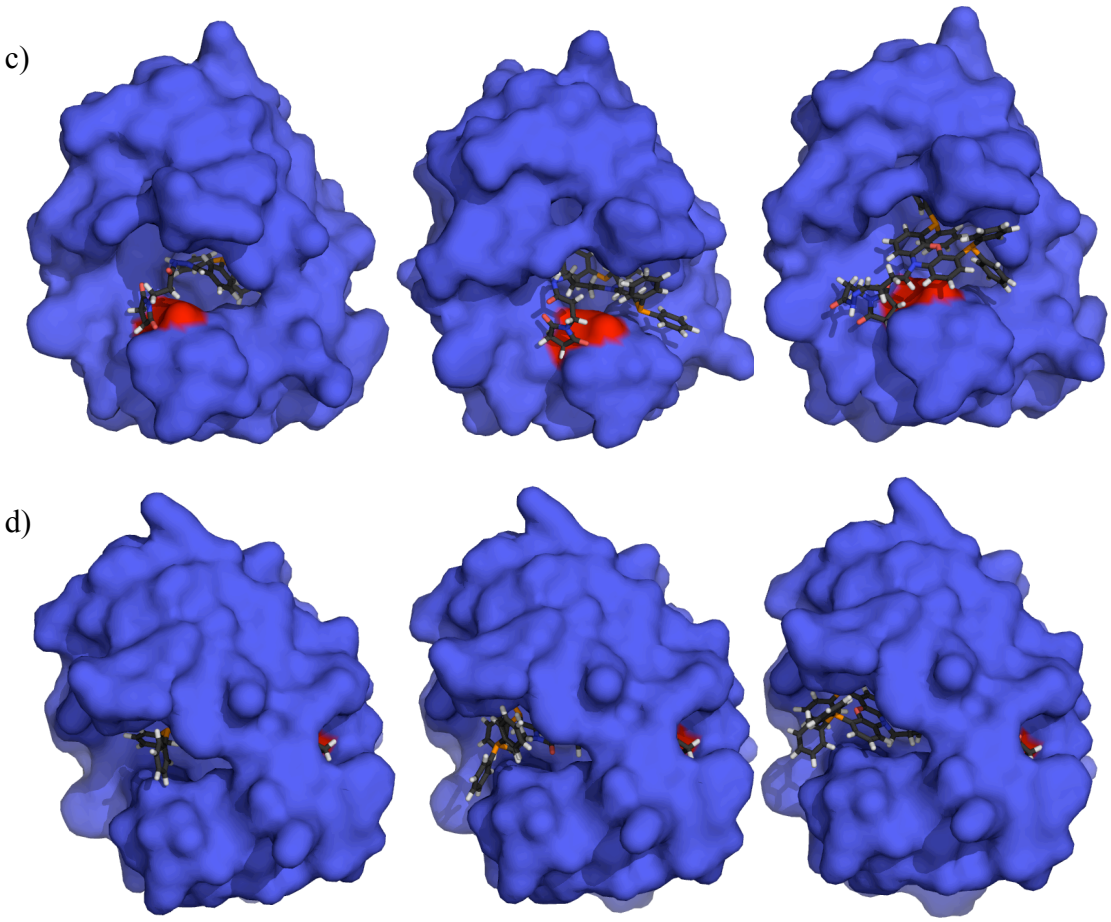


Figure 5. Selection of the top ranked results obtained by GOLD of a) SCP-2L E14C with from left to right: R-**ortho**, S-**ws** and S-**nix** b) SCP-2L W36C (with R-**para**, S-**ws** and R-**nix**) c) SCP-2L V83C (with R-**meta**, R-**ws** and R-**nix**) and d) SCP-2L A100C (with S-**para**, S-**ws** and S-**nix**) with the cysteine shown in red and R and S before the name of the ligand indicating the geometry of the chiral centre of the maleimide-sulfide bond.

A5 Crystal structures I1 and I8

A5.1 Crystallographic data obtained for I1

Below the crystallographic data is presented for **I1** (Figure 6 and Table 1). An overview of all bond angles and distances can be found on the accompanying CD.

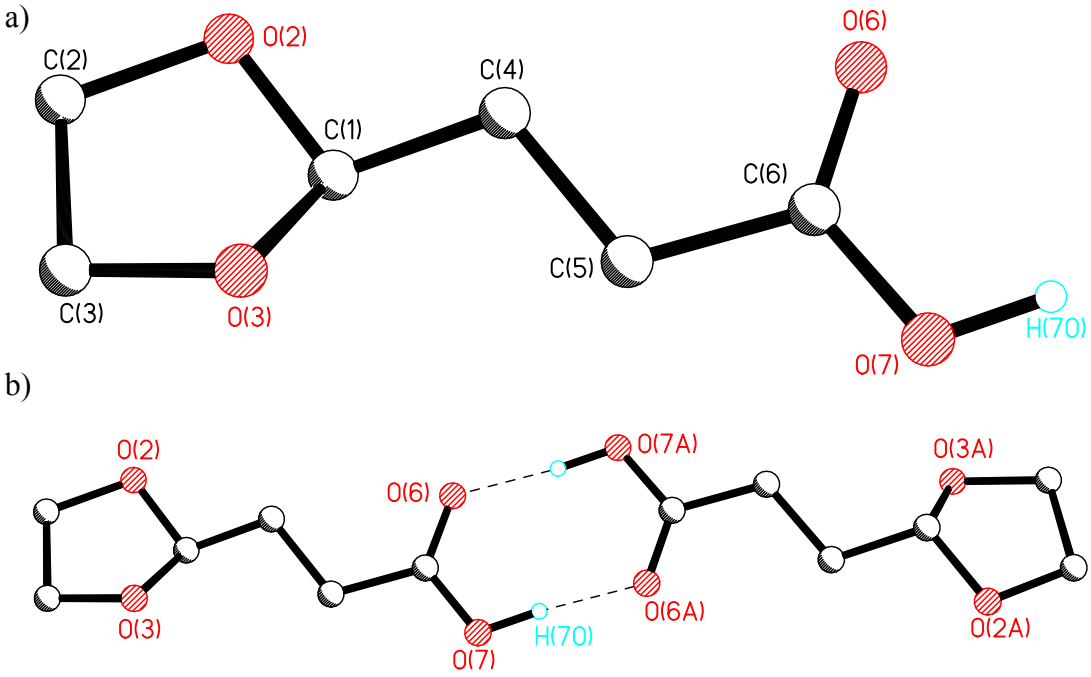


Figure 6. a) Crystal structure of **I1**. b) Crystal structure of **I1** showing the hydrogen bonding between H70 and O6.

Table 1. Crystal data and structure refinement for **I1** (St. Andrews reference = pdpk6).

Identification code	pdpk6	
Empirical formula	C6 H10 O4	
Formula weight	146.14	
Temperature	173(2) K	
Wavelength	1.54178 Å	
Crystal system	Monoclinic	
Space group	P2(1)/c	
Unit cell dimensions	a = 5.2617(7) Å	α = 90°.
	b = 7.4044(9) Å	β = 95.078(4)°.
	c = 17.841(2) Å	γ = 90°.
Volume	692.34(15) $Å^3$	
Z	4	
Density (calculated)	1.402 Mg/m^3	
Absorption coefficient	1.018 mm^{-1}	
F(000)	312	

Crystal size	0.150 x 0.150 x 0.030 mm^3
Theta range for data collection	4.98 to 67.46°.
Index ranges	-5<=h<=5, -8<=k<=8, -21<=l<=21
Reflections collected	5576
Independent reflections	1074 [R(int) = 0.0523]
Completeness to theta = 67.00°	86.6 %
Absorption correction	Multiscan
Max. and min. transmission	1.0000 and 0.9261
Refinement method	Full-matrix least-squares on F^2
Data / restraints / parameters	1074 / 1 / 96
Goodness-of-fit on F^2	1.056
Final R indices [I>2sigma(I)]	R1 = 0.0488, wR2 = 0.1359
R indices (all data)	R1 = 0.0493, wR2 = 0.1367
Extinction coefficient	0.037(7)
Largest diff. peak and hole	0.282 and -0.264 $e.Å^{-3}$

A5.2 Crystallographic data obtained for I8

Below the crystallographic data is presented for **I8** (Figure 7 and Table 2). An overview of all bond angles and distances can be found on the accompanying CD.

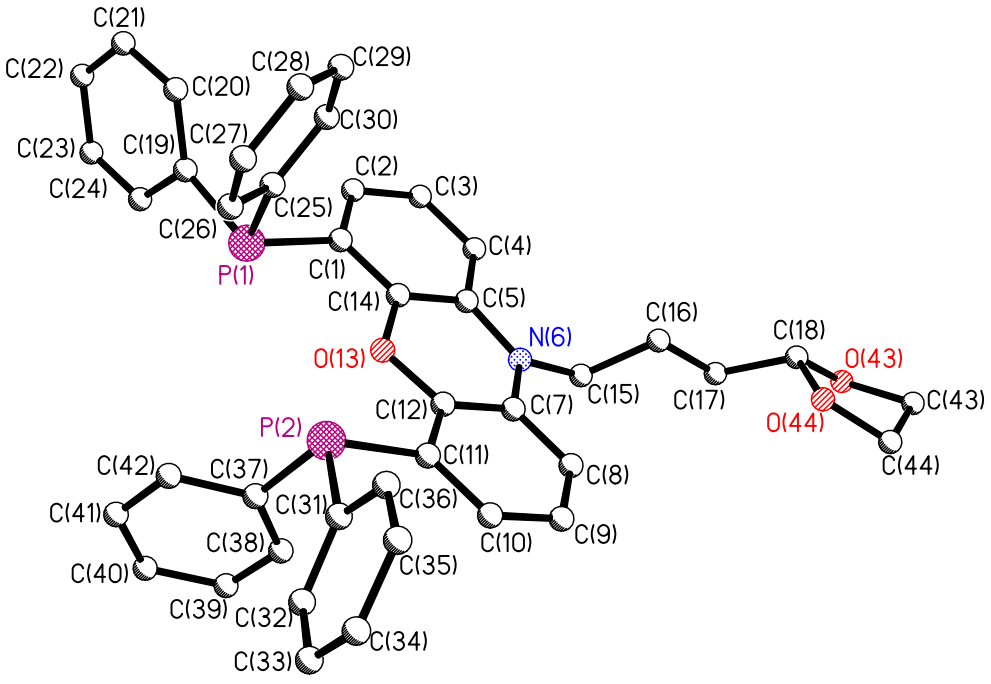
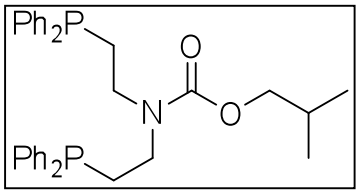


Figure 7. Crystal structure of **I8**.

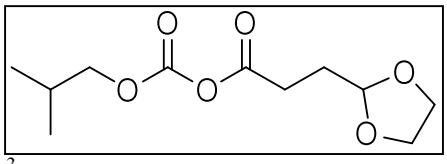
Table 2. Crystal data and structure refinement for **I8** (St. Andrews reference = pdpk7).

Identification code	pdpk7	
Empirical formula	C42 H37 N O3 P2	
Formula weight	665.67	
Temperature	93(2) K	
Wavelength	0.71073 Å	
Crystal system	Triclinic	
Space group	P-1	
Unit cell dimensions	a = 10.340(3) Å	α = 71.513(18)°.
	b = 11.455(2) Å	β = 89.22(3)°.
	c = 15.299(4) Å	γ = 78.64(2)°.
Volume	1682.6(7) $Å^3$	
Z	2	
Density (calculated)	1.314 Mg/m^3	
Absorption coefficient	0.172 mm^{-1}	
F(000)	700	
Crystal size	0.2000 x 0.2000 x 0.1000 mm^3	
Theta range for data collection	3.04 to 25.29°.	
Index ranges	-12<=h<=12, -9<=k<=13, -16<=l<=18	
Reflections collected	10625	
Independent reflections	5912 [R(int) = 0.0517]	
Completeness to theta = 25.00°	97.0 %	
Absorption correction	Multiscan	
Max. and min. transmission	1.0000 and 0.6424	
Refinement method	Full-matrix least-squares on F^2	
Data / restraints / parameters	5912 / 0 / 433	
Goodness-of-fit on F^2	0.990	
Final R indices [I>2sigma(I)]	R1 = 0.0676, wR2 = 0.1556	
R indices (all data)	R1 = 0.1144, wR2 = 0.1806	
Largest diff. peak and hole	0.960 and -0.307 $e.Å^{-3}$	

A6 ^{1}H-NMR spectroscopic data for the side products of the reaction between I1 and I2



Compound **I2** was a side-product in the reaction described in section 3.5.6.3. It was obtained as a 2:1 mixture with *iso*-butylcarbamate of **I2** from the alumina column described in that section. R_f=0.8 (Al_2O_3, 7:3 PE:EtOAc); ^{1}H-NMR (400 MHz, $CDCl_3$), δ 7.40-7.20 (m, 20H, Ar-C*H*), δ 3.73 (d, $^3J_{HH}$=6.7 Hz, 2H, -O-CH_2-CH(CH_3)$_2$), δ 3.35-3.12 (m, 4H, Ph_2P-CH_2-CH_2-), δ 2.30-2.08 (m, 4H, -CH_2-CH_2-N-), δ 1.77 (12tet, $^3J_{HH}$=6.7 Hz, 1H, -O-CH_2-C*H*(CH_3)$_2$) , δ 0.79 (d, $^3J_{HH}$=6.7 Hz, 6H, -O-CH_2-CH(CH_3)$_2$); ^{31}P-{^{1}H}-NMR (162 MHz, $CDCl_3$), δ -21.0 (s, 2P, -*P*Ph_2).



^{1}H-NMR (400 MHz, $CDCl_3$) δ 4.88 (t, $^3J_{HH}$=4.3 Hz, 1H, -CH_2-C*H*-(O-)$_2$), 3.93-3.75 (m, 6H, Ketal-CH_2's and -O-CH_2-CH(CH_3)$_2$), δ 2.38 (t, $^3J_{HH}$=7.7 Hz, 2H, OOC-CH_2-CH-), δ 1.95 (td, $^3J_{HH}$=7.7 Hz,$^3J_{HH}$=4.1 Hz, 2H, -CH_2-CH_2-CH-), δ 1.86 (12tet, $^3J_{HH}$=6.6 Hz, 1H, -O-CH_2-C*H*(CH_3)$_2$), δ 0.86 (d, $^3J_{HH}$=6.6 Hz, 6H, -O-CH_2-CH(CH_3)$_2$).

A7 Modification of M5 modified proteins with preformed transition metal complexes of phosphine aldehydes

A7.1 Conjugation of preformed rhodium complexes

The successful bioconjugation of several mono and diphosphines to SCP-2L V83C was demonstrated in the chapter 3. Seeking to extend the approach to the synthesis of metalloproteins, we explored the use of preformed transition metal complexes containing phosphine-aldehyde ligands for the second step in the coupling. This method is preferred over metal coordination following the conjugation of the ligand in an third conjugation step, as this will be more efficient because it will reduce the number of time consuming steps of washing and concentration of the protein. However, this requires the synthesis of monophosphine-rhodium complexes, which is not straightforward. Indeed, statistical mixtures of mono and bisphosphine rhodium complexes were obtained by using readily available [Rh(acac)$(CO)_2$], [Rh$(cod)_2$]BF_4, [Rh$(MeCN)_2$(cod)]BF_4 or [Rh$(nbd)_2$]BF_4 as metal precursors and phosphine-aldehyde **P3**. The separation of the different products from the mixtures was unsuccessful. The

synthesis of a monophosphine complex of **P3** was successful using palladium allyl chloride, however problems with the bioconjugation of the complex were encountered (see appendix A8). The synthesis of bisphosphine rhodium complexes is more straightforward, and although such complexes are unsuitable for artificial metalloenzyme-development, bisphosphine complexes might be interesting for future crosslinking applications.

The desired rhodium complexes of the diphosphine-aldehyde ligands **C4**, **P5** and **P6** were obtained without the problems encountered for the monophosphines. The complex obtained after the reaction of CDI activated **C4** with [Rh(acac)$(CO)_2$] provided a complicated second order ^{31}P- NMR spectrum. (Figure 8a). Simulation of the spectrum allowed the identification of two sets of signals of almost equal intensity (1:0.9, see A7.2 for more details). Following the analysis of the coupling constants, the signals were assigned to two different complexes with the phosphines in either *trans* or *cis* positions, with the former the dominant species. The modification of SCP-2L V83C with these complexes was unsuccessful, as analysis of the reaction by LCMS(ES$^+$) revealed only the presence of unmodified protein and a small fraction of protein modified with the oxidized ligand.

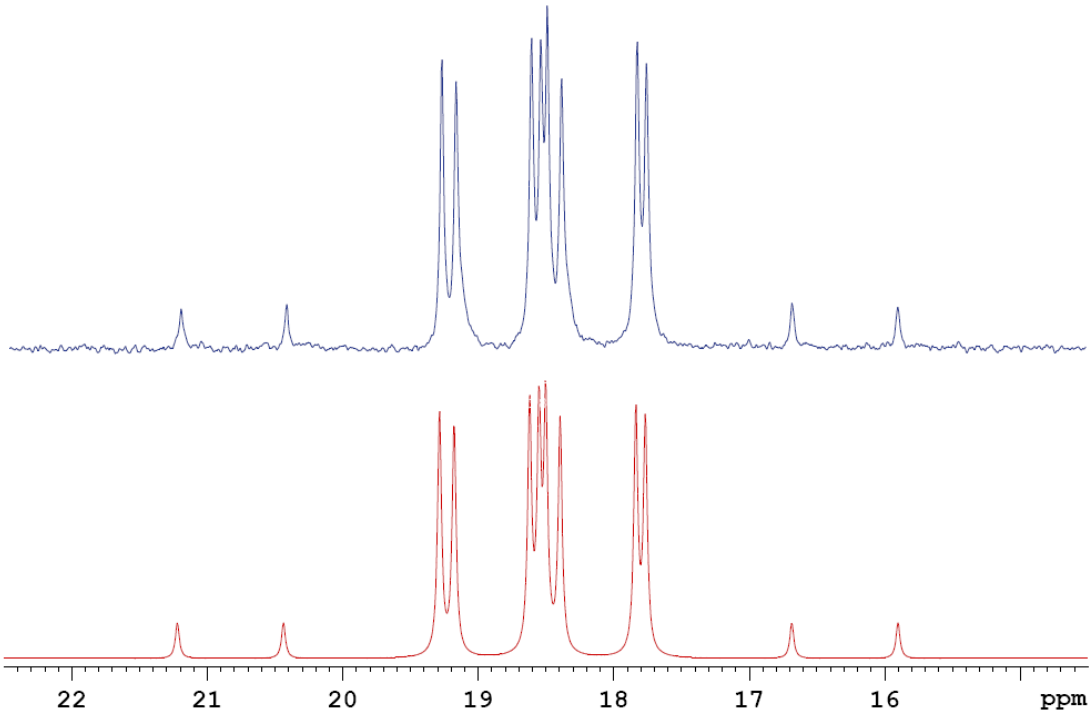


Figure 8. Experimental (blue) and simulated (red) spectra of formed complexes of [Rh(acac)$(CO)_2$] and phosphines a) CDI activated **C4** in DMF.

A complex of phosphine aldehyde **P5** with [Rh(acac)(CO)$_2$] generated in CDCl$_3$ gave a second order ^{31}P NMR spectrum similar to the complex obtained for **C4**. Also in this case analysis of the spectrum using simulation software revealed the presence of the *trans* and *cis* complexes with a ratio of 1:0.9 (see A7.2 for more details). The reaction of SCP-2L V83C-**M5** with this complex did not show any modification product upon analysis by LCMS(ES$^+$). Only the signal corresponding to unmodified SCP-2L V83C-**M5** was observed.

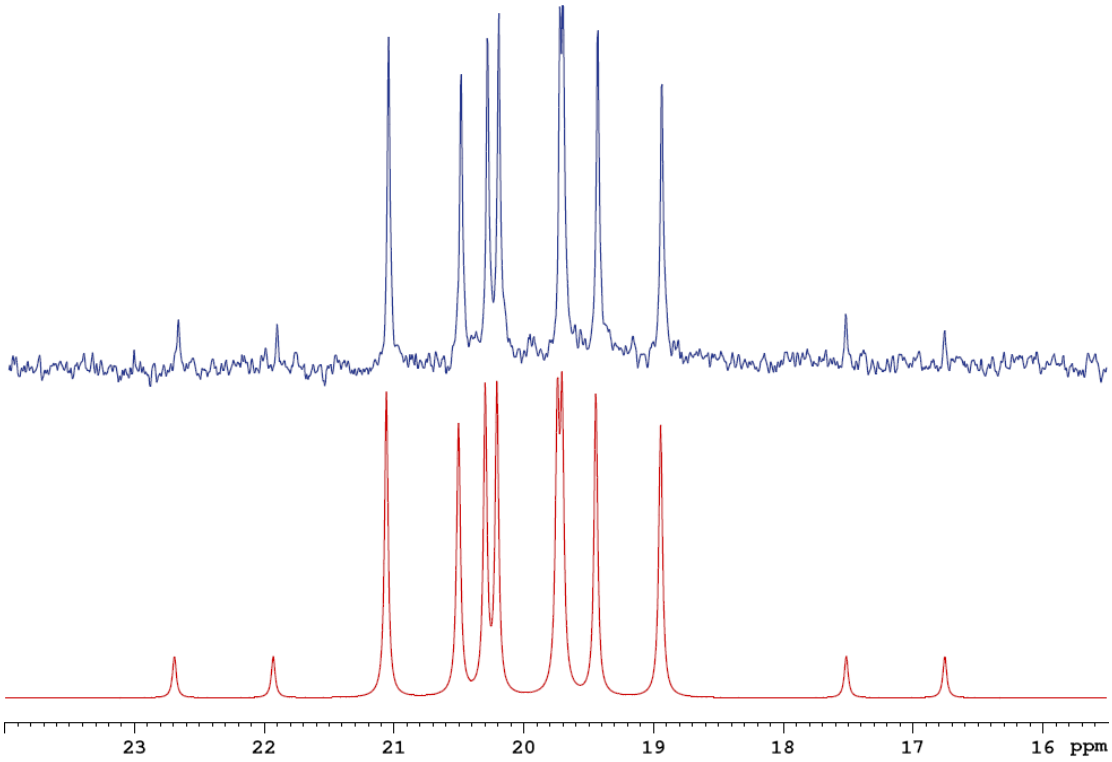


Figure 9. Experimental (blue) and simulated (red) spectra of formed complex of [Rh(acac)(CO)$_2$] and **P5** in CDCl$_3$.

A rhodium-**P6** complex was generated with [Rh(MeCN)$_2$(cod)]BF$_4$ in dichloromethane. Crystals suitable for X-ray analysis were obtained (see A7.3), showing a coordinating MeCN and cod molecule in addition to the diphosphine. The reaction of SCP-2L V83C-**M5** with this complex did not show any modification product upon analysis by LCMS(ES$^+$).

A7.2 Experimental data

*A7.2.1 Procedure for in situ complex formation with CDI activated **C4** and [Rh(acac)(CO)$_2$]*

A 25 mM solution of CDI activated **C4** was formed under argon atmosphere using the procedure described in chapter 3. 0.9 equivalents of [Rh(acac)(CO)$_2$] were added to this solution of CDI activated **C4** in DMF and the resulting mixture was mixed for about 10 minutes. The {^{1}H}-^{31}P-NMR (162 MHz) was recorded using an inner locking tube containing D_2O. In addition to the signal of unreacted **C4** a second order signal was observed between 15.8 and 21.3 ppm. Two species were identified using the DAISY second order simulation software of Topspin. The more abundant species, which was identified as a *trans* diphosphine rhodium complex because of the large phosphorus-phosphorus coupling constant ($^2J_{p,p}$=313.5 Hz), contained two phosphorus signals at 19.4 and 17.7 ppm with coupling constants of $^1J_{Rh,P}$=127.0 Hz and $^1J_{Rh,P}$=126.9 Hz. The second species, which was identified as a *cis* diphosphine rhodium complex because no phosphorus-phosphorus coupling was observed, contained two phosphorus signal at 18.9 and 18.1 ppm with coupling constants of $^1J_{Rh,P}$=101.6 Hz and $^1J_{Rh,P}$=101.9 Hz. This solution was used directly for protein modification experiments.

*A7.2.2 Procedure for complex formation with **P5** and [Rh(acac)(CO)$_2$]*

5.6 mg (21.8 µmol) [Rh(acac)(CO)$_2$] were added to a solution of 6.5 mg (21.8 µmol) **P5** in degassed $CDCl_3$ under argon atmosphere. The obtained solution was mixed for 10 minutes and the {^{1}H}-^{31}P-NMR (162 MHz) was recorded, which showed a second order signal between 16.7 and 22.8 ppm and some unidentified minor impurities at 10, 30 and 40 ppm. Two species were identified using the DAISY second order simulation software of Topspin. The more abundant, which was identified as a *trans* diphosphine rhodium complex because of the large phosphorus-phosphorus complex ($^2J_{p,p}$=354.6 Hz), contained two phosphorus signals at 20.7 and 18.7 ppm with coupling constants of $^1J_{Rh,P}$=123.5 Hz and $^1J_{Rh,P}$=122.9 Hz. The second species, which was identified as a *cis* diphosphine rhodium complex because no phosphorus-phosphorus coupling was observed, contained two phosphorus signal at 20.6 and 19.9 ppm with coupling constants of $^1J_{Rh,P}$=138.0 Hz and $^1J_{Rh,P}$=138.0 Hz. The solvent was

removed and the obtained solid diluted in degassed DMF to be used in coupling experiments.

*A7.2.3 Synthesis of [Rh**P6**(MeCN)(cod)]BF_4*

This procedure was performed using dry glassware and solvents under argon atmosphere. 16.67 mg (26.8 μmol) **P4** was dissolved in 1 ml dichloromethane. 10.19 mg (26.8 μmol) [Rh(MeCN)$_2$(cod)]BF_4 was dissolved in 2 ml dichloromethane and the two solution were mixed and stirred for one hour. The solvent was removed and the remaining brown residue washed with 5 x 1 ml ether. Crystals (~1 mg) were obtained by slow addition of 10 ml ether to a concentrated solution of the residue in dichloromethane.

General procedure for bioconjugation of preformed rhodium phosphine-aldehyde complexes with SCP-2L V83C-**M5**

A series of parallel reactions were performed using different equivalents of the complex (usually 1, 2, 3, 5, 10 and 20 equivalents). This was done by addition of the appropriate amount of a 25 mM solution of the complex in MeCN to 3.25 nmol SCP-2L V83C-**M5** in 250 μL of the appropriate buffer (usually 20 mM MES, 50 mM NaCl pH 6) in a 1.5 ml or 0.5 ml Eppendorf. The reactions were mixed overnight by slow continues inversion on a rotating wheel. The reaction mixtures were then centrifuged (16100 *g*, 10 minutes) to remove formed precipitate and excess complex. The solutions were directly used for analysis by LCMS(ES^+).

A7.3 Crystallographic data of [RhP6(MeCN)(cod)]BF_4

Below the crystallographic data is presented for [Rh**P6**(MeCN)(cod)]BF_4 (Figure 10 and Table 3). An overview of all bond angles and distances can be found on the accompanying CD.

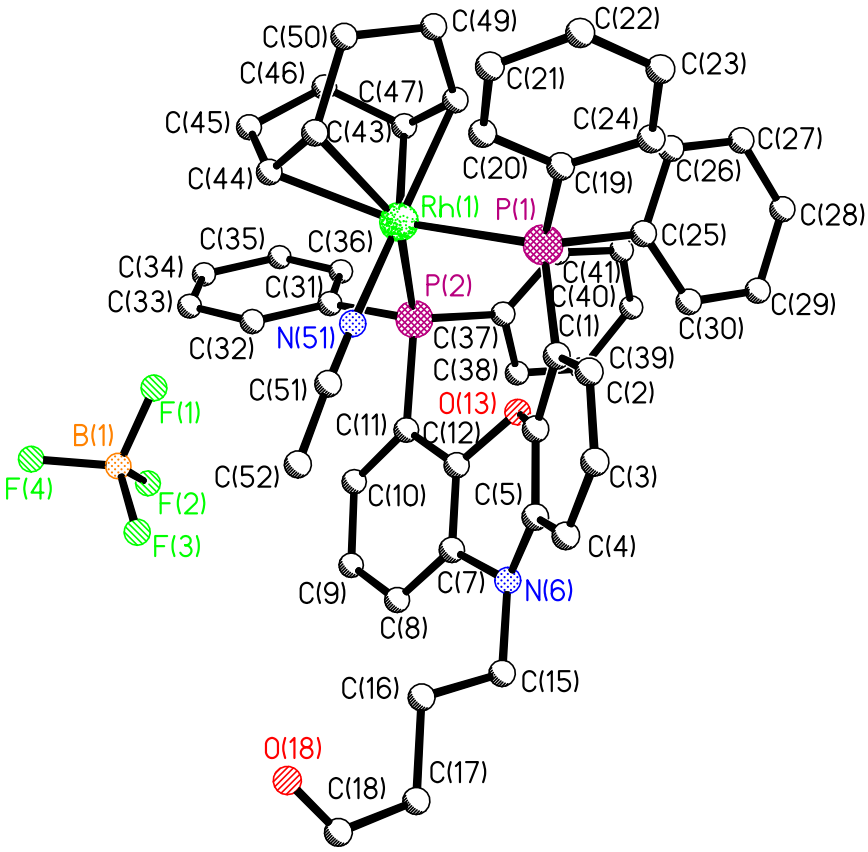


Figure 10. Crystal structure of [Rh**P6**(MeCN)(cod)]BF_4.

Table 3. Crystal data and structure refinement for [Rh**P6**(MeCN)(cod)]BF_4 (St. Andrews reference = pdpk10).

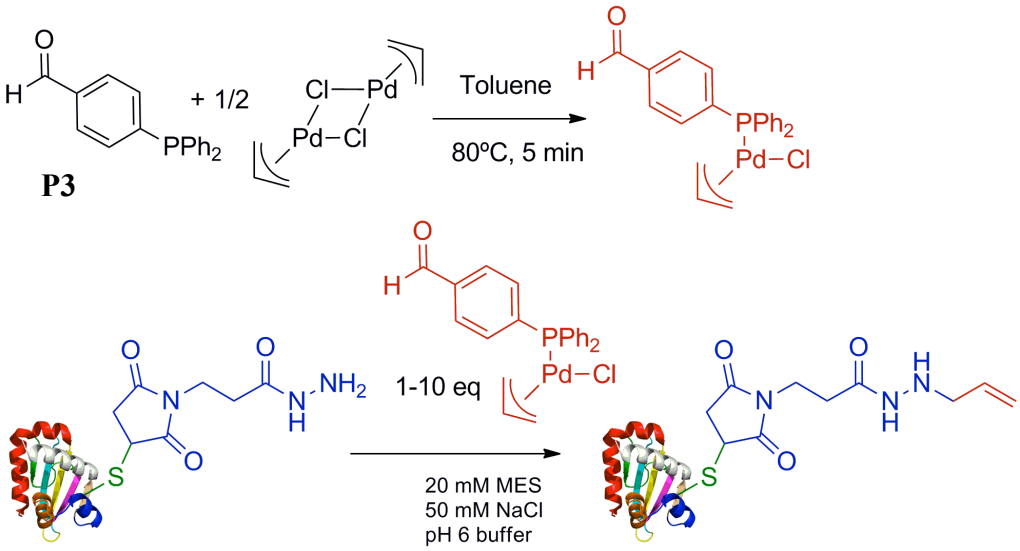
Identification code	pdpk10	
Empirical formula	C50 H48 B F4 N2 O2 P2 Rh	
Formula weight	960.56	
Temperature	93(2) K	
Wavelength	0.71073 Å	
Crystal system	Monoclinic	
Space group	P2(1)/n	
Unit cell dimensions	a = 10.620(5) Å	α = 90°.
	b = 32.224(14) Å	β = 108.377(8)°.
	c = 13.553(6) Å	γ = 90°.
Volume	4402(4) $Å^3$	
Z	4	

Density (calculated)	1.450 Mg/m^3
Absorption coefficient	0.521 mm^{-1}
F(000)	1976
Crystal size	0.20 x 0.10 x 0.03 mm^3
Theta range for data collection	2.03 to 25.43°.
Index ranges	-10<=h<=12, -38<=k<=38, -16<=l<=15
Reflections collected	44910
Independent reflections	8101 [R(int) = 0.0688]
Completeness to theta = 25.00°	99.9 %
Absorption correction	Multiscan
Max. and min. transmission	1.000 and 0.684
Refinement method	Full-matrix least-squares on F^2
Data / restraints / parameters	8101 / 0 / 560
Goodness-of-fit on F^2	1.106
Final R indices [I>2sigma(I)]	R1 = 0.0993, wR2 = 0.2756
R indices (all data)	R1 = 0.1089, wR2 = 0.2844
Largest diff. peak and hole	1.461 and -0.517 $e.Å^{-3}$

A8 Palladium-P3 complex formation and protein modification

A8.1 Results and discussion

The monophosphine complex shown in Scheme 1 was successfully obtained using $[Pd(allyl)Cl]_2$ and phosphine **P3** following a literature procedure involving a similar phosphine.[8] Several attempts were made for coupling of the complex to hydrazide modified SCP-2L V83C (SCP-2L V83C-**M5**). In each experiment the main product found after LCMS(ES^+) analysis was the protein modified with an allyl group. When a large excess of the complex was used an additional peak was observed corresponding to a double allyl modification. A small peak (20% intensity compared to the main peak) of the addition of the complex without the chlorine could be found in all experiments. The observed competition between modification with the complex and allyl addition makes this route ineffective for artificial metalloenzymes formation.



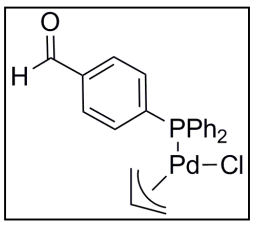
SCP-2L V83C-**M5**

SCP-2L V83C-**M5**-allyl (calc 13599.7 Da)
Found: LCMS (ES^+) 13600.0 Da

Scheme 1. Synthesis of palladium complex with phosphine aldehyde **P3** and conjugation to SCP-2L V83C-**M5**.

A8.2 Experimental data

*A8.2.1 Synthesis of the palladium allylchloride complex of **P3***



The following method is based on a literature procedure used for the synthesis of the triphenylphosphine analogue of this complex.[8] 52.14 mg (143 µmol) dimeric palladium allylchloride **P3** was dissolved in as little as possible warm dry toluene (~ 2ml, ~ 60ºC). To this yellow solution was slowly added an orange/brown solution containing 82.7 mg (285 µmol) *p*-diphenylphosphine benzaldehyde in 0.5 ml warm dry toluene. The solution was stirred for 5 more minutes at 80ºC and a white solid precipitated. The reaction mixture was cooled to 0ºC, the liquid was removed and the solid washed 3 times with 1 ml cold toluene. The solid was than dried *en vacuo* providing **P3** as a very light yellow powder. ^{1}H-NMR (400 MHz, $CDCl_3$), δ 9.98 (s, 1H, -C*H*O), δ 7.86-7.79 (m, 2H, Ph-C*H*), δ 7.74-7.65 (m, 2H, Ph-C*H*), δ 7.58-7.47 (m, 4H, Ph-C*H*), δ 7.46-7.35 (m, 6H, Ph-C*H*), δ 5.56 (m, 1H, C*H*), δ 4.74 (dt, J = 2.1 Hz, J = 7.2 Hz, 1H, C*H*H), δ 3.73 (dd, J = 9.8 Hz, J = 13.7 Hz, 1H, C*H*H), δ 3.10 (d, J = 6.6 Hz, 1H, C*H*H), δ 2.79 (d, J = 12.3 Hz, 1H, C*H*H); ^{13}C-NMR (101 MHz, $CDCl_3$), δ 191.7 (s, -*C*HO), δ 140.1 (d, $^1J_{CP}$ = 37 Hz, benzaldehyde-*p*-C_q-P), δ 137.2 (s, benzaldehyde-*ipso*-C_q), δ 134.4 (d, $^2J_{CP}$ = 12 Hz,

benzaldehyde-*m*-*C*H), δ 134.1 (d, $^{2}J_{CP}$ = 13 Hz, Ph-*o*-*C*H), δ 134.1 (d, $^{2}J_{CP}$ = 13 Hz, Ph-*o*-*C*H), δ 131.3 (d, $^{1}J_{CP}$ = 42 Hz, Ph-*ipso*-C_q-P), δ 131.0 (s, Ph-*p*-*C*H), δ 129.3 (d, $^{3}J_{CP}$ = 11 Hz, benzaldehyde-*o*-*C*H), δ 128.9 (d, $^{3}J_{CP}$ = 10 Hz, Ph-*m*-*C*H), δ 118.3 (d, $^{3}J_{CP}$ = 5 Hz, allyl-*C*H), δ 80.4 (d, $^{2}J_{CP}$ = 31 Hz, P-trans-allyl-*C*H_2), δ 61.4 (s, , P-cis-allyl-*C*H_2); ^{31}P-NMR (162 MHz, $CDCl_3$), δ 23.3 (s); Peak assignment based on literature data using the triphenylphosphine analogue.[9]

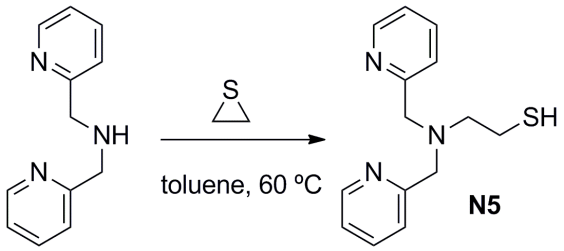
*A8.2.2 General procedure for bioconjugation of preformed phosphine aldehyde metal complexes with SCP-2L V83C-**M5***

A series of parallel reactions were performed using different equivalents of the complex (usually 1, 2, 3, 5, 10 and 20 equivalents). This was done by addition of the appropriate amount of a 25 mM solution of the complex in MeCN to 3.25 nmol SCP-2L V83C-X in 250 µL of the appropriate buffer (usually 20 mM MES, 50 mM NaCl pH 6) in a 1.5 ml or 0.5 ml Eppendorf. The reactions were mixed overnight by slow continues inversion on a rotating wheel. The reaction mixtures were then centrifuged (16100 *g*, 10 minutes) to remove formed precipitate and excess complex. The solutions were directly used for analysis by LCMS(ES^+).

A9 Nitrogen donor ligand synthesis

A9.1 Synthesis of *N*,*N*-dipicolyl-2-aminoethanethiol N5

Thiol containing **N5** was synthesised to provide a nitrogen ligand that can be introduced by disulfide bridge formation. The reaction between dipicolylamine with ethylene sulfide gave **N5** in quantitative yield (Scheme 2). Surprisingly, the NMR data we obtained for **N5** did not match that reported in literature.[10] Whereas no signals for the thiol proton or coupling constants to this proton by the neighbouring CH_2's were reported, in our ^{1}H-NMR spectrum of **N5** the signals belonging to the thiol proton including coupling constants could clearly be observed. An IR spectrum was recorded of **N5** to support the correct interpretation of our NMR data. A band at 2487.0 cm^{-1} was observed in the obtained IR-spectrum which is typical for a SH stretch band. When **N5** was left in $CDCl_3$ solution for more than 16 hours and the ^{1}H-NMR spectrum was recorded a small amount of the compound reported in literature appeared. This result demonstrated that **N5** does slowly transform into the disulfide bridge dimer of **N4** most likely by air exposure.



Scheme 2. Synthesis of dipicolylamine based ligand **N5**.

A9.2.Bioconjugation of N5 to SCP-2L V83C

In contrast to the maleimide conjugation reaction of SCP-2L V83C and A100C with **N4**, the modification reaction between **N5** and SCP-2L V83C did only lead to 50% formation of the modified protein, even when a large excess of **N5** was used (200 eq). Full conversion to SCP-2L V83C-**N5** could be achieved by addition of ligand **N5** as the iron(II)chloride adduct (1 equivalent). However, it is unclear whether the iron is still present inside the protein or is fully removed by washings. Because the addition of oxidative metal salts can give problems for catalytic application this is not a preferred modification procedure. Other oxidative additives to catalyse the reaction which can be removed efficiently after the reaction by dialysis or centrifugal concentration would be more suitable for this approach.

A9.3 Structural investigation into SCP-2L V83C-N5

Some studies into the effect of modification with **N5** on the structure of SCP-2L V83C were performed. The CD-spectrum (Figure 11), Watergate ^{1}H-NMR spectrum (Figure 12) were recorded. The CD-spectrum shows changes similar to those observed for the introduction of dipicolylamine based ligand **N4** (section 5.3.4). From both the NMR and CD spectra can be observed that the structure of SCP-2L V83C-**N5** seems to be very similar compared to that of SCP-2L V83C. This was confirmed by the fluorescent binding assay which gave a K_d of 0.20 (+/-0.05) μM for SCP-2L V83C-**N5**. This K_d value is of the same order of magnitude as the K_d of 0.18 (±0.07) μM found for SCP-2L V83C itself.

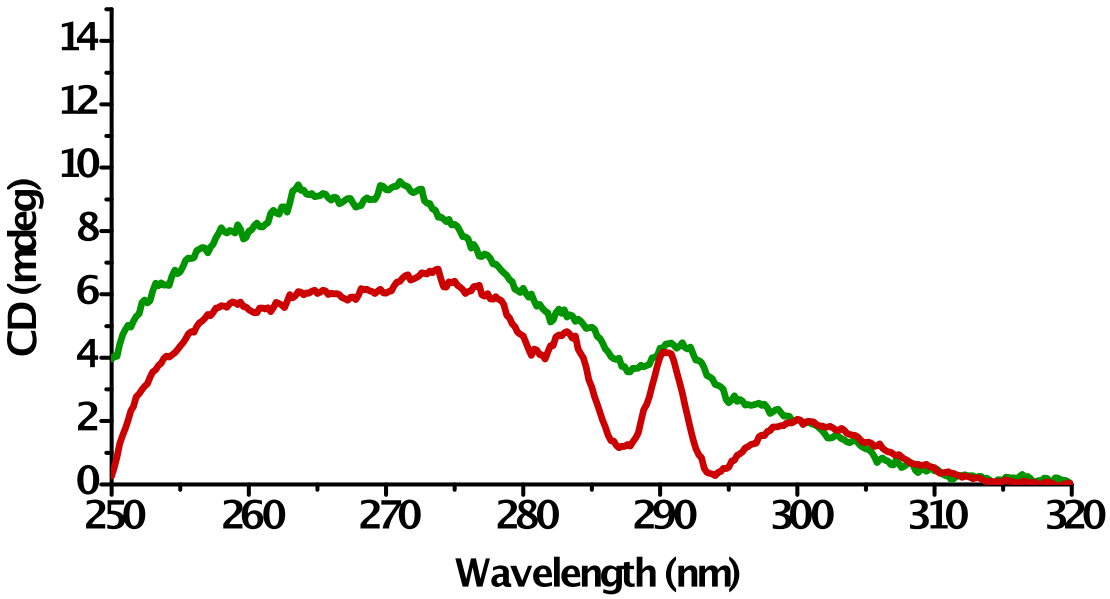


Figure 11. Near UV CD-spectrum of SCP-2L V83C (red) and SCP-2L V83C-**N5** (green).

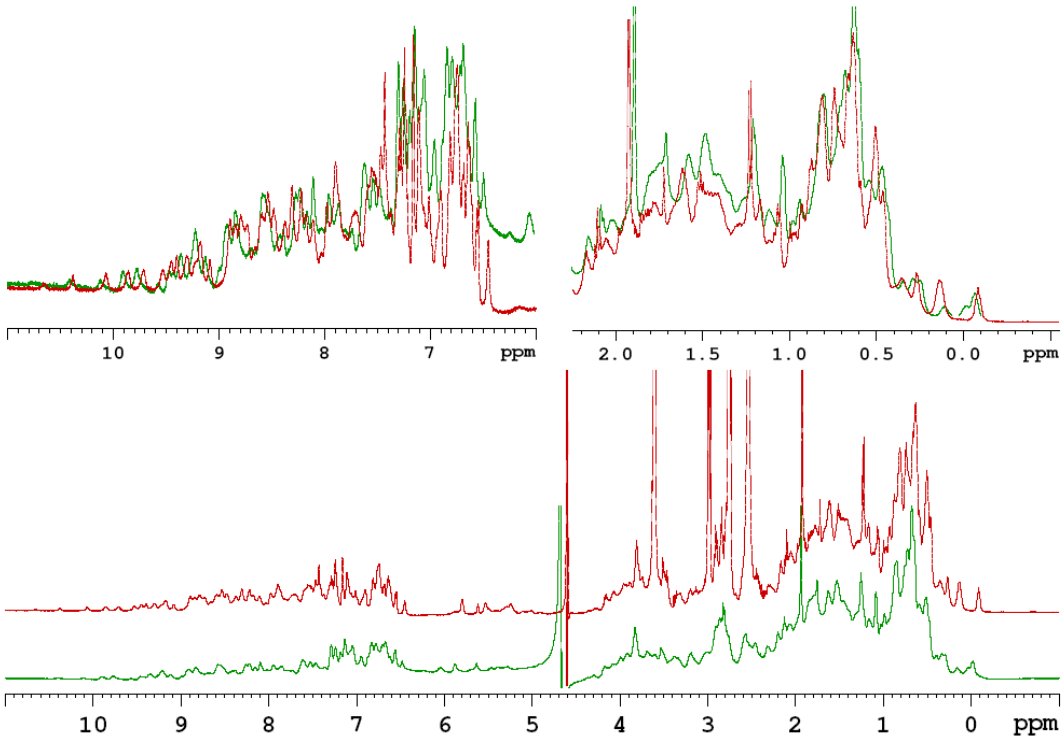
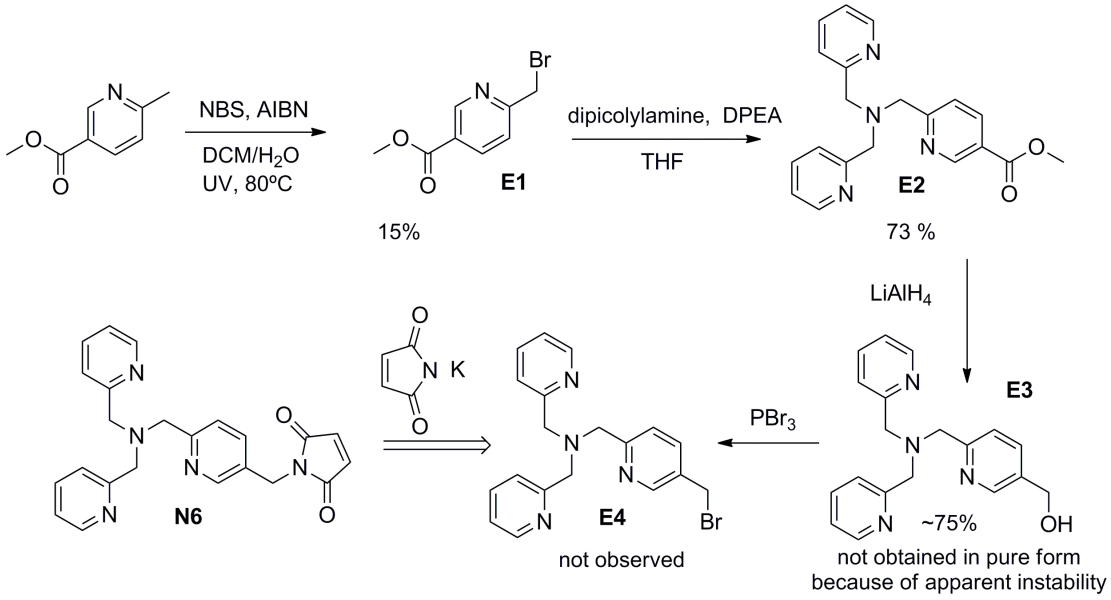


Figure 12. Watergate ^{1}H-NMR spectra of SCP-2L V83C (red) and SCP-2L V83C-**N5** (green) with overlays of the areas between11 and 6 ppm and 2.3 to -0.5 ppm.

A9.4 Synthesis towards mono functionalised tripicolylamine nitrogen donor ligands N6 and N7

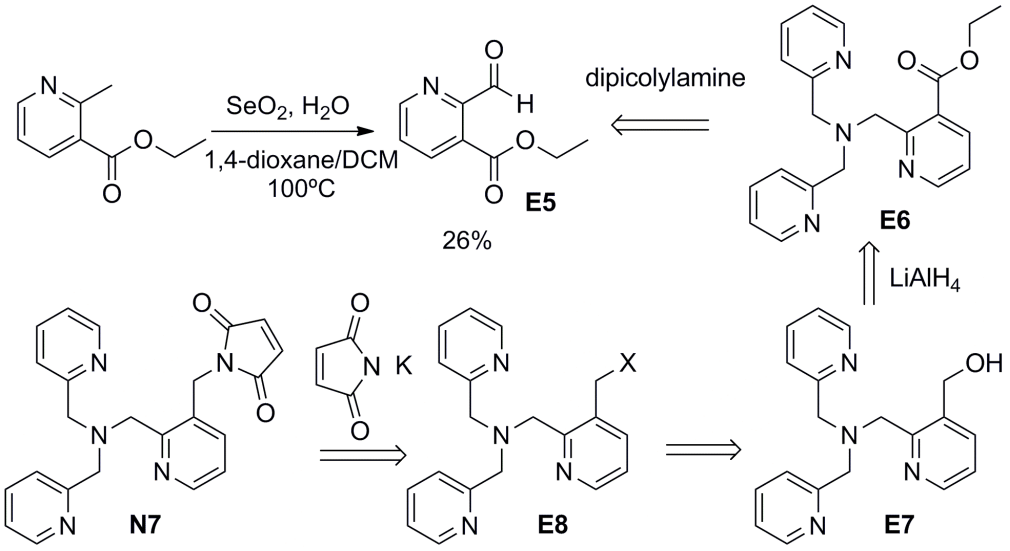
Synthetic routes towards additional nitrogen ligands containing tripicolylamine motifs for site selective introduction into proteins were also undertaken. The strategies outlined in Scheme 3 and Scheme 4 were pursued to introduce a maleimide attached in an 3- (**N6**) and 4- position (**N7**) on one of the pyridine rings.



Scheme 3. Synthesis towards ligand **N6**.

The synthetic strategy towards **N6** made use of a reported synthesis for **E3**.[11] The target ligand **N6** was thought to be accessible through bromination of methyl(6-methyl)nicotinate followed by nucleophilic substitution using potassium maleimide. The reported synthetic route to **E3** was performed via **E1** through bromination of methyl(6-methyl)nicotinate. This bromination was accomplished with NBS and tetrachloromethane affording product **E1** in 57% yield. However, tetrachloromethane is very toxic and therefore we tried to find an alternative method for this first step. Bedel *et al* reported the bromination of similar compounds with NBS in a mixture of water and dichloromethane.[12] The mono and dibrominated product ratios reported in the article could be reproduced by application of this method; however, the yield was very low (15%) due to the formation of what seemed to be insoluble polymeric products. **E2** was obtained in a 73% yield, which was higher than the reported 62%.[11b] **E3** was reported to be unstable,[11b] which prevented it from being obtained in a pure form. Small amounts of unidentified degradation products of **E3** were

observed even after column chromatography purification. Finally, **E3** was obtained with traces of impurities in about 75% yield. This mixture was used for the bromination step using PBr_3 to obtain **E4**. A mixture of different compounds was obtained but none could be identified as **E4**. Alternatively, **N6** should be accessible by formation of the tosylate or other compounds which transform the alcohol of **E3** into a good leaving group.

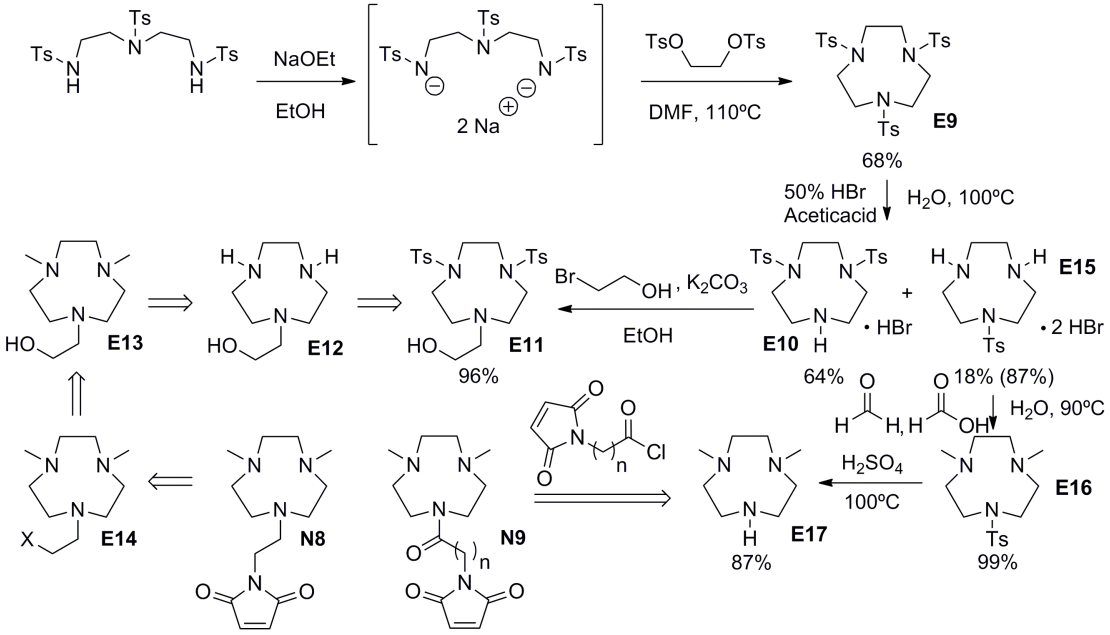


Scheme 4. Synthesis towards ligand **N7**.

The first step in the synthetic route towards nitrogen ligand **N7** was also performed. The formation of tripicolylamine structures like **E6** can be obtained by a reaction of aldehydes similar to **E5** with dipicolylamine.[13] These reactions were reported to provide the tripicolylamine motif in higher yields compared to the reaction with the bromides as used for the synthesis of **E2**. The most convenient method to obtain aldehyde **E5** seemed to be by selective oxidation of the substituted picoline using selenium dioxide according to a procedure applied by Behforous *et al* for which yields in the range of 40-90% were reported.[14] **E5** was obtained in 26% by application of this reaction with the corresponding carboxylic acid, the *N*-oxide and polymeric compounds being the main side products. Although the yield was low, it appeared better than the corresponding NBS-bromination reaction described above.

A9.5 Synthesis towards functionalised triazacyclononane nitrogen donor ligands N8 and N9

The synthetic strategy towards triazacyclone nitrogen ligands **N8** and **N9** is shown in Scheme 5.



Scheme 5. Synthesis towards ligands **N8** and **N9**. The value between brackets given for **E15** shows the obtained yield of the selective synthesis of this compound by a procedure reported in literature.[16]

In the first step the tritosylate **E9** is prepared by a reported procedure performed by various groups with variable yields.[15] In our hands a yield of 65% of **E9** was obtained. By application of a procedure that was supposed to provide the complete detosylation of **E9** a mixture of **E10** and **E15** (3.6:1) was obtained.[15d] The products **E10** and **E15** were easily separated by sequential crystallisation and were identified using $^{1}H/^{13}C$-NMR and crystallography (see appendices A9.6.7 and A9.7 for further information). Alternatively the monotosylate **E15** could also be obtained in one step using a selective detosylation step developed by Sessler *et al* in an 87% yield.[16] These partially protected triazacyclononane compounds offered several possibilities. The ditosylate **E10** can be mono functionalised. The functionalization had however to be able to withstand the harsh tosylate deprotection conditions. Therefore, **E10** was functionalised with 2-bromoethanol to provide the alcohol **E11** in 96% yield. From **E11, N8** should be accessible via detosylation to **E12**, methylation to **E13** and transformation of the alcohol to a leaving group **E14** followed by introduction of the

maleimide. The monotosylate **E15** was methylated with a mixture of formaldehyde and formic acid to provide **E16** in near-quantitative yield. Deprotection of **E16** with sulphuric acid gave **E17** in 87% yield. Coupling reactions of the maleimide acid chloride with n = 1 proved difficult due to troublesome product isolation and identification at the small scale at which the reactions were done. Modification with 3-maleimidopropionic acid (n = 2) using DCC/HOBT was also unsuccessful in the first two attempts.

A9.6 experimental data for the synthesis of nitrogen ligands N5-N9

*A9.6.1 Synthesis of N,N-dipicolyl-2-aminoethanethiol **N5***

This synthesis was performed according to a literature procedure.[10] Under an argon atmosphere 0.33 ml (5.56 mmol) ethylene sulfide was added dropwise to a solution of 0.65 ml (3.6 mmol) dipicolylamine in 1 ml of toluene. The resulting mixture was heated for two days at 65ºC. After this the volatiles were evaporated *in vacuo* to give 0.95 g (3.6 mmol) of **N5** as a yellow oil. This oil was stored under argon as exposure to air seemed to lead to decomposition, most likely in the form of the disulfide bridged dimer. ^{1}H-NMR (300 MHz, $CDCl_3$), δ 8.47 (ddd, $^{3}J_{HH}$=4.9 Hz, J_{HH}=1.8 Hz, J_{HH}=1.0 Hz, 2H, Py-*H*), δ 7.60 (dt, $^{3}J_{HH}$=7.6 Hz, J_{HH}=1.8 Hz, 2H, Py-*H*), δ 7.47 (td, $^{3}J_{HH}$=7.6, J_{HH}=1.0 2H, Py-*H*), δ 7.09 (ddd, $^{3}J_{HH}$=7.6 Hz, $^{3}J_{HH}$=4.9 Hz, J_{HH}=1.0 Hz, 2H, Py-*H*), δ 3.78 (s, 4H, Py-C*H*$_2$-N), δ 2.77-2.71 (m, 2H, N-C*H*$_2$-CH_2-SH), δ 2.65-2.56 (m, 2H, HS-C*H*$_2$-CH_2-), δ 1.55 (t, $^{3}J_{HH}$=7.3 Hz, 1H, -S*H*); ^{13}C-NMR (75.47 MHz, $CDCl_3$), δ 159.3 (C_q, Py*C*-CH_2), δ 149.1 (CH, Py*C*-H), δ 136.5 (CH, Py*C*-H), δ 123.1 (CH, Py*C*-H), δ 122.1 (CH, Py*C*-H), δ 60.2 (CH_2, Py-*C*H$_2$-N), δ 57.1 (CH_2, N-*C*H_2-CH_2), δ 22.5 (CH_2, CH_2-*C*H_2-SH); This NMR data does not completely correspond with the data previously reported for this compound, but as no S*H* signal or coupling of the neighbouring CH_2's to the SH-proton is reported, the literature data[10] most likely corresponds to the dimer. IR-spectrum (NaCl plate), 3051.0 cm^{-1} (m, CH stretch band), 3008.7 cm^{-1} (m, CH stretch band), 2933.8 cm^{-1} (m, CH stretch band), 2822.2 cm^{-1} (s, CH stretch band), 2487.0 cm^{-1} (w, SH stretch band).

*A9.6.2 General procedure for bioconjugation of **N5** to SCP-2L V83C*

A 25-50 μM solution of protein in buffer (20 mM MES 50 mM NaCl pH 6) was incubated with the corresponding amount of equivalents of the **N5** added from a 50

mM stock solution in DMF. Progress of the reaction was monitored using LCMS (ES^+) analysis. Typically about 50% conversion to the modified protein was observed after 16 hours. Alternatively an iron(II)chloride adduct of the ligand **N5** was used from a stock solution of the same concentration in water. Full conversion to SCP-2L V83C-**N5** (expected 13633.9 Da, found 13634.0 Da).was obtained after 16 h after analysis by LCMS(ES^+).

*A9.6.3 Synthesis of methyl(6-bromomethyl)nicotinate **E1***

A solution of 3.92 g (25.95 mmol) methyl(6-methyl)nicotinate, 5.08 g (28.55 mmol) NBS, 0,43 (2.6 mmol) AIBN in 50 ml water and 50 ml CH_2Cl was irradiated by a 300 W halogen lamp for three hours, while stirring vigorously and refluxing by the heat of the lamp. The organic layer was separated and dried with $MgSO_4$. The solvent was evaporated and the resulting mixture of starting material, monobrominated **E1** and dibrominated product was separated by column chromatography with PE:EtOAc 4:1 over neutralised (with Et_3N) silica, yielding 0.85 g (3.7 mmol 15%) **E1** as a red solid. ^{1}H-NMR (400 MHz, $CDCl_3$); δ 9.17 (d, $^4J_{HH}$ = 1.8, 1H, 1-Py-*H*), δ 8.32 (dd, $^3J_{HH}$ = 8.1, $^4J_{HH}$ = 1.8, 1H, 3-Py-*H*), δ 7.55 (d, $^3J_{HH}$ = 8.1, 1H, 4-Py-*H*), δ 4.60 (s, 2H, Py-CH_2-Br), δ 3.97 (s, 3H, -COOCH_3), similar to the data reported in literature.[11b, 17]

*A9.6.4 Synthesis of **E2***

This synthesis was performed according to a literature procedure.[11b, 17] A mixture of 0.76 g (3.7 mmol) 2-dipicolylamine, 0.74 g (3.22 mmol) methyl(6-bromomethyl)nicotinate **E1** and 1 ml (6 mmol) diisopropylethylamine in 40 ml THF was stirred for 24 hours. The resulting solution was filtered over celite providing a dark red solution. The THF was removed *in vacuo* and the resulting red/brown oil was dissolved in ether and again filtered through celite. Several crystallisations from ether at -20ºC gave a combined yield of 0.722 g (2.07 mmol, 64%) **E2** as a yellow solid. ^{1}H-NMR (400 MHz, $CDCl_3$); δ 9.1 (d, $^4J_{HH}$ = 1.8, 1H, 1-EPy-*H*), δ 8.55 (d, $^3J_{HH}$ = 4.1, 2H, 1-Py-*H*), δ 8.26 (dd, $^3J_{HH}$ = 8.1, $^4J_{HH}$ = 1.8, 1H, 3-EPy-*H*), δ 7.75-7.52 (m, 5H), δ 7.22-7.13 (m, 2H), δ 4.05-3.88 (m, 9H), similar to literature data.[11b, 17]

*A9.6.5 Synthesis of **E3***

This synthesis was performed according to a literature procedure.[11b] Under argon atmosphere, a solution of 0.722 g (2.07 mmol) **E2** in dry Et_2O was added drop wise to a solution of 0.2 g (4.6 mmol) $LiAlH_4$ in dry Et_2O cooled to 0ºC. The obtained mixture was stirred at room temperature for 18 hours after which TLC showed complete conversion. The mixture was again cooled to 0ºC and small quantities of water and a 10% NaOH solution were carefully added after which the mixture was stirred for an additional 2 hours. The organic layer was separated and the water layer extracted with Et_2O. The combined organic layers were filtered over celite, dried with Na_2SO_4, filtered and concentrated *en vacuo* resulting in 0.62 g (1.95 mmol, 95%) crude **E3** as a brown oil. ^{1}H-NMR (400 MHz, $CDCl_3$); δ 8.62-855 (m, 3H), δ 7.76-7.56 (m, 6H), δ 7.21-7.16 (m, 2H), δ 4.73 (s, 2H), δ 4.07-3.98 (m, 6H), similar to data reported in literature.[11b]

*A9.6.6 Synthesis of N,N',N''-1.4.7-tritosyl-1,4,7-triazacyclononane **E9***

This synthesis was performed according to a literature procedure.[15] Under argon atmosphere, 2.13 g (31.3 mmol) sodium etoxide in 20 ml dry EtOH was added to a hot slurry of 8.01 g (14.2 mmol) *N,N',N''*-1,4,7-tritosyl-1,4,7-triazaheptane in 20 ml dry EtOH . The ethanol was removed *en vacuo* and the resulting dark brown solid was dissolved in 50 ml dry DMF. This dark brown solution and a solution of 4.99 g (13.5 mmol) ethyleneglycolditosylate were added over 4½ hours to a dry DMF solution at 110ºC using a syringe pump. The reaction mixture was stirred for 4 more hours at this temperature after which it was cooled to room temperature. Water was added resulting in the precipitation of the crude product, that could be isolated by filtration and washing with water. Purification by column chromatography with PE:EtOAc 3:2 over alumina resulted in 5.66 g (9,45 mmol, 68%) **E9** as a white powder. R_f = 0.21; ^{1}H-NMR (400 MHz, $CDCl_3$), δ 7.71 (d, $^3J_{HH}$ = 8.1, 6H, Ts-Ar-*H*), δ 7.33 (d, $^3J_{HH}$ = 8.1, 6H, Ts-Ar-*H*), δ 3.43 (s, 12H, N-C*H*$_2$-C*H*$_2$-N), δ 2.44 (s, 9H, Ts-C*H*$_3$), similar to data reported in literature.[18]

*A9.6.7 Synthesis of. N,N'-ditosyl-1,4,7-triazacyclononane monohydrobromide salt **E10** and N-tosyl-1,4,7-triazacyclononane dihydrobromide salt **E15***

This synthesis was performed according to a literature procedure.[16, 19] A yellow suspension of 5.22 g (8.22 mmol) *N,N',N''*-1.4.7-tritosyl-1,4,7-triazacyclononane **E9**, 60 ml 50% HBr and 30 ml acetic acid was heated to 100ºC for 24 hours. Addition of diethylether and ethanol provided a white precipitate isolated by filtration providing 2.72 g (5.26 mmol, 64%) **E10**, which could be recrystallised to needles in a mixture of HBr, H_2O, EtOH and Et_2O (see section A9.7.1 for crystal structure). ^{1}H-NMR (300 MHz, D_6-DMSO), δ 9.07 (broad-s, 2H, -NH_2^+-), δ 7.72 (d, $^3J_{HH}$ = 8.2, 4H, Ts-Ar-*H*), δ 7.48 (d, $^3J_{HH}$ = 8.2, 4H, Ts-Ar-*H*), δ 3.47 (s, 8H, N-CH_2-CH_2-N), δ 3.33 (s, 4H, N-CH_2-CH_2-N), δ 2.41 (s, 6H, Ts-CH_3); ^{13}C-NMR (300 MHz D_6-DMSO), δ 144.5 (Ts-Ar-C_q), δ 133.9 (Ts-Ar-C_q), δ 130.5 (Ts-Ar-*C*H), δ 127.6 (Ts-Ar-*C*H), δ 51.6 (*C*H_2), δ 47.8 (*C*H_2), δ 45.0 (*C*H_2), δ 21.4 (Ts-*C*H_3). **E15** was obtained as glassy crystals from the filtrate 0.59 g (1.32 mmol, 18%) by slow solvent evaporation (see section A9.7.2 for crystal structure). ^{1}H-NMR (300 MHz, D_6-DMSO), δ 9.07 (broad-s, 4H, -NH_2^+-), δ 7.74 (d, $^3J_{HH}$ = 8.2, 2H, Ts-Ar-*H*), δ 7.49 (d, $^3J_{HH}$ = 8.2, 2H, Ts-Ar-*H*), δ 3.56 (s, 4H, N-CH_2-CH_2-N), δ 3.41 (broad-s, 8H, N-CH_2-CH_2-N), δ 2.20 (s, 3H, Ts-CH_3); ^{13}C-NMR (300 MHz D_6-DMSO), δ 144.5 (Ts-Ar-C_q), δ 132.6 (Ts-Ar-C_q), δ 130.2 (Ts-Ar-*C*H), δ 127.5 (Ts-Ar-*C*H), δ 46.8 (*C*H_2), δ 45.1 (*C*H_2), δ 42.8 (*C*H_2), δ 21.0 (Ts-*C*H_3).

*A9.6.8 Selective synthetic procedure to N-tosyl-1,4,7-triazacyclononane dihydrobromide salt **E15***

This synthesis was performed according to a literature procedure.[16] A reflux setup with a magnetic stirrer was connected to a base-trap with a sodium hydroxide solution. The flask was charged with 3.29 g (5.56 mmol) *N*,*N*',*N*''-1.4.7-tritosyl-1,4,7-triazacyclononane **E9**, 3.29 g (41,7 mmol) phenol freshly crystallised from petroleum ether 40-60ºC and 50 ml 30% HBr in HOAc. The resulting fuming black solution was heated to 90ºC for 36 hours, when a white precipitate had formed. The white precipitate was filtered off and washed with diethylether. The off-white powder was dried en vacuo resulting in 2.16 g (4.85 mmol, 87%) **E15**. More product could be obtained by crystallisation from the filtrate after addition of diethylether. ^{1}H-NMR corresponded to the product of the above reaction from which crystals were obtained.

A9.6.9 Synthesis of N-(2-ethoxy)-N',N''-ditosyl-1,4,7-triazacyclononane ***E11***

75.2 mg (0.6 mmol) 2-bromoethanol and a spatula-tip K_2CO_3 was added to 0.28 g (0.55 mmol) **E10** in 6 ml of ethanol. The mixture was refluxed under inert atmosphere till TLC showed complete conversion (3 days). Dichloromethane was added and separated and the organic layer was washed several times with water and dried on Na_2SO_4. The solvent was removed *in vacuo* yielding the crude product as a pale yellow solid. Column chromatography on silica with EtOAc:PE:Et_3N 19:1:1.5 afforded 0.254 g (0.53 mmol 96%) **E11** as a pale yellow solid. R_f=0.47; ^{1}H-NMR (300 MHz, $CDCl_3$), δ 7.58 (d, $^3J_{HH}$ = 8.3, 4H, Ts-Ar-*H*), δ 7.24 (d, $^3J_{HH}$ = 8.3, 4H, Ts-Ar-*H*), δ 3.53 (t, $^3J_{HH}$ = 4.9, 2H, N-C*H*$_2$-C*H*$_2$-OH), δ 3.34 (s, 4H, N-C*H*$_2$-C*H*$_2$-N), δ 3.18 (broad-s, 4H, N-C*H*$_2$-C*H*$_2$-N), δ 2.94 (t, J_{HH} = 4.6, 4H, N-C*H*$_2$-C*H*$_2$-N), δ 2.36 (s, 6H, Ts-C*H*$_3$); ^{13}C-NMR (300 MHz, $CDCl_3$), δ 143.7 (Ts-Ar-C_q), δ 135.1 (Ts-Ar-C_q), δ 129.9 (Ts-Ar-*C*H), δ 127.1 (Ts-Ar-*C*H), δ 60.4 (*C*H$_2$), δ 59.6 (*C*H$_2$), δ 55.2 (*C*H$_2$), δ 53.3 (*C*H$_2$), δ 52.9 (*C*H$_2$), δ 60.4 (*C*H$_2$), δ 21.5 (Ts-*C*H$_3$).

A9.6.10 Synthesis of N-tosyl-N'-N''-dimethyl-1,4,7-triazacyclononane ***E16***

This synthesis was performed according to a literature procedure.[20] A clear solution of 0.31 g (0.699 mmol) of **E15** in 1.5 ml 37% formaldehyde and 3 ml formic acid was heated to 90ºC for 20 hours. The solution was cooled to room temperature and HCl was added until pH = 1 was reached. All liquid was removed *in vacuo* followed by the addition of more HCl. This process was repeated three times yielding a yellow solid being the HCl salt of **E16**. 0.206 g (6.6 mmol, 96%) free amine of **E16** was obtained by flash column chromatography on silica with EtOAc:MeOH:Et_3N 8:1:1. R_f = 0.21 in EtOAc:PE:Et_3N 19:1:1; ^{1}H-NMR (300 MHz, D_3COD), δ 7.58 (d, $^3J_{HH}$ = 8.4, 2H, Ts-Ar-*H*), δ 7.29 (d, $^3J_{HH}$ = 8.4, 2H, Ts-Ar-*H*), δ 3.15-3.09 (m, 4H, N-C*H*$_2$-C*H*$_2$-N), δ 2.90-2.83 (broad-s, 4H, N-CH$_2$-C*H*$_2$-N), δ 2.90-2.70 (s, 4H, N-C*H*$_2$-CH$_2$-N), δ 2.32 (s, 9H, Ts-C*H*$_3$ and N-C*H*$_3$); ^{13}C-NMR (300 MHz D_6- D_3COD), δ 144.97 (Ts-Ar-C_q), δ 136.9 (Ts-Ar-C_q), δ 130.9 (Ts-Ar-*C*H), δ 128.3 (Ts-Ar-*C*H), δ 56.9 (2x *C*H$_2$), δ 51.3 (*C*H$_2$), δ 45.8 (N-*C*H$_3$), δ 21.4 (Ts-*C*H$_3$).

*A9.6.11 Synthesis of N,N-dimethyl-1,4,7-triazacyclononane **E17***

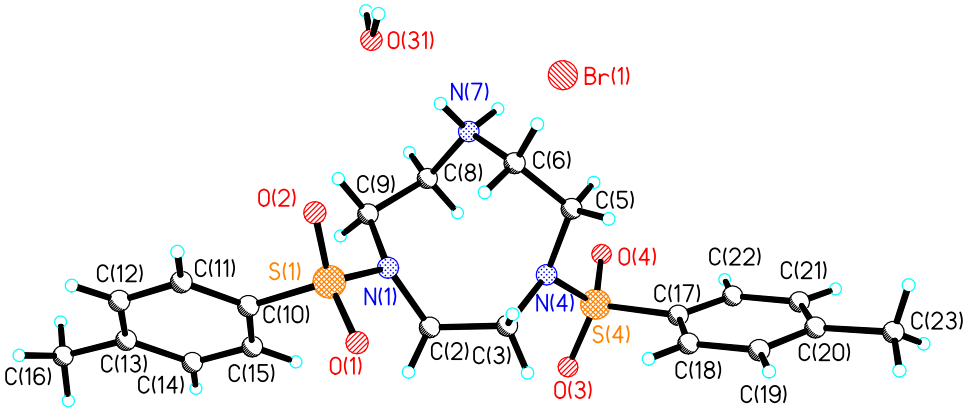
This synthesis was performed according to a literature procedure.[16, 20] A brown mixture of 0.55 g (1.76 mmol) *N*-tosyl-*N'*-*N''*-dimethyl-1,4,7-triazacyclononane **E16** and 5 ml sulphuric acid was heated to 120ºC for 2 days under argon atmosphere. The resulting black solution was cooled to 0ºC followed by the addition of a small amount of water. This mixture was made basic by careful addition of a concentrated sodium hydroxide solution. This mixture was extracted five times with chloroform. The combined organic layers where combined and the solvent was removed en vacuo providing 0.242 g (1.54 mmol, 87%) **E17** as a yellow oil. This yellow oil was stored under argon at -4ºC, since it was reported to decompose slowly at room temperature in air. This was confirmed by the observation of extra peaks in ^{1}H and ^{13}C and-NMR over time. 1H-NMR (400 MHz, D_3COD), δ 2.58-2.53 (m, 4H, N-C*H*$_2$-CH_2-N), δ 2.50-2.45 (m, 4H, N-C*H*$_2$-CH_2-N), δ 2.44 (s, 4H, N-C*H*$_2$-CH_2-N), δ 2.30 (s, 6H, N-C*H*$_3$); 13C-NMR (400 MHz, D_3COD), δ 54.9 (*C*H_2), δ 53.6 (*C*H_2), δ 46.1 (*C*H_2), δ 45.4 (N-*C*H_3); MS direct injection (ES+), m/z (relative intensity), 158.16 (M + H+, 100%), 180.16 (M + Na+, 52%), exact mass calculated for $C_8H_{19}N_3$ m/z 157.16, + H+ m/z 158.17, + Na m/z 180.15. Also a peak at m/z 170.17 was observed with 85% intensity, which could correspond to a decomposition product.

A9.7 Crystallographic data obtained for the crystals of E10 and E15

*A9.7.1 Crystallographic data of **E10***

Below the crystallographic data is presented for **E10** (Figure 13 and Table 4). An overview of all bond angles and distances can be found on the accompanying CD.

a)



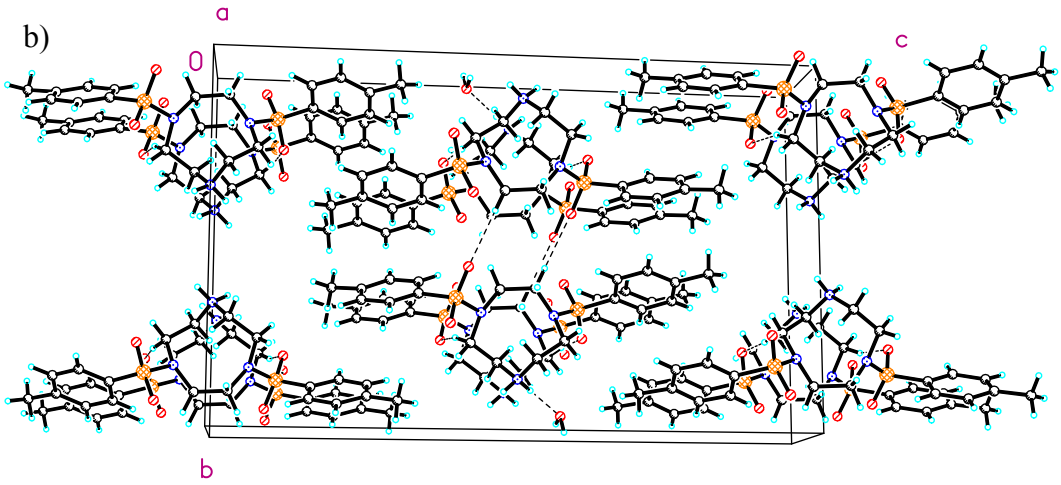


Figure 13. a) Crystal structure of **E10**. b) Crystal packing of **E10**.

Table 4. Crystal data and structure refinement for **E10** (St. Andrews reference = pdpk2).

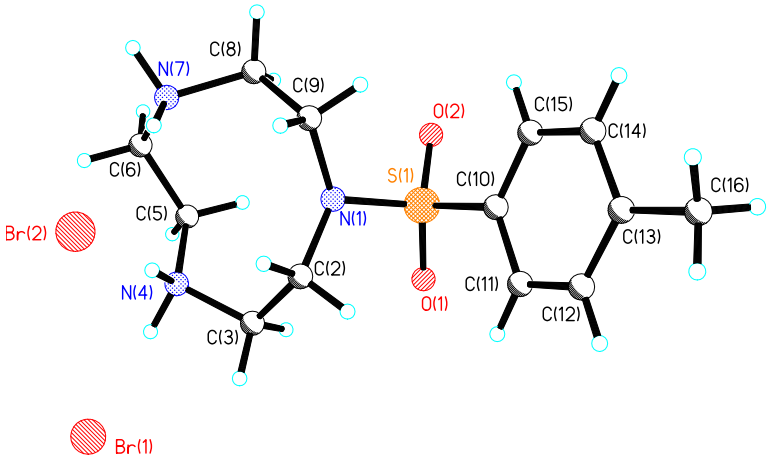
Identification code	pdpk2	
Empirical formula	C20 H30 Br N3 O5 S2	
Formula weight	536.50	
Temperature	93(2) K	
Wavelength	0.71073 Å	
Crystal system	Monoclinic	
Space group	P2(1)/n	
Unit cell dimensions	a = 5.788(2) Å	α = 90°.
	b = 15.674(7) Å	β = 92.034(7)°.
	c = 25.454(8) Å	γ = 90°.
Volume	2307.8(16) $Å^3$	
Z	4	
Density (calculated)	1.544 Mg/m^3	
Absorption coefficient	1.999 mm^{-1}	
F(000)	1112	
Crystal size	0.2000 x 0.0300 x 0.0100 mm^3	
Theta range for data collection	1.53 to 25.34°.	
Index ranges	-6<=h<=6, -18<=k<=18, -30<=l<=30	
Reflections collected	21254	

Independent reflections	4110 [R(int) = 0.0785]
Completeness to theta = 25.00°	98.1 %
Absorption correction	Multiscan
Max. and min. transmission	1.0000 and 0.6931
Refinement method	Full-matrix least-squares on F^2
Data / restraints / parameters	4110 / 4 / 297
Goodness-of-fit on F^2	1.093
Final R indices [I>2sigma(I)]	R1 = 0.0625, wR2 = 0.1333
R indices (all data)	R1 = 0.0795, wR2 = 0.1434
Largest diff. peak and hole	0.977 and -0.417 e.Å^{-3}

*A9.7.2 Crystallographic data of **E15***

Below the crystallographic data is presented for **E15** (Figure 14 and Table 5). An overview of all bond angles and distances can be found on the accompanying CD.

a)



b)

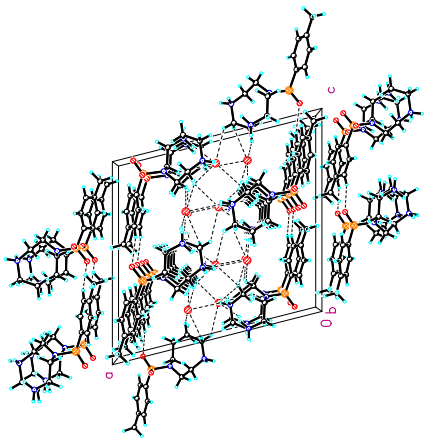


Figure 14. a) Crystal structure of **E15.** b) Crystal packing of **E15**.

Table 5. Crystal data and structure refinement for E15 (St. Andrews reference = pdpk1).

Identification code	pdpk1	
Empirical formula	C13 H23 Br2 N3 O2 S	
Formula weight	445.22	
Temperature	93(2) K	
Wavelength	0.71073 Å	
Crystal system	Monoclinic	
Space group	P2(1)/c	
Unit cell dimensions	a = 16.947(4) Å	α = 90°
	b = 6.3810(14) Å	β = 104.277(11)°
	c = 15.919(4) Å	γ = 90°
Volume	1668.3(7) $Å^3$	
Z	4	
Density (calculated)	1.773 Mg/m^3	
Absorption coefficient	4.992 mm^{-1}	
F(000)	896	
Crystal size	0.100 x 0.100 x 0.010 mm^3	
Theta range for data collection	2.48 to 25.35°.	
Index ranges	-19<=h<=20, -7<=k<=7, -16<=l<=18	
Reflections collected	8684	
Independent reflections	2773 [R(int) = 0.0662]	
Completeness to theta = 25.00°	91.3 %	
Absorption correction	Multiscan	
Max. and min. transmission	1.0000 and 0.5075	
Refinement method	Full-matrix least-squares on F^2	
Data / restraints / parameters	2773 / 4 / 209	
Goodness-of-fit on F^2	1.148	
Final R indices [I>2sigma(I)]	R1 = 0.0579, wR2 = 0.1148	
R indices (all data)	R1 = 0.0700, wR2 = 0.1209	
Extinction coefficient	0.0045(6)	
Largest diff. peak and hole	0.953 and -0.666 $e.Å^{-3}$	

A10 $N4CuCl_2$ crystal structure and EPR spectrum

A10.1 Crystal structure data of $N4CuCl_2$

Below the crystallographic data is presented for **N4**$CuCl_2$ (Figure 15 and Table 6). An overview of all bond angles and distances can be found on the accompanying CD.

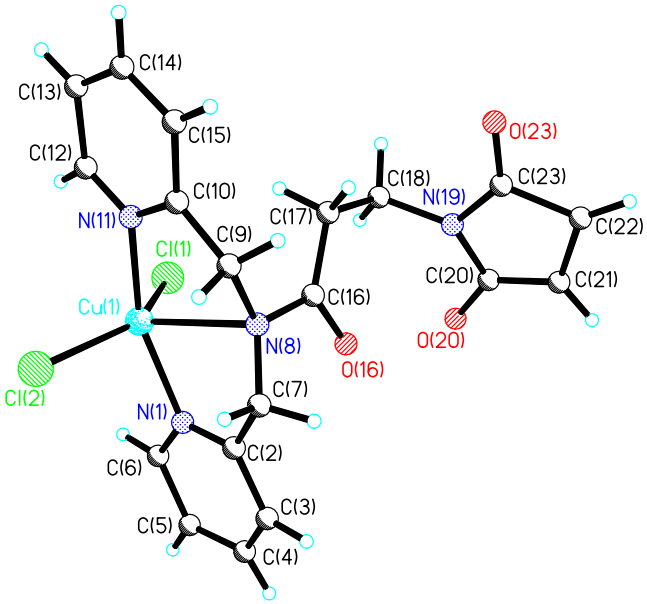


Figure 15. Crystal structure of **N4**$CuCl_2$.

Table 6. Crystal data and structure refinement for **N4**$CuCl_2$ (St. Andrews reference = pdpk5).

Identification code	pdpk5	
Empirical formula	C19 H18 Cl2 Cu N4 O3	
Formula weight	484.81	
Temperature	93(2) K	
Wavelength	0.71073 Å	
Crystal system	Monoclinic	
Space group	P2(1)/c	
Unit cell dimensions	a = 17.633(7) Å	α = 90°.
	b = 8.433(3) Å	β = 99.18(2)°.
	c = 13.478(5) Å	γ = 90°.
Volume	1978.4(12) $Å^3$	
Z	4	

Density (calculated)	1.628 Mg/m^3
Absorption coefficient	1.404 mm^{-1}
F(000)	988
Crystal size	0.1000 x 0.0100 x 0.0100 mm^3
Theta range for data collection	2.34 to 25.39°.
Index ranges	-17<=h<=21, -7<=k<=10, -14<=l<=15
Reflections collected	12110
Independent reflections	3584 [R(int) = 0.3939]
Completeness to theta = 25.00°	99.0 %
Absorption correction	Multiscan
Max. and min. transmission	1.0000 and 0.6759
Refinement method	Full-matrix least-squares on F^2
Data / restraints / parameters	3584 / 102 / 262
Goodness-of-fit on F^2	1.277
Final R indices [I>2sigma(I)]	R1 = 0.2570, wR2 = 0.5203
R indices (all data)	R1 = 0.4041, wR2 = 0.5815
Largest diff. peak and hole	3.529 and -1.825 $e.Å^{-3}$

A10.2 EPR spectrum N4CuCl$_2$

An additional advantage of the application of copper(II) complexes is the possibility to use EPR for their characterisation. This technique can be used to provide information on the surroundings of the metal centre and can also be applied to EPR-active transition metals in biomolecules.[21] Regarding the conjugation of copper-nitrogen ligand complexes to proteins, this technique can distinguish coordination of the copper to the desired nitrogen donor ligand environment from coordination to other potential coordination sites provided by the amino acid side chains. Initial experiments for this purpose were performed by recording the continuous wave (CW) X-band EPR spectrum of 0.1 mM **N4**CuCl$_2$ in 20 mM MES buffer pH 6 containing 20% glycerol. A spectrum was obtained by cooling the sample to 77 K with liquid nitrogen and optimisation of the power to 0.1 mW (Figure 16). A typical copper(II) spectrum was obtained in which hyperfine splitting for the nitrogen donor atoms can be observed in the downwards slope between 3200 and 3300 G. This spectrum could

be used as reference spectrum for EPR-spectra obtained for the proteins modified with copper-complexes of dipicolylamine based ligands.

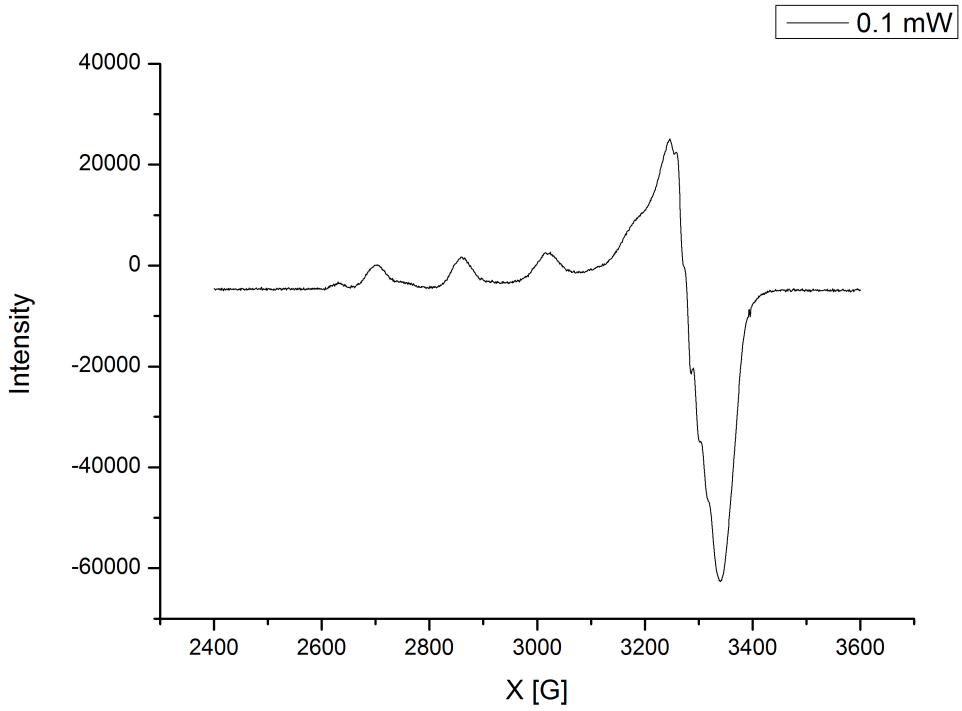


Figure 16. CW X-band EPR spectrum of 0.1 mM **N4**$CuCl_2$ in 20 mM MES buffer pH 6 containing 20% glycerol at 0.1 mW.

EPR spectra were recorded with the kind help and supervision of Dr. Olav Schiemann at the facilities of the University of St. Andrews.

A11 References

[1] www.pdb.org, *World Wide Protein Data Base (wwPDB)*, 01-04, **2010**.

[2] R. B. Kapust, J. Toezser, J. D. Fox, D. E. Anderson, S. Cherry, T. D. Copeland, D. S. Waugh, *Protein Eng.* **2001**, *14*, 993.

[3] W. A. Stanley, A. Sokolova, A. Brown, D. T. Clarke, M. Wilmanns, D. I. Svergun, *Journal of Synchrotron Radiation* **2004**, *11*, 490.

[4] a) R. den Heeten, B. K. Munoz, G. Popa, W. Laan, P. C. J. Kamer, *Dalton Trans.* **2010**, *39*, 8477; b) W. Laan, B. K. Munoz, R. den Heeten, P. C. J. Kamer, *ChemBioChem* **2010**, *11*, 1236.

[5] a) G. Jones, P. Willett, R. C. Glen, *J. Mol. Biol.* **1995**, *245*, 43; b) G. Jones, P. Willett, R. C. Glen, A. R. Leach, R. Taylor, *J. Mol. Biol.* **1997**, *267*, 727; c) M. L. Verdonk, J. C. Cole, M. J. Hartshorn, C. W. Murray, R. D. Taylor, *Proteins Struct., Funct., Genet.* **2003**, *52*, 609; d) J. W. M. Nissink, C. Murray, M. Hartshorn, M. L. Verdonk, J. C. Cole, R. Taylor, *Proteins Struct., Funct., Genet.* **2002**, *49*, 457; e) J. C. Cole, J. W. M. Nissink, R. Taylor, *Drug Discovery Series* **2005**, *1*, 379; f) M. L. Verdonk, G. Chessari, J. C. Cole, M. J.

Hartshorn, C. W. Murray, J. W. M. Nissink, R. D. Taylor, R. Taylor, *J. Med. Chem.* **2005**, *48*, 6504; g) M. J. Hartshorn, M. L. Verdonk, G. Chessari, S. C. Brewerton, W. T. M. Mooij, P. N. Mortenson, C. W. Murray, *J. Med. Chem.* **2007**, *50*, 726.

[6] J. Kopp, T. Schwede, *Nucleic Acids Res.* **2004**, *32*, D230.

[7] www.pymol.org, *Pymol, molecular visualization system*, 19-02, **2011**.

[8] N. Svensen, P. Fristrup, D. Tanner, P.-O. Norrby, *Adv. Synth. Catal.* **2007**, *349*, 2631.

[9] S. Fantasia, S. P. Nolan, *Chem.--Eur. J.* **2008**, *14*, 6987.

[10] N. Lazarova, J. Babich, J. Valliant, P. Schaffer, S. James, J. Zubieta, *Inorg. Chem.* **2005**, *44*, 6763.

[11] a) V. S. I. Sprakel, J. A. A. W. Elemans, M. C. Feiters, B. Lucchese, K. D. Karlin, R. J. M. Nolte, *Eur. J. Org. Chem.* **2006**, 2281; b) K. J. Humphreys, K. D. Karlin, S. E. Rokita, *J. Am. Chem. Soc.* **2002**, *124*, 6009.

[12] S. Bedel, G. Ulrich, C. Picard, *Tetrahedron Lett.* **2002**, *43*, 1697.

[13] Z. He, D. C. Craig, S. B. Colbran, *J. Chem. Soc., Dalton Trans.* **2002**, 4224.

[14] a) M. Hassani, W. Cai, D. C. Holley, J. P. Lineswala, B. R. Maharjan, G. R. Ebrahimian, H. Seradj, M. G. Stocksdale, F. Mohammadi, C. C. Marvin, J. M. Gerdes, H. D. Beall, M. Behforouz, *J. Med. Chem.* **2005**, *48*, 7733; b) M. Behforouz, W. Cai, M. G. Stocksdale, J. S. Lucas, J. Y. Jung, D. Briere, A. Wang, K. S. Katen, N. C. Behforouz, *J. Med. Chem.* **2003**, *46*, 5773; c) M. Behforouz, Z. Gu, W. Cai, M. A. Horn, M. Ahmadian, *J. Org. Chem.* **1993**, *58*, 7089.

[15] a) J. E. Richman, T. J. Atkins, *J. Am. Chem. Soc.* **1974**, *96*, 2268; b) R. Yang, L. J. Zompa, *Inorg. Chem.* **1976**, *15*, 1499; c) D. W. White, B. A. Karcher, R. A. Jacobson, J. G. Verkade, *J. Am. Chem. Soc.* **1979**, *101*, 4921; d) K. Wieghardt, W. Schmidt, B. Nuber, J. Weiss, *Chem. Ber.* **1979**, *112*, 2220; e) H. Koyama, T. Yoshino, *Bull. Chem. Soc. Jpn.* **1972**, *45*, 481.

[16] J. L. Sessler, J. W. Sibert, V. Lynch, *Inorg. Chem.* **1990**, *29*, 4143.

[17] Z. Tyeklar, R. R. Jacobson, N. Wei, N. N. Murthy, J. Zubieta, K. D. Karlin, *J. Am. Chem. Soc.* **1993**, *115*, 2677.

[18] O. Iranzo, T. Elmer, J. P. Richard, J. R. Morrow, *Inorg. Chem.* **2003**, *42*, 7737.

[19] a) A. J. Dickie, D. C. R. Hockless, A. C. Willis, J. A. McKeon, W. G. Jackson, *Inorg. Chem.* **2003**, *42*, 3822; b) K. Wieghardt, I. Tolksdorf, W. Herrmann, *Inorg. Chem.* **1985**, *24*, 1230.

[20] C. Flassbeck, K. Wieghardt, *Z. Anorg. Allg. Chem.* **1992**, *608*, 60.

[21] a) J. Podtetenieff, A. Taglieber, E. Bill, E. J. Reijerse, M. T. Reetz, *Angew. Chem., Int. Ed.* **2010**, *49*, 5151; b) K. Tanaka, A. Tengeiji, T. Kato, N. Toyama, M. Shionoya, *Science (Washington, DC, U. S.)* **2003**, *299*, 1212; c) G. H. Clever, S. J. Reitmeier, T. Carell, O. Schiemann, *Angew. Chem., Int. Ed.* **2010**, *49*, 4927.

Acknowledgements

I would never have been able to complete this thesis without the help of a vast number of people. Firstly I would like to state that it has been an amazing time in St. Andrews, which is mainly thanks to the people mentioned below! (Sorry non-Dutchies for the parts in Dutch).

The first to thank is Paul who created the concept of this project, found the neccary funding and most importantly let me work on it. *Ook tijdens het onderzoek stond je deur altijd open voor verhitte discussie over de vele problemen en mogelijkheden die langskwamen. Er zijn nog veel van jouw goede ideeën niet uitgewerkt dus ik hoop dat er nog vele mooie resultaten uit dit werk voortvloeien. Ook veel dank voor de tips over het positief presenteren van resultaten, iets wat zeker geholpen heeft bij latere sollicitaties, en ook geven van zelfvertrouwen in de vorm van positieve feedback. Verder waren er natuurlijk de gezellige bezoekjes waarvoor ook Gini (veel succes met het webdesign!) veel gratie voor verdient*.

I also have to thank Bob (Tooze) from Sasol for making this research possible and for being my secondairy supervisors together with David (Cole-Hamilton). Also Hans (de Vries) and David (O'Hagan) for the great discussion and critically overhaul of the project during the examination *(thanks for reading this monster of a thesis and suggestions for improvements!).*

The most important for this research was surely Wouter, for which I owe him many thanks. *Als een meer directe begeleider van dit project was je onmisbaar voor het succesvol volbrengen van dit onderzoek. Je hebt me ongelofelijk veel geleerd vanaf het produceren van eiwitten in bacteriën, het isoleren van DNA van agarose gels tot het corrigeren van mijn vernederlandste Engels schrijfwerk. Verder was het natuurlijk vaak gezellig in de Union, the Cellar of the London Nightclub.*

Many thanks to the other group members that came, went and several that were there for almost the entire time of my stay, for providing a nice work environment. Jason, *we started together and I will forever miss the great (and too frequent) Saturday afternoon coffee breaks in which everything got discussed!* The other PCJK-students with whom I shared a good part of my time with: Gina (and *Alina, who was also sort of part of our group*), Tanja (*nice pirate stories*), Christine, Michiel (*Dank, voor het helpen met een aantal zeer nuttige kleinigheidjes die ik nog nodig had op afstand*

tijdens het schrijfwerk), Johanneke, (*onze collega op afstand en het is mooi dat je ook proefschrift succesvol hebt afgerond!)*, Stephen and for a very short time Lorenz (*good luck with continuing this work and keep up the e-mail discussions!*). The PCJK-post-docs: Bert, (*Van jouw heb ik ongelovelijk veel geleerd in dat jaartje dat we tegelijk in St. Andrews waren en ook samen met Annemiek was het vaak gezellig. Veel plezeier met de kleine!)*, Gregorio (*also a bit as student*), Bianca (*Hopefully I get a chance to play some more golf with you and Angel*), Arnald (*thanks for your critical views on the research, I hope you will be reasonable to your own students!*), Debby (*Thanks for the good practical advice and I hope you overcame the shock of the "Oranjehuis" in Poznan that I dragged you into*), René (*altijd gezellig dat ene biertje!*), Sergei, Nikos, a little bit of Jenny en de speciale bezoeker Bauke. The students who I hope learned something from me: Craig and especially Yiteng (*Thanks for approaching you work with so much eager!*).

I have to thank several people that helped me with various experiments: Catherine, Sally and Alex from the MS department were I spent a lot of time, Melanja and Thomas from the NMR department, Alex (*Crystals and always funny e-mail subjects*), Ine (*vooral veel dank voor het tonen van de benodigde techniek voor HP-NMR en het "uitlenen" van spullen van Nolan*), Olav (*EPR*), Bobby (*what did you say?*), George and Colin, Peter (*GC's*), Baslav. Also the people with whom I shared the "biochemistry lab": Jane (*solved more small problems than I can recall*), Jude (*the shining presence*), Lei, Guogan, Arif, Helen and many more.

Also there are those who shared my hobbies outside chemistry and provided good times. Alistair, Duncan, Vit, Nic, Thomas, Simon and all the other football players. My golf buddies, Renald, Ruben, William (*en Jacorien, ik kan jullie toch wel in het Nederlands bedanken voor de gezellige tijd en geweldige maaltijden?*), Jan-Hendrik (*Mooi weer, Eden Course en een airshow dat waren goede tijden! Al had de Jubilee in de wind toch ook z'n charme, we moeten zeker nog eens de Old course samen afleggen*), Gary and Jürgen (*we might one day actually understand what they are saying about us in Itallian*). Also Marcin, Aga and later surely also Dominik for many nice evenings spent together. And many other for interesting nights out in Dundee and St. Andrews.

Naast al de mensen in Schotland moet ik ook veel mensen van het thuisfront bedanken. Tonnie, Rugé, Hauwert en de rest van the Soundabout voor het blijven creëren van goede muziek. Anouk, Pauline, Ferdinand, Bob, Mara, Patrick en

Matthijs voor de gezellige tijd. Mijn ouders voor de vele bezoeken en voor de uiterst belangrijke steun, natuurlijk ook Tomas (*vooral voor de discussies over hoe slecht Ajax dan weer speelde*) en de rest van de familie met in het bijzonder oma Born.
Finally I have to thank Micio and Penny who kept me company (and distracted) while writing this thesis and a very special thanks for Marzia for almost everything, this work would never have been possible without your continuous love and presence.

Thanks all,

Peter J. Deuss

St. Andrews, 04-04-2011