<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>Polyhedral features of cysteine: from protein functionalization to copper-binding</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Author(s)</strong></td>
<td>Cattani, Giada</td>
</tr>
<tr>
<td><strong>Publication Date</strong></td>
<td>2015-03-30</td>
</tr>
<tr>
<td><strong>Item record</strong></td>
<td><a href="http://hdl.handle.net/10379/4933">http://hdl.handle.net/10379/4933</a></td>
</tr>
</tbody>
</table>

Some rights reserved. For more information, please see the item record link above.
Polyhedral Features of Cysteine: from Protein Functionalization to Copper-binding

Giada Cattani
Polyhedric Features of Cysteine: from Protein Functionalization to Copper-binding

Thesis submitted in fulfilment of the Degree of Doctor of Philosophy (PhD) at the School of Chemistry

I declare that the work included in this thesis is entirely my own work and has not been previously submitted for a degree to this or any other academic institution

Giada Cattani, M.Sc.
ID No. 10233009
Born in Rome in 1985
Cover: Antiparallel double-helical assembly of PEGylated Plastocyanin (PDB ID 4R0O).
Summary

This thesis comprises two main projects which are about the cysteine chemistry involved in protein functionalization and copper-binding.

Plastocyanin functionalization via cysteine has been used to produce a conjugate for the aim of obtaining structural information of a self-assembled state directed by the conjugation. This aim was achieved with a PEGylated Plastocyanin by using NMR spectroscopy and X-ray crystallography. The first crystal structure of a PEGylated protein has been solved at a resolution of 4.2 Å. The electron density of the PEG chain is not visible suggesting a highly disordered structure, but the crystal packing is remarkable: a right-handed antiparallel double-helical assembly was observed.

In biological systems cysteine can be coordinated to copper, an essential trace element which is associated with neurodegenerative disorders and therefore its trafficking is regulated in cells. The preliminary NMR study of a novel cysteine rich copper-binding protein found by the group of Prof. Dennison in Newcastle University (UK) is presented. $^{15}$N-, $^{13}$C$^{15}$N- and $^{15}$N-selective amino acid labelled protein samples were produced to assign the backbone resonances and to study the binding of copper.
# Table of contents

List of abbreviations 7

1 - Introduction 9
   1.1 - Biological cysteine chemistry
   1.2 - Cysteine chemistry in protein nanostructures
   1.3 - Cysteine in copper-binding proteins
   1.4 - Outline of this thesis

2 - Covalent Modification of Solvent Exposed Cysteines 23
   2.1 - Introduction
   2.2 - Results and discussion
      2.2.1 - Optimization of protein production
      2.2.2 - Covalent modification of Plastocyanin D45C
      2.2.3 - Covalent modification of Plastocyanin P37C
      2.2.4 - Crystal structure of maleimide-functionalized Pc
   2.3 - Conclusions
   2.4 - Experimental section

3 - Structural Characterization of PEGylated Plastocyanin 65
   3.1 - Introduction
   3.2 - Results and discussion
      3.2.1 - NMR analysis
      3.2.2 - Crystallization conditions
      3.2.3 - Structure and assembly
   3.3 - Conclusions
   3.4 - Experimental section

4 - Preliminary Characterization of a Cysteine Rich Copper-binding Protein 95
   4.1 - Introduction
   4.2 - Results and discussion
      4.2.1 - Protein production for NMR analysis
      4.2.2 - Assignment of the apo-Csp1
      4.2.3 - Selective amino acid labelling
      4.2.4 - Preliminary study of copper-binding
   4.3 - Conclusions
   4.4 - Experimental section

5 - Discussion 123

Bibliography 131

Acknowledgements

Curriculum Vitae
List of abbreviations

ATP  adenosine triphosphate
BME  β-mercaptoethanol
BSA  bovine serum albumin
Csp1  cysteine rich copper-binding protein
Cys  cysteine
DCC  N,N'-dicyclohexylcarbodiimide
DCU  dicyclohexylurea
DEAE  diethylaminoethyl
DTT  dithiothreitol
DYANA dynamics algorithm for NMR applications
EDTA  ethylenediaminetetraacetic acid
ESI-TOF electrospray ionization time-of-flight
FPLC fast protein liquid chromatography
IPTG isopropyl-β-D-thiogalactopyranoside
LB  Luria Bertani medium
MM  minimal medium
MS  mass spectrometry
Mw  molecular weight
MWCO molecular weight cut-off
NMR  nuclear magnetic resonance
OD_{600} optical density measured at 600 nm
Pc  Plastocyanin
PEG  polyethylene glycol
PNIPAm poly(N-isopropylacrylamide)
Q  quaternary amine
SD  standard deviation
SDS-PAGE sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEC  size exclusion chromatography
Chapter 1

Introduction
This thesis comprises two main projects, both involved with cysteine chemistry. The first project is about cysteine as the route to functionalize proteins. The second project is about a novel cysteine rich copper-binding protein. The introduction chapter is guided by the desire to present an overview of the cysteine chemistry. This chapter will begin with a general description of the biological chemistry of cysteine and the focus will be on the post-translational modification of proteins via cysteine. Then, the application of cysteine chemistry in the development of protein nanostructures will be described. The last part of the introduction will be about copper binding proteins.

1.1 - Biological cysteine chemistry

Cysteine is a rare amino acid in proteins and its content increases with the complexity of the organism, reaching about 2% in intracellular mammalian proteins.\textsuperscript{1-3} Two previous studies\textsuperscript{2,3} analysed the cysteine occurrence and conservation during evolution. The first study\textsuperscript{2} based on protein sequences stored in the SwissProt database revealed that cysteine does not occur randomly, but it is preferentially organized into a Cys-X-Y-Cys motif (where X and Y stand for any amino acid). This motif will be discussed again at the end of this chapter, but here it has to be emphasized that this motif is associated with the biological cysteine chemistry. In fact, the two cysteines are found closest to each other in the tertiary structure allowed the binding of metals or the formation of a disulfide bond. This structural aspect was analysed in the second study\textsuperscript{3} by using protein structures from two structural databases (PDB and ModBase). This study related the cysteine conservation to two structure-based features, the clustering and the exposure of cysteine residues in the protein structure. It was found that the first feature was more conserved than the second one. This trend is relevant to the cysteine reactivity and hence to the biological functions of the cysteine. In fact, this residue has
different functions such as to protect the cell against oxidative stress, to stabilize the folded state of protein through cystine formation and to participate in electron transfer reaction.\textsuperscript{4-12} It should be mentioned, however, that in some cases a cysteine can cover more than one function. The description of all biological functions of cysteine listed above are beyond the scope of this section, which intends to describe three examples of covalent post-translational modifications of cysteine as discussed in more detail later. It is worth noting that not all biological cysteines are susceptible to covalent modifications which are determined mostly by the accessibility of the sulphydryl group within the tertiary structure of the protein and by the reactivity of the cysteine. The first determinant depends on the solvent exposed position of the cysteine and the second determinant depends on the influence of the neighbouring positively charged amino acids which stabilize the reactive thiolate anion.

The previous paragraph presented an overview of biological cysteine chemistry. In the following paragraph, the focus is on the covalent modifications of solvent exposed cysteine that occurs in the organisms. For this purpose, three examples have been chosen to serve as highlights. These reactions, shown in Figure 1, are S-glutathionylation, S-palmitoylation and S-farnesylation. An example of the effect of each modification on the modulation of protein-protein interactions will be described.

Glutathionylation is a post-translational modification of cysteine by the tripeptide (glutamate-cysteine-glycine) glutathione. This peptide can be reversibly bound to the sulphydryl of a cysteine residue via a thiol-disulfide exchange reaction (Figure 1A). The reversible protein-S-glutathionylation has the central roles in modulating protein interactions and in protecting proteins from irreversible cysteine oxidation, such as the formation of sulfenic or sulfonic acids.\textsuperscript{13-16} For example, glutathionylation modulates the actin polymerization.\textsuperscript{17,18} Actin, an ATP-
binding protein, is covalently modified by glutathione at a cysteine in the C-terminus leading to a decreased capacity to polymerize with respect to the unmodified actin. Spectrofluorometric analysis\textsuperscript{18} revealed that the glutathionylation regulates actin polymerization by inducing structural changes. An increase exposure of hydrophobic clusters and an increase rate of ATP exchange of the glutathionylated actin indicated changes of the actin conformation upon glutathionylation.

The modulation of actin polymerization by covalent modification of the cysteine was also reported with other small molecules, like lipids.\textsuperscript{19-21} Protein modification with lipids will be the subject of the following two examples of post-translational modification of biological cysteine, \textit{i.e.} S-palmitoylation and S-farnesylation.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure1.png}
\caption{Post-translational modifications of cysteine. (A) S-glutathionylation through glutathione (GSH), (B) S-palmitoylation through the activated form of palmitic acid by coenzyme A (CoA) and (C) S-farnesylation through the farnesyl pyrophosphate (PPi).}
\end{figure}

Palmitoylation is a reversible post-translational modification in which a cysteine is covalently modified by the palmitic acid, a saturated fatty acid. This reaction occurs as a nucleophilic substitution reaction after the activation of the fatty acid with the coenzyme A (Figure 1B). Four consequences are attributed to the S-palmitoylation: the conformational change of membrane proteins, the positive or negative regulation of protein association to the membrane, the promotion of
protein-protein complex formation and the prevention of other post-translational modifications such as ubiquitination.\textsuperscript{22-26} For example, the palmitoylation of tetraspanins, a family of membrane proteins, leads to a more stable structure of the tetraspanin network assembly.\textsuperscript{27-31} In general, palmitoylation contributes to tetraspanin-tetraspanin associations which are increased by the presence of divalent cations (such as Ca\textsuperscript{2+} and Mg\textsuperscript{2+}). A radiolabeling method using \textsuperscript{3}Hpalmitate was used to identify the palmitoylation sites and to prove the structure stabilization of the palmitoylated tetraspanin assembly.\textsuperscript{27} This study revealed that tetraspanins are palmitoylated at the juxtamembrane cysteines and that the interaction between palmitoylated tetraspanins remained stable in the presence of the metal-binding EDTA. Moreover, this study showed that the tetraspanin-tetraspanin association was dependent on palmitoylation because it was lost upon mutation of palmitoylation sites.

Farnesylation is a type of prenylation, an irreversible post-translational modification of cysteine by farnesyl pyrophosphate. This nucleophilic substitution reaction (Figure 1C) leads to the attachment of the farnesyl group, an unsaturated aliphatic chain, to the sulfhydryl group of the cysteine residue at the C-terminus. Although this modification is essential for the association of the protein to the cellular membranes, it has been found that farnesylation is important for an efficient protein-protein interaction.\textsuperscript{32,33} For example, the farnesylated peroxin Pex19p is essential in peroxisome biogenesis.\textsuperscript{34,35} Pex19p is a protein found in peroxisomes which are organelles present in eukaryotic cells. Peptide blots, two-hybrid analysis and fluorescence polarization titration showed that farnesylated Pex19p interacts with all peroxisomal membrane proteins (PMPs) more efficiently than the non-farnesylated Pex19p.\textsuperscript{35} This study was coupled with size exclusion and anion exchange chromatography analysis. The changes in chromatogram profiles between the farnesylated and the non-farnesylated Pex19p indicated that the farnesylation induced structural changes of Pex19p. This result
suggested that the efficient interaction between the farnesylated Pex19p and the PMPs is achieved through a conformational change in Pex19p upon farnesylation rather than through a direct interaction between the farnesyl group and the PMPs.

1.2 - Cysteine chemistry in protein nanostructures

Cysteine chemistry has been used to develop protein nanostructures as discussed in this section. Even if 90% of cysteine residues are buried, protein engineering has allowed the construction of proteins with solvent exposed cysteines in order to allow disulfide bonds or chemical modifications. Disulfide bond formation has been used to crystalize new packing arrangements or to cross-link proteins to form supramolecular assemblies also assisted by metal-binding. Chemical modification of proteins has not only been used to design protein nanostructures but also for other different purposes listed below. In this section, four examples of self-assembling protein systems will be described to serve as highlights.

Chemical modification of proteins is an expanding area in chemical biology. The amino acid cysteine is the most convenient target for selective modification owing to the strongly nucleophilic sulfhydryl side chain. Small molecules functionalized with thiol-reactive groups such as maleimide, iodoacetamide and sulfide have been widely used to covalently modify cysteine residues. These studies were focused on different purposes such as to study oligomerization process, to regulate the catalytic activity of an enzyme, to monitor conformational changes of a protein or to engineer supramolecular assembly. Among these purposes, the assembly of proteins has become of great interest because protein self-assembly is complementary to protein aggregation, which is relevant for the production of biopharmaceuticals. The self-assembly process is mediated via non-covalent intermolecular interactions such as
the hydrophobic effect, charge-charge interactions and hydrogen bonding. Controlled protein-protein assembly is used to study, understand and influence biological processes as well as to design new materials for nano-biotechnology. It has been demonstrated that proteins can be used in various ways to engineer complex nanostructures for different applications such as molecular electronics, tissue engineering and drug delivery. Next paragraphs will present the utilization of cysteine chemistry in self-assembling protein systems. Four examples of protein nanostructures will be described. The first involves the disulfide bond between proteins and the other three are about the protein functionalization via cysteine.

Disulfide bond formation has been applied to build and stabilize filamentous forms composed by either homo- or hetero-oligomers. For example, bacterial flagellin modified with a cysteine-containing loop was used to form self-assembling nanotube bundles. The flagellin forms naturally self-assembling nanotubes whereas the added cysteine-containing loop was used to bundle the nanotubes through disulfide bonds. Transmission electron and fluorescence microscopy allowed the visualization of this nanostructure. The flagellin nanotube bundles were assembled upon oxidation in air and disassembled upon addition of a reducing agent such as DTT.

Protein functionalization via cysteine has been applied to modify natural cages in self-assembling protein systems to produce a closed space used as a chemical reactor or container. For example, the cage of a heteroprotein assembly from bacteriophage T4 provided a catalytic reaction space (Figure 2). An engineered cysteine was covalently modified with an iron (III) protoporphyrin maleimide derivative. The introduced molecule did not change the protein structure assembly, but the catalytic reaction rate of the iron porphyrin-protein assembling system was six times greater than that of the free iron porphyrin maleimide in the same experimental conditions.
Protein functionalization with an iodoacetamide derivative of phenanthroline via cysteine has been used to construct assembled protein structures.\textsuperscript{74,75} Cytochrome $c_{b562}$ functionalized with phenanthroline iodoacetamide has been crystallized in the absence and in the presence of metals.\textsuperscript{75} In the three crystal structures, protein dimers were mediated solely by two phenanthroline groups or by the coordination of Ni(II) or Zn(II) with the two groups. In the absence of metals, two phenanthroline groups were buried in a loop region of the protein and formed $\pi$-stacking interactions with each other. In the presence of metals, one group was in the buried position and the other was in an extended conformation giving rise to porous crystal assemblies characterized by hexagonal channels.

Protein functionalization with synthetic polymers via cysteine has been used to direct protein assembly.\textsuperscript{76-78} mCherry, a fluorescent protein, was covalently modified at a solvent exposed cysteine with a maleimide derivative of PNIPAm, a thermoresponsive polymer.\textsuperscript{77} The nanostructure formation of the protein-polymer conjugate was controlled by altering pH and temperature. The resulting assembly was analysed by using biophysical techniques such as small-angle X-ray scattering and transmission electron microscopy. These measurements revealed the formation of nanostructures in a variety of morphologies such as micellar, cylindrical and lamellar forms.
Figure 2 Crystal structure (PDB ID 2Z6B) of protein assembly from bacteriophage T4. The natural cage (in the box) provided a catalytic reaction space. Ribbon representation of protein backbone with cysteines modified with maleimide derivative shown as sticks.71
1.3 - Cysteine in copper-binding proteins

In this section, an overview of biological copper chemistry and copper-binding proteins which coordinate this metal via cysteine will be presented. At the end, two copper-binding proteins involved in the copper regulation in prokaryotic organisms will be described.

Copper is an essential trace element in biological systems and it is primarily available in the cupric form in the extracellular space, while it is maintained in the cuprous form in the intracellular space. Biological copper is coordinated predominantly by the side chains of histidine, cysteine and methionine in agreement with the hard and soft acid/base concept.\textsuperscript{79-82} According to this concept, the cupric form is a borderline ion thus it is primarily coordinated to the imidazole group of histidine side chain, while the cuprous form is a soft ion therefore it is coordinated to the sulfur atoms of cysteine and methionine. Bertini \textit{et al.}\textsuperscript{81} analysed copper sites in proteins with known structures from the PDB database. The copper sites were grouped by the oxidation state of copper and then into broader categories based on structural or functional features. The three major categories are (1) copper sites of proteins associated with the copper transport/storage, (2) copper sites involved in stabilizing the tertiary or quaternary structure of proteins and (3) copper sites of proteins which participate in electron transfer reactions. This study also analysed the occurrence of specific copper-binding residues and each residue was associated to the protein secondary structure. Regarding to the cysteine residue, the percentage of copper-binding cysteine was: in category (1), 97 % of which 52 % in helices and 45 % in loops; in category (2), 80 % in helices; and in category (3), 27 % of which 8 % in sheets and 19 % in loops. For example, the copper sites of metallothionein, transcription factor CueR and plastocyanin, are within the three categories listed above, respectively. These proteins bind
copper prevalently via cysteine residues and the copper sites are shown in Figure 3.

Figure 3 Copper sites of (A) metallothionein, (B) transcription factor CueR and (C) plastocyanin. Ribbon representation of protein backbone with binding residues and copper ions shown as sticks and sphere, respectively. The PDB ID are 1RJU, 1Q05 and 2W88, respectively.

The previous paragraph summarized the key points of the copper chemistry and the coordination of this metal by cysteine. In the following paragraphs, two copper-binding proteins will be described. These proteins are involved in the copper regulation whose acquisition by bacterial cells is not yet clear. Scheme 1 illustrates a summary of copper regulation in prokaryotic organisms. In the intracellular space copper ions are coordinated to methanobactins, which have a high affinity for copper and reduce the toxicity of this metal. Alternatively the copper can bind different proteins called chaperones which have the primary role of transporting this metal to the target proteins such as the cytochrome c oxidase. P-type ATPase proteins are copper transporters that have the crucial role in eliminating excess copper. The evolution to eukaryotic organisms led to the development of more complex processes to regulate copper homeostasis. This is important to avoid high levels of copper which result in oxidative stress and neurodegenerative diseases. The detailed description of the copper trafficking is beyond the scope of this section, which intends to describe the structure of two bacterial copper-
binding proteins, methanobactin and chaperone. This is reported in the next two paragraphs.

![Scheme 1](image)

**Scheme 1** Copper regulation in prokaryotic organisms.

The methane-oxidizing bacterium *Methylosinus trichosporium* OB3b secretes two forms of methanobactins.\(^95-97\) Figure 4 illustrates the crystal structures of these proteins.\(^97\) One has a missing methionine residue in the C-terminus that has a small influence on the structure of the copper site. The copper ion is coordinated by the sulfurs of two enethiolate groups and by the nitrogens of two oxazolone rings in a distorted tetrahedral geometry. The molecule has a compact structure stabilized by a disulfide bridge, the reduction of which results in a decrease in the copper affinity.

![Figure 4](image)

**Figure 4** Crystal structure of methanobactins from *Methylosinus trichosporium* OB3b (A) with and (B) without the C-terminal methionine residue. The protein backbone is represented as sticks and the copper as a sphere. The PDB IDs are 2XJH and 2XJI, respectively.\(^97\)
A common feature of copper chaperones is the conserved Cys-X-X-Cys metal-binding motif on a loop region that binds copper via the thiols of the cysteine side chains.\textsuperscript{98,99} For example, the crystal structure of \textit{Synechocystis} chaperone Atx1\textsuperscript{98} showed that it can bind up to two equivalents of copper and adopt at least two dimeric conformations when metallated in agreement with NMR studies. The crystal structures of head-to-head and side-to-side dimers in the presence of two equivalents of copper are reported in Figure 5. A histidine residue on a loop region in the head-to-head dimer coordinates the copper and stabilizes the dimeric form while a chloride ligand is present in the side-to-side dimer which is hydrogen bonded by the histidine.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure5.png}
\caption{Crystal structures of Atx1 in the presence of two equiv of copper forming (A) head-to-head and (B) side-to-side dimers. A histidine is involved in copper (brown sphere) coordination directly and through a chloride ion (green sphere), respectively. Ribbon representation of protein backbone with binding residues shown as sticks. The PDB IDs are 2XMU and 2XMJ, respectively.\textsuperscript{98}}
\end{figure}
Chapter 1 - Introduction

1.4 - Outline of this thesis

Chapter 2 describes the production of two Plastocyanin mutants with an engineered surface cysteine residue. The conjugation reaction with different thiol reactive compounds is described together with the mass spectrometry and size exclusion chromatography analysis of the conjugated proteins.

Chapter 3 reports the structural characterization of PEGylated Plastocyanin. For this purpose, the NMR spectroscopy and X-ray crystallography were used. The crystal structure revealed a supramolecular self-assembled state in which the PEGylated protein adopts an antiparallel double-helical assembly. An NMR study under physiological conditions is also presented.

In Chapter 4, the preliminary NMR characterization of a novel copper-binding protein potentially involved in storing copper is described. The group of Prof. Dennison in Newcastle University have found this protein in the methane-oxidising bacterium *Methylosinus trichosporium* OB3b. It was called Csp1 and is presented herein.

In Chapter 5, the results of this thesis will be evaluated and the two main projects will be examined in the light of previous studies.

The goals of this thesis were to obtain structural information of a new self-assembled state of Plastocyanin directed by the cysteine conjugation on the protein surface and to study the binding of copper by Csp1 via cysteine. X-ray crystallography and NMR spectroscopy were central to this thesis.
Chapter 2

Covalent Modification of Solvent Exposed Cysteines
Abstract

The cysteine sulfhydryl group is the stronger nucleophile in a protein and this feature has been used to functionalize proteins. Here, two Plastocyanin (Pc) mutants were used. Firstly, the optimization of Pc purification was performed. Secondly, the covalent modification of the cysteine was tested with iodoacetic acid, two iodoacetamide compounds, four maleimide derivatives, an aziridine and a thiol containing reagents. The chemically modified forms of Pc were analysed by using ESI-MS. The masses are given in each analysis and they agree with the expected values. Among the tested reactions, the maleimide derivatives gave the better results. Some biophysical characterizations of functionalized Pc were also performed by using SEC. This analysis revealed interactions with the resin of the column and protein-protein aggregation. Finally, the objective of these covalent modifications was to obtain structural characterizations of functionalized Pc by using X-ray crystallography. In this chapter, the crystal structure of conjugated Pc with ethylphenyl maleimide will be briefly described. This crystal revealed the same crystal packing of the wild-type Pc crystal.
Chapter 2 - Covalent modification of solvent exposed cysteines

2.1 - Introduction

In this chapter, Plastocyanin (Pc) functionalization via cysteine will be described. Pc was chosen because it is a useful model protein for structural biology as presented in the second part of this introduction. After an overview of cysteine-reactive compounds, two assembled states of the 11 kDa Pc will be discussed.

Iodoacetamide and N-ethylmaleimide have been commonly used to covalently modify cysteine residues. However, other compounds are available such as aziridine and thiol containing reagents. These reactions have been extensively reviewed and the main characteristics of each are reported herein (Figure 1).\textsuperscript{100-103} Iodoacetamide and N-ethylmaleimide form covalent bonds with sulphydryls through nucleophilic substitution (S_N2) and Michael addition, respectively. These reactions are specific for sulphydryls when the modifying agents are used at pH 6.5-7.5 and in limiting quantities relative to the number of cysteine residues. The conjugation reaction with aziridine is a ring-opening process. The aziridine is specific for sulphydryls at slightly alkaline pH, however in aqueous solution the main side reaction is hydrolysis. The reaction with thiol reagents generates disulfide bonds over a broad range of pH. Among these reactions, the covalent modification with iodoacetamide and N-ethylmaleimide (and their derivatives) has been used extensively in a wide range of applications listed below.\textsuperscript{100,101} Targeting proteins \textit{in vivo} and \textit{in vitro}, modulation of biological processes, investigation of protein structure and function, discovery of active sites or study of protein-ligand interactions are some of these applications. For example, the effect of the chemical modification of cysteine residues on the activity of arginyl-tRNA synthetase was investigated by using iodoacetamide and N-ethylmaleimide.\textsuperscript{104} Iodoacetamide reacted with only one of the two solvent exposed cysteine residues with a 50 % loss of
activity, whereas N-ethylmaleimide reacted with both cysteines with more than 90% loss of activity.

![Chemical reactions](Image)

**Figure 1** Reactions of cysteine residue with (A) iodoacetamide, (B) maleimide, (C) aziridine and (C) thiol containing reagents.

In addition to these applications, covalent modification of the protein surface via solvent exposed cysteine residues is used increasingly to direct protein assembly. This was described in Chapter 1. In the following paragraphs, two systems developed to investigate protein assembly of Pc will be presented. These assembled states were achieved by using different strategies.

*Populus nigra* Pc has been assembled on gold surface via a disulfide (Cys21-Cys25) or a C-terminal thiol group and analysed by using several biophysical techniques such as scanning tunnelling microscopy and conductive atomic force microscopy. Reproducible images of Pc mutants self-assembled on gold surfaces were
obtained under various environmental conditions and the Pc dimensions agreed with the crystallographic structure.

*Phormidium laminosum* Pc has been crystallized in five different self-assembled states.\(^8\),\(^9\) The crystallization conditions influenced the protein-protein interactions and gave rise to substantially different crystal packing environments. Three crystal forms of the L14F mutant were obtained using the same precipitant (polyethylene glycol, PEG) with different metal ions (magnesium and zinc). Conformational changes and favorable intermolecular polar interactions facilitated the packing in these three structures. Two crystal forms of the Pc variant with an N-terminal methionine were obtained using zinc acetate in the presence or absence of PEG. Zinc ions were found at protein-protein interface, while the presence or absence of PEG led to the “closed” and “open” self-assembled Pc structures.

In this project, two *Phormidium laminosum* Pc mutants were covalently modified on the protein surface. Residues proline 37 (P37) and aspartic acid 45 (D45) were mutated to cysteine (C) to yield Pc P37C and Pc D45C, respectively. These cysteines are exposed to solvent at two different surface environments, a hydrophobic patch and a charge rich patch, respectively (Figure 2). These patches define a region of apolar residues around the copper binding site and a region formed by residues 44-47, respectively.
Chapter 2 - Covalent modification of solvent exposed cysteines

Figure 2 Electrostatic surface representation of *Phormidium laminosum* Plastocynin (PDB ID 2W88). The red, blue and white surfaces correspond to the anionic, cationic and uncharged residues, respectively. The position of the mutated residues is indicated. The figure was generated in PyMol (http://www.pymol.org).

The conjugation reactions of Pc mutants at the solvent exposed cysteines were performed with in-house synthesized and commercially available reagents (Scheme 1). Iodoacetic acid was used to assess the effective availability of cysteine prior to performing the conjugation reactions with other compounds. These reagents are two iodoacetamide derivatives (2 and 3), an aziridine derivative (4), a thiol compound (5) and four maleimide derivatives (6-9). The compounds are characterized by rigid (2, 4-7) and flexible (3, 8 and 9) substituents with hydrophobic (2-7) and hydrophilic (8 and 9) nature. These features are expected to enhance the possibilities for molecular recognition.\textsuperscript{111} For example, Arkin *et al.*\textsuperscript{112} found that a small molecule binds the interleukin-2 receptor on the protein surface at two subsites, which are distinguished by flexible and rigid conformations, respectively. The flexible region interacts with the hydrophobic portion of the ligand, whereas the rigid one interacts with the hydrophilic group of the ligand.
Here, the focus is on the chemical modifications of the solvent exposed cysteine in the Plastocyanin mutants. These reactions were performed to produce a conjugate for the second task, \textit{i.e.} to obtain structural information of a self-assembled state directed by the covalent modification of the Plastocyanin surface. This was achieved with the PEGylated Pc D45C mutant and it is described in Chapter 3.

\textbf{Scheme 1} Compounds used for covalent modification of Plastocyanin mutants: (1) iodoacetic acid; (2) 2-iodo-\textit{N}-(2-anthracenyl)acetamide; (3) \textit{N},\textit{N}-dihexyl-2-iodoacetamide; (4) 6-(N-aziridinyl)-2,3-dihydro-1H-pyrrolo[1,2-\textit{a}]benzimidazole-5,8-dione; (5) 2-thiobarbituric acid; (6) \textit{N}-(1-pyrenyl)maleimide; (7) \textit{(S)}-(−)-\textit{N}-(1-ethylphenyl)maleimide; (8) poly(\textit{N}-isopropylacrylamide) maleimide, 2 kDa; (9) methoxypolyethylene glycol maleimide, 5 kDa.
2.2 - Results and discussion

2.2.1 - Optimization of protein production

The preparation of high-purity protein samples is a skill of central importance to protein research. However, surface cysteines present significant challenges. For example, the initial purification attempts revealed a post-translational modification with glutathione, a thiol compound present in the cell extract. This result necessitated an optimization of the purification steps to obtain the target proteins.

Figure 3A illustrates the first purification step of Pc P37C by anion exchange chromatography using a DEAE column. This step was performed with a linear gradient of 0-100 mM NaCl and the separation was assessed by SDS-PAGE. This gel showed two major bands at ~11 and ~26 kDa in the impure fractions of DEAE. These proteins were separated on a SEC column equilibrated with 20 mM KH$_2$PO$_4$, 100 mM NaCl, pH 6.0. This is shown in Figure 3B. Two peak maxima were observed in the SEC chromatogram with elution volumes of ~68 and ~80 mL, respectively. The first peak corresponds to particles with Mw > 17 kDa and the second peak to particles with Mw < 17 kDa, according to the previously reported calibration curve. The pooled SEC fractions 65-70 and 78-85 were analysed by SDS-PAGE and ESI-MS. Before running the SDS-PAGE gel, a portion (3 mL) of the fractions 65-70 was incubated in 10 mM DTT at room temperature for two hours. This was done because it was initially assumed that the protein eluted at ~68 mL was the dimeric form (23 kDa) of Pc via disulfide bond. The SDS-PAGE gel showed a major band at ~26 kDa in the fractions 65-70, both in the presence and in the absence of DTT, and a major band at ~11 kDa in the fractions 78-85. This analysis revealed that the protein eluted at ~68 mL was not the Pc dimer and that the Pc monomer was eluted at ~80 mL, as expected for an 11.5 kDa protein. ESI-MS analysis gave a further confirmation. A major peak was obtained in the deconvoluted mass
Chapter 2 - Covalent modification of solvent exposed cysteines

spectrum (Figure 4A) of the SEC fractions 65-70. The experimental mass (28,908 Da) corresponded to the mass of β-lactamase (28,907 Da), an enzyme carried on the plasmid for Pc that provide resistance to β-lactam antibiotics like ampicillin. Two major peaks with a relative abundance of 100 and 30 % respectively were observed in the deconvoluted mass spectrum (Figure 4B) of the SEC fractions 78-85. The experimental mass of the first peak (11,556.00 Da) was in agreement with the predicted mass of Pc P37C (11,557.09 Da). The second peak (11,863.00 Da) suggested the post-translational modification of Pc with glutathione (11,862.40 Da).

The Pc D45C mutant gave similar results to those reported for Pc P37C. Figure 3C shows the SEC chromatogram of pooled fractions from the DEAE purification in 20 mM KH$_2$PO$_4$, 100 mM NaCl, pH 6.0. Two major peaks were observed in the SEC chromatogram with elution volumes of ~68 and ~80 mL, respectively. The pooled SEC fractions 66-69 and 78-81 were analysed by ESI-MS. A major peak was observed in the deconvoluted mass spectra (Figure 5) of both pooled SEC fractions. The experimental mass (28,906 Da) of the SEC fractions 66-69 was in agreement with the mass of β-lactamase (28,907 Da). The experimental mass (11,844.89 Da) of the SEC fractions 78-81 suggested the post-translational modification of Pc D45C with glutathione (11,844.43 Da).
Figure 3 Purification of Pc before optimization. (A) DEAE chromatogram of cell lysate containing Pc P37C eluted with a linear NaCl gradient (0-100 mM) and the SDS-PAGE gel of the DEAE fractions. (B) SEC chromatogram of the DEAE fractions containing Pc P37C eluted in 20 mM KH₂PO₄, 100 mM NaCl, pH 6.0 and the SDS-PAGE gel of the pooled SEC fractions 65-70 (lanes 1 and 2) and 78-85 (lane 3). Sample in lane 1 was incubated in 10 mM DTT for 2 hours at r.t. before loading to reduce the presumed Pc dimer. (C) SEC chromatogram of the DEAE fractions containing Pc D45C, the elution buffer was 20 mM KH₂PO₄, 100 mM NaCl, pH 6.0.
Figure 4 Deconvoluted mass spectra of the pooled SEC fractions (A) 65-70 and (B) 78-85 before optimization of purification steps of Pc P37C. The SEC chromatogram is shown in Figure 3B. The mass 28,908 Da in (A) correspond to β-lactamase (28,907 Da). The masses 11,556.00 and 11,863.00 Da in (B) correspond to the predicted mass of Pc P37C monomer (11,557.09 Da) and covalently modified with glutathione (11,862.40 Da).
Figure 5 Deconvoluted mass spectra of the pooled SEC fractions (A) 66-69 and (B) 78-81 before optimization of purification steps of Pc D45C. The SEC chromatogram is shown in Figure 3C. The mass 28,906 Da in (A) correspond to β-lactamase (28,907 Da). The mass 11,844.89 Da in (B) correspond to the predicted mass of Pc D45C covalently modified with glutathione (11,844.43 Da).
Optimization of purification steps was carried out on the P37C mutant and then the same procedures were applied to the D45C mutant. Figure 6A shows the DEAE purification step of Pc P37C using a linear gradient of 0-50 mM NaCl. The pooled DEAE fractions were loaded on a SDS-PAGE gel which revealed a major band at ~11 kDa corresponding to the Pc monomer (11.5 kDa). The band at ~26 kDa (corresponding to the 28.9 kDa β-lactamase) was not present. Therefore, a linear gradient of 0-50 mM NaCl on anion exchange chromatography was adopted to improve the first purification step. The pooled DEAE fractions were loaded on the SEC column equilibrated with 20 mM KH$_2$PO$_4$, 100 mM NaCl, pH 6.0. Two peak maxima were observed in the SEC chromatogram (Figure 6B) with elution volumes of ~67 and ~80 mL, respectively. The fractions 67 and 80 were loaded on a SDS-PAGE gel which revealed the band corresponding to Pc at ~11 kDa in both fractions. The higher elution volume (~67 mL) for the low molecular weight Pc (11.5 kDa) suggested the dimerization of Pc P37C via disulfide bond, probably caused to oxidation in air. Therefore, in the subsequent purifications of Pc the DTT was added to the SEC elution buffer to obtain the monomeric form and to avoid glutathionylation.
Figure 6 Optimization process of PcP37C purification. (A) DEAE chromatogram of cell lysate containing Pc and SDS-PAGE of the pooled DEAE fractions. Pc eluted with a linear NaCl gradient (0-50 mM). (B) SEC chromatogram of the pooled DEAE fractions and SDS-PAGE of the SEC fractions 67 (lanes 1) and 80 (lane 2). The SEC elution buffer was 20 mM KH$_2$PO$_4$, 100 mM NaCl, pH 6.0.
The purification of the Pc mutants following the reported optimization process gave a yield of ~10 mg per litre of culture determined by UV-vis spectroscopy assuming an extinction coefficient of $\varepsilon_{598} = 4.3 \text{ mM}^{-1}\text{cm}^{-1}$. Figure 7 shows the SEC chromatograms and the SDS-PAGE gels for both mutants. In both chromatograms a peak maximum was observed with elution volume of ~80 mL as expected for the 11.5 kDa Pc. The pooled SEC fractions were pure as judged by SDS-PAGE which shows a single band at ~11 kDa. ESI-MS analysis confirmed the presence of Pc monomer as the most abundant species. The experimental masses in the deconvoluted mass spectra (Figure 8) were 11,558.04 Da for Pc P37C (predicted mass: 11,557.09 Da) and 11,539.98 Da for Pc D45C (predicted mass: 11,539.12 Da).

**Figure 7** SEC chromatograms and SDS-PAGE gels of pooled SEC fractions of pure (A) Pc P37C (1.5 mM) and (B) Pc D45C (1.5 mM). The SEC elution buffer was 20 mM KH$_2$PO$_4$, 100 mM NaCl, 3 mM DTT, pH 6.0.
Figure 8 Deconvoluted mass spectra of (A) Pc P37C and (B) Pc D45C after optimization processes. The experimental masses 11,558.04 and 11,539.98 Da are in agreement with the predicted masses of Pc P37C (11,557.09 Da) and Pc D45C (11,539.12 Da), respectively.
The pure Pc mutants were reacted with the compounds shown in Scheme 1 under different conditions (see experimental section) after the reduction of the cysteine residue with DTT. The products were analysed by using ESI-TOF mass spectrometry to verify the Pc functionalization. Samples containing a mixture of Pc and functionalized Pc were observed by ESI-MS analysis. The relative abundance of each species in a mixture was estimated by the relative peak height from the deconvoluted mass spectrum according to previous studies.\textsuperscript{116,117} In the next two sections, the functionalization results for Pc D45C and Pc P37C are reported.

2.2.2 - Covalent modification of Plastocyanin D45C

The ESI-MS data of the conjugation products of Pc D45C are summarized in Table 1 and they are described below.

<table>
<thead>
<tr>
<th>Conjugate</th>
<th>Mw\textsubscript{calc} (Da)</th>
<th>Peak</th>
<th>Relative abundance (%)</th>
<th>Mw\textsubscript{exp} (Da)</th>
<th>Difference (Da)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pc-(1)</td>
<td>11,597.16</td>
<td>1</td>
<td>100</td>
<td>11,597.47</td>
<td>0.31</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>30</td>
<td>11,539.01</td>
<td>-58.15\textsuperscript{a}</td>
</tr>
<tr>
<td>Pc-(2)</td>
<td>11,772.39</td>
<td>1</td>
<td>100</td>
<td>11,539.25</td>
<td>-233.14\textsuperscript{a}</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>20</td>
<td>11,771.93</td>
<td>-0.46</td>
</tr>
<tr>
<td>Pc-(3)</td>
<td>11,764.49</td>
<td>1</td>
<td>100</td>
<td>11,539.15</td>
<td>-225.34\textsuperscript{a}</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>30</td>
<td>11,764.75</td>
<td>0.26</td>
</tr>
<tr>
<td>Pc-(5)</td>
<td>11,681.26</td>
<td>1</td>
<td>100</td>
<td>11,540.07</td>
<td>-141.19\textsuperscript{a}</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>40</td>
<td>11,682.29</td>
<td>1.03</td>
</tr>
<tr>
<td>Pc-(6)</td>
<td>11,836.43</td>
<td>1</td>
<td>100</td>
<td>11,836.61</td>
<td>0.18</td>
</tr>
<tr>
<td>Pc-(7)</td>
<td>11,740.34</td>
<td>1</td>
<td>100</td>
<td>11,739.72</td>
<td>-0.62</td>
</tr>
</tbody>
</table>

\textsuperscript{a} Mass corresponding to that of the modifying agent

Pc D45C conjugation via cysteine was first tested with (1). The deconvoluted mass spectrum (Figure 9) after this reaction showed a major peak with a mass of 11,597.47 Da in agreement with the predicted mass of the conjugated Pc (11,597.16 Da). A second peak with a mass of 11,539.01 Da and a relative abundance of 30 % revealed the presence of
unreacted monomeric Pc (predicted mass: 11,539.12 Da). This analysis proved the formation of the conjugated Pc with iodoacetic acid, however, unreacted monomer was found in low abundance (~20 %).

The reaction with (2) and (3) did not go to completion. Figure 10 show the deconvoluted mass spectra after these reactions. The more abundant species in the deconvoluted spectrum after the reaction with (2) was the Pc monomer. The conjugated Pc was present with a relative abundance of 20 % and its experimental mass (11,771.93 Da) corresponded to the predicted one (11,772.39 Da). The major peak in the deconvoluted mass spectrum after the reaction of Pc with (3) corresponded to the Pc monomer, whereas the conjugated Pc was present with a relative abundance of 30 % with the experimental mass (11,764.75) in agreement with the predicted one (11,764.49 Da). Even if the reactions with (2) and (3) were repeated a second time at a different pH according to literature protocols103 (see experimental section for details), the relative abundance of the conjugated Pc did not change in the deconvoluted mass spectra. The low abundances (~20 %) of conjugated Pc D45C with the iodoacetamide derivatives, compared to the result (~80 %) with the iodoacetic acid, can be explained by the low solubility of the tested iodoacetamide compounds in water. Even if these compounds were solubilized in DMSO, the reaction was performed in an appropriate buffer solution (see experimental section). A successive experiment with a cysteine sample suggested that a hindrance effect of the substituents on the amine group could prevent this conjugation reaction. This will be presented in the section 2.2.3.
Figure 9 Deconvoluted mass spectrum of the product of Pc D45C with iodoacetic acid. The experimental mass 11,597.47 Da is in agreement with the predicted mass of Pc covalently modified with iodoacetic acid (11,597.16 Da).

Figure 11 illustrates the deconvoluted mass spectrum after the reaction of Pc D45C with (5). Two major peaks with a relative abundance of 100 and 40 % respectively were observed in the deconvoluted mass spectrum. The first peak corresponded to the Pc monomer, whereas the experimental mass of the second peak (11,682.29 Da) was in agreement with the predicted mass of the conjugated Pc (11,681.26 Da). This analysis showed that Pc D45C was partially modified with the thiol compound. Despite using a more alkaline pH, the relative abundance of the conjugated Pc did not change in the deconvoluted mass spectrum.
Figure 10 Deconvoluted mass spectra of the products of Pc D45C with (A) aromatic and (B) aliphatic iodoacetamides. The masses 11,539.25 and 11,771.93 Da in (A) correspond to those of Pc monomer (11,539.12 Da) and conjugated Pc (11,772.39 Da), respectively. The masses 11,539.15 and 11,764.75 Da in (B) are in agreement with the predicted masses of Pc monomer and conjugated Pc (11,764.49 Da), respectively.
Chapter 2 - Covalent modification of solvent exposed cysteines

Figure 11 Deconvoluted mass spectrum of the reaction product of Pc D45C with thiol compound. The masses 11,540.07 and 11,682.29 Da correspond to the predicted masses of Pc monomer (11,539.12 Da) and conjugated Pc (11,681.26 Da), respectively.

The conjugation of Pc D45C was tested with the four maleimide derivatives. The deconvoluted mass spectra after the reaction of Pc with (6) and (7) are reported in Figure 12. A major peak was observed in both deconvoluted mass spectra. The experimental masses were 11,836.61 Da for the conjugated Pc with (6) (predicted mass: 11,836.43 Da) and 11,739.72 Da for the conjugated Pc with (7) (predicted mass: 11,740.34 Da). The ESI-MS analysis showed that Pc D45C was covalently modified with both maleimide derivatives. The conjugated products of the reaction of Pc D45C with (8) and (9) were loaded on the SEC column to analyze the size of the functionalized Pc with the polymer chains.
Figure 12 Deconvoluted mass spectra of the reaction products of Pc D45C with (A) pyrenyl and (B) ethylphenyl maleimides. The experimental masses 11,836.61 Da and 11,739.72 Da are in agreement with the predicted masses of the conjugated Pc: 11,836.43 and 11,740.34 Da respectively.
A peak was observed with elution volume of ~75 mL in the SEC chromatogram (Figure 13A) of the reaction product of Pc with (8) as expected for the ~13.5 kDa conjugated Pc. Two peak maxima were observed with elution volumes of ~55 and ~80 mL in the SEC chromatogram (Figure 13B) of the reaction product of Pc with (9). The first peak corresponded to the conjugated Pc D45C and the second one was consistent with the 11.5 kDa Pc monomer. The PEGylated Pc eluted from SEC with an apparent molecular weight of ~44 kDa and migrated slowly on the SDS-PAGE gel (Figure 13C). These features have been reported before for other PEGylated proteins.\textsuperscript{118,119} They are caused by the hydrodynamic properties of PEGylated proteins and by the interaction between SDS and PEG, respectively. The SEC fractions containing the conjugated Pc D45C with (8) and (9) were analysed by using ESI-MS. A large number of fragments due to the presence of the polymer chains were observed in both ESI-MS spectra (Figure 14) compared to the ESI-MS spectrum of Pc monomer.

![Figure 13 SEC chromatograms of Pc D45C reacted with (A) poly(N-isopropylacrylamide) and (B) methoxypolyethylene glycol maleimides (20 mM KH\textsubscript{2}PO\textsubscript{4}, 100 mM NaCl, 3 mM DTT, pH 6.0) (C) The PEGylated Pc migrates slowly on the SDS-PAGE gel due to interactions between SDS and PEG.](image-url)
Figure 14 ESI-MS spectra of Pc D45C reacted with (A) poly(N-isopropylacrylamide) and (B) methoxypolyethylene glycol maleimides. A large number of fragments due to the presence of the polymer chains is observed in both spectra compared to the ESI-MS spectrum of Pc monomer (C).
2.2.3 - Covalent modification of Plastocyanin P37C

The ESI-MS data of the conjugation products of Pc P37C are summarized in Table 2 and they are described below.

**Table 2** ESI-MS analysis of Pc P37C conjugates

<table>
<thead>
<tr>
<th>Conjugate</th>
<th>Mw&lt;sub&gt;calc&lt;/sub&gt; (Da)</th>
<th>Peak</th>
<th>Relative abundance (%)</th>
<th>Mw&lt;sub&gt;exp&lt;/sub&gt; (Da)</th>
<th>Difference (Da)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pc-(1)</td>
<td>11,615.13</td>
<td>1</td>
<td>100</td>
<td>11,615.28</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>20</td>
<td>11,557.26</td>
<td>-57.87&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Pc-(2)</td>
<td>11,790.36</td>
<td>1</td>
<td>100</td>
<td>11,557.28</td>
<td>-233.08&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>20</td>
<td>11,791.07</td>
<td>0.71</td>
</tr>
<tr>
<td>Pc-(3)</td>
<td>11,782.47</td>
<td>1</td>
<td>100</td>
<td>11,556.36</td>
<td>-226.11&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>40</td>
<td>11,782.24</td>
<td>-0.23</td>
</tr>
<tr>
<td>Pc-(4)</td>
<td>11,785.32</td>
<td>1</td>
<td>100</td>
<td>11,785.03</td>
<td>-0.29</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>40</td>
<td>11,557.57</td>
<td>-227.75&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Pc-(6)</td>
<td>11,854.40</td>
<td>1</td>
<td>100</td>
<td>11,855.46</td>
<td>1.06</td>
</tr>
<tr>
<td>Pc-(7)</td>
<td>11,758.32</td>
<td>1</td>
<td>100</td>
<td>11,758.76</td>
<td>0.44</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>40</td>
<td>11,555.98</td>
<td>-202.34&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<sup>a</sup> Mass corresponding to that of the modifying reagent

Pc P37C conjugation via cysteine was first tested with (1). The ESI-MS analysis of the reaction product revealed a major peak in the deconvoluted mass spectrum (Figure 15) with an experimental mass of 11,615.28 Da corresponded to the predicted mass of the conjugated Pc P37C (11,615.13 Da). A second peak with a relative abundance of 20 % suggested the presence of unreacted Pc monomer. This result proved that Pc P37C was covalently modified with iodoacetic acid and ~20 % of unreacted protein was present.

Figure 16 shows the deconvoluted mass spectra after the reaction of Pc P37C with (2) and (3). A major peak was observed in the deconvoluted mass spectrum after the reaction with (2) with an experimental mass in agreement with the predicted mass of Pc monomer. A peak with an experimental mass of 11,791.07 Da and a relative abundance of 20 % confirmed the presence of the covalently modified Pc with the aromatic iodoacetamide (predicted mass 11,790.36 Da). Two
major peaks with a relative abundance of 100 and 40 % respectively were observed in the deconvoluted mass spectrum of Pc reacted with (3). The experimental mass of the first peak corresponded to the Pc monomer. The second mass (11,782.24 Da) was in agreement with the predicted mass of the conjugated Pc (11,782.47 Da). This data showed that also the reaction of Pc P37C with the iodoacetamide derivatives did not go to completion, as observed with the Pc D45C mutant.

Figure 15 Deconvoluted mass spectrum of the reaction product of Pc P37C with iodoacetic acid. The masses 11,615.28 and 11,557.26 Da are in agreement with the predicted masses of Pc covalently modified with iodoacetic acid (11,615.13 Da) and Pc monomer (11,557.09 Da), respectively.
Figure 16 Deconvoluted mass spectra of the reaction products of Pc P37C with (A) aromatic and (B) aliphatic iodoacetamides. The masses 11,557.28 and 11,791.07 Da in (A) correspond to those of Pc monomer (11,557.09 Da) and conjugated Pc (11,790.36 Da), respectively. The masses 11,556.36 and 11,782.24 Da in (B) are in agreement with the predicted masses of Pc monomer and conjugated Pc (11,782.47), respectively.
Figure 17 Deconvoluted mass spectra of fractions at (A) 80 and (B) 100 mL from SEC (shown in box) of conjugated Pc P37C with aliphatic iodoacetamide. The SEC elution buffer was 20 mM KH$_2$PO$_4$, 100 mM NaCl, 3 mM DTT, pH 6.0. The masses 11,556.84 and 11,782.24 Da in (A) and (B) are in agreement with the predicted masses of Pc P37C monomer (11,557.09 Da) and conjugated Pc (11,782.47 Da), respectively.
Figure 18 ESI-MS spectrum of the reaction product of cysteine sample with aliphatic iodoacetamide. The experimental mass is in agreement with the predicted mass of alkylated cysteine (347.236 Da).

The reaction of Pc P37C with (3) was repeated and the product of this reaction was loaded on the SEC column (Figure 17) to analyze the elution profile of the conjugated protein. Two peak maxima were observed with elution volumes of ~80 and ~100 mL, respectively. The first peak corresponds to particles with Mw < 17 kDa and the second to particles with Mw ~2 kDa. The SEC fractions at 80 and 100 mL were analysed by ESI-MS. The deconvoluted mass spectrum (Figure 17A) of the first fraction showed a major peak with experimental mass of 11,556.84 Da corresponding to the predicted mass of Pc monomer. A major peak was observed in the deconvoluted mass spectrum (Figure 17B) of the second fraction with an experimental mass of 11,782.24 Da in agreement with the predicted mass of the conjugated Pc (11,782.44 Da). This data revealed that the SEC fraction at 100 mL contained the conjugated Pc P37C. The late elution volume of the conjugated Pc (11.8 kDa) was indicative of an interaction between the carbon chains of aliphatic iodoacetamide and the hydrophobic resin (composite of cross-linked agarose and dextran) of the column. Compound (3) was also tested with a cysteine sample in the same experimental conditions used for the Pc functionalization. The reaction product was analysed by ESI-MS (Figure 18) revealing an experimental mass of 347.237 Da which corresponded to the predicted mass of the alkylated cysteine (347.236 Da). This result proved that the conjugation of the cysteine sample with the aliphatic iodoacetamide went to completion, supporting the
Chapter 2 - Covalent modification of solvent exposed cysteines

hypothesis that a hindrance effect of the substituents on the amine group could prevent the protein functionalization.

Figure 19 illustrates the deconvoluted mass spectrum after the reaction of Pc P37C with (4). Two major peaks with a relative abundance of 100 and 40 % respectively were observed in the deconvoluted mass spectrum. The experimental mass of the first peak (11,785.03 Da) was in agreement with the predicted mass of the conjugated Pc (11,785.32 Da). The second peak corresponded to the Pc monomer. This data showed that Pc P37C was covalently modified with the aziridine derivative, however, unreacted monomer was found in low abundance (~30 %).

![Figure 19 Deconvoluted mass spectrum of the reaction product of Pc P37C with aziridine derivative. The detected masses 11,557.57 and 11,785.03 Da are in agreement with the predicted masses of Pc monomer (11,557.09 Da) and conjugated Pc (11,785.32 Da), respectively.](image)

Pc P37C conjugation was tested with the four maleimide derivatives. The deconvoluted mass spectra after the reaction of Pc with (6) and (7) are shown in Figure 20. A major peak was observed in the deconvoluted mass spectrum after the reaction with (6). The
Chapter 2 - Covalent modification of solvent exposed cysteines

Experimental mass 11,855.46 Da was in agreement with the predicted mass of the conjugated Pc (11,854.40 Da). The deconvoluted mass spectrum after the reaction with (7) showed two major peaks with a relative abundance of 100 and 40 %, respectively. The experimental mass of the first peak (11,758.76 Da) was in agreement with the predicted mass of the conjugated Pc (11,758.32 Da), while the second peak corresponded to the Pc monomer. This data proved that Pc P37C was covalently modified with (6) and (7). The reaction of Pc P37C with (7) was repeated and the product was loaded on the SEC column (Figure 21) to analyze the elution profile of the conjugated Pc. Two peak maxima were observed with elution volumes of ~65 and ~80 mL. These peaks correspond to particles with Mw ~30 and ~12 kDa respectively. The pooled SEC fractions 63-65 and 79-84 were analysed by ESI-MS (Figure 21). Both deconvoluted mass spectra showed a major peak with experimental masses of 11,757.39 Da for fractions 63-65 and 11,758.00 Da for fractions 79-84 which were in agreement with the predicted mass of the conjugated Pc (11,758.32 Da). This data revealed that both pooled SEC fractions contained the conjugated Pc P37C. The higher elution volume (~65 mL) for the 11.7 kDa conjugated Pc suggested a possible aggregation of the covalently modified protein. The peak maximum at ~80 mL was expected for the 11.7 kDa conjugated Pc. SDS- and native-PAGE analysis (Figure 22) of the pooled SEC fractions were also performed. The SDS-PAGE gel showed a major band at ~11 kDa in both fractions 63-65 and fractions 79-84, confirming the presence of Pc species. The protein profile of fractions 79-84 in the native-PAGE gel was similar both in the presence and in the absence of β-mercaptoethanol, but it was different to that of the monomeric Pc giving a further confirmation of the presence of the conjugated Pc.
Figure 20 Deconvoluted mass spectra of the reaction products of Pc P37C with (A) pyrenyl and (B) ethylphenyl maleimides. The mass 11,855.46 Da in (A) correspond to that of conjugated Pc (11,854.40 Da). The masses 11,555.98 and 11,758.76 Da in (B) are in agreement with the predicted masses of Pc monomer (11,557.09 Da) and conjugated Pc (11,758.32 Da), respectively.
**Figure 21** Deconvoluted mass spectra of fractions (A) 63-65 and (B) 79-84 from SEC (shown in box) of conjugated Pc P37C with ethylphenyl maleimide. The SEC elution buffer was 20 mM KH$_2$PO$_4$, 100 mM NaCl, 3 mM DTT, pH 6.0. The masses 11,757.39 and 11,758.00 Da in (A) and (B) are in agreement with the predicted mass of conjugated Pc (11,758.32 Da).
Chapter 2 - Covalent modification of solvent exposed cysteines

Figure 22 PAGE analysis of SEC fractions after the reaction of P<sub>c</sub> with ethylphenyl maleimide (A) SDS-PAGE of pooled fractions 63-65 and 79-84 (lanes 1 and 2 respectively). (B) Native-PAGE of pooled fractions 79-84 and P<sub>c</sub> monomer (Ref) without and with β-mercaptoethanol (BME). The SDS gel shows the presence of P<sub>c</sub> species and the native gel confirmed the formation of the conjugated P<sub>c</sub>.

P<sub>c</sub> P37C was reacted with (8) and (9) and the products were loaded on the SEC column to analyze the size of the functionalized P<sub>c</sub> with the polymer chains. A peak was observed with elution volume of ~75 mL in the SEC chromatogram (Figure 23A) of conjugated P<sub>c</sub> with (8) as expected for the ~13.5 kDa conjugated P<sub>c</sub>. Two peak maxima were observed with elution volumes of ~55 and ~80 mL in the SEC chromatogram (Figure 23B) of conjugated P<sub>c</sub> with (9). The first peak corresponded to the conjugated P<sub>c</sub> P37C and the second peak was expected for the 11.5 kDa P<sub>c</sub> monomer. The PEGylated P<sub>c</sub> eluted from SEC with an apparent molecular weight of ~44 kDa and migrated slowly on the SDS-PAGE gel (Figure 23C), as observed with the P<sub>c</sub> D45C mutant. The consistency of elution volume (55 mL) for both PEGylated P<sub>c</sub> mutants suggested that their molecular size is determined by the sizes of the native globular P<sub>c</sub> and the PEG, rather than by the PEGylation site. This result was comparable to previous observations with the 14.2 kDa α-lactalbumin. For example, the hydrodynamic radius (52.3 Å) of mono-PEGylated α-lactalbumin with a linear 20 kDa PEG was similar to that (53.0 Å) of the tetra-PEGylated protein with a linear 5 kDa PEG. The ESI-MS spectra of the SEC fractions containing the conjugated P<sub>c</sub> P37C with (8) and (9) showed a large number of fragments due to the
presence of the polymer chains compared to that of the Pc monomer, as observed with the Pc D45C mutant (Figure 14).

Figure 23 SEC chromatograms of Pc P37C reacted with (A) poly(N-isopropylacrylamide) and (B) methoxypolyethylene glycol maleimides (20 mM KH$_2$PO$_4$, 100 mM NaCl, 3 mM DTT, pH 6.0) (C) The PEGylated Pc migrates slowly on the SDS-PAGE gel due to interactions between SDS and PEG.

In the previous experiment, size exclusion chromatography has been used to analyze the size of the 16.5 kDa PEGylated Pc. It has been noted that the Mw of ~44 kDa obtained from the SEC elution profile was an apparent molecular weight, because the high elution volume was caused by the hydrodynamic properties of PEGylated proteins. In the next experiment, the size exclusion chromatography was used to examine if different counter-ions can affect the hydrodynamic properties of PEGylated Pc and modify the SEC elution profile. PEG can form complexes with cations, having small conformational differences.$^{122}$ For example, the hydrodynamic radius (12.9 nm) of a ~160 kDa PEG slightly decreased in aqueous 0.1 M KNO$_3$ (12.6 nm).$^{123}$ Counter-ions can screen the charge-charge interactions between proteins in solution. Increasing the ionic strength, the charge-charge interactions between proteins are disrupted and the proteins are eluted at different volumes from the SEC column.$^{114}$ Herein, the experiment was performed on six samples of
PEGylated Pc P37C (~0.2 mM each) in the presence of high (200 mM) concentrations of six different salts (MgCl$_2$, MgSO$_4$, NaCl, Na$_2$SO$_4$, NH$_4$Cl, (NH$_4$)$_2$SO$_4$). Higher salt concentrations were avoided to maintain a monophasic solution. In a biphasic solution the results can be different, because species can be partitioned differently in the two liquid coexisting phases. The chosen salts differ for their cationic radius ($\text{Mg}^{2+} < \text{Na}^+ < \text{NH}_4^+$) and for their ionic strength ($\text{NaCl} = \text{NH}_4\text{Cl} < \text{MgCl}_2 = \text{Na}_2\text{SO}_4 = (\text{NH}_4)_2\text{SO}_4 < \text{MgSO}_4$). Figure 24 illustrates the SEC chromatograms. A peak maximum at ~55 mL was observed in all experiments. This data showed that there was no effect of the tested counter-ions on the SEC elution profile. This result suggested that the SEC was not able to distinguish any modification of the hydrodynamic properties of PEGylated Pc in the presence of 200 mM salts.

**Figure 24** SEC chromatograms of PEGylated Pc Pc P37C in 20 mM KH$_2$PO$_4$, pH 6.0 containing 200 mM MgCl$_2$ (black), MgSO$_4$ (grey), NaCl (red), Na$_2$SO$_4$ (pink), NH$_4$Cl (blue) or (NH$_4$)$_2$SO$_4$ (cyan)
In the previous two sections, Pc functionalization verified by ESI-MS and some biophysical characterizations of functionalized Pc by using SEC were described. Among the tested reactions, the maleimide derivatives gave the better results. The iodoacetamide derivatives gave the worst result probably due to the low solubility of these compounds in water. Interestingly, the SEC analysis of the reaction product of Pc with dihexyl iodoacetamide revealed interactions with the hydrophobic resin of the column, whereas, the SEC chromatogram of functionalized Pc with ethylphenyl maleimide suggested a possible aggregation of the covalently modified protein. Finally, both PEGylated Pc mutants eluted at ~55 mL from the SEC column suggesting that their size is determined by the sizes of the Pc and the PEG, rather than by the PEGylation site. The next objective was to obtain structural information of a new self-assembled state of functionalized Pc by using X-ray crystallography. In the next section, the crystallization conditions and the crystal structure of conjugated Pc P37C with ethylphenyl maleimide will be briefly described. Further investigations on this crystal structure were not performed because the obtained self-assembled state was previously observed with the wild-type Pc crystal. In Chapter 3, the structural characterization of PEGylated Pc D45C will be described in more detail.

2.2.4 - Crystal structure of maleimide-functionalized Pc

Crystallisation drops containing the conjugated Pc P37C with ethylphenyl maleimide were prepared in-house using similar conditions used previously to grow Pc crystals. The crystallisation trials incorporated different combinations of PEG 8000, zinc acetate and sodium acetate to identify conditions that favoured protein crystallisation. Crystals were obtained in 6 % PEG 8000, 50 mM zinc acetate, 100 mM sodium acetate, pH 7.7. Data were collected at the European Synchrotron Radiation Facility by Dr. Valerie Pye (TCD) to a resolution of 2.65 Å and the crystal belonged to space group P3₁₁₂.
This crystal structure revealed the same crystal packing of the wild-type Pc crystal in which zinc-mediated Pc trimers were observed. Also in the crystal structure of the conjugated Pc P37C with ethylphenyl maleimide, the Pc trimers packed together via the hydrophobic patches to form the hexameric assembly (Figure 25A). After molecular replacement, a positive density in the F₀-Fₐ difference map was observed in the copper and zinc binding sites (Figure 25B). A positive density was also observed around the surface cysteines (Figure 25C) relating to the functionalization. The crystal packing did not change because the maleimides occupied a space that was available in the wild-type Pc crystal and the zinc ions were the dominant contributors to the packing.

Figure 25 Crystal structure of conjugated Pc P37C with ethylphenyl maleimide
(A) Hexameric assembly of trimers. The asymmetric unit comprises three molecules of Pc (magenta scale) packed together via the hydrophobic patches. Copper ions and Cys37 side chains are shown as spheres and sticks, respectively. (B) |Fo|-|Fc| map at the 2 σ level showing the zinc-mediated trimer and the copper binding sites in the asymmetric unit. The residues (Asp44, Asp45, His61) and (His39, Cys89, His92, Met97) bound to zinc and copper, respectively, are shown as sticks. (C) |Fo|-|Fc| map at the 1 σ level showing the maleimide modification at Cys37 (shown as stick) of a monomer. The figure was generated in PyMol (http://www.pymol.org).
2.3 - Conclusions

In this chapter the optimization of the purification of the Pc P37C and D45C mutants have been described. Initially, the mass analysis revealed the co-purification of a second protein that was identified as β-lactamase and the post-translational modification with glutathione. The protein contaminant was removed by using a step gradient of 50 mM NaCl on the DEAE column and the monomeric Pc was obtained by adding DTT to the SEC elution buffer.

The functionalization via cysteine of the two purified mutants was tested with iodoacetic acid (1), two iodoacetamide compounds (2 and 3), an aziridine derivative (4), a thiol compound (5) and four maleimide derivatives (6-9). ESI-MS analysis proved the formation of all the conjugated proteins. The better results were obtained with the maleimide compounds, while, 70-80 % of unreacted Pc was present after the reaction with the iodoacetamide compounds probably due to the low solubility of these reagents in water.

SEC was performed to analyse the elution profile of some conjugated Pc. It was observed that the Pc-(3) conjugate interacted with the resin of the column, the Pc-(7) conjugate formed aggregates and the elution profile of the Pc-(9) conjugate was not dependent on the PEGylation site and on the presence of 200 mM salt (MgCl$_2$, MgSO$_4$, NaCl, Na$_2$SO$_4$, NH$_4$Cl or (NH$_4$)$_2$SO$_4$).

Finally, a crystal of the conjugated Pc P37C with ethylphenyl maleimide was obtained to a resolution of 2.65 Å. The crystallization conditions were similar to those previously used to grow the wild-type Pc crystal. The crystal packing did not change: Pc trimers are mediated by zinc ions and the maleimide groups occupy a space that was available in the wild-type Pc crystal.
2.4 - Experimental section

**Protein production.** Plasmids were modified vectors of pET11PC \(^{85}\) that yielded Pc variants P37C and D45C, both with an N-terminal methionine. Pc was produced in *Escherichia coli* BL21(DE3) by using the auto-induction method developed by Studier.\(^{124}\) The expression method has been described previously.\(^{85,109,125}\) \(^{15}\)N-labeled protein was over-expressed by auto-induction using a two-step protocol. Cells were grown in LB medium, harvested and resuspended at four-fold the density in minimal medium (MM) according to Marley’s method.\(^{126}\) The MM culture for \(^{15}\)N-Pc contained 1g/L (\(^{15}\)NH\(_4\))\(_2\)SO\(_4\) as the sole nitrogen source. The MM cultures for \(^{15}\)N-Cys and \(^{15}\)N-Lys Pc were grown in the presence of 50 mg/L Cys and 100 mg/L of all other amino acids and the minimal medium was depleted of \(^{15}\)N source, except for the presence of the corresponding \(^{15}\)N-amino acid.\(^{127}\) Protein concentration was determined using UV-vis spectroscopy on a Lambda 35 (Perkin Elmer). The absorption spectrum was recorded at room temperature (22 °C) for detection of the band at 598 nm (\(\varepsilon = 4.3\) mM\(^{-1}\)cm\(^{-1}\)) of the oxidized form of the protein.\(^{85}\) Protein samples were analysed by electrophoresis on 15 % (w/v) polyacrylamide gels followed by Coomassie Brilliant Blue staining.

**Conjugation reactions.** Conjugation reactions were performed according to literature protocols.\(^{103}\) Pc was reduced with a ten-fold excess of DTT prior to the reaction with the ligands. The DTT was removed and a two-fold excess of ligands dissolved in DMSO or in 100 mM phosphate buffer, pH 7.2 or 7.8 was added to Pc. The conjugation reactions were performed overnight at 4 °C. The iodoacetamide derivative was protected from light to prevent the generation of iodine. The aziridine derivative was kindly provided by Dr. Karen Fahey (NUIG). Size exclusion chromatography was performed using 20 mM KH\(_2\)PO\(_4\), 100 mM NaCl, 3 mM DTT, pH 6.0 as an eluent on a fast performance liquid chromatography (FPLC) system equipped with a Superdex 75 prep grade column (GE Healthcare Life Sciences). The calibration curve was reported previously.\(^{114}\) Protein samples were desalted using 3 kDa MWCO centrifugal filter units (Millipore) for the mass analysis on a LCT Premier XE (Waters). Deconvoluted spectra were obtained using MassLynx v4.1 software.
These data were collected by Dr. Róisín Doohan or Ms. Marian Vignoles or Mr. Ger Fahy (NUIG). Molecular weights were predicted by using “Molecular Mass Calculator v2.02” at http://mods.rna.albany.edu/masspec/MoIE. The experimental masses in the deconvoluted spectra were compared with the value of average masses of unreacted and conjugated Pc. The height of the peaks in the deconvoluted spectra reflected the relative abundance of species.

**Synthesis of iodoacetamide derivatives.** The iodoacetamide derivatives were synthesized using known methods\textsuperscript{128,129} and commercially available reagents. TLC conducted on silica (1:1 toluene/ethyl acetate) and visualized with UV light: spots corresponded to isolated materials. Scheme 2 shows the reaction pathway. The 2-Iodo-N-(2-anthracenyl)acetamide (yellow solid) was isolated by filtration and washed with 5 % sodium bicarbonate and water and dried in vacuo. Mw calc. for C\textsubscript{16}H\textsubscript{12}INO (M-H) 359.9892 Da; found 359.9888 Da. The 2-Iodo-N-dihexylacetamide (yellow oil) was isolated with water/10 % NaOH and dried with rotary evaporator. Mw calc. for C\textsubscript{14}H\textsubscript{28}INO (M+H) 354.1290 Da; found 354.1295 Da.

\begin{center}
\includegraphics[width=\textwidth]{scheme2.png}
\end{center}

**Scheme 2** Reaction pathway of iodoacetamide synthesis
Chapter 3

Structural Characterization of PEGylated Plastocyanin
Abstract

The structural characterization of PEGylated Pc D45C was performed by using NMR spectroscopy and X-ray crystallography. NMR data revealed that the Pc functionalization caused a change of the chemical environment around the conjugated cysteine leaving intact the folding of the protein structure. NMR analysis in a physiological environment did not reveal any significant differences between Pc and PEGylated Pc, suggesting that the conjugated PEG did not influence Pc-protein interactions in this environment. The crystallographic structure of PEGylated Pc showed a right-handed antiparallel double-helical assembly of Pc. The structure is poorly packed with large cavities occupied by the PEGylated portion. The analysis of the PEG content in the crystal suggested that the space occupied by the conjugated PEG chains can be described by packed spheres each with a volume of $\sim 5 \times 10^4 \, \text{Å}^3$. In addition to providing a remarkable crystal packing of double-helical structures of Pc, the crystal data also indicated that the size ratio between PEG and Pc was important to obtain the supramolecular assembly of PEGylated Pc.
3.1 - Introduction

In this chapter, the supramolecular assembly of PEGylated Plastocyanin will be described. First, an overview of protein self-assembly systems will be presented. The focus will be on systems containing the polymer poly(ethylene glycol) (PEG).

Protein self-assembly systems are emerging as molecular devices with applications from medicine to biosensors. The design of supramolecular protein assemblies is challenging and has been achieved by using small molecule-protein interactions, engineering of fusion proteins or chemical modification of the protein surface. Supramolecular protein assemblies can be also promoted by macromolecular crowding agents. These agents affect the thermodynamics and kinetics equilibrium of protein folding, ligand binding and protein-protein interactions through the excluded volume effect. In the following paragraphs, two examples concerning the most common crowding agent PEG will be described. The first is about the α-synuclein fibrillation and the second is about the Plastocyanin crystallization.

Two independent studies showed that PEG stimulates in vitro fibril formation of the 14 kDa α-synuclein. Shtilerman et al. used PEG 20,000 over the concentration range 5-15 % and found an increase of the α-synuclein fibrillation rate. Similar observations were reported by Uversky et al. They found that α-synuclein fibrillation was accelerated by the presence of 2.5 % PEG 3350 and by an increase of the PEG concentration (up to 15 %).

Crowley et al. obtained the crystal structures of the 11 kDa Plastocyanin in the presence and absence of PEG. The presence of 5 % PEG 8000 dramatically altered the crystal structure assembly: Plastocyanin has a more "tightly" packed structure compared with a "loosely" packed structure obtained in the absence of PEG.
In addition to the use of PEG as a crowding agent, PEGylation is also extensively investigated due to the growing medical application of PEGylated proteins: the covalent attachment of a PEG chain can enhance the solubility and stability of therapeutic proteins.\textsuperscript{152-154} The first example of PEGylated protein was from Abuchowski \textit{et al.}\textsuperscript{155} who used bovine serum albumin (BSA) to control immunogenicity of the protein. This system was further applied to the generation of new classes of protein-polymer therapeutics inspired by supramolecular assemblies.\textsuperscript{156-162} For example, Lee \textit{et al.}\textsuperscript{158} generated nanoparticles containing a poly(histidine-\textit{co}-phenylalanine) core, a PEG inner shell and a conjugated BSA outer shell. These mixed micelles can be loaded with the drug doxorubicin and taken by cells. In this system the PEG and BSA shells are used to prevent potential immune responses.

PEGylated proteins are typically obtained by reacting activated PEG reagents with residue side chains, such as the sulfhydryl group of cysteine.\textsuperscript{56,163,164} These reactions lead to mixtures of modified and unmodified proteins or mixtures of PEGylated forms which necessitate purification.\textsuperscript{165-167} Size exclusion chromatography (SEC) is useful because the hydrodynamic volume (or viscosity radius) of PEGylated proteins increases. The hydrodynamic radius is a fundamental measure of molecular size which is determined by the molecular size and conformation. Fee \textit{et al.}\textsuperscript{119} developed a method to predict the viscosity radius of PEGylated protein in terms of a “grafted PEG layer”. According to this conformation, the predicted viscosity radii agreed well with the experimentally determined values using SEC. However, the conformation in which a PEG chain forms a layer over the protein surface cannot explain additional experimental data, reported below.\textsuperscript{121,168,169} Small angle neutron scattering revealed that a single conjugated PEG chain has similar dimensions to a free PEG chain in solution.\textsuperscript{169} Dynamic light scattering showed that the hydrodynamic size of a PEG-protein conjugate is equivalent to the sum of the diameters of
the free PEG and protein. Data from SEC indicated that the hydrodynamic volume of PEGylated proteins can be estimated as the sum of the corresponding volumes of the two macromolecules. All of these observations have suggested that the conformation of the conjugated PEG chain is a random-coil structure freely exposed to solvent.

Small-angle X-ray scattering (SAXS) and small angle neutron scattering (SANS) methods have been applied to determine the structural features of PEGylated proteins. Pai et al. used SANS to distinguish between a "dumbbell" and a "shroud" model (Figure 1). Using two proteins, the 14 kDa lysozyme and the 22 kDa human growth hormone each conjugated with a 20 kDa linear PEG, they found that only the random-coil form factor could describe the data and confirmed the "dumbbell" model. He et al. analysed the homodimeric human galectin (14.5 kDa per monomer) mono-conjugated with a 5 kDa linear PEG by SAXS and SANS. The data fit a model of an ellipsoid core with polymer chains attached to the surface and showed that the PEG adopted a “more compact conformation” with increasing concentration of the conjugate. Svergun et al. used SAXS to study the heterotetrameric hemoglobin (16 kDa per monomer) conjugated with two or six chains of a 5 kDa linear PEG. The reconstructed shapes illustrate either a "dual mushroom" or a "dual mushroom-to-brush" conformation. The authors reported that the majority of the PEG chains protruded away from the surface and a portion interacted with the protein. All these data have revealed that the native structure of proteins is not altered by PEGylation and it is likely the PEG chain adopts a random-coil conformation. The combination of SAXS and SANS methods has provided complementary structural information on the shape of PEGylated proteins. As the data from small-angle scattering techniques are time and ensemble-averaged, it is possible to acquire structural information for flexible systems.
Here, the structural characterization of PEGylated Pc is described. This study is relevant to determine the impact of PEGylation on the structure of the conjugated protein. The PEGylated Pc expressed, purified and characterized as described in Chapter 2 was studied by using NMR spectroscopy and X-ray crystallography. This study was focused on the Pc D45C because the first crystal structure of a PEGylated protein was obtained with this mutant. The PEGylated Pc P37C crystallization was not successful. The crystal structure, solved to a resolution of 4.2 Å, showed an antiparallel double-helical assembly of Pc. The volume occupied by a PEG chain and the concentration of PEG in the crystal was estimated. The comparison between these values and those of literature suggested that protein-protein association was promoted by PEGylation. These results are reported after the NMR analysis of PEGylated Pc: data were also acquired in human serum to understand if the conjugated PEG could influence protein interactions in a physiological environment.
3.2 - Results and discussion

3.2.1 - NMR analysis

The effect of PEGylation on the Pc D45C structure was evaluated by line-width analysis and by mapping the chemical shift perturbation of the backbone amide resonances. The 2D $^1$H-$^{15}$N HSQC spectrum was assigned according to previous data, except for the resonances belonging to Cys45 and Lys46. These resonances were assigned by using $^{15}$N-Cys and $^{15}$N-Lys Pc. In the following paragraphs, the analysis of the HSQC spectra of $^{15}$N-Cys and $^{15}$N-Lys Pc will be reported before the analysis of the HSQC spectrum of $^{15}$N-Pc.

![Figure 2](image.png)

**Figure 2** Spectral region from overlaid 2D $^1$H-$^{15}$N HSQC spectra of Pc (black), PEGylated Pc (orange) of (A) $^{15}$N-Cys and (B) $^{15}$N-Lys Pc. Spectra recorded in 20 mM KH$_2$PO$_4$, 100 mM NaCl, pH 6.0.

The 2D $^1$H-$^{15}$N HSQC spectra of unmodified and PEGylated $^{15}$N-Cys and $^{15}$N-Lys Pc are shown in Figure 2. Figure 2B was divided into
two regions for graphical reasons (there are no resonances in the region 8.5 (\textsuperscript{1}H) and 123 ppm (\textsuperscript{15}N) of the HSQC spectrum). Two resonances were observed in the HSQC spectrum of unmodified \textsuperscript{15}N-Cys Pc in agreement with the number of cysteine residues. The resonance at 7.7 ppm belonged to the Cys89 according to the previous assignment. The resonance at 8.8 ppm was assigned to Cys45. The PEGylated \textsuperscript{15}N-Cys Pc spectrum revealed three resonances. The Cys89 was unshifted whereas Cys45 was shifted and duplicated (the duplicated NH will be explained later). Seven resonances were observed in the HSQC spectrum of unmodified \textsuperscript{15}N-Lys Pc in agreement with the number of lysine residues. All resonances, except the resonance at 8.2 ppm, belonged to the lysine residues of the previous assignment. The resonance at 8.2 ppm was assigned to Lys46. The PEGylated \textsuperscript{15}N-Lys Pc spectrum revealed four unshifted resonances and three shifted resonances. Duplicated NH resonances for Lys53 and Lys58 can be distinguished in this spectrum. In total, six duplicated NH resonances (Phe43, Cys45, Lys53, Ala56, Asp57 and Lys58) were distinguished in the 2D \textsuperscript{1}H-\textsuperscript{15}N HSQC spectrum of PEGylated Pc (Figure 3). These duplicated resonances were indicative of two configurations and not to a slow exchange between two conformations. The sulphydryl group of cysteine coordinates the maleimide group via Michael addition and the reaction product possesses a chiral centre (with an absolute configuration which can be either R or S). The duplicated NH resonances are in the vicinity of the conjugated cysteine suggesting that the whole protein structure, in terms of the polypeptide backbone, did not change.

A region of 2D \textsuperscript{1}H-\textsuperscript{15}N HSQC spectra of unmodified and PEGylated \textsuperscript{15}N-Pc is shown in Figure 3. The majority of Pc resonances demonstrated small changes in line-width of the \textsuperscript{1}H resonances, increasing on average by 1 Hz (Table 1) for the PEGylated \textsuperscript{15}N-Pc. This result suggested that the molecular tumbling was weakly influenced by the conjugated PEG. The modest line-broadening can be explained by a
highly independent motion of the protein domain and the PEG chain as observed previously for interferon-α₂a (19 kDa) PEGylated with a branched 40 kDa PEG.

Figure 3 Spectral region from overlaid 2D $^1$H-$^{15}$N HSQC spectra of 0.3 mM Pc (black) and 0.3 mM PEGylated Pc (orange) in 20 mM KH$_2$PO$_4$, 100 mM NaCl, pH 6.0.
### Table 1

<table>
<thead>
<tr>
<th>Residue</th>
<th>¹H line-width (Hz)</th>
<th>¹H line-width (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pc</td>
<td>PEGylated Pc</td>
</tr>
<tr>
<td>F3</td>
<td>19.9</td>
<td>20.8</td>
</tr>
<tr>
<td>T4</td>
<td>19.5</td>
<td>20.7</td>
</tr>
<tr>
<td>V5</td>
<td>21.6</td>
<td>24.3</td>
</tr>
<tr>
<td>K6</td>
<td>18.0</td>
<td>19.1</td>
</tr>
<tr>
<td>G8</td>
<td>20.4</td>
<td>19.5</td>
</tr>
<tr>
<td>A9</td>
<td>16.0</td>
<td>17.1</td>
</tr>
<tr>
<td>D10</td>
<td>13.4</td>
<td>14.2</td>
</tr>
<tr>
<td>S11</td>
<td>15.5</td>
<td>16.2</td>
</tr>
<tr>
<td>G12</td>
<td>17.6</td>
<td>18.4</td>
</tr>
<tr>
<td>L13</td>
<td>16.3</td>
<td>17.2</td>
</tr>
<tr>
<td>L14</td>
<td>16.2</td>
<td>17.0</td>
</tr>
<tr>
<td>Q15</td>
<td>18.9</td>
<td>19.9</td>
</tr>
<tr>
<td>F16</td>
<td>20.4</td>
<td>20.0</td>
</tr>
<tr>
<td>E17</td>
<td>19.6</td>
<td>20.4</td>
</tr>
<tr>
<td>A19</td>
<td>13.3</td>
<td>13.5</td>
</tr>
<tr>
<td>N20</td>
<td>20.9</td>
<td>20.4</td>
</tr>
<tr>
<td>T22</td>
<td>21.3</td>
<td>22.4</td>
</tr>
<tr>
<td>V23</td>
<td>22.6</td>
<td>22.0</td>
</tr>
<tr>
<td>H24</td>
<td>21.0</td>
<td>21.7</td>
</tr>
<tr>
<td>G26</td>
<td>21.7</td>
<td>21.6</td>
</tr>
<tr>
<td>D27</td>
<td>13.7</td>
<td>15.8</td>
</tr>
<tr>
<td>V29</td>
<td>21.1</td>
<td>23.0</td>
</tr>
<tr>
<td>K30</td>
<td>21.5</td>
<td>21.1</td>
</tr>
<tr>
<td>W31</td>
<td>20.5</td>
<td>21.3</td>
</tr>
<tr>
<td>V32</td>
<td>22.5</td>
<td>22.1</td>
</tr>
<tr>
<td>N33</td>
<td>12.9</td>
<td>13.8</td>
</tr>
<tr>
<td>N34</td>
<td>20.4</td>
<td>20.3</td>
</tr>
<tr>
<td>K35</td>
<td>18.2</td>
<td>18.0</td>
</tr>
<tr>
<td>L36</td>
<td>17.0</td>
<td>19.0</td>
</tr>
<tr>
<td>H39</td>
<td>21.7</td>
<td>20.5</td>
</tr>
<tr>
<td>N40</td>
<td>16.9</td>
<td>17.0</td>
</tr>
<tr>
<td>I41</td>
<td>20.8</td>
<td>20.8</td>
</tr>
<tr>
<td>D44</td>
<td>17.0</td>
<td>20.2</td>
</tr>
<tr>
<td>Q47</td>
<td>21.0</td>
<td>21.5</td>
</tr>
<tr>
<td>G50</td>
<td>16.8</td>
<td>20.0</td>
</tr>
<tr>
<td>A51</td>
<td>19.6</td>
<td>18.6</td>
</tr>
</tbody>
</table>

**Average line-width (Pc) = 18.6 Hz**

**Average line-width (PEGylated Pc) = 19.6 Hz**

[^1]: The average line-width was calculated for 70% of the assigned resonances in CCPNmr. 174
In addition to line-broadening, chemical shift changes of the backbone amide resonances were observed. Figure 4A is a plot of the chemical shift changes for $^1$H and $^{15}$N resonances derived from the 2D $^1$H-$^{15}$N HSQC spectrum of PEGylated Pc with respect to the spectrum of unmodified Pc D45C. The largest observed perturbations were for the $^1$H of Cys45 (0.28 ppm) and for the $^{15}$N of Val48 (1.00 ppm). Mapping the chemical shift perturbations onto the crystal structure of Pc (Figure 4B) revealed that the amide resonances with significant (> 1SD) chemical shift changes ($\Delta \delta^1$H > 0.04 ppm and/or $\Delta \delta^{15}$N > 0.1 ppm) were in the vicinity of the covalently modified cysteine, with the perturbations of Thr86 and Val98 induced through the polypeptide backbone. This map also illustrated that the chemical shift changes were not spread widely over the protein surface suggesting that the native fold of Pc was maintained. Chemical shift perturbations can be explained by a solvation/desolvation effect. The hydrogen bonds of water molecules with the residues in the vicinity of the covalently modified cysteine might change: there is a new structure of water around these residues thus the environment is different and chemical shift perturbations are observed. Alternatively, the chemical shift perturbations localized in a region of the surface of Pc can be explained by assuming the "dumbbell" model (Figure 1A) of PEGylated proteins. If the PEG chain formed a layer over the protein surface according to the "shroud" model (Figure 1B), all the residues should have significant chemical shift perturbations. In conclusion, the reported NMR analysis did not show an assembled state of PEGylated Pc in solution.
Figure 4 (A) Plot of chemical shift changes measured for Pc backbone amides due to the PEGylation at C45. Pc residues are numbered from 2 to 105. (B) Residues for which the amide resonance experienced significant chemical shift change ($\Delta \delta^{1}H > 0.04$ ppm and/or $\Delta \delta^{15}N > 0.1$ ppm) are coloured orange on the surface representation of Pc. The maleimide is shown as purple sticks.

An investigation to understand if the conjugated PEG could have an effect on protein interactions in a physiological environment, such as human serum, was also performed. 2D $^{1}H$-$^{15}N$ HSQC spectra of Pc D45C and PEGylated Pc were recorded in serum and in the presence of the 66 kDa BSA. This protein was chosen because it is homologous (76 % sequence identity) to the human protein which is the most abundant protein in the human serum. The tested concentrations of BSA were 20, 40 (equal to the physiological concentration of albumin) and 80 g/L. The proteins in the mixtures analysed by NMR were
analysed on a SDS-PAGE gel (Figure 5). The 11.5 kDa $^{15}$N-Pc (lane 1) in the presence of 40 g/L BSA (lane 2) and in human serum (lane 3) was compared to the 16.5 kDa PEGylated $^{15}$N-Pc (lane 4) in the presence of 40 g/L BSA (lane 5) and in human serum (lane 6). Both protein samples presented the characteristic bands of the mixtures, *i.e.* the unmodified and PEGylated Pc in the presence of BSA (lane 7) and in human serum (lane 8).

**Figure 5** SDS-PAGE analysis of $^{15}$N-Pc and PEGylated $^{15}$N-Pc in BSA and in human serum. (lane 1) $^{15}$N-Pc; (lane 2-3) $^{15}$N-Pc in the presence of 40 g/L BSA and human serum, respectively; (lane 4) PEGylated $^{15}$N-Pc; (lane 5-6) PEGylated $^{15}$N-Pc in the presence of 40 g/L BSA and human serum, respectively; (lane 7) BSA; (lane 8) human serum.

2D $^1$H-$^{15}$N HSQC spectra of Pc and PEGylated Pc were acquired at pH 7.8 (equal to the pH of the human serum sample). These spectra showed a good dispersion with three resonances (Asp10, Gly50 and Gly94) belonging to loop regions broadened beyond detection as explained below. Rapid amide proton exchange with water causes backbone amide protons to be broadened at high pH. Accordingly, the amide proton resonances of residues in loop regions could not be observed. Additionally, in the HSQC spectra of PEGylated Pc the six duplicated NH resonances (Phe43, Cys45, Lys53, Ala56, Asp57 and Lys58) were observed as a result of the two possible absolute configurations of the chiral centre of the reaction product.
The majority of resonances of Pc and PEGylated Pc in BSA and serum demonstrated small changes in line-width of the $^1$H resonances, increasing on average by 1-5 Hz (Table 2). This result, compared to the approximately threefold line-width increase for the Pc bound to cytochrome $f$, suggested that the molecular tumbling of Pc D45C and PEGylated Pc in the physiological environment was not affected by a protein-protein binding event.

Table 2 $^1$H line-width measurements of the NH resonances extracted from HSQC spectra of Pc D45C and PEGylated Pc in 20 mM KH$_2$PO$_4$, 100 mM NaCl, pH 7.8 at 30 °C.  

<table>
<thead>
<tr>
<th>Protein</th>
<th>BSA (g/L)</th>
<th>Avg. line-width (Hz)</th>
<th>Avg. increase (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PcD45C</td>
<td>0</td>
<td>18.6</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>19.0</td>
<td>0.4</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>20.0</td>
<td>1.4</td>
</tr>
<tr>
<td></td>
<td>80</td>
<td>21.1</td>
<td>2.5</td>
</tr>
<tr>
<td>human serum</td>
<td></td>
<td>22.9</td>
<td>4.3</td>
</tr>
<tr>
<td>PEGylated PcD45C</td>
<td>0</td>
<td>19.6</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>21.1</td>
<td>1.5</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>21.7</td>
<td>2.1</td>
</tr>
<tr>
<td></td>
<td>80</td>
<td>22.9</td>
<td>3.3</td>
</tr>
<tr>
<td>human serum</td>
<td></td>
<td>24.8</td>
<td>5.2</td>
</tr>
</tbody>
</table>

$^a$The average line-width was calculated for 70 % of the assigned resonances in CCPNmr.$^{174}$

In addition to the line-width analysis, chemical shift analysis was performed. Not significant (< 0.04 ppm) average chemical shift changes were observed in the HSQC spectra of Pc D45C and PEGylated Pc (Figure 6) in the presence of BSA and in human serum. The largest average chemical shift perturbations were observed for His24, His61,
Gln63 and Gly105. This data suggested that Pc and PEGylated Pc did not interact with other proteins in the physiological environment. The NMR data can be explained by the protonation/deprotonation equilibrium of the histidine side chains. The pKₐ of His24 and His61 are near to the experimental pH thus it is equally probable for these residues to be protonated as deprotonated. The chemical shifts perturbations observed for Gln63 and Gly105 can be explained by their vicinity to His24 and His61 in the 3D protein structure. The pairs His24-Gly105 and His61-Gln63 are sufficiently close to interact with each other, hence a change of the environment (such as a protonation/deprotonation event) of one of the two residues in a pair may affect the chemical shift of both resonances in each pair. In conclusion, line-width and chemical shift analysis of Pc and PEGylated Pc in serum suggested that the conjugated PEG did not influence any interactions of Pc with other proteins in a physiological environment.

The NMR data of PEGylated Pc reported so far were enriched with other structural information obtained from X-ray crystallography. In the next sections, the crystallization conditions and the crystal structure will be described.
Chapter 3 - Structural characterization of PEGylated Plastocyanin

Figure 6 Spectral regions from overlaid 2D $^1$H-$^{15}$N HSQC spectra of (A) 0.3 mM Pc and (B) 0.3 mM PEGylated Pc (black) in 20 g/L BSA (cyan) and human serum (blue) in 20 mM KH$_2$PO$_4$, 100 mM NaCl, pH 7.8. Plot of the averaged chemical shift perturbations experienced by the backbone amide resonances of (A) Pc and (B) PEGylated Pc in the presence of 20 (●), 40 (□) and 80 (▼) g/L BSA and in human serum (◊). Pc residues are numbered from 2 to 105.
3.2.2 - Crystallization conditions

The initial crystallization-screening experiment was performed by choosing a commercial screen (Jena Bioscience, CS-153) available in our lab. The solution containing 3.15 M ammonium sulfate, 100 mM citric acid, pH 5 gave spherulites which are a good starting point for optimization.\textsuperscript{182} Subsequent screens were performed with zinc acetate, previously used in Plastocyanin crystallization.\textsuperscript{85} Different combinations of buffers (such as sodium acetate) and salts (ammonium sulfate and zinc acetate) were screened, but PEGylated Pc D45C did not crystallize in the presence of zinc acetate. This lack of crystals is likely explained by the fact that D45 is coordinated to a zinc ion in the crystal structure of the wild-type protein and this coordination cannot occur in the conjugated protein.\textsuperscript{85} A new strategy was performed using ammonium sulfate in the presence of potassium ferricyanide. Potassium ferricyanide was a reasonable choice because it is used to oxidize Pc. Crystallization trials using 5-30 mM potassium ferricyanide and 26 % ammonium sulfate in 100 mM sodium acetate at pH 6.3 were set up. The crystallization drops were free of precipitate at < 20 mM ferricyanide and showed a phase separation at > 20 mM ferricyanide. The optimization was achieved when a screen with 0-50 % ammonium sulfate and 30 mM potassium ferricyanide was performed. Phase separation occurred at < 30 % ammonium sulfate but spherulites and microcrystals were obtained at higher ammonium sulfate concentrations. Figure 7 illustrates two drops in which phase separation and spherulites were observed. The transition from phase separation to spherulites might be interpreted as a salting-out process.\textsuperscript{182} When the phase separation occurred, the protein preferentially was accumulated in one or the other phase, but when the salt concentration was increased, some of the water molecules solvated the salt ions. This induced stronger protein-protein interactions than solvent-protein interactions leading to the formation of spherulites.
Reproducible oxidized PEGylated Pc crystals were obtained in 48 % ammonium sulfate, 100 mM sodium acetate, 30 mM potassium ferricyanide, pH 6.3. A control drop lacking the protein was devoid of precipitate (Figure 8).

**Figure 7** Crystallization drops of PEGylated Pc in which (A) phase separation and (B) spherulites were observed. The reservoir solutions contained 20 and 50 % ammonium sulfate, respectively with 30 mM potassium ferricyanide in 100 mM sodium acetate, pH 6.3.

**Figure 8** (A) PEGylated Pc crystals were obtained with 48 % ammonium sulfate, 100 mM sodium acetate, 30 mM potassium ferricyanide, pH 6.3. (B) A control drop lacking protein was devoid of precipitate.
3.2.3 - Structure and assembly

The crystal structure of PEGylated Pc was solved to a resolution of 4.2 Å and belonged to space group P4\(_1\)2\(_1\)2\(_1\) (Table 3). The asymmetric unit comprised four monomers of Pc referred to as A, B, C and D (Figure 9A). Protein-protein interfaces were analysed in PISA and the crystal assembly was interpreted in the light of the volume occupied by a PEG chain and the concentration of PEG in the crystal as discussed in more details later.

**Table 3** Data collection, processing and refinement statistics of PEGylated Pc\(^a\)

<table>
<thead>
<tr>
<th>Data collection</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Light Source</td>
<td>Diamond, beamline I24</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.96861</td>
</tr>
<tr>
<td>Space group</td>
<td>P4(_1)2(_1)2(_1)</td>
</tr>
<tr>
<td>Cell constants</td>
<td>(a = b = 116.1) Å, (c = 172.1) Å</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>44.47-4.20 (4.31-4.20)</td>
</tr>
<tr>
<td>No. of unique reflections</td>
<td>9064 (662)</td>
</tr>
<tr>
<td>Multiplicity</td>
<td>5.6 (5.9)</td>
</tr>
<tr>
<td>I/σ(I)</td>
<td>5.4 (1.7)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.6 (100)</td>
</tr>
<tr>
<td>(R_{\text{merge}})</td>
<td>0.379 (1.152)</td>
</tr>
<tr>
<td>Solvent content (%)</td>
<td>72</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Refinement</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(R_{\text{factor}}) (%)</td>
<td>23</td>
</tr>
<tr>
<td>(R_{\text{free}}) (%)</td>
<td>27</td>
</tr>
<tr>
<td>No. of molecules in asymmetric unit</td>
<td>4</td>
</tr>
<tr>
<td>Average B, all atoms (Å(^2))</td>
<td>120.0</td>
</tr>
<tr>
<td>Ramachandran analysis(^c)</td>
<td></td>
</tr>
<tr>
<td>% residues in favoured regions</td>
<td>96</td>
</tr>
<tr>
<td>% residues in allowed regions</td>
<td>4</td>
</tr>
</tbody>
</table>

\(^a\)Values in parentheses correspond to those for the highest resolution shell;  
\(^b\)\(R_{\text{merge}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|/\sum_{hkl} \sum_i I_i(hkl); \(^c\)Calculated with MolProbity.\(^{183}\)
Figure 9  (A) |Fo|-[Fc| map at the 4 σ level showing the copper binding sites of the four PEGylated Pc monomers in the asymmetric unit. The backbone is shown as ribbon. The copper binding residues (His39, Cys89, His92 and Met97) are shown as sticks. (B) 2|Fo|-|Fc| map at the 1 σ level showing the maleimide modification for monomers A (purple) and C (green). The side chain of cysteines and the maleimide are shown as sticks. The backbone is shown as ribbon.
Figure 10 Assembly of PEGylated Pc in the crystal structure. (A) Right-handed double-helical structure, (B) helix-helix packing structure and (C) space group symmetry expansion viewed along the $a$ axis. Ribbon representation of protein backbone with helices in purple and green.
After molecular replacement with the structure of Pc (PDB ID 2W88) as the search model, a positive density in the $F_0 - F_C$ difference map was observed in the copper binding site of each monomer (Figure 9A). These electron densities belonged to copper ions which were not included in the search model. This result suggested that the structure solution was reasonable. Electron density around the solvent exposed cysteine residue of each monomer showed the presence of the maleimide group (Figure 9B). The absence of electron density for the conjugated PEG indicated that it was disordered and suggested a random-coil conformation. Space group symmetry expansion revealed a right-handed antiparallel double-helical assembly (Figure 10). The analysis and interpretation of this assembly are reported below.

The maleimide groups of monomers A and C in the asymmetric unit are at the interface of monomer-monomer contacts (Figure 9B) and the interface between these monomers involves an area $\sim215 \text{ Å}^2$ (Table 4). Comparable interfaces are observed in the neighbouring molecules ($\sim220 \text{ Å}^2$). In this case, the maleimide groups of monomers B and D are at the interface of monomer-monomer contacts due to the helix conformation. Table 4 lists all the interface areas. These data were generated by using PISA$^{184}$ and they were classified into two types according to the contribution of the interfaces to the crystal packing. Type 1 interfaces contribute to forming the double-helix (Figure 10A), those of type 2 contribute to helix-helix packing (Figure 10B). Type 1 interfaces have an average area of $440 \text{ Å}^2$. This area comprises the contacts between two consecutive monomers (pairs AD and BC in Table 4) and two crossed monomers (pair AB and CD in Table 4) in the helical strand. Type 2 interfaces have an average area of $250 \text{ Å}^2$. The reported interface area values, which resemble the contact areas found for transient cupredoxin dimers (ranging from 300 to 500 $\text{ Å}^2$),$^{185}$ are indicative of low specificity interfaces. The interfaces of transient protein complexes are flat and poorly packed favouring a fast dissociation.
Table 4 Interface areas involved in the PEGylated Pc crystal packing.\textsuperscript{a}

<table>
<thead>
<tr>
<th>Type\textsuperscript{b}</th>
<th>Monomer</th>
<th>Monomer</th>
<th>Symmetry operation</th>
<th>Interface area (Å\textsuperscript{2})</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>D</td>
<td>C</td>
<td>x,y,z</td>
<td>459.6</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>A</td>
<td>x,y,z</td>
<td>455.6</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>B</td>
<td>x,y,z</td>
<td>429.3</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>A</td>
<td>x,y,z</td>
<td>400.0</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>A</td>
<td>x,y,z</td>
<td>215.3</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>B</td>
<td>-x+1/2,-y+1/2,-z+1/4</td>
<td>439.2</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>A</td>
<td>-x+1/2,-y+1/2,-z+1/4</td>
<td>431.5</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>B</td>
<td>-x+1/2,-y+1/2,-z+1/4</td>
<td>219.7</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>B</td>
<td>y,x,-z</td>
<td>373.7</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>A</td>
<td>y,x,-z</td>
<td>217.2</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>D</td>
<td>y,x,-z</td>
<td>156.5</td>
</tr>
</tbody>
</table>

\textsuperscript{a}Data generated by using PISA\textsuperscript{184}; \textsuperscript{b}Interfaces which contribute to forming the double-helix are classified as type 1 and those which contribute to helix-helix packing as type 2.

The analysis of the interface areas involved in the PEGylated Pc crystal packing revealed the poorly packed structure shown in Figure 10C. Considering the disordered conformation and the space needed for the conjugated PEG, the space in the crystal not occupied by the double-helical assembly will be occupied by the PEG. Accordingly, the volume of a PEG chain in the crystal was calculated as follow. Knowing the solvent content and the volume (V\textsubscript{cell}) of a unit cell, the volume of solvent in a unit cell was calculated by multiplying V\textsubscript{cell} by the solvent content. The resulted value was dividing by 32 (the number of monomers in the unit cell) to obtained the volume occupied by a conjugated PEG chain (see experimental section for details).

The volume of a 5 kDa PEG chain in the crystal was compared to literature values. These results are reported in Table 5 and explained below. The first source from literature is the Fee data\textsuperscript{119}: the “volume of PEG layer” of mono-PEGylated proteins with PEG5000 was plotted as a function of the Mw of PEGylated proteins and a linear regression was plotted. The other sources from literature come from the knowledge of the size of a 5 kDa PEG chain: a PEG chain can be approximately
Chapter 3 - Structural characterization of PEGylated Plastocyanin

contained in a sphere of radius $R_h$ (hydrodynamic radius) or $R_g$ (radius of gyration) or with diameter $R_F$ (Flory dimension). These radii can be determined from the molecular weight ($M_w$) or the number of units per polymer ($N$) according to empirical equations. Alternatively, dynamic light (DLS) and small angle neutron scattering (SANS) can be used to measure respectively $R_h$ and $R_g$ in solution. In Table 5, it is worthy to note the compaction of PEG at an increased PEG concentration: SANS measurements suggested that the volume occupied by PEG is halved by going to a PEG solution 20 times more concentrated. (The concentration of PEG in the crystal of PEGylated Pc was calculated and interpreted as reported later. At the end of this section, the use of the Flory dimension to assess the conformation of a conjugated PEG chain will be also explained).

Table 5 Volumes of a 5 kDa PEG chain calculating from different sources.

<table>
<thead>
<tr>
<th>Source</th>
<th>Volume (Å$^3$)</th>
<th>Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extracted from Fee data$^a$</td>
<td>$9.8 \times 10^4$</td>
<td>[119]</td>
</tr>
<tr>
<td>$R_h = 0.1912 \ M_w^{0.559}; M_w = 5000$ Da</td>
<td>$4.6 \times 10^4$</td>
<td>[119]</td>
</tr>
<tr>
<td>$R_g = 0.2 \ M_w^{0.58}; M_w = 5000$ Da</td>
<td>$9.1 \times 10^4$</td>
<td>[187]</td>
</tr>
<tr>
<td>$R_F = 3.5 \ N^{0.6}; N = 115$</td>
<td>$11.4 \times 10^4$</td>
<td>[171]</td>
</tr>
<tr>
<td>$R_g = R_F/\sqrt{6} = (3.5 \ N^{0.6})/\sqrt{6}; N = 115$</td>
<td>$6.2 \times 10^4$</td>
<td>[186]</td>
</tr>
<tr>
<td>$R_h$ from DLS; [PEG5000] = 5.3 g/L</td>
<td>$4.5 \times 10^4$</td>
<td>[188]</td>
</tr>
<tr>
<td>$R_g$ from SANS; [PEG5000] = 1.4 g/L</td>
<td>$8.2 \times 10^4$</td>
<td>[170]</td>
</tr>
<tr>
<td>$R_g$ from SANS; [PEG5000] = 20 g/L</td>
<td>$4.5 \times 10^4$</td>
<td>[170]</td>
</tr>
</tbody>
</table>

$^a$ Data extracted from Table 2: the “volume of PEG layer” of mono-PEGylated proteins was plotted as a function of $M_w$ of PEGylated proteins. Linear regression: $y = 0.9171 \ x + 83274$ ($R^2 = 0.9861$) with $x = 16,500$ Da for PEGylated Pc D45C.

From literature, the volume range of a 5 kDa PEG chain was $4.5-11.4 \times 10^4$ Å$^3$. This range is roughly equivalent or two-fold higher than the volume calculated from the crystal of PEGylated Pc ($4.6-5.8 \times 10^4$ Å$^3$). This result suggested that a conjugated PEG chain is contained in a sphere of volume $\sim 5 \times 10^4$ Å$^3$ or folds in on itself to occupy this volume in the crystal. This result suggested that the conjugated PEG promoted
the double-helical assembly as a manifestation of excluded volume effect, which prohibits two polymers to occupy the same space.\textsuperscript{186}

The volume of solvent in a unit cell was used to calculate the concentration of PEG in the crystal as follow. The volume was converted from $\text{Å}^3$ to litres and the molarity concentration was calculated (see experimental section for details). The molarity was converted into g/L by multiplying it by the molecular weight of PEG. It was estimated that the unit cell contained 140-180 g/L of PEG, which is near the crossover concentration (200 g/L) from a dilute to a semidilute regime in a system containing proteins and PEG.\textsuperscript{148} In this system, the PEG chains start to overlap with a balance of repulsive and attractive forces between them. In the approximation of a spherical shape for each PEG chain, the dilute regime can be represented by spheres separated from each other. These spheres come closer each other and eventually come into contact as the concentration increases to the semidilute regime. This scenario suggests that the space occupied by the conjugated PEG chains in the crystal can be described by packed spheres each with a volume of $\sim 5 \times 10^4 \text{Å}^3$.

Previously, it has been noted that the volume of a 5 kDa PEG was inversely proportional to the concentration of PEG (Table 5). Also, it was found that the volume of 20 g/L PEG5000 ($4.5 \times 10^4 \text{Å}^3$) was roughly equivalent to the volume occupied by a PEG chain in the crystal ($\sim 5 \times 10^4 \text{Å}^3$) and it was calculated that the unit cell contained 140-180 g/L of PEG. These results suggested that the minimum volume occupied by the PEG was achieved in the crystal: a smaller volume would indicate that monomers of the PEG chain occupy the same space and this is prohibited by the excluded volume effect.\textsuperscript{186}

Finally, the conformation of a conjugated PEG chain can be assessed with the method of Svergun \textit{et al.} using the Flory dimension ($R_F$) and the distance between PEGylated sites on the surface ($D_G$).\textsuperscript{171} According to this method, the PEG polymer “fold in on itself over the grafted surface” for $D_G > R_F$ or “become extended due to steric
interactions between PEG chains” for \( D_G < R_F \). This method was hence applied to the PEGylated Pc crystal. The \( D_G \) between the cysteine side chains at the interface of monomer-monomer AC and BD is ~10 Å. This distance increases to 30-40 Å between two consecutive (AD and BC) and two crossed (AB and CD) monomers. Considering that the calculated \( R_F \) was approximately 60 Å, this suggests that the conjugated PEG chains become extended and the structure of PEGylated Pc can be described by the "dumbbell" model (Figure 1A).

In conclusion, the analysis of the crystal structure of the PEGylated Pc D45C suggests that the conjugated PEG chains are extended, occupying the space not occupied by the double-helical assembly of Pc. This structure suggests a mechanism for why, until now, it was not possible to crystallize PEGylated proteins. The space need of the conjugated PEG leads to the formation of a poorly packed structure assembly: the double-helical assembly of Pc are characterized by cavities as shown in Figure 10C. These cavities are occupied by the PEGylated portion having a molecular weight approximately half of the molecular weight of Pc. If the PEG is 20 or 40 kDa, the observed protein assembly will not be formed: the PEG would be too large to occupy the cavities and to allow the crystal structure formation of the PEGylated Pc.
3.3 - Conclusions

In this chapter the NMR and X-ray structural characterization of PEGylated Pc D45C has been described.

Chemical shift and line-width analysis of the HSQC spectra of Pc and PEGylated Pc revealed that the Pc functionalization caused changes in the chemical environment around the conjugated cysteine which did not influence the protein structure. NMR experiments in a physiological environment showed similar results for Pc and PEGylated Pc. Line-widths and chemical shift analysis revealed that there was not any Pc-protein interaction in this environment and that the conjugated PEG did not have any significant effect.

The crystal structure of PEGylated Pc was solved to a resolution of 4.2 Å. Electron density around the conjugated cysteine showed the presence of the maleimide group, while the absence of electron density for the PEG chain indicated that it was disordered and suggested a random-coil conformation. The crystal packing is remarkable: it is a double-helical assembly of a globular protein with large cavities occupied by the conjugated PEG. The calculated volume (4.6-5.8 × 10^4 Å^3) occupied by a PEG chain in the crystal was roughly equivalent or half the literature values (4.5-11.4 × 10^4 Å^3). The high concentration (140-180 g/L) of PEG in the crystal suggested that the space occupied by the PEG in the crystal can be described by packed spheres containing PEG chains. The conformation of a conjugated PEG chain assessed with the method of Svergun et al. suggested that the chains become extended and not fold on the protein surface. All these observations indicate that the PEGylated Pc structure can be described by a revised "dumbbell" model in which the PEG chain protrudes away from the Pc surface and occupies a sphere of volume ~5 × 10^4 Å^3 or folds in on itself to occupy this volume. The result is the formation of the observed protein assembly which only a conjugated PEG5000 can permit for size reasons.
3.4 - Experimental section

**NMR spectroscopy.** 0.3 mM \(^{15}\)N-labeled protein samples were prepared in 20 mM KH\(_2\)PO\(_4\), 100 mM NaCl, 10 % D\(_2\)O at pH 6.0 and 7.8. The pH was measured in Eppendorf before and after each NMR data acquisition and it was observed to be constant (± 0.05 pH). \(^{15}\)N-labelled Pc, purified from a single batch, was used for data collection in the presence of 20, 40 and 80 g/L bovine serum albumin (Sigma A4503, used without further purification) and human serum (Technopath, stored at –20 °C). The labeled protein was exchanged into water and divided into two samples of equal volume (200 µL) which were lyophilized using a centrifugal vacuum concentrator. One was dissolved in human serum and the other in the NMR buffer to which the required amount of solid BSA was added quantitatively. This analysis was repeated at the same experimental conditions with a PEGylated \(^{15}\)N-Pc sample prepared in the same way. 2D \(^1\)H-\(^{15}\)N HSQC spectra were acquired at 30 °C on a Varian 600 MHz equipped with a triple resonance cold probe. 2D spectra were obtained with 1024 and 64 points in the direct and indirect dimensions, respectively, with spectral widths of 16 ppm (\(^1\)H) and 40 ppm (\(^{15}\)N). Data were processed with NMRPipe and analyzed with CCPNmr. The chemical shift perturbation was analyzed for each backbone amide resonance by calculating the chemical shift differences (\(\Delta\delta\)) in the \(^1\)H and \(^{15}\)N dimensions. The average chemical shift perturbation was performed by using eq. (1).

\[
\Delta\delta_{avg} = \left[ (\Delta\deltaN^2/25 + \Delta\deltaH^2)/2 \right]^{1/2}
\]

(1)

In the case of duplicate peaks, \(\Delta\delta\) and \(\Delta\delta_{avg}\) values were calculated for each of the two signals separately.

**Protein crystallization.** Crystallization trials were performed by the hanging drop vapour diffusion method at 20 °C: drops were prepared by combining 1 µL of 0.2 mM protein in MilliQ water and 1 µL of reservoir solution. The PEGylated Pc structure was determined from a crystal grown in 48 % ammonium sulfate, 100 mM sodium acetate, 30 mM potassium ferricyanide, pH 6.3.
Chapter 3 - Structural characterization of PEGylated Plastocyanin

**X-ray diffraction and structure determination.** Oxidized PEGylated Plastocyanin D45C crystals were transferred to a reservoir solution supplemented with 20 % glycerol. Diffraction data were collected using the 20 µm beam at beamline I24 at the Diamond Light Source (UK) by Dr. Lutz Vogeley (TCD). Data were collected to a resolution of 4.2 Å from a single crystal. Data were automatically processed with XIA2 as implemented in CCP4\(^{190}\) using XDS for indexing, integrating, scaling and Aimless for merging. The structure of PEGylated Pc was solved by molecular replacement using Phaser\(^{191}\) with the structure of Plastocyanin (PDB ID 2W88) as the search model. Iterative model building (using Coot\(^{192}\)) and refinement (Refmac5 and Phenix\(^{193,194}\)) cycles were used. The coordinates for the maleimide were obtained from libcheck as implemented in Coot (ID LCY) and built into the model using Coot. Data collection, processing and refinement statistics are given in Table 1. Crystallographic coordinates have been deposited in the Protein Data Bank (ID 4R0O). Crystal interfaces were analyzed by using PISA\(^{184}\) and figures were prepared in PyMOL (http://www.pymol.org).

**Estimation of the PEG concentration and volume in the crystal.** The volume of a unit cell (V\(_{cell}\)) is equal to 2.3 \(\times\) 10\(^6\) Å\(^3\) and the volume occupied by PEG was calculated by multiplying V\(_{cell}\) by the solvent content (72 %)\(^{190}\) (V\(_{PEG,tot}\) = V\(_{cell}\) \(\times\) 0.72). An error of \(\pm\) 8 % was considered in this calculation. The resulting values were used to calculate the PEG concentration and the volume occupied by a conjugated PEG. V\(_{PEG,tot}\) was converted from Å\(^3\) to L (factor of 10\(^{-27}\)) and the molarity was calculated according to eq. (2). The number of molecules is 32, the number of monomers in a unit cell (Figure 11), and N\(_A\) is Avogadro’s number (6.022 \(\times\) 10\(^{23}\) mol\(^{-1}\)).

\[
M = \text{no. of molecules}/(N_A \times L) \tag{2}
\]

The molarity was converted into g/L by multiplying it by 5 \(\times\) 10\(^3\) g/mol (the molecular weight of PEG). The resulting concentration range was 140-180 g/L. The volume of a conjugated PEG (V\(_{PEG}\)) was calculated by dividing V\(_{PEG,tot}\) by 32. The resulting volume range was 4.6-5.8 \(\times\) 10\(^4\) Å\(^3\). All the volume and concentration values (Table 6) were rounded at the end of the calculation.
**Figure 11** Unit cell of the PEGylated Pc crystal which illustrates the 32 monomers. Each tetramer is coloured differently.

**Table 6** Volume and concentration values in the calculation of the PEG content in the crystal.

<table>
<thead>
<tr>
<th>V(_{\text{cell}}) (Å(^3))</th>
<th>Solvent (%)</th>
<th>V(_{\text{PEG, tot}}) (Å(^3))</th>
<th>V(_{\text{PEG, tot}}) (L)</th>
<th>[PEG](_{\text{tot}}) (M)</th>
<th>[PEG](_{\text{tot}}) (g/L)</th>
<th>V(_{\text{PEG}}) (Å(^3))</th>
</tr>
</thead>
<tbody>
<tr>
<td>64</td>
<td>1,486,022</td>
<td>1.49×10(^{-21})</td>
<td>0.0358</td>
<td>178.79</td>
<td>46,438</td>
<td></td>
</tr>
<tr>
<td>72</td>
<td>1,671,775</td>
<td>1.67×10(^{-21})</td>
<td>0.0318</td>
<td>158.93</td>
<td>52,242</td>
<td></td>
</tr>
<tr>
<td>80</td>
<td>1,857,528</td>
<td>1.86×10(^{-21})</td>
<td>0.0286</td>
<td>143.03</td>
<td>58,047</td>
<td></td>
</tr>
</tbody>
</table>
Chapter 4

Preliminary NMR Characterization of a Cysteine Rich Copper-binding Protein
Abstract

A novel cysteine rich copper-binding protein (Csp1) was produced and characterized by using NMR spectroscopy. A model of the Csp1 structure was built from the crystal structure of a homologue protein with unknown function. The structure of Csp1 consists of a four-helix bundle with 13 cysteines pointing toward the core of the protein. NMR and SEC results revealed that Csp1 is an oligomeric protein as the homologue. Multidimensional NMR experiments allowed the backbone and $C_\beta$ assignment of apo-Csp1. To aid this assignment selective amino acid (cysteine, alanine and lysine) labelling strategy was used. The chemical shift values of the assigned resonances were used to predict the secondary structure of Csp1. This result was consistent with the crystal structure of the homologue protein. Finally, Csp1 was titrated with copper and monitored by HSQC experiments to understand the mode of the binding of the copper on the basis of the model structure of Csp1. These data suggested that the copper was loaded into Csp1 through one end of the four-helix bundle.
4.1 - Introduction

In this chapter, the NMR characterization of the novel cysteine rich copper-binding protein Csp1 will be described. This protein, potentially involved in storing copper, is produced by the methane-oxidising bacterium *Methylosinus trichosporium* OB3b. Methane-oxidizing bacteria need copper for the active site of the particulate methane mono-oxygenase, an enzyme that catalyses the production of methanol from methane. *Methylosinus trichosporium* OB3b secretes two forms of methanobactins for the acquisition and delivery of copper, which structures have been described in Chapter 1.

In the following paragraphs, a brief overview of a class of small cysteine rich proteins (metallothioneins) will be presented. The focus will be on the NMR and X-ray structures of the yeast copper-metallothionein, useful for understanding the binding of the copper.

In the literature, there is a class of low molecular weight proteins (metallothioneins) found in both prokaryotic and eukaryotic organisms. A common feature of these proteins is that their sequences contain 15-30 % of cysteine residues arranged in Cys-X-Cys or Cys-X-X-Cys motifs, also present in Csp1. Metallothioneins can bind different metals such as copper, zinc and cadmium. These proteins have the highest affinity for copper(I) and may support roles in copper trafficking such as distribution or storage of this metal. In both roles, metallothioneins should be able to bind copper strongly and, for the distribution, they should quickly release copper when required. The 3D structure of metallothioneins can provide a starting model to understand the mechanism of the binding of the copper. Structural knowledge has been used to explain how metallothioneins bind copper. In the following two paragraphs, the NMR and X-ray structures of the yeast copper-metallothionein will be described.
Bertini *et al.* observed a highly refined NMR structure of the yeast copper-metallothionein which helped to understand structural features of the copper binding domain as discussed below. Ten cysteine residues were found in a region not accessible by the solvent. This feature stabilized the copper(I) by protecting it from oxidation. The cysteine sulfurs pointed towards the core of the protein and the defined position of all of the ten cysteine sulfurs allowed the arrangement of seven copper ions in the empty spaces in between them. There was not a single way to accommodate the copper ions in the empty spaces and several different arrangements of copper were obtained from DYANA calculations. All these arrangements were chemically plausible and consistent with the experimental NMR data. To explain these results, a fluctuation of the copper ions was suggested in the sense that a copper ion could change the cysteine sulphur to which it is attached.

Calderone *et al.* resolved the crystal structure of the yeast copper-metallothionein (shown in Figure 3A in Chapter 1) and suggested a mechanism for how the copper ions are loaded or released by the protein in solution as described below. The crystal structure contained ten cysteine residues bound to eight copper ions of which two are digonally coordinated to cysteines and the other six are trigonally coordinated. The two digonally coordinated copper ions are located at the opposite ends of the copper cluster. One of these copper ions is coordinated to a cysteine having a different conformation with respect to the corresponding cysteine in the NMR structure. The sulphurs of the side chain of these cysteines pointed in opposite directions when the X-ray structure was superimposed on the NMR structure. This result suggested that this copper was the copper ion missing in the NMR structure and that the cysteine pair bound to it can uptake or release the copper ion by side-chain movements. This mechanism can be applied for the other cysteine pair at the other end of the copper cluster.
Here, a preliminary NMR study of Csp1 is presented to investigate the protein structure and to learn about how the protein binds copper. Csp1 is predicted to be a four-helix bundle protein with 13 cysteine residues pointing towards the core of the monomer.\textsuperscript{195} In the PDB there is a crystal structure of a homologue protein (PDB ID 3LMF) of unknown function. The analysis with the Protein interfaces, surfaces and assemblies (PISA) service\textsuperscript{184} of this crystal structure suggested that it is a homotetramer (Figure 1A). The homologue protein contains 16 cysteine residues: the sequence identity is only 25\% and a model of the Csp1 structure was built by using SWISS-MODEL (Figure 1B).\textsuperscript{200} Csp1 was produced and characterized by using NMR spectroscopy. Multidimensional NMR experiments were recorded to determine the resonance assignment of the $^1$HN, $^{15}$N, $^{13}$C$_\alpha$, $^{13}$C$_\beta$ and $^{13}$CO nuclei for apo-Csp1. Isotopically labelled protein with $^{13}$C and $^{15}$N was overexpressed for this purpose. Moreover a selective amino acid (cysteine, alanine and lysine) labelling strategy was used to aid the assignment of the backbone resonances. The $^{15}$N- and $^{15}$N-Cys Csp1 samples were also titrated with copper and monitored by 2D $^1$H-$^{15}$N TROSY-HSQC experiment. The NMR data were used to predict the secondary structure of Csp1 from the chemical shifts of the assigned resonances and to yield a preliminary characterization of the copper binding to Csp1.
Figure 1 (A) Homotetramer of Csp1 homologue (PDB ID 3LMF). (B) Model structure of Csp1 with the 13 cysteine residues shown as spheres.
4.2 - Results and discussion

4.2.1 - Protein production for NMR analysis

$^{15}$N-labeled Csp1 was expressed in *E. coli* BL21(DE3) cells and purified by anion exchange chromatography. Figure 2 shows the Q chromatogram and the SDS-PAGE analysis of the last purification step. This gel showed a single electrophoresis band with molecular weight ~14 kDa corresponding to the mass of Csp1 (13 kDa). The ESI-TOF mass analysis revealed two major peaks in the deconvoluted mass spectrum. The experimental mass (12,734.10 Da) of the peak with a relative abundance of 100 % was in agreement with the predicted mass of $^{15}$N-Csp1 (12,734.29 Da). The second peak (12,865.59 Da) with a relative abundance of 40 % suggested the presence of the initiator methionine (mass increase of 131.30 Da).

![Figure 2](image.png)

*Figure 2* Last purification step: (A) Q chromatogram of $^{15}$N-Csp1. The protein was eluted with a linear NaCl gradient (0-250 mM) in 10 mM Tris, 1 mM DTT, pH 7.5. (B) SDS-PAGE of pooled fractions containing $^{15}$N-Csp1.

The 2D $^1$H-$^{15}$N HSQC spectrum showed a typical signal dispersion of a folded protein (Figure 3A). Pairs of resonances for three side chain amides were recognized to be in agreement with the primary structure (1 Gln, 2 Asn). Notably, most of the resonances were broader than expected for a 13 kDa protein. If Csp1 is a tetramer as the homologue protein (Figure 1A), it would have a molecular weight of 52
kDa. The 2D $^1$H-$^{15}$N TROSY-HSQC spectrum (Figure 3B) gave a first confirmation that Csp1 is an oligomeric protein. The TROSY spectrum, which is designed for higher (> 15-20 kDa) molecular weight proteins, resulted in a spectrum with greatly improved resolution. The number of resonances (112) was roughly in agreement with the expected total (118) and the observation of a single set of cross-peaks in this spectrum suggested that Csp1 is a symmetric oligomer. A TROSY-HSQC spectrum was also recorded at higher temperature (40°C instead of 30°C) giving some improvement in signal resolution and the sample remained perfectly soluble.

![Figure 3](image.png)

**Figure 3** (A) $^1$H-$^{15}$N HSQC spectrum of $^{15}$N-Csp1. The side chain amide resonances are connected by horizontal red bars. (B) $^1$H-$^{15}$N TROSY-HSQC spectrum of $^{15}$N-Csp1. Spectra recorded in 10 mM Tris, 100 mM NaCl, 1 mM DTT, 10 % D$_2$O, pH 7.5 at 30 °C.

Analysis by SEC gave a further confirmation of the homotetrameric form of Csp1 in solution (Figure 4). The chromatogram revealed a major peak at ~55 mL corresponding to particles with molecular weights > 44 kDa. This data supported the hypothesis that the protein is a homotetramer in solution because the Csp1 monomer was expected to elute in the 75-85 mL volume range.
Chapter 4 - Preliminary NMR characterization of Csp1

Figure 4 SEC chromatogram of $^{15}$N-Csp1 in 10 mM Tris, 100 mM NaCl, 1 mM DTT, pH 7.5. A major peak at 55 mL (corresponding to particles with Mw > 44 kDa) confirmed that Csp1 is a homotetramer in solution (52 kDa).

4.2.2 - Assignment of the apo-Csp1

The backbone and C$_\beta$ assignment of the apo-Csp1 was performed using TROSY-based NMR experiments. Figure 5 and 6 show regions of the HNCA, HNCO, HNCACB and CBCA(CO)NH spectra corresponding to three consecutive residues as an example to obtain sequential assignment. The connectivities obtained from such spectra are indicated with lines. The HNCA experiment correlates the intra-residue HN($i$)–N($i$)–C$_\alpha$($i$) and the inter-residue HN($i$)–N($i$)–C$_\alpha$($i$-1) connectivity. In this spectrum there are two resonances: the most intense corresponds to the C$_\alpha$ of residue $i$ and the less intense to the C$_\alpha$ of residue $i$-1 at the amine frequency of $i$. The HNCO experiment correlates the inter-residue HN($i$)–N($i$)–CO($i$-1) connectivity. In this spectrum each resonance corresponds to the CO of residue $i$-1 at the amine frequency of $i$. The HNCACB experiment correlates the intra-residue HN($i$)–N($i$)–C$_\alpha$($i$)–C$_\beta$($i$) and the inter-residue HN($i$)–N($i$)–C$_\alpha$($i$-1)–C$_\beta$($i$-1) connectivity. In this spectrum there are four resonances: the two most intense correspond to the C$_\alpha$ and C$_\beta$ of residue $i$ and the less intense to the C$_\alpha$ and C$_\beta$ of residue $i$-1 at the amine frequency of $i$. The
CBCA(CO)NH experiment correlates the inter-residue HN(i)–N(i)–Cα(i-1)–Cβ(i-1) connectivity. In this spectrum there are two resonances which correspond to the Cα and Cβ of residue i-1 at the amine frequency of i. The resonance assignment of ¹HN, ¹⁵N, ¹³Cα, ¹³Cβ and ¹³CO nuclei of the amino acids residues in the apo-Csp1 sequence was obtained with these connectivities.

The assignment of the apo-Csp1 was determined for 115 residues out of the 119 non-proline residues of Csp1. The backbone assignment was not determined for four residues in the N-terminus probably due to a relatively high exchange with solvent. In total, 98 % of ¹³Cα, 88 % of ¹³Cβ, and 95 % of ¹³CO were assigned. The Cβ chemical shifts of the cysteines (~27 ppm) were typical of reduced cysteine residues. Figure 7 shows the assigned ¹H-¹⁵N TROSY-HSQC spectrum of the apo-Csp1.

![Figure 5](image_url)

**Figure 5** Selected planes of the HNCA (green) and HNCO (purple) spectra showing an example to obtain sequential assignment. The strong and weak resonances in the HNCA spectrum correspond to the Cα of residues i and i-1, respectively. The resonance in the HNCO spectrum corresponds to the CO of i-1 residue. Horizontal and vertical lines show the connectivities obtained from such spectra. The resonance assignment of backbone ¹HN, ¹⁵N, ¹³Cα and ¹³CO nuclei was achieved with these connectivities.
Figure 6 Selected planes of HNCACB (blue/red) and CBCA(CO)NH (black) spectra. The blue and red resonances in the HNCACB spectrum correspond to the C$_\alpha$ and C$_\beta$, respectively. The resonances in the CBCA(CO)NH spectrum correspond to the C$_\alpha$ and C$_\beta$ of $i$-1 residue. Horizontal and vertical solid lines show the connectivities obtained from such spectra. The resonance assignment of $^1$HN, $^{15}$N, $^{13}$C$_\alpha$ and $^{13}$C$_\beta$ nuclei was achieved with these connectivities.
Figure 7 Assigned $^1$H-$^{15}$N TROSY-HSQC spectrum of the apo-Csp1. (A) HSQC spectrum recorded in 10 mM KH$_2$PO$_4$, 50 mM NaCl, 1 mM DTT, 10% D$_2$O, pH 7.0 at 30 °C. (B) Enlarged region of the box shown in A.
The chemical shift values of the $^1$HN, $^{15}$N, $^{13}$C$_\alpha$, $^{13}$C$_\beta$ and $^{13}$CO nuclei can be correlated with the protein secondary structure. The TALOS+ software estimates the torsion angles $\Phi$ and $\psi$ from the chemical shift of the $^1$H, $^{15}$N and $^{13}$C nuclei. The chemical shift values of a triplet $i-1$, $i$, $i+1$ are compared to empirical values from the database. If the $\Phi$ and $\psi$ values of the best ten triplets match, their average and SD are used as predicted values for the residue $i$. If the $\Phi$ and $\psi$ values do not match, no value is assigned. The TALOS+ software was used to predict the secondary structure of Csp1 from the backbone and C$_\beta$ chemical shift values. This program was run with all the protein structures included in the database. Figure 8 shows the predicted secondary structure of Csp1 as a function of the residue with the red bars corresponding to $\alpha$-helix structure and the height of the bars reflecting the probability of the prediction. Residues forming the four helices were 17-42, 47-69, 77-95 and 101-120, respectively. The predicted secondary structure of Csp1 was consistent with the crystal structure of the homologue protein (PDB ID 3LMF).

**Figure 8** Predicted secondary structure. The graph shows the probability for each residue assigned to be either $\beta$-sheet (positive) or $\alpha$-helix (negative) as determined with the TALOS+ software. Loop residues are not indicated. The height of the bars reflects the probability of the secondary structure prediction.
4.2.3 - Selective amino acid labelling

Csp1 was produced with specific amino acid labelling to aid the assignment of the TROSY-HSQC spectrum. O’Grady et al. reported a method to produce selective amino acid labelling by using prototrophic E. coli strains (labelling with cysteine was not tested). This approach involves growing the cells in the presence of the only chosen labelled amino acid. Here, in a first attempt E. coli was grown in minimal medium supplemented with only $^{15}$N-Cys, but the 2D $^1$H-$^{15}$N TROSY-HSQC spectrum (Figure 9) showed more than the expected 13 resonances. In this spectrum, in addition to the 13 cysteine residues, 50 % of the resonances matched to alanine and 20 % corresponded to other amino acids (including aspartate, glutamate and serine). This result suggested isotope scrambling (i.e. the conversion of the added labelled amino acid into others) for the labelled cysteine. In fact, the above mentioned amino acids can be interconverted in the biochemical pathways mediated by enzymes, i.e., desulfurase, synthase and transaminase.

![Figure 9](image-url)  
**Figure 9** 2D $^1$H-$^{15}$N TROSY-HSQC spectrum of Csp1 at 30 °C in 10 mM Tris, 100 mM NaCl, 1 mM DTT, 10 % D$_2$O, pH 7.5. E. coli cells were grown in the presence of 50 mg/L $^{15}$N-Cys as the sole nitrogen source.
To learn more about selective cysteine labelling, four Csp1 expression test cultures were supplemented with $^{15}$N-Cys and one unlabelled amino acid: alanine, serine, aspartate or glutamate. Figure 10 shows the 2D $^1$H-$^{15}$N TROSY-HSQC spectra of the purified Csp1 after expression by this new approach. In all these spectra, more than 13 resonances were observed and 50 % (this percentage drops to 30 % in the presence of the unlabelled alanine) of the resonances corresponded to alanine as observed previously. This result suggested that cysteine is readily metabolized into other amino acid types, especially into alanine. This latter reaction is catalysed by cysteine desulfurase and so far three different cysteine desulfurase have been identified in *E. coli*. Moreover, these data are consistent with a previous study in which it was observed that exogenous cysteine (> 10 mM) is toxic to *E. coli* and that the cell damage can be overcome by adding leucine, isoleucine, valine, and threonine to the culture.

When all 19 unlabelled amino acids were added to the minimal medium together with $^{15}$N-Cys, the TROSY-HSQC spectrum contained 13 resonances in agreement with the number of cysteine residues (Figure 11B). This sample was analysed by using ESI-TOF mass spectrometry and two major peaks were observed in the deconvoluted mass spectrum. The experimental mass (12,595.48 Da) with a relative abundance of 100 % was in agreement with the predicted mass of $^{15}$N-Cys Csp1 (12,596.69 Da). The peak with a relative abundance of 40 % and a mass of 12,727.07 Da suggested the presence of the initiator methionine (mass increase of 130.38 Da) as observed with the $^{15}$N-Csp1.

$^{15}$N-Ala and $^{15}$N-Lys Csp1 samples were also over-expressed. Figure 11 shows how the selective amino acid labelling strategy is a useful approach in simplifying the HSQC spectrum of a large protein. All resonances in the 2D TROSY-HSQC spectra of $^{15}$N-Cys, $^{15}$N-Ala and $^{15}$N-Lys Csp1 matched to the assigned resonances in the spectrum of $^{15}$N-Csp1.
Figure 10 2D $^{1}\text{H}$-$^{15}\text{N}$ TROSY-HSQC spectra of Csp1 at 30 °C in 10 mM Tris, 100 mM NaCl, 1 mM DTT, 10 % D$_2$O, pH 7.5. *E. coli* cells were grown in the presence of $^{15}$N-Cys and one unlabelled amino acid: (A) alanine, (B) serine, (C) aspartate and (D) glutamate.
Figure 11 2D $^1$H-$^{15}$N TROSY-HSQC spectra of Csp1 at 30 °C in 10 mM Tris, 100 mM NaCl, 1 mM DTT, 10 % D$_2$O, pH 7.5. (A) $^{15}$N labelled, (B) $^{15}$N-Cys labelled, (C) $^{15}$N-Ala labelled and (D) $^{15}$N-Lys labelled Csp1.
4.2.4 - Preliminary study of copper-binding

$^1$H-$^{15}$N HSQC experiments were used to characterize the binding of the copper by acquiring HSQC spectra of Csp1 after the addition of Cu(I). The apo form of $^{15}$N- and $^{15}$N-Cys Csp1 were titrated in 10 mM Tris, 100 mM NaCl, pH 7.5 with Cu(I) in the form of [Cu(CH$_3$CN)$_4$]PF$_6$. When copper was added to the apo form of Csp1, changes in the $^1$H-$^{15}$N TROSY-HSQC spectra were observed as discussed later.

Normally for a binding between protein and ligand, three different regimes of exchange between free and bound forms can be distinguished relative to the NMR time scale ($10^3$-$10^3$ s): fast, intermediate and slow. If the exchange is fast (Figure 12A) a gradual difference in the chemical shift of the resonance sensitive to the binding occurs during ligand addition. If the exchange is intermediate (Figure 12B) line-broadening of the peak occurs during the titration. If the exchange is slow (Figure 12C) two distinct peaks of different intensity appear at different chemical shifts. During a titration two changes can be hence observed: chemical shift and/or line-broadening. However, these changes can be caused either by direct interaction of the ligand at the binding site or by inducing conformational/allosteric change in the protein after the binding of the ligand. These two phenomena cannot be separated: it has been suggested that the ligand binding site can be identified from large changes in a single area and that large changes that do not map to a single area are indicators of conformational/allosteric change.

In the next two paragraphs, the titration experiments with $^{15}$N- and $^{15}$N-Cys Csp1 will be described. TROSY-HSQC spectra of $^{15}$N-Csp1 were relatively crowded after copper addition and the acquisition of TROSY-HSQC spectra of $^{15}$N-Cys Csp1 was useful in simplifying the spectrum of this large protein. Although there was no straightforward way to assign the bound spectrum from the free, a model about how the protein binds copper was proposed from the results obtained with the $^{15}$N-Cys Csp1.
Chapter 4 - Preliminary NMR characterization of Csp1

Figure 12 Representations of the three different exchange regimes between free and bound forms of a protein in the 2D $^1$H-$^{15}$N HSQC spectra during ligand (L) addition: (A) fast, (B) intermediate and (C) slow. Illustration adapted from a lecture of Prof. Cicero of Tor Vergata University, Rome.

The $^1$H-$^{15}$N TROSY-HSQC spectrum of the apo $^{15}$N-Csp1 had the greater signal dispersion with respect to copper-loaded forms (Figure 13). In the half Cu(I) loaded form line-broadening occurred and 40 % of the resonances clustered in the region 8 ($^1$H) and 120 ppm ($^{15}$N) of the HSQC spectrum. In the fully Cu(I) loaded form signal dispersion was again observed. The observation of line-broadening and chemical shift during the titration suggested a slow to intermediate exchange on the NMR time scale between free and bound Csp1 forms. This result was similar to that obtained with $^{15}$N-Cys Csp1 (Figure 14).
Figure 13 $^1$H-$^{15}$N TROSY-HSQC spectra of (A) apo, (B) half Cu(I) loaded and (C) fully Cu(I) loaded $^{15}$NCsp1 in 10 mM Tris, 100 mM NaCl, 10 % D$_2$O, pH 7.5 at 30 °C. (D) Overlaid HSQC spectra of apo (black) and fully Cu(I) loaded (green) $^{15}$NCsp1.
Figure 14 $^1$H-$^{15}$N TROSY-HSQC spectra of (A) apo, (B) half Cu(I) loaded and (C) fully Cu(I) loaded $^{15}$NCys-Csp1 in 10 mM Tris, 100 mM NaCl, 10 % D$_2$O, pH 7.5 at 30 °C. (D) Overlaid HSQC spectra of apo (black) and fully Cu(I) loaded (green) $^{15}$NCys-Csp1.
The 2D TROSY-HSQC spectrum of the apo $^{15}$N-Cys Csp1 contained 13 sharp resonances of the cysteine residues. At half titration, 19 resonances were observed. The addition of copper caused chemical shift and line-broadening of some of the 13 resonances present in the HSQC spectrum of apo-Csp1. In particular, there was small or no effect for eight resonances (belonged to Cys26, 87, 90, 94, 106, 110, 113 and 117) and large effect for five resonances (belonged to Cys33, 37, 51, 62 and 103). The HSQC spectrum of the fully Cu(I) loaded form also consisted of 13 peaks. These resonances were at different chemical shift compared to the apo-Csp1 and the line-width of the $^1$H resonances increased on average by 11 Hz with respect to the apo form (Table 1).

**Table 1** $^1$H line-width measurements of the NH resonances extracted from HSQC spectra of apo and fully Cu(I) loaded $^{15}$NCys-Csp1 in 10 mM Tris, 100 mM NaCl, 10 % D$_2$O, pH 7.5 at 30 °C.\(^a\)

<table>
<thead>
<tr>
<th>$^1$H line-width (Hz)</th>
<th>apo-Csp1</th>
<th>holo-Csp1</th>
</tr>
</thead>
<tbody>
<tr>
<td>34.8</td>
<td>47.8</td>
<td></td>
</tr>
<tr>
<td>31.6</td>
<td>39.4</td>
<td></td>
</tr>
<tr>
<td>20.0</td>
<td>50.4</td>
<td></td>
</tr>
<tr>
<td>36.0</td>
<td>48.5</td>
<td></td>
</tr>
<tr>
<td>36.9</td>
<td>37.6</td>
<td></td>
</tr>
<tr>
<td>36.2</td>
<td>40.0</td>
<td></td>
</tr>
<tr>
<td>31.4</td>
<td>54.1</td>
<td></td>
</tr>
<tr>
<td>40.1</td>
<td>51.1</td>
<td></td>
</tr>
<tr>
<td>36.5</td>
<td>48.7</td>
<td></td>
</tr>
<tr>
<td>35.1</td>
<td>40.1</td>
<td></td>
</tr>
<tr>
<td>31.6</td>
<td>34.9</td>
<td></td>
</tr>
<tr>
<td>32.5</td>
<td>55.5</td>
<td></td>
</tr>
<tr>
<td>35.2</td>
<td>35.9</td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)Average line-width (apo-Csp1) = 33.7 Hz
Average line-width (holo-Csp1) = 44.9 Hz

\(^\text{a}\) The average line-width was calculated for the 13 resonances using CCPNmr.\(^{174}\)
The titration data of the $^{15}$N-Cys Csp1 revealed that the resonances had altered chemical shifts and line-broadening upon copper addition suggesting a slow to intermediate exchange on the NMR time scale between free and bound Csp1 forms as observed with $^{15}$N-Csp1. However, the non-linearity of shifts for the cysteine residues from free to bound states might be due to an intermediate state. Williamson\textsuperscript{175} has argued that a non-linear shift from free to bound states can indicate a multiple binding interaction or a conformational change which is accompanied frequently by line-broadening. Accordingly, the non-linearity of shifts and the increased line-broadening observed during the titration of $^{15}$N-Cys Csp1 suggested that there was at least one additional state in addition to the exchange between free and bound states. The results obtained in the TROSY-HSQC spectrum of the half Cu(I) loaded $^{15}$N-Cys Csp1 suggested a model about how the protein binds copper. The five cysteines corresponding to the resonances that experienced large changes after copper addition were highlighted in the model structure of Csp1 (Figure 15). Four cysteines (33, 37, 51 and 103) were found to be grouped to one end of the four-helix bundle. The Cys62 was in the middle of the bundle far from the other four. A reasonable model that explains the observed results is presented below. The four grouped cysteines are probably the residues that bind the copper ions before the other cysteines suggesting that the copper was loaded from one end of the bundle. The large changes observed for the resonance of Cys62 are probably caused by a conformational change of the helix 47-69 (in which there is also Cys51) after the binding of the copper.
Figure 15 Structure model of Csp1 in the (A) apo, (B) half and (C) fully Cu(I) loaded forms. The cysteines corresponding to the resonances that experienced large changes after copper addition are shown in green.
4.3 - Conclusions

In this chapter the first NMR characterization of the novel copper binding protein Csp1 has been described. $^{15}$N-, $^{13}$C-$^{15}$N-, $^{15}$N-Cys, $^{15}$N-Ala and $^{15}$N-Lys Csp1 labelled samples were produced. The selective amino acid labelling strategy was performed using E. coli BL21 and not auxotrophic strains, which are used normally. The ESI-TOF mass analysis confirmed the predicted molecular weight of the apo-Csp1 and suggested the presence of a minor species containing the initiator methionine. Both HSQC spectra and SEC data indicated that the protein is a homotetramer in solution. The resonance assignment of $^1$HN, $^{15}$N, $^{13}$C$_\alpha$, $^{13}$C$_\beta$ and $^{13}$CO nuclei of the amino acids residues in the apo-Csp1 were successfully achieved with multidimensional TROSY-based experiments. The predicted secondary structure by using the experimental chemical shift values was consistent with the crystallographic secondary structure of the homologue protein. The preliminary results of the Cu(I)-titration of $^{15}$N- and $^{15}$N-Cys Csp1 samples suggested a slow to intermediate exchange on the NMR time scale between free and bound forms. The non-linearity of shifts and the increase of line-broadening in the HSQC spectra of $^{15}$N-Cys Csp1 after copper addition suggested another state in addition to the exchange between free and bound states. A model about how the protein binds copper was also proposed: according to the NMR results, the copper ions were loaded through one end of the four-helix bundle.
4.4 - Experimental section

**Protein production.** The plasmid pet29a for Csp1 was kindly provided by Prof. Dennison (Newcastle University, UK) and the protein was over-expressed in *E. coli* BL21(DE3).

$^{15}$N-, $^{15}$N-Cys, $^{15}$N-Ala and $^{15}$N-Lys Csp1 samples were over-expressed by auto-induction\(^\text{124}\) using a two-step protocol. Cells were grown in 0.5 L LB medium to OD\(_{600}=1.2\), harvested, resuspended in 1 L of minimal medium (MM) and grown for 24 hours at 30°C. The MM culture for $^{15}$N-Csp1 contained (\(\text{NH}_4\))\(_2\text{SO}_4\) as the sole nitrogen source. The MM cultures for the selective amino acids labelling were grown in the presence of 50 mg/L Cys and 100 mg/L of all other amino acids and the minimal medium was depleted of $^{15}$N source, except for the presence of the corresponding $^{15}$N-amino acid.\(^\text{127}\)

$^{13}$C, $^{15}$N-Csp1 was over-expressed using the two-step protocol in minimal medium containing (\(\text{NH}_4\))\(_2\text{SO}_4\) and \(^{13}\text{C}\)-glucose as the sole nitrogen and carbon source respectively and incubated for one hour prior to the addition of 1 mM IPTG. Culture growth was for 24 hours at 30°C.

The purification was performed by anion exchange chromatography (Q column) according to an established protocol from the group of Prof. Dennison. Csp1 was purified in two steps using a Q column. First step: linear gradient of 0-400 mM NaCl in 20 mM Tris, 1 mM DTT, pH 8.5. Second step: linear gradient of 0-250mM NaCl in 10 mM Tris, 1 mM DTT, pH 7.5. Protein concentration was estimated with 5,5′-dithiobis(2-nitrobenzoic acid) (DTNB, Ellman’s reagent) under denaturing conditions assuming an extinction coefficient of $\varepsilon_{412}=13.7 \text{mM}^{-1}\text{cm}^{-1}$ for the TNB\(^2-\) product.\(^\text{211}\) Protein samples were analysed by electrophoresis on 15 % (w/v) polyacrylamide gels followed by Coomassie Brilliant Blue staining. Protein samples were desalted for the ESI-TOF mass spectrometry analysis. Size-exclusion chromatography was performed with a Superdex 75 column (65 cm bed height) using buffer containing 10 mM Tris, 100 mM NaCl, 1 mM DTT, pH 7.5 as an eluent.

**NMR spectroscopy.** All spectra were acquired at 30 °C on a Varian 600 MHz equipped with a triple resonance cold probe, processed with NMRPipe and analyzed with CCPNmr.\(^\text{174,189}\) Backbone resonance assignments of the apo-
Csp1 were achieved with 3.5 mM $^{13}$C,$^{15}$N-Csp1 in 10 mM KH$_2$PO$_4$, 50 mM NaCl, 1 mM DTT, 10 % D$_2$O, pH 7.0. The following TROSY experiments were recorded. The HSQC spectrum was acquired with 1024 ($^1$H) and 64 ($^{15}$N) complex points with spectral widths of 19.2 and 42 ppm, respectively. The HNCA spectrum was acquired with 1024 ($^1$H), 48 ($^{13}$C) and 64 ($^{15}$N) complex points with spectral widths of 19.2, 36 and 42 ppm, respectively. The HNCO spectrum was acquired with 1024 ($^1$H), 32 ($^{13}$C) and 64 ($^{15}$N) complex points with spectral widths of 19.2, 13.2 and 42 ppm, respectively. The HNCA CB and CBCA(CO)NH spectra were acquired with 1024 ($^1$H), 64 ($^{13}$C) and 64 ($^{15}$N) complex points with spectral widths of 19.2, 73.2 and 42 ppm, respectively.

The Cu(I)-titration was performed by stepwise addition of 8 µL of freshly prepared 100 mM [Cu(CH$_3$CN)$_4$]PF$_6$ to ~0.6 mM Csp1 sample in 10 mM Tris, 100 mM NaCl, 10 % D$_2$O, pH 7.5. The pH was measured in Eppendorf before and after each NMR data acquisition and it was observed to be constant (± 0.05 pH). The NMR data was collected after five minutes from pH measurement. The 2D $^1$H-$^{15}$N TROSY-HSQC spectra were obtained with 1024 and 64 complex points in the direct and indirect dimensions, respectively, with spectral widths of 16 ppm ($^1$H) and 40 ppm ($^{15}$N).
Chapter 5

Discussion
Cysteine chemistry was the common topic of the two main projects presented in this thesis. Each project underlined two features of the cysteine in proteins: the protein functionalization and the copper-binding. The analysis of these two features were performed on Plastocyanin (Pc) and Csp1, respectively. Pc, one of the most characterized proteins, is a useful model for structural biology.\textsuperscript{185,212,213} The native fold of Pc results in a hydrophobic and a hydrophilic surface patches which influence protein-protein interactions with its natural partners. Pc can also form homodimeric complexes characterized by flat interfaces with a low geometric complementarity which promotes non-specific binding and fast dissociation. Csp1, a novel cysteine rich copper-binding protein, has been characterized by using NMR for the first time and the preliminary results are reported herein. This protein binds copper strongly and it is potentially involved in storing this metal. Similarly, low molecular weight cysteine-rich proteins (metallothioneins) that have high affinity for copper can be found in literature.\textsuperscript{81,198} In the next paragraphs the main findings for each project are provided and discussed.

The optimization in the production of two Plastocyanin mutants with an engineered solvent exposed cysteine residue was successfully achieved (Chapter 2). This was important for obtaining the target protein following the first purification attempts which had revealed the post-translational modification of cysteine with glutathione and the co-purification of β-lactamase. The covalent modification of the solvent exposed cysteine was tested with nine thiol reactive compounds: iodoacetic acid, two iodoacetamide compounds, four maleimide derivatives, an aziridine and a thiol containing reagents. The reactivity of cysteine was shown by the reaction with iodoacetic acid prior to performing the conjugation reactions with the other compounds. These compounds were chosen for their distinctive chemical characteristics: they are distinguished by rigid and flexible substituents with hydrophobic and hydrophilic properties which were expected to enhance
the possibilities for molecular recognition.\textsuperscript{111,112} ESI-TOF mass spectrometry confirmed Pc functionalization via cysteine with all compounds. Among the tested reactions, the maleimide derivatives gave the better results with the formation of 70-99 % of conjugated Pc. In contrast, the iodoacetamide derivatives gave the worst result with the formation of 20-30 % of conjugated Pc: this result was probably due to the low solubility of these compounds in water with also a hindrance effect of the substituents on the amine group as suggested by the formation of the alkylated product with a cysteine sample. SEC was performed to analyse the elution profile of the conjugated Pc with ethylphenyl maleimide and with dihexyl iodoacetamide. The early elution volume of the conjugated Pc with ethylphenyl maleimide suggested the formation of protein aggregates. The late elution volume of the conjugated Pc with dihexyl iodoacetamide was indicative of an interaction between the dihexyl substituent and the hydrophobic resin of the column. Finally, the formation of PEGylated Pc was confirmed by SEC, SDS-PAGE and ESI-MS analysis. The early SEC elution volume was caused by the hydrodynamic volume of the conjugate: the elution profile was not dependent on the PEGylation site and on the presence of 200 mM salt (MgCl\textsubscript{2}, MgSO\textsubscript{4}, NaCl, Na\textsubscript{2}SO\textsubscript{4}, NH\textsubscript{4}Cl or (NH\textsubscript{4})\textsubscript{2}SO\textsubscript{4}). The slow migration on the SDS-PAGE was caused by the interaction between SDS and PEG. The ESI-TOF mass spectrum showed a large number of fragments due to the presence of the conjugated PEG chain.

The next objective was to obtain structural characterizations of functionalized Pc by using X-ray crystallography. The first crystal structure was obtained with Pc conjugated with ethylphenyl maleimide to a resolution of 2.65 Å. This conjugate was crystallized using similar crystallization conditions of the wild-type Pc\textsuperscript{85} and the crystal packing did not change because the maleimides occupied a space that was available in the wild-type Pc crystal with the zinc ions being the
dominant contributors to the packing. More interesting results were obtained with the first crystal structure of PEGylated Pc (Chapter 3).

The crystal structure of PEGylated Pc was solved at a resolution of 4.2 Å and the asymmetric unit comprised four monomers of Pc. A confirmation of the reasonable solution after molecular replacement was the electron density of copper in the binding site. Electron density around the functionalized cysteine showed the maleimide modification and the absence of the electron density for the PEG suggested the disordered conformation of the polymeric chain. Space group symmetry expansion revealed a right-handed antiparallel double-helical assembly of Pc. The interfaces of monomer-monomer contacts were poorly packed with interface areas (ranging from 200 to 460 Å\(^2\)) similar to the contact areas found for transient cupredoxin dimers (ranging from 300 to 500 Å\(^2\)).\(^{184,185}\) The crystal structure was therefore poorly packed with large cavities occupied by the disordered PEG chain. The following observations arose from the analysis of the X-ray structure. The calculated volume (4.6-5.8 × 10\(^4\) Å\(^3\)) occupied by a PEG chain in the crystal was roughly equivalent or half the literature values (4.5-11.4 × 10\(^4\) Å\(^3\)).\(^{119,170,171,186-188}\) The high concentration (140-180 g/L) of PEG in the crystal suggested that the space occupied by the PEG in the crystal was described by packed spheres containing PEG chains.\(^{148,186}\) The conformation of a conjugated PEG chain assessed with the method of Svergun et al.\(^{171}\) suggested that the chain become extended and not fold on the surface of Pc. All these observations suggested a "dumbbell" model (Figure 1A, Chapter 3) to describe the PEGylated Pc structure: the PEG chain protrudes away from the Pc surface and occupies a sphere of volume \(~5 \times 10^4\) Å\(^3\) or, by revisiting the model, the extended PEG chain folds in on itself to occupy a sphere of volume \(~5 \times 10^4\) Å\(^3\). The result is the formation of the observed protein assembly which only a conjugated 5 kDa PEG can permit for size reasons: a PEG of 20 or 40
kDa would be too large to occupy the cavities and to allow the observed crystal structure assembly of the PEGylated Pc.

The PEGylated Pc was also characterized by using NMR spectroscopy. This analysis did not show an assembled state of PEGylated Pc. The modest line-width broadening (increased on average by 1 Hz) of PEGylated Pc as compared to Pc suggested independent motions of the protein and the conjugated PEG chain. The residues located around the modified cysteine experienced a significant chemical shift perturbation due to the Pc functionalization. To understand if the conjugated PEG could influence protein interactions in a physiological environment, HSQC spectra were acquired in human serum and in the presence of BSA, the homologue (76 % of sequence identity) of the human albumin which is the most abundant protein in serum. Similar minimal effects on chemical shift perturbations (< 0.04 ppm) and line-widths (increased on average by 1-5 Hz) were observed for both PEGylated Pc and Pc in the physiological environment, suggesting that there was not any Pc-protein interaction in this environment and that the conjugated PEG did not have any significant effect. The largest chemical shift perturbations were observed for His24, His61, Gln63 and Gly105, which are close each other (pairs His24-Gly105 and His61-Gln63) in the 3D structure. The side chains of His24 and His61 have a pKₐ near the experimental pH (equal to that measured in the human serum sample), hence protonation/deprotonation equilibrium of the imidazole group caused the observed chemical shift perturbations. The line-width analysis suggested that in the physiological environment there was not a binding event like in the complex with cytochrome f, the natural partner of Plastocyanin, resulting in an increase of the line-width of 26-27 Hz.

In the second main project, the structure of the novel cysteine rich copper-binding protein (Csp1) was modelled on the basis of a homologue protein (PDB ID 3LMF). Csp1 was predicted to be a tetrameric (52 kDa) four-helix bundle in its apo form (Chapter 4). SEC and HSQC data
supported this model structure. The SEC chromatogram revealed a major peak corresponding to particles with molecular weight > 44 kDa. The acquisition of a TROSY-HSQC spectrum, which is designed for large proteins (> 15-20 kDa), showed well resolved resonances as compared to the standard HSQC spectrum. Moreover, a single set of NH resonances indicated that Csp1 is a symmetric homotetramer in solution. The backbone and Cβ resonances assignment of the apo-Csp1 was achieved by using multidimensional TROSY-based experiments with a 13C, 15N-Csp1 sample. To aid the assignment of the 2D 1H-15N TROSY-HSQC spectrum, selective amino acid (15N-Cys, 15N-Ala or 15N-Lys) labelling samples were successfully produced using E.coli BL21 and not auxotrophic strains, which are used normally. The empirical chemical shifts were used to predict the secondary structure by using TALOS+ software which revealed four helix regions formed by residues 17-42, 47-69, 77-95 and 101-120, respectively, in agreement with the model structure. The preliminary titration results with 15N- and 15N-Cys Csp1 suggested a slow to intermediate exchange from free to bound states on the NMR time scale (10^{-3}-10^{-3} s). The non-linearity of shifts and the increase of line-broadening in the HSQC spectra of 15N-Cys Csp1 after copper addition suggested another state (for example a conformational change or a fluctuation of the copper during the titration) in addition to the exchange between free and bound states. Although there was no straightforward way to assign the bound spectrum from the free spectrum, a model about how the protein binds copper was proposed by mapping the resonances that experienced large changes onto the model structure of Csp1. According to this analysis, the copper ions were loaded through one end of the four-helix bundle.

Overall, in biological systems the cysteine residue is mainly involved in the stabilization of proteins structure and in the participation of catalytic and electron transfer reactions. The assembly of viral capsids, the cell regulation and the binding of biologically relevant
metals such as copper are the main important processes assisted by biological cysteine.

In biological systems, the post-translational modifications via cysteine can modulate protein-protein interactions. In the same way, the protein functionalization via cysteine has been used to develop protein nanostructures exploiting protein-protein interactions. In the first project, the study was focused on the Plastocyanin functionalization via cysteine. This approach is important for the emerging molecular devices which are exploiting protein self-assembly systems. These supramolecular structures are promoted by the covalent modification of the protein surface or by the non-covalent interaction between proteins and small molecules. In this regard, crystal structures are essential to explore protein-protein interactions and determine molecular recognition. The first X-ray structure of a PEGylated protein permitted the study of the contribution of a conjugated PEG chain to the protein assembly. This research underlined the ability of a heterogeneous system to self-assembly. Protein and polymer hybrid systems combined the propriety of high structure organization of the protein component to that of PEG. This assembled structure might help guide the generation of new supramolecular protein assembly.

In biological systems, the binding of the copper via cysteine can stabilize the protein structure, the copper can be transported/stored or participate in electron transfer reactions. This binding has been studied to learn more about the copper uptake and regulation in the organisms. In the second project, the study was focused on the preliminary NMR characterization of the novel cysteine rich copper-binding protein (Csp1). This project originated from the discovery of Csp1 by the group of Prof. Dennison (Newcastle University, UK) in Methylosinus trichosporium OB3b, a methane-oxidizing bacterium that needs copper in the active site of the enzyme for the methanol
In this organism, the acquisition and delivery of copper are regulated by methanobactins and the novel Csp1 is potentially involved in storing this metal. The model structure of Csp1 (Figure 1B in Chapter 3) is similar to the crystal structure of the yeast copper-metallothionein (Figure 3A in Chapter 1) and, similarly, the copper ions are loaded to one end of the bundle with a probable fluctuation of the copper ions after the copper loading. \(^{83,199}\)
Bibliography


(195) Biotechnology and Biological Sciences Research Council (BBSRC). The essential role of copper in bacterial methane oxidation: characterising a novel copper protein involved in storage and the soluble copper proteome. (http://www.bbsrc.ac.uk/pa/grants/AwardDetails.aspx?FundingReference=BB/K008439/1)


Acknowledgements

I wish to thank **all of the people** for their beneficial guidance and **Science Foundation Ireland** which funded my Ph.D.

Firstly, I am extremely grateful to my supervisor **Dr. Peter Crowley**. Thank you so much for your confidence in me and your support. Your advice helped me in the development and improvement of my research skills.

I am thankful to the **Crowley group members** for the useful discussions and for this I am very appreciative.

I am also grateful to **Dr. Róisín Doohan, Marian Vignoles and Ger Fahy** at the School of Chemistry for their assistance with the instrumentations and to **Dr. Ger Wall** at the Department of Microbiology who allowed me to use the microscope.

I am very thankful to **Dr. Lutz Vogeley** at the Trinity College Dublin for his help with crystal harvesting and data collection of the PEGylated protein.

I am very grateful to **Prof. Christopher Dennison** at the Newcastle University (UK) for the collaboration on Csp1 project and for his guidance.

I am also very grateful to **Prof. Nico van Nuland** at the Structural Biology Research Center in Brussel for his assistance with the NMR experiments.

Finally, I am very grateful to my **family and friends**. Thank you for your support, you have encouraged me to work on my project and follow my goals.
Giada Cattani was born on August the 13\textsuperscript{th} 1985 in Rome, Italy. In May 2010 she received the Laurea Specialistica (equivalent to M.Sc. degree) in Chemistry with 110/110 \textit{cum laude} in the University of Rome “Tor Vergata”. During her academic period in Rome, Giada worked in the organic lab with Prof.ssa Barbara Floris and in the NMR lab with Prof. Daniel O. Cicero. The projects focused on the oxidation reaction of alcohol with hydrogen peroxide catalysed by scandium triflate and the structural characterization of the HCV NS3 protease with the introduction of residual dipolar couplings. In September 2010 Giada joined the lab of Dr. Peter B. Crowley at the National University of Ireland, Galway to pursue a Ph.D. During this period, she followed several workshops and chemical research modules. She supervised students and undergraduate projects in the chemistry labs. Giada also became a Biochemical Society member in April 2012. In June 2012, she attended the FEBS course “Advanced Methods in Macromolecular Crystallization V” in Nove Hrady (Czech Republic), for which she was awarded the YTF grant. She attended the EUROMAR conference in Dublin (July 2012) and the TBSI symposium/NMR facility opening in TCD (April 2013). Giada presented posters of her work during these events. Her Ph.D. research involved the functionalization and characterization of Plastocyanin mutants and the characterization of a novel protein in collaboration with Prof. Christopher Dennison of Newcastle University, UK. The results of this research are presented in this thesis.