Bangor University

DOCTOR OF PHILOSOPHY

Probing factors controlling ligand binding and discrimination in Haem Proteins: Cloning, mutagenesis, spectroscopic and crystallographic studies of cytochrome c' from Alcaligenes xylosoxidans

Abdullah, Gaylany

Award date:
2010

Awarding institution:
Bangor University

Link to publication

Probing Factors Controlling Ligand Binding and Discrimination in Haem Proteins:

Cloning, Mutagenesis, Spectroscopic and Crystallographic Studies of Cytochrome c’ from *Alcaligenes xylosoxidans*

A thesis presented for the degree

of

Doctor of Philosophy

in the

School of Chemistry

by

Gaylany Hikmat Abdullah

PRIFYSGOL
BANGOR
UNIVERSITY

1884

Bangor University

Abstract:

A recombinant DNA vector was generated by cloning the AXCP gene into the pET26b(+) plasmid. Site directed mutagenesis was utilised to generate the desired mutants using the recombinant DNA plasmid as a template. The recombinant and all mutants were successfully expressed in the *E. coli* strain, BL21(DE3) and purified to not less than 95% using cation exchange chromatography.

At the haem binding site leucine in position 16 (L16) sits above the haem in the distal site blocking direct access to the unoccupied sixth coordination site of the haem. This residue is believed to a vital role in ligand binding and discrimination. To investigate the role of this residue, L16 was mutated to glycine (L16G), phenylalanine (L16F) and to tyrosine (L16Y). The side chain of tryptophan at position 56 (W56) blocks the solvent channel observed in type I cytochromes c′ which is the shortest route from the protein surface to the sixth coordinate position at the distal face of haem. This residue was mutated to glycine (W56G) to unblock the solvent channel and to tyrosine (W56Y) to investigate whether the side chain of tyrosine will stabilise the 6c adducts by providing H-bonding to the exogenous diatomic molecules upon their binding to the distal side of the haem. A double mutant; (L16F/W56G) was generated for further investigation of the effect of the channel.

Except for L16G, the ferric form of recombinant AXCP and all other mutants were sensitive to change in the solvent pH. At alkaline pH, the spin state changed from the quantum-mechanically admixed high-spin/intermediate-spin ($S={}^5/_2, {}^3/_2$) state to the pure high-spin ($S={}^5/_2$).

The spectroscopic studies of NO titration with the ferrous forms of proteins showed that the recombinant AXCP has formed a stable 5c-NO via the transient 6c-NO adduct. Except for L16G, all mutants have formed stable 5c-NO adducts but the transient 6c-NO adducts were undetectable in the timescale of the spectroscopic experiments. Among them, W56Y exhibited the highest reactivity towards NO. Interestingly, L16G generated a stable 6c-NO adduct accompanied by an enhanced reactivity towards the NO. This is the first reported 6c-NO adduct for AXCP.

The ferrous forms of recombinant, L16F, L16Y and the double mutant L16F/W56G were oxidised upon the CO titration and did not bind CO. This was attributed to the decrease in the oxidation potential of these mutants upon the CO addition. Therefore, the excess concentration of reductant was kept in the reaction mixture to keep the oxidation potentials high enough preventing the oxidation. In this case all of them have formed a stable 6c-CO adducts. Other mutants; L16G, W56G and W56Y generated the 6c-CO adducts. These results show the effect of W56 and L16 on the oxidation potentials.

The crystal structures of recombinant AXCP at 1.15 Å, L16G at 1.18 Å, L16Y at 0.98 Å, and the double mutant L16F/W56G at 0.96 Å resolution were determined. The crystal structure of the recombinant AXCP exhibited a high degree of similarity to that of native protein. The overall structures of the mutants were similar to the recombinant AXCP. In contrast to the crystal structures of recombinant and all the other mutants, L16G was already in the reduced form and 6c with histidine and a water molecule in the fifth and sixth coordination positions to haem. This is the first reported structure of AXCP that is 6c without the introduction of any exogenous ligands.

Table of contents

Chapter 1 Introduction and Literature Review

Chapter 2 Protein Crystallography

Chapter 3 Materials and Methods

Chapter 4 Cloning, expression, purification and generation of cytochrome c' mutants

Chapter 5 Electronic Spectroscopy studies of recombinant cytochrome c` and Mutants

Chapter 6 The crystal structures

List of Figures

CHAPTER 1

CHAPTER 2

CHAPTER 3

CHAPTER 4

CHAPTER 5

CHAPTER 6

List of Tables

Abbreviations

5c	Penta-coordinate
6c	Hexa-coordinate
Å	Angstrom
Amp	Ampicillin
AX	Alcaligenes xylosoxidans
AXCP	Alcaligenes xylosoxidans
bp	Base pare
CAPS	*N*-cyclohexyl-3-aminoprpanesulfonic acid
CCHL	Cytochrome c haem lyase
Ccm	Cytochrome c maturation
CCP4	Collaborative computational project No. 4
cGMP	Cyclic guanosine monophosphate
CLS	Canadian light source
Cm	Chloramphenicol
CT	Charge transfer
cycP	Cytochrome c
Da	Dalton
dNTP	Deoxyribonucleotide triphosphate
DT	Dithionite
EPR	Electronic paramagnetic resonance
FT	Fourier transform
GTP	Guanosine triphosphate
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
Hmc	High molecular weight cytochrome
H-NOX	Haem-nitric oxide and/or oxygen-binding domain
HS	High spin state
IPTG	Isopropyl-β-D-thiogalactopyranoside
kDa	Kilo Dalton
Km	Kanamycin
LB	Luria Bertani broth
MAD	Multiple-wavelength anomalous dispersion
MCD	Magnetic circular dichroism
MES	2-Morpholinoethanesulfonic acid
MHTPC	MES, HEPES, TAPS, PIPES, CAPS cocktail buffer

MIR	Multiple isomorphus refinement
MOSP	3-(N-morpholino)propanesulfonic acid
MR	Molecular replacement
NCBI	National centre for biotechnology information
NCIMB	National collection of industrial, marine and food bacteria
NMR	Nuclear magnetic resonance
OD	Optical density
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PDB	Protein data bank
PEG	Polyethyleneglycol
PEG	Polyethylene glycol
pelB	*E. coli* sequence leader
pET	Plasmid for expression by T7 RNA polymerase
Ps	Picoseconds
QS	Quantum mechanically admixed high-spin/intermediate spin state
r.m.s	Root mean square
r.m.s.d	Root mean square deviation
RMS	Root mean square
RMSD	Root mean square deviation
SAD	Single-wavelength anomalous dispersion
SDS	Sodium dodecyl sulphate
Sec	Type II secretion pathway
sGS	Soluble guanylate cyclase
SLS	Swiss light source
SR	Synchrotron radiation
SRS	Synchrotron radiation source
TAPS	N-Tris(hydroxymethyl)methyl-3-aminopropanesulfonic acid
TB	Terrific broth media
TEMED	*N*-*N*-*N*`-*N*`-tetramethylethylenediamine
T_m	Melting temperature
UV	Ultra violate
Vis	Visible

Acknowledgments

I would like to thank many people who have helped me over the past few years. I am indebted to my supervisor Dr. Loretta M. Murphy for her unfailing encouragement and advice throughout my PhD. Without her support and optimism, the accomplishment of this thesis would have not been possible. Thanks due to Prof. Peter Williams who let me work in the molecular biology laboratory in the School of Biological Sciences during the course of the lab work. Special thanks go to the School of Chemistry staff for their support, especially Gwynfor Davies, Mike Lewis and John Charles.

I am extremely grateful for the continuous support of Prof. Samar Hasnain. He provided me with many helpful suggestions, important advice and constant encouragement during the course of this work. Without you I was still far from the cutting-edge and exciting science "protein crystallography", thank you for letting me work within your group at both Daresbury Laboratory and Liverpool University, thank you for every second of your valuable time you spent with me. Special gratitude goes to Prof. Robert Eady for his pleasant scientific discussions especially his comments and feedbacks on the spectroscopy chapter, which indeed improved the quality of this thesis. Special thanks go to Dr. Svetlana Antonyuk who taught me advanced in crystallography and helped in the crystallographic data collection and analysis. Sincere thanks to Dr. Richard Strange for taking intense scientific interest in this study. His valuable suggestions and instructions have certainly made the quality of this thesis better. Many thanks to Dr. Mark Ellis whose patience made me taste the magnificent of the high tech world. Thanks due to Dr. Mike Hough for his feedback and suggestions on the crystallography chapter. My heartiest thanks go to my parents, who helped me financially and encouraged me to concentrate on my study.

Finally I would like to express a very special thank to my wife Suhad Asa'd. Although she was a molecular biology PhD student herself, she has helped me to concentrate on completing this thesis and supported me mentally during the course of this work. Suhad, you have been a very supportive wife and a kind mother to our children Talan and Muhammad. Without your help and encouragement, this study would not have been completed. I thank you for everything.

Chapter 1: Introduction and Literature Review

1.1 Haemoproteins:

Haemoproteins are an extensive group of proteins whose members widely differ in function, structure, and other physicochemical properties and are essential in variety of biological functions.[1] These functions include ligand transport,[2] electron transfer,[3,4] and the metabolism of substrates.[5] Haemoproteins also act as biosensors or signal transducers.[6,7,8] They are found to function in gene regulation[9] and play a role in programmed cell death (apoptosis) process.[10,11]

Haemoproteins have various topologies and folds. All haemoproteins have in common that they contain an iron-containing porphyrin macrocycle prosthetic group. They may contain one or more porphyrin cofactors (haem). For example, cytochrome c′ from *Rhodocyclus gelatinosus*[12] is a *mono*haem protein, while the c_7-type cytochrome is *tri*haem[13], NapC/NirT family[14] are *tetra*haem proteins, *octa*haem hydroxylamine oxidoreductase[15] have eight haem groups, and the high molecular mass cytochrome c (Hmc) is a 16-haem-containing protein.[16] Haemoproteins may also contain other types of prosthetic group e.g. the Cu_A and Cu_B sites in cytochrome c oxidase.[17] The ubiquitous nature of haemoproteins, their wide diversity in function and structure, and their intense colour, which allows for easy identification and application of diverse spectroscopies, has entertained scores of scientists for the last century.

Haem or iron-porphyrin complexes are the versatile and ubiquitous active centres of haemoproteins.[18] Haems are characterised by their four five-membered pyrrole rings. The four rings of the macrocycle are labelled clockwise A-D starting with the first two symmetric rings with regard to the ring substituents (see figure 1.1). The four pyrrole nitrogens of the haem macrocycle bind iron strongly.[19]

According to the porphyrin macrocycle substitutions, most naturally occurring haems can be classified into three groups, type-a, type-b, and type–c haems. 5-aminolevulinic acid is the first intermediate in the haem biosynthesis pathway and a precursor of protoporphyrin IX. Ferrochelatase enzymes then incorporates Fe(II) into the protoporphyrin IX to generate protohaem IX, or haem b, which is the most basic

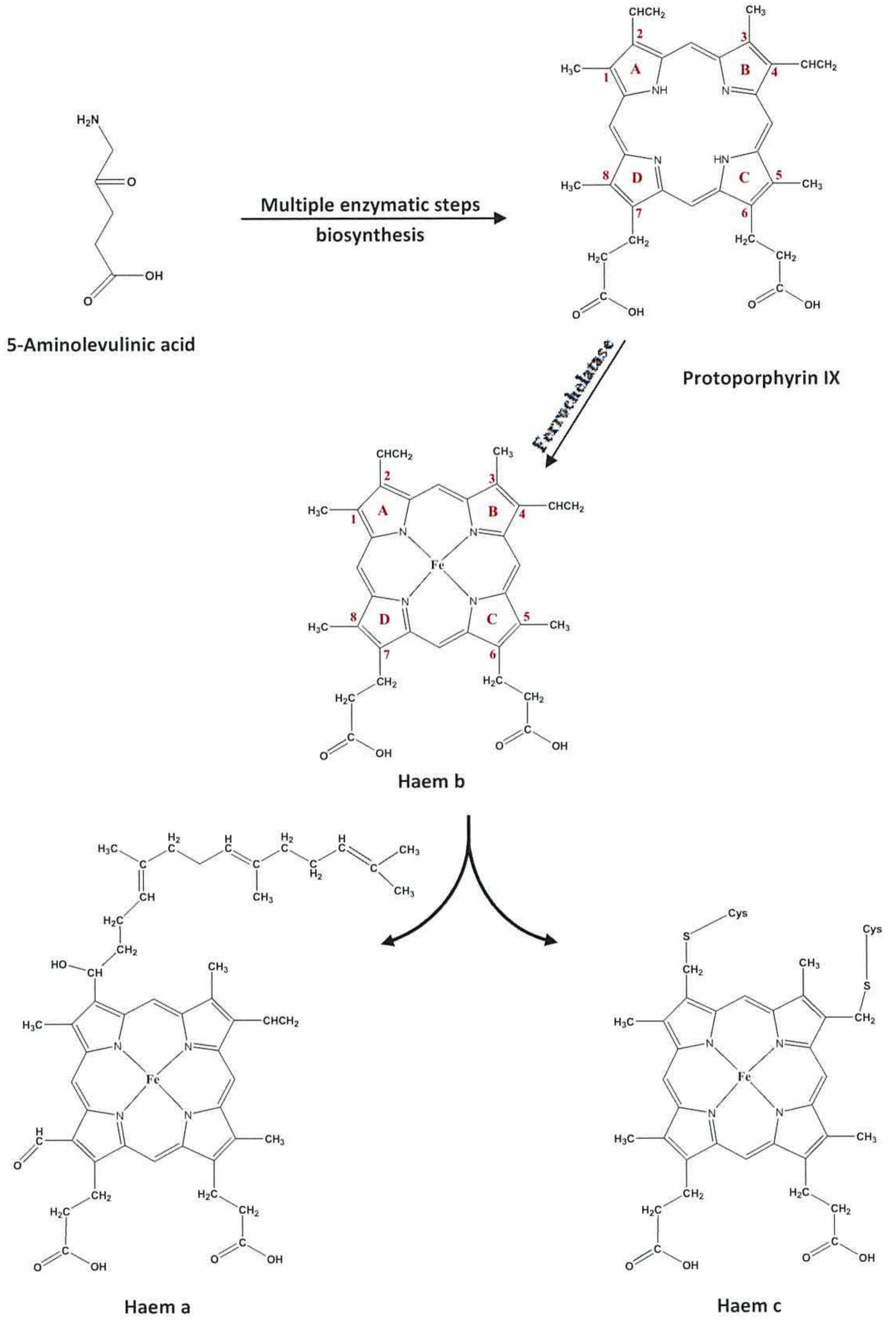


Figure 1.1: Chemical structures of haem a, b, c, and d. Protoporphyrin IX synthesised from the precursor 5-aminolevulinic acid and Fe(II) is consequently added by ferrochlatase to form a protohaem IX (haem b). Haems a, c, and d are derivatives of protohaem. Structures were drawn using ChemDrawUlta 7.0

haem.[20,21,22] Haem b has four methyl groups in positions 1, 3, 5, and 8, two vinyl groups in positions 2 and 4, and two propionic acid groups in positions 6 and 7 (Figure 1.1).

In haem a, the methyl side chain at position number eight of haem b is oxidised to a formyl group, and a hydroxylfarnesylethyl group has been attached to the vinyl group at position number two.[23] The structure of haem a was not determined until the mid-1970s[24] since it is significantly larger, more hydrophobic and less stable than naturally abundant protohaem.[25]

Haem c, such as found in c-type cytochromes, is formed when only one or both vinyl substituents in positions 2 and 4 are used to form covalent linkage to the apoprotein by addition of the thiol function of cysteine residues, such that two thioether bridges are formed. This introduction will focus on a c-type cytochrome which is the subject of this thesis.

1.2 Cytochromes:

Cytochromes (meaning "cellular pigments") are commonly utilised in electron transfer reactions. The valence charge of the iron atom in the haem co-factor determines their redox activity (Fe^{3+} to Fe^{2+}).[26] The cytochromes are usually named after the type of haem they contain, such as cytochromes a, cytochromes b, and cytochromes c. The numerical subscript next to the haem type is used for distinguishing different proteins with the same type for example; cytochromes a_1 and a_3 used to distinguish between a-type cytochromes,[27,28] b_1 and b_2 used to distinguish between b-type cytochromes,[29,30] in the same manner c_1, c_3, used to distinguish between c-type cytochromes.[31,32] Some cytochromes name are based upon the α-band wavelength such as cytochrome c-554.[33] The iron ion in all electron transfer cytochromes, except for cytochrome c′, is hexa-coordinate (6c). Independent of the haem type, the haem iron axial ligand varies, but the most common ligand is histidine (Table 1.1).

Proteins	Haem type	Axial ligand	Function
Cytochromes c			Electron transfer
Class I	c	His/Met	
Class II	c	His/Met	
Class III	c	His	
Class IV	c	His/His	
Cytochromes b			
Myoglobin	b	His	Oxygen storage
Haemoglobin	b	His	Oxygen transport
Flavohaemoproteins	b	His	Oxygen sensing
CooA	b	His	CO-sensing transcription activator
FixL	b	His	Oxygen sensor
Cytochromes a			
Cytochrome c oxidase	a	His	Cytochrome c Oxidation

Table 1.1: A table showing some examples of cytochromes, their functions, the type of haem they hold and the axial ligand to the haem iron.

1.3 Cytochromes c:

In all cytochromes c type, the haem is covalently attached to the polypeptide via two, or very rarely one, thioether bonds between vinyl groups of the haem and the reduced cysteine's thiol group in the apocytochrome (Figure 1.2). In the case of two thioether linkages,[34] the presence of the sequence motif CXXCH (where C is cysteine, H is histidine and X can be any amino acid except cysteine) in the apocytochrome provides thiol groups of reduced cysteine residues for the formation of the thioether linkages with the haem's vinyl groups.[35] The imidazole of the histidine often serves as the fifth axial ligand to the haem iron. This motif changes to F/AXXCH in the case of a single thioether linkage.[36,37] The sixth ligand can be methionine (His/Met) or histidine (bis-His). The other four ligands are provided by nitrogen atoms of the tetrapyrrole ring. Some bacterial cytochromes c could exhibit different sequence motifs, such as CXXXCH[38] and CXXXXCH.[39] The axial ligand may vary as well, for instance, the nitrite reductase NrfA contains the unusual CXXCK haem binding site,[40] in which the axial ligand is lysine (K) instead of the usual histidine residue.

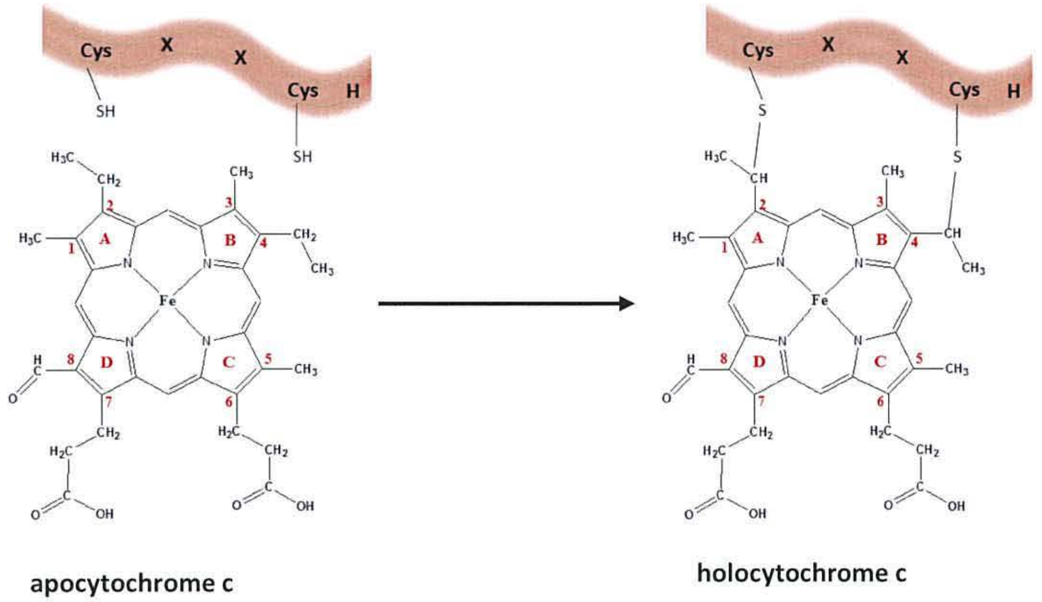


Figure 1.2: The CXXH sequence motif, showing the formation of thioether bonds by a stereospecific addition of reduced cysteine thiol groups to the apocytochrome c to protoporphyrin IX producing holocytochrome c.

1.3.1 Cytochrome c biosynthesis:

All cytochromes c reside in a different location than the site of synthesis of their apoproteins and haem co-factor. The formation of thioether bond(s) between the reduced cysteine thiol groups and the haem's vinyl group(s) in a posttranslational process is the key stage of the formation of holocytochrome c, a process known as cytochrome c maturation.[41] This bond formation does not occur spontaneously under physiological conditions, is tightly coupled to the folding of apcytochrome c and the transport of apocytochrome c and the haem prosthetic group to their site of function. Based on the occurrence of specific haem attachment components, three pathways have been recognised, termed, system I, II, and III, all of which lead to the covalent attachment of haem to the polypeptide.[42] All the systems have in common four general steps within the maturation process, translocation and the correct folding of the apocytochrome c, translocation of haem group, reduction of the cysteine thiol groups, and the formation of thiol linkages.

System I also known as CCM (cytochrome c maturation) is common in α and γ-protobacteria microbial models, where it was initially discovered, land plants, and in some red algae and protozoa mitochondria.[43, 44] System I is best understood in *Escherichia coli (E. coli)*, where it consists of nine membrane and periplasmic assembly proteins (CcmABCDEFGH and DsbD). They all carry out their functions in the periplasmic space.[45] *Alcaligenes xylosoxidans* cytochrome c′ (AXCP) which is the subject protein in this thesis is synthesised in the periplasm by this system, therefore it will be discussed in detail hereafter.

Since the cytochrome c maturation process occurs in the bacterial periplasmic space, while both apocytochrome c and haem substrates are synthesised in cytoplasm, the maturation process requires the following: the synthesis of the apocytochrome c and haem substrates, delivery of the substrates to the periplasmic space and maintenance under a reduced state, and then the haem ligation reaction. These steps are summarised below and represented in (Figure 1.3):

1- Translocation of apocytochrome c: The apocytochrome c is synthesised in bacterial cytoplasm and the N-terminal sequence (leading sequence) directs the protein to its location of function in the periplasm. Once in the periplasm the N-terminal sequence is cleaved by a leader peptidase. The precursor is then translocated into the periplasmic space before haem attachment by the type II secretion pathway. [41]

2- Translocation of haem: The haem is synthesised in bacterial cytoplasm and has to be transported to the periplasmic space where it assembles with the apocytochrome. The haem chaperone CcmE is the key player in Ccm system. Experiments in *E. coli* showed that haem is covalently but transiently bound to CcmE, therefore, haem-bound-CcmE (holo-CcmE) is thought to be a transient intermediate in the haem translocation.[46] The haem attachment to CcmE is dependent on CcmC.[47] CcmA and CcmB belong to the ATP-binding cassette (ABC) transporter, and CcmA has been shown to hydrolyse ATP.[48] Therefore it has been suggested that CcmA and CcmB are required to release haem from the haem chaperone CcmE by coupling the energy obtained from ATP hydrolysis. CcmD is a mediator of complex formation between CcmC and CcmE.[49]

3- Haem ligation: CcmF and CcmH are involved in the transfer of haem from holo-CcmE to apocytochrome c, including the covalent haem attachment step to produce the product holocytochrome.[50] For the bonds to occur between the cysteines of apocytochrome CXXCH motif and the haem, the thiol group of the cysteines have to be reduced.[45] The cysteine thiolates of the apocytochrome at haem binding site form an intramolecular disulfide bond that must be reduced prior to covalent haem attachment.[41] Therefore, thioredoxin-like protein CcmG and thiol-disulfide transporter DsbD are proposed to facilitate the reduction of this intramolecular disulfide bond.[51, 52]

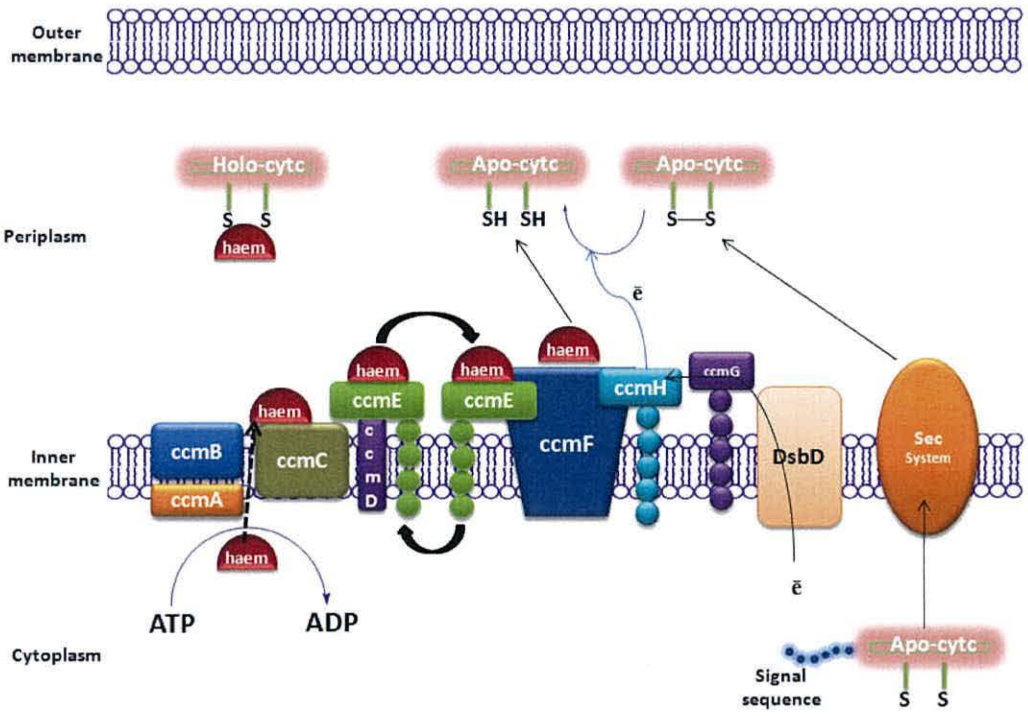


Figure 1.3: Cytochrome c biogenesis system I (the Ccm system) in *E. coli*. The Apo-cytochrome c (apo-cyt c) is synthesised in cytoplasm, the signal sequence directs the apo-cyc c to the periplasm where it will cleaved by the action of leader peptidase enzyme, and then transported to the periplasmic space by the type II secretion pathway (sec system). Haem has to be transported to periplasm. Haem is transported by CcmA and CcmB which are parts of the ABC transporter, and then covalently binds to the CcmE forming holo-CcmE by the aid of CcmC and CcmD. CcmF and CcmH helps the transport of haem from holo-CcmE to the apoproteins. For the formation of thioether bonds between haem and apo-cyt c, thiol groups of the apo-cyt c has to be reduced, CcmG and DsbD are involved in this reduction. The figure adapted from references [26, 52, 53]

The other two systems are less complicated than system I. System II known as the CCS (cytochrome c synthesis), occurs in chloroplasts, some Gram positive[54] and Gram negative bacteria,[55] and cyanobacteria.[26] This system is less well understood than system I at the molecular level. It is very likely to consist of four proteins[56] (Figure 1.4). CcsX also called ResA is a thioredoxin-like protein; it is similar in structure to CcmG[57] and CcdA, similar in function to DsbD; together reducing a disulfide bond in the apocytochrome CXXCH motif. The haem attachments and possibly haem delivery function is catalysed by CcsA and CcsB.[53]

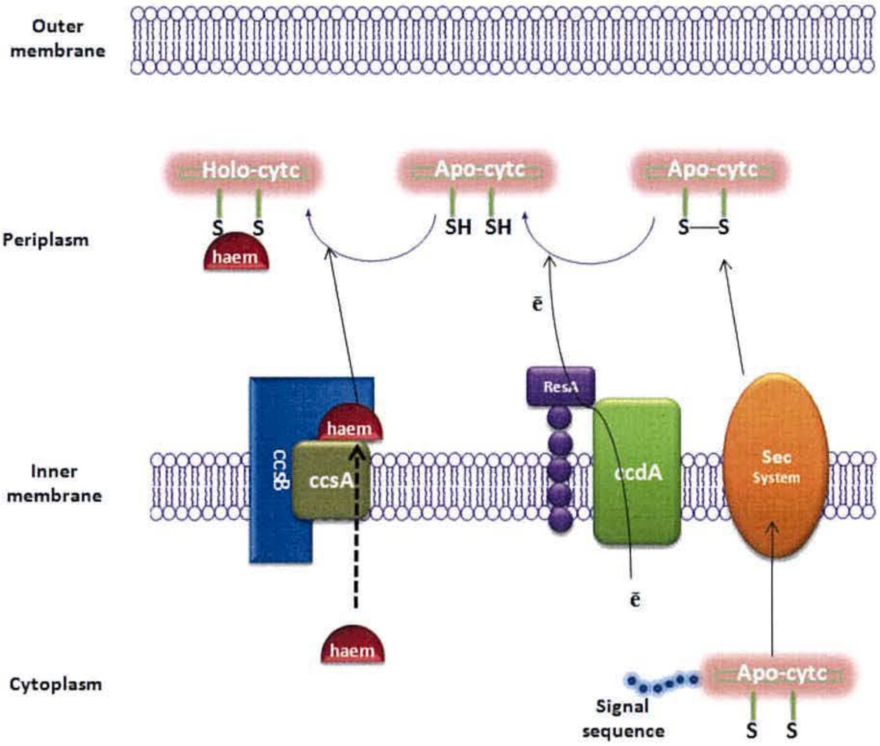


Figure 1.4: Cytochrome c biogenesis II in Gram-negative and Gram positive bacteria, chloroplasts, and cyanobacteria.

System III, also referred to as CCHL (cytochrome c haem lyase), is the simplest known maturation system. It is restricted to the mitochondria of fungi, animals, apicomplexan parasites, diatoms and green algae,[58] where CCHL (also known as holocytochrome synthases) are the central components of biogenesis. Cytochrome c_1 haem lyase[59] (CC_1HL) and CCHL[60] of yeast were found to be involved in the maturation of cytochrome c_1 and cytochrome c respectively.

1.3.2 Classification:

On the basis of sequence similarities, c-type cytochromes have been grouped into four classes:[61]

Class I includes the monohaem mitochondrial and bacterial cytochromes whose haem is attached close to the N-terminus. The haem Fe is hexacoordinate with histidine in the fifth coordinate position and methionine, which is about 40 residues further on toward the C-terminus, in its sixth coordinate position. The total number of amino acids ranges between 80 and 120.

Class II includes the high-spin (also known as class IIa) and low spin (know as class IIb) cytochromes c′. They are monohaemoproteins with the haem covalently bound to the C-terminus. The haem is pentacoordinate with histidine in the iron's fifth coordinate site while the sixth coordinate position is vacant and may be available for binding small exogenous molecules such as CO, NO, or CN^-. The protein usually exists as a dimer with about 110-130 amino acids for each monomer. Cytochrome c′ from *Alcaligenes xylosoxidans* (AXCP) which is the subject of this thesis is regarded as class II c-type cytochrome.

Class III comprises the multihaem cytochromes with bis-histidine axial ligation. They are characterised by the presence of variable number of helices, little β sheet content and only 30-40 amino acid residues per haem group in their structures.[62]

1.4 Cytochrome c′:

Cytochromes c′ are small c-type cytochromes found in photosynthetic,[63] denitrifying,[64] nitrogen-fixing,[65] and sulphur-oxidising bacteria.[66] They are a distinct family of class II cytochromes, although several species undergo an unusual ligand-controlled dimer dissociation.[67] They are usually isolated as soluble homodimers composed of two identical subunits of approximately 130 residues. Each monomer contains a c-type haem within a four α-helix bundle.[68] The first cycP was isolated from a photosynthetic bacteria, *Rhodospirillum rubrum* in 1954 and was named pseudohaemoglobin.[69] The covalent attachment mode of the c-type haem serves to classify these molecules as c-type cytochromes, and the 'prime' superscript indicates that the haem-iron atom is high-spin and also discriminates these proteins from the structurally dissimilar low-spin mitochondrial proteins. The haem iron in cycP is pentacoordinate (5c) with only a single histidine axial ligand (proximal pocket), while the sixth iron coordination is vacant (distal pocket) and surrounded by bulky hydrophobic residues which form a so called hydrophobic cage with only hindered access of ligands to the haem iron's sixth coordinate.[70] This vacant distal pocket directs the protein to play a role in ligand binding rather than the electron transport role of the hexacoordinate (6c) cytochromes c.[71] All cycP proteins characterised to date form complexes with CO and NO but do not generate stable O_2 adducts[71] and they show much lower affinity for exogenous ligands when compared to globins[67] probably due to the steric constraints around the haem distal pocket.

The vacant distal pocket of different cytochromes c′ is relatively buried and packed with hydrophobic residues such as leucine, methionine or phenylalanine, while the proximal pocket is close to the protein's surface with the histidine ligand exposed to solvent. Based on the nature of the distal hydrophobic pocket residues, cytochromes c′ have been divided into two types:[72] Type 1 consists of aromatic residues such as Phe and Tyr above the distal haem coordination site, combined with a solvent exposed channel to the distal pocket such as those from *Chromatium vinosum* and *Rhodobacter capsulatus*. In contrast, Type 2 cytochromes c′ possess aliphatic residues such as Leu or Met above the distal haem coordination site with no direct access to the distal pocket. The solvent channel in this type is blocked by other residues making the distal haem coordinate inaccessible; AXCP is an example of this type (Figure 1.5).

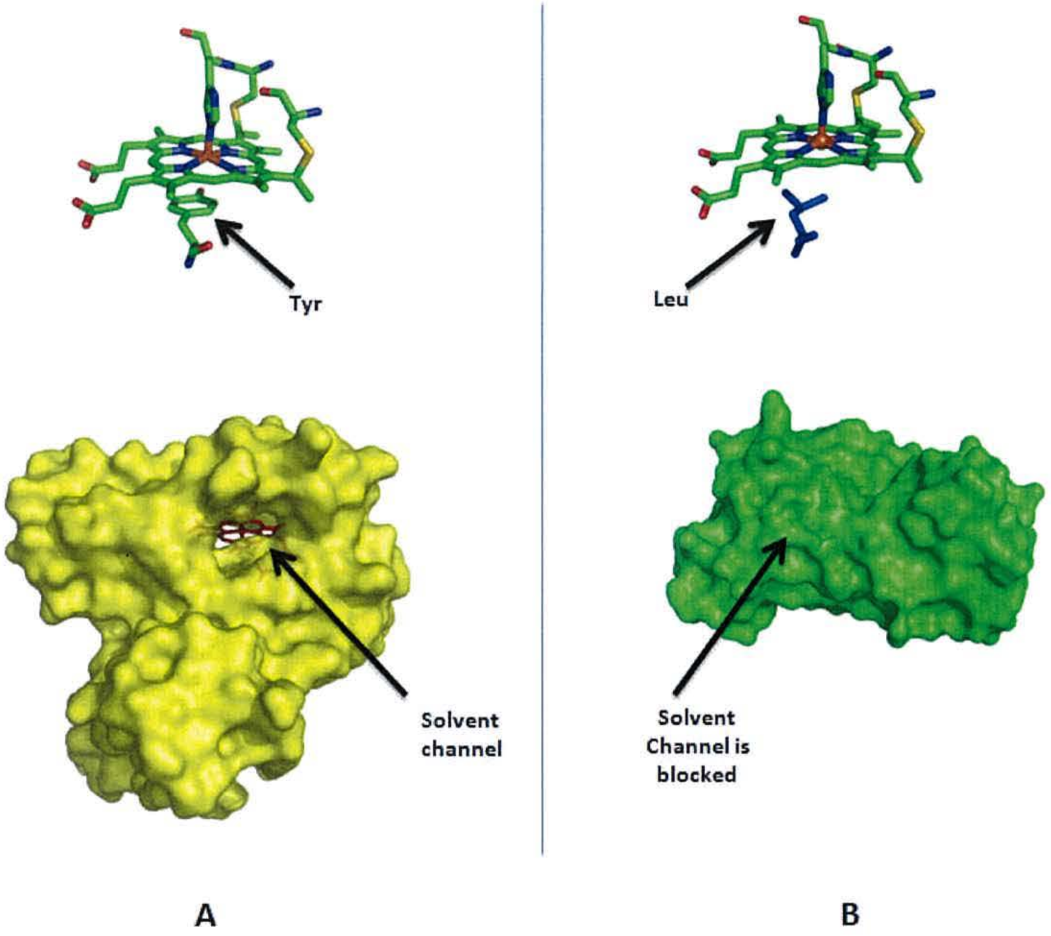


Figure 1.5: (A) Type I cytochrome c′ from *C. vinosum* (PDB 1BBH) showing the aromatic residue in the distal pocket of the haem iron (above) and the channel by which the distal pocket is exposed to the solvent (below). (B) is showing type II cytochrome c′ from *A. xylosoxidans* (PDB 1E83), the aliphatic residue Leu is occupying the distal haem pocket, the solvent channel is blocked in this type. Figures have been created with PyMOL.[73]

Although cytochromes c′ have been isolated for more than half a century[69] and have been extensively characterised, their physiological function is still somewhat unclear[71] since even in closely related organisms the level of expression in response to oxygen/nitrate level is different. Recently, evidence for possible roles for cytochrome c′ in protecting bacteria from nitrosative stress[74] and in NO-shuttling has emerged.[75]

1.5 Crystal Structures:

To date, the crystal structures from the diverse bacterial species that have been solved and deposited in the RSCB Protein Data Bank[76] are summarised in (Table 1.2).

Bacterial Source	PDB code	Resolution (Å)
Alcaligenes xylosoxidans	1CGO (oxidised)	1.80
Alcaligenes denitrificans	1CGN	2.15
Alcaligenes xylosoxidans	1E83 (oxidised form)	2.05
Alcaligenes xylosoxidans	1E84 (reduced form)	1.90
Alcaligenes xylosoxidans	1E85 (NO-bound)	1.35
Alcaligenes xylosoxidans	1E86 (CO bound)	1.95
Rubrivivax gelatinosus	1JAF	2.50
Rhodobacter capsulatus	1CPQ	1.72
Rhodobacter sphaeroides	1GQA	1.80
Phodopseudomonas palustris	1MQV	1.781
Phaeospirillum molischianum	2CCY	1.67
Allochromatium vinosum	1BBH	1.80

Table 1.2: The Protein Data Bank (PDB) deposited crystal structures of cytochrome c′ from different bacterial species with their resolutions and codes

The unusual feature of cytochromes c′ is the amino acid sequence identity among cytochromes c′ is around 20 % in most cases. But the x-ray crystal structures (Figure 1.6) showed that their overall folding pattern (classic four-helix bundle), haem attachment to a CXXCH motif near the carboxyl terminus [77] and the haem environment are structurally homologous.[78]

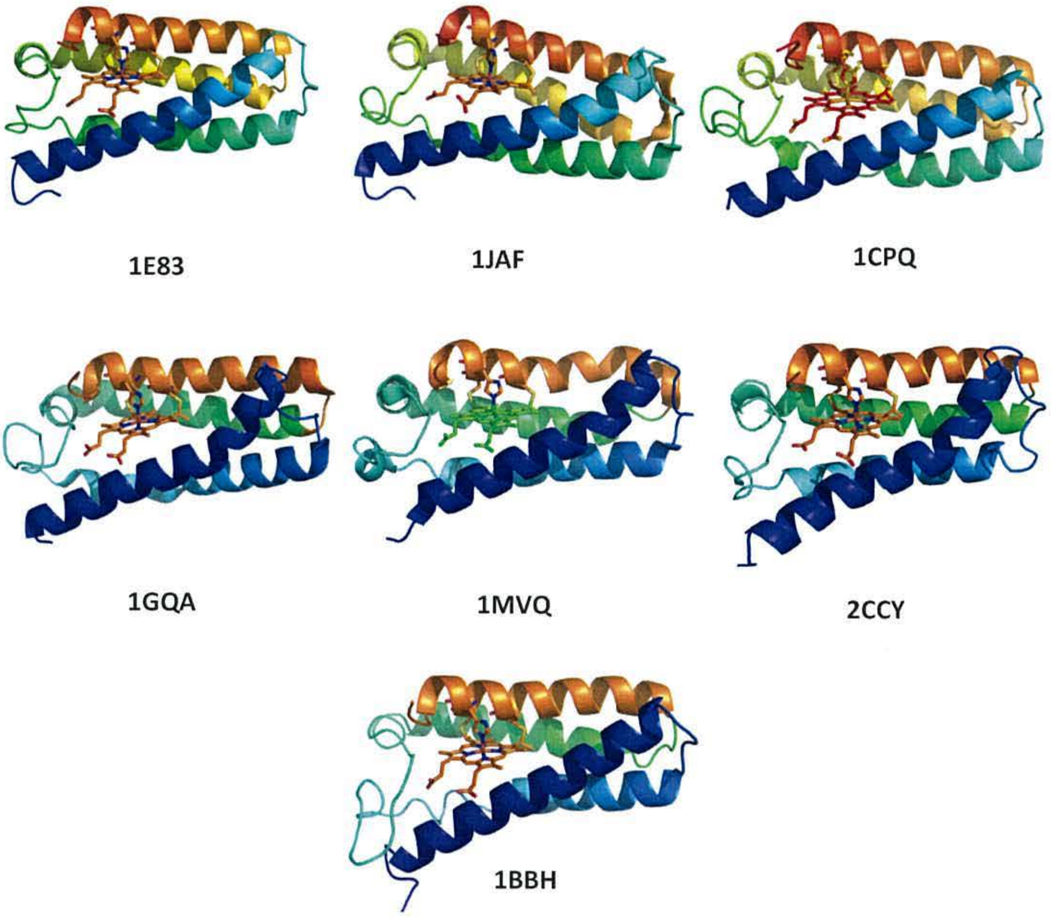


Figure 1.6: The crystal structures of cytochromes c′ from different bacterial species. In spite of their low amino acid sequence identity, they exhibit the same classic four-helix bundle, and the haem attachment to CXXCH motif to the C-terminus site. Legends are PDB codes, for relevant bacterial species refer to table 1.2.

Since, the study of AXCP is the objective of this work; it will be described in more detail.

1.5.1 Alcaligenes xylosoxidans (AX):

First described in 1971,[79] formerly known as *Pseudomonas denitrificans*, but nowadays reclassified as *Alcaligenes xylosoxidans* or *Alcaligenes sp.* NCIMB 11015.

It is an aerobic, motile, Gram-negative bacillus which has been isolated from the normal flora of the ear and gastrointestinal tract in humans.[80] Infections caused by *AX* include bacteremia, meningitis, biliary tract infection, urinary tract infection, pneumonia, and osteomyelitis.[80]

1.5.2 Overall AXCP crystal structure description:

The first ferric AXCP crystal structure was solved in 1996 at 1.8 Å resolution[81] (PDB code 1CGO), while the ferrous crystal structure was not solved until 2000 (PDB code 1E84) when the NO- and CO-bound crystal structures were solved as well.[82] (PDB codes 1E85 and 1E86 respectively).

AXCP is a dimeric protein and is made up of two identical polypeptide chains of 14 kDa. The subunits are arranged in head-to-tail fashion. The major parts of the dimer interface lie between residues Tyr11, His22, Leu18, and Ala15 from helix A of one subunit to His22, Tyr11, Ala15, and Leu18 of the other subunit. Each subunit consists of a left-twisted anti-parallel four-α-helix bundle, labeled A-D (Figure 1.7). There are a few cross-linking hydrogen bonds between helices. Helices are connected by two short loops (A-B and C-D), and one longer one (loop B-C). Loop (A-B) consist of only five residues (33-37), and the (C-D) loop can hardly described as a loop because residue number 100 is the last residue of helix C and residue 101 is the first of helix D. The one longer loop (B-C) comprises 21 residues (55-75) and it helps to provide some of the packing around the haem. Since the A-B and C-D loops are shorter than B-C loop, the helix axes spatially diverge from the double-connected end towards the single-connected end, where the haem binding pocket is formed. Each subunit incorporates a protohaem IX molecule as a prosthetic group, which is covalently bound to helix D in CXXCH motif via thioether bonds between the two cysteine residues and the haem vinyl groups close to the C-terminus [81] and is partially exposed to solvent.[82] The iron haem is five coordinate (5c) with His120 as the fifth axial ligand in the proximal pocket, while the sixth coordination site is vacant, and surrounded by bulky hydrophobic residues.

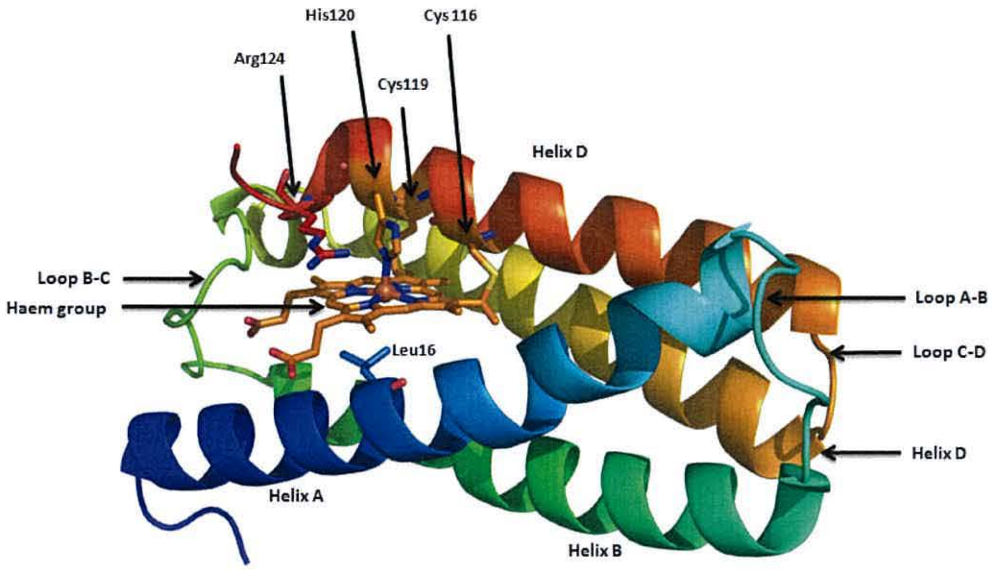


Figure 1.7: The overall structure of the AXCP monomer (PDB code 1CGO). Loops and helices are shown. Two cysteine residues, Cys116 and Cys119, and the histidine residue His120 are shown which illustrate the CXXCH motif. Leu16 is shown, located in the haem's distal pocket.

Both distal and proximal haem sites of AXCP have very different structural features. The distal haem pocket is deeply buried within the protein with no direct access to solvent. By contrast, the proximal haem face is much closer to the protein surface, with the proximal His120 ligand being solvent-exposed.

1.5.3 Oxidised and reduced AXCP structure:

The least squares superposition of the oxidised (PDB code 1E83) and reduced (PDB code 1E84)[82] structures at 2.05 and 1.9 Å resolution respectively, revealed very little difference in the overall structure of both structure with an (r.m.s.) deviation of 0.12 Å overall main chain atoms. The major observable change is in the position of the side chain of Arg124 (Figure 1.8). In the oxidised structure, the guanidinium group of Arg124 is parallel to the **A** pyrrole ring of the haem and perpendicular to the imidazole ring of His120. In contrast, in the reduced structure this group is parallel to the imidazole ring of His120 and perpendicular to the haem plane. It is appears that the side chain of this Arg124 senses the oxidation state of the iron haem, but the mechanism is still obscure. However, it has been proposed that the positive charge of the Arg124 overlaps with the negative electron density of the haem π system.[82]

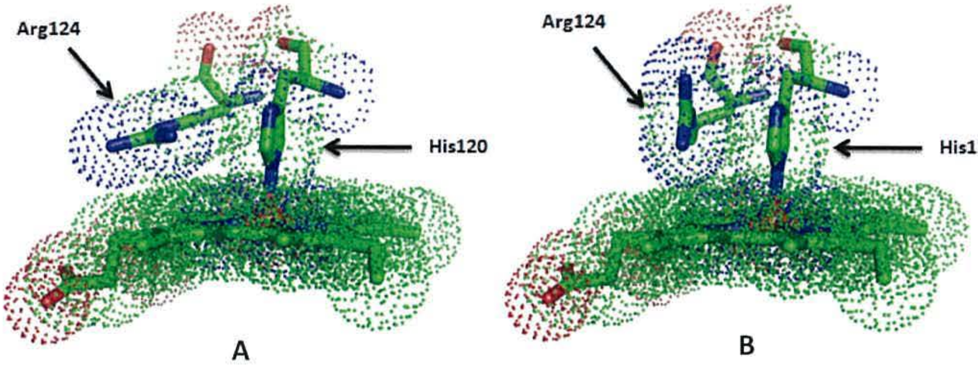


Figure 1.8: (A) Oxidised (PDB code 1E83) and (B) reduced (PDB code 1E84) structures of AXCP showing the guanidinium group of Arg124. The guanidinium group of this residue is parallel and perpendicular to the haem plane in the case of oxidised and reduced forms respectively

1.6 Spectroscopic properties of *AXCP*:

When compared to all other haemoproteins, cytochromes c′ have novel spectroscopic and magnetic properties[83] making AXCP of particular interest to be further characterised.

The iron in haemoproteins exist in two oxidation states, either ferric Fe(III) or ferrous Fe(II). The electronic structure of Fe(III) is $4s^2$ $3d^5$. In hexa-coordinate octahedral systems, iron is coordinated to six ligands. The crystal field of the six ligands splits the degeneracy of the five *3d* orbitals on the metal ion. Two of the *3d* orbitals ($3d_{x^2-y^2}$ and $3d_{z^2}$ – also the called e_g orbital set) point directly toward the ligands as shown in (Figure 1.9), while the other three orbitals ($3d_{xy}$, $3d_{xz}$, and $3d_{yz}$ - also known as t_{2g}) lie between ligands. Upon the ligand binding, the energy Fe(III) metal ion's e_g orbitals increases much more than the energy of t_{2g} orbitals.

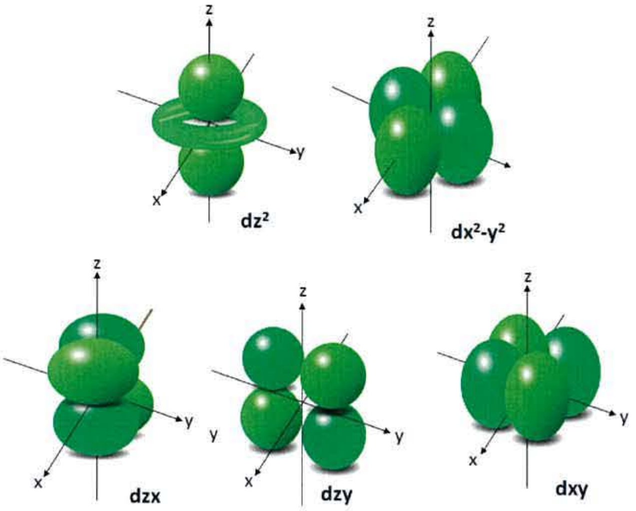


Figure 1.9: The spatial arrangement of *3d*-orbitals. Axis x and y lie in the haem plane, while z axis is equivalent to the axial ligand. Orbitals $d_x{}^2{-}_y{}^2$ and dxy are in haem plane whereas the other three orbitals lie above and below the plane.

1.6.1 Electronic spin states of AXCP:

In porphyrin systems, since the lobes of the d_{xy} and $d_x{}^2{-}_y{}^2$ orbitals lie in the porphyrin plane, and the d_{xz}, d_{yz}, and $d_z{}^2$ orbitals are above and below the plane, the spin-state is sensitive to the ligand field strength of the iron axial ligand. Accordingly, the electronic ground state of ferric haemoproteins could either be low-spin (S=1/2), mid-spin (S=3/2), high-spin (S=5/2), or quantum mechanically admixed high- and mid-spin (S=3/2, S=5/2) state. These are schematically shown in (Figure 1.10).

The low electronic spin state is commonly found in six-coordinate haem compounds where two axial ligands complete the coordination environment of the haem ferric ion. They are characterised by an energy separation between $d_{(xy,xz)}$ and $d_z{}^2$ orbitals (Δ_1) larger than the spin pairing energy (P) (Figure 1.10). A large number of six-coordinate, low-spin ferric haem complexes have been synthesized, and many six-coordinate, low-spin ferric haemoproteins have been isolated, whereas, no low-spin five-coordinate ferric haem protein has been characterised.[83]

A mid-spin state occurs when Δ_1 is smaller than P while the energy gap between $d_z{}^2$ and $d_x{}^2{-}_y{}^2$ (Δ_2) is larger than P (Figure 1.11). While several cases of mid-spin model complexes are known, no example of pure, or nearly pure, haemoproteins is currently available.[83]

High-spin states arise when both Δ_1 and the energy separation between $d_{x^2-y^2}$ and d_{z^2} (Δ_2) are smaller than P (Figure 1.11). This spin state has been found in a large number of five-coordinate complexes having a moderately strong anionic ligand. A small number of six-coordinate complexes having two neutral, weakly basic axial ligands are known. Many five-coordinate, high-spin ferric haemoproteins are also known.[83]

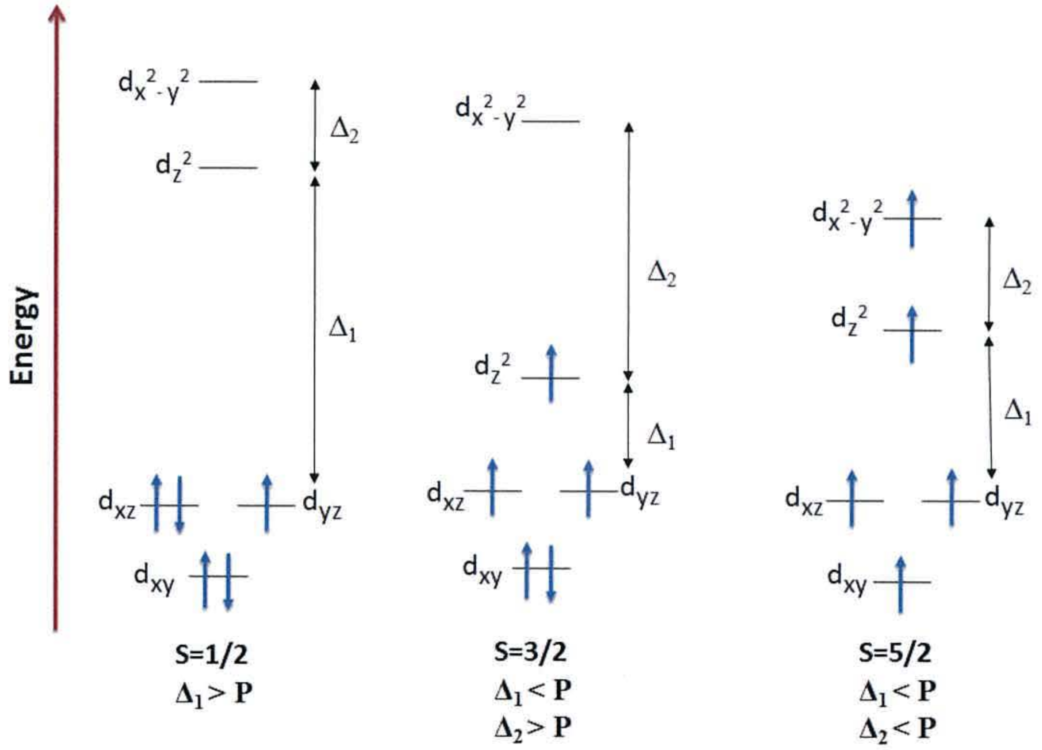


Figure 1.10: Schematic distribution of *d*-orbitals giving rise to different electronic spin ground states in five coordinate Fe(III) haemoproteins. The caption is taken from reference 84 and the figure is re-drawn from the same reference. P is spin pairing energy, Δ_1 refers to the energy separation between d_{xy},d_{xz} and d_z^2, and Δ_2 is the energy gap between the d_z^2- $_y^2$ and d_z^2.[84]

The spin state of the ferric cytochromes c′ has been the subject of controversy. Initially, magnetic susceptibility measurements suggested a thermal equilibrium between the low-spin (S=1/2) and the high-spin (S=5/2) states.[85] Afterwards, near infrared, near infrared magnetic circular dichroism (MCD), Mössbauer, and NMR data were interpreted as a significantly high-spin (S=5/2) state at neutral pH.[86,87,88] On the other hand, based on the anomalous EPR spectra, Maltempo[89] explained the unusual magnetic properties of ferric cytochromes c′ as arising from a quantum mechanical admixture of a high-spin (S=5/2) and a mid-spin (S=3/2) states.

The quantum mechanically admixed med-spin (S=3/2) and high-spin (S=5/2) state, which is very unusual in biological systems,[90] has been found to be common in cytochrome c′ from various bacterial sources at neutral pH.[91, 92] This mechanical admixed spin-state arises from the contribution from two or more different spin states. When the energy separation of mid- and high-spin states approaches the magnitude of spin pairing energy, the electronic ground state is best described as a quantum mechanical admixture of two spin states.[83] Unlike thermal admixtures of spin states in which protein molecules can be monitored in two magnetically distinguishable pure spin states, mechanically admixed spin state protein would correspond to a single magnetic species with magnetic properties distinct from high-mid spin states.[89] The presence of this spin-state has revealed the weak axial ligand and saddle-shaped haem. Scheidt and Reed[93] proposed that any effect which weakens the axial ligand field of a high-spin ferric haemoprotein may bring the spin state into the admixed-spin state. Cheng *et al*[94] attributed the existence of quantum mechanically admixed spin states to the saddle-shaped haem deformation. But both suggestions seem not to be sufficient to cause the QS state.[90]

The contribution of the mid-spin state to the QS vary from 10% to 60% depending on the bacterial source from which the cytochrome c′ is derived and also depends on the pH of the solution, decreasing as the pH increases. Cytochrome c′ isolated from *Rhodobacter capsulatus* and *Chromatium venosum* exhibits a large mid-spin contribution to an admixed ground state ranging from 40 to 57%, while cytochromes c′ from *Alcaligenes xylosoxidans, Rhodospirillum rubrum,* and *Rhodospirillum molischiatum* appears to have a smaller mid-spin contribution of around (10-15%).[83,]**Error! Bookmark not defined.**

The mid-spin contribution to the QS decreases as the pH of the solution increases, resulting in high-spin ferric ion at pH 11.0. Therefore the haem centre exhibits a pH-dependent spin state equilibrium between high-spin and QS-state. Deprotonation of the axial His ligand at moderate alkaline pH values (7-9) may cause the formation of a strong anionic ligand, and the consequent Fe(III) displacement from the haem plane which cause a destabilization of $d_z{}^2$ and a stabilization of $d_x{}^2{}_y{}^2$ which in turn causes a transition from mid-spin (S=3/2) to quantum admixed-spin state, and then to high-spin state (S=5/2). It has also been suggested that the negative charge on the proximal ligand in cytochrome c′ could be stabilised ionically by a nearby positively charged residue, as with AXCP, where His120 is closely associated with Arg124.

1.7 Ligand binding and discrimination:

Sensing and discriminating between diatomic molecules such as nitric oxide (NO), carbon monoxide (CO), and oxygen is a key biochemical feature present in all life forms, and this binding is predominantly mediated by haem-based sensors.[95] The common feature shared by all haem-based sensors is the need not only to recognise its target ligand, but also to discriminate against other ligands of the same size and shape.

In animals discrimination between diatomic molecules is essential for a wide range of physiological processes such as vascular homeostasis,[96,97]platelet aggregation, neuronal signaling,[98] host defense,[99] and stress responses.[100] While in Bacteria, this gas sensing is crucial for aerotaxis, combating the host immune response,[101] using CO as the energy source,[102] protecting bacteria from nitrosative[74] stress and in NO-shuttling[75] and inducing certain genes under unaerobic conditions.[103]

Diatomic molecules interact with the haem iron forming either five co-ordinate (5c-) or six co-ordinate (6c-) adducts. Interaction of these adducts with the surrounding protein residues are critical determinants of ligand discrimination.[104] For instance in other types of haemoproteins, mutagenesis studies have shown that the H-bonding between the polar Fe-O bond and distal residue His64 of the haem in myoglobin (the oxygen storage protein in muscle tissue) is a key factor promoting binding of O_2 over NO and CO.[105] Some members of H-NOX (Haem-Nitric oxide and/or oxygen-binding domain) family[106] which include bacterial proteins from aerobic and anaerobic organisms form a tight complex with O_2 while others, like soluble guanylate cyclase (sGC) do not bind O_2 but do form a 5c Fe-NO adduct by selectively excluding O_2. sGC catalyses the cyclisation of a guanosine 5'-triphosphate (GTP) to guanosine 3',5'-cyclic monophosphate (cGMP). The presence of a hydrogen bonding network in the distal pocket of O_2-binding H-NOX proteins provided by distal pocket residue tyrosine or other residue capable of forming a strong hydrogen bond to bound O_2 is critical, while the lack of such a distal pocket hydrogen bond in sGC eliminates O_2 as a ligand, making the sGC H-NOX domain selective for NO instead.[107] The origin of this selectivity has been shown to be kinetic; a distal pocket tyrosine in O_2 binding H-NOX proteins markedly slows down its dissociation rate, whereas the dissociation rate is very fast in the case of sGC due to absence of tyrosine or an alternative H-bond donor in the distal pocket.[107]

There is a strong similarity in the distinctive reactivity and spectroscopic properties of the haem centre in AXCP and sGC.[108] Ferrous AXCP has been reported to bind NO and CO but not O_2, just like sGC.[109] NO forms five-coordinate (5c) ferrous-NO complex with sGC[110] and AXCP.[64] In addition both proteins exhibit a much reduced affinity to cyanide, and in the ferric state both AXCP and sGC show Soret maxima in 390-400 nm region.

The crystallographic characterisation of AXCP showed the exciting discovery that exogenous ligands can bind to the iron from both sides of the haem face. While CO binds to iron at the distal face of haem, producing a six-coordinate carbonyl (6c-CO) complex, NO exhibited a novel and unexpected proximal five-coordinate nitrosyl (5c-NO) complex with iron by displacement of the proximal His120 ligand.[82]

1.7.1 NO bound structure:

NO is a free radical, small, short lived, and highly reactive gaseous molecule, that readily and quickly diffuses through the plasma membrane and reacts with haem iron. However, spectroscopic studies have shown that, depending on its concentration, NO can form both 6c- and 5c- complexes with AXCP haem iron.[111]

At physiological pH values the His-Fe bond of the reduced AXCP is cleaved on NO binding to form a novel five-coordinate haem-nitrosyl (5c-NO) adduct.[112] This finding has been confirmed by the x-ray crystal structure of reduced AXCP NO-bound adduct at 1.35 Å resolution (PDB code 1E85).[82] The crystal structure of this complex showed the unexpected presence of NO at the proximal haem pocket in place of the His120 residue. The mechanism by which the 5c-NO complex is formed will be discussed in detail in section 1.7.2.

Least square superposition of oxidised and NO-bound AXCP has shown a small change in the overall structure with an (r.m.s) deviation of 0.14 Å. The main change is the displacement of His120 to one side by 111° around C_α-C_β bond. In its new position, His120 is stabilised by hydrogen bonding to a water molecule (Figure 1.11). The NO molecule is bound to iron at proximal side and has been modeled in two configurations. The first molecule forms hydrogen bonding with NH1 of Arg124, which in turn changes its configuration back to the conformation similar to that observed for the oxidised structure. No significant changes occur to distal pocket residues.

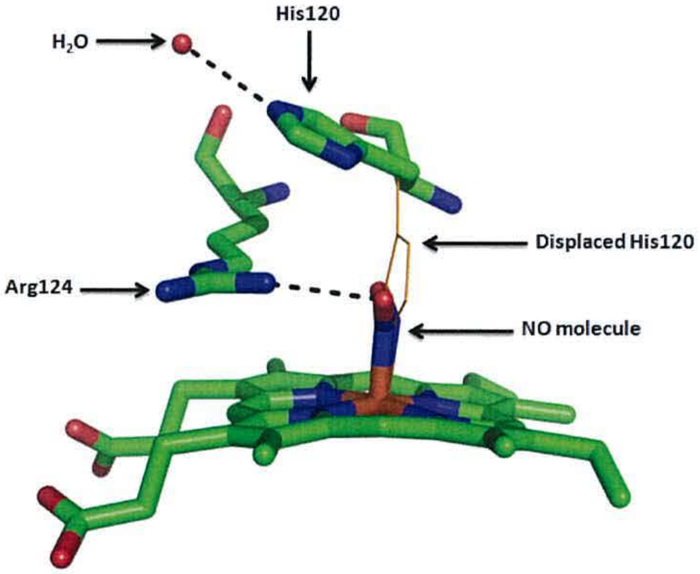


Figure 1.11: Five-coordinate-NO adduct of AXCP (PDB code 1E85). The NO molecule has displaced His120 (coloured in red, PDB code 1E83) to form an NO-adduct at the proximal haem site. Two configurations of NO molecule has been modeled, one of them is hydrogen bonded to guanidinium group of Arg124. His120 has been stabilised by the hydrogen bonding to a water molecule. Figure created using PyMOL.

1.7.2 Mechanism of 5c-NO adduct formation:

Based on x-ray crystallographic analysis and spectroscopic results of AXCP, it was proposed[113] that the formation of the 5c-NO complex occurs via a distal 6c-NO intermediate in two NO-dependent reactions as follows: the first NO molecule binds to the sterically crowded distal face of the ferrous haem to give the 6c-NO complex, and the repulsive trans-effect of NO[114] weakens the proximal Fe-His bond. In this reaction the rate determining step which, in principle, involves Fe-NO bond formation, together with the high- to low-spin transition of the haem iron and its motion from out-of-plane to in-plane. A second NO molecule displaces the proximal His and binds to the proximal face of the haem forming a transient 6c-dinitrosyl complex. Release of the His ligand prior to attack of the second NO molecule has recently been proposed.[71] The second NO molecule repels the distally bound one to leave the iron's sixth coordinate position, producing a 5c-NO complex. It has been shown that the distal side of the haem controls the initial NO binding, while the proximal side controls the release of NO, with the ability of trapping and gating NO, with virtually unidirectional release of NO.[115]

In order to understand how the protein structure of AXCP controls NO dynamics, Resonance Raman and time resolved spectroscopy at the haem level together with molecular computational dynamics study at the entire protein structure level have recently been performed.[115,116] The study revealed that NO dissociation from the 5c-NO adduct yields the 4c-heam adduct. The side chain of displaced His120 and Arg124 in the proximal haem pocket confines NO close to the iron so that an ultrafast (7 picoseconds) and complete (~99%) geminate rebinding occurs. Whereas the proximal His120 does not rebind to the haem iron because His120 does not bind on the timescale of NO geminate rebinding. His120 rebinding to haem calculated to be (~100 ps).

On the bases of recent picoseconds kinetic data and on the previous kinetics and X-ray studies, the following mechanism of 5c-NO complex formation and dissociation has been proposed (Figure 1.12):[116]

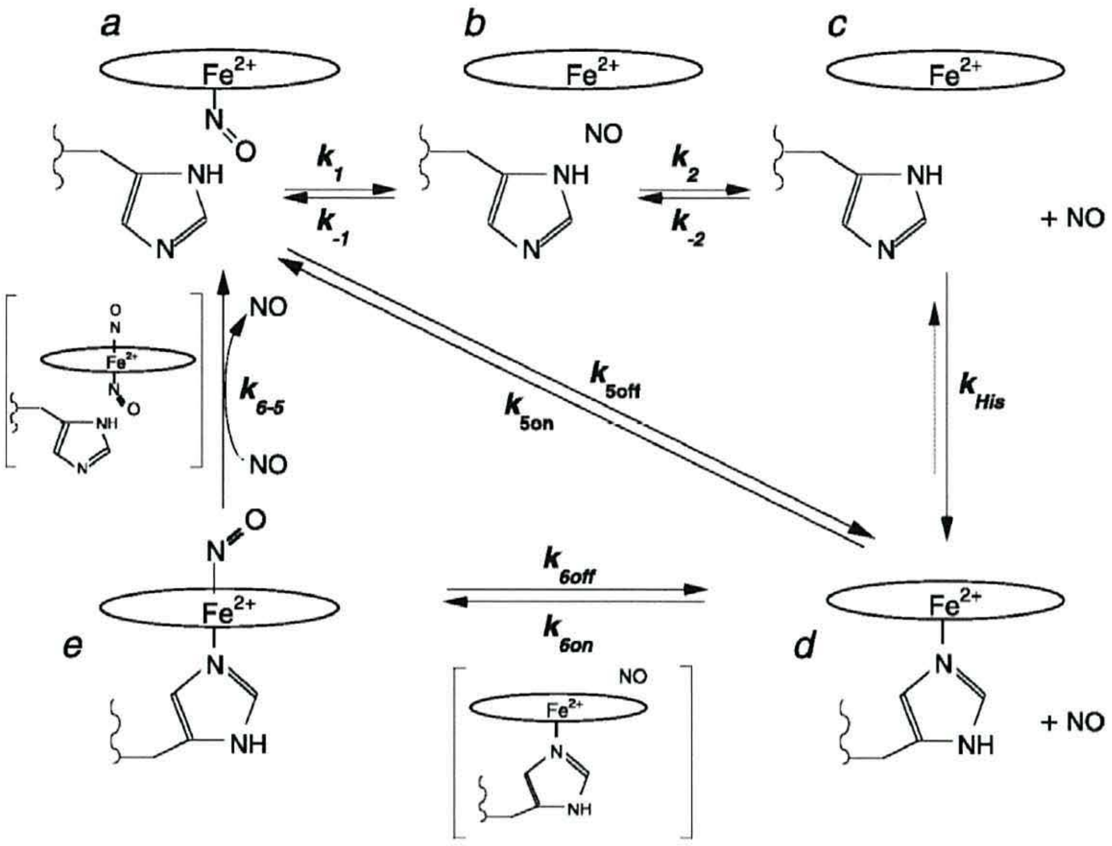


Figure 1.12: The proposed mechanism of 5c-NO adduct formation based on the recent picoseconds kinetic data and on the previous kinetics and X-ray studies. cited from Kruglik 2007.[116]

Photodissociation of the 5c-NO adduct (achieved with an excitation pulse at 564 nm whose duration was ~40 fs with a repetition rate of 30 Hz) releases the NO molecule and produces a 4c-haem adduct (k_1). Geminate rebinding of ~99 % of the released NO molecule rebinds to the 4c-haem adduct on picoseconds (~7 ps) time scale (k_{-1}). The remaining (~1%) population of NO which escape from the haem pocket cannot rebind directly to the proximal side of the haem but releases to the solution (k_2). The presence of a high energy barrier for NO access to the haem proximal pocket makes NO rebinding to the haem unfavorable (k_{-2}). The kinetic data and structural dynamics calculation showed that His120 rebinding to 4c-haem (K_{His}) cannot occur when NO is still located within the proximal haem pocket, but may occur after the release of NO to the solvent in ~100 ps, which is far from the time scale of NO geminate binding (~7 ps). Due to the reactivity of the 4c-haem, the back process (short arrow) has a low possibility to occur. On the basis of the previous kinetics study (which has been discussed earlier in detail), NO binds to the haem's distal face of the 5c-His, which represents the resting state of AXCP, to form a transient 6c-NO-His adduct (6_{on}); this binding weakens the Fe-His bond, and the attack of the second NO molecule displaces His120 to form a transient 6c-dinitrosyle adduct ($k_{6\text{-}5}$). The binding of the second NO causes the first NO molecule to leave the distal pocket of the haem, thereby forming the 5c-NO adduct.

1.7.3 CO bound structure:

Both spectroscopic and crystallographic studies showed that CO binds to the crowded distal face of the haem in AXCP. The crystal structure of AXCP CO-bound complex has been solved (PDB code 1E86).[82] The structure showed that CO binds to the distal face of the reduced haem, producing a six-coordinate CO-bound complex with an almost linear configuration with an Fe-C-O angle of 167°. Upon CO binding, Leu16, which occupies the distal haem face, is displaced to one side by a 134° rotation around C_α-C_β bond. This rotation displaces the pyrrole ring A of the haem group by 0.86 Å towards the proximal pocket, which in turn flattens the haem group. The Arg124 configuration remains perpendicular to the haem plane, which is similar to that found in the reduced structure.

Least square superposition of the CO bound structure (PDB 1E86) to the reduced structure (PDB 1E84) revealed an r.m.s deviation of 0.6 Å, indicating a significant change in the overall AXCP structure upon the CO binding.

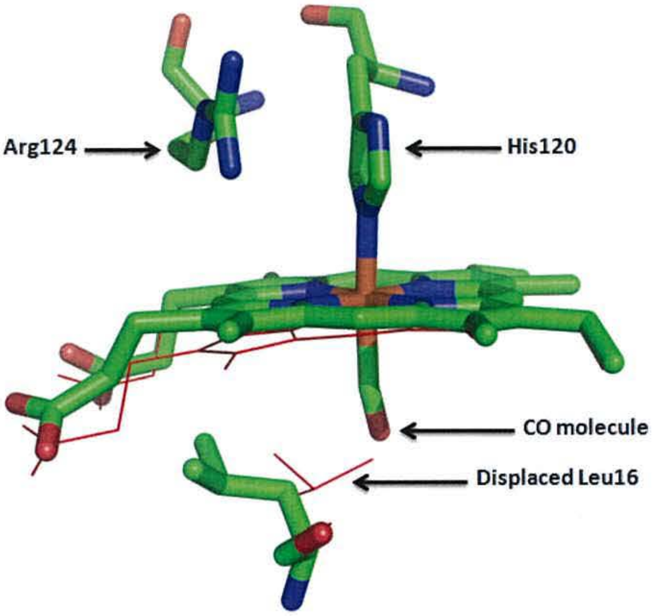


Figure 1.13: Six-coordinate-CO adduct of AXCP (PDB code 1E86). CO group binds to sixth coordinate of the iron haem in a linear manner. This attachment has rotated Leu16 to a side and flattened the haem group while Arg124 has conserved its configuration compared to the reduced structure (coloured in red PDB code 1E84). Figure created using PyMOL.

1.8 Electronic absorption spectroscopy of *AXCP*:

The porphyrin macrocycle is an extensively conjugated system that gives rise to intense electronic absorption bands. This property makes it possible that haemoproteins are easily identified and that a great deal of information may be extracted using a relatively simple technique such as UV-Vis spectroscopy, which has been a powerful analytical tool in haem protein research.

Ferric and ferrous AXCP exhibits four distinct absorption bands; these are, an intense band called Soret band, a weaker band named α/β band and very weak band known as charge transfer band (CT); finally, an additional useful band appears at 280 nm due to aromatic residues. The ratio of absorbance at 280 nm to the absorbance at the Soret maximum can be used as a purity index of the protein. The electron transitions responsible for the optical absorbance of ferricytochromes are illustrated in Figure (1.14) and discussed below:

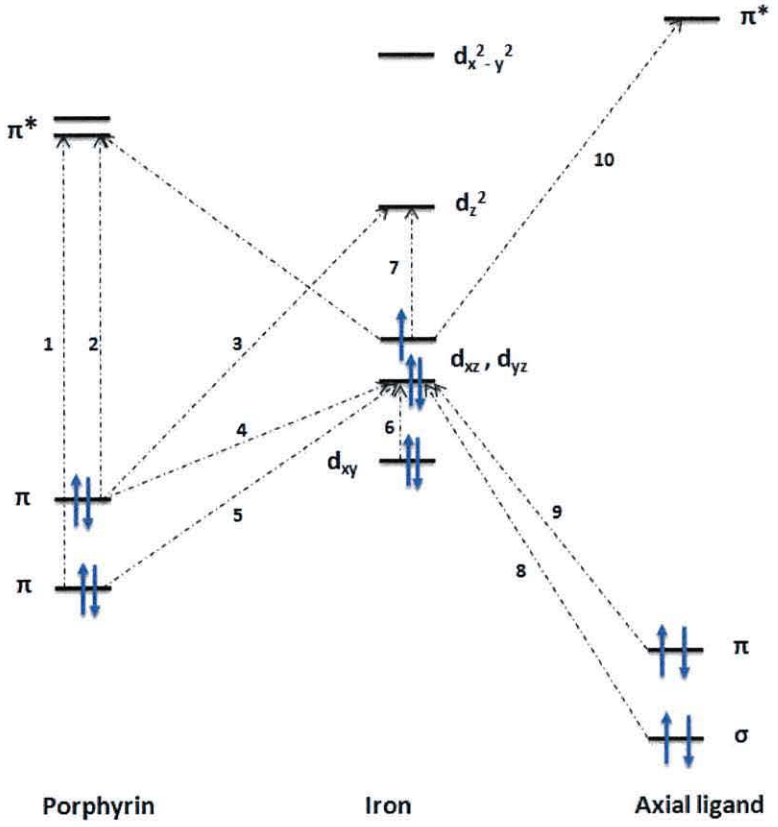


Figure 1.14: A schematic diagram of the electron transitions responsible for the electronic absorption spectra in ferricytochromes. Numbers indicate the orbital transitions. Adapted from Moor and Pettigrew (1990).[117]

The intense Soret band arises from (π→π*) transition of the porphyrin ring (transitions 1 and 2). This transition is also accompanied by a weaker one, the α/β band. The intensity of both bands is correlated. The bands are highly dependent on the iron's oxidation state and the coordination of the haem group at the peptide. The charge-transfer band at ~640 nm arises from transitions 3, 4, and 5. This transition is only observed in the high-spin ferric and 5c- cytochromes indicating that it is sensitive to the spin state[118] and also to the coordination number.[83] It is believed to be due to *d-d* transition which comes from transitions 6 and 7 which are weak and usually not observed in the UV-Vis region. Transition 11 is forbidden in this region and other transitions (8, 9, and 10) are uncertain.

A typical spectrum of ferric AXCP is shown in (Figure 1.15) to illustrate the bands explained above.

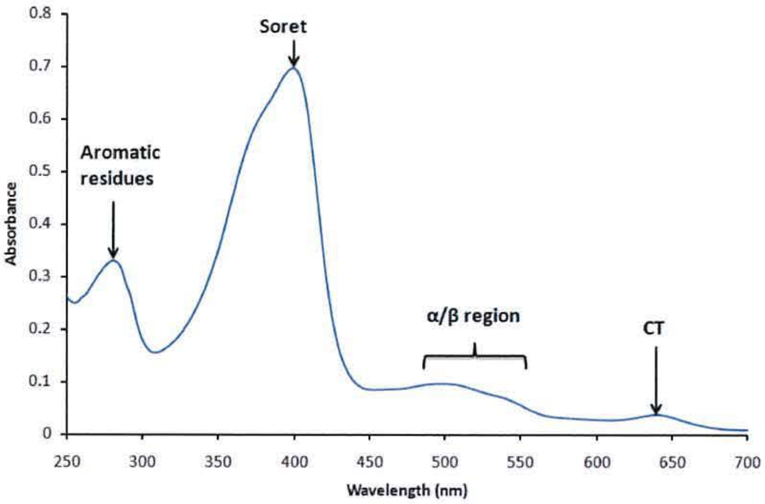


Figure 1.15: A typical electronic absorption spectra of cytochrome c′ from *Alcaligenes xylosoxidans*. Four distinct bands can be noticed; the Soret which arises from ($\pi \rightarrow \pi^*$) of the porphyrin ring accompanied by a weaker band α/β region band, Charge Transfer (CT) band which is related to the spin state and the coordination number of iron, and finally a band at around 280 nm arises from the absorption of the aromatic residues.

At pH 7.2 the ferric-AXCP exhibits a Soret band at 401 nm with a shoulder at a lower wavelength at 380 nm and two band at 535 and 498 nm related to α and β bands, respectively. A shift to the longer wavelength is observed in the ferrous form at 425 nm with a shoulder at 435 nm accompanied by a band at 553 nm at β region and 565 nm at α region.[119] The NO-bound AXCP showed the Soret band at 414 and 395 nm with bands at 484 and 566 nm, corresponding to β and α bands respectively. A sharp and intense Soret at 418 nm is observed in the case of CO-bound AXCP, the α/β region showing bands at 535 and 566 nm. These absorption bands are summarised in (Table 1.3):

AXCP	Soret	α	β
Ferric	401, 380sh	498	535
Ferrous	425, 435sh	553	565
NO-bound	395	484	566
CO bound	417	535	566

Table 1.3: Electronic spectral data for ferric, ferrous, NO- and CO-bound ferrocytochrome c′ from *Alcaligenes xylosoxidans*.

1.9 Scope and Aims of this work:

The distal pocket of AXCP is deeply buried within the protein with no direct access to solvent. The aim of this project was to probe the effect of the distal pocket residues in ligand binding and discrimination especially L16 and W56. L16 which is located in the distal face of haem, sterically hinders the access to the distal pocket. W56 apparently blocks the solvent channel that leads to the distal pocket. In order to understand the effect of these residues on the spectral properties and ligand binding and discrimination, a program of site-directed mutagenesis was used to mutate these two residues. Initially, L16 was mutated to glycine (L16G), phenylanalnine (L16F) and tyrosine (L16Y). In order to unblock the solvent channel W56 was mutated to glycine (W56G). With the purpose of investigating the effect of H-bonding on ligand binding and discrimination, W56 was mutated to tyrosine (W56Y). A double mutant was also studied in this work; L16 and W56 were mutated at the same time to (L16F/W56G).

To produce the native AXCP and all the above mutants in a reasonable amount, a reliable overexpression system and purification method has to be established.

1.9 References:

---

[1] Messerschmidt A., Huber R., Wieghardt K., and Poulos T., (2001) "Handbook of metalloproteins", Wiley, New York.

[2] Antonini E., Brunori M. (1971) "Hemoglobin and myoglobin in their reactions with ligands" North-Holland, Amsterdam.

[3] Gray H. B., Winkler J. R. (1996) "Electron transfer in proteins" *Annual Review of Biochemistry* **65**, 537-561.

[4] Simonneaux G., and Bondon A. (2005) "Mecahnism of electron transfer in heme proteins and models: the NMR approach" *Chemical Review* **105**, 2627-2646.

[5] Bernard Meunier, de Visser S. P., and Shaik S. (2004) "Mechanism of oxidation reactions catalyzed by cytochrome P450 enzymes" *Chemical Review* **104**, 3947-3980.

[6] Paternolli C., Ghisellini P., and Nicolinin C., (2007) "Nanostructuring of heme-proteins for biodevice applications" *IET Nanobiotechnology* **2**, 22-26.

[7] Gilles-Gonzalez MA, Gonzalez G., (2005) "Heme-based sensors: defining characteristics, recent developments and regulatory hypotheses" *Journal of Inorganic Biochemistry* **99**, 1-22.

[8] Ge B., Meyer T., Schöning M.J., Wollenberger U., and Lisdat F., (2000) "Cytochrome c′ from *Chromatium vinosum* on gold electrodes" *Electrochemistry Communications* **2**, 557-561.

[9] Lanzilotta W.M., Schuller D.J., Thorsteinsson M.V., Kerby R.L., Roberts G.P., Poulos T.L., (2000) "Structure of the CO sensing transcription activator CooA" *Nature Structural Biology* **7**, 876-880.

[10] Parone P.A., James D., Martinou J.C., (2002) "Mitochondria: regulating the inevitable" *Biochimie* **84**, 105-111.

[11] Orrenius S. (2004) "Mitochondrial regulation of apoptotic cell death" *Toxicology Letters* **149**, 19-23.

[12] Archer M., Banci L., Dikaya E., and Romão (1997) "Crystal structure of cytochrome c′ from *Rhodocyclus gelatinosus* and comparison with other cytochromes c′" *Journal of Biological Inorganic Chemistry* **2**, 611-622.

[13] Pokkuluri P.R., Londer Y.Y., Duke N.E.C., Erickson J., Pessanha M., Salgueiro C.A., and Schiffer M., (2004) *Protein Science* **13**, 1684-1692.

[14] Roldan M.D., Sears H.J., Cheesman M.R., Ferguson S.J., Thomson A.J., Berks B.C., and Richardson D.J., (1998) "Spectroscopic characterization of a novel multiheme c-type cytochrome widely implicated in bacterial electron transport" *Journal of Biological Chemistry* **273**, 28785-28790.

[15] Igarashi N., Moriyama H., Fujiwara T., Fukumori Y., and Tanaka N., (1997) " The 2.8 Å structure of hydroxylamine oxidoreductase from a nitrifying chemoautotrophic bacterium *Nitrosomonas europaea*" *Nature Structural Biology* **4**, 276-284.

[16] Florens L., and Bruschi M., (1994) " Recent advances in the characterization of the hexadecahemic cytochrome c from *Desulfovibrio*" *Biochimie* **76**, 561-568.

---

[17] Tsukihara T., Aoyama H., Yamashita E., Tomizaki T., Yamaguchi H., Shinzawa-Itoh K., Nakashima R., Yaono R., and Yoshikawa S., (1995) "Structures of metal sites of oxidised bovine heart cytochrome c oxidase at 2.8 Å" *Science* **269**, 1069-1074.

[19] Chapman S. K., Daff S., and Munro A. W. (1997) "Haem: the most versatile redox centre in biology" *Structure and Bonding* **88**, 39-70.

[20] Perotti C., Fukuda H., DiVenosa G., McRobert A.J., Batlle A., and Casas A., (2004) "Porphyrin synthesis from ALA derivatives for photodymanic therapy. *In vitro* and *in vivo* studies" *British Journal of Cancer* **90**, 1660-1665.

[21] Dailey H.A., (2002) "Terminal steps of haem biosynthesis" *Biochemical Society Transactions* **30**, 590-595.

[22] Sellers V.M., Wu CK., Dailey T. A., and Dailey H.A., (2001) "Human ferrochelatase: characterization of substrate-iron binding and proton-abstracting residues" *Biochemistry* **40**, 9821-9827.

[23] Mogi T., Saiki K., and Anraku Y. (1994) "Biosynthesis and functional role of haem O and haem A" *Molecular Microbiology* **14**, 391-398.

[24] Caughey W. S., Smythe G. A., O'Keeffe D. H., Maskasky J. E., and Smith M. L. (1975) "Heme A of cytochrome c oxicase. Structure and properties: comparison with hemes B, C, and S and derivatives" *Journal of Biological Chemistry* **250**, 7602-7622.

[25] Yamashita E., Aoyama H., Yao M., Muramoto K., Shinzawaltoh K., Yoshikawa S., and Tsukihara T. (2005) "Absolute configuration of the hydroxyfarnesylethyl group of haem A, determined by X-ray structural analysis of bovine heart cytochrome c oxidase using methods applicable at 2.8 Å resolution" *Acta Crystallographica* **D61**, 1373-1377.

[26] Hamel P., Corvest V., Giegé P., Bonnard G. (2009) "Biochemical requirements for the maturation of mitochondrial c-type cytochromes" *Biochimica et Biophysica Acta* **1793**, 125-138.

[27] Matsushita K., Shinagawa E., Adachi O., and Ameyama M., (1990) "Cytochrome a_1 of *Acetobacter aceti* is a cytochrome ba functioning as ubiquinol oxidase" *Proceedings of the National Academy of Science USA* **87**, 9863-9867.

[28] Argade P.V., Ching Y.C., Rousseau D.L., (1984) "Cytochrome a_3 structure in carbon monoxide-bound cytochrome oxidase" *Science* **225**, 329-331.

[29] Smith J. M. A., Aqirk A. V., Plank R. W. H., Diffin F. M., Ford G. C., and Harrison P. M., (1988) "The identification of *Escherichia coli* bacterioferritin and cytochrome b_1" *Biochemical Journal* **255**, 737-740.

[30] Capeillére-Blandin C., and Albani J., (1987) "cytochrome b_2, an electron carrier between flavocytochrome b_2 and cytochrome c" *Biochemical Journal* **245**, 159-165.

[31] Stonehuerner J., O'Brient P., Gerent L., Millett F., Steid J., Yu L., and Yu C., (1985) "Identification of the binding site on cytochrome c_1 for cytochrome c" *The Journal of Biological Chemistry* **260**, 5392-5398.

[32] Bruschi M., Leroy G., Bonicel J., Campese D., and Dolla A. (1996) "The cytochrome c_3 superfamily: amino acid sequence of a diemeric octahaem cytochrome c_3 (M(r) 26,000) isolated from Desulfovibrio gigas" *Biochemical Journal* **320**, 933-938.

---

[33] Nomenclature Committee of the International Union of Biochemistry (NCIUB) (1991) "Nomenclature of electron-transfer proteins – recommendations 1989" *European Journal of Biochemistry* **200**, 599-611.

[34] Stevens J. M., Daltrop O., Allen J. W. A., and Ferguson S. J. (2004) "C-type cytochrome formation: chemical and biological enigmas" *Accounts of Chemical Research* **37**, 999-1007.

[35] Ambler R. P. (1991) "Sequence variability in bacterial cytochromes c" *Biochimica et Biophysica Acta* **1058**, 42-47.

[36] Fülöp V., Sam K. A., Ferguson S. J., Ginger M. L., and Allen J. W. A. (2009) "Structure of a trypanosomatid mitochondrial cytochrome c with heme attached via only one thioether bond and implications for the substrate recognition requirements of heme lyase" *The FEBS Journal* **276**, 2822-2832.

[37] Allen J. W. A., Ginger M. L., Ferguson S. J. (2004) "Maturation of the unusual single-cysteine (XXXCH) mitochondrial c-type cytochromes found in trypanosomatids must occur through a novel biogenesis pathway" *Biochemical Journal* **383**, 537-542.

[38] Juêngst A., Wakakbayashi S., Matsubara H., and Zumft W. G. (1991) "The nirSTBM region coding for cytochrome cd$_1$-dependent nitrite respiration of *Pseudomonas stutzeri* consists of a cluster of mono-, di-, and tetraheme proteins" *FEBS Letters* **279**, 205-209.

[39] Hartshorne S., Richardson D. J., and Simon J. (2006) "Mutiple haem lyase genes indicate substrate specificity in cytochrome c biogenesis" *Biochemical Society Transactions* **34**, 146-149.

[40] Eaves D. J., Grove J., Staudenmann W., James P., Poole R. K., White S. A., Griffiths I., and Cole J. A. (1998) "Involvement of products of the *nrfEFG* genes in the covalent attachment of haem c to a novel cysteine-lysine motif in the cytochrome c$_{552}$ nitrite reductase from *Escherichia coli*" *Molecular Biology* **28**, 205-216.

[41] Thöny-Meyer L. (1997) "Biogenesis of respiratory cytochromes in bacteria" *Micrbiology and Molecular Biology Review* **61**, 337-376.

[42] Kranz R., Lill R., Goldman B., Bonnard G., and Merchant S. (1998) "Molecular mechanism of cytochrome c biogenesis: three distinct systems" *Molecular Microbiology* **29**, 386-396.

[43] Spielewoy N., Schulz H., Grieneberger JM., Thöny-Meyer L., Bonnard G. (2001) "CCME, a nuclear-encoded heme-binding protein involved in cytochrome c maturation in plant mitochondria" *Journal of Biological Chemistry* **276**, 5491-5497.

[44] Meyer E. H., Giegé P., Gelhaye E., Rayapuram N., Ahuja U., Thöny-Meyer L., Grieneberger JM., and Bonnard G. (2005) "Cytochrome maturation pathway in *Arabidopsis* mitochondria, interacts with apocytochrome c" *Proceeding of the National Academy of Sciences of USA* **102**, 16113-16118.

[45] Allen J. A., Daltrop O., Stevens J. M., Ferguson S. J. (2003) "C-type cytochromes: diverse structures and biogenesis systems pose evolutionary problems" *Philosophical Transactions of the Royal Society* **358**, 255-266.

[46] Thöny-Meyer L. (2003) "A heme chaperone for cytochrome c biosynthesis" *Biochemistry* **42**, 13099-13105.

[47] Schulz H., Fabianek R. A., Pellicioli E. C., Hennecke H., and Thöny-Meyer L. (1999) "Heme transfer to the heme chaperone CcmE during cytochrome c maturation requires the CcmC protein, which may function independently of the ABC-transporter CcmAB" *Proceeding of the National Academy of Science of the USA* **96**, 6462-6467.

---

[48] Christensen O., Harvat E. M., Thöny-Meyer L., Ferguson S. J. (2007) "Loss of ATP hydrolysis activity by CcmAB results in loss of c-type cytochrome synthesis and incomplete processing of CcmE" *FEBS Journal* **274**, 2322-2332.

[49] Ahuja U., and Thöny-Meyer L. (2005) "CcmD is involved in complex formation between CcmC and the heme chaperone CcmE during cytochrome c maturation." *Journal of Biological Chemistry* **280**, 236-243.

[50] Ren Q., Ahuja U., and Thöny-Meyer L. (2002) "A bacterial cytochrome c heme lyase." *The Journal of Biological Chemistry* **277**, 7657-7663.

[51] Setterdahl A. T., Goldman B. S., Hirasawa M., Jacquot P., Smith A. J., Kranz R. G., and Knaff D. B. (2000) "Oxidation-reduction properties of disulfide containing properties of the *Rhodobacter capsulatus* cytochrome c biogenesis system." *Biochemistry* **39**, 10172-10176.

[52] Allen J. W., Jackson A. P., Rigden D. J., Willis A. C., Ferguson S. J., Ginger M. L. (2008) "Order within a mosaic distribution of mitochondrial c-type cytochrome biogenesis systems?" *The FEBES Journal* **275**, 2385-2402.

[53] Feissner R., Richard-Fogal C. L., Frawley E. R., and Kranz R. G. (2006) "ABC transporter-mediated release of a haem chaperone allowes cytochrome c biogenesis." *Molecular Microbiology* **61**, 219-213.

[54] Le Brun N. E., Bengtsson J., and Hederstedt L (2000) "Genes required for cytochrome c synthesis in *Bacillus subtilis*" *Molecular Microbiology* **36**, 638-650.

[55] Feissner R. E., Richard-Fogal C. L., Frawley E. R., Loughman J. A., Earley K. W., and Kranz R. G. (2006) "Recombinant cytochrome c biogenesis systems I and II and analysis of haem delivery pathways in *Escherichia coli*." *Molecular Microbiology* **60**, 536-577.

[56] Feissner R. E., Beckett C. S., Loughman J. A., and Kranz R. G. (2005) "Mutations in cytochrome assembly and periplasmic redox pathway in *Bodetella pertussis*." *Journal of Bacteriology* **187**, 3941-3949.

[57] Crow A., Acheson R. M., Le Brun N. E., and Oubrie A. (2004) "Structural basis of redox-coupled protein substrate selection by cytochrome c biosynthesis protein ResA." *Ournal of Biological Chemistry* **279**, 23654-23660.

[58] Bernard D. G., Gabilly S. T., Dujardin G., Merchant S., and Hamel P. P. (2003) "Overlapping specificities of the mitochondrial cytochrome c and c_1 heme lyases" *The Journal of Biological Chemistry* **278**, 49732-49742.

[59] Zollner A., Rödel G., and Haid A (1992) "Molecular cloning and characterization of the *S. cerevisiae* CYT2 gene encoding cytochrome-c_1-heme lyase" *European Journal of Biochemistry* **207**, 1093-1100.

[60] Dumont M. E., Ernst J. F., Hampsey D. M., and Sherman F. (1987) "Identification and sequence of gene encoding cytochrome c heme lyase in the yeast *Saccharomyces cerevisiae*" *EMBO Journal* **6**, 235-241.

[61] Ambler R. P. (1991) "Sequence variability inbacterial cytochromes c" *Biochimica et Biophysica Acta* **1058**, 42-47.

[62] Frazão C., Sieker L., Sheldrick G., Lamzin V., LeGall J., and Carrondo M. A. (1998) "*Ab initio* structure solution of a diemeric cytochrome c_3 from *Desulfovibrio gigas* containing disulfide bridges" *Journal of Biological Inorganic Chemistry* **4**, 162-165.

---

[63] Meyer T. E., Kamen M. D. (1982) "New perspectives on c-type cytochromes."*Advances in Protein Chemistry* **35**, 105-212.

[64] Iwasaki H., Yoshimura T., Suzuki S., Shidara S. (1991) "Spectral properties of *Achromobacter xylosoxidans* cytochrome c′ and their NO complexes." *Biochimica et Biophysica Acta* **1058**, 79-82.

[65] Yamanaka T., and Imai S. (1972) "A cytochrome cc′-like haemoprotein isolated from *Azotobacter vinelandii.*" *Biochemical and Biophysical Research Communications* **14**, 150-154.

[66] Schmidt T. M., and DiSpirito A. A. (1990) "Spectral characterization of c-type cytochromes purified from *Beggiatao alba.*" *Advance of Microbiology* **154**, 453-458.

[67] Kassner R. J. (1991) "Ligand binding properties of cytochromes c′." *Biochimica et Biophysica Acta* **1058**, 8-12.

[68] Ramirez L. M., Axelrod H. L., Herron S. R., Rupp B., Allen J. P., Kantardjieff K. A. (2003) "High resolution crystal structure o ferricytochrome c′ from *Rhodobacter sphaeroides.*" *Journal of Chemical Crystallography* **33**, 413-424.

[69] Vernon L. P., and Kamen M. D. (1954) "Hematin compounds in photosynthetic bacteria." *Journal of Biological Chemistry* **211**, 643-662.

[70] Cusanovich M., and Gibson Q. (1973) "Anomalous ligand binding by a class of high spin c-type cytochromes." *The Journal of Biological Chemistry* **248**, 822-834.

[71] Pixton D. A., Petersen C. A., Franke A., van Eldik R., Garton E. M., Andrew C. R. (2009) "Activation parameters for heme-NO binding in *Alcaligenes xylosoxidans* cytochrome c′: the putative dinitrosyl intermediate forms via a dissociation mechanism." *Journal of the American Chemical Society* **131**, 4846-4853.

[72] Tahirov T. H., Misaki S., Meyer T. E., Cusanovich M. A.,Higuchi Y., and Yasuoke N. (1996) "High-resolution crystal structure of two polymorphs of cytochrome c′ from purple phototropic bacterium *Rhodobacter capsulatus.*" *Journal of Molecular Biology* **259**, 467-479.

[73] DeLano, W.L. (2002) "The PyMOL Molecular Graphics System" DeLano Scientific, Palo Alto, CA, USA. **http://www.pymol.org**

[74] Huston W. M., Andrew C. R., Servid A. E., McKay A. L., Leech A. P., Butler C. S., and Moir J. W. B. (2006) "Heterologous overexpression and purification of cytochrome c′ from *Rhodobacter capsulatus* and a mutant (K42E) in the dimerization region. Mutation does not alter oligomerization but impacts the heme iron spin state and nitric oxide binding properties." *Biochemistry* **45**, 4388-4395.

[75] Choi P. S., Grigoryants V. M., Abruña H. D., Scholes C. P., Shapleigh J. P. (2005) "Regulation and function of cytochrome c′ in *Rhodobacter sphaeroides* 2.4.3" *Journal of Bacteriology* **187**, 4077-4085.

[76] Berman H. M., *et al.* (2000) "The protein data bank" *Nucleic acid research* **28**, 235-242.

[77] Tahirov T. H., Misaki S., Meyer T. E., Cusanovich M. A., Higuchi Y., and Yasuoka N. (1996) "High-resolution crystal structures of two polymorphs of cytochrome c from purple phototropic bacterium*Rhodobacter capsulatus.*" *Journal of Molecular Biology* **259**, 467-479.

[78] Ambler R. P., Bartsch R. G., Daniel M., Kamen M. D., McLellan L., Meyer T. E., and van Beeumen J. (1981) "Amino acid sequences of bacterial cytochromes c′ and c-556." *Proceedings of the National Academy of Sciences of the USA* **78**, 6854-6857.

---

[79] Yabuuchi E., and Oyama A. (1971) "Achromobacter xylosoxidans n. sp. from human ear discharge" *Japanese Journal of Microbiology* **15**, 477-481.

[80] Tsay R. W., Lin L. C, Chiou Liao J. C., Chen C. H., Liu C. E., Young T. G. (2005) "*Alcaligenes xylosoxidans* bacteremia: clinical features and microbiological characteristics of isolates" *Journal of microbiology, Immunology, and Infection* **38**, 194-199.

[81] Dobbs A. J., Anderson B. F., Faber H. R., Baker E. N. (1996) "Three-dimensional structure of cytochrome c′ from two *Alcaligenes* species and the implication for four-helix bundle structures" *Acta Crystallographica* **D52**, 356-368.

[82] Lawson D. M., Stevenson C. E. M., Andrew C. R., Eady R. R. (2000) "Unprecedented proximal binding of nitric oxide to heme: implications for guanylate cyclise" *EMBO Journal* **19**, 5661-5671.

[83] Weiss R., Gold A., and Terner J. (2006) "Cytochromes c′: Biological model for the $S=3/2$, 3/5 spin state admixture?" *Chemical Reviews* **106**, 2550-2579.

[84] Benini S., Rypniewski W. R., Wilson K., and Ciurli S. (2008) "High resolution crystal structure of *Rubrivivax gelatinosus* cytochrome c′" *Journal of Inorganic Biochemistry* **102**, 1322-1328.

[85] Ehrenberg A., and Kamen M. D. (1965) "Magnetic and optical properties of some bacterial hame proteins" *Biochimica et Biophysica Acta* **102**, 333-340.

[86] Rawlings J., Stephens P. J., Nafie L. A., and Kamen M. D. (1977) "Near-infrared magnetic circular dichroism of cytochrome c′" *Biochemistry* **16**, 1725-1729.

[87] Emptage M. H., Zimmermann R., Que L.J., Münch E., Hamilton W. D., and Orme-Johnson W. H. (1977) "Mössbauer studies of cytochrome c′ from *Rhodospirillum rubruml*" *Biochimica et Biophysica Acta* **495**, 12-23.

[88] Banci L., Bertini I., Turano P., and Oliver M. V. (1992) "NOE and two-dimensional correlated 1H-NMR spectroscopy of cytochrome c′ from *Chromatium vinosum*" *European Journal of Biochemistry* **204**, 107-112.

[89] Maltempo M. M. (1974) "Magnetic state of an unusual bacterial heme protein" *The Jounal of Chemical Physics* **61**, 2540-2547.

[90] Smulevich G., Feis A., and Howes B. D. (2005) "Fifteen years of Raman spectroscopy of engineered heme containing peroxidaese: what have we learned?" *Accounts of chemical research* **38**, 433-440.

[91] Fujii S., Yoshimura T., Kamada H., Yamaguchi K., Suzuki S., Shidara S., and Takakuwa S. (1995) "Electron paramagnetic resonance studies of ferric cytochrome c′ from photosynthetic bacteria" *Biochimica et Biophysica Acta* **1251**, 161-169.

[92] Othman S., Richaud P., Verméglio A., and Desbois A (1996) "Evidence for proximal histidin interaction in the structure of cytochromes c′ in solution: A Resonance Raman study" *Biochemistry* **35**, 9224-9234.

[93] Scheidt W. R., Reed C. (1981) "Spin-state/stereochemical relationship in iron porphyrins: Implication for the hemoproteins" *Chemical Reviews* **81**, 543-555.

[94] Cheng R. J., Chen P. Y., Gau P. R., Chen C. C., and Peng S. M. (1997) "Control of spin state by ring conformation of iron(III) porphyrins. A novel model for the quantum-mixed intermediate spin state of ferric cytochrome c′ from photosynthetic bacteria" *Journal of American Chemical Society* **119**, 2563-2569.

---

[95] Rodger K. R. (1999) "Heme-based sensors in biologival systems" *Current Opinion in Chemical Biology* **3**, 158-167.

[96] Ignarro L. J. (2002) "Nitric oxide as a unique signalling molecule in the vascular system: A historical review" *Journal of Physiology and Pharmacology* **53**, 503-514.

[97] Moncada S. Higgs E. A. (2006) "The discovery of nitric oxide and its role in vascular biology" *British Journal of Pharmacology* **147**, S193-S201.

[98] MacMicking J., Xie Q., and Nathan C. (1997) "Nitric oxide and macrophage function" *Annual Review of Immunology* **15**, 323-350.

[99] Tsoyi K., Lee T. Y., Lee Y. S., Kim H. J., Seo H. G., Lee J. H., and Chang K. C. (2009) "Heme-oxygenase-1 induction and carbon monoxide-releasing molecule inhibit lipopolysaccharide (LPS)-induced high-mobility group box 1 release in vitro and improve survival of mice in LPS- and cecal ligation and puncture-induced sepsis model in vivo" *Molecular Pharmacology* **76**, 173-182.

[100] Lahiri S., Roy A., Baby S. M., Hoshi T., Semenza G. L., and Prabhakar N. R. (2006) "Oxygen sensing in the body" *Progress in Biophysics and Molecular Biology* **91**, 249-286.

[101] Roberts G. P., and Youn H., Kerby R. L. (2004) "CO sensing mechanism" *Microbiology and Molecular Biology Review* **68**, 453-473.

[102] Ascenzi P., Bocedi A., Leoni L., Visca P., Zennaro E., Milani M., and Bolognesi M. (2004) "CO sniffing through heme-based sensor proteins" *IUBMB Life* **56**, 309-315.

[103] Bailey_Serres J., and Chang R. (2005) "Sensing and signalling in response to oxygen deprivation in plants and other organisms" *Annals of Botany (Oxford)* **96**, 507-518.

[104] Xu C., and Spiro T. G. (2008) "Ambidentate H-bonding by heme-bound NO: structural and spectral effects of –O versus –N H-bonding" *Journal of Biological Inorganic Chemistry* **13**, 613-621.

[105] Quillin M. L., Arduini R. M., Olson J. S., and Jr G. N. P. (1993) "High resolution crystal structure of distal histidin mutants of sperm whale myoglobin" *Journal of Molecular Biology* **234**, 140-153.

[106] Lyer L. M., Anantharaman V., and Aravind L. (2003) "Ancient conserved domains shared by animal soluble guanylyl cyclises and bacterial signalling proteins" *BMC Genomics* **4**, 5.

[107] Boon E. M., Huang S. H., Marletta M. A.. (2005) "A molecular basis for NO selectivity in soluble guanylate cyclase" *Nature Chemical Biology* **1**, 53-59.

[108] Stone J., and Marletta M. A. (1994) "Soluble guanylate cyclase from bovine lung: Activation with nitric oxide and carbon monoxide and spectral characterization of the ferrous and ferric states" *Biochemistry* **33**, 5636-5640.

[109] Meyer T. E., and Kamen M. D. (1982) "New perspectives on c-type cytochromes" *Advances in Protein Chemistry* **35**, 105-212.

[110] Vogel I. K. M., Hu S. Z., Spiro T. G., Dierks E. A., Y A. E., and Burstyn J. N. (1999) "Variable forms of soluble guanylyl cyclase: protein-ligand interactions and the issue of activation by carbon monoxide" *Journal of Biological Inorganic Chemistry* **4**, 804-813.

[111] Andrew C. R., George S. J., Lawson D. M., and Eady R. R. (2002) "Six- to five- coordinate heme-nitrosyl conversion in cytochrome c′ and its relevance to guanylate cyclase" *Biochemistry* **41**, 2353-2360.

---

[112] Yoshimura T., Suzuki S., Nakahara A., Iwasaki H., Masuko M., and Matsubara T. (1986) "Spectral properties of nitric oxide complexes of cytochrome c′ from *Alcaligenes sp.* NCIMB 11015." *Biochemistry* **25**, 2436-2442.

[113] Lawson D. M., Stevenson C. E. M., Andrew C. R., George S. J., and Eady R. R. (2003) "A two-faced molecule offers NO explanation: the proximal binding of nitric oxide to haem" *Biochemical Society Transactions* **31**, 553-557.
[114] Traylor T. G., and Sharma V. S. (1992) "Why NO?" *Biochemistry* **31**, 2847-2849.

[115] Yoo B. K., Lamarre I., Martin J. L., Andrew C. R., Nioche P., and Negrerie M. (2008) "Understanding the NO-sensing mechanism at molecular level" *EU-Korea Conference on Science and Technology,* Springer Proceeding in Physics (Springer Berlin), Ed: Yoo S. D., **124**, 517-524.

[116] Kruglik S. G., Lambry J. C., Cianetti S., Martin J. L., Eady R. R., Andrew C. R., and Negrerie M. (2007) "Molecular basis of nitric oxide dynamics and affinity with *Alcaligenes xylosoxidans* cytochrome c′" *The Journal of Biological Chemistry* **282**, 5053-5062.

[117] Moor G. R., and Pettigrew G. W. (1990) "Cytochromes c: evolutionary, structural and physicochemical aspects" Springer-Verlag, New York.

[118] Spiro T. G (1983) "The resonance Raman spectroscopy of metalloporphyrins and haemoproteins" in Iron Porphyrins, Part two, Edited by Leve A. and Gary H. B., Adison Wesley, London Amestrdam, 89-159.

[119] Shidara S., Iwasaki H., Yoshimura T., Suzuki S., and Nakahara A. (1986) "Purification and some properties of cytochrome c′ from a strain of *Achromobacter xylosoxidans*" *The Journal of Biochemistry* **99**, 1749-1752.

Chapter 2: Protein Crystallography

2.1 Introduction:

The primary tool for determining the atomic structure of macromolecules is the diffraction of x-ray radiation by crystals. The crystalline protein which can be considered as regular arrays of molecules interact with short wavelength electromagnetic radiation causing diffraction. The diffraction pattern of a substance then can be used for the reconstruction of the electron density. If the resolution of the diffraction pattern is sufficient, the atomic coordinates of the molecule can be obtained. The diffraction of x-ray radiation with a crystalline matter has been mathematically described Bragg in 1918.[1]

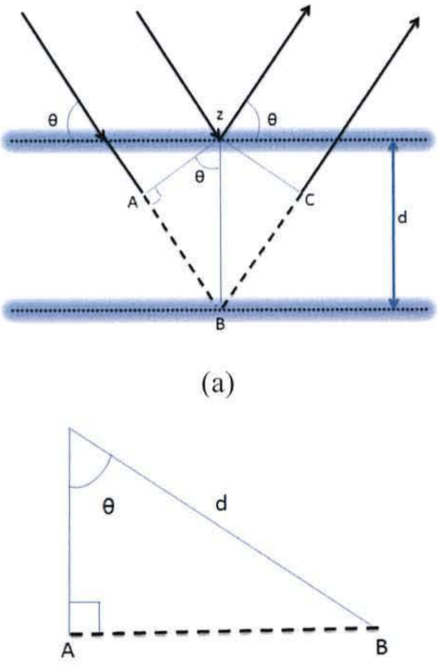
At the end of most imaging methods, an image of the sample will be elucidated. In crystallography what will be obtained at the end of the experiment is the diffraction pattern of the image which is related to the Fourier Transform (FT) of the object. In the real life, when light is scattered from an object, it scatters as what can be described mathematically as a FT. Then the lens of the eyes cornea takes the scattered light effectively does an inverse FT to turn it back into the image of the object on the retina. To probe atomic positions in a protein a very short wavelength radiation should be used which can resolve the distances between atoms, *ca* 1 Å which is comparable to the wavelength of x-rays. The need to use x-ray introduces a fundamental problem arising from the failure of getting the phase information of the data. During the course of collecting the information from the FT on a detector as the diffraction pattern, half of the information, the "phase" information, from the FT is lost, while the other half of the information which is not lost is the "amplitude" information. The loss of phase information is a famous problem in crystallography named a (phase problem). The phase problem must be solved before the electron density of the protein can be calculated. In later sections of this chapter we will discuss briefly why the phase information lost and how to solve this problem.[2]

Since the principles of protein x-ray crystallography have been extensively discussed and explained in many textbooks, [3,4] only the basic principles will be discussed herein:

2.1 Basic Theory of Protein Crystallography:

W.L Bragg[1] showed that the diffraction of an x-ray from a crystal can be described as the scattering from parallel planes of molecules, and can be mathematically described as a reflection from sets of equivalent parallel planes of atoms in the crystal. Constructive interference between the scattered x-ray from successive planes will take place only if the path difference between these rays is equivalent to an integral number of wavelengths:

A set of equivalent, parallel planes with index (hkl) and interplanar distance (d_{hkl}) produces a diffracted beam when x-ray of wavelength λ incide on the planes at an angle θ and are reflected out at the same angle only when it meets the Bragg's condition (Figure 2.1): $n\lambda = 2\,d\,sin\theta$



(a)

(b)

Figure 2.1: (a) Description of Bragg's law using the reflection geometry. The lower beam must travel the extra distance (AB + BC) to continue travelling parallel and adjacent to the upper beam. This extra distance must be an integral (n) multiple of the wavelength (λ) ($n\lambda = AB + BC$) for the phase of the two beams to be the same. In (b) we can see that ($AB = 2d\ sin\theta$), and because ($AB = BC$), the first equation can be written as ($n\lambda = 2AB$), by substitution of this one into the second equation produces Bragg's equation ($n\lambda = 2d\ sin\theta$).

The goal of protein x-ray crystallography is to calculate the electron density (**ρ**) at every position (*x, y, z*) in the unit cell of a crystal. The electron density map can be derived from the structure factors ***F****(hkl)* and their phase angles ($\propto_{\boldsymbol{hkl}}$) of each reflection.

$$\boldsymbol{F}(h\,k\,l) = \sum_{j=1}^{N} f_j \exp\left(2\pi i\,(hx_j + ky_j + lz_j)\right) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2.1)$$

f_j is the scattering factor of j^{th} atom,

h, k, l defining the coordinates in reciprocal space (reflections) x_j, y_j, and z_j is the coordinated for the j^{th} atom.

The electron density **ρ(x y z)** in the unit cell volume (***V***) is the FT of the structure factor and the reverse is also true:

$$\rho(x\,y\,z) = \frac{1}{V}\sum_h \sum_k \sum_l \boldsymbol{F}(h\,k\,l) \exp\left[-2\pi i(hx + ky + lz)\right] \ldots\ldots\ldots\ldots\ldots (2.2)$$

Since the structure factor can be written as:

$$\boldsymbol{F}(hkl) = |F_{hkl}| \exp[i \propto_{hkl}] \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2.3)$$

F_{hkl} is the amplitude that is related to the intensity of a diffracted beam and directly measured on the detector. The intensity of a reflection with Miller indices *hkl* ($\boldsymbol{I_{hkl}}$) is proportional to $|\boldsymbol{F_{hkl}}|^2$.

$\propto_{hkl}$ is the phase of the structure factor for each reflection.

i is the imaginary number $(-1)^{1/2}$.

Substitution of equation 2.3 in 2.2 gives:

$$\rho(x\,y\,z) = \frac{1}{V}\sum_h \sum_k \sum_l |F_{hkl}| \exp[-2\pi i(hx + ky + lz) + i \propto_{hkl}] \ldots\ldots (2.4)$$

In principle calculating the electron density $\boldsymbol{\rho(x\,y\,z)}$ at every position $(\boldsymbol{x\,y\,z})$ in the unit cell appears straightforward. However, there is a problem; although $|\boldsymbol{F_{hkl}}|$s can be calculated from the intensities $\boldsymbol{I_{hkl}}$, the phase of the structure factor cannot be derived straightforwardly from the diffraction pattern. A famed problem arises which is called, the "phase problem". Fortunately, several methods have been developed to solve this problem; they have been discussed in (section 2.7)

2.3 Methodology:

There are some prerequisites before a macromolecular crystal structure determination can be performed; such as the availability of sufficient amounts of highly purified protein and crystals of sufficient diffraction quality for a high resolution structure determination.[5] Protein overproduction and purification has been discussed in (chapter 4) and principles of growing x-ray quality crystals and the data collection and processing will be discussed in the following sections.

2.4 Data collection and processing:

As soon as a suitable protein crystal has grown, the next step is to check whether it diffracts well enough to be used for the three-dimensional structure solution. This test involves fishing the crystal from its mother liquor, immersing it into a proper cryoprotectant and mounting the crystal in a cryogenic nitrogen stream on an x-ray source. Finding a suitable cryoprotectant is an important part of the experiment. Once a diffracting crystal is found, the actual process of data collection can be commenced.

2.4.1 X-ray radiation sources:

X-ray radiation sources are usually either a rotating anode generator in the home laboratory or an electron storage ring producing synchrotron radiation at a national laboratory.

X-rays from a rotating anode are generated by bombarding an anode made up of a pure elemental metal, usually copper or molybdenum, with a beam of electrons generated from a hot filament by a high voltage electric field. This bombardment results in driving an electron from an atomic orbital of the metal resulting in a vacancy in the orbital, which is then filled by the relaxation of an electron from a higher energy shell and emission of an x-ray photon. The energy of the emitted radiation depends on the electronic structure of the emitting element. Thus, a Cu anode produces 1.54 Å x-rays, while Mo produces 0.709 A x-rays.

An important x-ray source, which is especially useful for crystallographic measurements, is provided by synchrotron radiation. In general they consist of a linear accelerator (LINAC), a booster synchrotron and a doughnut shaped storage ring. Electrons generated in the electron gun are directed into the LINAC where they are energised, and then into the second part; the booster, where electrons are further energised to nearly the speed of light. Finally, they are injected as 'bunches' of electrons into the storage ring. The stored electrons are guided in a 'race track' shaped path using strong magnetic fields provided by dipole magnets. Each time the electron bunches change their direction (accelerate) they emit electromagnetic rays, called synchrotron light. Third generation synchrotron radiation sources are capable of producing high brilliance x-rays (e.g. the DIAMOND Light Source emits 3.8 x 10^{15} photons. s^{-1}. mm^{-3}) using insertion devices. These devices consist of arrays of dipole magnets with alternating polarities placed in straight section of the storage ring. They can be designed to improve specific characteristics of the synchrotron radiation; for example, a wavelength shifter can extend the spectral range to shorter wavelengths, wigglers increase the intensity, and undulators increase the brilliance and produce a quasi-monochromatic beam.[6] The synchrotron light is directed along beamlines into experimental hutches, where the experiments take place. Synchrotrons are capable of producing electromagnetic radiation from infra-red energies to hard x-rays.

There are three reasons that make synchrotron radiation especially useful for crystallographic measurements. First, synchrotrons are 6-8 orders of magnitude more intense than home sources,[7] which means diffraction experiments can be done more rapidly (e.g. exposure times of less than one second are possible). However, it should be noted that home sources can now achieve x-ray fluxes that are comparable to bending magnets on 2nd generation SR sources, for example the Rigaku FRE+. Second, it produces well collimated rays, which, in principle, means complex diffraction patterns (featuring many reflections) from dramatically large complexes can be well resolved.[8,9] The third advantage is the range of x-ray photon energies or wavelengths produced allow selection of suitable wavelengths for anomalous diffraction from heavy atoms in the macromolecule which can be used to solve the phase problem.[10]

2.4.2 Growing the x-ray quality crystals:

2.4.2.1 Principles:

The first requirement for protein structure determination is to grow suitable crystals of highly purified sample. Crystals are a periodic array of protein molecules, where each element of the smallest repeating unit (asymmetric unit) (Figure 2.2a, an example with one protein molecule in the asymmetric unit), assemble neighbouring molecule in a geometrical way related by symmetry operations of the space group (Figure 2.2b) forming a unit cell (Figure 2.2c), which in turn assembled by repeated deposition in a periodic manner to form a lattice (Figure 2.2d) and hence the crystal.

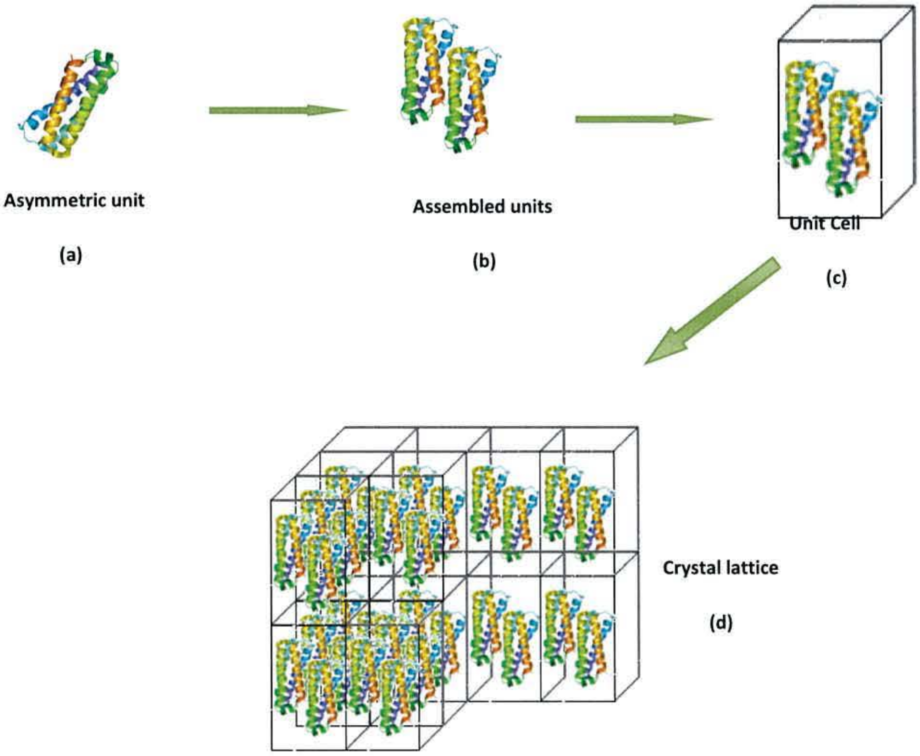


Figure 2.1: The smallest repeating unit of the crystal, named the asymmetric unit (a), assembles itself with its neighbouring molecule in a geometrical way related by symmetry operation (b) forms a unit cell (c) which is assembled by repeated deposition in a periodic manner to for a lattice (d) and hence the crystal.

Crystallisation is usually achieved by changing physical parameters affecting the solubility of the protein in order to reach a state of supersaturation. Factors include changing the protein concentration, temperature, pH, ionic strength and the presence of precipitating agents, such as ammonium sulphate or polyethyleneglycol (PEG).

To obtain good structural data, crystals need to be single and have dimensions of at least 10 μm^3, preferably much larger. However, smaller crystals are becoming more viable with technology advances at latest generation SR sources. Crystal formation occurs in two steps; nucleation and growth. Nucleation sites for crystal growth are formed by chance collisions of molecules in solution, forming molecular aggregates. This happens when protein and/or the precipitant concentrations are higher than the concentrations required for slow precipitation, as shown in the blue region of phase diagram (Figure 2.3). Once nuclei have been formed, the next step starts; the actual crystal growth. The step-wise attachment of new molecules to the surface of a growing crystal happens because the binding energy of molecules to the growing crystal surface is larger than if they attached to a flat surface.[11] The formation of too many nuclei results in the formation of too many small crystals. To prevent this, the supersaturation point must be reduced to a lower level. Also, crystals should grow slowly to allow the molecules to order themselves into a crystalline lattice and to reach a maximum degree of order in their structure. In a highly supersaturated solution the molecules may aggregate to form an amorphous precipitate.

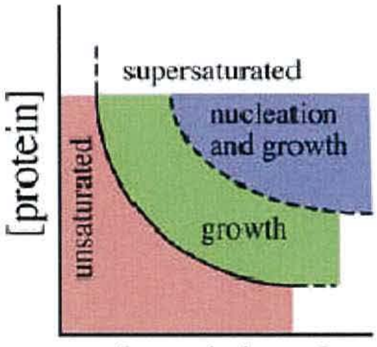


Figure 2.3: Phase diagram for crystallisation mediated by a precipitant. The red region represents concentrations of protein and precipitant at which the solution is not saturated with protein, so neither nucleation nor growth occurs. The green and blue regions represent unstable solutions that are supersaturated with protein. Conditions in the blue region support both nucleation and growth, while conditions in the green region support growth only. Picture and caption taken from Rhodes.[3]

The ideal crystallisation strategy has been illustrated in (Figure 2.4). It starts with a condition that corresponds to the blue region of the phase diagram (Figure 2.3). Once nuclei are formed, they move into the green region, where crystals start to grow but the nucleation process stops.

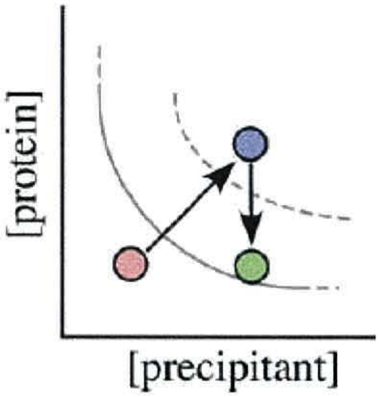


Figure 2.4: An ideal strategy for growing large crystals is to allow nucleation to occur under conditions in the blue region, then to move to conditions in the green region until crystal growth ceases. Picture and caption taken from Rhodes.[3]

Not all crystals can diffract x-rays to give good structural data. That is why the production of x-ray quality crystal is the rate limiting step in the structure determination even once protein have been solubilised and purified.[12] A statistical survey of 29 structural genomics projects from the beginning of 1998 to the 25th September 2009 when this thesis was submitted (Table 2.1). The table shows that from about 154,870 cloned proteins 24% of them were successfully expressed and purified, 8.8% crystallised, and only 4.6% of them diffracted in the x-ray beam.

Status	Records	% of Cloned ProteinTargets
Cloned	**154870**	**100**
Expressed	**110799**	**71.5**
Soluble	**43522**	**28.1**
Purified	**38444**	**24.8**
Crystallised	**13640**	**8.8**
Diffracting crystals	**7196**	**4.6**

Table 2.1: Statistical survey on the structural genomics from the January 1st 1998 to December 25th 2009. Numbers taken from http://targetdb.pdb.org, shows the sum of results from 29 structural genomics/proteomic projects worldwide as of 25th September 2009. (The numbers for each individual project vary as they are dealing with different genomes at different times).

2.4.2.2 Crystallisation techniques:

Many techniques have been developed to grow diffraction quality crystals. These techniques have been described extensively by Giegé and McPherson[13] and other authors.[14] Here the vapour diffusion technique will be discussed as it is a widely used technique and has been used to grow crystals in this work.

In the vapour diffusion crystallisation method, a volume of 1 μL – 10 μL of the macromolecule to be crystallised is mixed with an equal volume of the precipitant solution (precipitant, buffer and additives) in a drop, giving a precipitant concentration about 50% of that required for the protein crystallisation. The drop is then either suspended underneath a siliconised microscope glass cover slip, which is sealed with grease (this is called hanging drop method; Figure 2.5a), or sat on a support above the reservoir if the solution has a low surface tension; this is called the sitting drop method (Figure 2.5b). Because the precipitant is the major solute present, vapour diffusion results in net transfer of water from the drop to the reservoir until the precipitant concentration in the drop is the same as its concentration in the reservoir. This leads to the volume decrease of the drop, so that the concentration of all components in the drop increases, moving the conditions diagonally into the nucleation region. As the nuclei form, the protein concentration drops, moving the condition vertically into the growth region (Figure 2.3).

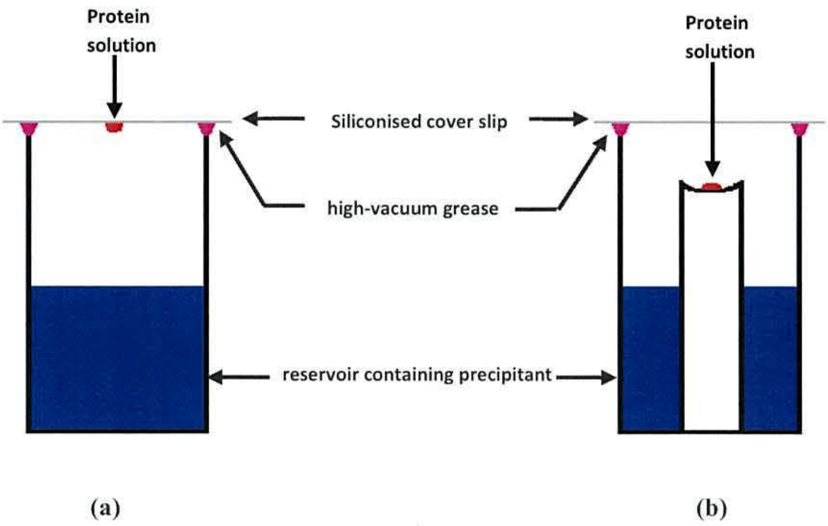


Figure 2.5: Vapour diffusion crystallisation method. An equal amount of the reservoir (blue) solution which contains the precipitant, buffer and additives, and a highly purified protein solution are mixed in a drop on a siliconised cover slip and the cover slip inverted and placed above the well (a), the method hence called hanging drop, or it is sat on a support (b), which is called the sitting drop method.

2.5 Data collection:

Water molecules comprise between 30 - 75% of the total crystal content. Therefore, protein crystals are often very fragile and should be handled with caution. Crystals are often harvested, examined, and mounted for crystallography within their mother liquor. This will prevent the crystal from dehydration and hence cracking.

Data collection starts by fishing and transferring a single crystal from its mother liquor with a minimum amount of mother liquor, using thin-walled capillaries of glass or quartz, or more often, as in this work, by using fiber loops. The loop is then mounted onto a goniometer, and the crystal centred where it will be exposed to the x-ray beam.

2.5.1 X-ray radiation damage:

When an x-ray beam is incident on a typical crystal, approximately 98% of the beam passes straight through without interacting, and then is absorbed by the beamstop which is located downstream of the crystal. The remaining 2% interacts in one of three ways: firstly, by the photoelectric effect (84%), in which a synchrotron photon ejects a lower shell electron from an atom causing ionisation and other processes; secondly, by Compton scattering (8%), which is inelastic and incoherent and adds background to the diffraction images; thirdly by Thomson scattering (8%), which is elastic and coherent and results in the observed diffraction pattern.[15] The ionisation process results in the formation of radicals,[16] which may then cause damage to the protein; this kind of damage is called "radiation damage". For example, radicals interact with the protein backbone, producing a stable carbonyl radical. The stability of the carbonyl radical makes carbonyl groups of the peptide bond act as a major trapping centre for electrons produced by the x-rays. Electrons then migrate along the peptide chain through the hydrogen bonds between peptide units. Affected side chains then become the place for structural damage, for example, cleavage of disulphide bridges, decarboxylation of glutamates and aspartates, loss of the tyrosine residue's hydroxyl group and the cleavage of the C-S bond of methionine.[17,18] Online spectroscopy (UV-Vis and Raman) can be used for monitoring the radiation damage[19, 20] at the time of data collection.[20] Radiation damage affects a number of generally observable parameters; increase in the unit cell size, decreasing diffraction intensities especially at high resolution, increase in Wilson *B* values, increase in R_{meas}, and increase in the mosaicity.

2.5.2 Cryocrystallography:

To reduce the effect of the radiation damage, cryo-techniques were developed. That is, data collection is performed at low temperatures, most often at 100K. The advantage of collecting data at cryotemperature is to significantly reduce the mobility of most radicals formed as a result of the photoelectric effect. This extends crystal lifetime in the beam, and hence enables a complete dataset to be collected from a single crystal.

A gradual decrease in the temperature to the cryo-level causes crystalline ice formation within or around the protein crystal. Crystalline ice affects the diffraction in two ways; first, it disrupts internal order due to expansion, and second, it interferes with the diffraction pattern from the protein. Hence, the crystal must be cooled so fast that the water molecules in the solvent channels are in glassy or vitreous rather than the crystalline state at the end of the process.[15] Here, crystallographers faced another problem - the water vitrification time. Micrometer sized drops of pure water can vitrify with cooling on a time scale of the order of 10^{-5} s, [21] which is an impossible requirement for solvent in crystal channels. The most common method for overcoming this problem is to equilibrate the crystal to replace some waters with a solution containing a cryoprotective agent. Using the correct cryoprotectant, which is determined experimentally, extends the vitrification time to 1-2 s, and this can be achieved by flash-cooling the crystal to 100 K.

There are two types of cryoprotectant; those that penetrate into solvent channels such as glycerol, and those that coat the crystal surface and do not penetrate into the solvent channels, such as oils which are especially useful for crystals with narrow solvent channels. The most commonly used cryoprotectant is glycerol in the ratio of 20-30% (v/v). Cryoprotectant should not degrade the crystal, therefore, the components of the mother liquor should be considered. In order to introduce the cryoprotectant needed for flash cooling, several methods can be followed. These are serial transfer into increasing strength of cryoprotectant, brief transfer before flash cooling, dialysis and growth in cryoprotectant. Soaking time can be from 1 s to days. The best proof that a suitable cryoprotectant has been found for a particular crystal is the diffraction quality that is obtained.

2.5.3 Mounting the crystal:

After fishing the crystal from its mother liquor using a suitable sized fiber loop, and soaking in the correct cryoprotectant, the crystal is mounted onto the goniometer and flash cooled to 100 °K by a liquid nitrogen gas stream. The process must be rapid in order to prevent crystalline ice formation, and to ensure that water in the solvent channels cools as a vitreous liquid.

2.5.4 X-ray diffraction data collection:

The x-ray data quality plays the most important role in making all subsequent steps easier, and leads to more precise atomic models. Therefore, prior to the full data collection from a crystal, the diffraction quality should be checked. The check is normally achieved by collecting two diffraction images at 90° degree (φ) apart. The advantages of performing this check are first, to check problems like, splitting of spots due to the presence of more than one lattice, disorders, the diffraction intensity and resolution limits, or even there may be no diffraction at all. Twinning may not be noticeable until the end of image processing or even until there are difficulties during refinement of the structure. Split crystals give two superimposed diffraction patterns. A high mosaicity or statistical disorders can be noticed in disordered crystals, in which spots appear to be sharp in parts of the image and smeared in others in a regular pattern. If the crystal did not diffract at all, the crystal-to-detector distance should be reassessed, or it may worthwhile to do the check at ambient temperature. A further advantage of the test is to determine the unit cell parameters and space group and to plan a strategy for collecting a complete data set. The strategy option in HKL2000 software[22] has been used in this work to optimise the total number of images and the oscillation degree of rotation between successive images to ensure a complete data set is recorded. In this work, crystals were cryocooled and data were collected at three different synchrotron light sources; beamline 10.1 at Synchrotron Radiation Source at Daresbury (SRS) in the United Kingdom, beamline XO6SA at Swiss Light Source (SLS) in Switzerland, and at beamline 08ID-1 at Canadian Light Source in Canada. Detectors used were; Mar 225 CCD at SRS and Canadian light source, and Pilatus at SLS. The data collection procedure will be discussed in more detail in Chapter 6.

2.6 Data processing:

The raw data from a set of diffraction images need to be processed or reduced to give data that can be used to calculate the structure factors. At the end of the data reduction a list of indices (*hkl*) with their associated amplitudes (the square root of the intensities) and estimate of their errors will be obtained. Many data processing packages are available, such as HKL2000, Mosflm,[23] d*TREK,[24] and an automated program XDS.[25]

The Mosflm program has been used to process the diffraction data in this work and it will be described here. The procedure can be divided into three steps[26], autoindexing, refining and integrating images.

First step - autoindexing: in this stage the unit cell parameters, the orientation of crystal and the mosaicity are estimated. The program suggests a list of possible space groups based on a penalty function. Subsequent steps will show whether the chosen space group was correct or not. The success of autoindexing depends on experimental parameters such as the x-ray wavelength, crystal-to-detector distance and most importantly an accurate direct beam position. For autoindexing Mosflm uses two images from the data set, separated by 90°, as mentioned previously when discussing data collection strategy with HKL2000.

Second step – refining: In this stage some parameters will be refined, such as the direct beam position, crystal to detector distance, unit cell parameters, small rotations of the detector about a vertical and horizontal axis, the relative scale factor applied to the detector Y coordinate and also the crystal mosaicity.

Third step - integration: the objective of this step is to measure the total intensity of each reflection on each image. This can be achieved by predicting the position in the digitised image of each Bragg reflection present on that image and then estimating its intensity after subtraction of the background and an error estimate of the intensity. The position of each reflection on the detector will be predicted by using the refined unit cell parameters and the crystal orientation. Then the program will compare the predicted to the observed position. If the fit is not good it will refine again. Usually, the detector parameters, crystal orientation and the mosaic spread will be refined during the integration, while the unit cell parameters will be fixed.

2.7 Scaling and merging:

The experimental differences in intensities during the data collection make the scaling process necessary in order to put the data on a consistent scale. These differences are due to factors related to the incident and the diffracted beam and also factor related to the detector. The incident beam-intensity varies over time in both home and synchrotron x-ray sources, but this variation is more serious in home sources. When a crystal rotates, more or less of the crystal may be in the beam resulting in the variation in the diffracted beam. If the x-ray beam hits a long thin crystal along the long axis this will give much stronger diffraction pattern than shooting it width-ways, where the crystal has fewer unit cells in the beam. Another factor which needs to be scaled is the radiation damage, because it causes the structural change over the time and hence the changes the diffracted intensities. Different regions of the detector may have different sensitivity of response and also may have defective regions.

These factors are accounted for in the scaling part of the data processing, where symmetry related copies of a given reflection on different images and the intensities of partially recorded reflections are compared, and scale factors optimised to make them the same. Scaled intensities are then merged together and reduced to a set of unique intensity values for each reflection.

The SCALA program,[27] part of the CCP4 program suit[28] (Collaboration Computational Project, Number 4) was used in this thesis for scaling and merging of reflections. SCALA implements various scaling models:[29]

$$g_{hl} = (\text{Scale}\ \boldsymbol{factor}) \times (\boldsymbol{Temperature}\ \text{factor}) \times (\text{Absorption}\ \boldsymbol{corrections}) \times (\text{Tails}\ \boldsymbol{correction}) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (2.5)$$

g_{hl} is the inverse scale factor for an observation I_{hl}; that is, lth observation of reflection h.

To accomplish data scaling and merging, SCALA does three passes through the data. First, scales are estimated initially, and then the scale parameters are refined. Second, discrepancies are analysed and the estimated standard deviations are adjusted. In the last pass the scales are applied, agreements are analysed and the output file is written.

The TRUNCATE program[30] part of the CCP4 program suit was used to estimate the structure factors from the scaled intensities. The overall temperature factor was estimated by the Wilson method.[31]

2.8 Obtaining phases of the diffraction data:

As described in section 2.1, the goal of protein crystallography is to solve the three dimensional structure of the protein, by calculating the electron density map of the molecule in the unit cell of the crystal lattice. The electron density as described mathematically in the same section is derived from the measurable amplitudes of the structure factors (measurable because it can be computed from the reflection intensity) and the initially unknown phase. The diffraction intensities, which are relatively easy to measure, contain far less information than do the phases,[32] which are more difficult to obtain. Obtaining the phase angles of each reflection is the major problem in protein crystallography, the so called phase problem.

2.9 Phasing methods:

There are some common methods used for overcoming this problem. These include the heavy-atom method, also called (multiple) isomorphous replacement (MIR), exploiting the anomalous scattering (anomalous dispersion) of x-rays, and molecular replacement. The initial solution of the phase problem given by these methods is only an estimate of phases. These initial phases may be improved by density modification methods until they good enough to start model building. Then the process of refining the model begins.

The molecular replacement method has been used to solve the phase problem in this thesis; therefore it will be discussed in more detail. However, the other two methods will be explained briefly.

In isomorphous replacement, crystals of the protein are soaked in solutions of heavy ions, producing heavy atom derivative crystals. Data are collected from both the native and the derivative crystal. The derivative data will show differences in their diffraction patterns compared to native due to the presence of the heavy atoms. By mathematical comparison of the Patterson maps of native and its derivatives, it is possible to determine the location of the heavy atoms in the unit cell, allowing by the mean of the Harker construction, an approximate determination of the phase of each reflection. Most commonly used heavy ions are Hg, Pt, or Au. Such ions bind to one or a few specific sites on the protein. The heavy atom must not disturb crystal packing or the conformation of the protein; i.e derivative crystals must be isomorphous with the native crystals. Also, there must be measurable changes in at least a modest number of reflection intensities. The resolution of derivative data needs not to be as high as that of native data.

Multi- and single-wavelength anomalous dispersion (MAD/SAD) are widely used methods for solving the phase problem. In this method, instead of collecting data from multiple derivative crystals to obtain an estimate of phases as in MIR, data are collected at different x-ray wavelengths using the native crystal only. The wavelengths are selected close to an absorption edge of e.g. intrinsic metal atoms or Se substituted for S atoms in Cys residues. By this means the anomalous diffraction that occurs above the metal or Se absorption edge can be optimised and exploited to solve the phase problem. In some cases the weak anomalous signal from intrinsic S atoms can be used for this purpose.

2.9.1 Molecular replacement:

Molecular replacement (MR)[33] is a method to obtain initial phases of an unknown structure (target model) using a homologous protein with known structure (search model). The principle of MR is that the phases for the reflections of the target model are borrowed from known phases of the search model, and an initial map is then calculated from these borrowed phases using the experimentally observed amplitudes from the target. This can be achieved by performing three rotational transformation (α, β, γ) and three translational transformations (t_x, t_y, t_z) searches, which superimpose the search model in the target unit cell. The search is usually performed in two separate steps. In the rotational step, the spatial orientation of the search and target models with respect to each other is determined, while in the translational step both molecules are superimposed in the unit cell. Performing the search in two steps, rather than trying to search for the orientation and position at the same time, dramatically reduced the time taken to find the solution[34]. The basic principle of the molecular replacement method can be understood by regarding the Patterson function of a protein crystal structure.

The most powerful tool in determining the coordinates is a Fourier sum called Patterson function $P(u,v,w)$. The coordinates (u,v,w) locate a point in Patterson map, in the same way coordinates (x,y,z) locate a point in an electron density map.

$$P(u,v,w) = \int_v \rho\,(x,y,z)\,\rho(x+u,y+v,z+w)dv \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad 2.6$$

This can be expressed as:

$$P(u,v,w) = \frac{1}{V}\sum_h \quad \sum_k \quad \sum_l |F_{hkl}|^2 cos2\pi\,(hu+kv+lw) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots 2.7$$

The Patterson map is a vector map between atoms. Vectors in the real structure shows up as vectors from the origin to the maxima in Patterson map. If the pairs of atoms are belong to the same molecule, then the corresponding vectors are relatively short and their end points are found not too far from the origin in the Patterson map; in this case they called self-Patterson vectors. If there were no intermolecular vectors cross-Patterson vectors, this inner region of the Patterson map would be equal for the same molecule in different crystal structures, apart from a rotation difference. For homologous molecules, it is not exactly equal but very similar. Therefore, the self-Patterson vector can supply us with the rotational relationship between the known and unknown molecule structure. From the cross-Patterson vectors, the translation required for moving the molecule to their correct position in the unit cell can be derived.

2.9.1.1 The Rotation function:

The comparison of orientations between target and the search model can be done by various functions in Patterson space, such as a product rotation function or correlation coefficient. The Patterson product rotation function (RF) can be expressed as:

$$RF(\boldsymbol{R}) = \int_{u,v,w} P_{target}\,(u,v,w)\,P_{search}\,[(u,v,w)\ \times \boldsymbol{R})]du\,dv\,dw \ldots\ldots\ldots\ldots\ldots\ 2.8$$

$P_{target}(u,v,w)$ is the product of the target model crystal Patterson at points (u,v,w)

$P_{search}(u,v,w)$ is the rotated search model Patterson at points (u,v,w)

$\boldsymbol{R}$ is the rotation angle.

The Patterson product rotation function evaluates the correlation between Patterson maps for the target model and for the search model in various orientations. For this orientation search, which is often called rotation search, the computer is looking for large values of the search model Patterson function $P_{search}(u,v,w)$ at locations corresponding to peaks in Patterson map of the target protein $P_{target}(u,v,w)$. The rotation function is zero where one of Pattersons has a peak and the other does not. In contrast the rotation function value is large when two Pattersons have coinciding peaks. So the integral product will be very large (maximum) if there are many coincident peaks in the two maps. These maxima should tell us the best orientation for placing the search model in the unit cell of the desired protein, the target model. Near the maxima, the rotation search can be repeated as smaller angular intervals to refine the orientation.

2.9.1.2 Translation function:

After the search model has been oriented in the correct way with the rotation function, the translation is required to overlap one molecule or subunit onto the other in real space with the translation function. In this search, only the cross-Patterson vectors are necessary. Most translation functions involve a comparison between the observed structure factor amplitudes and those calculated based on the search model. These methods have been discussed by Tong L.[35] Since the automated program for molecular replacement, MolRep[36] a part of the CCP4 program suit,[28] has been used in this thesis, the Patterson-correlation translation function used in MolRep will be discussed briefly. This method is the most commonly used translation search indicator. It is based on the correlation between the observed (from the target model) and the calculated (from the search model) Patterson maps, which has been derived by Crowther and Blow (1967).[37]

$$TF(\boldsymbol{S}) = \int_{u,v,w} P_{target}\,(u,v,w)\,P_{search}\,[(u,v,w)\ \times \boldsymbol{S})]du\,dv\,dw \ldots\ldots\ldots\ldots\ldots 2.9$$

Where $(\boldsymbol{S})$ is a translation vector

$P_{target}(u,v,w)$ is the product of the target model crystal Patterson at points (u,v,w)

$P_{search}(u,v,w)$ is the rotated search model Patterson at points (u,v,w)

Once the search model is correctly oriented and positioned in the unit cell, an initial phase can be calculated (borrowed). These phases together with the observed structure factors can be used to calculate an electron density map.

More recently, new methods for rotation and translation functions have been developed that are not based on Patterson methods, but on maximum likelihood methods.[38] These take account of errors in the data and model using sophisticated statistical methods and improve the success rate of molecular replacement when data and/or the model are poor.

Since molecular replacement depends on the presence of homologous proteins and the number of macromolecular structures solved and deposited in the Protein Data Bank (PDB)[39] is about 50,000,[40] the molecular replacement method is probably the most widely used method in solving phase problem of the new macromolecules.

Recently, many automatic molecular replacement pipelines have been developed. They can solve structures automatically with a minimum user intervention, such as MrBUMP,[41] JCSG MR pipeline,[42] and BALBES.[43]

2.9.2 Model building and refinement:

Model building is an interpretation of the currently available electron density, while, refinement is the adjustment of the built model to fit better to the experimental data. A crucial point here is that a density map computed from the refined model is generally better than the map obtained from the same model before the refinement. This then allows for an even better model to be built, provided the initial model is not faulty. Thus, refinement is needed to improve the outcome of model building by generating a better electron density map and model building is needed to provide a model in the first place and to provide stereochemical restraints for the subsequent refinement to proceed smoothly.

Model refinement consists of two different stages. Bias generated by incomplete models and radius of convergence are important considerations at early stages of refinement, because the models are usually incomplete, contain significant errors in atom parameters and may carry errors from misinterpretation of poorly phased electron density maps. During this stage of the process, the primary concern is to determine how the model of the chain tracing and conformation of the residues should be described. In later stages, after the description of the model has been determined, the objective is to determine accurate estimates of the values of the parameters which best explain the observed data.

In order to avoid distorted geometries, stereochemical restraints and constraints must be taken into account and appropriately weighted through iterative cycles. The agreement between the amplitudes of two sets of the structure factors can be given by the R-Factor:

$$R = \frac{\sum_{hkl}\left||F_{obs}(hkl) - |k|F_{calc}(hkl)|\right|}{\sum_{hkl}|F_{obs}(hkl)|} \quad \ldots\ldots\ldots 2.10$$

$|F_{obs}(hkl)|$ is the absolute value of the observed structure factors.

$|F_{calc}(hkl)|$ is the absolute value of the calculated structure factors.

While k is the scale factor between observed and calculated structure factors.

If the observed and calculated intensities agree with each other, the differences in the numerator are small, and the sum of the differences is small compared to the sum of the intensities themselves, so R is small.

Despite the use of stereochemical restraints, it is possible to over-fit or misfit the diffraction data. An incorrect model can be refined to very good R values[44]. Thus, a reliable and unbiased indicator of the accuracy has been proposed,[45] the Free-R factor, which is a statistical quantity that measures the agreement between observed and computed structure factor amplitudes for a set of reflections that is omitted in the modelling and refinement processes.

$$R_T^{free} = \frac{\sum_{hkl\in T}\left||F_{obs}(hkl)| - k|F_{calc}(hkl)|\right|}{\sum_{hkl\in T}|F_{obs}(hkl)|} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots..2.11$$

R_T^{free} refers to the "test set T", a set of reflections (usually 5-10% of the observed data) which is omitted from refinement. The Free-R factor is calculated alongside the R-factor, and should remain within a few percent of each other (typically 2-5% for high to medium resolution data) of each other during the refinement process. A decrease in R factor and increase or no change in Free-R factor indicates over-fitting or mis-fitting the diffraction data.

Several programs have been developed for refinement, such as REFMAC5,[46] PHENIX,[47] SHELX,[48] TNT,[49] and CNS.[50] REFMAC5 a part of the CCP4 program suit has been used in refining all structures shown in this thesis. Refmac5 uses a maximum likelihood method, which states that the best hypothesis (*i.e* set of parameters, which includes estimates of the errors) on the evidence of the data is the one that explains what has in fact been observed with the highest probability[51]. If the model is changed to make these observations more probable, the likelihood increases, indicating that the model is in agreement with data. Estimates of errors included in the probabilities are from the observed data and the model itself. As the model gets better during the refinement process, its errors gets smaller, so the probabilities become sharper. The sharpening of probabilities increases the likelihood, and the model fits better to the observed data, followed by a decrease in the values of R-factor and hence by a decrease in R-free value. In order to add free atoms in the current likelihood-gradient map and to remove atoms that lie in low density map, the ARP/wARP[52] program has been used alongside REFMAC5. The refinement process is described in more detail in chapter 6.

2.9.3 Structure validation:

R_{free} and R_{factor} as important validation parameters are monitored during the refinement process. Root-mean-square deviation (rmsd) from stereochemical standards (ideal bond distances, bond angles, dihedral angles, and van de Waals contact distances) are also monitored during the refinement. Rmsd from standard stereochemistry indicates how much the model departs from geometrical parameters that are considered typical, or represent chemical common sense based on previous experience. Good quality medium to high resolution structures are expected to have a rmsd(bond) of about 0.02 Å, when this number becomes too high, it means that something might be wrong with the model. These parameters might raise questions about the quality of the structure but would not identify the source of possible errors. To trace them, PROCHECK[53] a part of CCP4 program suit and MOLPROBITY[54] have been used to look for these features.

The Ramachandran plot[55] provides good indicators of the accuracy of the protein model. The values of the main chain dihedral angles in a protein can only take on a limited range of values. As glycine residues lack any side chains, these can adopt a wider range of φ and Ψ angles. The φ-Ψ values of other residues should fall within the allowed values. If there were a considerable number of outliers, then the structure is probably in need of improvement.

MolProbity[54] runs on a java-enabled browser from the well-documented homepage. The notable feature is web-based identification and viewing of potential problems in the model. Hydrogen atoms are added to assist in the contact analysis. Poor φ-Ψ angles are flagged. Poor rotatamers are flagged but cannot be fixed interactively on line. Potential Asn, Gln, and His flips are visualized. Bad contacts are visualized with colour-coded dots, plus a prioritised list.

2.10 References:

---

[1] Bragg W.L., (1913) "The Diffraction of Short Electromagnetic Waves by a Crystal", *Proceedings of the Cambridge Philosophical Society*, **17**, 43–57.

[2] McCoy A. (2008) "A protein crystallography overview lecture" *BCA/CCP4 Summer School XV*, Oxford University 7th-13th September.

[3] Rhodes G (2006) "Crystallography made crystals clear, a guide for users of macromolecular models" 3rd Ed., Elsevier Inc.

[4] Wu Z. (2008) "Lecture notes on computational structural biology" World scientific publishing Co. Pte. Ltd., Singapore.

[5] Schwede T., Peitsch M. (2008) "Computational structural biology, methods and applications" World scientific publishing Co. Pte. Ltd., Singapore.

[6] Helliwell J. R. (2006) "Synchrotron crystallography" *International Tables of Crystallography* **F**, 155-166.

[7] Brock J. D., and Sutton M (2008) "Materials science and x-ray techniques" *Materials Today* **11**(11), 52-55.

[8] Gouet, P., Diprose, J. M., Grimes, J. M., Malby R., Burroughs J. N., Zientara S., Stuart D. I., and Mertens P. P. C. (1999) "The highly ordered doublestranded RNA genome of bluetongue virus revealed by crystallography" *Cell* **97**, 481–490.

[9] Ban, N., Nissen, P., Hansen, J., Moore, P. B., and Steitz, T. A. (2000) "The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution" *Science* **289**, 905.

[10] Garman E., and Sweet R. M. (2007) "X-ray data collection from macromolecular crystals" *Methods in Molecular Biology* **364**(2) 63-93. Edited by: Sylvie Doublié, Humana Press Inc., Totowa, New Jersy, USA.

[11] Chernov A. A. (2003) "Protein crystals and their growth" *Journal of Structural Biology* **142**, 3-21.

12 Chayen N. E. (2004) "Turning protein crystalisation from an art into a science" *Current Opinion in Structural Biology* **14**, 577-583.

[13] Giegé R., and McPherson A. (2006) "Crystallization" *International Tables of Crystallography* **F**, 81-93.

[14] Drenth J (2007) "Principles of protein x-ray crystallography" 3rd Ed., Springer Science and Business Media, LLC, New York, USA.

[15] Garman E. E., and Owen R. L. (2006) "Cryocooling and radiation damage in macromolecular crystallography" *Acta Crystallographica* D**62**, 32-47.

[16] O'Neill P., Stevens D. L., and Garman E. F. (2002) "Physical and chemical considerations of damage induced in protein crystals by synchrotron radiation: a radiation chemical perspective" *Journal of Synchrotron Radiation* **9**, 329-332.

[17] Burmeister W. P. (2000) "Structural changes in a cryo-cooled protein crystal owing to radiation damage" *Acta Crystallographica* D**56**, 328-341.

---

[18] Fioravanti E., Vellieux F. M. D., Amara P., Madern D., and Weik M. (2007) "Specific radiation damage to acidic residues and its relation to their chemical and structural environment" *Journal of Synchrotron Radiation* **14**, 84-91.

[19] Soothworth-Davies R. J., and Garman F. E. (2007) "Radioprotectant screening for cryocrystallography" *Journal of Synchrotron Radiation* **14**, 73-83.

[20] Beitlich T., Kühmel K., Schulze-Briese C., Shoeman R. L., and Schilchting I. (2007) "Cryoradiolytic reduction of crystalline heme proteins: analysis by UV-Vis spectroscopy and x-ray crystallography" *Journal of Synchrotron Radiation* **14**, 11-23.

[21] Johari G. P., Hallbrucker A., and Mayer E. (1987) "The glass-liquid transition of hyperquenched water" *Nature* **330**, 552-553.

[22] Otwinowski Z., and Minor W. (1997) "Processing of X-ray Diffraction Data Collected in Oscillation Mode" Eds. Carter C.W., Jr. and Sweet R. M., Academic press, New York. *Methods in Enzymology* **276**, 307-326,

[23] Leslie A. G. W. (1992) "Recent changes to the MOSFLM package for processing film and image plate data" *joint CCP4/ESF-EACMB Newsletter on protein crystallography*, No. **26**

[24] Pflugrath J. W. (1999) "The finer things in x-ray diffraction data collection" *Acta Crystallographica* D**55**, 1718-1725.

[25] Kabsch W. J. (1993) "Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants" *Journal of Applied Crystallography* **26**, 795-800.

[26] Leslie A. G. W. (2006) "The integration of macromolecular diffraction data" *Acta Crystallographica* D**62**, 48-57.

[27] Evans P. (1993) "Data reduction: data collection and processing" *Proceedings of the CCP4 study weekend.* Data collection and processing, 114–123.

[28] CCP4 (1994) "The CCP4 Suite: programs for protein crystallography" *Acta Crystallographica*, D**50**, 760-763.

[29] Evans P. (2006) "Scaling and assessment and data quality" *Acta Crystallographica* D**62**, 72-82.

[30] French G. S., and Wilson K. S. (1978) "On the treatment of negative intensity observations" *Acta Crystallographica* A**34**, 517.

[31] Wilson A. J. C. (1949) "The probability distribution of x-ray intensities" *Acta Crystallographica* **2**, 318-321.

[32] Cowtan K. "Book of Fourier" http://www.ysbl.york.ac.uk/~cowtan/fourier/fourier.html

[33] Rossmann M. G. (2001) "Molecular replacement – historical background" *Acta Crystallographica* D**57**, 1360-1366.

[34] Evans P., McCoy A. (2008) "An introduction to molecular replacement" *Acta Crystallographica* D**64**, 1-10.

[35] Tong L. (2006) "Translation functions" *International Tables for Crystallography* **F**, 275-278.

---

[36] Vagin A., and Teplyakov A. (1997) "Molrep: and automated program for molecular replacement" *Journal of Applied Crystallography* **30**, 1022-1025.

[37] Crowther R. A., and Blow D. M. (1967) "A method of positioning a known molecule in an unknown crystal structure" *Acta Crystallographica* B**25**, 2571-2580.

[38] Read R. J. (2001) "Pushing the boundaries of molecular replacement with maximum likelihood" *Acta Crystallographica* D**57**, 1373-1382.

[39] Berman H. M., et al. (2000) "The protein data bank" *Nucleic Acid Research* **28**, 235-242.

[40] Hasnain S. S., and Wakatsuki S. (2008) "Biophysical methods: structure, function and dynamics studies of macromolecular assemblies using electrons, laser, neutrons and x-rays" *Current Opinion in Structural Biology* **18**, 577-580.

[41] Keegan R. M., and Winn M. D. (2007) "Automated search-model discovery and preparation for structure solution by molecular replacement" *Acta Crystallographica* D**63**, 447-457.

[42] Schwarzenbacher R., Godzik A., and Jaroszewski L. (2008) "The JCSG MR pipeline: optimized alignments, multiple models and parallel searches" *Acta Crystallographica* D**64**, 133-140.

[43] Long F., Vagin A. A., Young P., and Murshudov G. N. (2008) "BLABES: a molecular-replacement pipeline" *Acta Crystallographica* D**64**, 125-132.

[44] Brändén C., and Jones T. (1990) "Between objectivity and subjectivity" *Nature* **343**, 687-689.

[45] Brünger A. T. (1992) "Free R value: a novel statistical quantity for assessing the accuracy of crystal structure" *Nature* **355**, 472-475.

[46] Murshudov G. N, Vagin A. A., and Dodson E. J. (1997) "Refinement of macromolecular structures by the maximum-likelihood method" *Acta Crystallographica* D**53**, 240-255.

[47] Adams P. D, Grosse-Kunstleve R. W., Hung L., Ioerger T. R., McCoy A. J., Moriarty N. W., and Read J. R. (2002) "PHENIX: building new software for automated crystallographic structure determination" *Acta Crystallographica* D**58**, 1948-1954.

[48] Sheldrick G. M. (2008) "A short history of SHELX" *Acta Crystallographica* A**64**, 112-122.

[49] Tronrud, D. E. (1997) "The TNT Refinement Package. in Macromolecular Crystallography" Eds. Carter C.W., Jr. and Sweet R. M., Academic press, New York. *Methods in Enzymology* **277**, 306-319.

[50] Brüger A. T., et al.(1998) "Crystallography and NMR system: A new software suit for macromolecular structure determination" *Acta Crystallographica* D**54**, 905-921.

[51] McCoy A. J. (2004) "Liking likelihood" Acta Crystallographica D**60**, 2169-2183.

[52] Lamzin V. S., and Wilson K. S. (1993) "Automated refinement of protein models" *Acta Crystallographica* D**49**, 129-147.

[53] Laskowski R. A., MacArthur M. W., Moss D. S., and Thornton J. M (1993) "PROCHECK: a program to check the stereochemical quality of protein structures" *Journal of Applied Crystallography* **26**, 283-291.

[54] Davis I. W., et al.(2007) "MolProbity: all-atom contacts and structure validation for proteins and nucleic acids" *Nucleic Acid Research* **35**, W375-W383.

---

[55] Ramachandran G. N., Ramakrishnan C., and Sasisekhran V (1963) "Stereochemistry of polypeptide chain configurations" *Journal of Molecular Biology* **7**, 955.

Chapter 3: Materials and Methods

3.1 Introduction to gene cloning:

To investigate and study any protein in any organism it has to be mass-produced, in order to do that its gene or part of its gene needs to be identified. Once the gene has been identified it can be isolated from the genomic DNA and inserted into a self-replicating genetic element called vector producing recombinant DNA. The vector can be a bacterial plasmid or a viral phage. This process is called recombinant DNA technology or gene cloning. Recombinant DNA is then introduced to an appropriate host cell to produce a desired protein. In the following sections an outline of the theory behind gene cloning is described, followed by a description of the specific cloning and expression system used for the production of recombinant cytochrome c′ from *Alcaligenes xylosoxidans*.

3.1.1 DNA structure:

DNA is a double stranded molecule of polydeoxyribonucleotide, each strand contains many monodeoxyribonucleotides which composed of one molecule of nitrogen bases: adenine (**A**), thymine (**T**), cytosine (**C**) or guanine (**G**), and a molecule of pentose sugar of 2`-deoxyribose and a phosphate group. Monomers are linked together by phosphodiester bonds between the 3`-hydroxyl of one pentose to the 5`-hydroxyl of the next.[1] Two strands are linked by hydrogen bonds between complimentary nitrogen base pairs; **A-T** base pair held together by two H- bonds whereas **C-G** base pair by a three H-bonds. Unlike most polymers, DNA chains are paired in antiparallel manner in which the 5`-end of one strand is paired with the 3`-end of the other strand. The hydrophilic deoxyribo-phosphate of each strand is on the outside of the molecule while the hydrophobic nitrogen bases located inside perpendicular to the axis of the helix. This double helical structure prevents the relaxation of torsional stress by rotation about a single covalent bond, and the stacking of the bases on top of each other gives the DNA an unusual large flectional rigidity (Figure 3.1).[2,3,4]

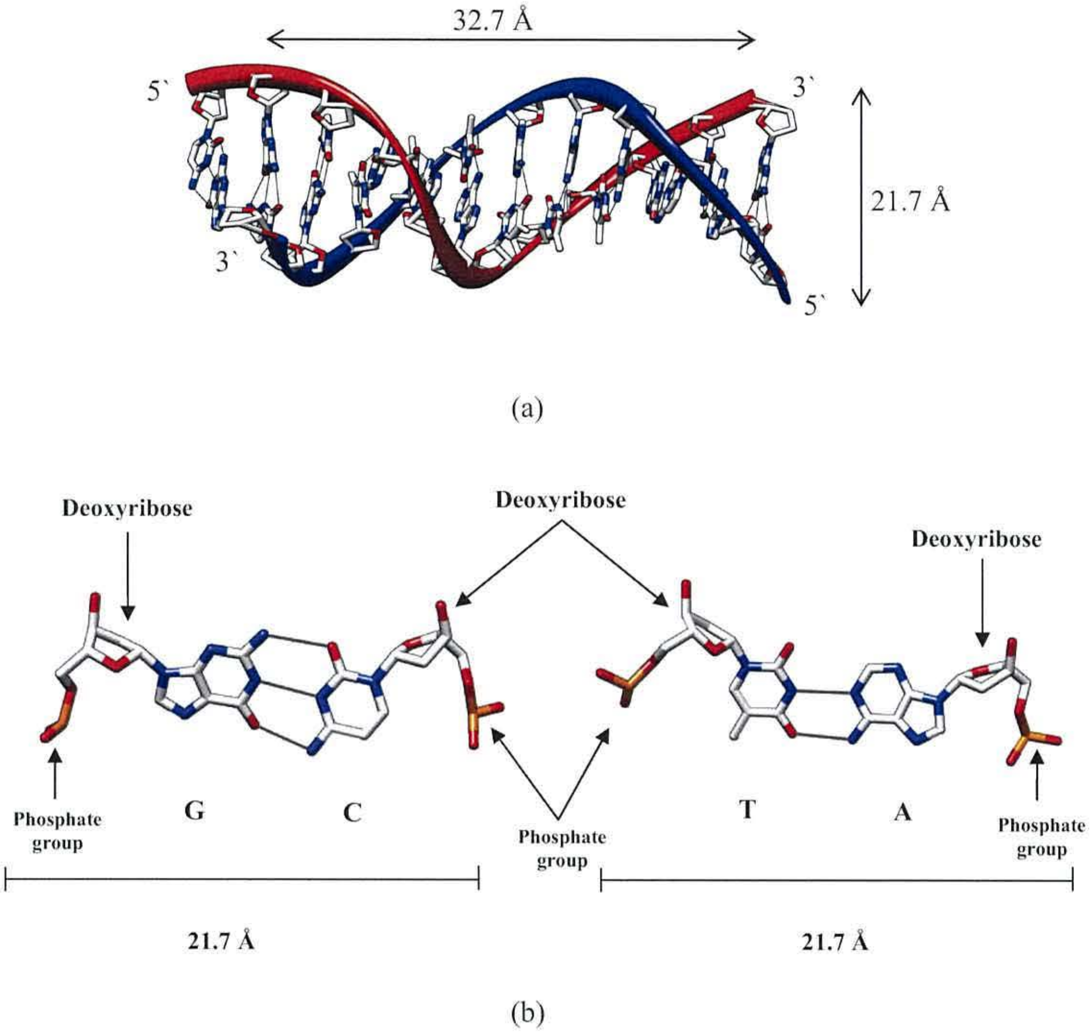


Figure 3.1: A representation of (a) DNA double helix modelled from pdb 1D65[4] above showing the 3` and 5` terminals with nitrogen bases and their corresponded hydrogen bonds (b) base pairing showing the triple hydrogen bond between nitrogen bases Guanine (G) and Cytosine (C) and a double hydrogen bond between Thymine (T) and Adenine (A). Sugar molecule of deoxyribonucleic acid and the phosphate group has been shown. The molecular graphics created by UCSF Chimera[5] all distances shown is for the same pdb file.

3.1.2 The Gene:

In prokaryotes the complete sequence of the DNA makes a circular chromosome which carries the total genetic information needed by that organism; this is referred to as the genome. Genomic DNA varies enormously in size from 5224 base pairs in Simian virus (SV40) [6] to about 2.91 billion base pairs in human.[7] A region within the genomic DNA which encodes the production of a particular protein or responsible for a particular organism's characteristic is called gene.[8] The gene encodes the synthesis of protein in two main steps; first the transcription of the gene into complementary RNA molecules (mRNA, tRNA, rRNA) each of which has a specific function. The second step, called translation, starts when the mRNA is read from the 5`-end to the 3-end in blocks of three nucleotide bases (codons). The codons are decoded by tRNA, which interacts with mRNA and translated to protein on rRNA.

The correspondence between a particular codon and the amino acid is specified by genetic codes. As there are four different nucleotides and each codon is comprises of three nucleotides; there will be sixty four different codons (i.e sixty four different tRNAs) each of which have an anticodon to recognise codons in the mRNA (Table 1).[9] Thus there are frequently different codons available to specify one of the twenty different amino acid monomers utilised to make up proteins. The *in vivo* protein synthesis is much more complicated than what we have described here, and it has been covered by many text books[1,8,10] and will not be discussed in detail here.

	U	C	A	G	
U	**UUU Phe**	**UCU Ser**	**UAU Tyr**	**UGU Cys**	**U**
	UUC Phe	**UCC Ser**	**UAC Tyr**	**UGC Cys**	**C**
	UUA Leu	**UCA Ser**	**UAA Term**	**UGA Term**	**A**
	UUG Leu	**UCG Ser**	**UAG Term**	**UGG Trp**	**G**
C	**CUU Leu**	**CCU Pro**	**CAU His**	**CGU Arg**	**U**
	CUC Leu	**CCC Pro**	**CAC His**	**CGC Arg**	**C**
	CUA Leu	**CCA Pro**	**CAA Gln**	**CGA Arg**	**A**
	CUG Leu	**CCG Pro**	**CAG Gln**	**CGG Arg**	**G**
A	**AUU Ile**	**ACU Thr**	**AAU Asn**	**AGU Ser**	**U**
	AUC Ile	**ACC Thr**	**AAC Asn**	**AGC Ser**	**C**
	AUA Ile	**ACA Thr**	**AAA Lys**	**AGA Arg**	**A**
	AUG Met	**ACG Thr**	**AAG Lys**	**AGG Arg**	**G**
G	**GUU Val**	**GCU Ala**	**GAU Asp**	**GGU Gly**	**U**
	GUC Val	**GCC Ala**	**GAC Asp**	**GGC Gly**	**C**
	GUA Val	**GCA Ala**	**GAA Glu**	**GGA Gly**	**A**
	GUG Val	**GCG Ala**	**GAG Glu**	**GGG Gly**	**G**

Table 3.1: The genetic code, the orange AUG correspond to Methionine an initiator of protein synthesis, UAA, UAG and UGA specify termination of protein synthesis which known as stop codons. (RNA codons)

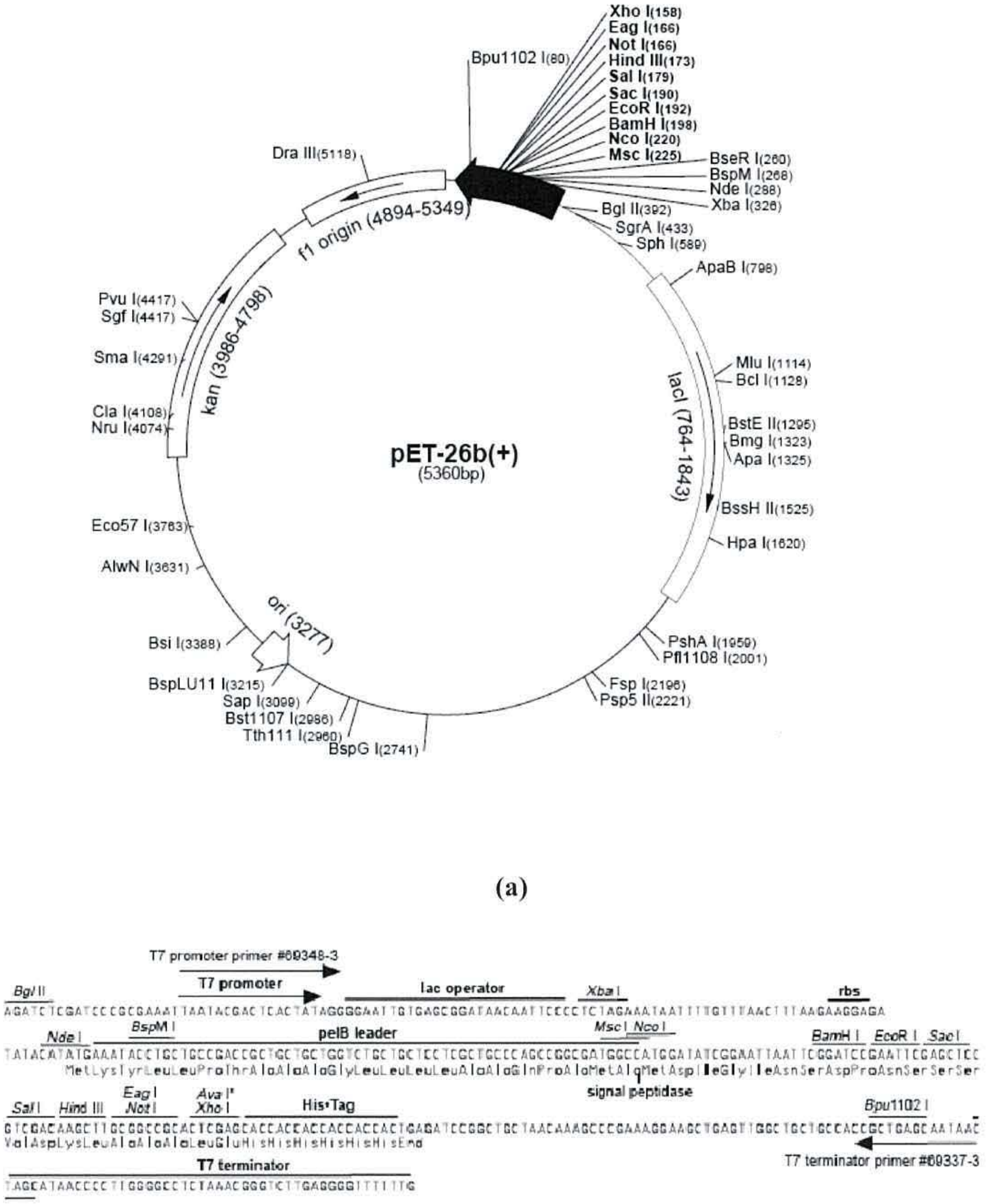
3.1.3 Cloning Vectors:

A DNA molecule that carries a foreign segment of DNA is called a cloning vector. These vectors can be plasmids (as in this work), bacteriophages, viruses, or even small artificial chromosomes.[10] Vectors all contain a sequence enabling them to replicate autonomously called the origin of replication, an antibiotic resistance gene as well as at least two regions (restriction sites) that can be cut by restriction endonucleases to permit the insertion of exogenous DNA fragment.[11] The vector and the exogenous DNA should be cut by the same restriction enzymes: this allow the exogenous DNA fragment to be ligated to the correct restriction site. Of course the number of copies of the exogenous genomic DNA fragments may not be sufficient to be cloned to a vector; therefore, polymerase chain reaction technique can be used to amplify the number of the specific gene copies.

3.2 Molecular biology methods:

3.2.1 The plasmid vector:

pET 26b(+) (Novagen) was used as a plasmid vector to clone the *Alcaligenes xylosoxidans* cytochrome c′ gene (AXCP). Reasons behind choosing this vector were; first it carries an N-terminal *pelB* signal sequence (Figure 3.2) for potential periplasmic localization, plus optional C-terminal His•Tag® sequence which gives another option of the purification. Another reason for choosing pET 26(+) was the kanamycin (Kan) resistance as ampicillin (Amp) combined with cytochrome c′ could be toxic for the host system.



(a)

(b)

Figure 3.2: The pET 26b(+)expression vector (Novagen), (a) the vector map and (b) the cloning and expression regions. The vector has *pelB* leader for the periplasmic localization and the C-terminal His-tag that provide another protein purification option.

Since the cytochrome c′ needs to be matured,[12] an additional plasmid (pEC86) which carries cytochrome c maturation gene (Ccm ABCDEFGH) was also co-transformed into *E. coli.* This plasmid was previously constructed from pACYC184 [12] (Figure 3.3) and was a kind gift from Dr. Thöny Meyer through Dr Julie Stevens from Oxford University.

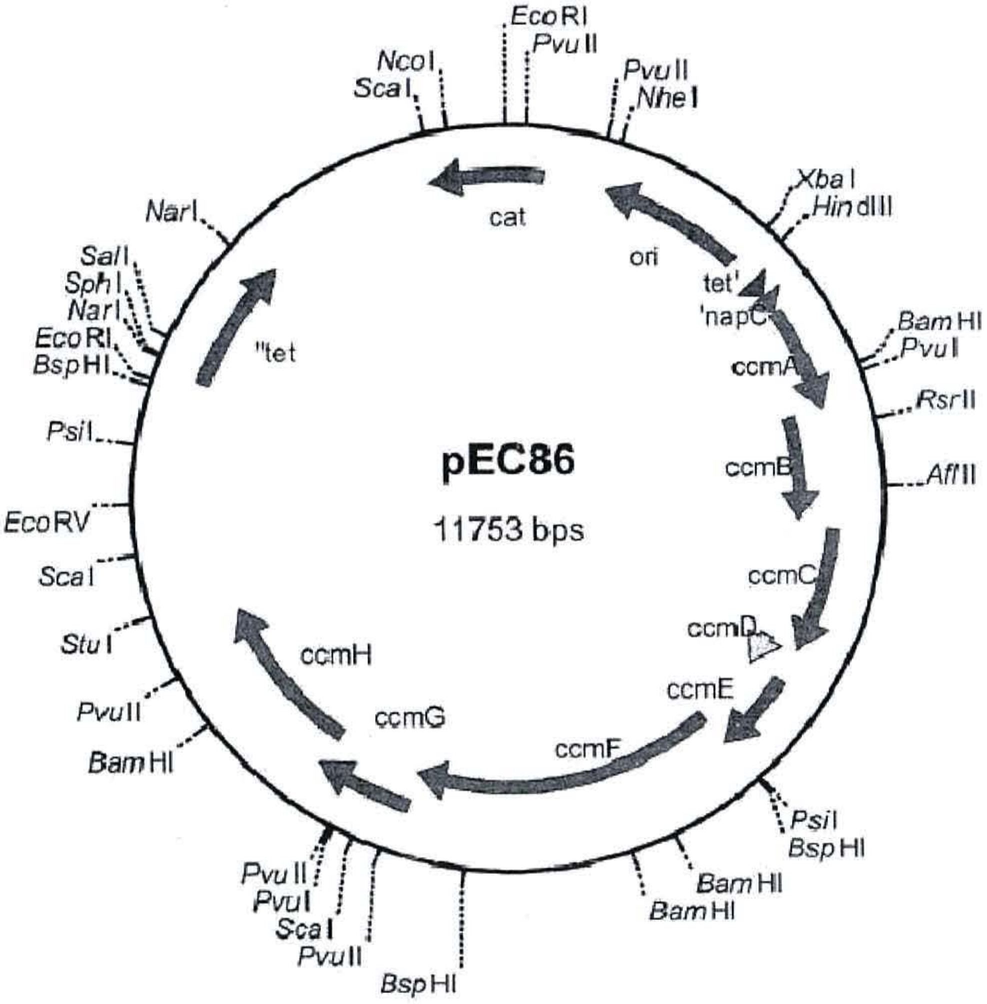


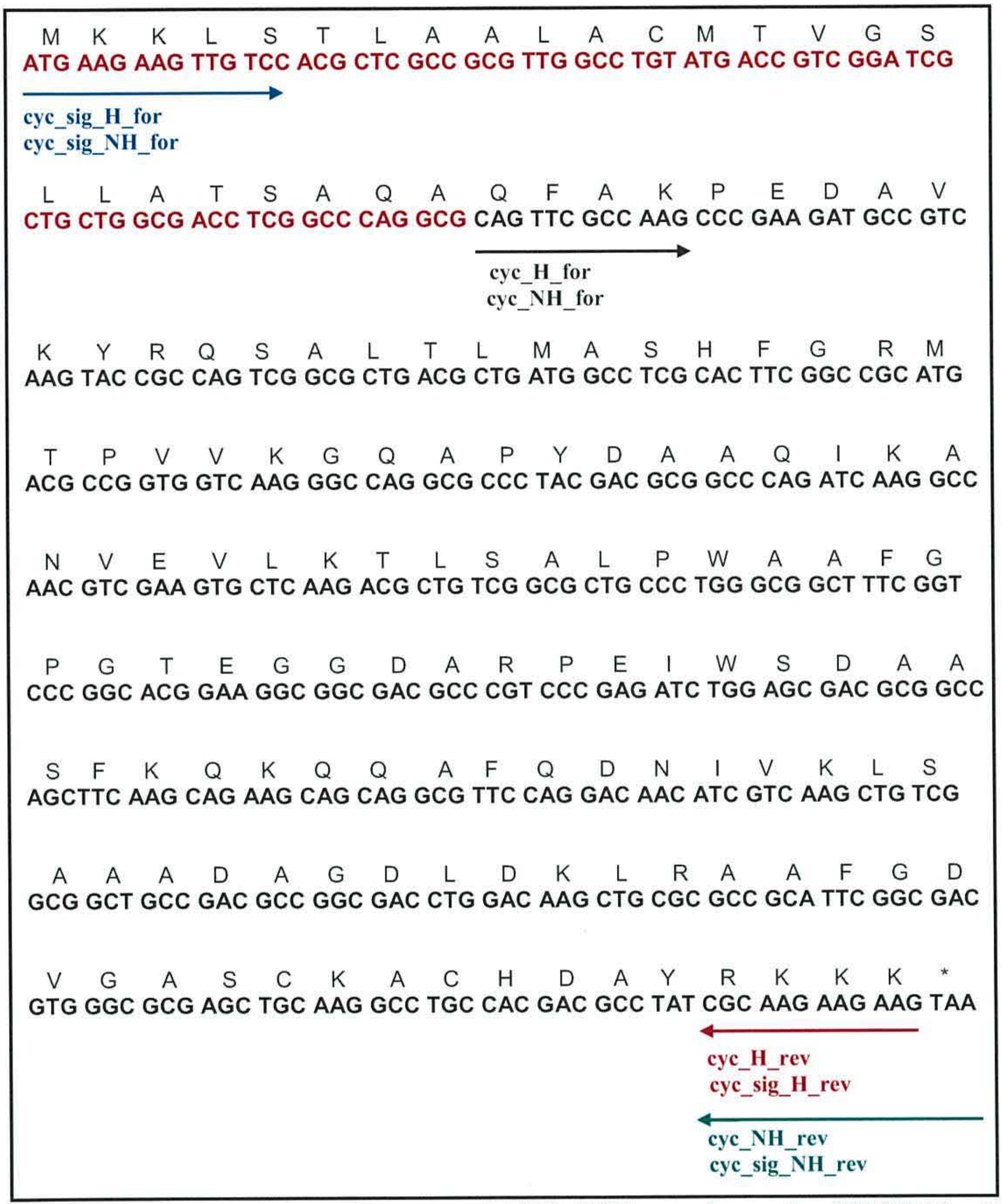
Figure 3.3: pEC86 plasmid carries cytochromes c maturation genes (Ccm ABCDEFGH) which are necessary for the cytochrome c′ maturation and mass-production.

3.2.2 Construction of the expression plasmid:

To PCR amplify the cytochrome c′ from *Alcaligenes xylosoxidans* strain NCIMB11015 (NCBI accession number EF472535.1) (Figure 3.4), several pairs of primers were designed (Table 3.2). The KOD hot start DNA polymerase kit (Novagen) was used for the PCR polymerisation. The cloning strategies were: to clone AXCP into the plasmid vector with and without AXCP signal peptide, and with and without the His-tag in each case. The first pair of primers (Table 3.2) were used to amplify AXCP without its signal peptide but with the His-tag in its C-terminal, (DNA fragment labelled- *pETC-HS*). In order to include the His-tag in the protein, the stop codon (TAA) was removed. The second pair of primers designed to generate the DNA fragment labelled *pETC-NHS*, which is AXCP without a signal peptide and without the His-tag, in this case its stop codon was included. The restriction sites have been engineered as appropriate to facilitate the cloning process (as shown in different colours in table 3.2). The amplified fragments were cloned into the pET 26b(+)using the NcoI, HindIII, and BamHI restriction enzymes as appropriate (Table 3.2). In both plasmids the protein was directed to the periplasm by the *pelB* signal sequence. Since no successful results were obtained from theses two strategies, we chose to carry into the next strategy which was to include its signal peptide with and without the His-tag, so the third and fourth pairs were designed to amplify AXCP with its signalling peptide either with or without the His-tag, generating the DNA fragments *pETC-SIG-HS* and *pETC-SIG-NHS* respectively.

Primer	**Nucleotide sequence (5`→ 3`)**	**Restriction enzyme**	**T_m °C**	**Note**
Cyc_H_For	GCC **CCA TGG** ATG CAG TTC GCC AAG C	NcoI	64	Designed to clone AXCP gene without its signal peptide but with the His-tag.
Cyc-H-rev	ACC GGC **AAG CTT** CTT CTT GCG ATA GGC	HindIII	63	
Cyc-NH-for	GCC **CCA TGG** ATG CAG TTC GCC AAG C	NcoI	64	Designed to clone AXCP gene without its signal peptide and without the His-tag.
Cyc-NH-rev	CGC GGT ACC GGC **GGA TCC** TTA CTT CTT CTT	BamHI	66	
Cyc-sig-H-for	AGA TAT ACA CCC **CAT ATG** AAG AAG TTG TCC ACG CTC	NdeI	64	Designed to clone AXCP with its signal peptide and the His-tag.
Cyc-sig-H-rev	ACC GGC **AAG CTT** CTT CTT GCG ATA GGC	HindIII	63	
Cyc-sig-NH-for	AGA TAT ACA CCC **CAT ATG** AAG AAG TTG TCC ACG CTC	NdeI	64	Designed to clone AXCP with its signal peptide but without His-tag.
Cyc-sig-NH-rev	CGC GGT ACC GGC **GGA TCC** TTA CTT CTT CTT	BamHI	66	

Table 3.2 : Oligonucleotide primers used for the amplification and cloning the AXCP gene into the expression vector. Tm is the melting temperature of the primers. Restriction sites have been engineered to facilitate the cloning process, Red: NcoI, Orange: HindIII, Blue: BamHI, Grey: NdeI



M K K L S T L A A L A C M T V G S
ATG AAG AAG TTG TCC ACG CTC GCC GCG TTG GCC TGT ATG ACC GTC GGA TCG

cyc_sig_H_for
cyc_sig_NH_for

L L A T S A Q A Q F A K P E D A V
CTG CTG GCG ACC TCG GCC CAG GCG CAG TTC GCC AAG CCC GAA GAT GCC GTC

cyc_H_for
cyc_NH_for

K Y R Q S A L T L M A S H F G R M
AAG TAC CGC CAG TCG GCG CTG ACG CTG ATG GCC TCG CAC TTC GGC CGC ATG

T P V V K G Q A P Y D A A Q I K A
ACG CCG GTG GTC AAG GGC CAG GCG CCC TAC GAC GCG GCC CAG ATC AAG GCC

N V E V L K T L S A L P W A A F G
AAC GTC GAA GTG CTC AAG ACG CTG TCG GCG CTG CCC TGG GCG GCT TTC GGT

P G T E G G D A R P E I W S D A A
CCC GGC ACG GAA GGC GGC GAC GCC CGT CCC GAG ATC TGG AGC GAC GCG GCC

S F K Q K Q Q A F Q D N I V K L S
AGCTTC AAG CAG AAG CAG CAG GCG TTC CAG GAC AAC ATC GTC AAG CTG TCG

A A A D A G D L D K L R A A F G D
GCG GCT GCC GAC GCC GGC GAC CTG GAC AAG CTG CGC GCC GCA TTC GGC GAC

V G A S C K A C H D A Y R K K K *
GTG GGC GCG AGC TGC AAG GCC TGC CAC GAC GCC TAT CGC AAG AAG AAG TAA

cyc_H_rev
cyc_sig_H_rev

cyc_NH_rev
cyc_sig_NH_rev

Figure 3.4: DNA and amino acid sequence of AXCP (NCBI accession number EF472535.1) with primers annealing positions. The blue arrow represents cyc_sig_H_forward and cyc_sig_NH_forward, black arrow represents cyc_H_forward and cyc_NH_forward primers, red arrow represents cyc_H_reverse and cyc_sig_H_reverse while the green arrow represents cyc_NH_reverse and cyc_sig_NH_reverse primer annealing position.

3.2.3 Polymerase chain reaction:

The KOD hot start DNA polymerase enzyme (Novagen) was chosen to PCR amplify the cytochrome c′ gene from *Alcaligenes xylosoxidans* genomic DNA. The KOD hot start combines high fidelity and fast extension speed and effectively inhibits primer degradation during setup at ambient temperature due to exonuclease activity. This enzyme quickly and accurately amplifies genomic and phage/plasmid DNA targets, GC-rich targets are also efficiently amplified.[13] The PCR reaction mixture contains 1.5 mM $MgCl_2$, 0.2 mM each of dATP, dCTP, dGTP, dTTP (dNTPs) with 0.02U/μL KOD hot start DNA polymerase enzyme. 0.3μM of the forward and the reverse primers were used with 10 ng DNA template.

In general PCR compose of five temperature change steps; the first step is called the initialisation step, in this step the reaction mixture will be heated to 95 °C for 2 minutes to activate the thermostable polymerase enzyme. Next is the denaturation step, which is the first regular cycling step, in which the reaction mixture was heated to 95 °C for 20 seconds. This step causes the separation of the double stranded DNA template and primers by heat-disruption of H-bonds between complementary bases. The third step is annealing the primers to the single stranded DNA by lowering the reaction temperature to the T_m of the primers used in the reaction (Table 3.2) for 10 seconds. The last step of the PCR cycling events is the extension step; in this step the reaction temperature changed to 70 °C which is the optimum temperature of the polymerase enzyme used for 90 seconds. At this step new DNA strands complimentary to the DNA template strand synthesised by adding complimentary dNTPs to the template in 5' to 3' direction. To ensure that any remaining single stranded DNA is fully extended, a single step was performed at 70 °C for 10 minutes, this step is called final elongation step.

The PCR machine (Techgene) was programmed as the following:

Step 1- Initialisation step: 95 °C for 2 minutes to activate the polymerase enzyme.

Step 2- Denaturation step: 95 °C for 20 seconds to separate the DNA strands.

Step 3- Annealing step: Lowest primer T_m °C for 10 seconds (see Table 3.2).

Step 4- Extension step: 70 °C for 90 second to add-up complimentary dNTPs to the template.

Step 5- Final Elongation step: 70 °C for 10 minutes to enable any uncompleted double strand to be completed.

The thermal cycler machine was then programmed to repeat step 2-4 for 25 times.

3.2.4 PCR product purification:

PCR product has been purified from primers, nucleotides, polymerase, salts and other enzymatic reactions by using QIAquick PCR purification kit (Qiagen, UK). The PCR product was diluted five-fold with PB buffer, then centrifuged in QIAquick column at 13000 rpm for 1 min. for 1 minute at room temperature. Apart from single and double stranded PCR products all other components will be eluted through the column, and then the column was further washed with 0.75 ml buffer PE to remove any remaining impurities leaving only the double stranded DNA fragments. Then the double stranded DNA was eluted with 50μL of nuclease-free water.

3.2.5 Agarose gel electrophoresis:

To visualize and assess the PCR product, agarose gel electrophoresis was used. 1% (w/v) agarose was prepared by melting 1g of agarose in 100mL of 1x Tris-borate-ethylenediaminetetra-acetic acid (TBE) buffer pH 8.2 in the microwave oven for 5 minutes. The melted agarose was stirred and 5μL of ethidium bromide stock solution (10mg/mL) was added to the gel. Ethidium bromide can insert into the DNA strand and forms fluorescent complexes and these can be viewed under UV light. The solution mixture then poured into a gel rack, a comb inserted at one side of the gel and allowed to cool down at room temperature. The gel, together with rack was placed into a tank and enough 1x TBE buffer was added to the tank to cover the gel. To 10 μL PCR product, 2 μL of loading buffer (0.25% (w/v) bromophenol blue in 40% (w/v) sucrose dissolved in nuclease-free water) was added and then loaded into the well. Bromophenol blue enables tracking the progress of the electrophoresis while the high density of sucrose prevents the sample being lost from the sample well. A 1 kbp DNA ladder (Promega) was routinely run in another lane in the gel to calibrate the size of the base pairs in the PCR products (Figure 3.5). A current of (80 mA) was applied for 45 minutes. Agarose gels were visualized under UV light in a dark room. Once band(s) on the correct molecular weight were detected on the gel, the resultant solutions from the PCR were purified (section 3.2.4) or the fragments extracted from the agarose gel and purified (section 3.2.6).

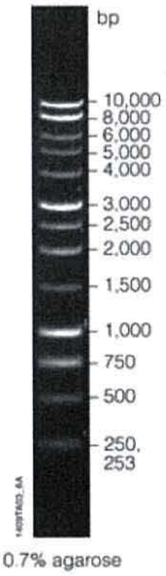


Figure 3.5: The DNA ladder from (Promega) with fragment sizes ranging from 250 to 10,000 base pairs used to indicate the corresponding sizes of the DNA fragments from the PCR product.

3.2.6 Gel extraction and purification of DNA fragments:

In some cases the desired gene was extracted from the gel using QIAquick gel extraction kit (Qiagen, UK). After the identification of the fragment on the agarose gel under the UV light, the corresponding band was isolated, and the DNA was extracted from the gel and purified from salts and stain by following the manufacturer's protocol.[14]

3.2.7 Nucleic acid digestion:

To yield complementary ends, both purified PCR product and the plasmid vector were double digested with the proper restriction enzymes separately. NcoI and HindIII restriction enzymes were used to generate *pETC-HS*, while to generate *pETC-NHS* NcoI and BamHI (both from Promega) were used, for *pETC-SIG-HS* both NdeI and HindIII were used and finally NdeI and BamHI were used to construct *pETC-SIG-NHS*. Each recombinant DNA construction reaction was performed in the final volume of 10 μL. To double digest the plasmid; 2 μL of pET 26b(+), 1 μL of each enzyme, 1 μL of 10x appropriate supplied buffer and 5 μL nuclease-free water were mixed and incubated at 37 °C for 1 hour. The purified PCR product was vacuum dried and resuspended in 7 μL nuclease-free water and 1μL of each enzyme and 1 μL of 10x appropriate supplied buffer mixed and incubated at 37 °C for 1 hour. The restriction digested plasmid and the PCR product were mixed and purified together using QIAquick PCR purification kit (Qiagen, UK) (see section 3.2.4).

3.2.8 DNA ligation reaction:

In order to join complementary ends of the gene and the plasmid to produce the recombinant DNA, purified product from the above (section 3.2.7) was dried under vacuum for 45 minutes, and resuspended in 8μL of nuclease-free water. 1μL of 10x supplied buffer (300 mM Tric-HCl pH 7.8, 100 mM $MgCl_2$, 100 mM DTT and 10 mM

ATP) and 1μL T4 ligase (Promega) were added to the resuspended DNA and incubated at 15 °C overnight. T4 ligase catalyses the ligation between the 5`-phosphate and 3`-hydroxyl group of adjacent nucleotides to form phosphodiester bond in the presence of Mg^{2+} and adenosine triphosphate (ATP) as the energy yielding cofactor.[15]

3.2.9 Checking the cloned gene size and sequencing:

In order to check whether the *AXCP* gene was inserted into the pET26 (+) plasmid and the recombinant plasmid DNA has been successfully constructed, 7 μL of the eluted solution from section (3.3.5) was double digested with 1 μL of each appropriate restriction enzyme and 1 μL of suitable supplied buffer at 37 °C for 1 hour and then 2 μL of loading buffer was added to the digested mixture and run on 1% (w/v) agarose gel electrophoresis alongside with the DNA molecular weight marker (section 3.2.5). This allows the size (length) of the cloned gene and the original plasmid to be determined. Once the correct length was observed on the gel, 15 μL of the same sample from (section 3.3.5) was sent to (Eurofins MWG Operon, Germany) for sequencing.

3.2.10 Site-directed mutagenesis:

Site directed mutagenesis is a powerful tool for the study of structure-function relationship, in which a mutation can be created at a defined site of plasmid DNA. Site directed mutagenesis is performed by, firstly designing two synthetic oligonucleotide primers where a change at least one codon from that observed in the native DNA is made. Thus the DNA polymerase extends the primers on the template plasmid during temperature cycling (PCR). This results in the production of mutated plasmid DNA. As DNA from almost all common *E. coli* strains is Dam methylated; parental DNA was then subject to digestion by DpnI, which targets sequence 5`-Gm^6ATC that is specific for methylated and hemimethylated DNA [16] leaving the mutated plasmid DNA undigested (as it is not methylated).

To perform site-directed mutagenesis QuickChange® Site-Directed mutagenesis kit (Stratagene) kit was used. Wild-type recombinant DNA plasmid was used as template

DNA with a set of mutagenic oligonucleotide primers (Table 3.3) specially designed to direct the mutations (in the case of the double mutant (L16F/W56G) mutated plasmid L16F was used as a template). The melting temperature (T_m) was calculated according to the following equation:[17]

$$T_m = 81.5 + 0.41\ (\%GC) - 675/N - \%\ mismatch$$

N is the primer length in bases and the values for %GC and % mismatch are integers. The manufacturer recommends the selected point of mutation to be in the middle of the primer with 10-15 unchanged bases each side, and should have a T_m of ≥78 °C and contain 25-45 bases with a minimum GC content of 40%.

Practically, 50 μL of reaction mixture was composed of a final concentration of 150 ng of each mutagenic primers, 10 ng plasmid DNA template, 1 μL of dNTP mix and 8% dimethyl sulfoxide 1x reaction buffer (10 mM KCl, 10 mM $(NH_4)_2SO_4$, 20 mM Tris-HCl pH 8.8, 2mM $MgSO_4$, 0.1% Triton X-100 and 5μg nuclease-free bovine serum albumin) then 1 μL of Pfu Turbo DNA polymerase (2.5 U/μL) was added and was subject to PCR in the thermal cycler machine which was set as the following:

Step 1- polymerase activation at 95 °C for 30 seconds.

Step 2- Template denaturation at 95 °C for 1 minute.

Step 3- Primers annealing at 55 °C for 1 minute.

Step 4- Extension at 68 °C for 6 minutes.

Steps from 2-4 was repeated for 16 cycles, at the end of last cycle the reaction mixture was kept at 4 °C.

The PCR product then digested at 37 °C for 1 hour by adding DpnI enzyme then transformed to XL1-Blue super competent cells (section 3.3.3) and the plasmid was isolated (section 3.3.5) and sent to (Eurofin MWG Operon, Germany) for sequencing to check whether the desired mutation occurred or not.

Primer	Sequence (5`-3`)
L16F	***For:*** G TAC CGC CAG TCG GCG **TTC** ACG CTG ATG GCC TCG CAC
	Rev: GTG CGA GGC CAT CAG CGT **GAA** CGC CGA CTG GCG GTA C
L16G	***For:*** G TAC CGC CAG TCG GCG **GGC** ACG CTG ATG GCC TCG CAC TTC GGC
	Rev: GCC GAA GTG CGA GGC CAT CAG CGT **GCC** CGC CGA CTG GCG GTA C
L16Y	***For:*** AGC TGC AAG GCC TGC **TAT** GAC GCC TAT CGC AAG
	Rev: CTT GCG ATA GGC GTC **ATA** GCA GGC CTT GCA GCT
H120A	***For:*** AGC TGC AAG GCC TGC **GCC** GAC GCC TAT CGC AAG
	Rev: CTT GCG ATA GGC GTC **GGC** GCA GGC CTT GCA GCT
H120N	***For:*** TGC AAG GCC TGC **AAC** GAC GCC TAT CGC
	Rev: GCG ATA GGC GTC **GTT** GCA GGC GTT GCA
H120F	***For:*** AGC TGC AAG GCC TGC **TTC** GAC GCC TAT CGC AAG
	Rev: CTT GCG ATA GGC GTC **GAA** GCA GGC CTT GCA GCT
W56G	***For:*** CTG TCG GCG CTG CCC **GGC** GCG GCT TTC GGT CCC
	Rev: GGG ACC GAA AGC CGC **GCC** GGG CAG CGC CGA CAG
W56Y	***For:*** CTG TCG GCG CTG CCC **TAC** GCG GCT TTC GGT CCC
	Rev: GGG ACC GAA AGC CGC **GTA** GGG CAG CGC CGA CAG
L16F/W 56G double mutation	***For:*** CTG TCG GCG CTG CCC **GGC** GCG GCT TTC GGT CCC
	Rev: GGG ACC GAA AGC CGC **GCC** GGG CAG CGC CGA CAG

Table 3.3: Site-directed mutagenic oligonucleotide primers. mutated nucleotides shown in red, primers were designed on the base on the wild-type DNA plasmid, the last pair of primers show double mutation in which L16F plasmid DNA has been used as a template.

3.3 Microbiology methods:

3.3.1 Bacterial strains, growth media and antibiotics:

Bacterial strains used in this study were BL21(DE3) (Promega) for the over-expression of the protein, DH5α (Invitrogen) for cloning and general plasmid amplifications and XL1-Blue super competent cells (Stratagene) for the site directed mutagenesis experiments.

Luria Bertani broth, LB, (10 g of NaCl, 5 g of yeast extract and 10 g of peptone per liter of deionised water and autoclaved at 121 °C for 15 minutes under pressure prior to use) was used for routine growth of bacteria and over expression.

To grow bacterial colonies on a solid medium, 1% (w/v) of agar was added to the LB medium, autoclaved, and appropriate sterile filtered antibiotics to the final concentration of 30 μg/mL of kanamycin (Km) and/or 20 μg/mL of chloramphenicol (Cm) were added after cooling the medium down to 50 °C and then poured into plates of 100 mm of diameter. Kanamycin was dissolved in deionised water because of its slight solubility in water, while chloramphenicol was prepared in methanol.

SOC medium was used for the recovery of the *E. coli* competent cells after the heat shock transformation prior plating cells on the antibiotic containing LB agar plates. A 100 mL of this medium was prepared as the following: a solution of 1 litre containing 20.0 g of tryptone, 5.0 g of yeast extract, 0.5 g of NaCl was prepared in deionosed water and then autoclaved (solution A). Prior to use, 10 mL of filter sterilised 1 M $MgCl_2$ and 1 M $MgSO_4$ was added to (solution A) producing (solution B). Immediately before the use, 2 mL of filter sterilised 20% (w/v) glucose was added to 98 mL of solution B.

The highly enriched medium of Terrific Broth media, TB, (24 g of yeast extract, 12 g of tryptone and 4 mL of glycerol dissolved in 850 mL of water autoclaved at 121 °C for 15 minutes, and then adjusted to 950 mL with 100 mL filter sterilized solution containing (0.17 M KH_2PO_4 0.72 M K_2HPO_4). In this medium, tryptone and yeast extract provide necessary nutrients for excellent growth of recombinant strains of *E. coli*. Potassium phosphates provide potassium for cellular system and buffer the medium (pH 7.2 ± 0.2) to prevent cell death due to drop in pH. Finally, glycerol is added as a carbon and energy source, unlike glucose, glycerol is not fermented to acetic acid. TB was used for the overproduction of recombinant AXCP and all mutants.

3.3.2 Preparation of E. coli competent cells:

Competent cells of *E. coli* strains DH5α and BL21(DE3) were prepared according to the method of Sambrook *et al.*[18] A single colony was selected from the LB plate and incubated into (10mL) LB liquid medium, grown overnight at 37 °C and sub-cultured (1:100) into a final volume. The sub-culture was incubated at 37 °C to an optical density (OD_{550nm} of 0.35). The cells were chilled on ice for (5 min) and then centrifuged at (8000 rpm) for (5 min) at (4 °C). The supernatant was removed and the cell pellet was resuspended in (2/5) volume of competent cell buffer 1 which composed of potassium acetate (30 mM), rubidium chloride (100 mM), calcium chloride (10 mM), manganese chloride (50 mM) and glycerol (15% v/v). The resuspended cells were chilled on ice for (5 min) then centrifuged as before. The supernatant was removed and the cell pellet was resuspended in competent cell buffer 2 which composed of MOPS (10 mM), calcium chloride (75 mM), rubidium chloride (10 mM), and glycerol (15% v/v). After 15 min period on ice the cells were aliquoted (200 μL) into pre-chilled microcentrifuge tubes and stored at (-80 °C).

3.3.3 Heat shock transformation of E. coli competent cells:

Transfection of *E. coli* with plasmid DNA involves in two main steps; binding of the DNA to the $CaCl_2$ treated cell (competent) surface at 0 °C and the insertion of DNA to the cell cytosol by heat-pulse from 0 °C to 42 °C. Practically, to bind the DNA to the competent cell surface, 2 μL DNA plasmid of recombinant or mutants were added to 200μL of competent *E. coli* DH5α, mixed gently and placed on ice for 30 minutes. Cells then were subject to a heat-pulse from 0 °C to 42 °C by placing the tube in a 42 °C hot plate for exactly 45 seconds to let the DNA plasmid enters the cell cytosol. To recover the *E. coli,* the tube has returned onto ice for 2 minutes and 500 µL of SOC medium containing no antibiotics was added and incubated at 37 °C for 1 hour. This step allows the synthesis of enough plasmid-encoded antibiotic resistance prior to plating cells on appropriate antibiotic-containing LB agar plates.

3.3.4 Selecting transformed cells:

Aliquots of 50 µL, 100 µL and 200 µL were spread onto Petri dishes of LB agar plates containing kanamycin (30 µg/mL) and incubated overnight. This medium will allow bacteria carrying the recombinant plasmid to grow, as kanamycin resistance gene is an integral part of pET 26b(+), while untransformed cells will die on the same plate. Single colonies, which grew were then transferred to 5 mL of liquid LB medium containing (30 µg/mL) kanamycin, to further ensure only resistant cells grow, an overnight culture incubation at 37 °C was completed, resulted in high-copy plasmid.

3.3.5 Extraction and purification of high-copy plasmids:

To extract the high copy plasmid, QIAprep spin miniprep kit (Qiagen, UK) has been used and the manufacturer's protocol was followed. Cells from the overnight culture (section 3.3.4) were harvested and suspended and lysed under alkaline condition in 250 µL buffer (P2) in the presence of RNAse (250 µL of buffer P1) which catalyses the cleavage RNA's phosphodiester bonds. Buffer (P2) consist of SDS and NaOH, SDS releases the cell contents by solubilising the phospholipid and protein components of the cell while NaOH denatures chromosomal DNA, plasmid DNA and proteins. The lysate is then neutralised and adjusted to high salt binding concentration by adding 350 µL buffer (N3), the mixture was thoroughly and gently mixed, the high salt concentration precipitates denatured chromosomal DNA, proteins, cellular debris and SDS while the renatures the plasmid DNA, the solution was then centrifuged for 10 minutes at 13,000 rpm, a compact white pellet formed. The supernatant was then transferred to the supplied spin column and centrifuged at the same speed for 1 minute to remove RNA, cellular proteins and metabolites while plasmid DNA remains on the silica membrane. Supplied columns have a silica membrane for selective adsorption of plasmid DNA in high-salt buffer and elution in low-salt buffer. Silica membrane was washed with 750 μL buffer PE, centrifuged for 1 minute and the flow through discarded, the spin column then placed in a clean 1.5 mL eppendorf. Plasmid DNA was then eluted with 30 μL nuclease-free water and collected by centrifuging for 1 minute, thus the resulting solution containing pET 26b(+)with the cloned cycP gene.

3.4 Cytochrome c′ overproduction:

1 μL of the recombinant DNA plasmid or the mutant plasmid and 1 μL of pEC86 were cotransformed into 200 μL of *E. coli* BL21(DE3) competent cells (Novagen) as a high level expression *E. coli* strain (section 3.3.3). Then aliquots of 25, 50, and 100 μL were spread on the LB agar plates containing 20μg/mL chloramphenicol and 30 μg/mL kanamycin then incubated overnight at 37°C. A single colony was picked and streaked onto a fresh LB agar plate containing the same antibiotics at the same concentrations and incubated overnight at 37 °C. Next day a single colony was picked from the plate and transferred into 5 mL of freshly prepared liquid LB medium containing both antibiotics and incubated overnight. The content then was sub-cultured (1:100) into a final volume of 500 mL in liquid LB. 50 mL of the overnight culture was transferred into 950 mL TB medium containing both antibiotics and incubated at 37°C with shaking at 180 r.p.m. After 5 hours 330 μL of 1M hemin (Sigma-Aldrich) was added, shaking was continued at 37 °C. 24 hours later 1 mL of the metal ion master mix (2 mM Ni^{2+}, 2 mM Co^{2+}, 10 mM Zn^{2+}, 10 mM Mn^{2+} and 50 mM Fe^{3+}) was added to the medium, and continued shaking on 37 °C for 96 hours then cells were harvested.

3.5 Extraction and Purification of Recombinant and mutant Cytochrome c′:

3.5 .1 Preparation of the crude extract:

Cells were harvested from the culture by centrifugation for 10 minutes at 8000 r.p.m and 4 °C. Four different techniques were compared to break open the cells;, mechanical lysis either French press or sonication and chemical lysis like osmotic shock and BugBuster® master mix (Novagen)

3.5.1.1 Mechanical lysis:

3.5.1.1a French press:

The operating principle of the French press is to pass a suspension from a compartment under high pressure, through a small hole, to a compartment of lower pressure. The sharp pressure gradient created disrupts biological membranes very efficiently. Cells were resuspended in 30 mL of 30 mM Tris-HCl pH 8, a pressure of 20,000 psi was supplied, the suspension was cooled on ice prior to passage through the French press. After passing the whole cells through the French press three times, the soluble fraction was separated by centrifugation at 11500 r.p.m for 45 minutes at 4 °C. The supernatant, containing the crude extract was collected for further analysis.

3.5.1.1b Sonication:

This method is used to disrupt biological membranes of bacteria by exposing them to high frequency sound waves. After suspending cells in 30 mM Tris-HCl buffer pH 8, the suspension solution was then sonicated four times for 30 seconds, to break open cells whilst avoiding overheating the solution. The solution was then centrifuged at 11500 r.p.m for 45 minutes at 4 °C. The supernatant was separated and collected for further analysis.

3.5.1.2 Chemical lysis:

3.5.1.2a BugBuster® master mix:

To chemically lyse cells; BugBuster® master mix (Novagen) was used. It is composed of Benzonase® Nuclease that reduces the viscosity of extract caused by liberation of DNA, and highly specific rLysozyme^{TM} which hydrolyses *N*-acetylmuramide linkages in the peptidoglycan layer of the cell wall which gently disrupted the cell wall and liberates soluble proteins.

After harvesting, cells were resuspended in room temperature BugBuster master mix solution by gentle vortexing using 5 mL reagent per gram of wet cell paste. Cell suspension incubated on a shaking platform at a slow setting for 20 minutes at room temperature. insoluble cell debris were then removed by centrifugation at 11500 r.p.m. BugBuster® master mix

3.5.1.2b Osmotic shock:

Treating bacteria with a high concentration of sucrose makes them dessicate through osmosis, and then a rapid exchange with a sucrose free buffer causes water to rush back into the cells with enough force to break open the cell wall. The pellet from 1 L culture was resuspended in 100 mL of osmotic solution (30mM Tris-HCl, 20% (w/v) sucrose and 1mM EDTA) and incubated at room temperature for 10 minutes with continuous shaking, then centrifuged for 10 minutes at 8000 r.p.m and 4 °C. The pellet was resuspended in 100 mL of ice-cold 5 mM $MgSO_4$, incubated on ice for 10 minutes with shaking and centrifuged again for 10 minutes at 8000 r.p.m and 4 °C. The supernatant should contain the periplasmic fraction and the pellet contains the rest of cells. The pellet can then be resuspended in a low osmotic pressure buffer which releases the intracellular contents.

3.5.2 Dialysis:

The crude extract was dialyzed against 5000x deionised water for 48 hours with changing the water every 4 hours. The dialysis tube of (MWCO 10000 Da) was used in this experiment; the tube has been boiled in 10 mM EDTA for 10 minutes and then rinsed well with deionised water prior to transferring of the crude extract. Dialyzed solution was then centrifuged at 20000 r.p.m for 60 minutes at 4 °C. The supernatant was the subject to further purification using cation exchange chromatography.

3.5.3 Cation exchange chromatography:

Carboxymethyl cellulose CM-52 (Sigma-Aldrich) was prepared as described by the manufacturer. The dry resin was suspended in 5 volumes of distilled water and allowed to settle for 45 minutes. The settled volume of the resin was measured, this was used as the column volume (CV) for the purpose following steps. The suspension was filtered and resuspended in 2 CV of 0.1 M NaOH containing 0.5 M NaCl for 10 minutes. The slurry was then filtered on Buchner funnel while applying gentle suction and washed with 2 CV of the above solution. This step was repeated using 0.5 M NaCl only, and repeated again using 0.1 M HCl containing 0.5 M NaCl. Finally the same step repeated using deionised water. The resin was washed using 5-10 CV deionised water or until the effluent pH was 5 or greater, resuspended again in 2 CV of 1 M NaCl and the pH of slurry adjusted to 7-8 with NaOH and then filtered and washed using 5 CV of deionised water passed through the resin on the filter. The resin was then resuspended in 2 CV of 400 mM MES [2-(N-Morpholino)ethanesulfonic acid)] buffer pH 6 (Melford) and then filtered and resuspended again in 2 CV of 40 mM of the same buffer. Finally the resin was packed into the column (model XK 16 – Pharmacia biotech)

3.5.4 Protein purification:

The dialysed crude extract was loaded onto the column, washed with 2 CV deionised water and then with 2 CV of 40 mM MES buffer pH 6 to remove unbound proteins. Cytochrome c′ bound on the top of the column as a dark brown band. Pure AXCP was eluted by applying a 2CV linear gradient of NaCl between 50 mM – 200 mM in 40 mM MES buffer pH6 and collected in 15 mL falcon tubes.

3.5 Sample concentration:

Purified proteins were concentrated to ca (40mg/mL) using 10,000 Da MWCO Vivaspin 2 (Sartorius Stedim Biotech).

3.6 Sodium Dodecyl Sulphate-PolyacrylAmide Gel Electrophoresis (SDS-PAGE):

SDS-PAGE first described by Laemmli[19] was used to asses the purity of protein samples and in combination with the use of molecular weight marker, allowed the estimation of its molecular weight.

The glass plates were cleaned with ethanol. The 15% acrylamide resolving gel was prepared with double de-ionised (ddH_2O) (3.4 ml), degassed acrylamide (30%, 4.0 ml) (BioRad), gel buffer; Tris (tris (hydroxy methyl) aminomethane) pH 8.8 (2.5 ml), and SDS (10%, 100 µl). The stacking gel was prepared in the same way except that the pH of the gel buffer was 6.8. Immediately prior to the gel being poured, ammonium persulphate (APS) (10%, 100 µl) and N,N,N',N'-tetramethylethylene diamine (TEMED) (10 µl) were added to the appropriate gel solution. TEMED initiates gel polymerization once swirled. The resolving gel was poured into the plate until it was 1 cm from the top, then water-saturated ethanol was poured on top of the gel to prevent the gel from contact with the air and subsequent crust formation. When the resolving gel was set, the ethanol was washed away with distilled water. The staking gel was poured onto the resolving gel, and a comb was pushed into the stacking gel. The samples were mixed with loading buffer, made up with: electrophoresis buffer (1.0 ml), glycerol (3.0 ml), 0.5% bromophenol blue (0.2 ml), mercaptoethanol (0.8 ml) and dH_2O (5.0 ml), and then heated to 95 °C for 5 min to denature the protein. These samples were then loaded onto the gel, which was run at 150 V for 45 min in an electrolyte solution of 1 x running buffer. The gel was then stained with Coomassie solution for 30 min made up with Coomassie Blue (1.25 g), methanol (500 ml), and acetic acid (100 ml), made up to (1 L) with distilled water. The gel was then de-stained with a mixture of acetic acid (50 ml) and industrial methylated spirits (IMS) (100 ml), made up to 500 ml with distilled water.

3.7 Biochemical methods:

3.7.1 Reduction and oxidation of samples:

Proteins were reduced using sodium dithionite (Sigma-Aldrich). A cocktail buffer – designated MHTPC buffer (50 mM MES, 50 mM HEPES, 50 mM TAPS, 50 mM PIPES and 100 mM CAPS) titrated to pH 6 was bubbled with N_2 to displace dissolved oxygen. An appropriate amount of concentrated protein was resuspended in the cocktail buffer inside a quartz cuvette to a final volume of 1 mL. The cuvette was sealed using a rubber septum and purged with high-purity nitrogen to displace the oxygen that trapped above the solution. 200 mM of sodium dithionite (DT) was prepared in deoxygenated MHTPC, and bubbled with high-purity nitrogen again. Aliquots of DT were added to the super-sealed cuvette using a Hamilton syringe (Hamilton), mixed with a magnetic flea until it was fully reduced. Any extra DT remaining in the cuvette was removed by vortexing the cuvette so the solution contacts the rubber septa and any excess DT will react with diffused oxygen in the rubber. Ultraviolet and visible (UV/Vis) spectroscopy was used to check the reduction and excess DT removal process.

L16G mutant needed to be oxidised to remove the ligand bound to the haem iron. 100μL of concentrated mutant sample was oxidised using 100 μL of 200 mM potassium ferricyanide prepared in MHTPC and kept in room temperature for 30 minutes. The oxidised protein was then passed down on a P6DG column to remove the oxidant in excess. The protein then eluted with 500 μL of MHTPC buffer pH 6.

3.7.2 pH titrations:

To make the buffer composition same for all pH range used for the titration, a stock 5x cocktail buffer of 250 mM MES, 250 mM HEPES, 250 mM TAPS, 250 mM PIPES and 500 mM CAPS (5x MHTPC) was prepared. Prior to use, diluted to 1x and pH adjusted at 5, 6, 7, 8 and 9. To study the pH effect on recombinant and mutants 1 μL - 2 μL of the sample was diluted to 1 mL with 1x MHTPC buffer at desired pH, stirred and the UV-Vis spectrum was recorded using Perkin Elmer UV/VIS spectrophotometer Lambda 25.

3.7.3 NO titration:

NO saturated water was prepared as previously described by Barbieri *et al.*[20] For further purification of NO gas (BOC) from higher nitrogen oxides, NO gas was mixed and stored over 1 mM NaOH under a positive pressure (10 cm H_2O). The saturated solution of NO was prepared in a 10 mL assay bottle fitted with rubber septa, the assay bottle was flushed with high-purity nitrogen for 5 minutes, 50 mL of NO gas was removed by syringe from the storage vessel and injected into the assay bottle, vented to atmospheric pressure by syringe needle. This process was repeated again then 2 mL of deionised water previously flushed with high-purity nitrogen for 5 minutes was injected to the assay bottle, vented during the injection. To achieve equilibrium of NO into the liquid phase, the bottle was shaken, this will give a saturated solution of 1.94 mM NO at 25 °C.[21]

Aliquots of NO saturated H_2O were successively added to 1 mL of reduced protein solution in the super sealed quartz cuvette under nitrogen by 10 μL gas tight syringe. The solution of NO and protein was mixed with a magnetic stirrer and the UV-Vis spectrum was recorded Perkin Elmer Lambda 35 spectrophotometer.

3.7.4 CO titration

CO saturated H_2O was prepared in the same way as for NO saturated H_2O except that further purification and storage of CO was in deionised water instead of 1 mM NaOH.

3.7.5 UV-Vis spectrophotometry:

UV-Vis spectrophotometry was used to determine the purity, assess the concentration and to characterize the recombinant and mutant cytochrome c′. All spectrophotometric experiments were performed at room temperature using Perkin Elmer Lambda 35 spectrophotometer. The protein purity was estimated by dividing A_{400nm} by A_{280nm} (A_{400}/A_{280}) Concentrations were determined using a molar extinction coefficient for the Soret band of 80 mM^{-1} cm^{-1} for ferric cytochrome c′.[22]

3.8 References:

[1] Lehninger A.L., Nelson D.L, and Cox M.M (1997) "Principles of Biochemistry" 2nd Ed., Worth publishers, Inc.

[2] Strick T, Allemand JF, Croquette V and Bensimon D (2000) "Twisting and stretching single DNA molecules" *Progress in Biophysics and Molecular Biology* **74,** 115-140.

[3] Yakovchuk P., Protozanova E. and Frank-Kamenetskii M. D. (2006)"Base-stacking and base-pairing contributions into thermal stability of the DNA double helix" Nucleic Acids Research **34**(2) 564-574.

[4] Edwards, K.J., Brown, D.G., Spink, N., Neidle, S. (1992) "Molecular structure of the B-DNA dodecamer d(CGCAAATTTGCG)2. An examination of propeller twist and minor-groove water structure at 2.2 A resolution." *J Mol Biol* **226**(4), 1161-1173.

[5] Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004) "UCSF Chimera - A Visualization System for Exploratory Research and Analysis." *J. Comput. Chem.* **25**(13), 1605-1612.

[6] Fiers W, Contreras R, Haegemann G, Rogiers R, Van de Voorde A, Van Heuverswyn H, Van Herreweghe J, Volckaert G, Ysebaert M (1976) "complete nucleotide sequence of SV40 DNA" Nature **273**(5658), 113-120.

[7] Venter J.C., *et al* (2001) "The sequence of human genome" Science **292**(5507), 1304-1351.

[8] Mathews C.K., Holde K.E., Ahern K.G (2000) "Biochemistry" 3rd Ed., Addison Wesley Longman, Inc. 100-105

[9] Knight R.D., and Landweber L.F. (1998) "Rhyme or reason: RNA–arginine interactions and the genetic code" Chemistry and biology 5(9), R215-R220.

[10] Horton H.R., Moran L.A., Scrimgeour K.G., Perry M.D. and Rawn J.D (2006) "Principles of biochemistry" 4th Ed., Pearson Education, Inc.

[11] Hadi S.H., and Yuan R. (1974) "Complementation in vitro by mutant restriction enzymes from Escherichia coli K" The journal of biological chemistry 249(14), 4580-4586.

[12] Arslan E., Schulz H., Zufferey R, Künzler P and Thöny-Meyer L (1998) "Overproduction of Bardyhizobium japonicum c-type cytochrome subunit of the cbb3 oxidase in Escherichia coli" Biochemical and biophysical research communications 251, 744-747.

[13] KOD hot start DNA polymerase handbook (2006)

[14] QIAguick® spin handbook March (2008)

[15] Protocols and application guide (1996) , 3rd Ed., Promega corporation.

[16] Weiner M.P., Costa G.L., Schoettlin W., Cline J., Mathur E. and Bauer J.C. (1994) "Site-directed mutagenesis of double-stranded DNA by the polymerase chain reaction" Gene, 151(1-2), 119-123.

[17] QuickChange Site-directed mutagenesis kit handbook (2004), Stratagen.

[18] Sambrook J., Russell D.W (2001) "Molecular cloning, a laboratory manual" 3rd Ed., cold spring harbour laboratory press, New York.

[19] Laemmli U.K. (1970) "Cleavage of Structural Proteins During the Assembly of the Head of Bacteriophage T4", Nature, 227, 680-685.

[20] Barbieri S., Murphy L.M, Sawers R.G., Eady R. R., Hasnain S.S. (2008) "Modulation of NO binding to cytochrome c′ by distal and proximal haem pocket residues" Journal of biological inorganic chemistry 13, 531-540.

[21] Zacharia I. G., Deen W. M. (2005) " Diffusivity and solubility of nitric oxide in water and saline" Annals of biomedical engineering 33(2), 214–222

[22] Cusanovich M. A., Tedro S. M. and Kamen M.D. (1970) "Pseudomonas denitrificans cytochrome cc′ " Archives of biochemistry and biophysics, 141, 557-570.

Chapter 4: Cloning, expression, purification and generation of cytochrome c′ mutants

4.1 Introduction:

The generation of an expression clone that is capable of producing the protein of interest in a soluble form in high yield is the rate-limiting step in protein production. Protein expression in heterologous systems often fails, is expressed as insoluble aggregates, or cannot be purified by standard methods. Thus, returning to the cloning stages to make a new construct with a different purification tag, moving the target gene from one vector to another; or to express the protein in a different host are required to enable the system to express the protein. That is why protein expression pipe lines can be time consuming.[1]

As explained in chapter one, apo-cytochrome c polypeptide reaches the periplasm by *Sec*-dependent translocation using a signal peptide, where the signal peptide must be cleaved. Post-translational protein modification then occurs, which leads to the cytochrome c maturation. The maturation means the conversion of a linear peptide into a three dimensionally structured polypeptide and the covalent attachment of the haem group.[2] This covalent attachment occurs in the periplasmic space through the formation of two covalent bonds between two cysteine residues of a CXXCH motif in the protein and two vinyl groups of haem.[3] This attachment is a difficult process and is not straightforward.[4] It requires the involvement of cytochrome c maturation (Ccm) proteins, which are eight specific proteins, organised in a membrane protein complex[2] named CcmABCDEFGH. Among them only the structure of the apo form of CcmE is known.[5] These have been cloned into pEC86 plasmid[6] and co expressed with the constructed plasmids that produce apo-cytochrome c′.

4.2 Construction of plasmid vectors:

The expression vector pET 26b(+) was chosen for reasons explained in section (3.2.1). Using the above vector, four different recombinant plasmids have been constructed as described in section (3.2.2) in order to obtain a higher expression level of the cytochrome c′. The first attempt was to produce AXCP using the *E. coli* sequence leader (*pelB*) which was incorporated to the plasmid pET26(b) while the AXCP signal sequence was excluded, and the His-tag was integrated for purification purposes, giving the new plasmid construct named pET-H. The second attempt was the same as pET-H but without the His-tag. The plasmid was then named pET-NH. In the third variation, the AXCP signal peptide and the His-tag were included and the new plasmid vector named pET-HS. The last construct was the same as pET-HS but with the His-tag excluded, the construct was then named pET-NHS.

4.2.1 Construction of pET-H:

In this construct, the AXCP gene without its signal peptide was cloned between two restriction sites NcoI and HindIII using two primers, cyc_H_for and cyc_H_rev. This will express a protein with the *pelB* leader which is already inserted in the expression vector pET 26b(+) (see chapter 3, figure 3.2). To include the His-tag for purification purposes, the reverse primer was designed to exclude the stop codon (TAA) of the AXCP gene, so the expression continues to include histidine residues in the C-terminus. The forward primer cyc_H_for was designed with NcoI restriction site and cyc_H_rev with HindIII restriction site (see table 3.2 material and methods). The PCR was performed in two tubes as described in section (3.2.3) and fragments of 381 base pairs (bp), corresponding to the AXCP gene without its signal peptide and the stop codon from *gene* was removed. To visualise the PCR product, contents of both tubes were run on 1% (w/v) agarose gel as described in section (3.2.5). Fragments of ~400 bp were detected on the gel under UV illuminator (Figure 4.1)

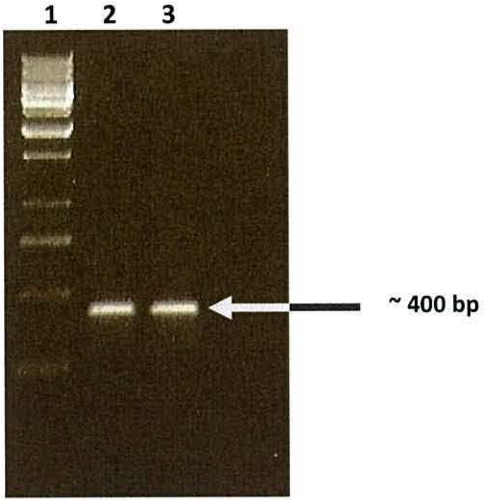


Figure 4.1: 1% (w/v) agarose gel electrophoresis visualised under UV illuminator. Lane 1 is the DNA molecular marker, Lanes 2 and 3 are two replications of the amplified AXCP gene by PCR. Fragments of ~400 bp have been detected using cyc_H_for and cyc_H_rev primers.

The contents of both tubes were mixed and then purified using QIAquick PCR purification kit (Qiagen, UK) as described in detail in (section 3.2.4).

In order to clone the AXCP gene fragments obtained from the PCR into the expression vector pET 26b(+). Complementary ends of the gene and the plasmid needed to be produced. To do so, the purified PCR product and the vector were digested separately in two different tubes using restriction enzymes NcoI and HindIII, as described in section (3.2.7). The complementary ends on the gene fragments and on the plasmid vector were produced. They were mixed and further purified together (see section 3.2.4). The complementary ends were joined using T4 ligase enzyme with proper buffer and incubation overnight at 16 °C (see section 3.2.8).

The ligation product was heat shock transformed into DH5α competent cells (see section 3.3.3). Aliquots of different volumes were plated onto the LB agar Petri dishes containing kanamycin (30μg/mL) and incubated overnight at 37 °C. Since the kanamycin resistance gene is integrated as part of the pET 26b(+), only transformed bacteria carrying the recombinant plasmid grew, while untransformed ones died on the antibiotic LB-agar plate (see section 3.3.4). For plasmid preparation, the day after, 10 single colonies were transformed from LB-agar plates into 10 different tubes containing 5 mL liquid LB medium and the kanamycin at the same concentration and incubated over night at 37 °C. Cells were then harvested and plasmids were extracted (see section 3.3.5). A 7 µL aliquot of extracted plasmids was double digested with NcoI and HindIII. The digested product was then run on 1% (w/v) agarose gel electrophoresis to ensure the insertion of the gene into the plasmid (see section 3.2.9) (figure 4.2).

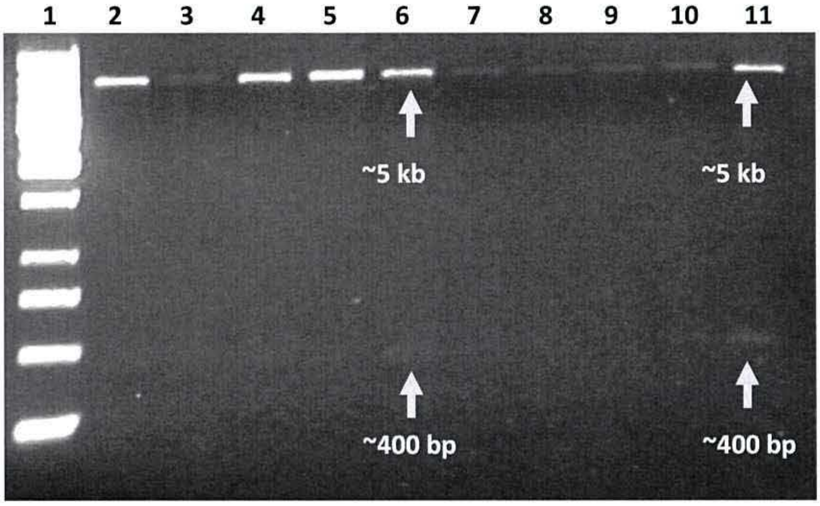


Figure 4.2: 1% (w/v) agarose gel electrophoresis of the double digested recombinant plasmid. Digested with both enzymes NcoI and HindIII. Lane 1 is the DNA molecular weight marker. Lanes 2-11 are digested recombinant plasmids. Lane 6 and 11 showed two bands, one at the expected molecular weight of the AXCP gene (400 bp) and another at about (5 kbp) corresponds to the digested vector pET 26b(+).

Plasmids of clone 6 and 10 show two bands, one at ~400 bp corresponding to the AXCP gene and another at ~5 kb corresponding to the pET 26b(+) plasmid. These results demonstrate the successful cloning of the AXCP gene into the pET 26b(+) vector in clones 6 and 11. These two clones were sent for sequencing to Eurofins MWG operon, Germany. The sequencing result confirmed the success of the AXCP gene cloning into pET 26b(+) without mutation. In the other clones the AXCP gene may either not have been cloned into the vector or the bands at ~400 bp are very faint so they were undetectable under the UV illuminator.

4.2.2 Construction of pET-NH:

Construction of a recombinant plasmid that carries the *pelB* leader of the vector but without the His-tag was the aim of this construct.

Two primers have been used to PCR amplify the AXCP gene from the *A. xylosoxidans;* cyc_NH_for with NcoI restriction site and cyc_NH_rev with BamHI restriction site (see table 3.2 material and methods). The forward primer was designed to include *pelB* leader sequence of the vector, and the reverse primer designed to include the stop codon in order to exclude the His-tag sequence. The PCR was performed as three replicates as described in (section 3.2.3). Fragments of 384 bp were isolated. 1% (w/v) agarose gel electrophoresis was run for all three replications (see section 3.2.5). Multiple bands were detected for all replicates on the gel (Figure 4.3).

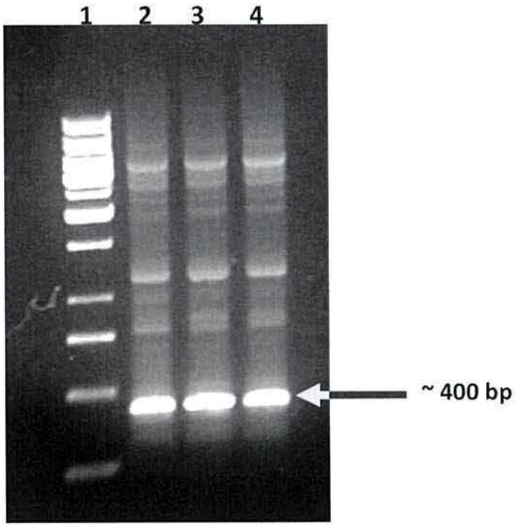


Figure 4.3: 1% (w/v) agarose gel electrophoresis. Lane 1 is the DNA molecular marker, Lanes 2-3 are replications of the PCR. Fragments of ~400 bp correspond to AXCP gene have been detected using cyc_NH_for and cyc_NH_rev primers, however, the gel showed multiple fragments.

Despite several attempts to eliminate the multiple fragments and optimise the PCR reaction, the background fragments were still visible (figures not shown). Different magnesium ion (Mg^{2+}) concentrations were used to increase the rate of dissociation of polymerase enzyme from the template by destabilising the enzyme-nucleic acid interaction.[7] Different template concentrations and melting temperatures were used as well. Desired fragment at ~400 bp (Figure 4.3) was extracted from the gel using the gel extraction kit (Qiagen, UK), as described in (section 3.2.6) and run on the agarose gel again (Figure 4.4). It shows a single fragment revealing the presence of the required gene only.

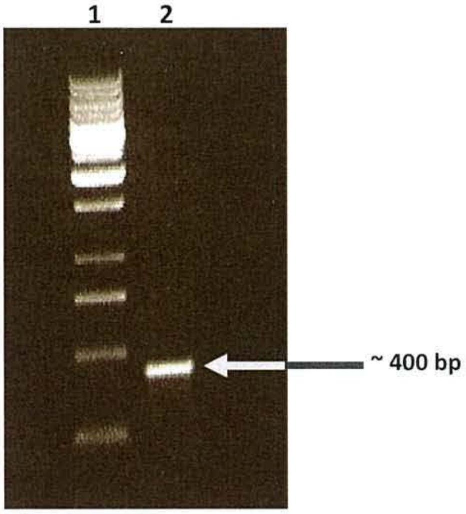


Figure 4.4: 1% (w/v) agarose gel electrophoresis of the gel-extracted and purified fragment from the gel in the previous figure.

To clone the gene into pET 26b(+) and produce the recombinant plasmid, the purified PCR product (fragments) and the pET 26b(+) were double digested using restriction enzymes NcoI and BamHI as described in (section 3.2.7), purified (see section 3.2.4) and ligated overnight at 16 °C using T4 ligase enzyme (section 3.2.8). These steps have already been discussed in the previous section.

The recombinant plasmid was then heat-shock transformed into DH5α competent cells (see section 3.3.3), plated on LB agar containing kanamycin (30μg/mL) and grown overnight at 37° (section 3.3.4). The next day, a few colonies were observed. Single colonies were picked and inoculated into 5 mL of LB medium containing kanamycin (30μg/mL), and grown overnight at 37 °C. The high-copy plasmids were extracted and purified as described in (section 3.3.5). Then, to check the incorporation the AXCP gene into the plasmid, a fraction of purified plasmid from each clone was double digested with both restriction enzymes; NcoI and BamHI, and 1% (w/v) agarose gel electrophoresis was run (section 3.2.9) (Figure 4.5).

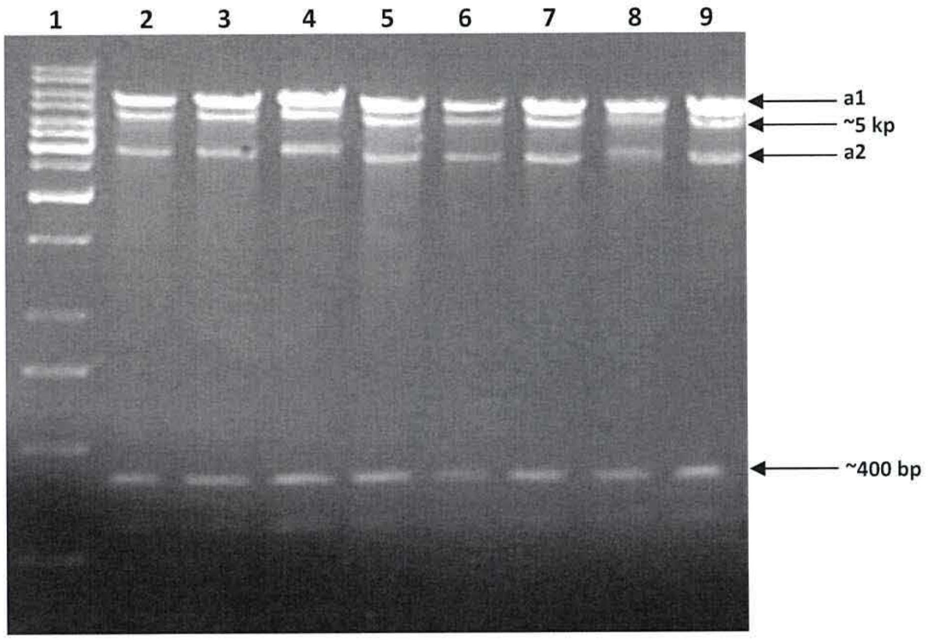


Figure 4.5: 1% (w/v) agarose gel electrophoresis of the double digested recombinant plasmid. Digested with both enzymes NcoI and BamHI. Lane 1 is the DNA molecular weight marker. Lanes 2-9 are digested recombinant plasmids. At all lanes five bands have been detected, one at (~400 bp) the expected molecular weight of the AXCP gene, another one at about (5 kbp) corresponds to the digested vector pET 26b(+), bands (a1) and (a2) are undigested pET 26b(+).

Figure 4.5 shows that in all selected clones, the AXCP gene was successfully incorporated into the expression vector pET 26b(+) (Lanes 2-9). A band at ~400 bp corresponds to the AXCP gene and another at ~ 5 kb belongs to the digested pET 26b(+) vector. Since the double stranded DNA can migrate on agarose gel in at least two forms, depending on how compact the molecule is, two extra bands have appeared (band **a1** and **a2**). Band (**a1**) represents the closed circular DNA that is undigested pET 26b(+) while (**a2**) represents the super coiled molecule, which is more tightly packed, and hence migrates as a smaller molecule. When the circular DNA is digested on both strands, the molecule is linear and more relaxed and hence shows a single migrating fragment.

Recombinant plasmids from two clones were sent to Eurofins MWG operon, Germany for sequencing. The result confirmed the successful cloning of AXCP gene without mutations.

4.2.3 Construction of pET-SIG-HS:

Expression of cytochrome c′ using its signal peptide with the His-tag in its C-terminus was the reason for constructing pET-HS recombinant plasmid. In this case, the *pelB* leader of the pET 26b(+) was removed. Instead, the signal sequence of the cytochrome c′ was included to direct the protein to the periplasmic space. Two primers were designed; the forward primer, named cyc_sig_H_for and the reverse one, named cyc_sig_H_rev (Table 3.2). The forward primer designed with NdeI restriction site, while the reverse one designed with HindIII restriction site, while the stop codon of the AXCP has not been included in order to integrate the His-tag in its C-terminus. The AXCP gene with its signal peptide was PCR amplified using above primers as described in (section 3.2.3) and represented in (Figure 3.4). A fragment of 456 base pairs was isolated (Figure 4.6).

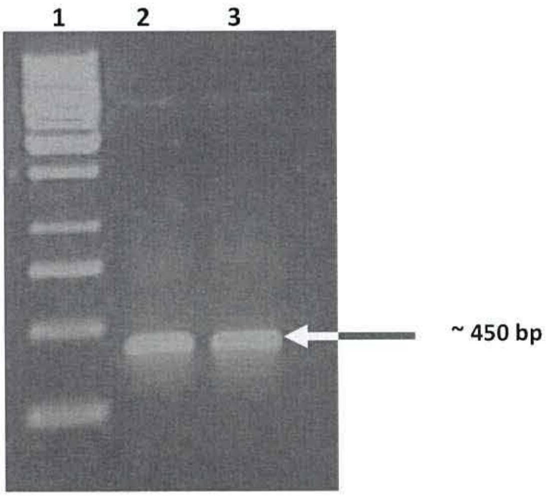


Figure 4.6: 1% (w/v) agarose gel electrophoresis of PCR product using cyc_sig_H_for and cyc_sig_H_rev primers. Lane 1 is DNA molecular weight marker, lanes 1 and 2 are two replicates of the PCR product of AXCP gene with its sequence leader. Bands at ~450 bp represent the expected molecular weight of the gene.

In order to obtain a high purity sample, the fragment at ~450 bp was extracted from the gel as described in section 3.2.6. The extract and the expression vector pET 26b(+) were double digested with both restriction enzymes NdeI and HindIII (see section 3.2.7) then purified together and vacuum dried then ligated at 16 °C overnight using T4 ligase enzyme (see section 3.2.8). The ligated recombinant DNA was heat-shock transformed into the DH5α competent cells (see sections 3.3.3) and plated on kanamycin (30μg/mL) LB agar plates and grown at 37 °C overnight. 12 single colonies were picked from the plate and inoculated into a 5 mL of LB containing the kanamycin (30μg/mL) and grown overnight at 37 °C (Section 3.3.4). Cells were then harvested and the plasmid was extracted (see section 3.3.5). The extracted recombinant DNA was double digested using both restriction enzymes NdeI and HindIII to check the success of the cloning (3.2.9) and run on 1% agarose gel (Figure 4.7). Four clones (Lanes 3, 7, 10 and 11) showed the successful cloning. Plasmids from clones 10 and 11 were sent to Eurofins MWG operon, Germany for sequencing. The sequencing results showed the successful cloning of the AXCP gene.

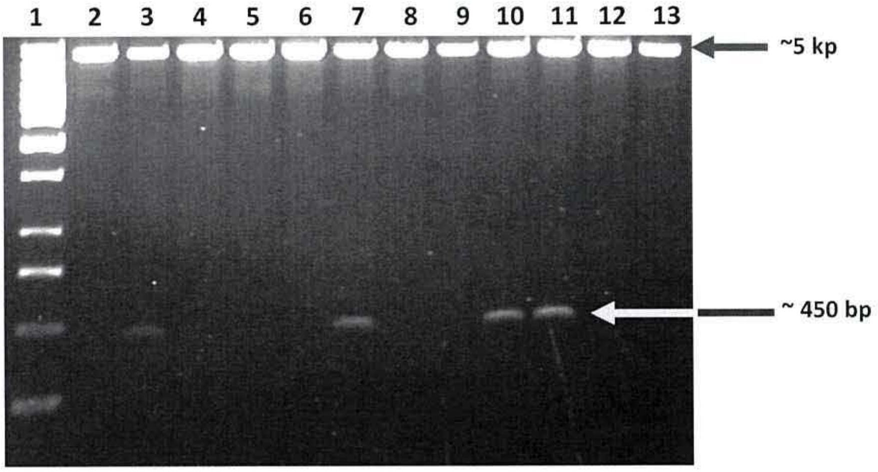


Figure 4.7: 1% (w/v) agarose gel electrophoresis of the double digested recombinant DNA plasmid. Digested with NdeI and HindIII restriction enzymes. Lane1 is the DNA molecular marker and Lanes (2-13) are from the selected colonies. Lanes (3, 7, 10 and 11) showed two band representing AXCP gene with its signal peptide at (~450 bp) and another one at (~5 kb) represents the digested pET 26b(+).

4.2.4 Construction of pET-NHS:

The last variation for constructing recombinant DNA plasmids was pET-NHS. This plasmid was constructed to carry the cytochrome c′ gene and its signal sequence without the His-tag sequence. Hence, two primers were designed; cyc_sig_NH_for as a forward primer with NdeI restriction site, and cyc_sig_NH_rev as a reverse primer with BamHI (Table 3.2). The stop codon TAA of the AXCP gene was included in order to stop the translation of the His-tag sequence to histidine residues in its C-terminus during the protein expression.

The PCR was performed using the primers as described in (section 3.2.3) to amplify the AXCP gene with its signal sequence (see figure 3.4). As was expected, a fragment at ~450 bp which represents the AXCP gene size plus the signal sequence was detected on 1% (w/v) agarose gel (Figure 4.8)

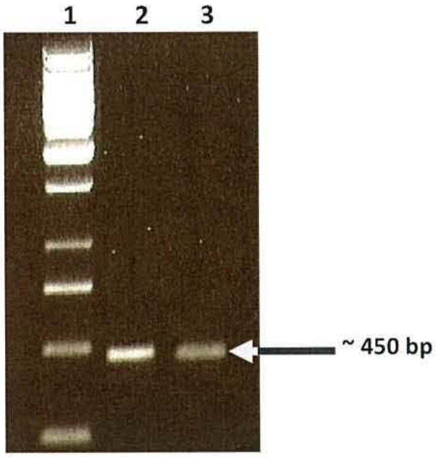


Figure 4.8: 1% (w/v) agarose gel electrophoresis of PCR product using cyc_sig_NH_for and cyc_sig_NH_rev primers. Lane 1 is DNA molecular weight marker, lanes 1 and 2 are two replicates of the PCR product of AXCP gene with its sequence leader. Bands at ~450 bp represent the expected molecular weight of the gene.

No multiple bands were detected on the gel, hence there was no need for the gel extraction. Contents of both tubes were mixed and then purified using QIAquick PCR purification kit (Qiagen, UK) as described in (section 3.2.4). The purified gene and the expression vector pET 26b(+) were digested separately using restriction enzymes NcoI and BamHI at 37 °C for 45 min (see section 3.2.7). Digested products of the gene and the vector were mixed and the purified together using QIAquick PCR purification kit (Qiagen, UK) (section 3.2.4). The purified products were vacuum dried and then ligated using T4 ligase enzyme overnight at 16 °C. The ligated recombinant plasmid was then heat-shock transformed into competent cells; DH5α. Aliquots were plated on kanamycin (30 μg/mL) LB agar plates and grown overnight at 37 °C. The day after, a few colonies had grown, 9 of them were selected and inoculated to 5 mL of LB containing (30 μg/mL) kanamycin and incubated overnight at 37 °C. Plasmids were then extracted from harvested cells of the overnight culture, and double digested with restriction enzymes NdeI and BamHI. The digested product was run on 1% (w/v) agarose gel to check whether the correct gene was cloned or not (Figure 4.9).

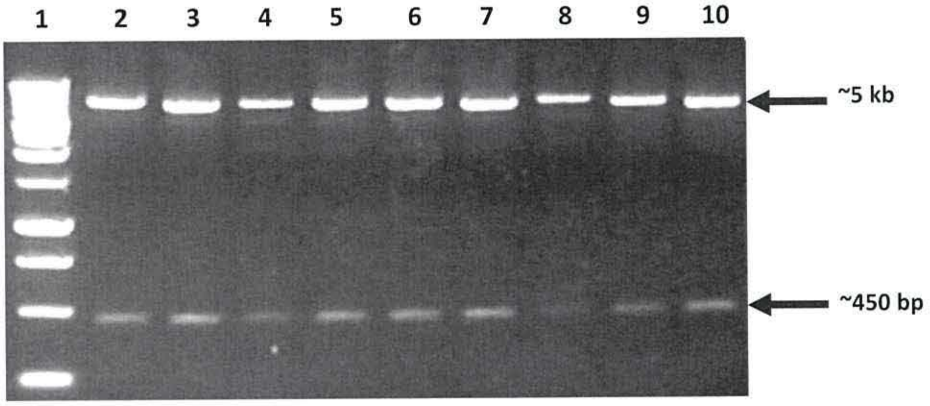


Figure 4.9: 1% (w/v) agarose gel electrophoresis of the double digested recombinant DNA plasmid. Digested with NdeI and BamHI restriction enzymes. Lane1 is the DNA molecular marker and Lanes (2-10) are from the selected colonies, all clones showed two band representing AXCP gene with its signal peptide at (~450 bp) and another one at (~5 kb) represents the digested pET 26b(+).

All plasmids showed two band at expected molecular weights; ~450 bp and 5 kb represents AXCP gene with the signal peptide and pET 26b(+) respectively. Clones 3 and 7 were sent for sequencing. The sequencing results confirmed the correct cloning of the gene with its signal peptide.

4.3 Cytochrome c′ over-production.

pEC86 which carries the cytochrome c maturation genes (CcmABCDEFGH) is essential for the covalent attachment of haem to the apo-cytochromes in the periplasm.[8] Cytochrome c′ from *A. xylosoxidans* could not be expressed or correctly folded without a maturation system present.[9] Therefore, all expression attempts in this thesis were carried out in the *E. coli* strain BL21 (DE3) transformed with pEC86. Hence, competent cells of high level expression *E. coli* BL21 (DE3) transformed with pEC86 were prepared following the protocol described in section 3.3.2 and were then transformed with the earlier specified constructed plasmids. The plasmid pEC86 was a kind gift from Dr. Thöny-Meyer through Dr. Julie Stevens – Department of Biochemistry at Oxford University.

4.3.1 Overproduction with pET-H and pET-NH:

High level expression *E. coli* strain BL21 (DE3)-pEC86 competent cells were heat-shock transformed (see section 3.3.3) with pET-H plasmid which carries the His-tag and the *E. coli* signal peptide (*pelB* leader). Since the plasmid pEC86 is chloramphenicol (Cm) resistant and pET-H is kanamycin (Km) resistant, all growth media were containing both antibiotics at concentrations of (20μg/mL) Cm and (30μg/mL) Km.

For a small scale experiment, three tubes containing 10 mL of LB with appropriate antibiotics were inoculated with 100 μL of an overnight culture. Ferriprotoporphyrin IX chloride (Hemin chloride) was added to all of them to final concentration of 10 μg/mL. The first tube was grown until an OD_{600} reached 0.6 at 37 °C, and then induced for 4 hours with 0.25 mM IPTG with continuous shaking at 180 rpm. The two other tubes were left to grow for 24 and 48 hours from the inoculation. The negative control was prepared by transforming BL21 (DE3)-pEC86 strain with pET 26b(+) and grown for 48 hours.

To visualise the recombinant cytochrome c′ overproduction, 1 mL of the culture was centrifuged at 8000 rpm. for 10 minutes. Cells were resuspended in 200 μL SDS-loading buffer, heated to 95 °C for 5 minutes, centrifuged again at 13000 rpm for 2 minutes, and 10 μL from the supernatant was run on 12% (w/v) acrylamide gel (Figure 4.10)

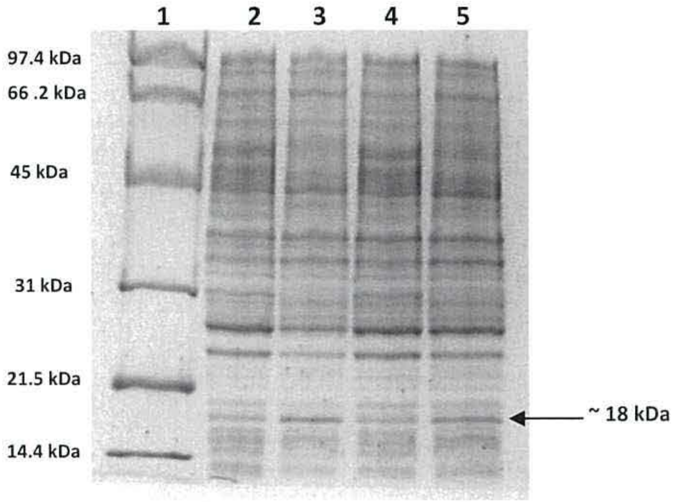


Figure 4.10: 12% (w/v) Acrylamide SDS-PAGE electrophoresis of recombinant cytochrome c′ overproduction in *E. coli* BL21 (DE3) strain in LB media at 37 °C. Lane 1 is the molecular weight marker. Lane 2 is the negative control where BL21 (DE3) strain transformed with pET 26b(+) and pEC86. Lane 3 is BL21 (DE3)-pEC86 transformed with pET-H and induced for 4 hours with 0.25 mM IPTG. Lanes 4 and 5 are BL21 (DE3)-pEC86 and pET-H grown for 24 and 48 hours respectively without induction at. A band at ~18 kDa represents unprocessed AXCP.

Although the molecular weight of cytochrome c′ is 28 kDa as a dimer, the loading buffer of SDS-PAGE denatures the protein and prevents the electrostatic interactions between the two monomers, and hence, cytochrome c′ appears as a band at 14 kDa on the polyacrylamide gel.

It is obvious from Figure 4.10 that no band at 14 kDa has appeared, while an intense band at ~18 kDa is clear. Since the pET-H plasmid was constructed to produce cytochrome c′ using the *E. coli* signal sequence *pelB* in its N-terminus, which is ca. 3 kDa in molecular weight, and the His-tag in its C-terminus about 1 kDa in molecular weight, the band at 18 kDa may represents unprocessed AXCP which still carries the signal peptide and the His-tag. The SDS-PAGE showed similar profile when the same experiment was performed at different temperatures at 30 and 25 °C and different inducer concentration 0.1 mM (data not shown).

The same series of experiments were carried out using the plasmid pET-NH which was constructed to produce cytochrome c′ with the *E. coli* signal sequence but without the His-tag. SDS-PAGE showed a faint band at ~14 kDa (Lane 4 and 5 in figure 4.11) which could indicate the production of mature cytochrome c′ after 24 and 48 hours respectively. While no similar band was noticeable when induced by IPTG (Lane 2) nor in the negative control (Lane 3). At the same time, the band for unprocessed AXCP at ~ 18 kDa was evident as well (Lanes 2, 4 and 5).

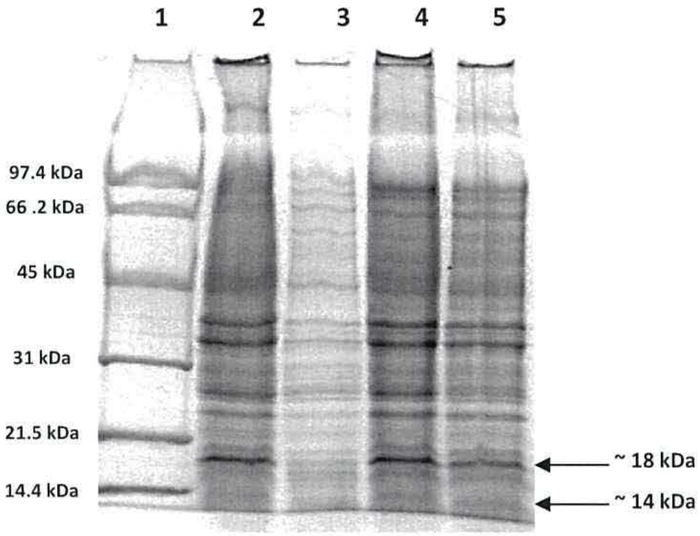


Figure 4.11: 12% (w/v) Acrylamide SDS-PAGE electrophoresis of recombinant cytochrome c′ overproduction in *E. coli* BL21 (DE3) strain in LB *medium* at 37 °C. Lane 1 is the molecular weight marker. Lane 2 is BL21 (DE3)-pEC86 transformed with pET-NH and induced for 4 hours with 0.25 mM IPTG. Lane 3 is the negative control where BL21 (DE3) strain transformed with pET 26b(+) and pEC86. Lanes 4 and 5 are BL21 (DE3)-pEC86 and pET-NH grown for 24 and 48 hours respectively without induction at. A band at ~18 kDa represents unprocessed AXCP and the band at ~14 kDa could be the mature AXCP.

The faint band at the correct molecular weight may suggest the production of processed and mature AXCP. In order to confirm the production of the mature and processed AXCP, and to check whether the protein is soluble or not, a larger scale experiment was performed.

The large scale experiment was done as follows. To scale-up the experiment to 500 mL culture a single colony from the freshly plated BL21 (DE3)-pEC86-pET-NH was picked. Transformed into three different 5 mL of LB medium containing Cm and Km. Media then incubated at 37 °C overnight with shaking at 180 rpm. They were used to inoculate three different 500 mL of LB culture in three different 2.5 L flasks containing 10 μg/mL hemin chloride and appropriate antibiotics. After the inoculation, the first flask was left to grow at 37 °C and 180 rpm. until the OD_{600} of 0.6 and then induced for 4 hours by adding 0.25 mM IPTG. The rest were left to grow for 24 and 48 hours with a continuous shaking at 180 rpm and 37 °C. Cells were then harvested by centrifugation for 10 min. at 8000 rpm and 4 °C. To prepare the crude extract, they were resuspended in 30 mM Tris-HCl buffer pH 8 and sonicated (see section 3.5.1.1b). The solution was centrifuged at 11500 rpm. for 45 min. at 4 °C and SDS-PAGE was run for the both the soluble fraction and the cell debris (Figure 4.12).

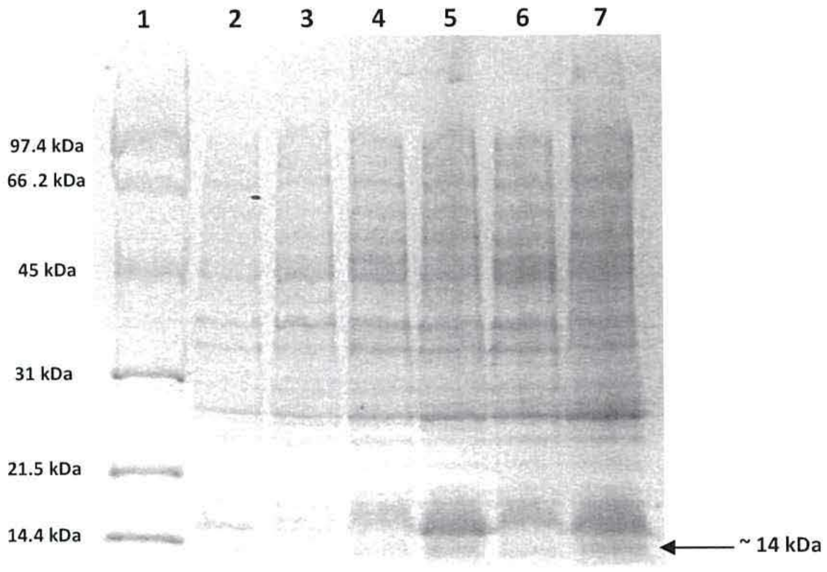


Figure 4.12: 12% (w/v) Acrylamide SDS-PAGE electrophoresis for BL21(DE3)-pEC86-pET-NH. Lane 1 is the molecular weight marker. Lane 2 is the soluble fraction and Lane 6 is cell debris of induced for 4 hours with 0.25 mM IPTG. Lane 3 is the soluble fraction and lane 6 is cell debris of 24 hours growth. Lane 4 and 7 are the soluble fraction and cell debris respectively of the 48 hours growth.

Cell debris fractions showed a band at ~ 14 kDa (Lanes 5, 6 and 7) while there was no band at that position for the soluble fraction. The same experiment was repeated at 30 and 25 °C, but unfortunately the production of soluble AXCP was not detected, even when the medium was changed to the rich nutrient content medium Terrific Broth (TB) at 37, 30 and 25 °C (data not shown). Hence, different methods of cell breakage were tried (see sections 5.5.1.1 and 3.5.1.2) to extract cytochrome c′ from the periplasmic membrane as a soluble protein. SDS-PAGE gels were run for both the soluble fraction and cell debris from all methods, and results showed no evidence for soluble AXCP (data not shown). This leads to the conclusion that although the

AXCP may have been produced in the cytoplasm, either it has not been transported to the periplasm for the incorporation of haem, hence the signal peptide is still attached, or the apo-cytochrome c′ was misfolded. In either case the protein will be insoluble.

4.3.2 Overproduction with pET-SIG-HS and pET-NHS:

These two recombinant plasmids were constructed to carry cytochrome c′ signal sequence with the His-tag in its C-terminus (pET-SIG-HS) and without any His-tag (pET-SIG-NHS). A similar series of experiments as described in the previous section was repeated using these two plasmids.

Bacterial strain BL21 (DE3)-pEC86 was transformed with 2 μL pET-HS, grown overnight on LB agar plate containing Cm (20μg/mL) and Km (30 μg/mL). For the small scale experiment, a single colony was transferred into three different liquid LB media containing the same amount of antibiotics and hemin chloride (10 μg/mL) and shaken at 180 rpm. at 37 °C. The first tube was induced for 4 hours with 0.25 mM IPTG after the OD_{600} reached 0.6. The second and the third one were left to produce AXCP without any induction for 24 and 48 hours respectively. The negative control was prepared by transforming BL21 (DE3)-pEC86 with pET 26b(+) for 48 hours. 1 mL of the inoculated growth *medium* was harvested from each tube after the periods indicated above. 200 μL SDS loading buffer was added to cells and heated for 5 minutes at 100 °C then centrifuged again at 13000 rpm for 2 minutes. 5 μL of the resulting supernatant was loaded into the polyacrylamide gel alongside with the molecular weight marker. The transformation and characterization experiments outlined above were repeated using pET-NHS plasmid.

Except for the negative control (lane 5), SDS-PAGE gels for the first set of experiments using pET-HS (Figure 4.13) exhibited a band at ~19kDa (lanes 2, 3, and 4) which indicates unprocessed AXCP, another band at ~ 15 kDa (lanes 2, 3, and 4) likely represents the processed and mature AXCP with the His-tag attached to its C-terminus.

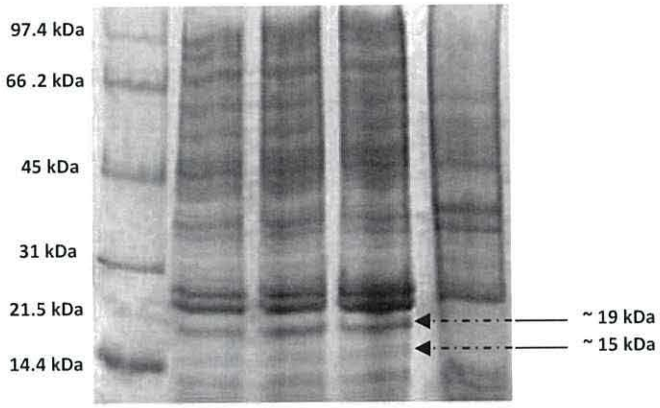


Figure 4.13: 12% (w/v) Acrylamide SDS-PAGE electrophoresis for BL21(DE3)-pEC86-pET-HS grown in LB *medium*. Lane 1 is the molecular weight marker. Lane 2 represents induced cells for 4 hours with 0.25 mM IPTG. Lane 3 and 4 are for the production of AXCP for 24 and 48 hours respectively and Lane 5 is the negative control which is BL21 (DE3)-pEC86-pET 26b(+).

The band at ~15 kDa was a promising indication for the production of mature AXCP. Thus the growth conditions were scaled-up to check whether the AXCP produced is soluble or not. Cells were grown and expressed in three different batches of 500 mL of LB medium containing the same amount of antibiotics and the source of haem. The first batch was induced for 4 hours using IPTG after the OD_{600} was reached 0.6 and batches 2 and 3 were grown for 24 and 48 hours respectively at 37 °C and 180 rpm and cells then harvested. The crude extracts were prepared as described in section (3.5.1). and the SDS-PAGE gel was run. Despite using all periplasmic extraction methods, no soluble AXCP was detected on the polyacrylamide gel (data not shown).

Changing the temperature to 30 and 25 °C and the expression medium from LB to TB did not make any difference in the production of the soluble form of AXCP (data not shown). This could be due to the misfolded AXCP arising from the presence of the His-tag in its C-terminus which was incorporated for the purification purposes.

The final attempt was the production of AXCP using its signal peptide but without the His-tag, assuming that the His-tag was the problem in the correct cytochrome c′ folding and hence the production of the soluble protein. The transformation and characterization experiments outlined above for pET-SIG-HS were repeated by using the pET-NHS plasmid. This plasmid was designed to produce AXCP with its own signal peptide but without the His-tag. SDS-PAGE gels for the small scale experiment showed the same profile as for figure 4.13. Changing the medium to the rich nutrient content medium TB has shown a different profile (Figure 4.14). The band at 14 kDa (Lane 3 Figure 4.14) results when cells are grown for 48 hours without using the inducer. This band is very likely to indicate the production of the mature and processed AXCP. Induction by IPTG and the expression for 24 hours without the inducer showed no band at 14 kDa. Bands at ~ 18.5 kDa in lanes 2 and 5 indicate the production of unprocessed AXCP.

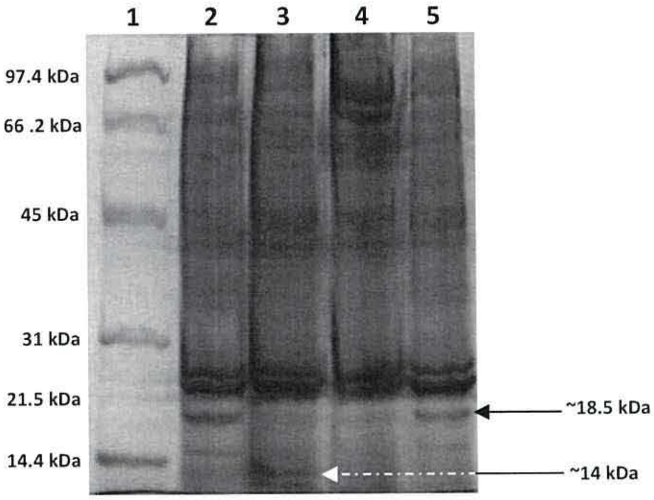


Figure 4.14: 12% (w/v) Acrylamide SDS-PAGE electrophoresis for BL21(DE3)-pEC86-pET-NHS grown in Terrific Broth *medium* (TB). Lane 1 is the molecular weight marker. Lane 2 represents expression for 24 hours. Lane 3 for 48 hours. Lane 4 is the negative control which is BL21 (DE3)-pEC86-pET 26b(+). Lane 5 is the induced cells for 4 hours with 0.25 mM IPTG. The band at ~ 14 kDa in lane 3 represents the mature and processed AXCP, and the band at 19 kDa represents unprocessed protein.

Again, the band at ~14 kDa in lane 3 (Figure 4.14) was an indication that AXCP had been produced. Again a large scale experiment was performed. In 500 mL TB containing proper antibiotics and the haem source, cells were induced for 4 hours with 0.25 mM IPTG (tube 1), expressed for 24 (tube 2) and 48 hours (tube 3) alongside with the negative control (tube 4). Cells were then harvested, sonicated, and centrifuged at 11,500 rpm for 45 min at 4 °C. Interestingly, the soluble fraction after 48 hours expression had a red coloured extract. This was the first red coloured extract seen so far. This was an additional indication, supporting the results of the SDS PAGE, and confirming the production of mature AXCP.

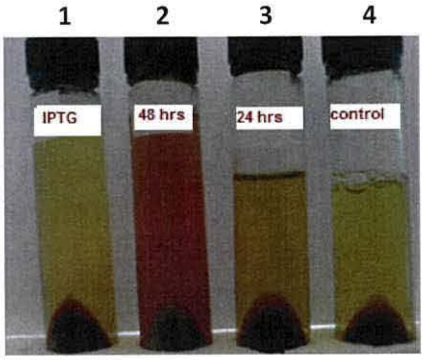


Figure 4.15: A picture shows the soluble fraction of the crude extract (supernatant) and cell debris after cell breaking using the sonicator. It is obvious from the red extract in (tube 2) which represents the expression and maturation of AXCP at 48 hours that AXCP was processed and matured successfully. Induction for 4 hours using IPTG (tube 1) and expression for 24 hours (tube 3) did not show the expected protein. Tube 1 represents the negative control.

Samples of the soluble fraction of were analysed by SDS-PAGE to double check the solubility of the recombinant AXCP. Indeed, the polyacrylamide gel for the red supernatant (figure 4.16) which is the crude extract of cells grown for 48 hours, showed an intense band at ~14 kDa (Lane 5, figure 4.16) while the growth for 24 hours showed a very faint band at the same position (Lane 3 figure 4.16), and induction by IPTG did not result in the production of the protein (lane 4) which is almost the same migration profile for the negative control (Lane 2).

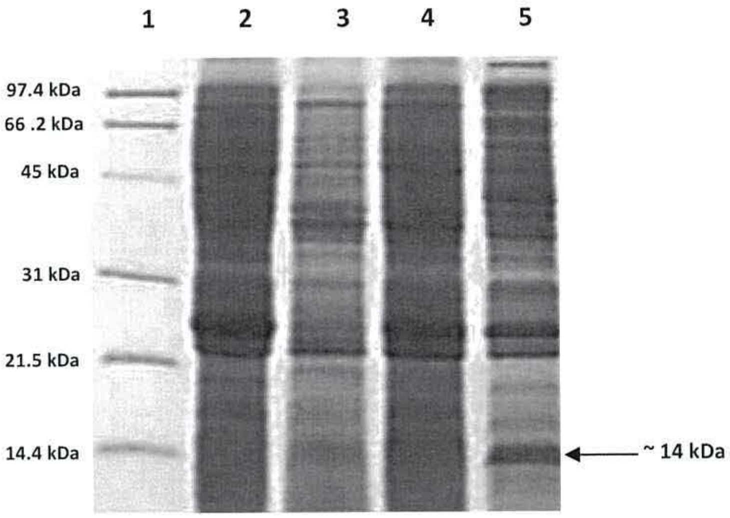


Figure 4.16: 12% polyacrylamide gel electrophoresis for the soluble fractions of the negative control (lane 2), growth and protein expression for 24 hours (lane 3) and for 48 hours (lane 5) and (lane 4) is for induction for 4 hours using 0.25 mM IPTG. Lane 1 is the molecular weight marker. The intense band at ~14 kDa in lane 5 is likely to be the processed and mature haem containing recombinant cytochrome c′.

From the above results it was concluded that the recombinant cytochrome c′ can be processed and matured with the correct incorporation of the haem to the apo-protein only when it produced with its own signal sequence and without any His-tag.

Finally, the preliminary expression conditions have been established. Further optimisation of conditions required expression of sufficient amounts of the recombinant cytochrome c′.

4.4 Overexpression conditions optimisation:

In order to extract the mature cytochrome c′ located in the periplasmic space, which is the space between the inner cytoplasmic membrane and external outer membrane, four different methods were compared. French press (kindly provided by Structural Biology Lab (SBL) at SRS – Daresbury) and sonication as mechanical lysis were used (see section 3.5.1.1). Osmotic shock and the BugBuster®master mix (Novagen) as a chemical lysis were used as well (see section 3.5.1.2).

Cells from the 48 hours growth were harvested by centrifugation at 8000 rpm for 10 minutes at 4 °C. They were subject to the lysis by methods described in sections 3.5.1. SDS-PAGE gels were run for the soluble fraction for each lysis method. They all show bands at the correct molecular weight position for AXCP at ~14 kDa as in (lane 5, figure 4.16) indicating the production of the mature recombinant AXCP. The UV-Vis spectra was recorded for the soluble fraction (Figure 4.17).

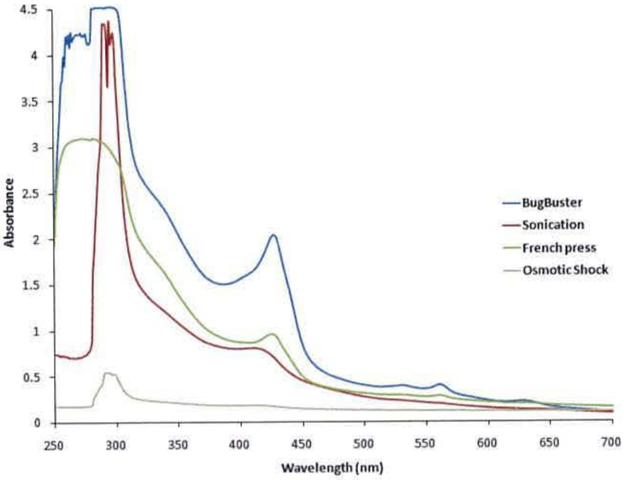


Figure 4.17: The electronic spectra of extracted unpurified cytochrome c′ using four different periplasmic extraction methods. BugBuster®master mix (Novagen) showed the highest level of the extraction, followed by the French press, Sonication and Osmotic shock.

The electronic spectra of unpurified protein (Figure 4.16) revealed that the extraction efficiency was in the order of BugBuster®master mix > French press > Sonication > Osmotic shock. Since the aim was to express cytochrome c′ in sufficient quantities for crystallisation and other studies, BugBuster®master mix was not the best choice due to its relatively high price. Osmotic shock was not a good choice due to its inefficiency. The French press was not a suitable option, as it was not available at Bangor University, therefore sonication was chosen as a cell disruption and protein extraction method in this work.

Further optimisation choices were the duration of the expression and the addition of extra nutrients, like metal ions at low concentrations. The optimum condition was described in section 3.4. Parameters that were varied the optimisation experiments are summarised in Table 4.1. In all methods, recombinant AXCP was purified as described in (sections 3.5.2 and 3.5.3). Cytochrome c′ concentrations were calculated by applying Beer-Lambert law using the molar extinction coefficient of 80 mM^{-1}. cm^{-1} for the Soret maxima band.[10]

Media	Additives	Temperature (°C)	Duration (h)	Yield (mg/L)
LB	IPTG	37, 30, 25	4	None
	IPTG + Hemin	37, 30, 25	4	None
	Hemin	37, 30, 25	24, 48	None
	Hemin + metal ion master mix	37, 30, 25	24, 48	None
TB	IPTG	37, 30, 25	4	None
	IPTG + Hemin	37, 30, 25	4	None
	Hemin	37	24	5
			48	23
			72	36
			96	40
		30	24	4
			48	12
			72	18
			96	24
	Hemin + metal ion master mix	37	24	6
			48	28
			72	43
			96	62
		30	24	5
			48	28
			72	36
			96	45

Table 4.1: Optimisation conditions for the production of recombinant AXCP. No AXCP was produced at any conditions in LB medium. The highest yield was in TB medium when the source of haem as (hemin) and the metal ion master mix were added and cells grown at 37 °C. The agitation was 180 rpm. in all conditions. Protein concentrations were calculated by applying Beer-Lamber law using the extinction coefficient of 80 mM^{-1} cm^{-1} at 400 nm. Metal ion master mix is a mixture of metal ions added to the media after 24 hours of growth, metal ions and final concentrations were: Ni^{2+} 2x10^{-3} mM, Co^{2+} 2x10^{-3} mM, Zn^{2+} 1x10^{-2} mM Mn^{2+} 1x10^{-2} mM and Fe^{3+} 5x10^{-2} mM.

Table 4.1 shows that recombinant AXCP cannot be expressed in LB medium even after the addition of metal master mix at different temperatures. Changing the growth medium from LB to a highly enriched TB yielded a reasonable amount of recombinant AXCP (see section 3.3.1 for the media compositions) (Figure 4.18). TB contains an extra amount of peptone and yeast extract which provide necessary nutrients and cofactors for growth of recombinant *E. coli* strains. The yeast extract concentration is increased to elevate cell yields. Potassium phosphate buffer prevented the drop in pH (and hence cell death) and provide potassium for cellular systems. Glycerol acts as a carbohydrate source and, unlike glucose, it is not fermented to acetic acid.

The addition of Ni^{2+} $2x10^{-3}$ mM, Co^{2+} $2x10^{-3}$ mM, Zn^{2+} $1x10^{-2}$ mM Mn^{2+} $1x10^{-2}$ mM and Fe^{3+} $5x10^{-2}$ mM increased the production of AXCP by about two fold, owing to the role of divalent metal ions in the promotion of enzymatic reactions[11] and hence the increase in the protein yield.

This expression method was used to overproduce mutants generated in this work. In all cases the yield was sufficiently high enough for the structural and other studies.

Figure 4.18: Recombinant AXCP crude extract and cell debris of the 96 hours growth in TB (a and b) and in LB (c and d). The protein has not been produced in LB medium, while in the TB it is obvious the AXCP was mass-produced.

4.5 Cytochrome c′ purification:

Growing x-ray quality crystals for the structural studies necessitated the use of highly purified proteins.[12] Recombinant AXCP and all other mutants were purified using the method described in (section 3.5.3). Since the AXCP could not be produced with the Hig-tag, and because the calculated isoelectric point of AXCP is 9.21, [13] cation exchange chromatography was chosen. After harvesting cells from the growth medium by centrifugation at 800 rpm for 10 min. at 4 °C, the crude extract was obtained using sonication (section 3.5.1.1b), and cell debris were removed by centrifugation at 11500 rpm for 45 min. at 4 °C. The soluble fraction was dialysed against 5000x deionised water (section 3.5.2) for 48 hours using dialysis tube of MWCO 10 kDa to remove the buffer salts, low molecular weight proteins and any extra hemin chloride that remained in the crude extract. Dialysed protein was then centrifuged at 20000 rpm for 1 hour and the supernatant was loaded onto the carboxymethyl cellulose CM-52 (Sigma-Aldrich). The resin was activated at pH 6 and the protein was purified as described in section 3.5.3. Although it was relatively slow, the gravity flow of the loaded supernatant was the most efficient way for the cation exchange. Negatively charged cytochrome c′ (pI 9.21) binds to positively charged carboxymethyl cellulose and appears as a red to reddish-brown band depending on the type of the mutant. The column was then washed with deionised water and then with buffer and the pure protein was eluted by a salt gradient from 50-200 mM NaCl (Figure 4.19).

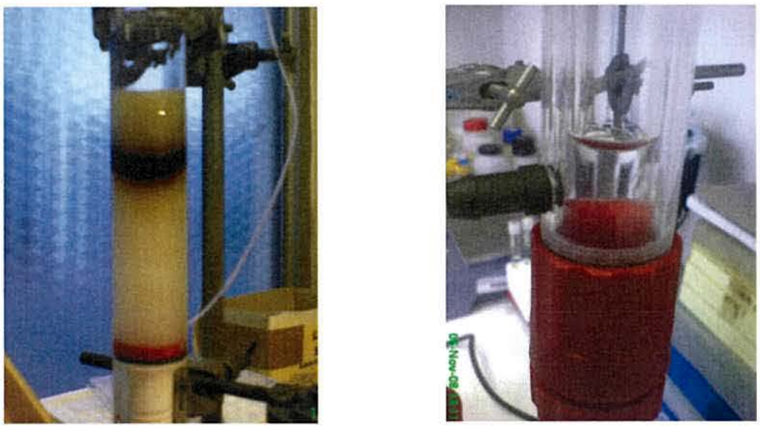


Figure 4.19: Purification of recombinant L16G mutant using CM-Cellulose cation exchange chromatography. Negatively charged protein binds to positively charged CM-cellulose and appears as red or reddish-brown (depending on the mutant type) circle bound to the stationary phase of the column.

Fractions were collected and the purity was checked by SDS-PAGE (Figure 4.20). The intense band at ~14 kDa in lane 2 revealed that the AXCP was overproduced although there are still other lower and higher molecular weight bands observed which belong to other proteins produced at the same time, while lane 3 shows only one intense band at ~14 kDa while other bands have been disappeared owing to pure AXCP. Lane 4 showed same bands as in lane 2 except the intense band at ~14 kDa. A faint band at ~14 kDa can be seen in lane 4 which may be due to the production of other proteins with the same molecular weight as AXCP.

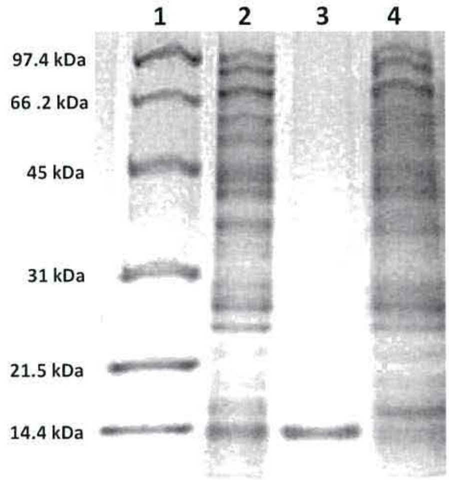


Figure 4.20: 12% SDS-PAGE for the dialysed crude extract (lane 2), pure AXCP from carboxyl cellulose column (lane 3) and the effluent (lane 4) compared to the molecular weight marker (lane 1). The crude extract showed an intense band at ~14 kDa referred to overexpressed recombinant AXCP with many other bands at higher and lower molecular weights. Purified protein in lane 2 showed only one band at ~14 kDa revealed to highly purified AXCP. The effluent lane 3 showed all bands present in lane 2 and a faint band at ~14 kDa which may be belongs to other proteins which have about the same molecular weight.

The purity index was measured by calculating the ratio of the Soret band maximum absorbance of the recombinant or mutants to the maximum absorbance of the protein band at 280 nm[14] and was 2.86 for recombinant AXCP (Figure 4.21).

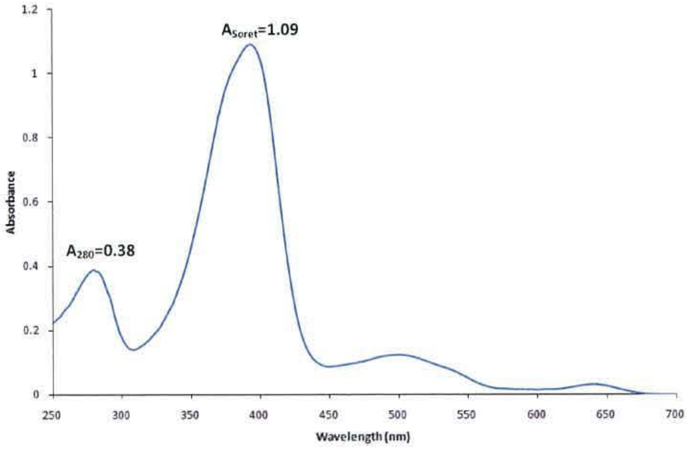


Figure 4.21: The electronic spectrum of ferric recombinant cytochrome c′ showing an intense band at the Soret maxima at 395 nm and a lower band at 280 nm due to protein. The ratio between these two bands gives the purity index of the purified AXCP which is equals to 2.86.

4.6 Generation and overproduction of mutants:

Residues close to the iron haem were selected (see section 5.1). They all have been mutated using site directed mutagenesis and pET-NHS plasmid as a template DNA vector. Mutations have been confirmed by sequencing the resultant vectors. Mutants were then overproduced, purified and concentrated to the desired concentration in the same way as for recombinant native AXCP (Figure 4.22).

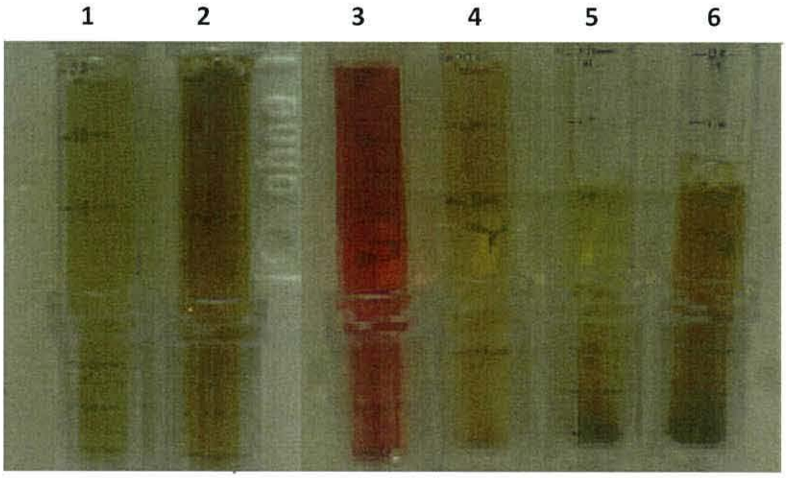


Figure 4.22: A figure shows the purified cytochrome c′ mutants. 1: W56G/L16F, 2:W56Y, 3: L16G, 4: L16Y, 5: W56G, and 6: L16F..

The mutant plasmids of H120A, H120F and H120N were generated successfully and their correct sequences were confirmed by sequencing (see section 3.2.10). The SDS-PAGE analysis for the whole cells has confirmed the production of mutants (Figure 4.23). But despite of many attempts, it was not possible to purify them using the purification methods described in this work or by using ammonium sulphate fractionation.

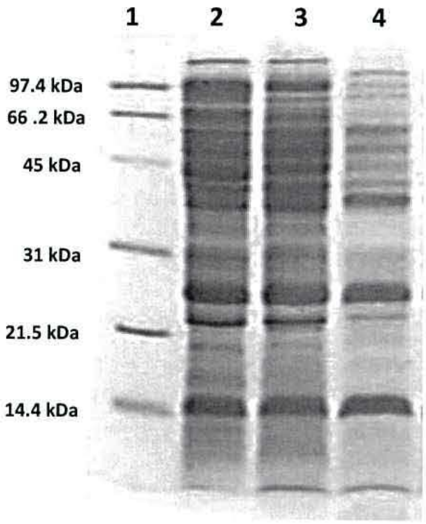


Figure 4.23: 12% SDS-PAGE showing the production of H120A (lane 2), H120F (lane 3) and H120N (lane 4) run alongside with the molecular weight marker (lane 1). However, the gel shows that they have been overexpressed but they could not get purified.

The purity index and the final yield of recombinant and mutants are summarised in Table 4.2.

Protein	Yield (mg/$L_{culture}$)	Purity index (A_{Soret}/A_{280})
Recombinant native	62	2.86
L16F	60.3	2.5
L16G	54.5	6.65
L16Y	54.6	2.72
W56G	74.3	2.80
W56Y	68.8	2.48
L16F/W56G	84.2	3.16

Table 4.2: The final yield in mg/liter culture and the purity index which is the Soret band maximum absorbance of the recombinant or mutants to the maximum absorbance of the protein band at 280 nm. They all have been produced in large quantities not less than 54.6 mg/L and are highly purified with the purity index not less than 2.48.

4.7 Conclusions:

Growing the x-ray quality crystals and the spectroscopic studies require the production of highly purified protein with sufficient quantities. In this work, satisfactory quantities of mature recombinant native and mutant holocytochrome c′ from *A. xylosoxidans* were produced with high purity index (A_{Soret}/A_{280}) not less than 2.48. Four different plasmids were constructed aiming the production of highly pure AXCP in sufficient quantities. The successful overproduction was achieved by using pET-NHS plasmid. In this plasmid the AXCP gene was cloned between the two restrictions sites NdeI and BamHI of the pET 26b(+) vector to produce AXCP with its signal peptide and without any His-tag. From the published literature the highest level of AXCP reported so far is 20-30 mg/L culture[15] while in this work 54 – 84 mg/L culture was produced.

The production of mature cytochromes c was improved by the use of plasmid pEC86.[8] This plasmid expresses the *E. coli* cytochrome c maturation genes CcmABCDEFGH constitutively. Therefore, we did not attempt to produce AXCP without the use of pEC86 plasmid.

The overproduction of large amounts of the native and mutant cytochromes c′c′ was highly depending on the type of the growth media, induction time, presence of the haem source. Using LB medium did not show the production of AXCP, while the nutrients enrich medium TB has shown the production of significant amounts.

However the outer membrane of Gram-negative bacteria is not permeable to hemin,[16] addition of 10μg/mL hemin chloride as the haem source has significantly increased the amount of AXCP production. Furthermore, addition of metal ion master mix increased the quantities the product. This may be due to the promotion in the enzymatic activity by these metal ions.[11]

Induction for a period of 4 hours by IPTG did not result in the production of the AXCP. Hence, the auto-induction was employed; its duration seems to have a significant effect on protein mass production. 12 hours of auto-induction has not caused the production of the mature AXCP, while longer periods revealed the maturation of the apo-protein. The longer the induction duration the higher the final yield was. This may be due to the time required for the processing and maturation of the protein. The higher yield at longer induction duration was linked to the production of NO[9] by bacteria as each cell of *E. coli* generates approximately one molecule of NO in one minute[17] and cells then take benefit from the NO.

As mentioned above, four different plasmids were constructed (see section 3.2.2) and they have been co-transformed with pEC86. The following is the overall conclusion for each mutant:

Production of AXCP with the *pelB* leader of the pET 26b(+) in the N-terminus with the His-tag (pET-H) and without the His-tag (pET-NH) at its C-terminus: In both cases, SDS-PAGE analysis showed an intense band at ~18 kDa indicating that the produced AXCP was mostly unprocessed owing to inefficient cleavage of the *pelB* by the signal peptidase. A band at the correct molecular weight at 14 kDa was present in cell debris only while it was not detectable in the soluble part of the crude extract. This may be due to the failure of the haem insertion into the apo-cytochrome c′, since this insertion is essential for c-type cytochrome folding[18]. Attempts to slow down the overproduction rate by decreasing the growth temperature and addition of an extra amount of the haem source to the growth medium showed no significant change. Depending on that, we conclude that *pelB* leader which is the *E. coli* signal sequence, may not direct AXCP to the periplasm or, however, even if it could direct it, the signal peptide remained uncleaved due to the inability of signal peptidase to cleave it. This will leave a large fraction of the protein unprocessed.

- Replacing the *pelB* leader by the AXCP signal sequence without the His-tag (pET-NHS) led to the successful production of mature holocytochrome c′only after 48 hours. Incorporation of the His-tag (pET-HS) for the purification purposes prevented the formation of soluble protein.
- pET-NHS was used as a template for the generation of different mutant plasmids; L16F, L16G, L16Y, W56G, W56Y, L16F/W56G, H120A, H120F, and H120N. Except for the H120 mutation series, all othet mutants have been successfully overproduced and purified.
- However the SDS-PAGE analysis of H120 mutation series revealed the production of these mutants, but they could not be purified. Since the covalent attachment of the haem to the apo-cytochrome c is necessary for the correct apo-protein folding, the apo-protein of these mutants may have been overproduced but

remained unfolded. Further investigations and using other purification methods are desired to get these interesting mutants matured and purified.

4.8 References:

---

[1] Esposito D., Garvey L. A., Chakiath C. S (2009) "Gateway cloning for protein expression" *Methods in Molecular Biology* **498**, 31-54.

[2] Thöny-Meyer L. (2002) "cytochrome c maturation: a coplex pathway for a simple task?" *Biochemical Society Transactions* **30,** 633-638.

[3] Stevens J. M., Uchida T., Datrop O., and Ferguson S. J. (2005) "Covalent cofactor attachment to proteins: cytochrome c biogenesis" *Biochemical Society Transactions* **33,** 792-795.

[4] Stevens J. M., Daltrop O., Aleen J. W. A., Ferguson S. J. (2004) "c-type cytochrome formation: chemical and biological enigmas" *Accounts of Chemical Research* **12,** 999-1007.

[5] Enggist E., Thöny-Meyer L., Guntert P., and Pervushin K. (2002) "NMR structure of the haem chaperon CcmE reveals a novel function motif" *Structure* **10,** 1551-1557.

[6] Thöny-Meyer L., Fischer F., Künzler P., Ritz D., and Hennecke H. (1995) "*Escherichia coli* genes required for cytochrome c maturation" *Journal of Bacteriology* **177**, 4321-4326.

[7] Coleman W., and Tsongalis (2006) "Molecular diagnostics for the clinical laboratorian" Humana Press Inc., Totowa, New Jersey, pp.48-51.

[8] Arslan E., Schulz H., Zufferey R., Künzler P., and Thöny-Meyer L. (1998) "Overproduction of the *Bradyrhizobium japonicum* c-type cytochrome subunits of the ccb_3 oxidase in *Escherichia coli*" *Biochemical and Biophysical research communications* **251**, 744-747.

[9] Barbieri S. (2007) "Mutagenesis, spectroscopy, and crystallography of recombinant cytochrome c′ from *A. xylosoxidans*" *PhD thesis,* University of Wales, Bangor.

[10] Cusanovich M.A., Tedro S. M., and Mamen M. (1970) "*Pseudomonas denitrificans* cytochrome cc′1" *Archives of Biochemistry and Biophysics* **141,** 557-570.

[11] Mordasini T., Curioni A., and Andreoni W. (2003) "Why do divalent metal ions either promote or inhibit enzymatic reactions" *The Journal of Biological Chemistry* **278,** 4381-4384.

[12] Asherie N., Ginsberg C., Greenbaum A., Blass S., and Knafo S. (2008) "Effect of protein purity and precipitant stereochemistry on the crystallization of Thaumatin" *Cystal Growth and Design* **8,** 4200-4207.

[13] Iwasaki H., Shidara S., Sato H., Yoshimura T., Suzuki S., and Nakahara A. (1986) "Cytochrome c′ isolated from *Achromobacter xylosoxidans* GIFU 1055" *Plant Cell Physiology* **27,** 733-736.

[14] Suquet C., Savenkova M., and Satterlee J.D. (2005) "Recombinant PAS-heme domains of oxygen sensing proteins: High level production and physical characterization" *Protein Expression and Purification* **42,** 182-193.

[15] Barbieri S., Murphy L. M., Sawers R. G., Eady R. R., Hasnain S. S. (2008) "Modulation of NO binding to cytochrome c′ by distal and proximal haem pocket residues" *Journal of Biological Inorganic Chemistry* **13,** 531-540.

[16] Stojiljkovic I., Hantke K. (1994) "Transport of haemin across the cytoplasmic membrane through a haemin-specific periplasmic binding-protein-dependent transport system in *Yersinia enterocolitic*" *Molecular Biology* **13,** 719-732.

---

[17] Nioche P., Berka V., Vipond J., Minton N., Tsai A., and Raman C. S. (2004) "Femtomolar sensitivity of a NO sensor from *Clostridium botulinum" Science* **306,** 1550-1553.

[18] Mavridou D. A. I., Braun M., Thöny-Meyer L., Stevens J. M., Ferguson S. J. (2008) "Avoidance of the cytochrome c biogenesis system by periplasmic CXXCH motifs" *Biochemical Society Transactions* **36,** 1124 – 1128.

Chapter 5: Electronic Spectroscopy studies of recombinant cytochrome c' and mutants

5.1 Introduction:

Cytochrome c' from *Alcaligenes xylosoxidans* (*AXCP*) like other members of class II cytochromes has novel spectroscopic and ligand binding properties. It is characterised by a split Soret band. *AXCP* forms stable complexes with carbon monoxide (CO) and nitric oxide (NO) with the CO complex being hexa-coordinate (6c) while it forms a penta-coordinate (5c) complex with NO. However it discriminates against dioxygen (O_2) by its inability to form a stable complex with it. Kinetic studies show that NO reacts with ferrous state of *AXCP* initially to form a transient 6c-NO-His adduct which then reacts in an NO concentration dependent manner to cleave His-Fe bond to form a stable 5c-NO adduct.[1]

Another novel property of *AXCP* is the electronic state of iron. In contrast to most other known haem proteins, where iron is either pure low spin (S= 1/2) or high spin (HS) (S=5/2) species, in *AXCP* iron has been shown to be in a quantum-mechanical admixture (QS) of intermediate spin (S=3/2) and high spin (S=5/2) states where the contribution of the intermediate spin state is 10-15%.[2,3] The spin state equilibrium between HS and QS is pH dependent, with HS species predominating at high pH. This is rationalised as being due to the delocalisation of the electrons of the iron towards the coordinated ligand induced by protonation/deprotonation or hydrogen bonding of Nε1 histidine axial ligand[4] or by the electron-withdrawing effect[5] of the guanidinium group of arginine 124. In both cases the iron axial ligand weakens and leads to HS predominating.

In this work, residues close to the haem iron pocket have been mutated to different residues which have been selected with the aim of understanding their role in ligand binding/discrimination and in perturbing the haem iron spin state. Targeted residues were leucine 16 (L16), tryptophan 56 (W56) and the axial ligand histidine 120 (H120) (Figure 5.1). The mutation strategies have been presented in (Figure 5.2).

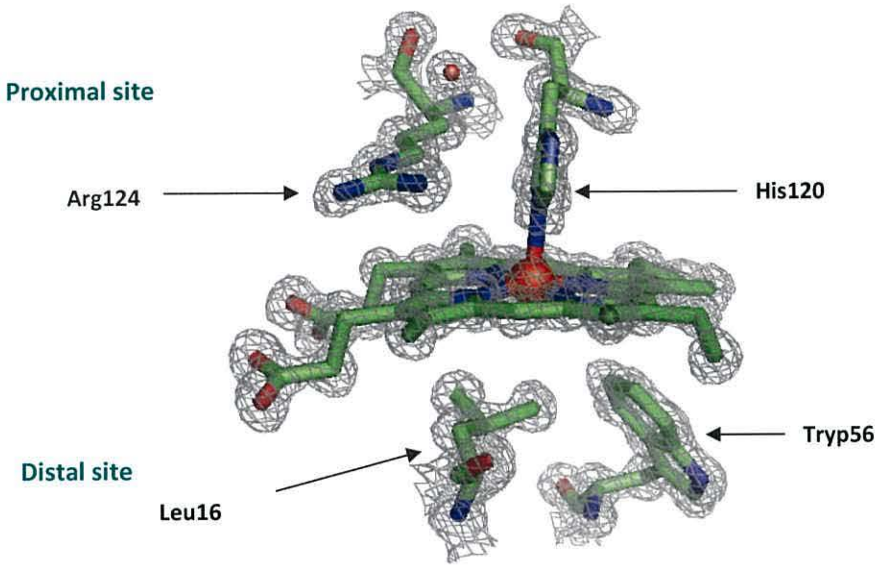


Figure 5.1: Electron density of recombinant cytochrome c' from *Alcaligenes xylosoxidans* at 1.15 Å resolution showing the key amino acid residues which have been mutated in the work. Distal pocket residue leucine (Leu16) potentially provides steric hindrance to ligand binding to the distal face of the haem, tryptophan (Tryp56) located in the distal pocket of the haem may provide the steric hindrance to some extent and at the same time blocks a solvent channel preventing the direct access to the haem. Proximal pocket residues histidine (His120) is bound to the iron's fifth coordination position and arginine (Arg124) sit above the porphyrin ring in parallel manner to the ring at the proteins ferric state.

The distal and proximal haem environments of *AXCP* have very different structural features. The distal haem pocket is deeply buried within the protein with no direct access to solvent. Leucine 16, which crowds the distal face of haem, is proposed to be responsible for the discrimination of the ligand.[6] Furthermore, its steric hindrance is anticipated to have a role in destabilising the 6c-NO-His adduct.[6] In contrast, the proximal face is much closer to the protein surface, with the proximal His 120 ligand being solvent-exposed.

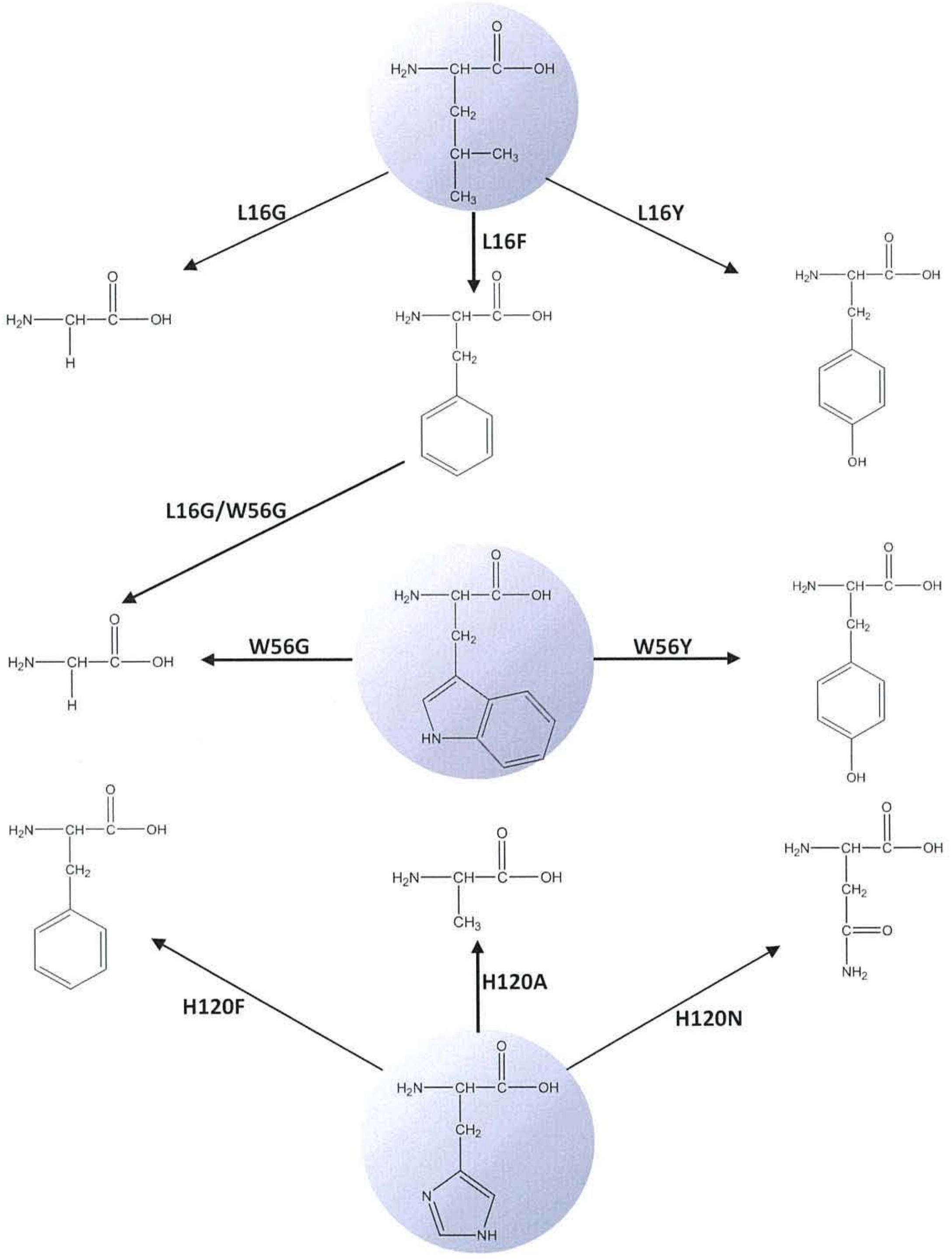


Figure 5.2: The mutation strategy. leucine 16 (L16) has been mutated to glycine (L16G), tyrosine (L16Y), phenylalanine (L16F) and tryptophan 56 (W56) to glycine (W56G) and tyrosine (W56Y), histidine 120 (H120) to alanine (H120A), phenylalanine (H120F) and to asparagine (H120N), and (L16F/W56G) is the double mutant where L16 mutated to F and W56 to G at the same time.

The solvent channel observed in type 1 cytochromes c' such as *C. vinosum and R. capsulatus* is the shortest route from the protein surface to the 6th coordinate position at the distal pocket of haem. In *AXCP*, the channel is blocked by the side chain of tryptophan 56 (Figure 5.3).

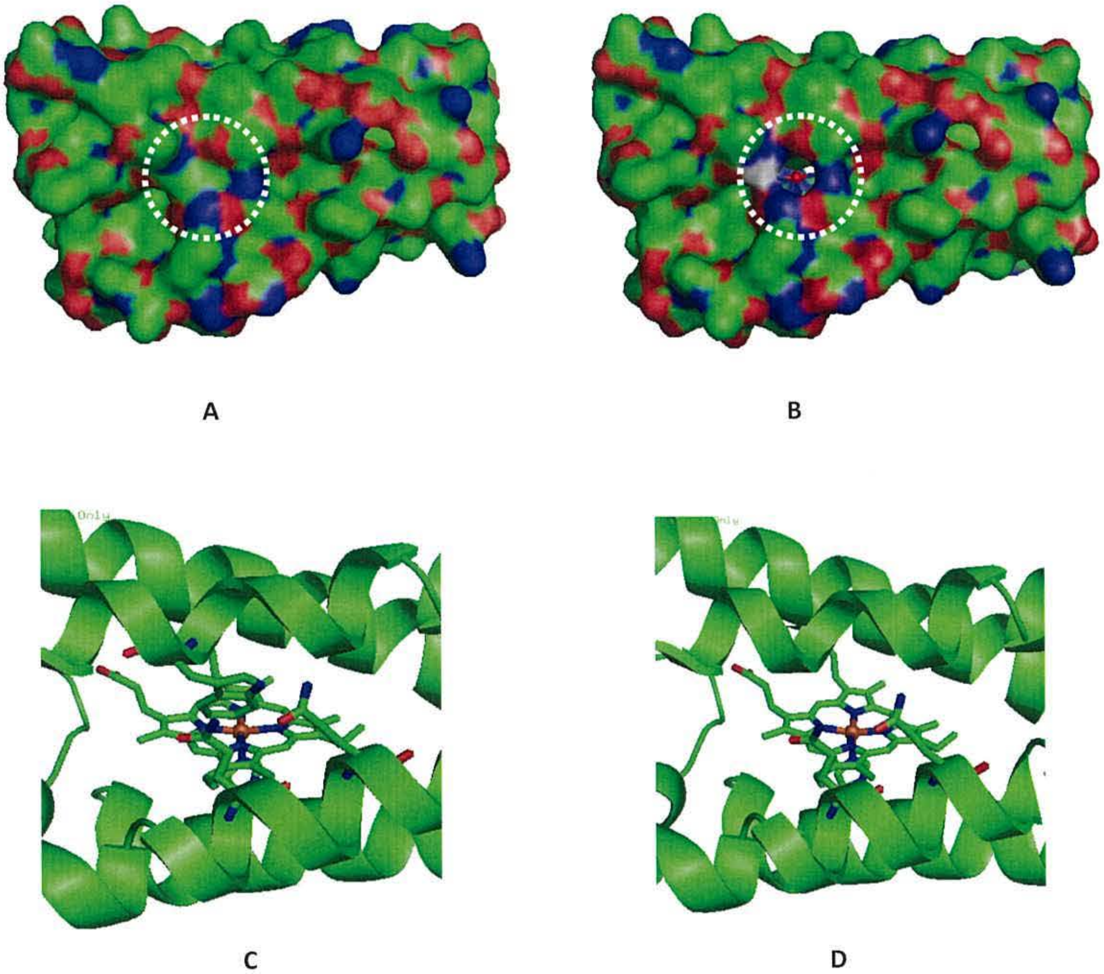


Figure 5.3: Cytochrome c' van der Waals surfaces of (A) recombinant and (B) proposed W56G mutant show that replacing tryptophan 56 by glycine (circled) will unblock the solvent channel hence the haem iron will be exposed to the solvent and this may alter the ligand binding affinity. For clarity, (C) recombinant and (D) W56G mutant have been represented by stick and cartoon style. Image prepared with PyMOL using the crystal structure of recombinant *AXCP* solved in this thesis at 1.15Å.

Replacing W56 with glycine is hypothesized to unblock the channel and make the haem iron's sixth coordinate site to be more exposed to the solvent. Residues glutamine 83 and glutamine 87 near to the protein surface may also block access to this channel. In this work W56 has been mutated to glycine (W56G) to determine whether there is a clear change in access to the distal face of haem, as well as to the discrimination between NO and CO. Upon binding of O_2, CO and NO, tyrosine 70 in the signal transducer sensor protein HemAT (haem-based aerotactic transducer) was found to play the major role toward ligand recognition and discrimination.[7] Tyrosine 70 in HemAT is equivalent to W56 in *AXCP;* therefore, W56 was mutated to tyrosine to find out whether tyrosine will play any role in ligand binding and discrimination.

L16, W56 and His120 were successfully mutated to appropriate residues (section 3.2.10) and the proteins were overexpressed as described in (section 3.4).

Except for the H120 series, other mutants were successfully overproduced, isolated and purified. However, the soluble fraction of the crude extract showed a band corresponding to the H120 mutant series at 14 kDa on SDS-PAGE, but they could not be purified (section 4.6). On top of the purification method which was carried out successfully for all other mutants (section 3.5), different purification methods have been tried, but were unsuccessful.

In this work, UV-Vis spectroscopic studies have been implemented to get an insight into the effect of the key residues on the electronic spin state and their role in controlling ligand binding and discrimination.

5.1.1 Generating ferrous forms:

A novel method for generating *AXCP* ferrous forms has been described here. This method showed its reliability in the reduction of haem iron of *AXCP* anaerobically without using an anaerobic chamber and desalting columns to remove excess reductant from the reaction mixture. This method can be used for reduction of other metalloproteins.

Recombinant native *AXCP* and all other mutants were reduced as described in (section 3.7.1). The reduction process was monitored spectrophotometrically by measuring the UV-Vis spectra in the range between 250nm -700nm after each addition of the reductant; sodium dithionite (DT). Figure 5.4 shows the reduction process of the recombinant *AXCP* as an example. This method was applied to all other mutants.

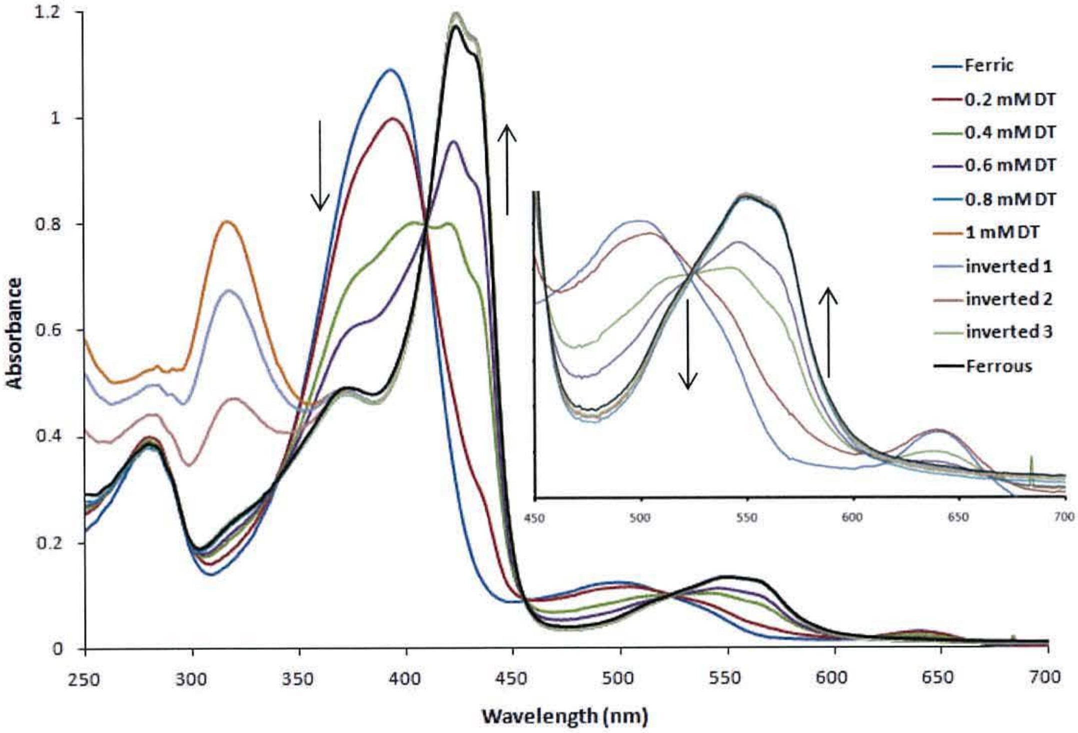


Figure 5.4: Reduction process of 13.5 μM recombinant native *AXCP* at pH 6. DT is the reductant; sodium dithionite. Inverted 1, 2, and 3 represents the inversion of the cuvette once a time so the reaction mixture contacts the rubber septa and DT reacts with the oxygen on the rubber to decompose excess DT.

The reduction was achieved anaerobically in a quartz cuvette super sealed with rubber septa as described in (section 3.7.1). The reduction caused a change in the protein colour from almost colourless (at concentration of 13.0 μM) to a faint yellow colour. The excess amount of DT was monitored by appearance of a peak at 317 nm. If the reduced form of the *AXCP* was required for ligand binding discrimination studies, excess DT was removed from the reaction mixture by inverting the cuvette carefully so the reaction mixture touches the surface of rubber septum. After each inversion, the level of DT was monitored by measuring the electronic spectra. A decrease in a 317 nm band was observed revealing the decrease in DT level in the reaction mixture. DT has a strong reduction property and decomposes rapidly in aqueous media under oxygen consumption as in the following equation.

$$Na_2S_2O_4 + O_2 + H_2O \rightarrow NaHSO_4 + NaHSO_3$$

This will allow any excess of DT to react rapidly with the oxygen present on the rubber surface.

Addition of 0.2 μM DT (Figure 5.4) decreased the Soret band intensity without change in the peak position. The charge transfer band (CT) at 640 nm which was reported to be sensitive to spin state[8] and irons coordination number[5] remained the same even in its intensity. This peak particularly was only seen in high-spin[8] five coordinate[5] c-type haem proteins. At a concentration of 0.4 μM DT, most features of both ferric and ferrous *AXCP* were observed; a shift of the Soret band from 394 to 401 nm, 380 nm (sh) has become more predominant, a peak at 420 nm has appeared with more well defined peak at 434 nm (sh). The β peak has shifted to 541 nm and in the signal intensity of the CT peak at 640 nm decreased. These revealed the presence of both ferric and ferrous species at the same time. Ferrous features were predominant at the concentration of 0.6 μM DT. At concentrations of 0.8 and 1.0 μM DT only features related to the ferrous form were observed.[9] The characteristic feature, a split Soret band with a peak at 424 nm and another at a lower intensity at 434 nm are attributed to a nondegenerate transition arising from electronic nonequivalency in the x and y directions of the porphyrin plane[10] with a shoulder at lower wavelength at 373 nm. The α/β beak is now centred at 550 nm and 565 nm, finally the CT band at 640 nm has vanished. The Resonance Raman studies of *AXCP* have proved that the ferrous form is in the 5c high spin state.[11]

The ferric species of recombinant *AXCP* showed an electronic absorption spectrum that was identical to previously reported spectra for wild-type *AXCP* indicating the integrity of the haem environment in the synthesized protein. The ferrous species also displayed an identical spectrum to previously published spectra[12] confirming the success of the method we followed for the reduction.

5.1.2 The spectral properties of ferric and ferrous AXCP mutants:

The spectral properties of ferric and ferrous forms of the recombinant *AXCP* were discussed in the previous section as part of the reduction process (figure 5.3).

5.1.2.1 Spectral properties of L16G:

Replacing leucine by glycine (L16G) displayed a dramatic change in the electronic absorption spectra (Figure 5.5) if compared to the recombinant *AXCP* absorption spectra (Figure 5.4). Once the L16G protein was isolated and purified, an absorption spectrum was recorded. Interestingly, it possessed a narrow and intense absorption maximum at 418 nm and a shoulder at 397 nm. This is a red shift by 24 nm with respect to the Soret band of the recombinant *AXCP*. Bands of α/β region were at 536 and 565 nm, a red shift by

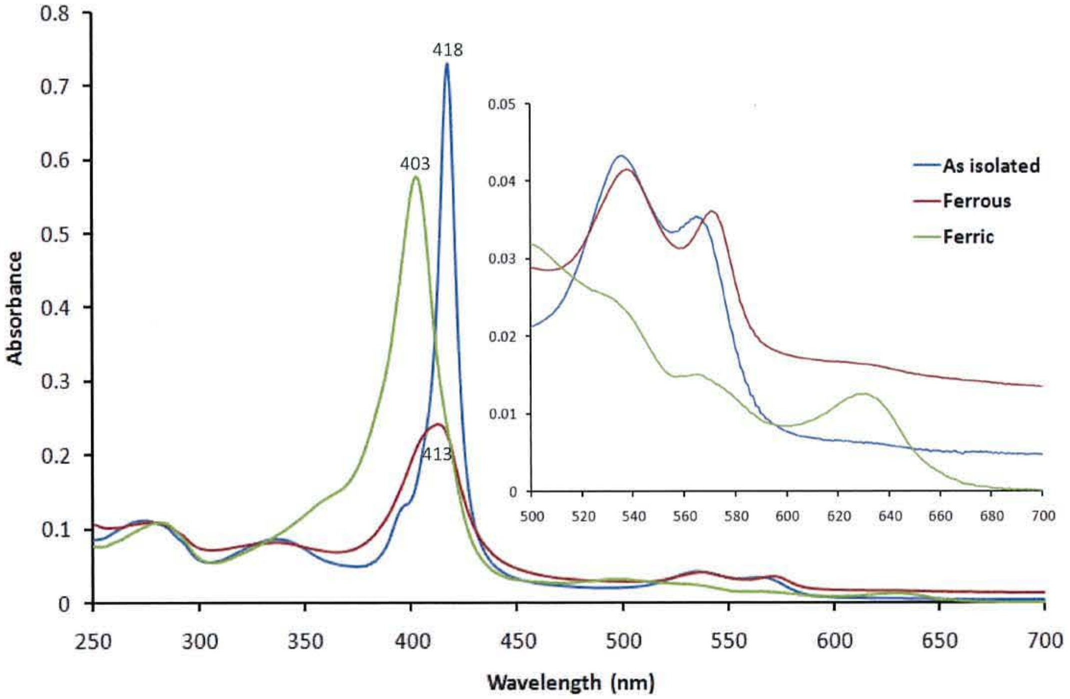


Figure 5.5: Electronic spectra of reduced, oxidised and as isolated L16G mutant (4.0 µM) at pH 6. Calculated extinction coefficients are 185, 141, and 62.9 $mM^{-1}cm^{-1}$ for as isolated, ferric and ferrous proteins respectively

40 and 30 nm compared to the equivalent region of the spectrum of recombinant *AXCP* respectively. The Soret band at 418 nm and 397 nm (sh) of this mutant is identical to that reported previously for 6c-CO-ferrous *AXCP* [11] may signify that the haem iron is hexa-coordinate (6c). The crystal structure of this mutant which has been solved in this thesis at resolution of 1.18 Å confirmed that L16G is 6c with a water molecule bound to the distal site of the haem group and Hist 120 at the proximal site of the haem iron, and the iron is in the reduced form (see figure 6.15 in chapter 6).

In order to better understand the effect of replacing leucine by glycine on the ligand binding and discrimination, and to observe the oxidised and the 5c-reduced form, any bound ligand has to be removed (see section 3.7.1).

The ferric form exhibited a Soret band at 403 nm and 364 (sh) that is shifted to a longer wavelength by 7 and 16 nm compared to the recombinant *AXCP* respectively. Two distinct peaks in α/β region at 496 and 536 nm were not significantly different from those of the recombinant while a peak at 571 nm was observed - which is unique for this mutant and has not been seen among other mutants in this and any previously published work.[13] It was shown that for the 6c haemproteins containing an imidazole as the fifth ligand, the CT ranges from 600-637 nm.[14] Depending on those spectral observations, L16G is expected to be 6c.

The ferrous form's spectrum displayed a single Soret band at 414 nm with a shoulder at 336 nm, which is again is unique for this mutant, as all other mutants in this work and previous published literature exhibit a split Soret band in their ferrous forms.[12,13] The α/β region showed two bands at 538 and 571 nm, while a band at 496 nm which was seen in the ferric form has vanished.

5.1.2.2 Spectral properties of L16F:

The haem iron of the L16F mutant was already in the oxidised state when isolated and purified. It was reduced as described previously in section 5.1.1. The electronic spectrum (Figure 5.6) shows the oxidised and reduced forms of the mutant. The Soret band at 401 nm displayed by the oxidised form shifted to a split Soret band at 426 nm and a shoulder at longer wavelength at 435 nm in the reduced form, while the shoulder at the shorter wavelength has shifted from 380 nm to 375 nm. These shifts of the Soret band were accompanied by a red shift in α/β region from 498 and 535 nm in the oxidised form to 547 and 565 when reduced. The CT band at 640 nm has disappeared from the ferrous form's spectrum this may be due to the change in the spin state of the haem iron.

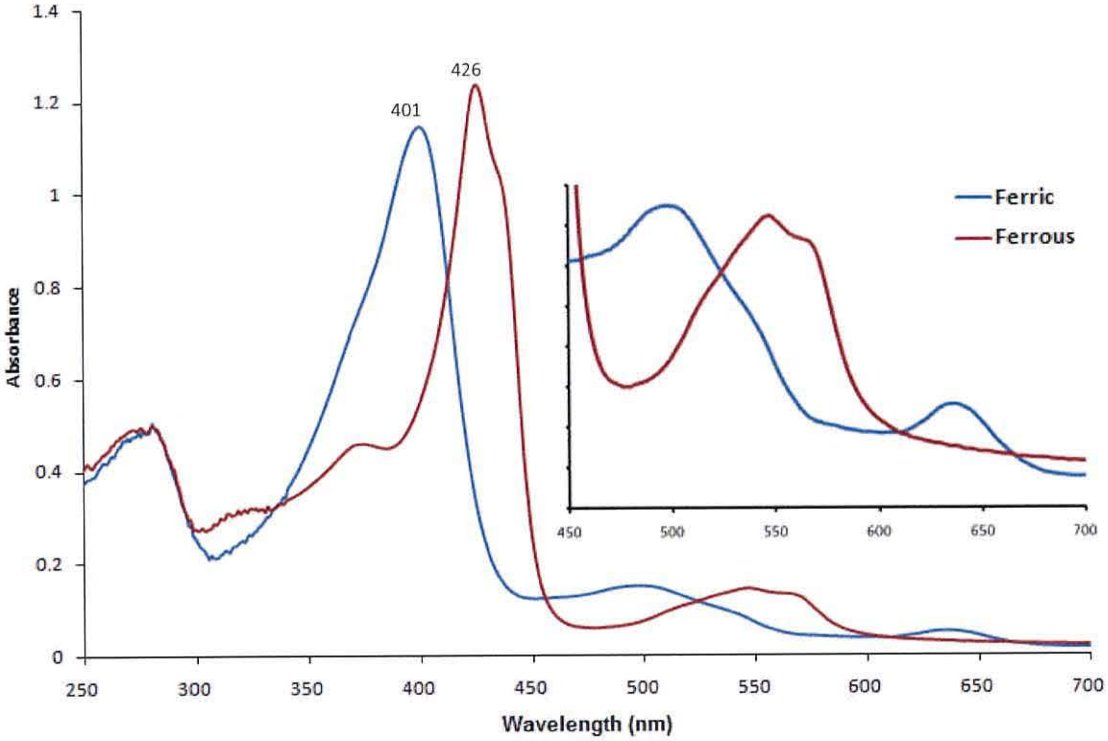


Figure 5.6: Electronic spectra of (17.0 μM) ferric and ferrous L16F mutant at pH 6. The calculated extinction coefficients are 66.96 and 71.0 $mM^{-1}.cm^{-1}$ for the ferric and ferrous forms respectively.

Comparing ferric and ferrous forms of the L16F mutant with recombinant *AXCP*, ferric L16F has exhibited a red shift of the Soret band from 394 nm to 401 nm. Concomitantly, the other bands displayed no changes (Figure 5.11). While ferrous L16F showed a slight shift in the split Soret band from 424, 434 nm (sh) to 426, 435 nm (sh). The shoulder at the shorter wavelength shifted from 374 nm to 375 nm. (Figure 5.12).

5.1.2.3 Spectral properties of L16Y:

The electronic spectra of ferrous and ferric haem iron of L16Y mutant are shown in figure 5.7. Comparing ferric and ferrous forms of this mutant, the Soret band at 404 nm in the case of ferric form has changed to split Soret bands at 424 nm with a shoulder at 437 nm in the ferrous form, and the shoulder at 373 nm band has became more predominant. The α/β region displayed a red shift of both bands from 498 and 535 nm in ferric form to 548 and 566 nm respectively when reduced. Moreover, they became more distinctive. Finally the CT band has vanished upon the reduction revealing the change in the spin state from admixture to the low spin state.

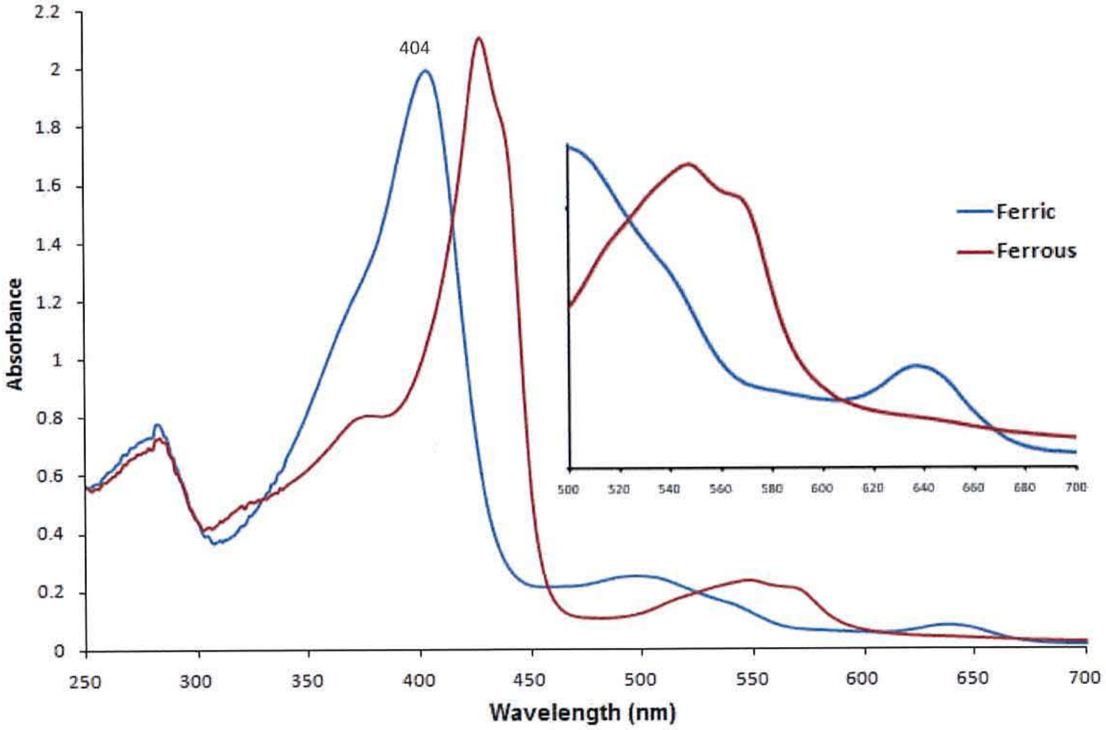


Figure 5.7: Electronic spectra of 25.6 μM ferric and ferrous L16Y mutant at pH 6. Extinction coefficients are 77.0 and 86.11 mM^{-1}.cm^{-1} for ferric and ferrous respectively.

Compared to the ferric recombinant protein, the Soret band of ferric L16Y shifted from 394 to 404 nm and the shoulder from 380 to 373, while bands in α/β region remained unchanged (Figure 5.11). Ferrous L16Y exhibits a small shift in wavelength maxima of the major bands in the spectrum compared to the ferrous form of recombinant *AXCP* (Figure 5.12); the split Soret band shifted to 428, 437 nm (sh) from 424, 434 nm (sh) and the shoulder at the lower wavelength shifted to 373 from 374 nm of the recombinant.

5.1.2.4 Spectral properties of W56G:

The UV-Vis spectra of the ferric and ferrous W56G mutant (Figure 5.8) showed that the Soret band of the ferric W56G at 404 nm changed to a split band and shifted to 423, 432 nm (sh). A distinctive shoulder at a lower wavelength has appeared at 369 nm. The α/β region of the ferric form showed a broad band at 490 and another at 535 nm, while these bands became more pronounced at 548 and 563 nm with a shoulder at 511 nm in the case of the ferrous form. Again the CT band at 634 nm has disappeared upon the reduction.

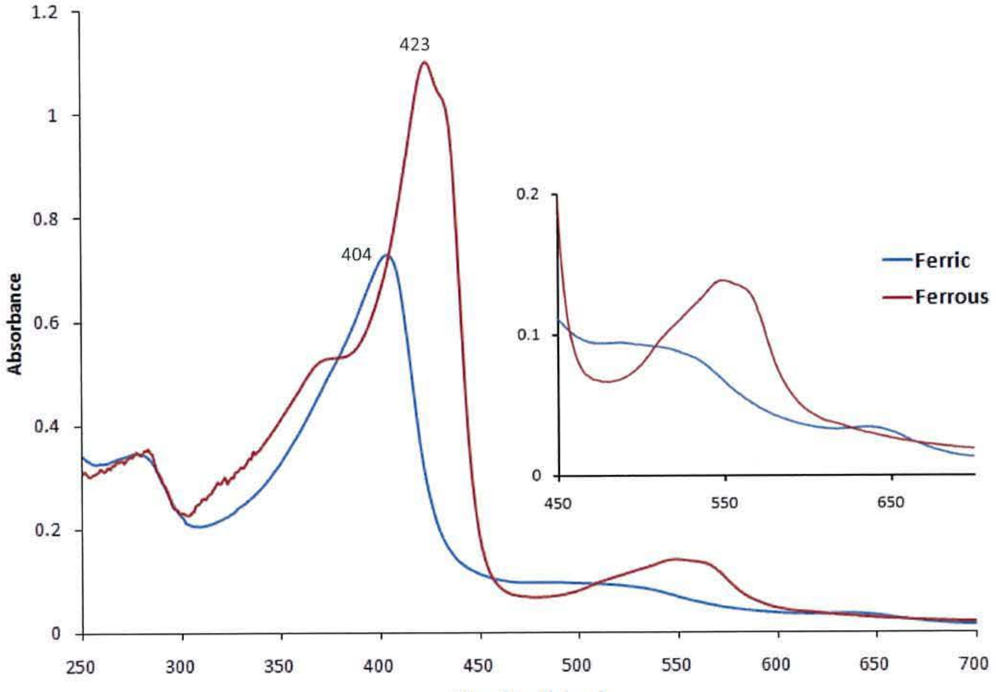


Figure 5.8: Electronic spectra of 13.5 μM ferric and ferrous W56G mutant at pH 6. The calculated extinction coefficients are 60.0 and 88.7 $mM^{-1}.cm^{-1}$ respectively.

The ferric W56G when compared to ferric recombinant *AXCP* showed a red shift of the Soret band to 404 nm and the shoulder at 380 nm ablated. A small shift has been noticed at α/β region to 494 and 539 nm, where bands are observed at 498 and 535 nm in recombinant *AXCP*. Finally, the CT band shifted to 634 nm (Figure 5.11). The Soret band of ferrous W56G remained unchanged compared to the ferrous recombinant at 424 and 434 nm, but the shoulder has shifted to 369 nm from 374 nm. Apart from appearance of a new shoulder at 511 nm, the 550 and 565 nm band of the α/β region bands stayed same (Figure 5.12).

5.1.2.5 Spectral properties of W56Y:

Reducing the ferric form of W56Y shifted the Soret maxima from 400 to a split band at 425 nm and 434 nm (sh), the shoulder at 380 nm shifted to 372 nm and became more pronounced. The α/β region bands shifted from a broad band centred at 500 and a shoulder at 548 nm to well-defined band at 549 and 568 nm respectively. The CT band at 640 nm has vanished upon the reduction (Figure 5.9).

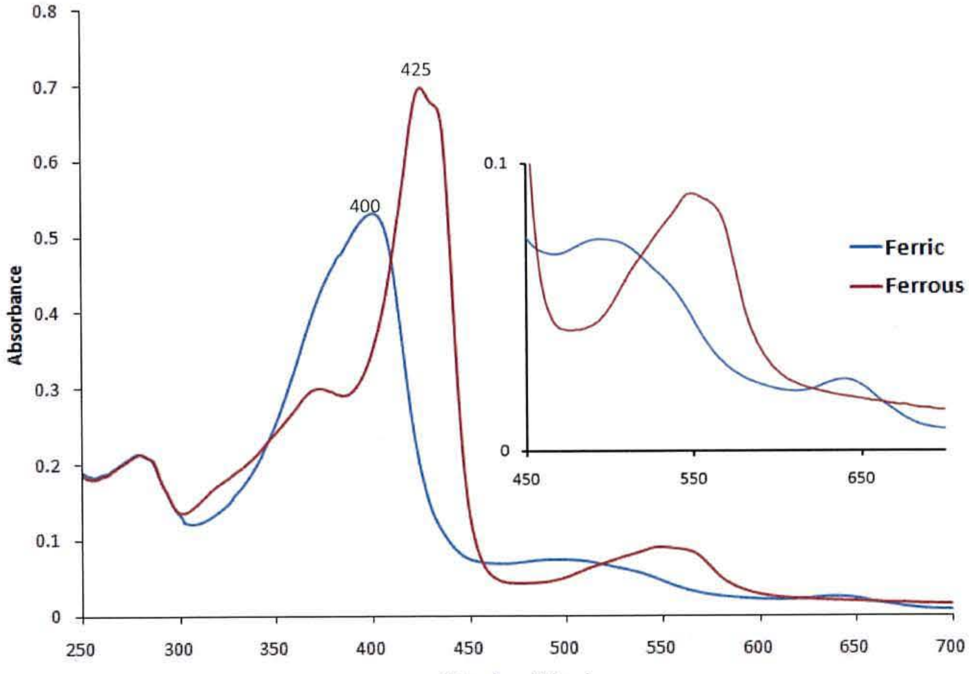


Figure 5.9: Electronic spectra of 8.0 μM ferric and ferrous W56Y mutant at pH 6. Extinction coefficients for the ferric and the ferrous forms are 70.3 and 80.0 $mM^{-1}.cm^{-1}$ respectively.

Replacing tryptophan 56 by tyrosine does not affect the Soret band in its ferrous form compared to the ferrous recombinant *AXCP*. However, the shoulder at shorter wavelength of the ferrous form of W56Y has shown a slight shift to 372 nm from the 374 nm observed for the reduced form of the recombinant *AXCP*. Similarly, the α/β region does not show any significant shifts in comparison the ferrous forms (Figure 5.12). While, in the ferric form the Soret band shifted to 400 nm from 394 and the α/β peaks are shifted to 500 and 548 nm from 498 and 535 nm of the recombinant, respectively (Figure 5.11).

5.1.2.6 Spectral properties of L16F/W56G:

Mutating two amino acids; leucine 16 to phenylalanine and tryptophan 56 to glycine at the same time, showed a split Soret at 400 nm with a shoulder at 380 nm. Upon the reduction, these bands have shifted to 426, 435 nm (sh) and to 375 nm (sh) and the CT band has disappeared (Figure 5.10).

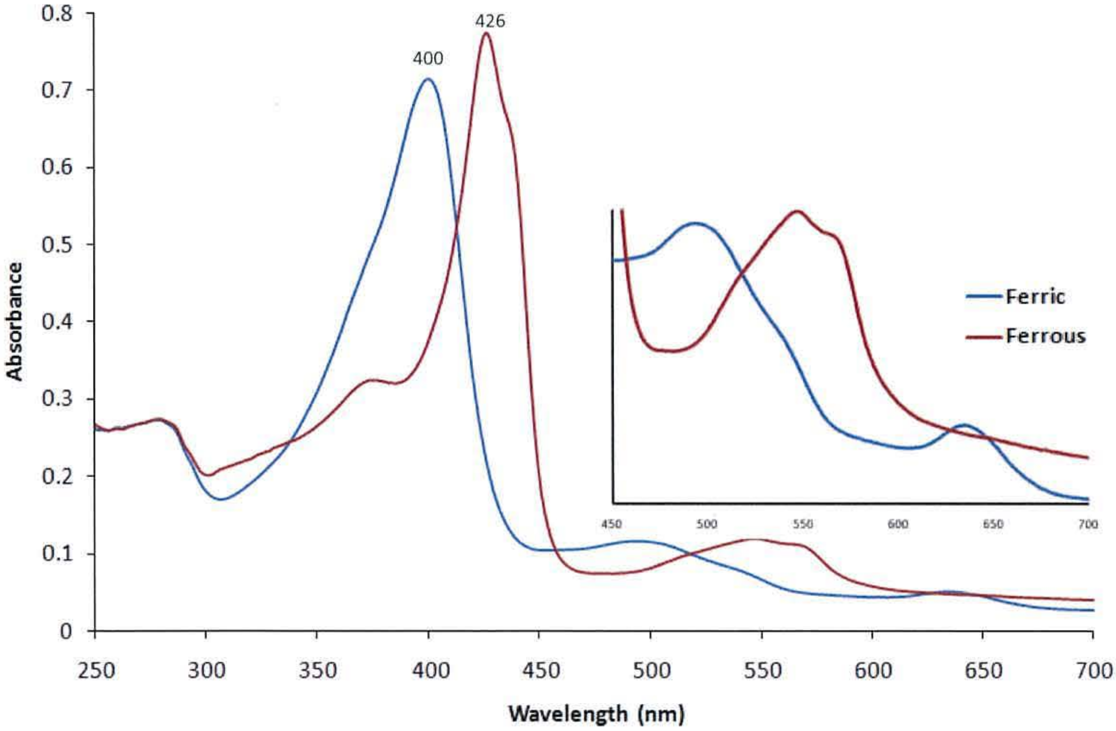


Figure 5.10:Electronic spectra of 9.5 μM ferric and ferrous L16F/W56G mutant at pH 6. The extinction coefficients are 74.48 and 80.15 $mM^{-1}.cm^{-1}$ for the ferric and the ferrous forms respectively.

Relative to the recombinant *AXCP*, the spectrum of the ferric L16F/W56G mutant protein exhibited a red shift of the Soret band to 400 nm. The α/β region bands shifted to 494 and 542 nm, and the CT band has shifted to 635 nm (Figure 5.11). The Soret band of the ferrous form of the double mutant protein shifted to a longer wavelength by 2 nm to 426 nm and the shoulder by 1 nm to 435 nm. The α/β region shifted to 546 and 569 nm (Figure 5.12).

For clarity, ferric forms of the recombinant native and all mutants were superimposed and depicted in (Figure 5.11) and the ferrous forms in (Figure 5.12). In addition all the spectral data of ferric forms are summarised in (Table 5.1) and the ferrous forms summarised in (Table 5.2).

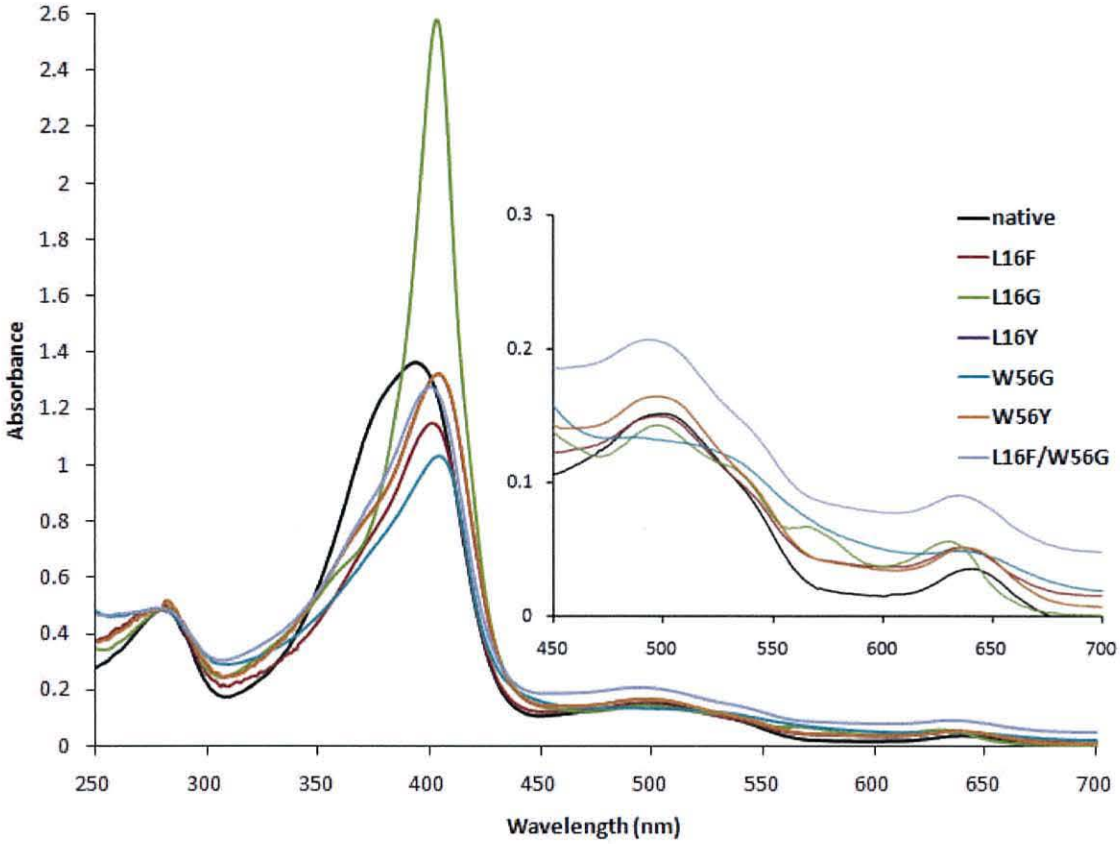


Figure 5.11: Electronic spectra of ferric recombinant, L16F, L16G, L16Y, W56G, W56Y and the double mutant L16F/W56G at pH 6.

Ferric form	Soret λ_{max} nm	α/β band nm	CT nm	Extinction coefficient $mM^{-1}.cm^{-1}$
Recombinant native	380 sh, 394	498, 535 sh	640	80.0
L16F	380 sh, 401	498, 535 sh	636	66.96
L16G	364 sh, 403	496, 536 sh, 571 sh	631	185
L16Y	373 sh, 404	498, 535 sh	640	77.07
W56G	404	490, 535 sh	634	60.0
W56Y	380 sh, 400	500, 548 sh	639	70.30
L16F/W56G	380 sh, 400	494, 542 sh	635	74.5

Table 5.1: Electronic spectral data of ferric recombinant *AXCP* and mutants at pH 6. CT is the charge transfer band.

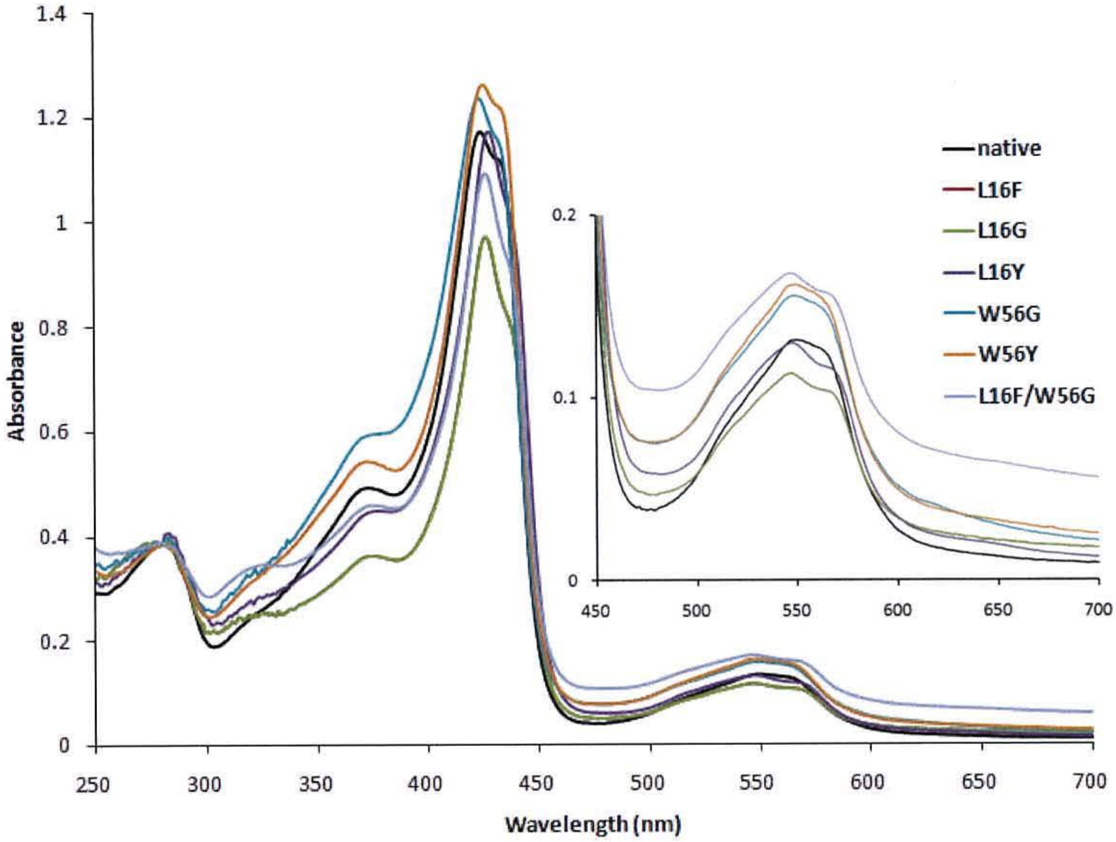


Figure 5.12: Electronic spectra of ferrous recombinant, L16F, L16G, L16Y, W56G, W56Y and the double mutant L16F/W56G at pH 6.

Ferrous form	**Soret λ_{max} nm**	**α/β λ_{max} nm**	**CT nm**	**Extinction coefficients mM^{-1}.cm^{-1}**
Recombinant native	374 sh, 424, 434	550, 565 sh	-	86.0
L16F	375 sh, 426, 435	547, 565 sh	-	71.04
L16G	336 sh, 409 sh, 413	538, 571 sh	-	62.9
L16Y	373 sh, 428, 437	548, 565 sh	-	86.11
W56G	369 sh, 423, 432	511 sh, 550, 565 sh	-	88.7
W56Y	372 sh, 425, 434	549, 568 sh	-	80.06
L16F/W56G	375 sh, 426, 435	546, 569 sh	-	80.15

Table 5.2: Electronic spectral data of ferrous recombinant *AXCP* and mutants at pH 6. CT is the charge transfer band.

5.2 pH titrations of ferric and ferrous recombinant and mutants *AXCP*:

The spectral property of ferric *AXCP* is highly dependent on the solvent pH.[15] The QS state changes to the pure high spin state upon an increase in the pH. This can be observed as the appearance of a shoulder between 370 and 380 nm and a shift of the Soret band to 400-407 nm at alkaline pH, and changes in the band intensities at 500 and, 535 nm. The band at 643 nm is characteristic of the presence of the high-spin state. Unlike the ferric form, the electronic absorption spectrum of the ferrous form of recombinant *AXCP* is insensitive to pH.

In this work, pH titrations were carried out for both ferric and ferrous states of recombinant *AXCP* and all mutants to find out whether distal pocket residues play any role in pH dependent behaviour. Certainly, the ferrous forms of the recombinant and all other mutants showed no change in the band positions. They were identical to those reported in Table 5.2 over the pH range 5-9, (data not shown). Hence, only ferric forms have been presented and discussed here. All titrations were carried out anaerobically as described in section 3.7.2.

5.2.1 pH titration of recombinant AXCP:

The pH dependence of the electronic spectra over the pH range 5-9 of the ferric recombinant *AXCP* are shown in Figure 5.13. At pH 5, the spectra show features due to the Soret band at 394 nm, a broad α/β region band at 500 nm with a shoulder at 535 nm and also a band at 640 nm corresponding to a HS feature. As the pH was increased from pH 5-7, the Soret and β band exhibited a red shift while the position of the α band was constant at 535 nm initially as a weak shoulder but becoming more prominent at pH 8 and 9. At pH 9 the Soret maxima shifted from 394 to 405 nm with a small increase in intensity and a distinct shoulder at 373 nm has appeared. The β band at 500 nm became less intense and shifted to 505 nm while the α band (shoulder) at 535 nm became more intense. From pH 5-9 the CT peak at 640 nm increased in intensity, consistent with an increase in the % of the HS species. These changes in the electronic absorption spectra suggest that there is a change in the spin state from admixture state at acidic and neutral pH to a pure high-spin state at alkaline pH. The change involves deprotonation of the his 120 axial ligand to the haem iron.

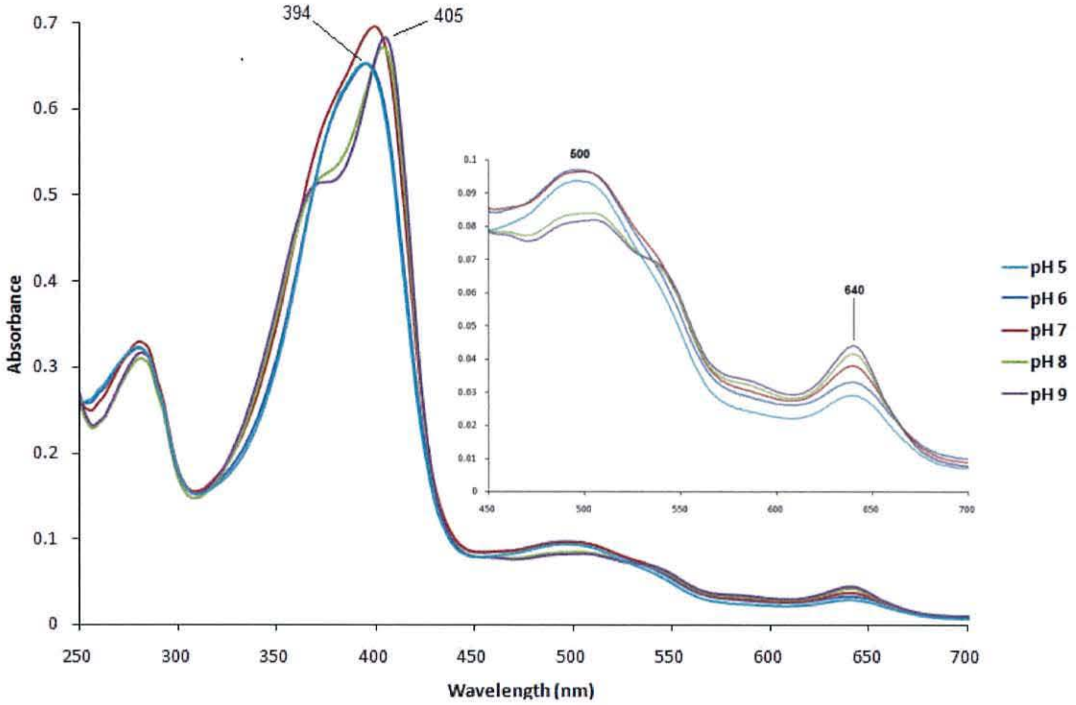


Figure 5.13: The pH dependent electronic spectra of 8.1 µM ferric recombinant *AXCP*

5.2.2 pH titration of L16F mutant:

At pH 5 and 6, L16F exhibited a Soret maximum of 401 nm and α/β region bands at 497 and 535 nm with the CT band at 637 nm. With an increase in the pH, all bands showed a red shift (Figure 5.13). At neutral pH the Soret band shifted from 401 to 403 nm and the beta band from 497 to 498 nm while the alpha band remained constant. The CT band shifted to 639 nm. At alkaline pH significant changes were observed. The Soret maxima band shifted to 407 nm. The β band to 502 nm with a decrease in its intensity, and the α band became more predominant at 535 nm. The CT band centred at 640 nm, with an increase in its intensity indicating the increase in the HS components. At pH 8, a distinct shoulder appeared at 377 nm and shifted to 372 nm at pH 9. A small band at 583 nm was observed at pH 8 and 9. These changes imply the change in the spin state of the iron haem from admixture spin state to pure high spin state at alkaline pH. However the change was less significant compared to the recombinant *AXCP* protein.

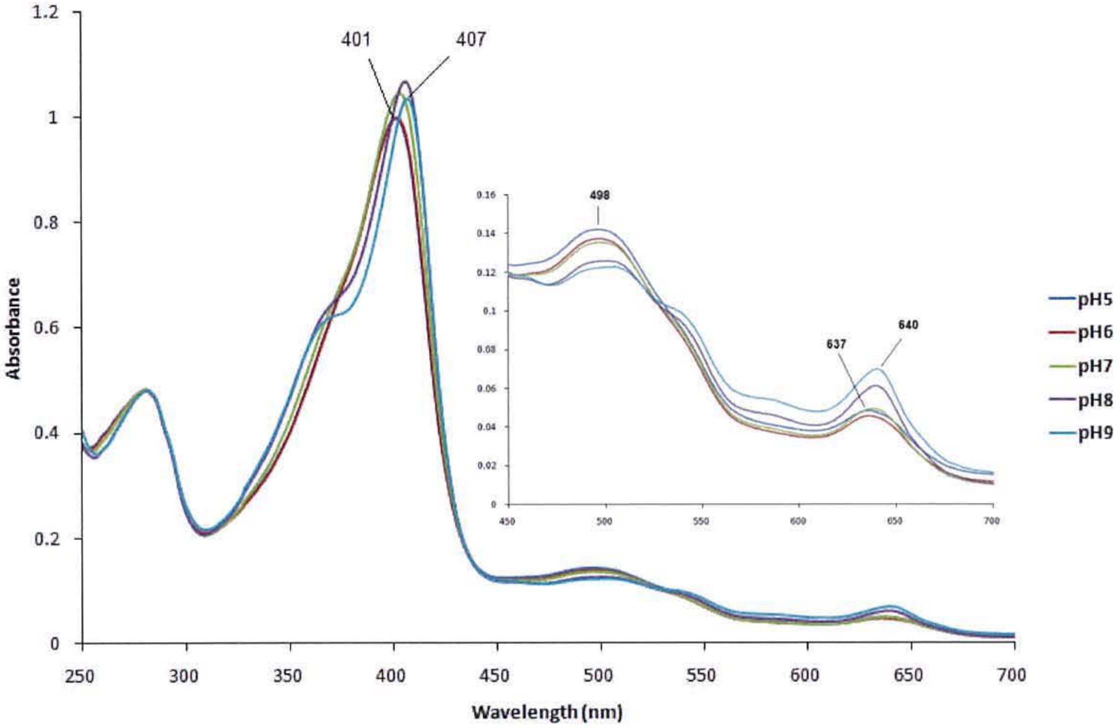


Figure 5.14: The pH dependent electronic spectra of 18.5 μM ferric L16F.

5.2.3 pH titration of L16G mutant:

In the case of L16G, no major change was observed in the position of the Soret band whereas the band intensity was decreased from pH 5 to pH 9 (Figure 5.15). A shoulder at 364 nm was observed over the pH range 5-9, while it was more predominant at pH 8 and 9. Although the position of the β band remained constant at 503 nm, its intensity decreased at pH 9. The α band was positioned at 535 nm with a shoulder at 567 nm over the pH range of 5-8, there was a slight shift of the shoulder from 567 to 571 nm at pH 9. The CT band shifted from 628 nm at pH 5 to 636 nm at pH 9. These electronic spectral observations suggest that the spin state of the ferric form of this mutant is insensitive to the change in the pH.

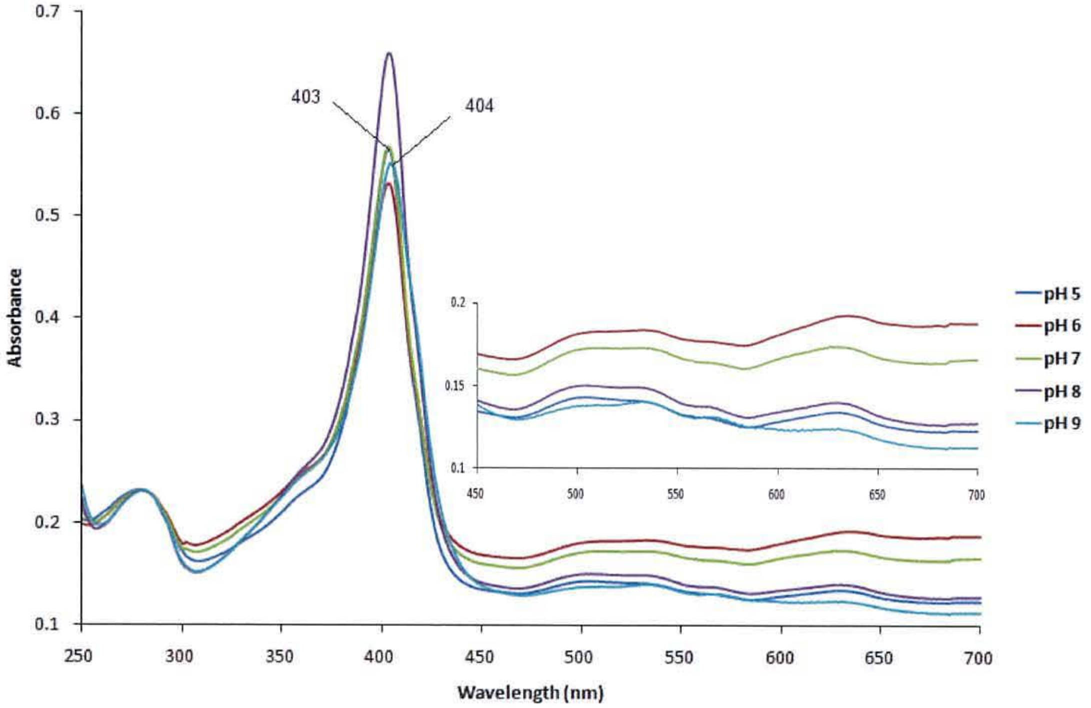


Figure 5.15: The pH dependent electronic spectra of 9.5 μM ferric L16G.

5.2.4 pH titration of L16Y mutant:

L16Y (Figure 5.16) exhibited a Soret maximum at 404 nm at pH 5 and pH 6 while it has shifted to 405, 408 and 409 nm at pH 7, 8 and 9 respectively, accompanied by an increase in the band intensity. At pH 8, a shoulder at a lower wavelength observed at 373 nm this was shifted to 369 nm at pH 9. The β band at pH 5 and 6 was at 498 nm, and shifted to 500, 505 and 507 nm at pH 7, 8 and 9 respectively with decrease in its intensity in the same order. The α band maxima at 539 nm was constant over the pH range, with an increase in its intensity at pH 8 and 9. The band at 583 nm which was also observed at pH 8 and 9 in the case of L16F, is predominant now over the pH range 5-9 while it was more pronounced at pH 8 and 9. CT band shifted from 636 nm at pH 5 to 641 nm at pH 9 with an increase in its intensity. The change in the electronic spectra of this mutant protein is due to increase in the HS species at alkaline pHs. The increase in the HS species may have been caused by the deprotonation of his 120 axial ligand at alkaline pH which in turn weakens the his-Fe axial band.

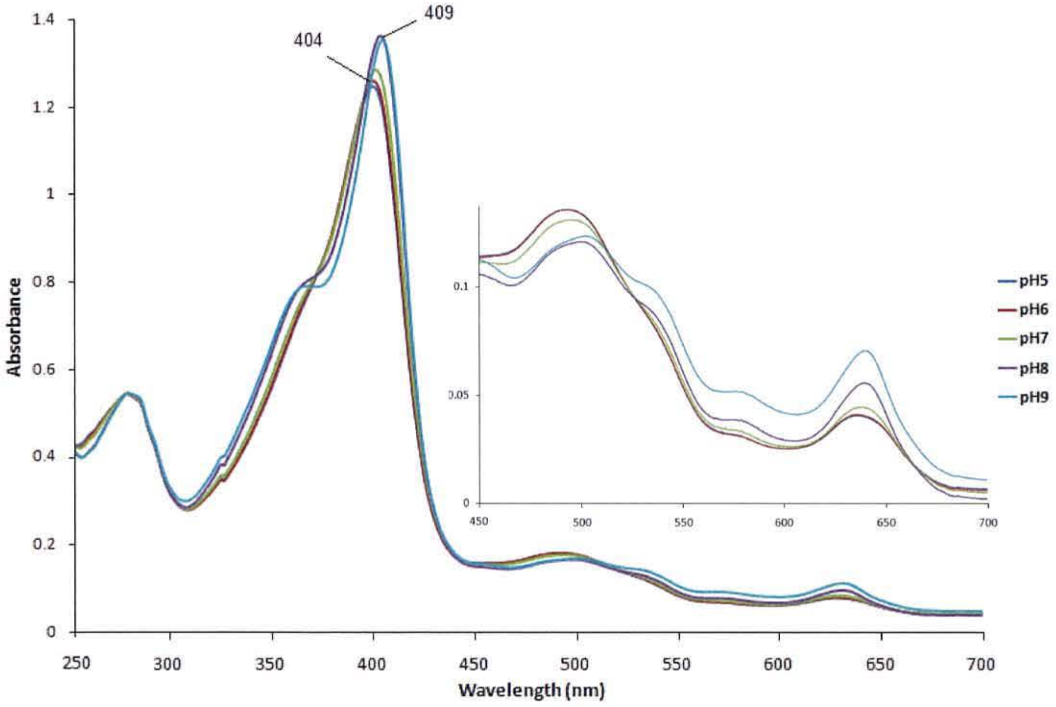


Figure 5.16: The pH dependent electronic spectra of 17.0 μM ferric L16Y.

5.2.5 pH titration of W56G mutant:

Replacing W56 by glycine does not cause any changes in the position of the Soret maximum, but there were minor changes in the intensity (Figure 5.17). However at pH 7 and 8 it has shifted to 406 nm and at pH 9 to 407 nm with appearance of a distinct shoulder at 373 nm at pH 8 and 9. The position of the α/β bands remained constant at 487 and 519 nm over the pH range 5-9 but became less pronounced at pH 9. There was a band observed at 535 nm at pH 8 which has shifted to 539 nm and increased in intensity and at pH 9. The CT band has revealed a slight red shift upon increase in the pH, which has shifted from 635 nm at pH 5 to 639 nm at pH 9 accompanied by increase in the band intensity. These observations showed that the changes are less predominant over the pH range 5 to 9 when compared to the recombinant *AXCP* protein. This confirms the effect of W56 on the spin state of the haem.

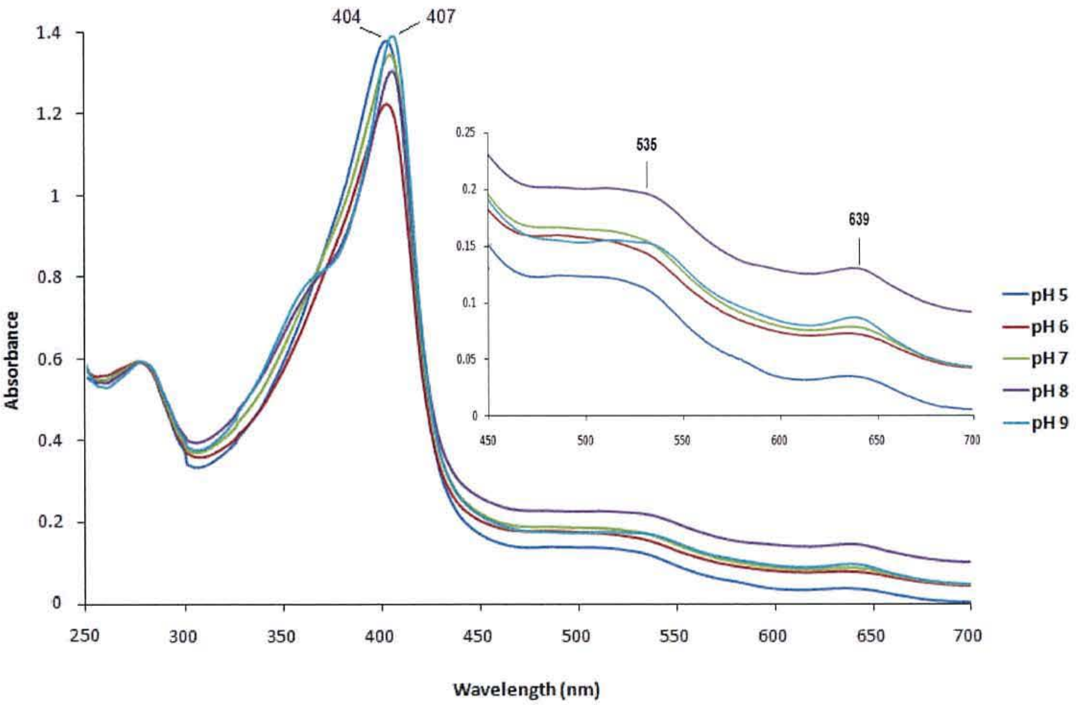


Figure 5.17: The pH dependent electronic spectra of 17.0 μM ferric W56G.

5.2.6 pH titration of W56Y mutant:

The key spectral features of the W56Y mutant were same at pH 5 and 6 (Figure 5.18). The Soret band was observed at 400 nm with a slight decrease in the intensity at pH 6, the β at 493 nm, α at 548 nm and the CT at 638 nm. At the neutral pH the Soret band has shifted to 403 nm. The β band at 494 nm has slightly decreased in intensity and developed a very small shoulder at 504 nm. The α band remained at the same wavelength at 548 nm, while the CT slightly shifted to 639 nm. Significant changes were observed at alkaline pH: the Soret band has shifted to 405 nm at pH 8 and to 406 nm at pH 9.Aa distinct shoulder at 372 has appeared in both cases. The β band has shifted from 493 to 490 nm and became less pronounced, while the peak at 504 nm has shifted to 510 nm and became more predominant. The α band became more prominent and was centred at 538 and 539 nm at pH 8 and 9, respectively. A significant increase in the CT band intensity has been observed and shifted to 640 nm.

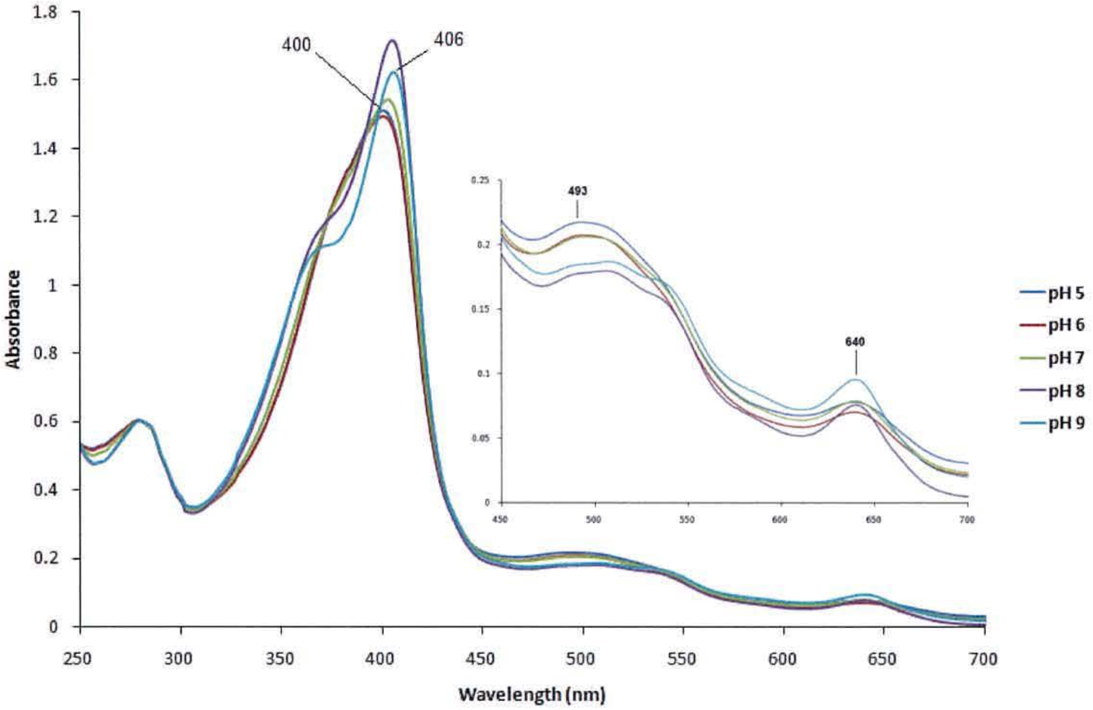


Figure 5.18: The pH dependent electronic spectra of 22.0 μM ferric W56Y.

5.2.6 pH titration of the double mutant L16F/W56G:

Again the key spectral features at pH 5 and 6 were same. The Soret maximum was at 400 nm. The β/α band at 494 and542 nm and the CT band was centred at 635 nm with a slight increase in its intensity at pH 6 (Figure 5.19). At pH 7, the Soret band shifted to 401 nm, and β band to 497 nm while the α band remained unchanged at 542 nm with the CT band slightly shifted to 636 nm. At pH 8 the Soret shifted to 404 nm and a distinct shoulder appeared at 376 nm. The Soret band has shifted even more to 406 nm accompanied by a slight increase in the band intensity at pH 9, and the shoulder has shifted to 373 nm. CT band increased in the intensity and centred at 640 nm at pH 9. A minor shoulder at 580 nm is observed in the pH range 5-9 with an increase in its intensity with increasing the pH.

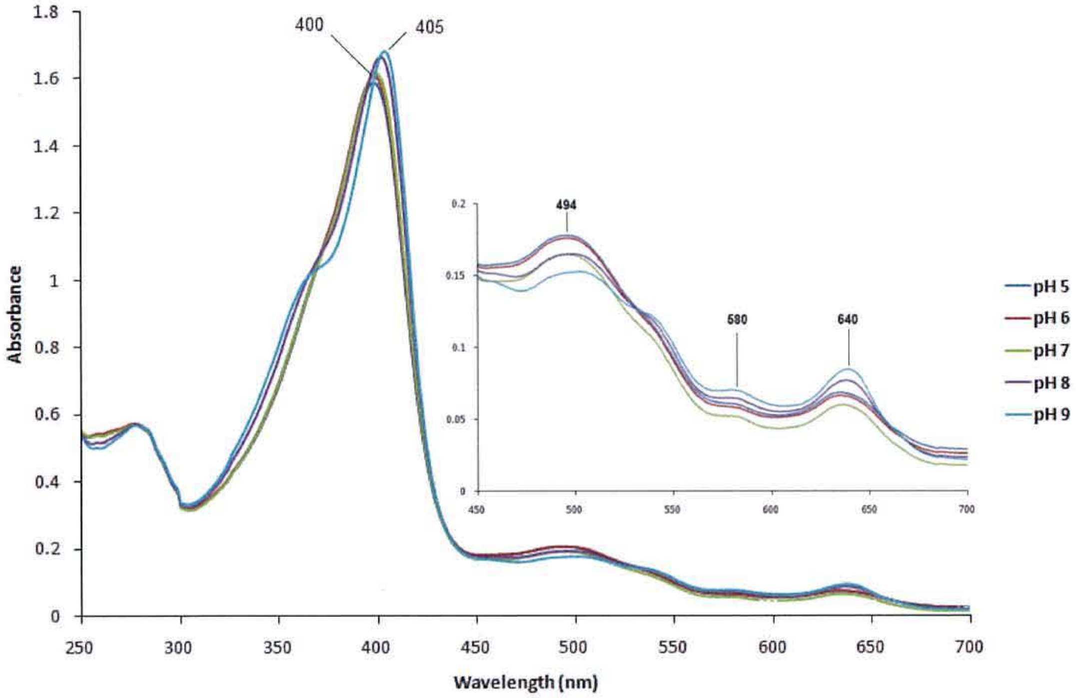


Figure 5.19: The pH dependent electronic spectra of 20.3 μM ferric double mutant L16F/W56G.

For simplicity and comparison, the above observation are summarised in (Table 5.3), which shows the Soret, α, β, and CT band in the pH range 5-9 for the recombinant and all other mutants.

- **Soret band- ferric**

pH	Recomb.	L16F	L16G	L16Y	W56G	W56Y	L16F/W56G
5	394	401	364sh, 403	404	404	400	400
6	380 sh, 395	401	364sh, 403	404	404	400	400
7	380 sh, 399	403	364sh, 403	405	406	403	401
8	374 sh, 404	377 sh, 406	364 sh, 403	373 sh, 408,	373 sh, 406,	372 sh, 405	376 sh, 404
9	373 sh, 405	372 sh, 407	364 sh, 404	369 sh, 409,	373 sh, 407,	372 sh, 406	373 sh, 405

- **β band**

pH	Recomb.	L16F	L16G	L16Y	W56G	W56Y	L16F/W56G
5	500	497	503	498	487	494	494
6	496	497	508	499	487	494	494
7	498	498	503	500	487	494sh , 504	497
8	505	501	503	505	487	490 sh, 510	501
9	505	502	503 sh	507	516	490 sh,510	494 sh, 507

- **α band**

pH	Recomb.	L16F	L16G	L16Y	W56G	W56Y	L16F/W56G
5	535 sh	535 sh	535, 567	539 sh, 583	519	548 sh	542 sh
6	535 sh	535 sh	535, 567	539 sh, 583	519	548 sh	542 sh
7	535 sh	535 sh	537, 567	539 sh, 583	519	548 sh	542 sh
8	535	535	535, 567	539, 583	519, 535 sh	538	542
9	535	535	535, 571sh	539, 583	519 sh, 539	539	542

- **CT band**

pH	Recomb.	L16F	L16G	L16Y	W56G	W56Y	L16F/W56G
5	640	637	628	636	635	638	635
6	640	637	629	637	638	639	635
7	640	639	631	640	638	639	636
8	640	639	631	641	638	640	638
9	640	640	636	641	639	640	640

Table 5.3: Summarised electronic spectral data of the ferric form of recombinant, L16F, L16G, L16Y, W56G, W56Y and the double mutant L16F/W56G observed at different pH range 5-9. The table shows the electronic absorption maxima of the Soret, α, β, and Charge Transfer (CT) bands.

5.3 Nitric oxide titrations:

Nitric oxide (NO) is a small, short-lived, and highly reactive gaseous molecule. Unlike most radicals, NO does not dismute nor covalently dimerise; it readily diffuses through cell membranes and reacts with the haem iron and acts as a signal transmitter in physiological pathways[16]. Haemoproteins usually form a 6c-NO-His complex having histidine at the proximal side and NO at the distal side of the haem. But in *AXCP*[17] and guanylate cyclase (sGS)[18] NO replaces the histidine ligand to give a five coordinate iron (5c-NO). In the *AXCP*, the mechanism of 5c-NO formation was found to be in two steps and concentration dependent. Initial binding of NO to the distal position of the haem (6c-NO-haem) leads to the cleavage of the Fe-His bond and the formation of 5c-NO complex. The crystal structure of the 5c-NO adduct showed that NO is bound to the proximal face of the haem and forms a H-bond with R124. Resonance Raman studies[11] indicate a positively polarised environment of proximally bound NO, consistent with its binding close to R124, implicating the role of this residue in ligand discrimination. Mutation of R124 does not result in the abolition of NO binding .[12] However, it has been shown recently[19] that the proximal preference for NO is mainly due to the steric destabilisation of the L16 residue that crowds the distal face of haem.

5.3.1 NO titrations with recombinant native cytochrome c':

The saturated solution of nitric oxide was prepared as described in section 3.7.3. Ferrous forms of proteins were prepared following steps outlined in section 3.7.1 which was discussed in detail in section 5.1. Aliquots of the saturated solution were titrated against the ferrous form of the protein in a total volume of 1 mL. The titration process was achieved in the super sealed quartz cuvette. After each addition a spectrum was recorded. The addition of nitric oxide solution was terminated when no further change in the spectrum was observed compared to the previously recorded spectrum.

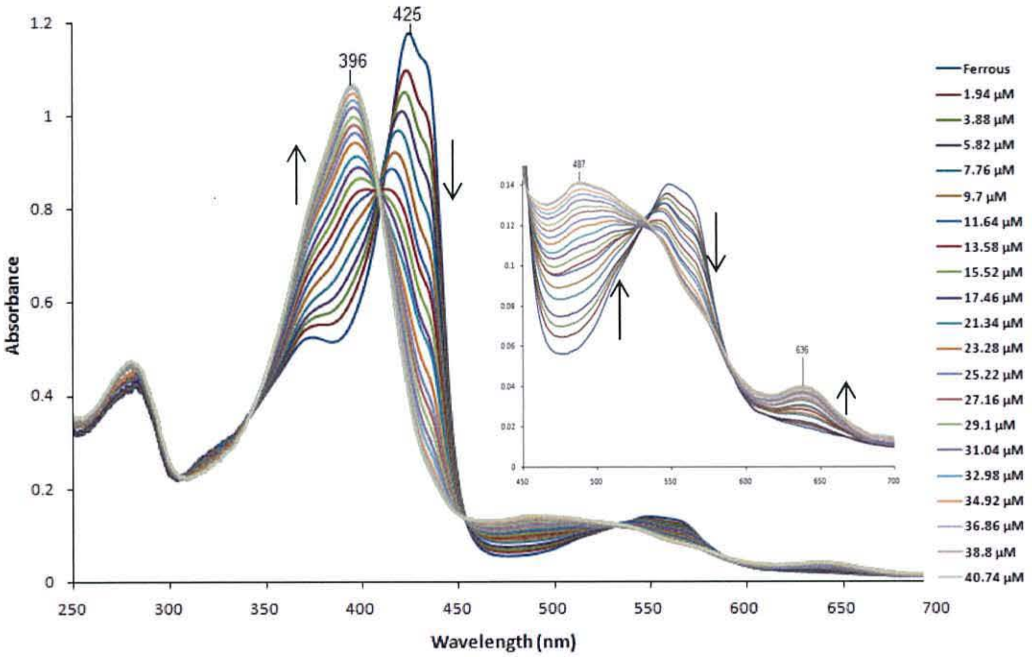


Figure 5.20: Electronic spectra of 14.5 µM ferrous recombinant native *AXCP* titration at pH 6 against NO. The extinction coefficient is 63.8 $mM^{-1}.cm^{-1}$.

Aliquots of 1 µL saturated NO solution were titrated against 14.5 µM recombinant ferrous *AXCP*, and the spectrum was recorded after each addition and shown in Figure 5.20. The final concentrations of NO in the cuvette are shown as the titre points. Since the 6c-NO-His intermediate *AXCP* has a relatively long lifetime in the presence of sub-stoichiometric NO concentrations, it can be measured spectroscopically.[20] While in the presence of an excess amount of NO the 5c-NO form has formed rapidly. Hence, the NO solution was added gradually to trap the 6c intermediate. This intermediate was observed when the concentration of the NO solution reached 11.64 µM (trace the 6th addition in figure 5.20) characterised by the 414 nm peak which is characteristic for 6c-NO-His.[21,1] The addition of saturating concentrations of NO lead to the formation of 5c-NO *AXCP* . This was evidenced by the Soret maxima band located at 396 nm and the CT band at 636 characteristic for the 5c-NO adduct. The expanded region from 450-700 nm shows a broad band at 487 nm, β band at 536 nm and α band at 565 nm which are characteristic of 5c-NO haem protein.[22]

5.3.2 NO titration with L16F, L16Y and L16G mutants

Titrations of L16F, L16Y and L16G with NO are shown in (Figures 5.21-23). From the data presented the characteristic peak at 416 nm was not observed and hence no evidence for the presence of the 6c-NO-His intermediate was detected. L16F exhibited a shifted Soret peak from 426 nm to 401 nm, and the Soret peak of L16Y shifted from 427nm to 403 nm directly without observing any evidence for the presence of 6c-NO-His haem. Bands at 634 and 637 in the case of L16F and L16Y respectively which are sensitive the coordination number, confirms the formation of the 5c-NO.

A peak at 401 nm in the case of L16F and 403 nm in the case of L16Y have been observed in the 5c-His ferric forms of these two mutants as shown in Table 4.1. Thus it is inferred that the NO complexes of L16F and L16Y form 5c-NO adducts. The five coordinate nature of the iron in these mutants has been further confirmed by the determination of their crystal structures which are presented in this thesis (see chapter 6).

Replacing L by F (Figure 5.21) and Y (Figure 5.22) has changed the reactivity towards NO. On the timescales that these spectroscopic experiments have been conducted it has not been possible to observe a 6c-NO-His intermediate. This may be interpreted either as a much faster inter-conversion of the 6c-NO-His to the 5c-NO or the formation of the final 5c-NO adducts is proceeded via a 6c-NO-His intermediate.

According to the crystal structure of L16Y, solved in this work at 0.98 Å resolution (see chapter 6), the phenol ring of the tyrosine is stacked parallel to the haem plane and in van der Waals contact. It appears that NO, CO or other ligands binding to the sixth haem iron ligand site must perturb the tyrosine side chain. On the other hand, tyrosine 16 is tightly constrained by neighbouring residues, hence there must be other changes to accommodate this movement. No crystal structure is available for L16F, but the double mutant crystal structure of L16F/W56G resolved to 0.96 Å resolution in this work (see chapter 6) showed the same observations as for L16Y regarding the position and orientation of phenylalanine side chain in L16F. Ren *et al* [23] have predicted that cytochromes c' having an aromatic residue at this position will dissociate into monomers upon binding haem ligands. Accordingly, L16Y and L16F monomer may dissociate while binding to NO.

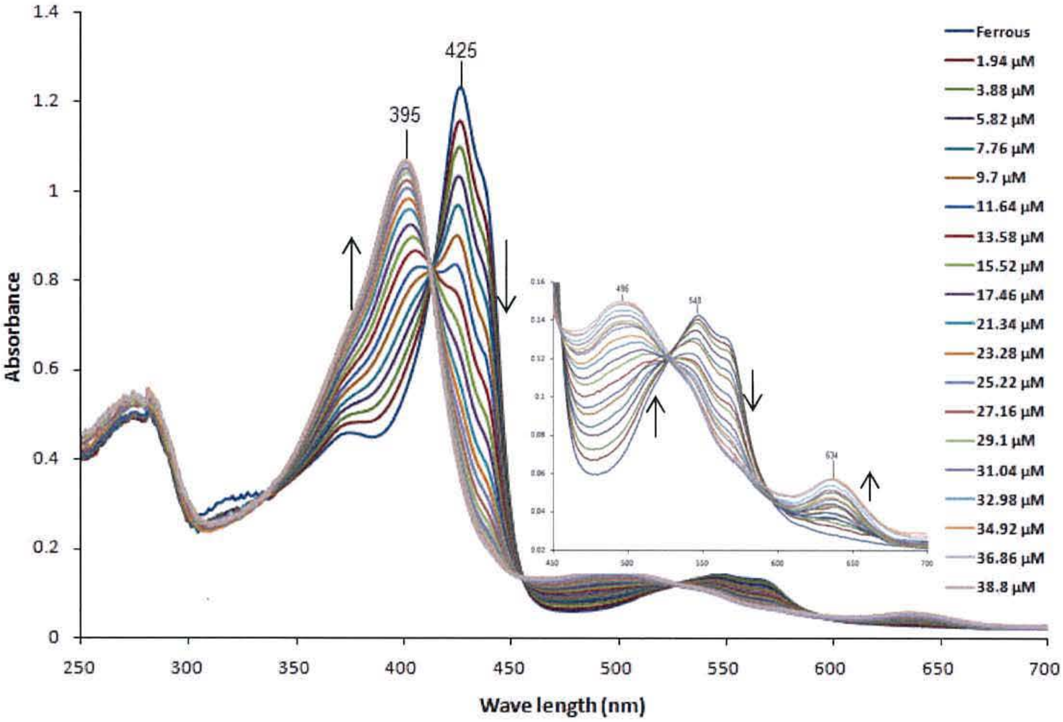


Figure 5.21: Electronic spectra of titration of mutants 17 µM L16F with NO at pH 6. The calculated extinction coefficient is 57.7 $mM^{-1}.cm^{-1}$.

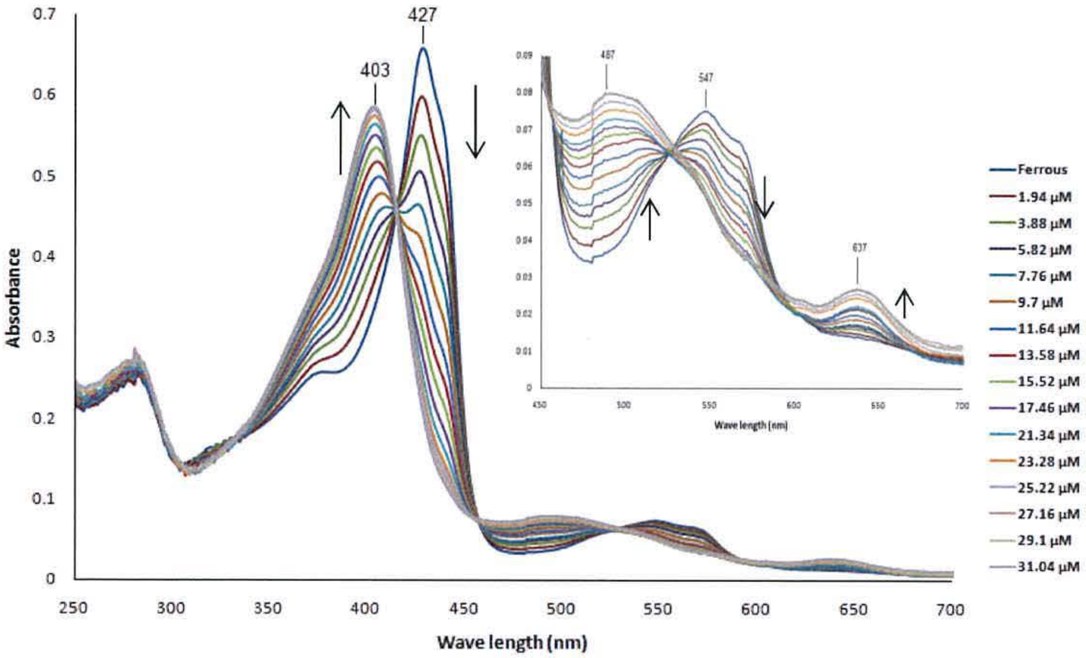


Figure 5.22: Electronic spectra of titration of mutants 8.1 μM L16Ywith NO at pH 6. The calculated extinction coefficient is 63.8 mM^{-1}.cm^{-1}.

Very interestingly, substitution of leucine by glycine (Figure 5.23) resulted in the formation of a stable 6c-NO-His adduct. This is in contrast to the recombinant and other mutant proteins reported so far where they form a 5c-NO species. The crystal structure of this mutant in its native state showed that it is 6-coordinate with a water molecule bound to the iron's sixth coordinate (6c-H_2O-His) (see chapter 6). For NO titration purposes, any bound molecule to the haem sixth coordinate of L16G has to be removed to make the distal side available for NO binding. This was achieved by the oxidation of the L16G mutant protein using an oxidising agent; sodium ferricyanide (see section 3.7.1). The excess oxidising agent was removed on a desalting column. Oxidised 5c mutant protein then reduced prior to the NO titration. Addition of a saturated solution of NO resulted in the formation of an intense and sharp peak at 416 nm confirming the 6c-NO-His formation. Peaks at 538 and 571 nm which had already been seen in the ferrous form did not change, while the peak at 631 nm has been shifted to 629 nm. L16G is the only mutant reported so far to form a stable 6c-NO-His. This clearly shows the importance of the residues present in the distal pocket in driving the 5c-NO-His formation. Substituting leucine for the much smaller glycine in the distal pocket made this side less crowded and hence more exposed to the exogenous molecules. This interesting finding confirms the vital role of leucine in the ligand binding and discrimination.

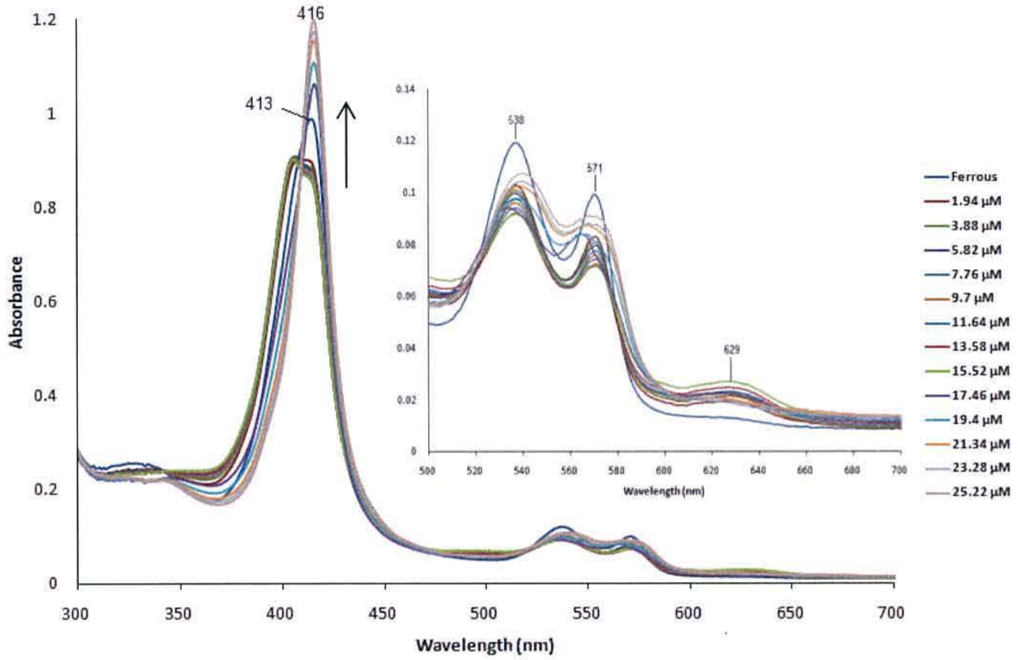


Figure 5.23: Electronic spectra of titration of 6.5 µM L16G against NO at pH 6. The calculated extinction coefficient is 51.5 mM^{-1}.cm^{-1}.

5.3.3 NO titrations with W56G, W56Y and the double mutant L16F/W56G:

Addition of aliquots of 1.94 mM NO to the 13.5 µM W56G solution decreased the intensity of the 423 nm band accompanied by disappearance of the shoulder at 369 nm (Figure 5.24). These two bands have vanished upon the addition of further NO solution until any further spectra changes were observed at a concentration of 64 µM NO. At the end of the titration, a band at 397 nm specific for the 5c has appeared, alongside with broad bands at 487 and 636 nm.

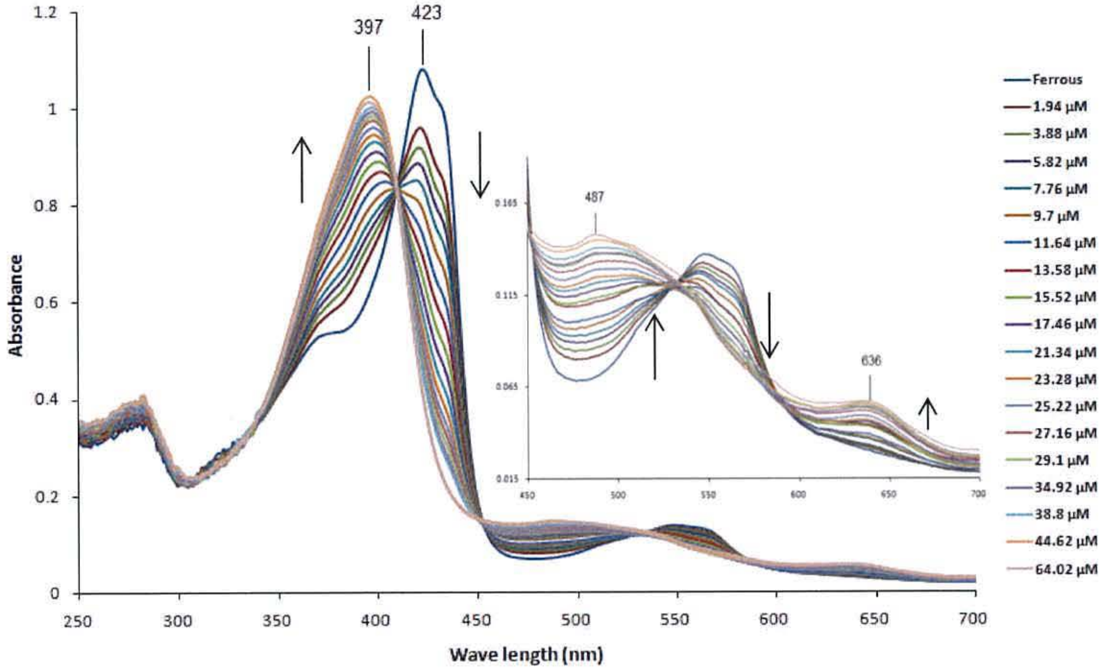


Figure 5.24: Electronic spectra of mutants 13.5 µM W56G titrated with NO at pH 6. The calculated extinction coefficient is 70.8 mM^{-1}.cm^{-1}.

W56Y exhibited faster reactivity compared to native during titration with NO (Figure 5.25). The first 1 µL aliquot of 1.94 mM NO solution caused a significant decrease in the absorption intensity at 426 nm with appearance of a band at 402 nm. Addition of a further aliquot of 1.94 mM NO caused the appearance of a band at 395 nm and a broad band at 482 and 541 nm. In contrast to the recombinant and all other mutants no band at 630 nm region was observed.

Comparing the spectra of these two mutants, W56Y required 25 µM NO solution to reach the saturation while W56G required 64 µM NO solution.

Similarly to L16F, L16G and L16Y, the electronic spectra of W56G and W56Y showed no peak at 416 nm suggesting that no 6c-NO-His has been formed. This may be due to the fast inter-conversion of 6c-NO-His to 5c-NO adduct.

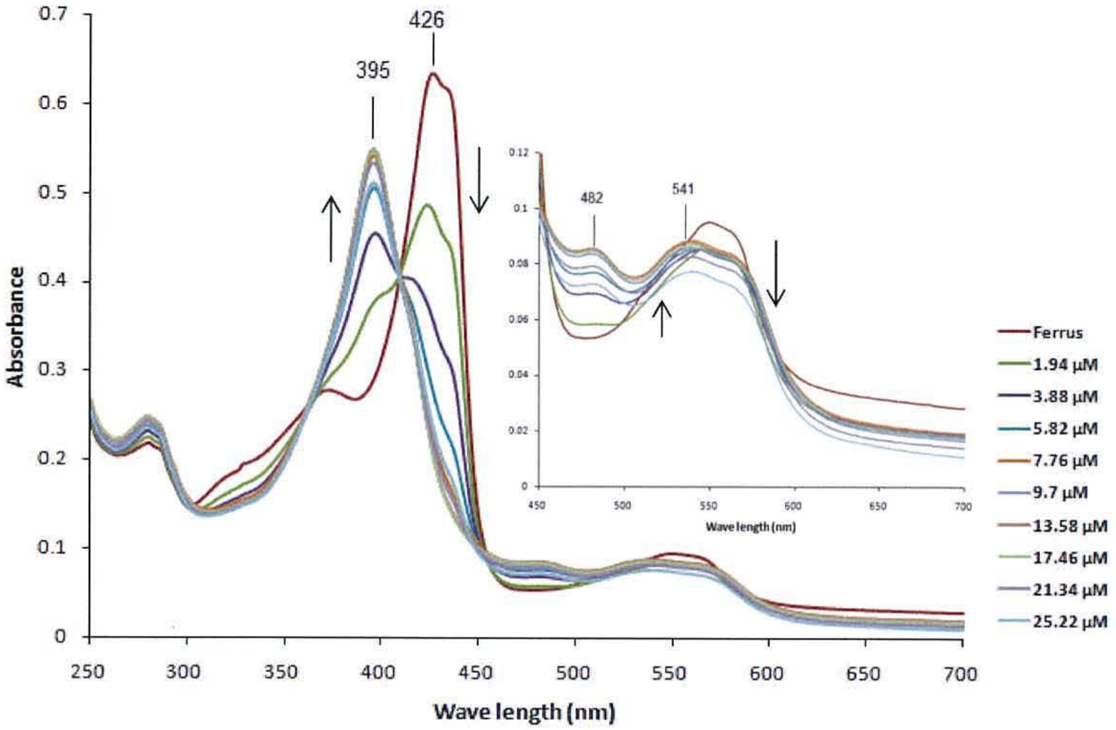


Figure 5.25: Electronic spectra of mutants 9.0 µM W56Y titrated with NO at pH 6. The calculated extinction coefficient is 62.8 mM^{-1}.cm^{-1}.

The above observations showed the effect of mutations of W56 on ligand binding and affinity. Replacing W56 by glycine which hypothesised to provide an extra solvent channel and hence increase its reactivity towards the NO, did not make a significant change in the reactivity. There also were not any indication that that a 6c-NO-His intermediate was formed. Therefore, the mutation W56G has perturbed the solvent channel in a manner which is yet to be fully understood.

Surprisingly, W56Y exhibited a significantly increased reactivity to NO binding compared to recombinant and all other mutants. The inference is that the factors controlling NO binding in the distal pocket are more complex than purely steric effects and that Leu16 has a more important role to play in controlling NO binding to distal side.

The 6c-NO-His formation was not detectable during the NO titration of the double mutant L16F/W56G (Figure 5.26), indicating a fast inter-conversion from 6c-NO-His to 5c-NO or it is not formed at all. Addition of aliquots of 1.94 mM NO solution to the protein solution caused a significant decrease in the Soret maxima 426 nm intensity, and shoulder at a shorter wavelength at 406 nm has appeared. Further additions made the Soret band disappear, and the shoulder to shift to 400 nm upon the saturation which is characteristic for 5-coordibate L16F/W56G (see table 5.1). At the same time a band at 486 nm and another at 634 nm has appeared.

For both L16F/W56G and L16F, the 6c intermediate was undetectable. However the double mutant seems to have more reactivity towards NO. This higher reactivity may be due to that: the mutation of W56 to glycine has exposed the distal pocket of the haem to the exogenous ligands probably by providing the proposed extra solvent channel. Moreover, replacing L16 by phenylalanine played a positive role in ligand binding reactivity.

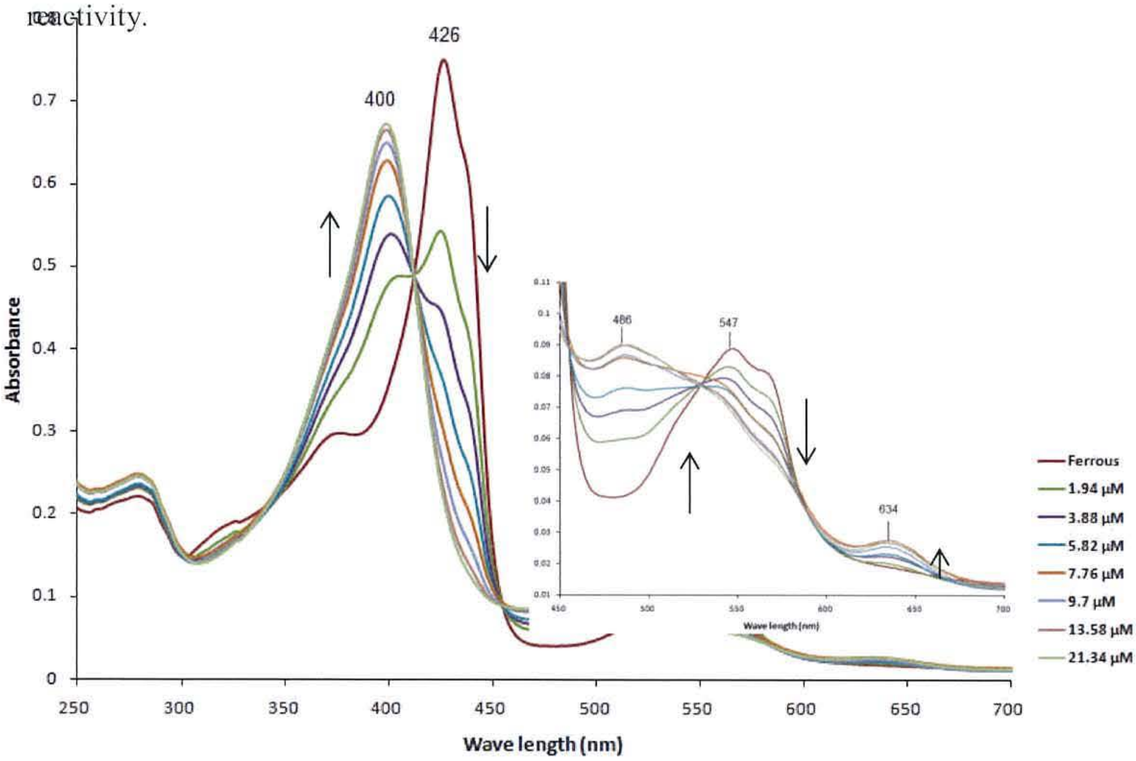


Figure 5.26: Electronic spectra of titration of 13.5 μM double mutant L16F/W56G with NO. The calculated extinction coefficient is 77.3 $mM^{-1}.cm^{-1}$.

5.4 Carbon monoxide titrations:

To probe the effect of distal pocket residues on CO binding and discrimination, the reduced forms of *AXCP* was titrated against the saturated solution of CO at pH 6. The additions were performed gradually to monitor the kinetics of CO binding. Carbon monoxide saturated solution was prepared as described in (section 3.7.4), and reduced forms were prepared as in section 3.7.1. The titration process has been illustrated in section 3.7.3.

5.4.1 CO titration with recombinant AXCP:

Addition of CO to ferrous *AXCP* generated the electronic absorption bands at 418, 535, and 565 nm (Figure 5.27) characteristic of a 6c-CO-His complex [11, 24] and another band at 636 nm. The spectrum shows that the binding affinity of CO to the haem iron is much lower in comparison to the binding affinity of NO. The magnetic circular dichroism (MCD) study of the CO-binding[24] has confirmed the relatively low CO-affinity for *AXCP* since unreacted ferrous form was still observed in the reaction mixture. Overall, the low CO-affinity observed herein and in the literature is explained by steric hindrance of the distal pocket. Since CO yields a 6c-CO-His complex in which CO is bound to the sixth, and his 120 to the fifth coordination site. CO has to accommodate itself into the crowded side. This accommodation requires rearrangements of distal pocket residues.[6]

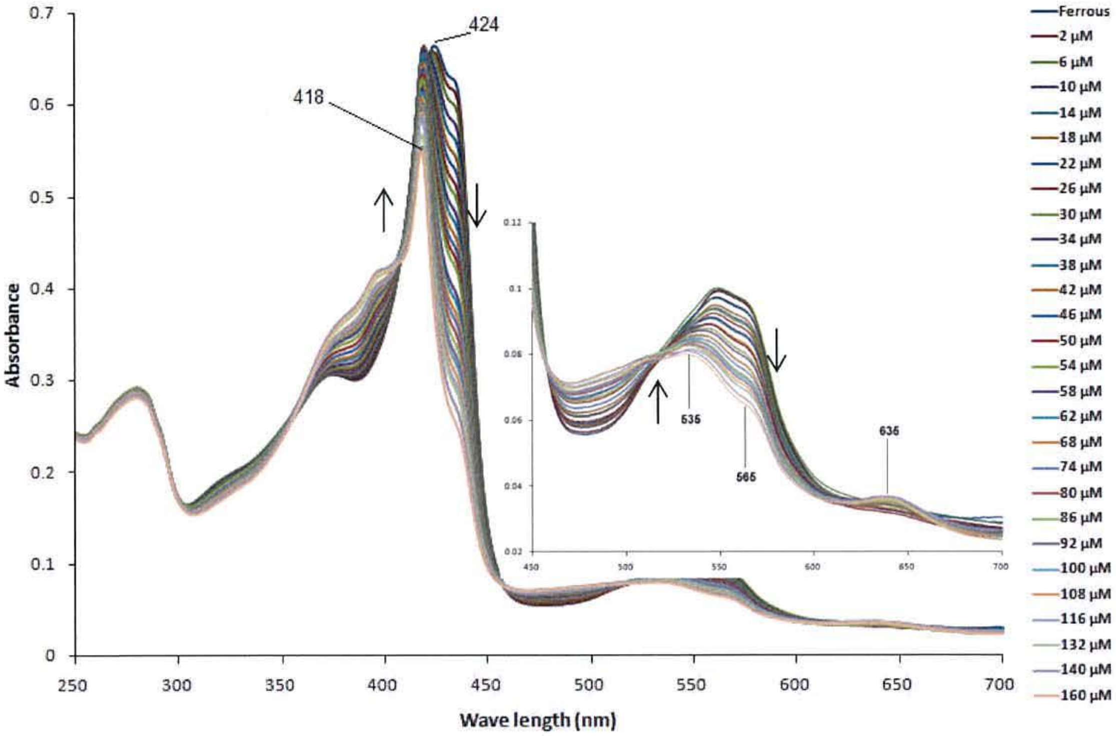


Figure 5.27: Electronic spectra of titration of 10.5 μM ferrous recombinant native *AXCP* with CO at pH 6.

The presence of the 418 nm peak (Figure 5.27) indicates the formation of 6c-CO-His adduct, however its relative intensity is much lower than that published for the 6c-CO-His adduct,[11] which should be some threefold more intense than the ferrous/ferric Soret peak. This is likely to be because CO decreases the oxidation potential of haem iron, therefore the haem iron undergoes oxidation upon the addition of further CO solution (personal communication with Professor Robert Eady). To confirm the idea of the decreased oxidation potential, the same experiment was repeated with excess reductant in the reaction mixture to keep the protein's oxidation potential high enough to stay reduced during the course of CO titration. For this purpose aliquots of 25 mM of ascorbate were added until the protein was fully reduced. Then aliquots of 1 mM CO were added to the reaction mixture and the spectra were recorded (Figure 5.28).

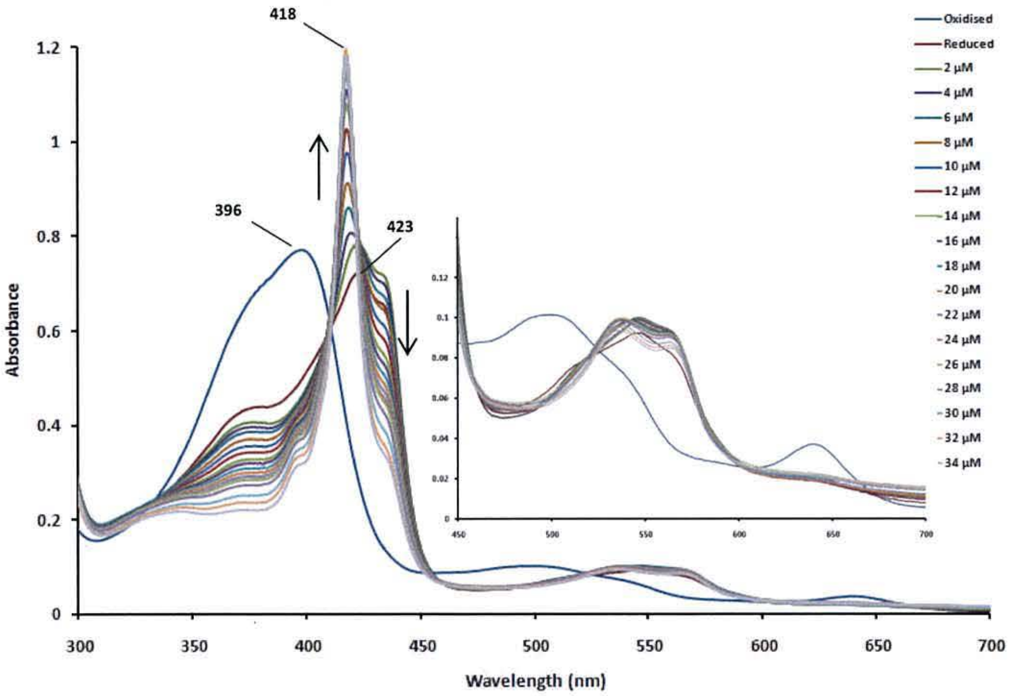


Figure 5.28: Electronic spectra of titration of 11.0 µM recombinant native *AXCP* with CO in the presence of excess reductant at pH 6. The calculated extinction coefficient is 97.8 $mM^{-1}.cm^{-1}$.

Indeed, a sharp and intense peak was observed. However, this experiment showed that CO has decreased the oxidation potential of the protein.

The crystal structure of the CO bound form of *AXCP* at 1.35 Å resolution [6] shows that CO binds to the distal pocket in a slightly bent form (Fe-C-O angle of 167°), with histidine bound to the proximal side of the haem iron producing a 6c-CO-His haem. When CO binds to the haem iron, the distal pocket undergoes a significant rearrangement. Leucine 16 residue which is located under the iron atom displaces to one side by 134° rotation around the $C\alpha$—C_β and induces the flattering of the porphyrin ring by pushing the pyrrole ring upward into the proximal pocket.

5.4.2 CO titration with L16F:

Addition of 1μL of 1 mM CO solution decreased the Soret band and the α/β region band intensity. Small additions caused relatively big changes in the band intensity, so it was necessary to add CO in very small aliquots to track the formation of possible 6c or 5c adducts. In contrast to the recombinant *AXCP*, addition of only 21 μM CO saturated the ferrous solution of L16F, implying a higher affinity to CO relative to the recombinant native *AXCP*. Bands at 402, 499 and 635 nm (Figure 5.29) have already been observed in the ferric 5c-His L16F except the shoulder of the Soret band at 380 nm and α band at 535 nm has vanished (Table 5.1). No band at 418 nm characteristic for 6c-CO-His complex was observed. This may be due to the formation of 5c-CO adduct instead of the expected 6c-CO-His adduct. Since the addition of CO may decrease the oxidation potential of cytochrome c' (personal communication with Professor Robert Eady), addition of CO may have oxidised the protein therefore the spectrum associated with oxidised 5c-Fe was observed.

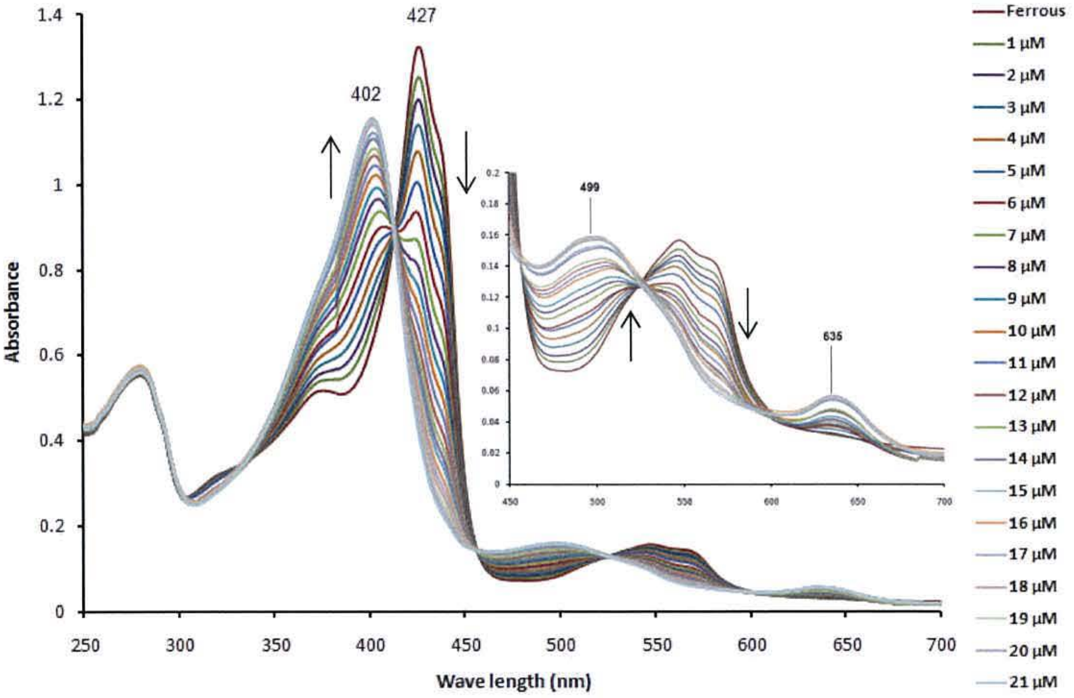


Figure 5.29: Electronic spectra of titration of 20.3 μM L16F with CO at pH 6.

To keep the oxidation potential high, excess ascorbate was added to the reaction mixture. Aliquots of 1 mM CO were added and the spectra were recorded (Figure 5.30). However, the spectra have exhibited very different features from Figure 5.29, but no band at 418 nm characteristic of the 6c-CO adduct was observed. There are some changes in the spectra; first, the change in the Soret peak position from 428 to 421 nm; and second, the loss in the intensity of the absorbance by about 50%.

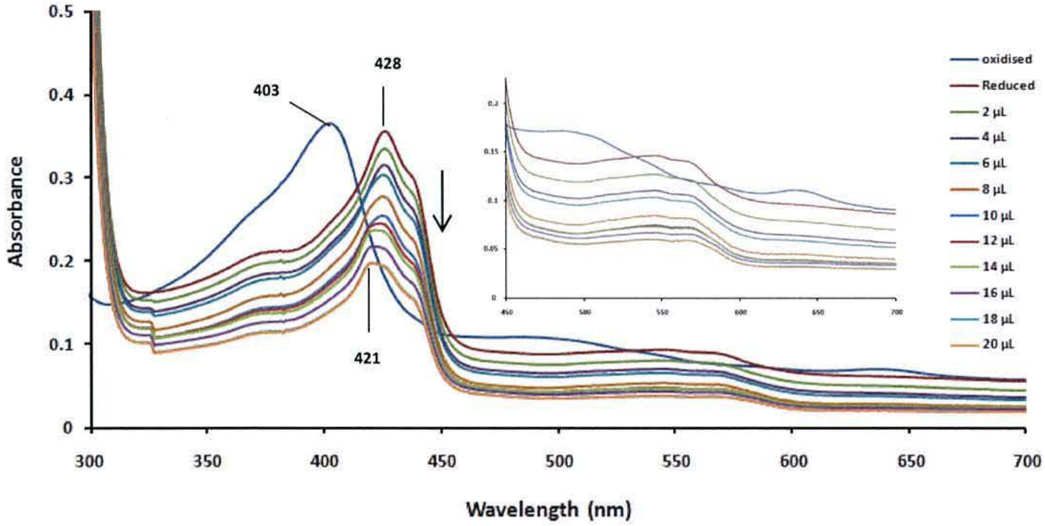


Figure 5.30: Electronic spectra of titration of 6.5 µM L16F with CO in the presence of excess reductant at pH 6. The calculated extinction coefficient is 21.1 $mM^{-1}.cm^{-1}$.

Although no crystal structure of L16F has been solved so far, the crystal structure of the double mutant L16F/W56G showed that the aromatic ring of the phenylalanine sits parallel to the haem plane in van der Waals contact. Apparently, this large, occluding hydrophobic moiety blocks the easy access to the distal pocket. Thus, binding of CO to the sixth coordinate is less favourable because of two reasons. First, because of the steric hindrance of the large and hydrophobic pocket; second, even if it does bind to the iron's sixth coordinate, it will be in a transient state and undetectable. Formation of the 6c-CO adduct should rotate the phenylalanine's side chain just like leucine 16 in the case of native recombinant *AXCP*[6] which is unlikely to happen in the case of phenylalanine because of its size and hydrophobicity. Therefore a 5c-CO adduct may have formed instead of 6c-CO adduct which indeed requires further investigations to confirm that.

5.4.3 CO titration with L16Y:

In contrast to L16F, CO binding to the haem iron of L16Y has showed a band at 418 nm which is characteristic to the 6c-CO-His adduct and the CO-affinity has decreased dramatically (Figure 5.31). The intensity of the band at 418 nm is much lower than that already been observed for the recombinant native (Figure 5.28) and from the previously reported spectra[11] owing to the decrease in the oxidation potential upon CO binding.

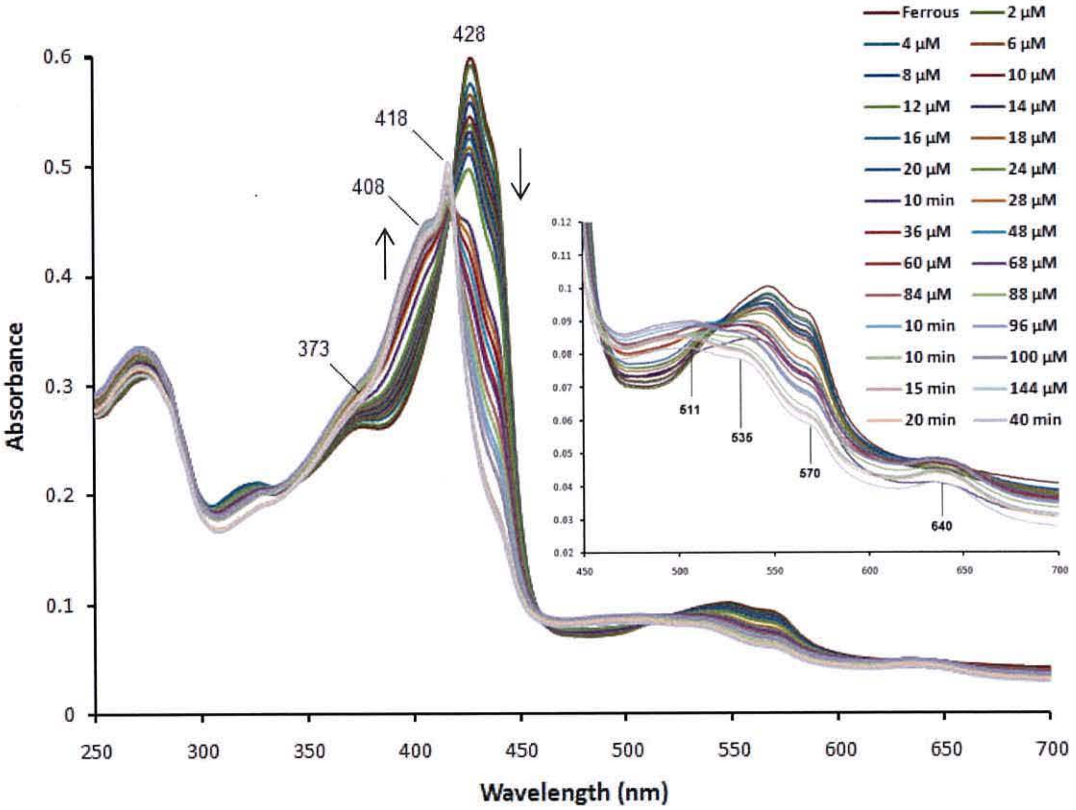


Figure 5.31: Electronic spectra of titration of 12.5 mM L16Y with CO at pH 6.

Apart from the decrease in the band intensity, addition of the first 24 μL of 1 mM CO did not affect the Soret and α/β band positions (Figure 5.31). After 10 minutes from the addition of 24 μM CO, a band at 418 nm has emerged with 428 nm band. The spectrum was measured after 15 and 20 minutes as well (data not shown) but no change was observed due to the equilibrium state. Addition of 88 μM CO made the Soret band at 428 nm to shift while a shoulder at 408 nm has appeared. At the end of the titration (with waiting for time intervals indicated on Figure 5.31) the band at 418 nm became more predominant and shoulders at 408 and 373 has emerged together with the appearance of bands at 511, 535, 570 nm at α/β region and the CT band at 640 nm.

Again to stabilise the ferrous form of this mutant, excess ascorbate as a reductant was added to the reaction mixture followed by addition of aliquots of 1 mM CO solution. The spectra (Figure 5.32) exhibited a sharp and intense band at 418 nm corresponding to 6c-CO adduct.

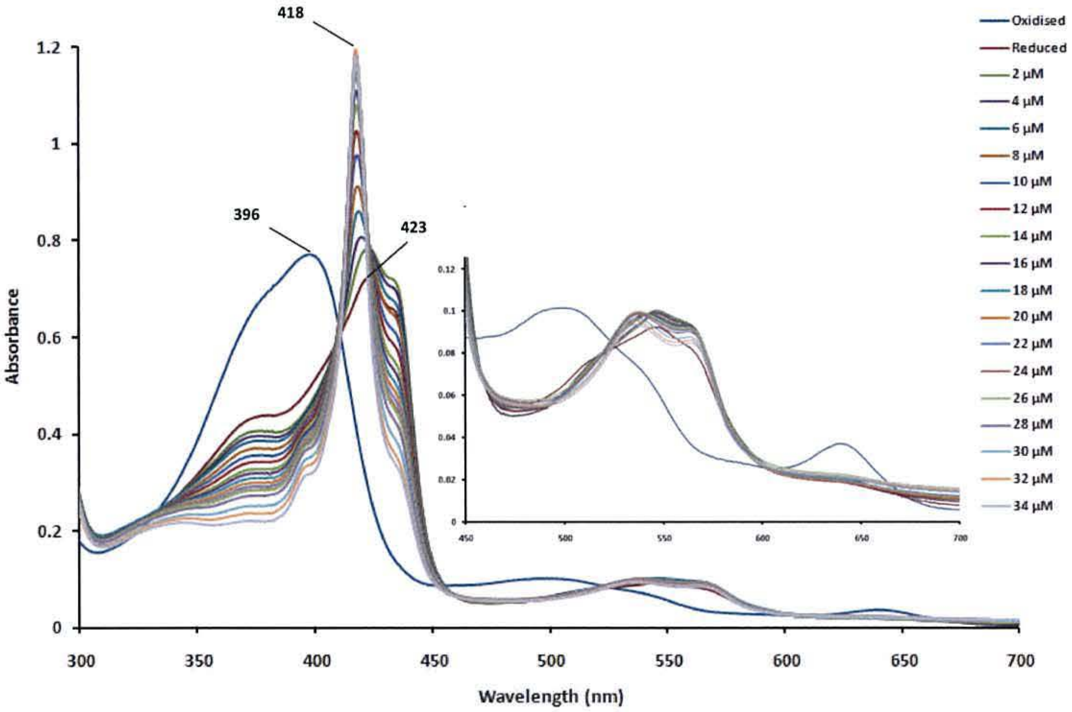


Figure 5.32: Electronic spectra of titration of 12.5 μM L16Y with CO in the presence of excess reductant at pH 6. The calculated extinction coefficient is 118.3 $mM^{-1}.cm^{-1}$.

The crystal structure of L16Y which has been resolved to (0.98 Å) resolution in this work has revealed that the phenol ring of tyrosine sits parallel to the porphyrin ring plane in van der Waals contact, and blocks the direct access to the sixth coordinate of iron, hence the binding of small molecules to the distal side of the iron must lead to the movement of the phenol ring of the tyrosine. In contrast to L16F, the phenol group of tyrosine could accommodate itself within the distal pocket and stabilise in its new position probably by the formation of hydrogen bond between the hydroxyl group of the side chain and a water molecule or surrounding residues.

5.4.4 CO titration with L16G:

Addition of CO to the ferrous form of L16G generated electronic absorption bands at 418, 536 and 564 nm (Figure 5.33), which has the same spectral pattern as the mutant after isolation and purification (Figure 5.5). This confirms that L16G was already 6c when isolated, and has been confirmed by its crystal structure at 1.18 Å resolution. As has been explained earlier in sections 4.1.2 and 4.2, for the titration purposes, any ligated ligand in its sixth coordination has been removed (see section 3.7.1) and the titrated against the CO. L16G showed a high CO-affinity, and readily bound to any available amounts of CO, so there was no need to wait for the reaction to be completed. This is due to the less crowded distal pocket and hence the sixth coordination of the iron haem is more exposed to the exogenous ligands.

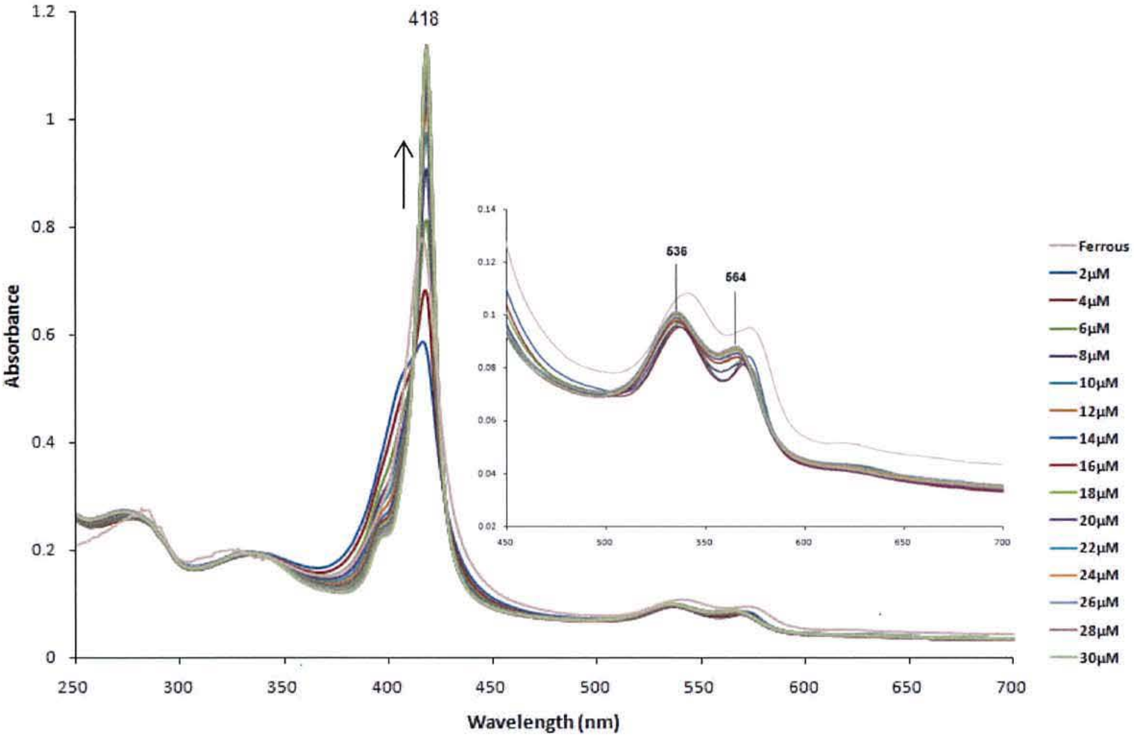


Figure 5.33: Electronic spectra of titration of 6.5 μM L16G with CO at pH 6. The calculated extinction coefficient is 121.5 $mM^{-1}.cm^{-1}$.

5.4.5 CO titration with W56G:

Addition of CO at very low concentrations changed the electronic spectra significantly (Figure 5.34) indicating high CO-affinity. Adding aliquots of 1 mM CO to the ferrous form of W56G made the Soret band at 423 nm to shift to 418 nm, while shoulders at both sides 369 and 432 nm remained constant with a big decrease in the band intensity. This confirmed the formation of the 6c-CO-His complex. Addition of another 1 μM resulted in a slight shift of the band to 417 nm with an increase in the band intensity. Upon further addition the band at 417 nm remained constant in intensity and the shoulders at 369 and 432 became less predominant. Bands at α/β region have shifted to 535 and 565 nm.

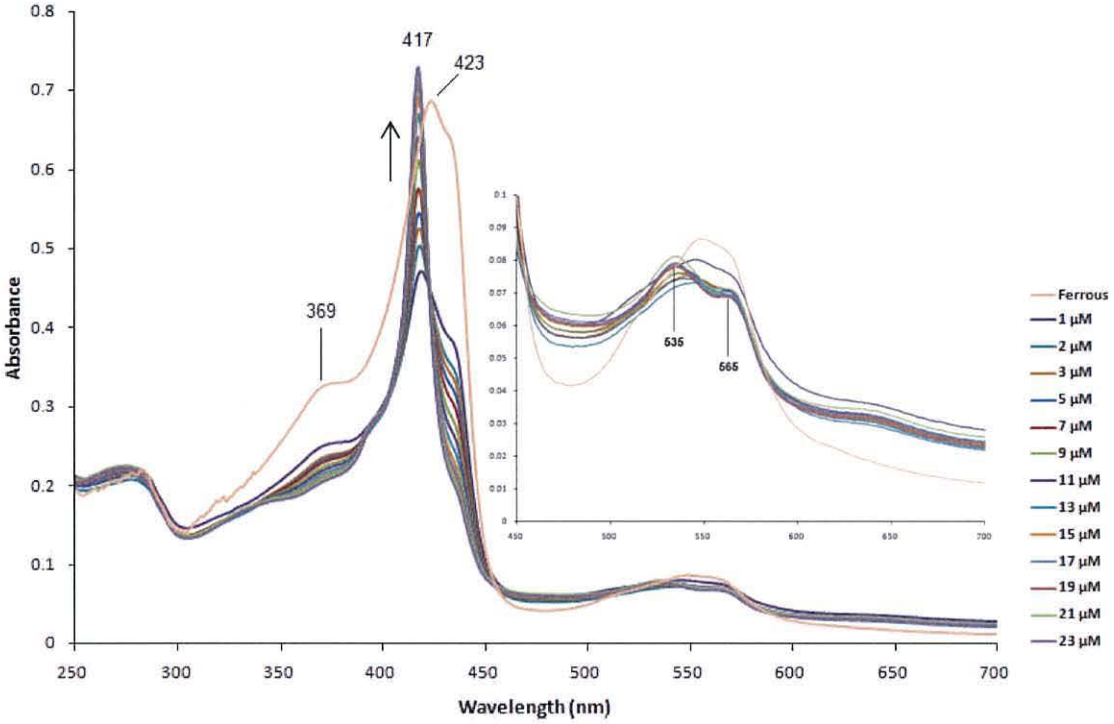


Figure 5.34: Electronic spectra of titration of 8.5 μM W56G with CO at pH 6. The calculated extinction coefficient is 98.7 $mM^{-1}.cm^{-1}$.

As was proposed, replacing W56 by G has increased the protein affinity and reactivity to CO. This high reactivity may be due to the direct exposure of distal site of the haem iron to the solvent when W56 was replaced by glycine. Another reason could be the change in the hydrophobicity and the steric hindrance of distal site of the haem.

5.4.6 CO titration with W56Y:

Addition of CO to ferrous W56Y produced the electronic absorption bands at 418, 535, and 565 nm characteristic of a 6c-CO-His complex and two shoulders at 379 and 399 nm (Figure 5.35). Addition of 1 μL of 1mM CO caused a significant decrease of the Soret band intensity at 425 nm. Adding 3 μL CO shifted the Soret band to 420 nm with a small decrease in the band intensity, while adding 7 μM CO produced a band at 418 nm characteristic to 6c-CO-His complex and the shoulder at 434 nm became less predominant. Further additions produced a sharp and intense band at 418 nm accompanied by the formation of another two shoulders at a lower wavelength 379 and 399 nm. Bands in the α/β region are the same as for W56G at 535 and 565 nm. An extra band at 640 nm seen in the spectra was not observed in W56G. These observations demonstrated a relatively high affinity of CO for W56Y compared to the recombinant *AXCP* but lower than W56G. This may be due to the possible hydrogen bonding between carbonyl oxygen of the CO and the hydroxyl group of tyrosine.

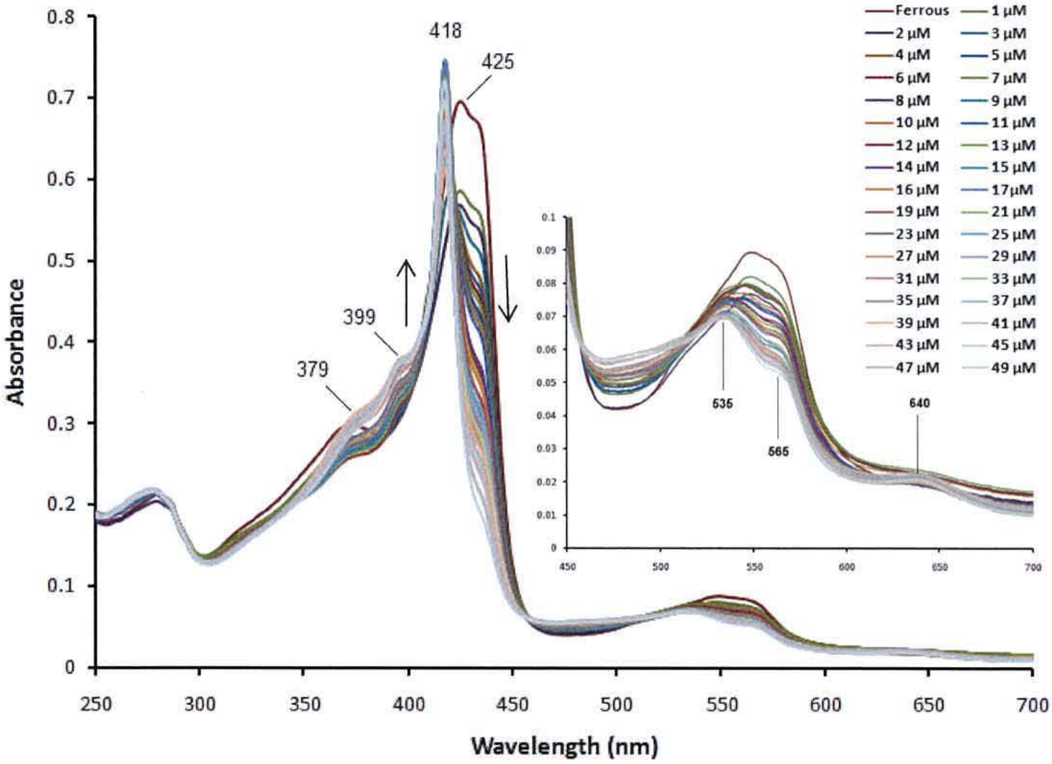


Figure 5.35: Electronic spectra of titration of 9.0 µM W56Y with CO at pH 6. The calculated extinction coefficient is 94.4 mM^{-1}.cm^{-1}.

5.4.7 CO titration with the double mutant L16F/W56G:

The CO-reactivity is low in the case of the double mutant L16F/W56G relative to L16F and recombinant *AXCP*. Except in terms of the band intensity, addition of 8 μM CO did not shift the Soret band (426 nm). Addition of 10 μL of 1mM CO caused a blue shift of the Soret band to 420 nm. The shoulder at 435 became less predominant. The band characteristic of the 6c-CO-His at 418 nm was detected at 15 μM CO together with a shoulder at 412 nm. Further additions caused the band at 418 nm to become less predominant. There was a shift of the 412 nm band by 1 nm upon addition of an extra 1 μM CO (Figure 5.36) until the band has centred at 401 nm. No further changes were detected upon addition of more than 39 μM CO (data not shown) and hence the titration was stopped and spectra produced were deleted for the clarity. At the level of 39 μM CO, bands at 401, 396, 565 and 633 nm were produced. These observations showed that the haem iron was undergoing the oxidation rather than the CO binding due to decrease in the oxidation potential upon CO addition (Personal communication with Professor Robert Eady).

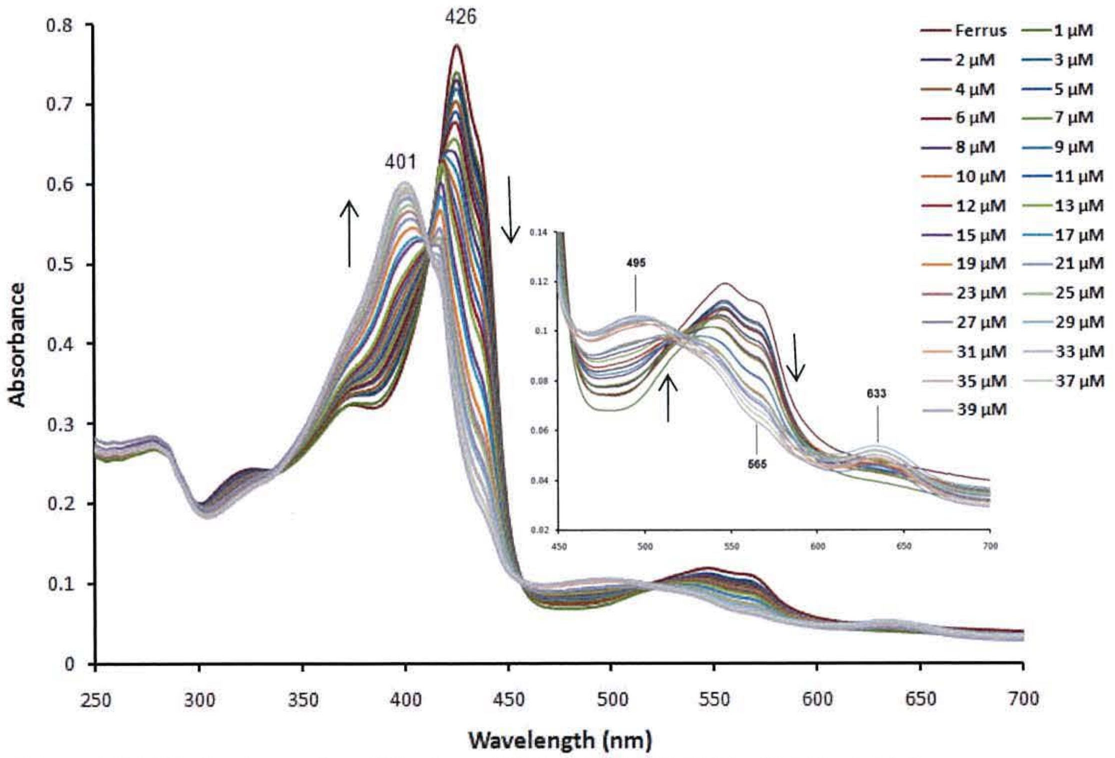


Figure 5.36: Electronic spectra of titration of 15.0 μM L16F/W56G with CO at pH 6.

In order to stabilise the ferrous form during the CO titration, excess ascorbate was added to the reaction mixture. Aliqouts of 1 mM CO solution were added and the spectrum was recorded after each addition (Figure 5.37). The band at 418 nm confirms the formation of the 6c-CO adduct. In contrast to the single mutant L16F, which does not bind CO, the double mutant L16F/W56G does bind the CO and formed a 6c-CO adduct.

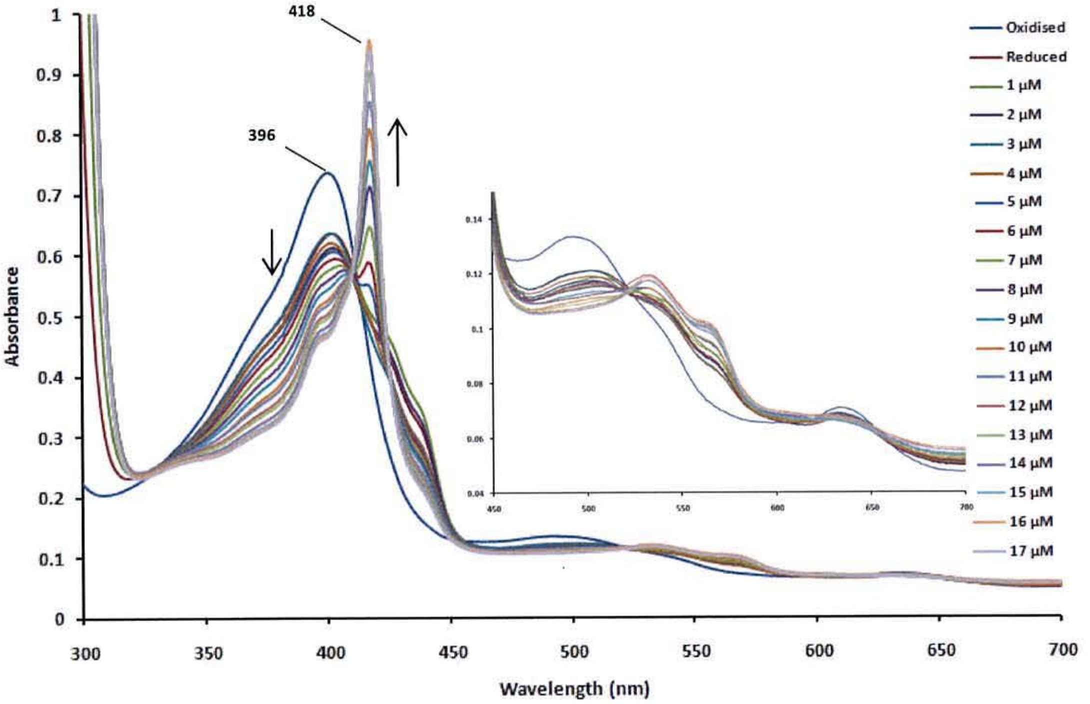


Figure 5.37: Electronic spectra of titration of 15.0 µM of L16F/W56G mutant with CO in the presence of excess reductant at pH 6. The calculated extinction coefficient is 62.77 $mM^{-1}.cm^{-1}$.

The crystal structure of this mutant (see chapter 6) solved at 0.96 Å explains why the double mutant does bind CO while the single mutant does not. Mutation of tryptophan 56 to a small residue glycine may have affected the formation of the 6c-CO adduct in two ways, first, by making more space for F16 to rotate upon the CO binding, second, by opening the proposed solvent channel (Figure 5.3).

5.5 Conclusions:

5.5.1 Reduction process:

A new method for the anaerobic reduction of cytochrome c' inside a cuvette has been described. The method was applied in experiments described in this thesis to reduce the native recombinant and its mutant proteins. It was very straightforward and showed its efficiency and reliability in the reduction process. This method can be adapted to reduce other metalloproteins without the need of the anaerobic chamber such as the glove-box.

5.5.2 Ferric and ferrous forms of recombinant and mutants:

At pH 6, ferric forms of the recombinant and mutants show a Soret band with λ_{max} range of 394 to 404 nm and a shoulder at lower wavelength ranged from 364– to 380 nm. The α/β band ranged from 494 to 500 nm and a shoulder from 535 to 548 nm. While the charge transfer band was at 640 nm for recombinant and all mutant proteins except for L16G which was at 631 nm indicating the presence of HS species.

Ferrous forms at the same pH revealed a distinctive electronic spectrum. Except the L16G mutant, the recombinant and other mutant proteins exhibited a split Soret band ranged from 423 to 426 nm. A shoulder was observed in the range from 432 to 437 nm, and the shoulder at lower wavelength became more pronounced and ranged from 369 to 375 nm. Both α and β bands exhibited a red shift with an increase in their intensities. The CT band has vanished in the ferrous forms suggesting the change in the spin state from high to low spin state.

Replacing leucine 16 by phenylalanine (**L16F**) and by tyrosine (**L16Y**) induced a red shift on the Soret band, while the α/β band remained unchanged, with a blue shift of the CT band in the case of L16F.

Replacing tryptophan 56 by glycine (W56G) and tyrosine (W56Y) revealed the same effect as for L16F and L16Y on the Soret maxima. W56G shifted the CT band to a lower wavelength while the shift was trivial in W56Y mutant. The β band shifted to a lower wavelength in W56G and to higher in W56Y mutant. The α band was shifted to a higher wavelength in W56Y mutant.

The electronic spectra of L16F/W56G mutant combined the effect of L16F and W56G on the same spectra. It has affected both the Soret maximum and the α/β band.

A very different electronic spectrum was observed when leucine 16 replaced by glycine (L16G). The spectrum showed a sharp and an intense band at 418 nm characteristic for 6c adducts. The α/β region showed the same features as for its ferrous form, while no CT band was observed. The coordination number and the spin state were confirmed by the crystal structure of L16G in its relaxing state resolved at 1.18 Å resolution. The structure showed that a water molecule was in the sixth coordination site of the haem, and the guanidinium group or Arg124 was perpendicular to the haem plane.

5.5.3 Ligand binding:

Even though, the behaviour of the recombinant and all mutants upon the NO and CO binding has been discussed under each relevant figure, a comparison between NO and CO binding to the haem iron and overall conclusion is summarised here:

Addition of CO to the recombinant, L16F, L16Y and the double mutant L16F/W56G proteins decreased the oxidation potential of the haem iron. consequently, instead of the CO complex formation, proteins were oxidised. Except for L16F, Stabilisation of ferrous forms has lead to the formation of 6c-CO adducts.. This stabilisation was achieved by preventing the decrease in the oxidation potential of proteins by keeping excess amount of the reductant during the CO titrations.

In contrast to the recombinant and all other mutants, L16F does not show any evidence for the 6c -CO adduct formation. Instead, the formation of 5c-CO has been proposed. Further investigations required to confirm this proposal. Relative to recombinant *AXCP*, the affinity of L16F towards NO remained the same and a 5c-NO has been produced.

Surprisingly, and in contrast to L16F, L16Y does bind CO and formed a 6c- CO adduct. The affinity to NO was also very low and led to the formation of 5c-NO adduct.

The W56G mutation may have provided a solvent channel and may also has made the distal face less crowded. The ferrous form was stable upon the CO addition, and a 6c-CO adduct was formed. The increase in the NO and CO affinity of the protein was observed. However, the affinity towards CO was more significant than to NO.

Relative to the recombinant *AXCP* protein, W56Y formed 5c-NO and 6c-CO adduct more rapidly. Similar to W56G, the oxidation potential does not decreased upon the CO addition, therefore ferrous forms were stable during the titration process. This may be due to the decrease in the distal pocket's hydrophobicity which may retain the solvent and hence diatomic molecules close to the distal face.

Providing an extra solvent channel by mutating W56 to G and replacing Leu 16 by a bulky and a hydrophobic amino acid F at the same time; L16F/W56G has changed the L16F behaviour to bind CO. This mutant has formed a stable 6c-CO adduct. This was attributed to the flexibility of F16 movements upon the CO binding to the distal site of the haem. the probably of providing an extra solvent channel to the distal haem site was another reason for the reactivity of this mutant. The electronic spectra showed the formation of 5c-NO adduct.

Very interestingly, L16G was isolated as a reduced protein. The crystal structure solved at 1.18 Å showed a water molecule is bound to its sixth coordination site. A powerful oxidising agent was used to oxidise and remove the bound water from the iron's sixth coordinate. More interestingly and in contrast to the recombinant and all other mutants, evidences of the 6c-Fe-NO adduct formation were observed. The change in the distal pocket environment has led to the formation of 6c-NO-Fe adduct. This confirms the idea that L16 enforces the breakage of Fe-His bond which in turn results in the formation of 5c-NO-Fe adduct not the 6c-NO-Fe adduct.

In general, these observations show the importance of L16 modulating ligand binding and discrimination of cytochrome c'. W56 mutations confirmed the critical role of this residue in the ligand binding and discrimination cytochrome c'.

5.6 Reference:

---

[1] Andrew C. R., George S. J., Lawson D. M., and Eady R. R. (2002) "Six- to five-coordinate heme-nitrosyl conversion in cytochrome c' and its relevance to guanylate cyclase" *Biochemistry* **41**, 2353-2360.

[2] Maltempo M. M. (1974) "Magnetic state of an unusual bacterial heme protein" *The Journal of Chemical Physics* **61**, 2540 – 2547.

[3] Fujii S., Yoshimura T., Kamada H., Yamaguchi K., Suzuki Sh., Shidara S., Takakuwa S. (1995) "Electronic paramagnetic resonance studies of ferric cytochrome c' from photosynthetic bacteria" *Biochimica et Biophysica Acta* **1251**, 161-169.

[4] Weber P. C. (1982) "Correlation between structural and spectroscopic properties of the high spin heme protein cytochrome c'" *Biochemistry* **21,** 5116 – 5119.

[5] Weiss R., Gold A., and Terner J. (2006) "Cytochromes c': biological model for the S = 3/2, 5/2 spin-state admixture" *Chemical Review* **106,** 2550-2579

[6] Lawson D., Stevenson C. E. M., Andrew C. R., Eady R. R. (2000) "Unprecedented proximal binding of nitric oxide to heme: Implications of guanylate cyclase" *The EMBO Journal* **19,** 5661-5671.

[7] Pinakoulaki E., Yoshimura H., Daskalakis V., Yoshioka S., Aono S., and Varotsis C. (2006) "Two ligand-binding sites in the O_2-sensing signal transducer HemAT: Implications for ligand recognition/discrimination and signalling" *Proc. Natl. Acad. Sci. USA* **103,** 14796-14801.

[8] Spiro T. G (1983) "The resonance Raman spectroscopy of metalloporphyrins and haemoproteins" in Iron Porphyrins, Part two, Edited by Leve A. and Gary H. B., Adison Wesley, London Amestrdam, 89-159.

[9] Shidara S., Iwasaki H., Yoshimura T., Suzuki S., and Nakahara A. (1986) "Purification and some properties of cytochrome c' from a strain of *Achromobacter xylosoxidans*" *Journal of Biochemistry* **99**, 1749-1752.

[10] Strekas T. C., and Spiro T. G. (1974) "Resonance-raman evidence for anomalous heme structures in cytochrome c' from Rhodopseudomonas palustris" *Biochimica et Biophysica Acta* **351**, 237-245.

[11] Andrew C. R, Green E. L., Lawson D. M., and Eady R. R. (2001) "Resonance Raman studies of cytochrome c' support the binding of NO and CO to opposite sides of the heme: Implications for ligand discrimination in heme-based sensors" *Biochemistry* **40,** 4115-4122.

[12] Barbieri S., Murphy L. M., Sawers R. G., Eady R. R., and Hasnain S. S. (2008) "Modulation of NO binding to cytochrome c' by distal and proximal haem pocket residues" *Journal of Biological Inorganic Chemistry* **13**, 531-540.

[13] Barbieri S. (2007) "Mutagenesis, spectroscopy and crystallographic studies of recombinant cytochrome c' from *Alcaligenes xylosoxidans*" *PhD thesis*, University of Wales, Bangor.

[14] Smulevich G (1998) "Understanding heme cavity structure of peroxidises: comparison of electronic absorption and Resonance Raman spectra with crystallographic results" *Biospectroscopy* **4**, S3-S17.

---

[15] Yoshimura T., Suzuki S., Nakahara A., Iwasaki H., Masuko M., Matsubara T. (1985) "Identification of heme axial ligands of cytochrome c' from *Alcaligenes sp.* N.C.I.B 11015" *Biochimica et Biophysica Acta* **831,** 267-274.

[16] Lancaster J. Jr (1996) "Nitric oxide: principles and actions" 1-18. Academic Press, San Diego, USA.

[17] Yoshimura T., Suzuki S., Nakahara A., Iwasaki H., Masuko M., and Matsubara T. (1986) "Spectral properties of nitric oxide complexes of cytochrome c' from *Alcaligenes* sp. NCIB 11015" *Biochemistry* **25,** 2436-2442.

[18] Zhao Y, Brandish P. E., Ballou D. P., and Martella M. A (1999) "A molecular basis for nitric oxide sensing by soluble gualnylate cyclase" Proc. Natl. Acad. Sci. USA **96**, 14753-14758.

[19] Marti M. A., Capece L., Crespo A., Doctorovich F., and Estrin D. A. (2005) "Nitric oxide interaction with cytochrome c' and its relevance to guanylate cyclase. Why does the iron histidin bond break?" *Journal of American Chemical Society* **127**, 7721-7728.

[20] Yoo B., Lamarre I., Martin J., Andrew C. r., Nioche P., and Negrerie M. (2008) "Understanding the NO-sensing mechanism at molecular level" *EU-Korea Conference on Science and Technology, Ed: Yoo S. D* **124,** 517-524.

[21] Lee B., Usov O. M., Grigoryants V. M., Myers W. K., Shapleigh J. P., and Scholes C. P. (2009) "The role of arginine 127 at the proximal NO-binding site in determining the electronic structure and function of 5-coordinate NO-Heme in cytochrome c' of *Rhodobacter sphaeroides*" *Biochemistry* DOI 10.1021/bi900833f.

[22] Yoshimura T., Suzuki S and Nakahara A. (1996) "Spectroscopic characterisation of nitrosylheme in nitric oxide complexes of ferric and ferrous cytochrome c' from photosynthetic bacteria" *Biochemica et Biophysica Acta* **1292**, 39-46.

[23] Ren Z., Meyer T., and McRee D. E. (1993) "atomic structure of cytochrome c' with unusual ligand-controlled dimer dissociation at 1.8 Å resolution" *Journal of Molecular Biology* **234**, 433-445.

[24] Suzuki S., Nakahara A., Yoshimura T., Iwasaki H., Shidara S., and Matsubara T. (1988) "Spectral properties of carbon monoxide or cyanide complexes of cytochromes c' from denitrifying bacteria" *Inorganica Chimica Acta* **153,** 227-233.

Chapter 6: The crystal structures

6.1 Introduction:

In order to understand how proteins function at the atomic level, we need to know their structures. The use of x-ray crystallography to find out their structures has been the most effective technique.[1] The first step in the determination of protein structure via x-ray crystallography is growing high quality crystals which can diffract x-rays to high-enough resolution to see the detailed atomic structure. Growing such crystals has always been a challenge and the bottleneck to structure determination.[2] The major problem of crystallisation is either the failure to grow crystals at all or, getting crystals that are not of high enough quality to enable structure determination (Table 2.1). Proteins usually can be made to form crystals when exposed to appropriate conditions, however, the optimisation of these conditions still remains the challenging step in getting single diffracting crystals.[3]

In this work, the x-ray quality crystals have been grown and data were collected at atomic resolution. The crystal structure of recombinant native cytochrome c′ is solved at resting state (oxidised) at 1.15 Å resolution. The crystal structure of three mutants are presented here as well; L16G at 1.18 Å, L16Y at 0.98 Å, and the double mutant L16F/W56G at 0.96 Å resolution.

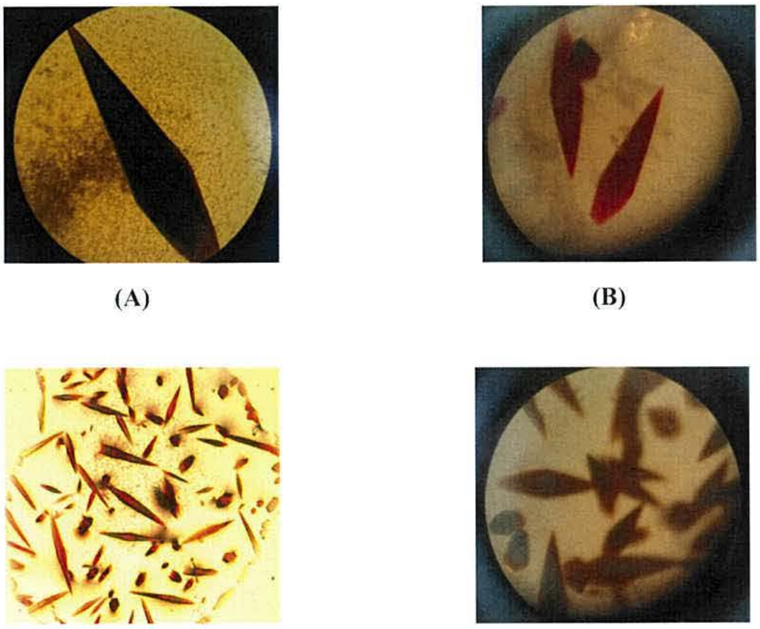
The diffraction data are regarded as atomic if the resolution is at 1.2 Å or lower, and at this limit a majority of intensities are above the 2σ level.[4] The 1.2Å limit is shorter than the average bond length within the crystallised molecule, so peaks in the Fourier synthesis corresponding to individual atoms do not overlap.

At atomic resolution it is possible to determine structures with the highest accuracy, which enables very precise modelling. At that resolution, various subtle features become apparent and, although these may have been visible even at lower resolution, they can be modelled with higher confidence at atomic resolution. For example alternative conformations can be modelled at even lower resolution than 2 Å, but they become more clearly distinguishable at atomic resolution. In some cases many internal hydrogen atoms can be located which is vital in order to better understand enzymatic mechanisms. However, hydrogens cannot be refined as independent atoms; they usually are refined as 'riding' on their parent atoms. These features enable probing the chemistry of the active site of enzymes, metal prosthetic groups, and ligand binding via hydrogen bonds.[5] Although in recent years the number of x-ray crystal structures of proteins refined at atomic resolution has increased, the number of metalloproteins refined at that resolution is still small, owing to the variety of factors including the diffraction limit because of the crystalline disorder[6].

6.2 Crystallisation:

Proteins of recombinant AXCP, L16F, L16G, L16Y, W56G, W56Y, and the double mutant L16F/W56G were purified as described in section 3.5.4, and concentrated to ca 40 mg/mL as described in section 3.5.5. Protein concentrations were determined using UV absorption measurements at the Soret band of recombinant native protein using molar extinction coefficient at the Soret band of 80 $mM^{-1}.cm^{-1}$, [7] and mutant concentrations were calculated relative to the recombinant native peak at 280 nm.

Crystals were obtained by the hanging drop vapour diffusion technique (section 2.3.2.2) against a reservoir containing ammonium sulphate buffered with 100 mM HEPES at pH 7.5. Protein and ammonium sulphate concentrations were screened for optimisation. Protein concentrations screened ranged from 10 - 30 mg/mL, and the reservoir concentration ranged from 1.4 – 2.6 M at 21 °C. However, crystals grew at protein concentrations higher than 16 mg/mL in ammonium sulphate concentrations greater than 1.8 M at 21 °C after 48 hours; the optimum condition for getting single, large, and less fragile crystals was 26 mg/mL of protein, 2.2 M ammonium sulphate at 21 °C in 100 mM HEPES buffer pH 7.5. Figure 6.1 shows crystals of recombinant and L16Y, L16G and L16F/W56G. All proteins crystallised in space group $P6_522$, with approximate cell parameters $a = b = 53$ Å, $c = 182$ Å and $\alpha = \beta = 90°$, $\gamma = 120°$ when measured at 100 K.



(A) (B)

(C) (D)

Figure 6.1: Crystals of recombinant (A), L16G (B), L16Y (C), and L16F/W56G (D) grown at 26 mg/mL of proteins and 2.2 M ammonium sulphate in 100 mM HEPES buffer pH 7.5 at 21 °C

6.3 Crystallographic Data Collection:

For each protein (native or mutant) a single crystal was fished from the drop using a suitable size fiber cryo-loop, and soaked for about 1 minute into the cryoprotectant solution which was made from the mother liquor with 40 % sucrose, and immediately mounted onto the goniometer where it was flash cooled to 100 K under the nitrogen stream.

Data from crystals of oxidised recombinant AXCP were collected on beamline 10.1 at a wavelength of 0.98 Å at Synchrotron Radiation Source (SRS) at Daresbury – Cheshire in the UK using the Mar225 CCD detector. Data from crystals of the mutants L16G and L16Y were collected on beamline 08ID-1 at Canadian Light Source (CLS) at a wavelength of 0.953 Å using the same detector type as in SRS. The crystallographic data from the double mutant L16F/W56G were collected at 0.725 Å on beamline XO6SA at Swiss Light Source (SLS) using Pilatus-6M detector. Details of data collection strategies and parameters are summarised in (Table 6.1).

In order to collect data at atomic resolution, it is crucial to collect a complete dataset at both high and low resolution limits. Data at SRS and CLS were collected in three different passes; high, medium and low resolution limits to accurately estimate the strongest intensities at atomic resolution. The strong, low resolution reflections are necessary for the electron density map interpretation and to modulate the Fourier maps more significantly. The weak, highest resolution reflections contain information about the finest details of the structure.[8] Since the reflection intensities of the high resolution limit were very weak in the case of recombinant native AXCP, the exposure time was increased with no attenuation in the beam intensity (Table 6.1), while low resolution data were collected with the beam attenuation of 80% by slitting down. The crystal to detector distance was changed from 250 and 135 mm at low and intermediate resolution respectively to 65 mm at high resolution limit with a lower oscillation range. For L16G and L16Y, the exposure time remained the same but with beam attenuation (using aluminium foil) of 75 and 45% in low and medium resolution passes respectively. The crystal to detector distance was changed from 300 mm to 200 mm and then to 100 mm when the data were collected at low, medium and high passes respectively. Data collection parameters and strategies were summarised in (Figure 6.1). Finally for the double mutant L16F/W56G the data were collected at atomic resolution at the SLS. The data collection process at atomic resolution was achieved in only one pass due to the use of Pilatus-6M detector in the beamline XO6SA which has a larger dynamic range compared to the CCD detectors.[9]

Parameters	Recombinant (oxidised)			L16G				L16Y				L16F/W56G
	High-pass	*Med-pass*	*Low- pass*	*High-pass*	*High-pass FLAT*	*Med-pass*	*Low- pass*	*High-pass*	*High-pass. 20° tilted*	*Med-pass*	*Low- pass*	*One pass*
Crystal-detector's distance (mm)	65	135	250	100	100	200	300	100	100	200	300	250
Resolution limit (Å)	0.98	1.403	2.33	0.929	1.197	1.903	2.707	1.173	1.144	1.903	2.707	0.96
Oscillation range (ψ°)	0-60	0-60	0-60	0-60	0-60	0-60	0-60	0-60	0-60	0-60	0-60	0-60
Oscillation angle ($\Delta\psi$°)	0.1	0.5	1	0.2	0.2	0.5	1	0.2	0.2	0.5	1	0.5
Number of images	600	120	60	450	300	120	60	300	300	120	60	120
Counting time/image (sec./image)	10	1	1	1	1	1	1	1	1	1	1	0.25
Beam attenuation %	0	0	80	0	0	45	75	0	0	45	75	0
Total time (sec.)/pass	6000	120	60	450	300	120	60	300	300	120	60	30
X-ray wavelength (Å)	0.98			0.953				0.953				0.725
Detector	Mar CCD			Mar CCD				Mar CCD				Pilatus 6M
Beamline	10.1 Synchrotron Radiation Source (SRS) – UK			08ID-1 Canadian Light Source (CLS) - Canada				08ID-1 Canadian Light Source (CLS) - Canada				XO6SA at Swiss Light Source (SLS) - Switzerland

Table 6.1: Summary of data collection strategy and parameters. In order to collect a complete dataset at highest possible resolution, data collected at CLS &SRS were collected at three different passes; high, medium and low resolution. To get even higher resolution, the detector was tilted at 20° in the case of L16G and L16Y mutants. Data collected at SLS using Pilatus detector did not necessitate collection at different passes.

A sample of AXCP diffraction pattern is shown in (Figure 6.2).

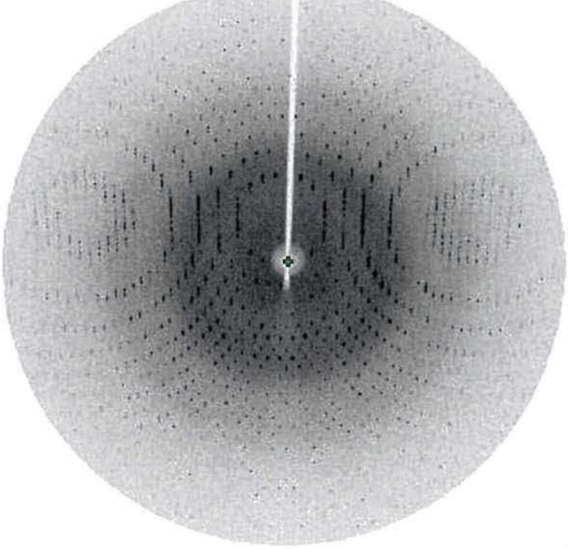


Figure 6.2: X-ray diffraction pattern of recombinant cytochrome c′ from *Alcaligenes xylosoxidans*, which has crystallised in space group $P6_522$. Mutants L16G, L16Y and the double mutant L16F/W56G showed similar quality diffraction patterns and crystallised in the same space group.

6.4 Data processing:

The x-ray diffraction intensities of recombinant and all other mutants were indexed, refined and integrated with *MOSFLM* software.[10] Detailed data processing with *MOSFLM* has been discussed in section 2.5.

The integrated data in the MOSFLM MTZ file was merged and scaled with the *SCALA* program,[11] part of the CCP4 program suite.[12] The temperature factor (B-factor) was estimated using the *TRUNCATE* program[13] via the Wilson plot[14] part of the CCP4 program suite. The number of molecules in the asymmetric unit and its solvent content were estimated by calculating the Matthews coefficient using the CCP4 program suite. In all datasets, the outer resolution limits were based on the data completeness and the mean *I/σ(I)* ratio.

6.4.1 Data processing results:

6.4.1.1 Oxidised recombinant AXCP:

The resulting dataset of oxidised recombinant AXCP contained a total of 692928 reflections of which 53237 were unique. The overall completeness of the data set in the range of 41.16 – 1.15 Å was 97.0% with a completeness of 94.4% in the outer resolution shell (1.18 – 1.15 Å) (from this point numbers in parenthesises represent the outer shell). The multiplicity was 13.0 (7.3) and the R_{merge} was 11.8 (57.0) %. The B-factor calculated from Wilson plot using *TRUNCATE* program was 6.4. The crystal contains one molecule in the asymmetric unit (that is half of the dimer), with a solvent content of 53.8 % (v/v). Data collection and processing details are summarised in Table 6.2.

Data collection	Oxidised
Resolution range (Å)	41.16 - 1.15
Outer Shell (Å)	(1.18 – 1.15)
Space group	$P6_522$
Unit cell parameters (Å)	53.4, 53.4, 180.91
Total No. of reflections	692928 (27534)
No. of unique reflections	53237 (3761)
Multiplicity	13.0 (7.3)
R_{merge} (%)	11.8 (57.0)
Completeness (%)	97.0 (94.4)
Mean I/σ(I)	16.7 (3.10)
Solvent content (% v/v)	53.78
Wilson B-factor (Å^2)	6.4

Table 6.2: Data collection and processing statistics of the recombinant AXCP in oxidised states. Numbers in parenthesises represent statistics of the outer resolution shell.

6.4.1.2 L16G and L1Y and the double mutant L16/W56G

For the L16G mutant, a total number of 266651 (27328) reflections were collected, of which 46403 (6573) were unique in the resolution range of 45.64 – 1.18 Å (1.24 – 1.18 Å) from a total of 930 images (Figure 6.1). The data completeness was 89.3 % (88.5 %) with the mean *I/σ(I)* of 13.9 (2.7) with multiplicity of 5.7 (4.2). The overall R_{merge} value was 6.6% and 40.1% in the outer shell. The temperature factor estimated via the Wilson plot was 9.4 $Å^2$.

In the case of the L16Y mutant, 790 images were recorded in three different passes. In the resolution range of 41.34 – 1.08 (1.04 – 0.98), 462227 (30980) reflections were recorded, 60511 (5853) of them were unique. The completeness was 90.4 (61.4) % with the mean *I/σ(I)* of 21.6 (5.9) and the multiplicity was 7.6 (5.3). The data showed R_{merge} value of 5.8% (22.4%). The B-factor was 8.05 $Å^2$ calculated as mentioned above.

The x-ray data of the double mutant L16F/W56G were collected at the SLS using a Pilatus detector which does not necessitate data collection in different passes. Only 120 images were collected in the resolution range of 32.29 – 0.96 Å (1.01 – 0.96 Å). A total number of 451712 (37094) reflections were collected of which 90854 (11479) were unique. The data were 97.5 % (86.6 %) complete. The mean *I/σ(I)* was 23.6 (2.7) with multiplicity of 5.0 (3.2). The Wilson B-value was 5.5 $Å^2$.

Crystals of all mutants contain one molecule in the asymmetric unit. They had a solvent content of 54.2 and 54.3, and 53.4 % (v/v) in the case of L16G, L16Y, and the double mutant L16F/W56G respectively.

Data are summarised in (Table 6.3).

Data collection	**L16G**	**L16Y**	**L16F/W56G**
Resolution range (Å)	45.64 – 1.18	41.34 –0.98	32.29 – 0.96
Outer Shell (Å)	(1.24 – 1.18)	(1.04 – 0.98)	(1.01 – 0.96)
Space group	$P6_522$	$P6_522$	$P6_522$
Unit cell parameters (Å)	53.4, 53.4, 182.39	53.5, 53.5, 181.59	53.1, 53.1, 181.2
Total No. of reflections	266651 (27328)	462227 (30980)	451712 (37094)
No. of unique reflections	46403 (6573)	60511 (5853)	90854 (11479)
Multiplicity	5.7 (4.2)	7.6 (5.3)	5.0 (3.2)
R_{merge} %	6.6 (40.09)	5.8 (22.4)	5.8 (39.0)
Completeness %	89.3 (88.5)	90.4 (61.4)	97.5 (86.6)
Mean I/σI	13.9 (2.7)	21.6 (5.9)	23.6 (2.7)
Solvent content %	54.18	54.25	53.41
Wilson B-factor $Å^2$	9.4	8.05	5.5

Table 6.3: Data collection and processing statistics of L16G and L16Y and L16F/W56G mutants. Numbers in parentheses represent statistics of the outer shell.

6.5 Structure solution and refinements:

6.5.1 Recombinant native AXCP:

In order to obtain the initial phases for the diffraction data of oxidised recombinant AXCP, molecular replacement was used (see section 2.8.1). The *MolRep* program[15] as part of the CCP4 program suite was employed for that purpose. The crystal structure of cytochrome c′ at resolution of 2.05 Å (PDB code 1E83)[16] was used as the search model. Data between the resolution range of 41.16 to 3.0 Å were used. Since the asymmetric unit contains one molecule, only one monomer was searched for. This gave a solution with an R_{factor} of 42.7% and a correlation coefficient of 56.6%. The structure then refined using *Refmac5* program[17] (see section 2.8.2).

To refine the structure, ten cycles of maximum likelihood rigid body refinement by *Refmac5* were performed and resulted in an R_{factor} and R_{free} of 38.7%. At this stage 5% of reflections were omitted from the refinement to calculate the R_{free} to monitor the progress of refinement and to check that the model is not over-fitted. Ten cycles of restrained refinement (positional and individual isotropic temperature factors) in the resolution range of 41.16 to 3 Å with the restraints weighting factor of 0.03, yielded an R_{factor} of 24.5% and R_{free} of 29.8%. At this stage, and prior to further cycles of refinement, the model was rebuilt using the molecular graphics program *Coot.*[18] Data up to 2 Å were then added to the refinement which increased the number of reflections used to 10386, the matrix weighting term was increased to 0.15, and R_{factor} and the R_{free} were increased to 26.0% and 30.7% respectively. Further restrained refinements were performed. When the solvent and all hydrogen atoms were added, the R_{factor} improved to 13.9%, the R_{free} to 18.5%, and the correlation coefficient to 95.8%, with the weighting term set at 0.2. The program ARP/wARP,[19] part of the CCP4 program suite, was used to identify solvent peaks and hence addition of solvent molecules into those peaks. More data were added gradually, refined and manually rebuilt in *Coot* as described before. After each data addition, increases in the R values were observed, but they were decreased by further refinement and increasing the matrix weighting term. Anisotropic displacement parameters (ADPs) refinement was performed after the addition of all data up to 1.15 Å. At this resolution, 50656 reflections were used. The weighting term of 4.0 was used. Alternative conformations were modelled during building with *Coot* and in some regions where residues were poorly defined in the electron density map, a lower occupancy was given to their side chains.

At the end of refinement, 352 water molecules and one sulphate (SO_4^{2-}) ion were added. 38 residue side chains were modelled as double conformations. The final R_{factor} was 12.6% and the R_{free} was 14.9%. The small difference between the R_{factor} and the R_{free} indicates that the model was not under determined.

6.5.2 Mutants:

To solve the crystal structure of L16G, L16Y and the double mutant, the same processes as described above were followed. The only exception was the use of the fully refined crystal structure of oxidised recombinant AXCP at 1.15 Å as a search model in the case of all mutants. In all cases, 5% of the used reflections were set aside to calculate the R_{free} as a parameter to monitor the refinement process in order not to over-fit the model.

MolRep[15] was employed to find the out the initial phases using the fully refined oxidised recombinant AXCP. Again only one monomer was searched for in all cases.

The weighting term was set at 0.03. The solution gave the $R_{factors}$ of 40.6, 39.3 and 42.8% for L16G, L16Y and L16F/W56G respectively and correlation coefficients of 64.7, 64.5 and 58.4%, again respectively. Subsequently 10 cycles of restrained refinement were performed for all three datasets up to 3 Å with individual isotropic temperature factors, and the same weighting term of 0.03 resulted in decrease of the R_{factor} to 25.5, 25.8 and 26.4% with R_{free} of 29.5, 30.9, and 31.7% for L16G, L16Y and the double mutant L16F/W56G respectively.

All side chains were adjusted to the electron density maps and the mutated residues were changed appropriately using the graphic interface program *Coot.*[18] In the case of L16G, leucine was omitted and replaced by glycine in the model. In the case of L16Y, there was a clear blob in the position number 16 indicating the mutation of leucine to tyrosine. For the double mutant L16F/W56G there was an obvious change in the electron density map at mutation sites, with leucine 16 replaced by phenylalanine and tryptophan 56 replaced by glycine.

The minimum resolution limit was increased to 2 Å and the matrix weighting term to 0.3 with further isotropic temperature factor refinement. Initial increases in R values were noticed as more reflections were added. Side chains were fitted to the electron density using the *Coot* program and water molecules were added using ARP/wARP, and the models were rebuilt again. This resulted in the drop of R values. At this reolution the $R_{factors}$ dropped to 17.1, 18.5 and 19.2% for L16G, L16Y and the double mutant respectively and R_{frees} to 19.1, 21.3, and 25.6% again respectively. The Root Mean Square Deviation for bonds (RMS bond) was 0.015 Å for L16G, 0.016 Å for the double mutant and 0.01 Å for L16Y. The minimum resolutions were increased stepwise and models were refined to their final resolution. In each step R values and RMS bonds were monitored. At a resolution of 1.5 Å the structures were refined with riding hydrogen atoms, which led to a decrease in R values. When all data were added, structures were refined using anisotropic temperature factors.

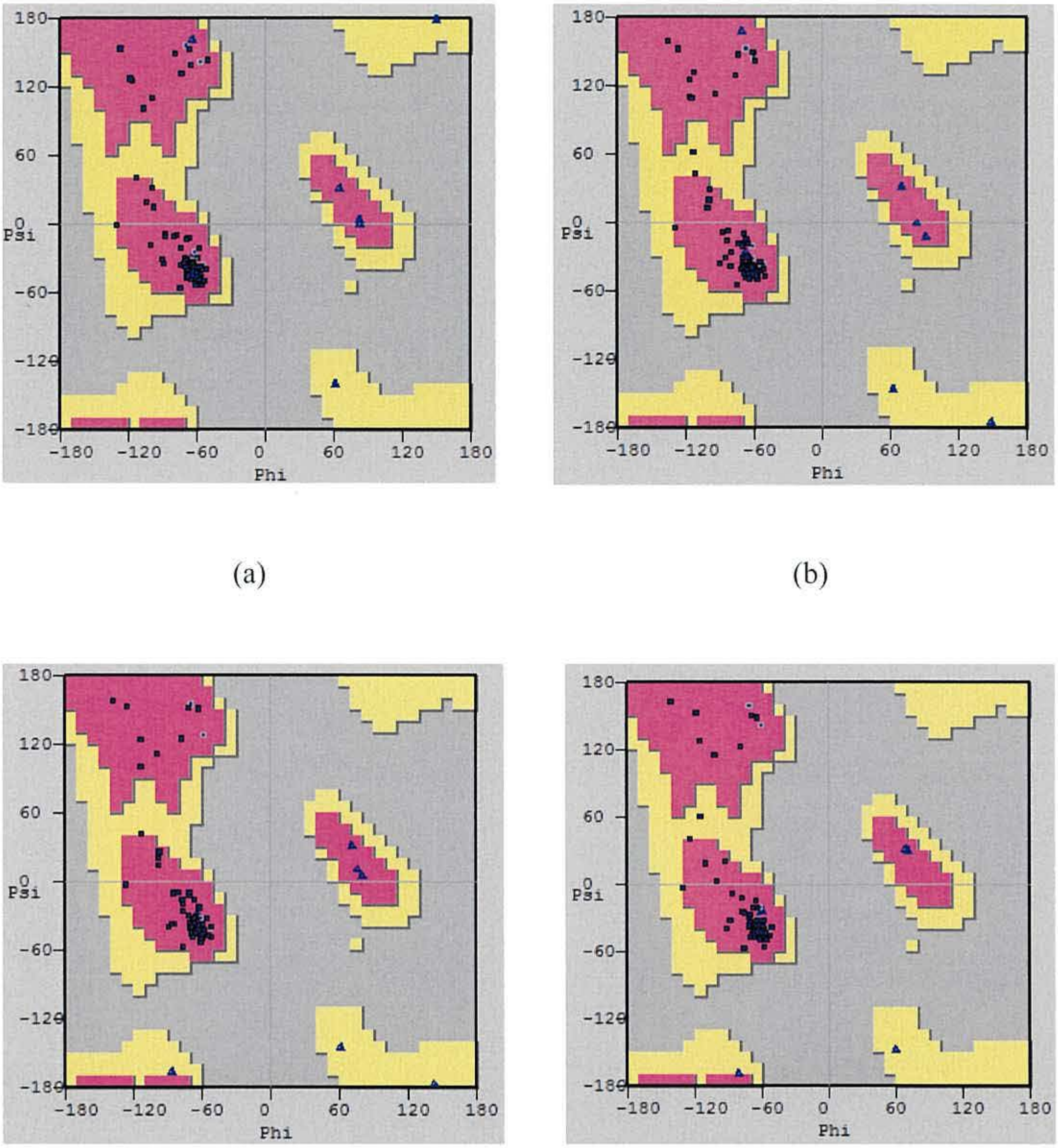
L16G was refined to 1.18 Å. At the end of refinement, the R_{factor} was 16.6% with R_{free} of 19.5%, the rms-bond length was 0.020 Å; 232 water molecules were added. L16Y was refined to 0.98 Å, with R_{factor} and R_{free} of 13.9 and 15.8 respectively and rms-bond of 0.019 Å, and 288 water molecules were added. The double mutant

L16F/W56G, was refined to 0.96 Å, with R_{factor} of 13.1 and R_{free} of 14.4% and the rms-bond of 0.019 Å, with addition of 322 water molecules to the model.

6.6 Structure validation:

The *PROCHECK* program,[20] a part of CCP4 program suite, and MOLEPROBITY[21] (see section 2.8.3) were used to identify regions or amino acids that may exhibit problem geometries. Both programs provide a detailed check on the stereochemistry of the protein structure. Rotamer outliers, main chain bond angles and lengths, and other parameters were looked at. After fixing any observed problem, the revised coordinate structure was saved and passed through the *Refmac5* for further refinement cycles.

The dihedral angles φ against ψ of amino acid residues were visualised by a Ramachandran plot[22] which is employed in *Coot.* For the oxidised form of recombinant AXCP, 99.1% of 110 non-glycine residues were located in preferred regions, and 0.9% of them were found in allowed regions. Since leucine 16 was mutated to glycine in the case of L16G mutant and tryptophan 56 was mutated to glycine in the case of the double mutant L16F/W56G, 109 non-glycine residues were found in the structure. Of these 109 non-glycine residues, 98.2% and 98.1% occupied positions in the preferred region in the case of L16G and L16F/W56G respectively, with 1.8% and 1.9% in allowed regions. The Ramachandran plots for all proteins are shown in (Figure 6.4). No outliers were detected in any structure. These values with other quality parameters as well as the refinement R values are listed in (Table 6.3).



(a) (b)

(c) (d)

Figure 6.3: Ramachandran plot of oxidised forms of recombinant AXCP (a), L16G (b), L16Y (c) and the double mutant L16F/W56G (d). Plots were produced using the graphic program *Coot*.

	Oxidised	L16G	L16Y	L16F/W56G
Resolution (Å)	1.15	1.18	0.98	0.96
Refinement				
R_{factor} (%)	12.5	16.6	13.9	13.1
R_{free} (%)	14.7	19.6	15.8	14.4
Ramachandran				
Residues in preferred region(%)	99.07	98.23	99.09	98.08
Residues in allowed region (%)	0.93	1.77	0.91	1.92
RMSD-				
Bond (Å)	0.020	0.020	0.018	0.019
Angle (deg)	1.932	1.971	1.805	1.867
DPI (Å)	0.028	0.029	0.018	0.017

Table 6.4: Refinement parameters of recombinant AXCP and mutants.

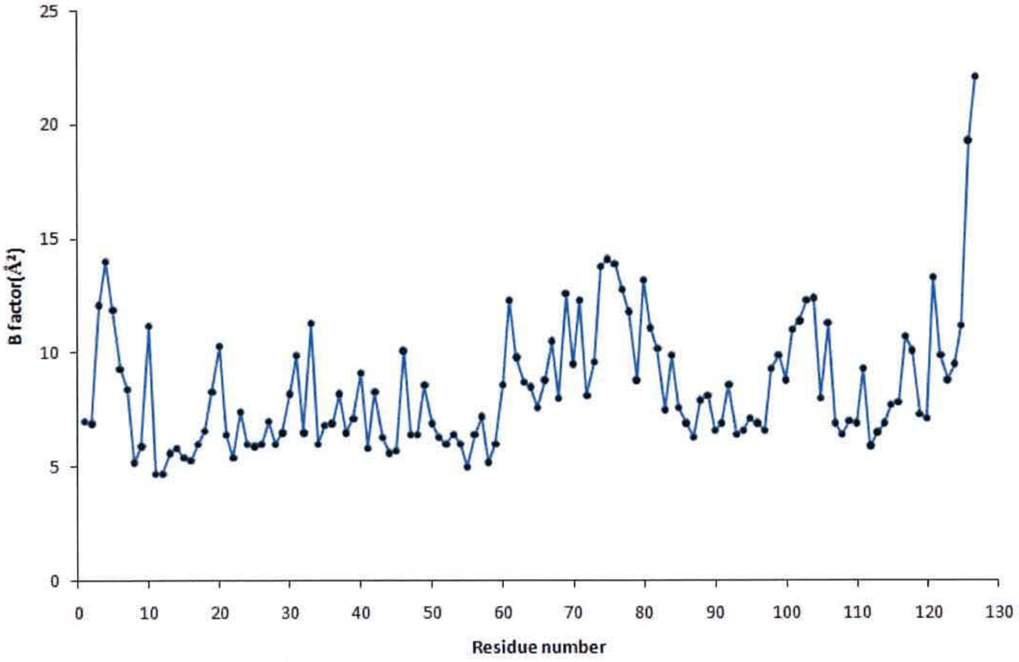
6.7 Recombinant *AXCP* crystallographic structure description at 1.15Å:

Since the crystal structures of recombinant AXCP which have previously been reported were from different expression systems,[16,23,24] it was considered appropriate to determine the crystal structure in our new expression system.

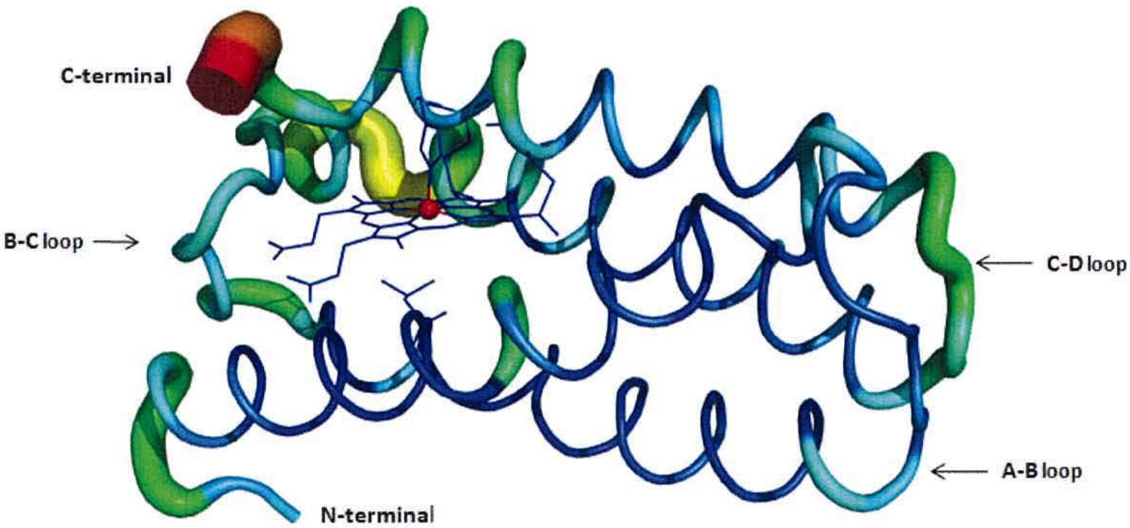
The overall crystal structure of recombinant AXCP expressed in our system was similar to that described by Dobbs *et al*[23] but it will be described here again in detail because of the greater accuracy due to the higher resolution.

The crystal structure at 1.15 Å contains a total number of 127 residues, 352 water molecules, one haem group and one sulphate group. Altogether 46 side chains and 17 main chain residues were modelled with alternate conformations. The fully refined model showed the R_{factor} and R_{free} of 12.5 and 14.7% respectively.

The Wilson B factor[14] was 6.4 $Å^2$. The average B-factor for the fully refined structure was 8.44 $Å^2$, (was 7.48 $Å^2$ for main chain and was 9.38 $Å^2$ for the side chains). The average B-factor for the solvent was 19.8 $Å^2$. The mean values of B are plotted against each residue in Figure 6.4.A. For more clarification the same data are shown graphically in (Figure 6.4 – B) using the program PyMOL.[25]



(A)



(B)

Figure 6.4: The plot of average temperature factor against each residue (A), and the graphical representation of the temperature value (B) shows that the model is more floppy in loops and in solvent exposed long side chain amino acid residues, these residues are represented in green and thick lines. The graph was created with PyMOL[25].

Many of the amino acid residues that are exposed to solvent have disordered side chains, especially the long side chains of Lys 4, Glu 71, Asp 75, Asp 101, Asp 103, Lys 117, Asp 121, and Lys 126, showing high B factors. These are represented as green and thick lines in (Figure 6.4 – B). However, apart from the last two residues; Lys126 and Lys127, all other residues had B factors under 14 Å^2. The low B factors indicate that AXCP as a whole is tightly organised and well ordered.

The crystal structure at 1.15 Å resolution of cytochrome c′ showed that it consists of four left hand twisted α-helix bundles labelled A to D, and one haem group covalently bound to helix D via Cys116 and 119, forming the CXXCH motif. Helix A is formed by residues 4-32, helix B by residues 37-59, while helix C is formed by residues 76-100, and finally helix D by residues 102-124 (Figure 6.5).

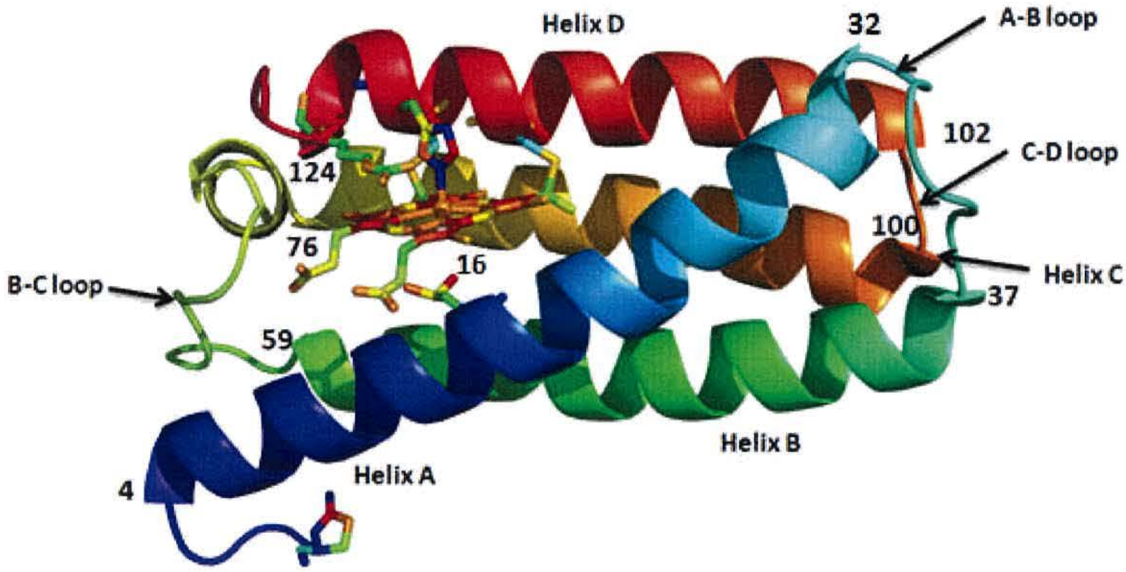


Figure 6.5: The overall structure of the recombinant cytochrome c′ at 1.18 Å, Helices and loops have been shown, numbers indicating the start and the end residue of helices and loops.

The crystal structure shows that most parts of these helices are stabilised by a classic Pauling-Corey-Branson hydrogen bonded model.[26] The (r + 4 → r) hydrogen bond, in which every main-chain N-H group donates a hydrogen bond to the main-chain C=O group of the amino acid four residues earlier is predominant in the structure.

Helix A begins with two (r + 3 → r) hydrogen bonds, or 3_{10} helix, 4O→7N and 5O→8N, and then the hydrogen bonding changes to (r + 4 → r) pattern from residue 6 - 21 while residues 22 - 25 does not involved in interhelical hydrogen bonding.

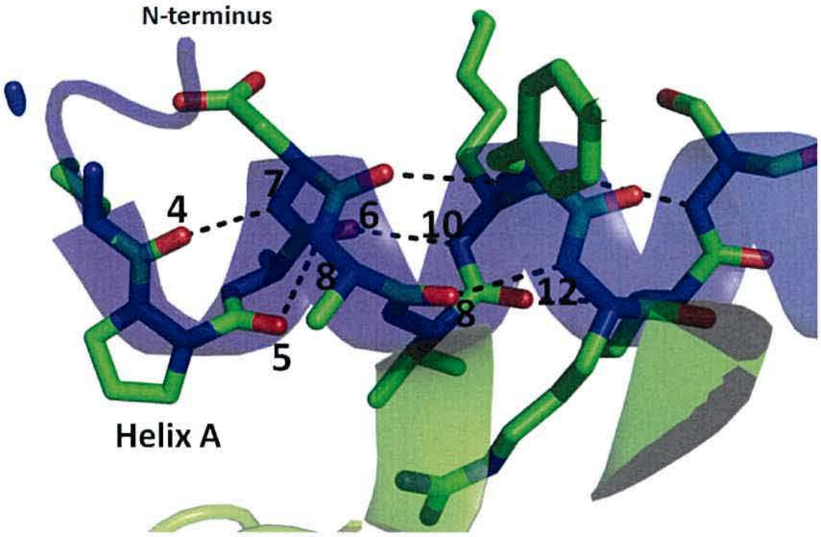


Figure 6.6: A cartoon representation of helix A, starts with the (r + 3 → r) hydrogen bonding pattern and then changes to (r + 4 → r) pattern.

This pattern changed to 3_{10} helix (r + 3 → r) again in helix A's c-terminus from residue 23-32 (i.e O23→N26, O24→N27, O26N29, O27→N30 and O29→N32) (Figure 6.7). Since the nitrogen atom of proline is part of the rigid ring, and rotation about N-C_{α} bond is not possible, Pro28 introduces a destabilising kink in the α helix. It is obvious from the same figure that Pro28 bends the helix through an angle of 30°. In addition, the nitrogen atom of a Pro residue in a peptide linkage has no substituent hydrogen to participate in hydrogen bonding with other residues.

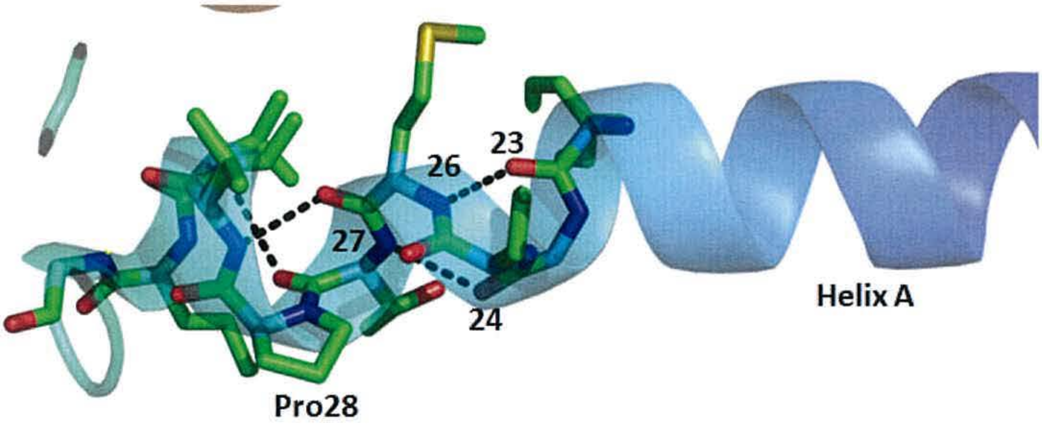


Figure 6.7: A cartoon representation of helix A, shows Pro28 residue which introduces a destabilising kink around the α helix.

In B-C loop the 3_{10} pattern changed to (r + 2 → r) pattern, where N71 forms a hydrogen bond with two residues earlier O69, N72 with O70 and N73 with O71 (Figure 6.8).

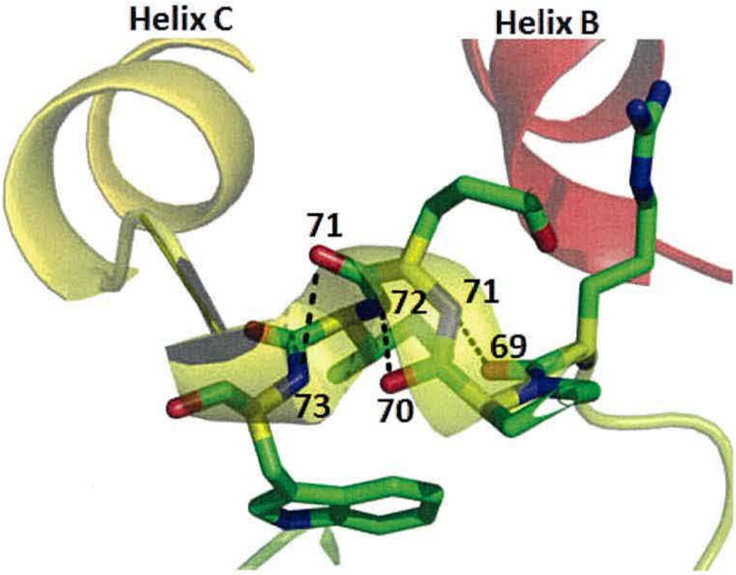


Figure 6.8: A cartoon representation showing the change in the pattern from 310 to (r + 2 → r) pattern

The N-cap hydrogen bonding interactions stabilise the structure. All four helices have N-caps. In helix A, Glu6 hydrogen bonds with its own peptide NH, and Asp7 hydrogen bonds with the peptide NH of Lys4. In helix B, Asp37, hydrogen bonds with the peptide NH of Gln40.; Asp101, hydrogen bonds with NH of Lys104 in the Helix D N-terminus. In the case of helix C, Ser78 hydrogen bonds with the peptide NH of a second conformation of Lys75 seen at 1.15 Å resolution. These bonds are shown in Figure 6.9. These interactions play a role in the stabilisation of the bundles.

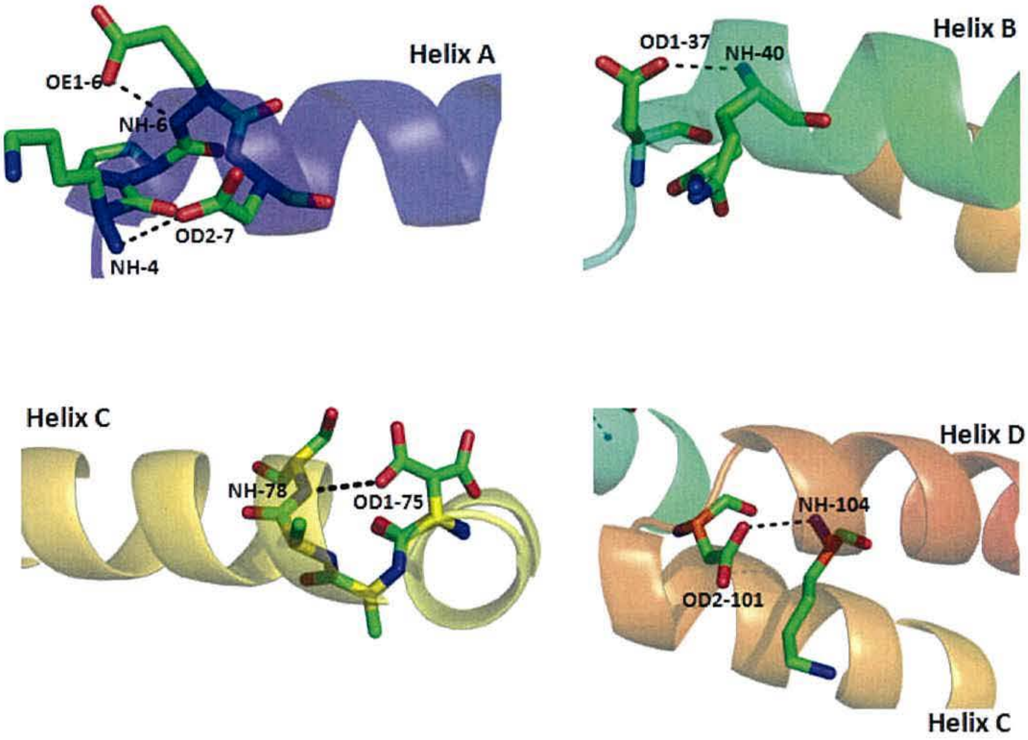


Figure 6.9: A cartoon representation showing N-caps in helices A, B, C, and D. These bonds stabilising the bindles.

Other bonds that stabilise the structure are the cross-linking and the salt bridge interactions. An important group of interactions are the network of hydrogen bonds by Arg12 on helix A. NH1 group of the guanidinium is hydrogen bonded to the carbonyl oxygen atom of residues 58 and 63 and the oxygen atom of hydroxyl of Thr63, OG1; these residues are all located in B-C loop, and NH2 hydrogen bonded to carbonyl oxygen atom of residues 63 and 65 from the same loop, while the NE hydrogen bonded to the propionate O atom of the D ring O1D. NH2 group of Glu13 hydrogen bonded to the propionate oxygen atom of the D ring O2D (Figure 6.10).

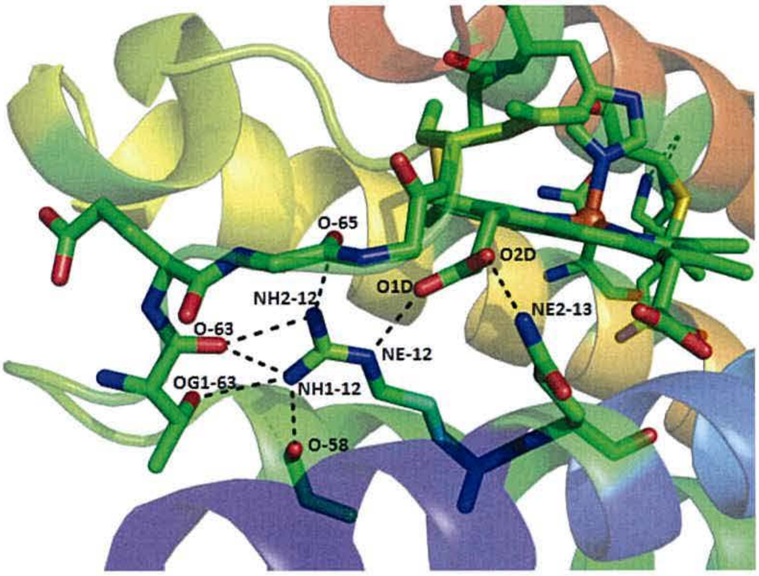


Figure 6.10: A stick and cartoon representation showing the location of the network of hydrogen bonds linking helix A to the haem group.

The other cross-linking hydrogen bonds include the hydrogen bond between the carbonyl oxygen atom of Asn44 on helix B and His22 on helix A, where OD1-44 is hydrogen bonded with ND1-22, and ND2-22 is hydrogen bonded to O-22 (Figure 6.11-A). Another hydrogen bond is between the NE1 group of Trp73 and the oxygen atom of the carbonyl group of Gly60; O-60, this hydrogen bond stabilises the conformation of the B-C loop (Figure 6.11-B). Finally, a salt bridge is formed between Lys92 on helix C and Asp111 on helix C linking the two helices (Figure 6.11-C). Hydrophobic residues on the surface of the four helices provide most of the interactions required for the bundle stabilisations.[23]

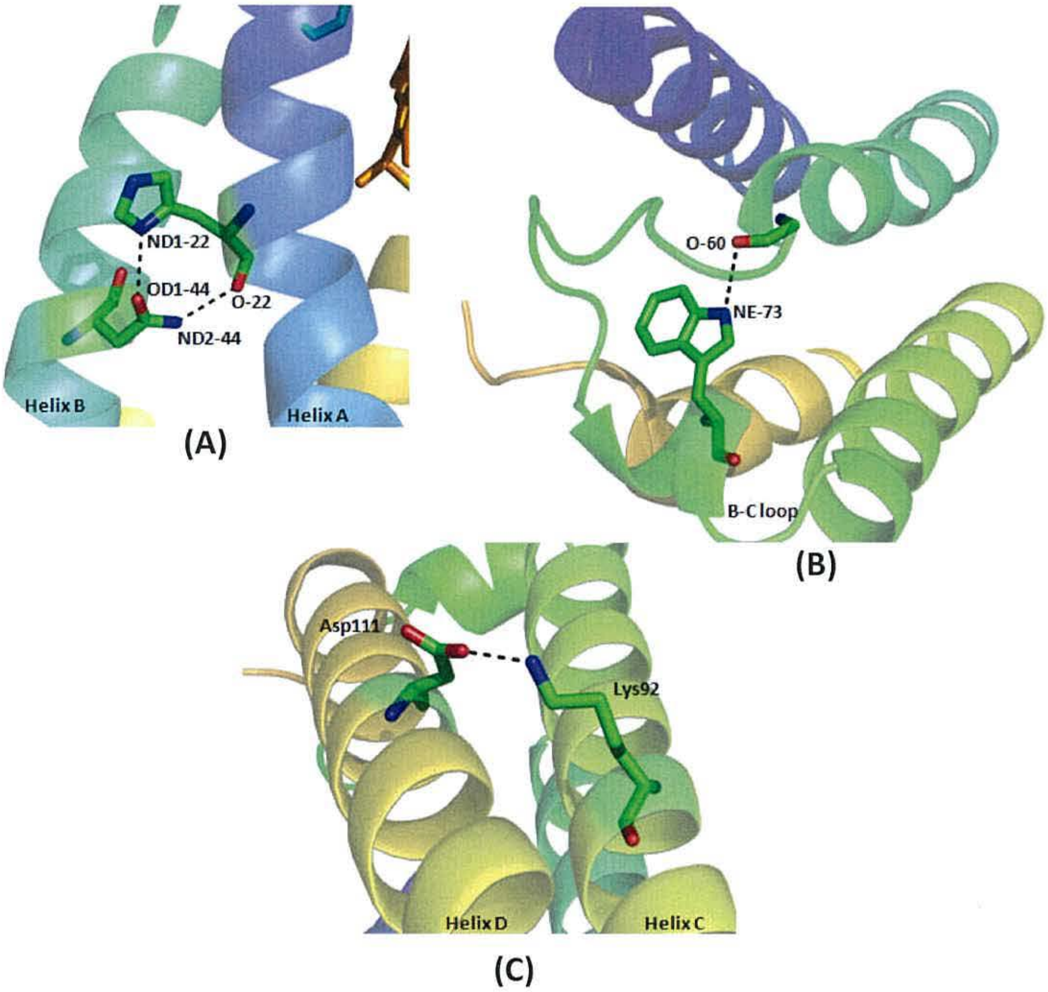


Figure 6.11: A stick and cartoon representation showing the cross-linking hydrogen bond between His22 and Asn44 linking helices A and B (A), Gly60 and Trp73 hydrogen bonding stabilises the B-C loop conformation (B), and the salt-bridge between Lys92 and Asp111 linking helices C and D.

6.7.1 Haem environment and geometry:

The haem group is located at the C-terminus of the protein. Helices A, C and D with the long B-C loop are directly involved in packing around the haem. They provide 22 residues with side chains that pack against or interact with the haem group. Helix A provides Val19, Arg12, Gln13, Leu16, Thr17, Met19, Ala20, and Phe23, helix C provides Phe79, Lys82, Gln83, and Phe86, while helix D provides the covalent attachment to the haem group via residues Cys116 and Cys119. These two residues are separated by two other residues, forming the characteristic CXXCH motif, which is a fingerprint for c-type cytochromes.[27] His120 is found to be an axial ligand to the haem iron in the proximal position,[16,23] the guanidinium group of Arg124 from the same helix packs the plane-to-plane with the haem group with presumed overlap of the positive charge of arginine with the negative charge of the haem π system.

Apart from the covalently bound residues Cys116, Cys119, the His120, and hydrogen bonded residues Arg12 and Gln13, all other interactions are hydrophobic. The most extensive interactions are made by Leu16 which covers the sixth iron coordination site and the aromatic side chains of Trp56, Phe59, Phe79, and Phe86.

The geometry of the haem group in recombinant AXCP is compared to haem group in the mutants in sections 6.11.

6.8 Comparison between the crystal structure of recombinant *AXCP* at 1.15Å resolution and the published structure at 2.05Å resolution:

The fully refined structure of recombinant AXCP at 1.15Å resolution is compared with the structure of the same protein refined at 1.8 Å resolution[23] and with the 2.05Å resolution crystal structure published in the protein data bank[28] (PDB code 1E83[16]).

The average bond length of Fe to the pyrrole N atoms is 2.05, 1.99 and 2.02 Å in the case of 1.15, 1.8, and 2.05 Å resolution structures respectively. The bond length to the fifth ligand, NE2 of His120, is 2.13, 2.02, and 1.94 Å with respect to the 1.15, 1.8, and 2.05 Å. The iron atom displacement is the same in all resolutions; it is 0.28 Å from the plane of the four pyrrole N atoms towards the His120 ligand. Finally the distance of the water molecule to the ND1 atom of the His120 is 3.2 Å. The guanidinium group of Arg124 lies parallel to the haem plane and perpendicular to the imidazole ring of His120 (Figure 6.12). The average distance of Arg124 to the haem plan is 3.73, 3.7, and 3.97 Å in the case of 1.15, 1.8 and 2.05 Å resolution crystal structures respectively, while the distance between the ND1 of His120 and NH1 group of Arg124 is 3.88, 3.75, and 3.5 Å in the same manner. These distances are summarised in (table 6.5).

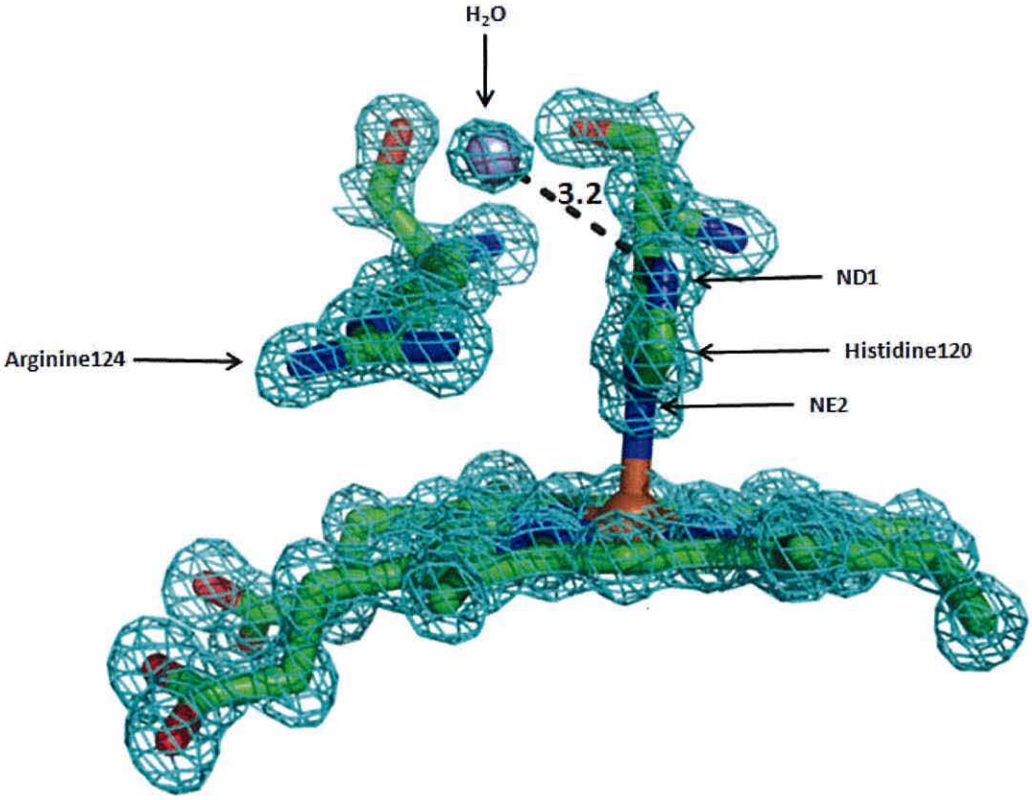


Figure 6.12: The electron density map of recombinant native AXCP showing the haem group, histidine120 and arginine124, and a water molecule.

Crystallographic Resolution	1.15 Å	1.8 Å [23]	2.05Å [16]
		Distances Å	
Fe-NA	2.05	1.97	1.94
Fe-NB	2.05	1.95	2.07
Fe-NC	2.05	1.99	2.07
Fe-ND	2.05	2.05	2.01
Average Fe-N atoms	2.05	1.99	2.02
Fe-NE2(His120)	2.13	2.02	1.94
Arg124-haem plane	3.73	3.70	3.97
ND1(His120)-NH1(Arg124)	3.88	3.75	3.5

Table 6.5: comparison of the haem geometry and Fe-ligand distances of the crystal structures of recombinant native AXCP solved in this thesis at 1.15 and the previously reported structures at 1.8, and 2.05 Å.

6.9 Description of the crystallographic structures of L16G, L16Y and the double mutant L16F/W56G:

Detailed structure solution and refinement processes of mutants have been discussed previously.

L16G was refined to 1.18 Å resolution with the R_{factor} and R_{free} of 16.6 and 19.6% respectively. No electron density was detected in the region within the distal haem pocket where the Leu16 side chain would have been expected on the basis of the native crystal structures (6.15). This provides structural confirmation that the leucine residue has been mutated to glycine and supports the results of the cloning and mutagenesis studies presented in Chapter 4. Both mutants L16Y and L16F/W56G were refined to even higher resolutions; 0.98 and 0.96 Å respectively with the R_{factor} and R_{free} of 13.9, 15.8% and 13.1 and 14.4% correspondingly (Table 6.4). In the case of L16Y large peak in the electron density was seen in the distal pocket of the haem, confirming a mutation from L16 to Y16. The same peak of the electron density was detected in the distal pocket for the double mutant, showing it has been replaced by phenylalanine, while the electron density around the original position of tryptophan vanished, showing it was replaced by glycine. These changes in the electron densities show the success of the mutatgenesis experiments. The electron density maps of all mutations are shown and discussed in the next section.

The crystal structure of all mutants each comprises 127 amino acid residues, and one haem group. The number of water molecules and the numbers of alternative amino acid conformations were different. In the case of L16G, 326 water molecules were added, and 45 residues were modelled with alternative configurations, 15 of which from the main chain, and 30 were side chains. The final structure of L16Y contains 296 water molecules, while 85 residues where modelled with alternative conformations, again, 16 of which were from the main chain, and 69 were side chains. The double mutant composed of 350 water molecules, 26 main chains and 73 side chain residues that were modelled with alternative conformations.

L16G showed the highest average temperature factor, the average B-factor for the protein was 12.6 $Å^2$; (11.4 and 13.9 $Å^2$ for the main chain and side chains respectively), while the B-factor calculated by the Wilson plot[14] was 11.2 $Å^2$. The protein B-factor for L16Y mutant was 8.9 $Å^2$ (8.2 and 9.7 $Å^2$ for the main chain and side chains respectively), and the Wilson B-factor was 7.2 Å^2. The double mutant L16F/W56G showed the lowest B-factor value, 7.4 Å^2 (6.6, 8.2 Å^2 main chain and side chains respectively). The B-factor for waters were 28.5, 21.0 and 19.3 Å^2 in L16G, L16Y and L16F/W56G respectively. These data are summarised in (Table 6.6):

Parameter	Recombinant	L16G	L16Y	L16F/W56G
B-factor (Å^2)				
Wilson	6.4	8.3	7.2	5.5
Protein	8.4	12.6	8.9	7.4
Main Chain	7.5	11.4	8.2	6.6
Side Chains	9.4	13.9	9.7	8.2
Waters	19.8	28.5	21.0	19.3
No. Waters	352	326	296	350
Alt. conformations				
Main Chain	13	15	16	25
Side chains	43	30	69	68

Table 6.6: A summary of the main chain, side chain, and waters B-factor, and the number of water molecules and alternative conformations exhibited by the fully refined crystal structures of recombinant native (1.15 Å), L16G (1.18 Å), L16Y (0.98 Å) and the double mutant L16F/W56G (0.96 Å).

Comparison between the temperature factors of each residue in different mutants are shown in Figure 6.14, and for the clarification, the same data are shown graphically in Figure 6.15 A-C.

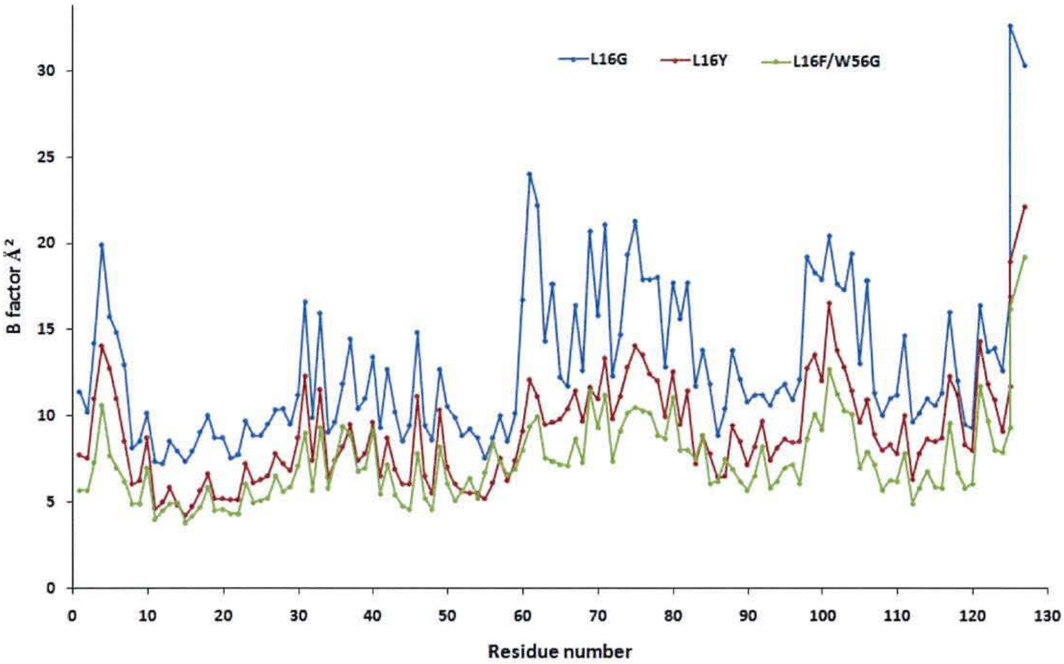


Figure 6.13: The variability of the average amino acid temperature factors of mutants; L16G, L16F, and the double mutant L16F/W56G.

Figures 6.13 and 6.14 show that all mutants have the higher temperature factor for last two residues, Lys126 and Lys127. Comparing B-factors of all three mutants, the B-C loop showed a higher B-factor in the case of L16G.

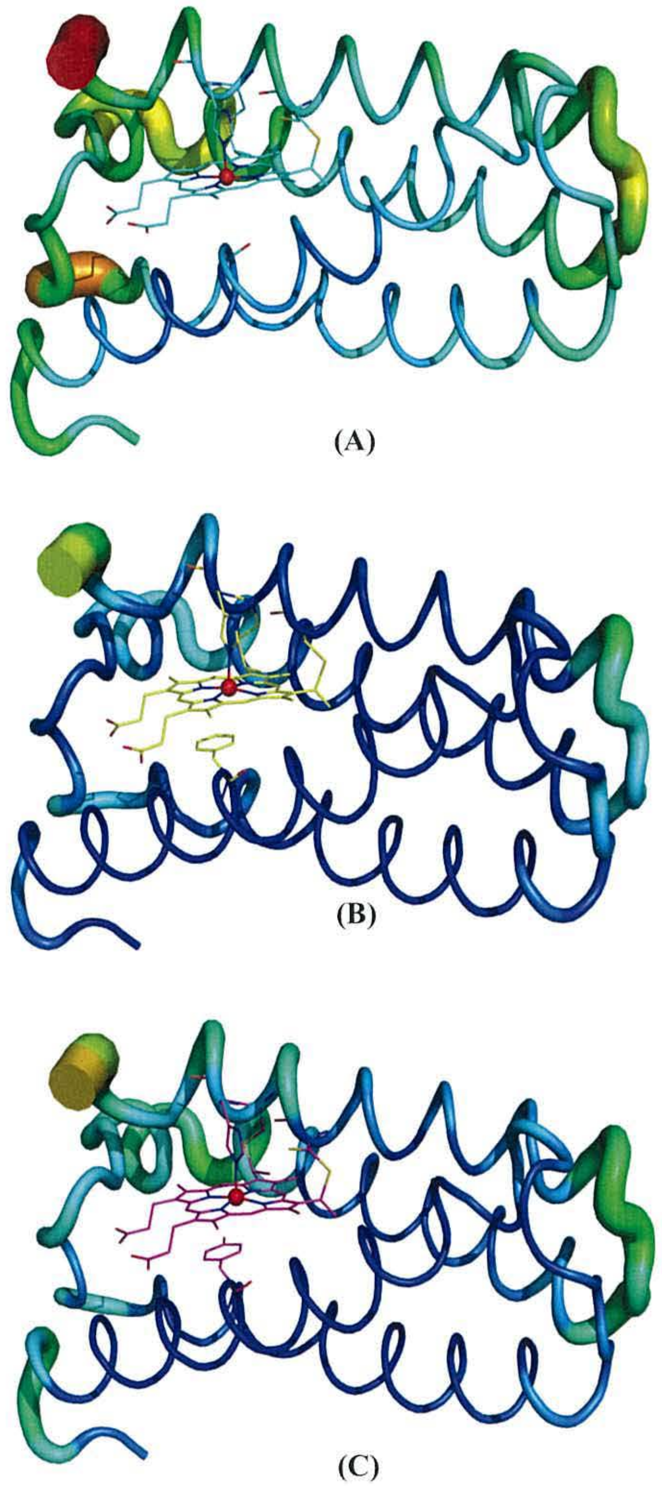


Figure 6.14: The graphical representation of the temperature values of each amino acid residue of the double mutant L16F/W56G (A), L16F (B) and L16Y (C). The green and thick lines represent the higher temperature values in those regions, thinner and blue lines represent lower temperature values. They all were drawn relative to the recombinant native AXCP temperature values showed in figure 6.4. The graph was created with PyMOL

6.9.1 The crystallographic structure of L16G:

The crystal structure of the mutant L16G (Figure 6.15) at 1.18 Å resolution showed that the Fe atom is six-coordinate with a water molecule as a sixth ligand. L16G is the first cytochrome c′ mutant reported so far to be six coordinate when isolated and purified with a water molecule in iron's sixth coordination position. The only other six coordinate iron moiety in the AXCP, that has been observed is with the addition of the exogenous CO to AXCP as characterised by Lawson *et al.16* The guanidinium group of Arg124 lies parallel to His120 and perpendicular to the haem plane suggesting that this mutant is already in the reduced form, and this has been confirmed by the UV-Vis spectra and the visible red colour of this mutant (Figure 4.5 - chapter 4). Cys116 has been modelled with alternate conformations denoted by (cycteine116-B in figure 6.15), which is probably caused by the exposure to x-ray radiation during the course of atomic resolution data collection.[29,30]

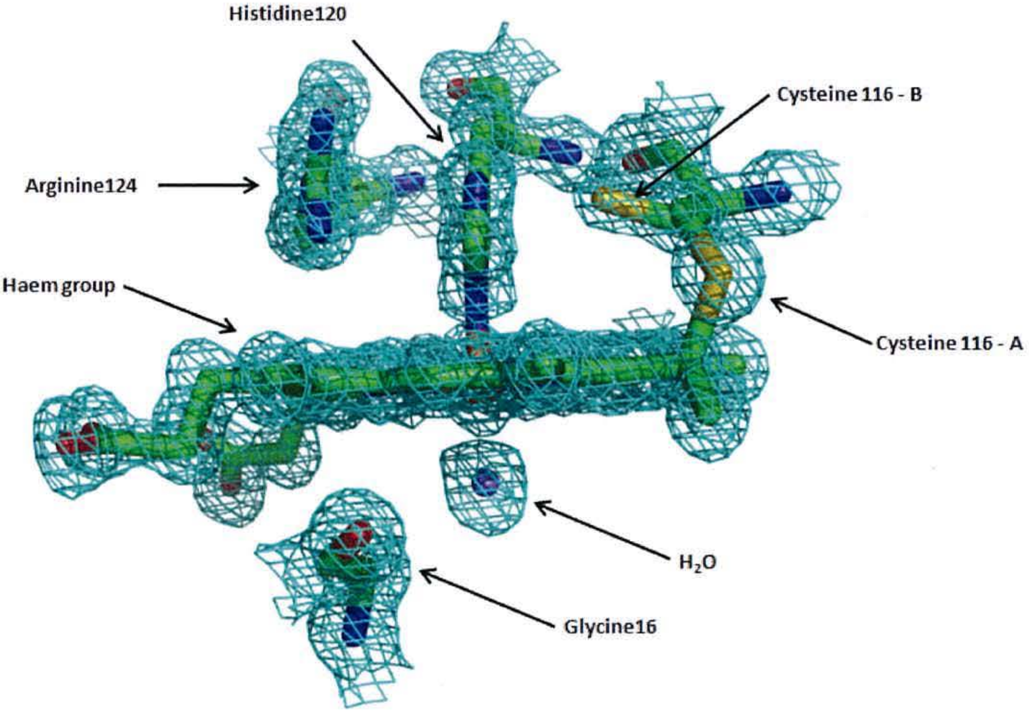


Figure 6.15: The electron density map of L16G (as isolated) at 1.18 Å resolution showing a clear electron density around glycine 16 confirming the success of the mutagenesis and around the water molecule bound to the haeme's sixth coordinate. Arginine 124 is perpendicular to the haem plane indicating that this mutant is already reduced when isolated. Cysteine 116 was modelled in two alternative conformations (cysteine 116-A and cysteine 116-B).

6.9.2 The crystallographic structure of L16Y:

The crystal structure of L16Y at 0.98 Å showed that the phenol ring of the tyrosine residue is parallel to the haem plane and within van der Waals contact (Figure 6.16). Arg124 has been modelled with alternative conformation suggesting the partial photoreduction of the haem iron by 50%, the photoreduction and the change in the oxidation state can occur rapidly at the metal centre.[29,30] The UV-Vis spectra of this mutant (Figure 4.7 - Chapter 4) showed that this mutant was oxidised when isolated, hence the partial reduction of the haem iron may be caused by the x-ray beam. Cys116 has been modelled with alternate conformation as well, also probably caused by the x-ray radiation.

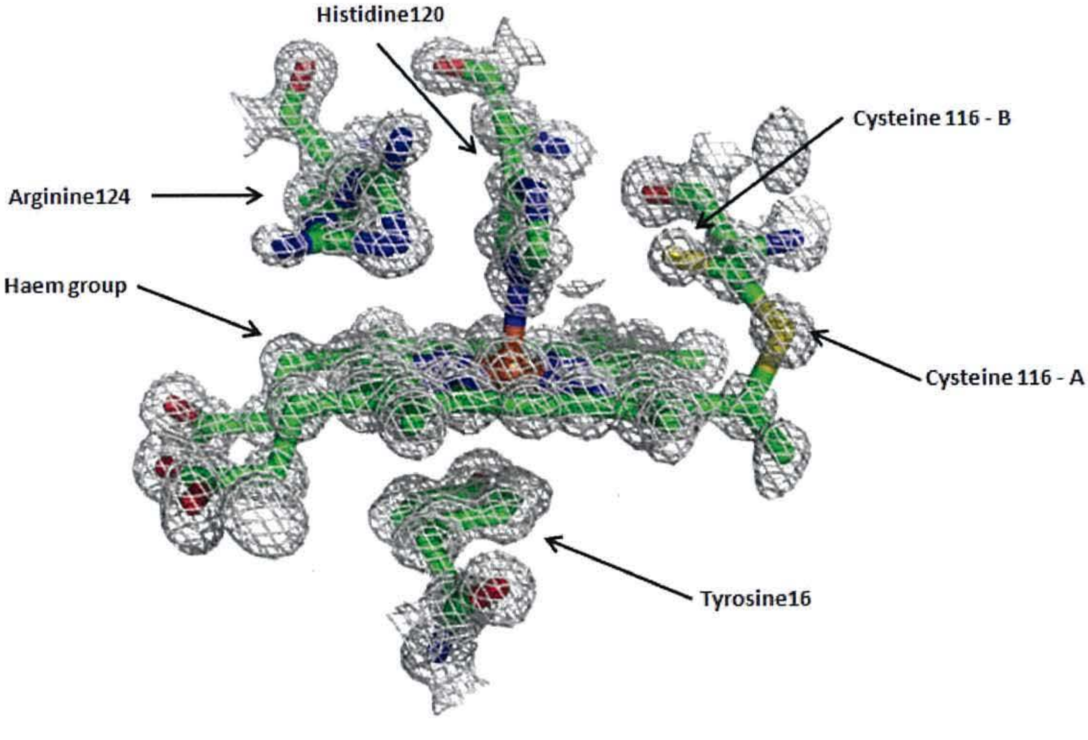


Figure 6.16: The electron density map of L16Y at 0.98 Å resolution showing a clear electron density around tyrosine16 confirming the success of the mutagenesis. Cysteine116 and arginine124 are modelled in two alternative conformations resulted from the photoreduction of the metal centre by the x-ray radiation during the course of atomic resolution data collection.

6.9.3 The crystallographic structure of the double mutant L16F/W56G:

The phenol ring of the phenylalanine residue of the double mutant L16F/W56G solved at 0.96 Å also stacks parallel to the haem group (Figure 6.17). The UV-Vis spectra (Figure 4.10 – chapter 4) showed that this mutant was oxidised when isolated. However, these data were collected at the SLS-Switzerland, which uses a Pilatus-6M detector that allows data collection at atomic resolution in only one pass, and hence the crystal is in the x-ray beam for a shorter period of time. Arg124 was modelled with alternate conformations again suggesting the partial reduction of the haem iron by 30%. This may be due to the higher x-ray radiation energy (0.725 Å) used in this beamline. Cys116 was modelled with only one conformation.

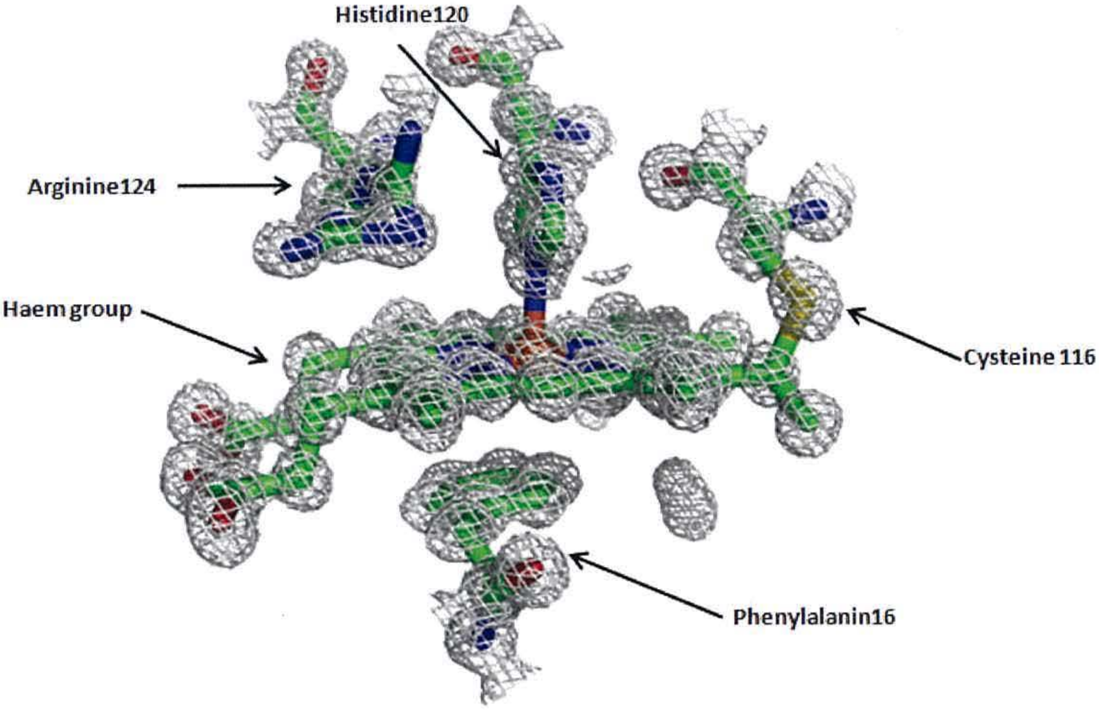


Figure 6.17: The electron density map of the double mutant L16F/W56G at 0.96 Å resolution showing a clear electron density around tyrosine16 confirming the success of the mutagenesis. Only arginine 124 is modelled in two alternative conformations resulted from the photoreduction of the metal centre by the x-ray radiation during the course of atomic resolution data collection.

6.10 Crystal structure comparisons:

The mutation of Leu16 and Try56 to different amino acid residues did not affect the overall structure of the protein (Figure 6.18).

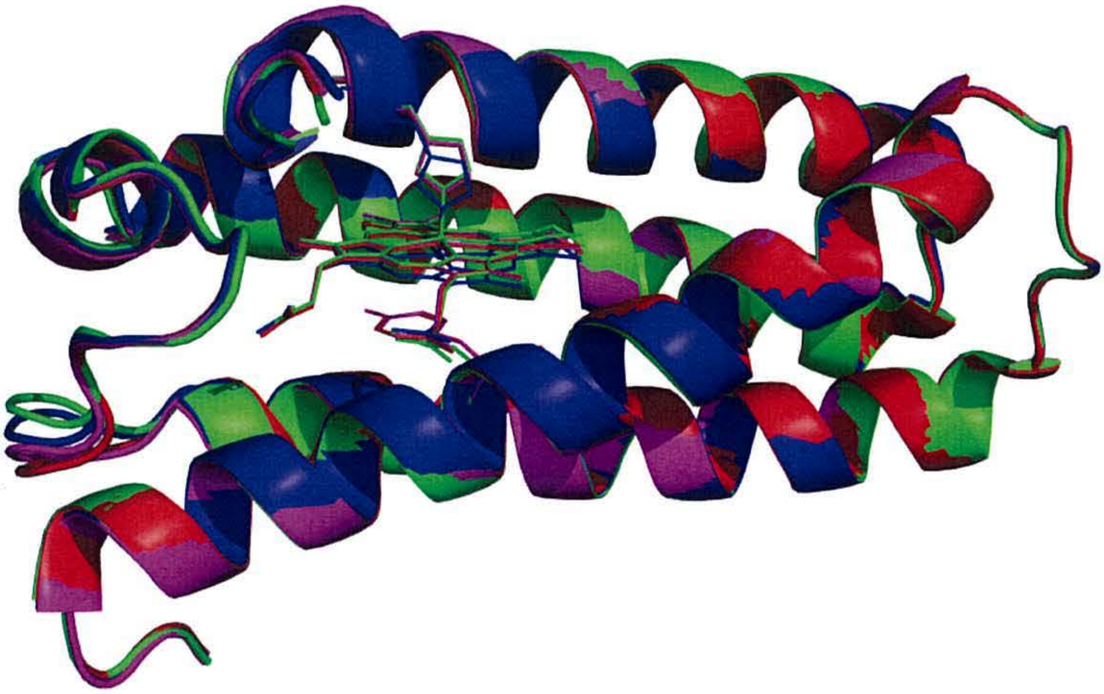


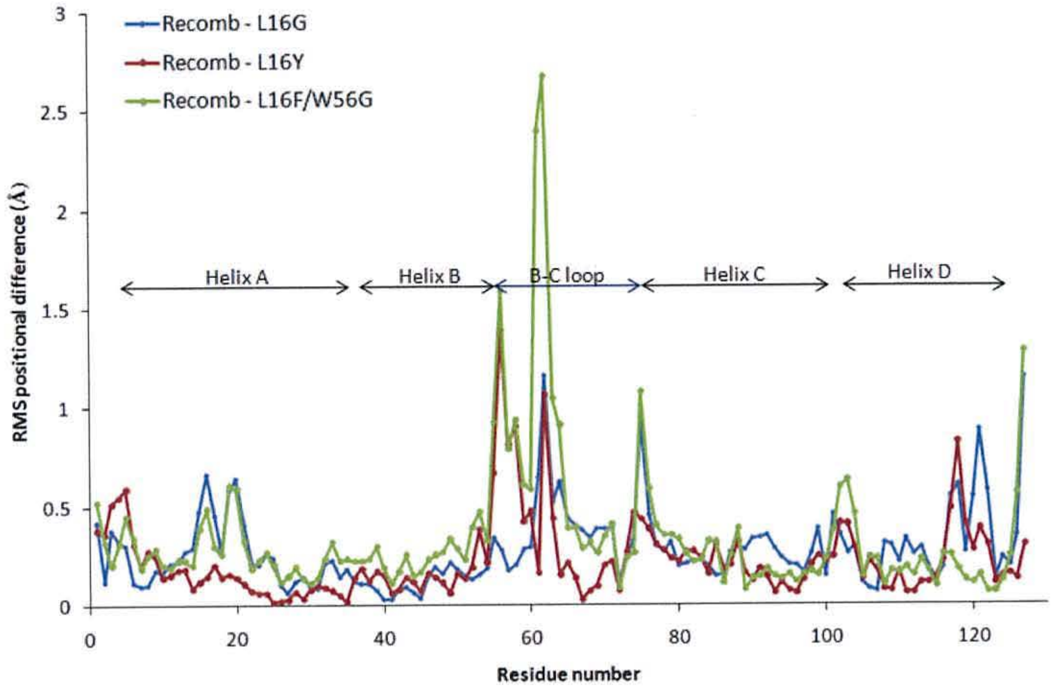
Figure 6.18: Ribbon representation of superimposed recombinant native AXCP (green) to L16G (blue), L16Y (magenta), and the double mutant L16F/W56G (red).

Superimposing the carbon α atoms of the mutant structures on the equivalent native recombinant AXCP carbon α atoms using the PyMOL[25] molecular graphic system revealed the RMS positional differences of 0.28, 0.38 and 0.50 Å for L16G, L16Y and L16F/W56G. These differences were smaller when mutants were compared to each other. The RMS differences of L16G compared to L16Y and L16F/W56G were 0.14 and 0.16 Å respectively, while the RMS difference was 0.14 Å when L16Y and L16F/W56G were compared (Table 6.7).

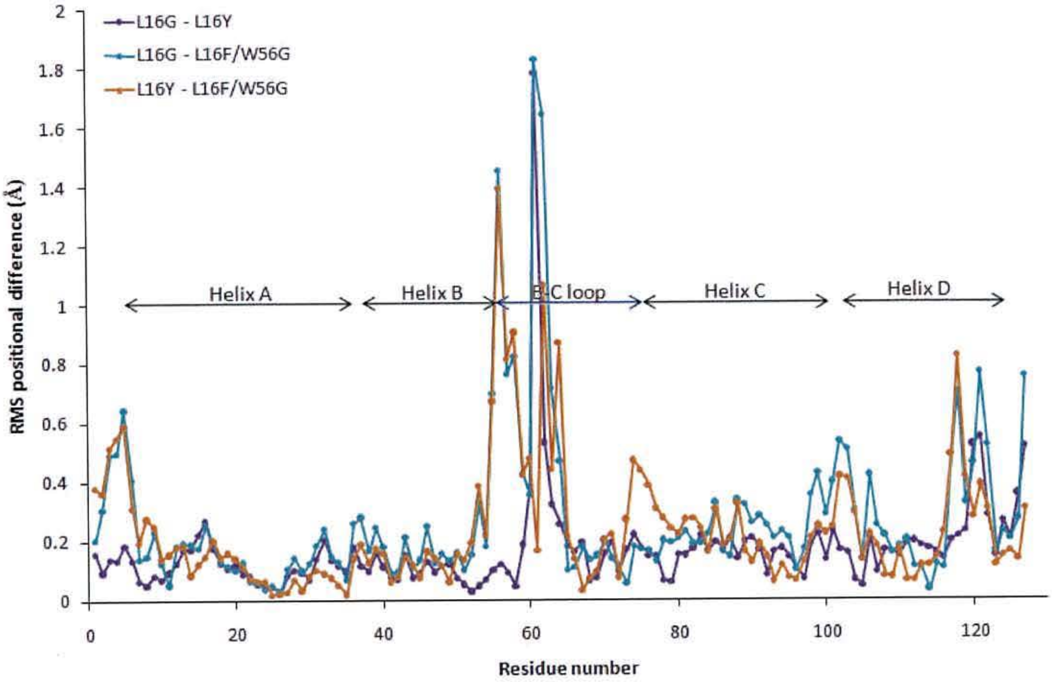
Models superimposed	RMS deviation from C_α (Å)
Recombinant (1.15 Å) – L16G (1.18 Å)	0.24
Recombinant (1.15 Å) – L16Y (0.98 Å)	0.19
Recombinant (1.15 Å) – L16F/W56G (0.96 Å)	0.22
L16G (1.18 Å) – L16Y (0.98 Å)	0.14
L16G (1.18 Å) – L16F/W56G (0.96 Å)	0.16
L16Y (0.98 Å) – L16F/W56G (0.96 Å)	0.14

Table 6.7: RMS positional deviations for the C_α atoms of the four structures, numbers in the brackets indicate the crystal structure resolution.

Using the *Superpose*[31] program part of the CCP4 program suit,[12] recombinant AXCP at 1.15 Å was aligned to the 1.18, 0.98 and 0.96 Å resolution crystal structures of L16G, L16Y, and L16F/W56G mutants. The mutant structures were also superposed together. The resultant RMS positional difference of individual residual C_α atom at equivalent positions was plotted against the residue number; recombinant against mutants is shown in (Figure 6.19 A) and mutants between themselves are shown in (Figure 6.19 B). It is obvious from both figures that for all mutants the RMS values are higher than the average value and are all located at the N-terminus of helix A, C and D, at the C-terminus of helix D and in the A-B and B-C loops. The highest differences were calculated for residues in the B-C loop and in the mutation points.



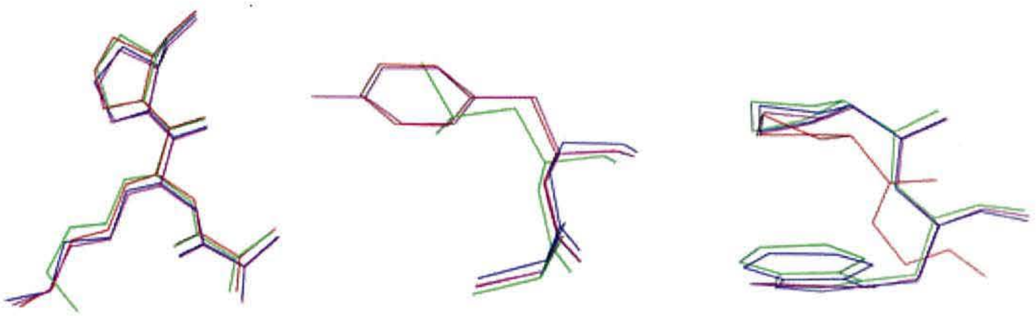
(A)



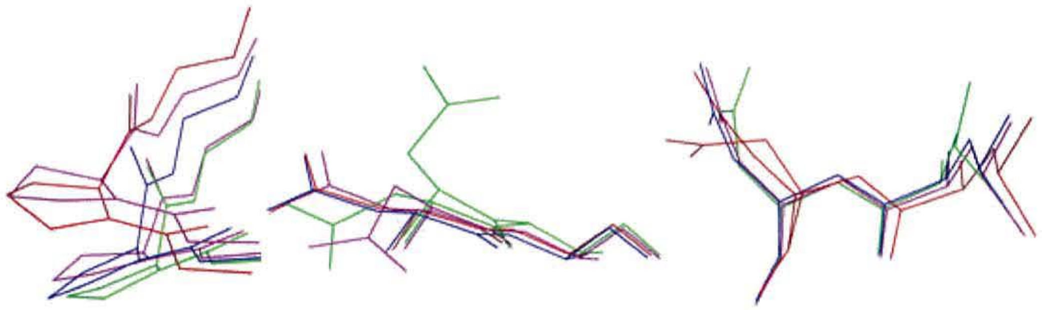
(B)

Figure 6.19: RMS positional difference plot between equivalent C_α atoms of all residues of 1.15 Å resolution crystal structure of recombinant AXCP with mutants; 1.18 Å L16G, 0.98 Å L16Y and 0.96 Å resolution L16F/W56G (A) and mutants between themselves (B)

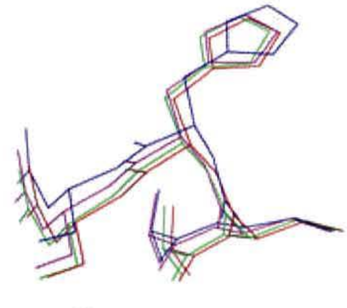
Residues with an RMS difference values significantly higher than its average values are presented in (Figure 6.20).



Residues 3-5 **Residues 15,16** **Residues 55,56**



Residues 60-62 **Residues 74,75** **Residues 102,103**



Residues 118-120

Figure 6.20: Residues with considerably higher RMS positional difference than the average values. recombinant native AXCP (green) to L16G (blue), L16Y (magenta), and the double mutant L16F/W56G (red).

6.11 The effect of Leu16 mutation on the overall haem geometry:

In order to find out the effect of mutating Leu16 on the haem group geometry, all atoms of the haem group of recombinant AXCP and other mutants were superimposed using *Superpose* program. Superimposing the haem group of recombinant AXCP to L16G, L16Y, and L16G/W56G showed the RMS positional difference value of 0.33, 0.21, and 0.13 Å respectively with the haem. Superimposing the haem group of L16G to L16Y and L16F/W56G revealed the RMS positional difference of 0.17 and 0.24 Å respectively, while superimposing the haem group of L16Y to L16F/W56G showed an RMS positional difference of 0.1 Å, which is the smallest difference among all structures. These data are summarised in (Table 6.8) and showed in figures thereafter. These positional differences are considerably higher than those previously reported for R124K and R124E mutants,[24] explaining the considerable effect of replacing Leu16 on the overall haem geometry.

Haem groups superimposed	RMS deviation of the haem group (Å)
Recombinant (1.15 Å) – L16G (1.18 Å)	0.33
Recombinant (1.15 Å) – L16Y (0.98 Å)	0.21
Recombinant (1.15 Å) – L16F/W56G (0.96 Å)	0.13
L16G (1.18 Å) – L16Y (0.98 Å)	0.17
L16G (1.18 Å) – L16F/W56G (0.96 Å)	0.24
L16Y (0.98 Å) – L16F/W56G (0.96 Å)	0.10

Table 6.8: RMS positional deviations for the haem group atoms of the four structures calculated by using *superpose* program, part of CCP4 program suit, numbers in the brackets indicate the crystal structure resolution of the relevant protein.

The individual haem group atoms are shown in (Figure 6.21), and the plot of each atom of the haem group against the RMS positional difference of the equivalent atom is shown in (Figures 6.22 A and B).

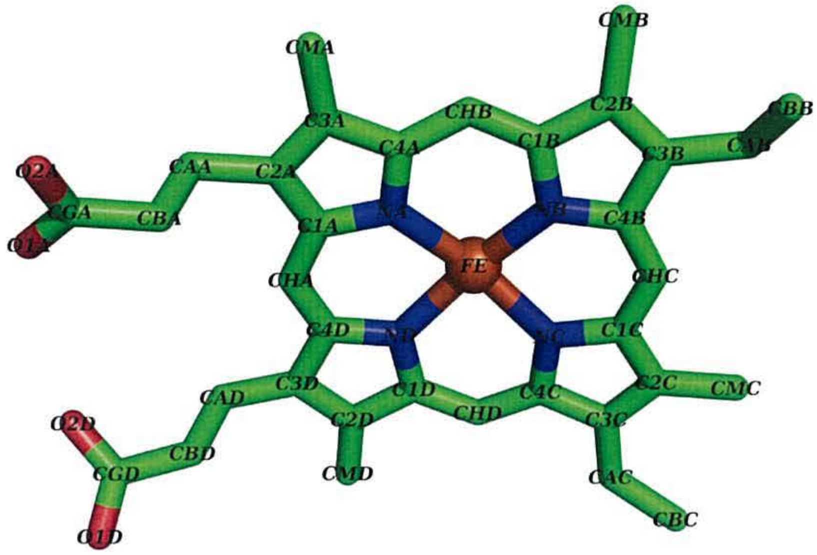


Figure 6.21: The haem group, showing the individual atom name.

The effect, of replacing Leu16 by other residues on the RMS positional difference of the haem group was in order of glycine > tyrosine > phenylalanine. The greatest effect was found to be on atoms located at the periphery of the haem group especially atoms CHA, CGA, O2A and other atoms like CMC, C3C, CBC (Figures 6.22 A and B The (Fe atom) was also observe to displaced in the same order.

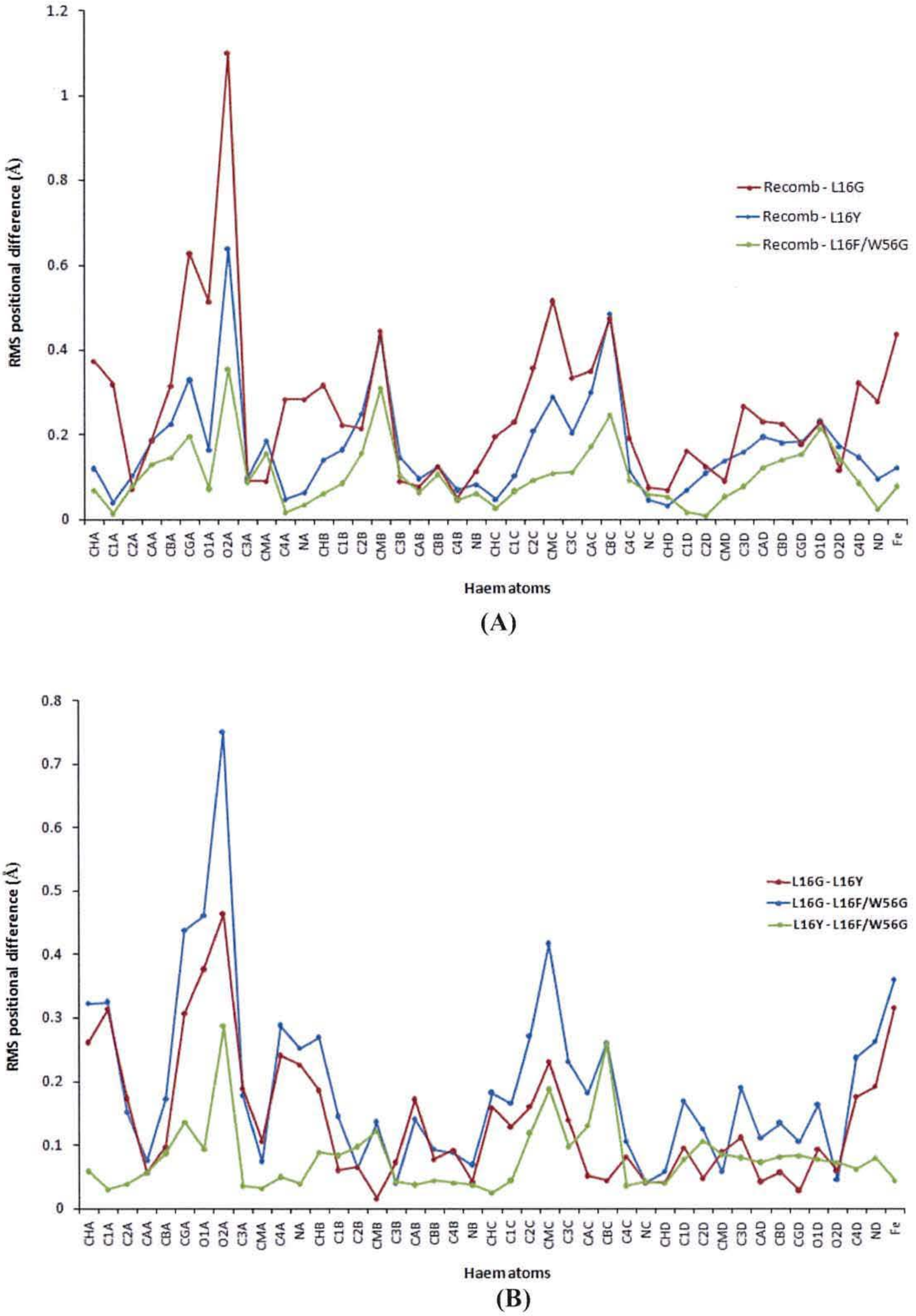


Figure 6.22: RMS positional difference plot between equivalent atoms of the 1.15 Å resolution of recombinant AXCP with 1.18 Å L16G, 0.98 Å L16Y and 0.96 Å resolution L16F/W56G crystal structures (A) and mutants between themselves (B)

6.11.1 Haem group comparison of recombinant native AXCP and L16G mutant:

Unexpectedly, the iron atom of L16G was found to be 6-coordinate with His120 in the fifth and a water molecule in its sixth-coordinate; this is the first reported six-coordinate cytochrome c′ mutant (Figure 6.15). Aligning the haem group of this mutant (1.18 Å) to recombinant native AXCP (1.15 Å) resolution shows that the iron atom is displaced towards the sixth ligand by 0.4 Å. This displacement can be confirmed by the RMS positional deviation plot (Figure 6.22 - A). The NE2-Fe bond is now perpendicular to the haem plane with NE2-Fe-CHC angle of 89.4°, and the haem group has flattened with the NA-Fe-ND angle of 179.0° (Figure 6.23). Bond lengths between the iron atom and the N-atoms of the four pyrrole rings of the haem have been measured (Table 6.9).

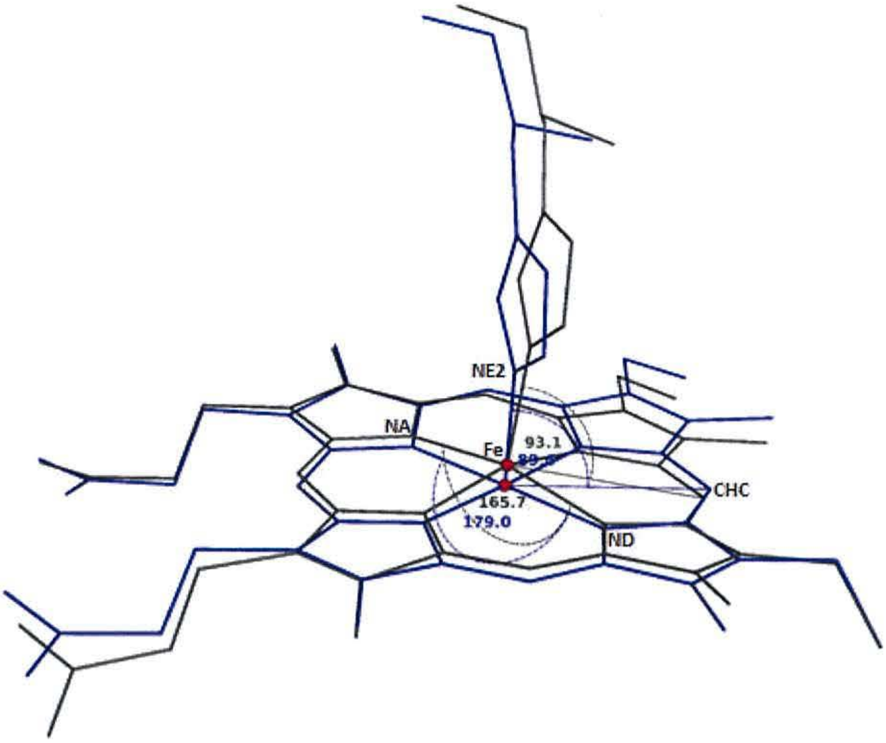


Figure 6.23: Superposition of L16G (blue) and native (grey) haem group. The haem displacement and the haem flattening can be observed.

Distances Å	Recombinant native	L16G	L16Y	L16F/W56G
Fe-NA	2.05	2.00	2.06	2.03
Fe-NB	2.05	2.04	2.06	2.04
Fe-NC	2.05	2.03	2.05	2.03
Fe-ND	2.05	2.01	2.06	2.03
Fe-NE2(His120)	2.13	2.07	2.12	2.11

Table 6.9: comparison n of the haem Fe-ligand distances between recombinant native AXCP *(1.15 Å)*, L16G (1.18 Å), L16Y (0.98 Å) and the double mutant L16F/W56G (0.96 Å) crystal structures solved in this thesis.

6.11.2 L16Y and L16F/W56G:

The haem iron of L16Y and in the double mutant L16F/W56G was found to be five-coordinate. Superimposing the haem group of L16Y (0.98 Å) and L16F/W56G (0.96 Å) to the haem group of recombinant AXCP (1.15 Å) revealed a small iron atom displacement, which was 0.1 Å in both cases. There also were very small bond length differences between iron atom and N-atoms of the four pyrrole rings (Table 6.7). The NE2-Fe bond of L16Y and L16F/W56G is shorter by 0.01 and 0.02 Å respectively from the same bond of the recombinant AXCP. The NE2-Fe-CHC angle is 91.8° and 91.9° for L16Y and L16F/W56G consequently (Figure 6.24). The haem group of both mutants has deviated to the planarity if compared to the recombinant AXCP, where the NA-Fe-ND angle has changed to 167.9 and 168.1° in L16Y and L16F/W56G respectively, comparing to 165.7° of the recombinant AXCP.

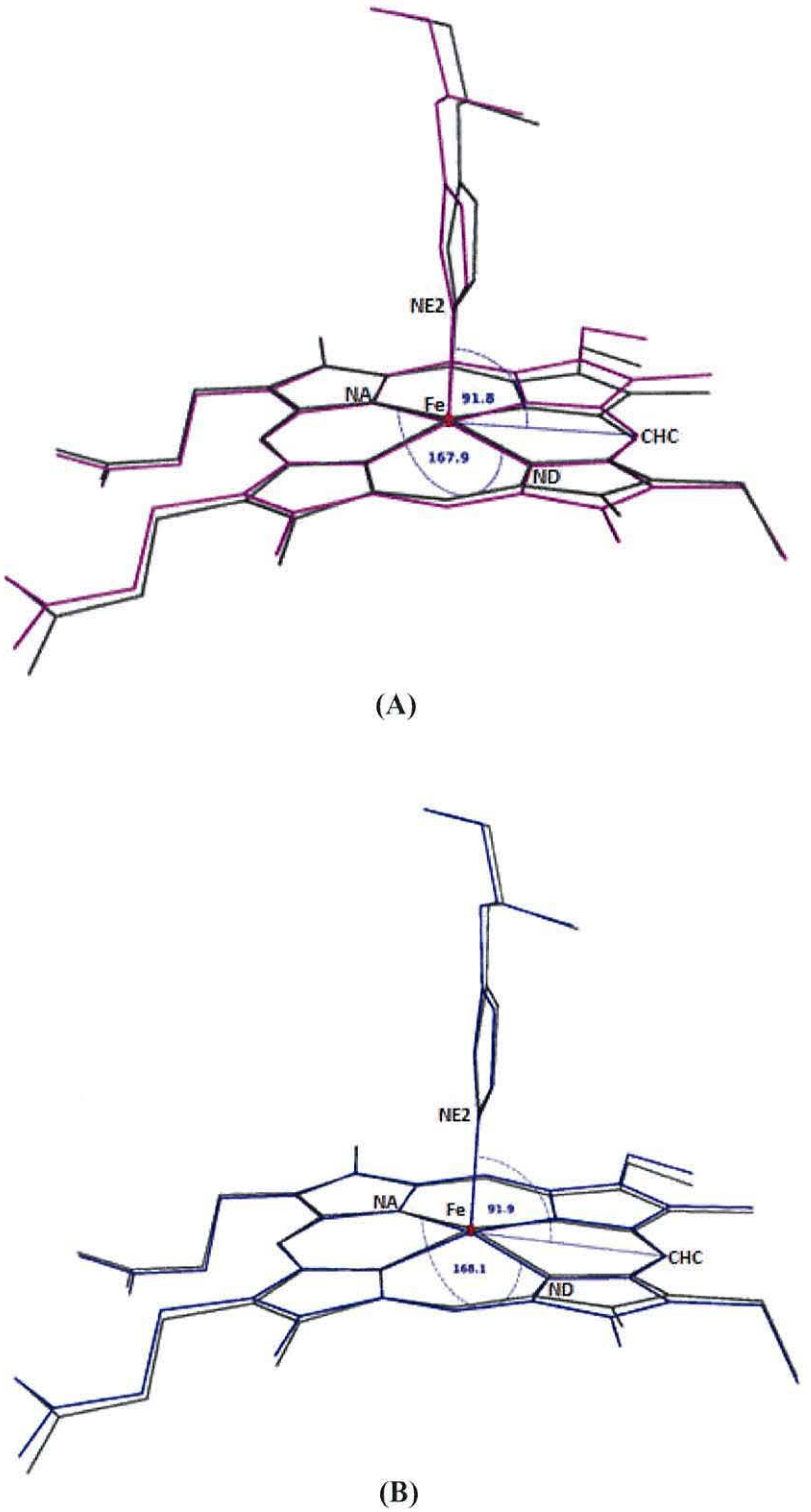


Figure 6.24: superposition of L16Y (blue) (A) L16F/W56G (blue) (B) and native (grey) haem group. The haem displacement and the haem flattering are show.

6.12 Conclusions:

The crystal structures of recombinant native AXCP and four mutants were solved at atomic resolution. Mutations were made on residues close to the distal face of the haem iron in order to probe their effects on ligand binding and discrimination. The overall conclusion on each mutant is discussed below:

Recombinant native AXCP in its resting state was solved at a resolution of 1.15 Å with R_{factor} and R_{free} of 12.5 and 14.7% respectively. The overall structure of the recombinant was similar to the native protein. It consists of 352 water and 127 residues and one haem group. Residues are arranged in four left hand twisted α-helix bundle. The haem group is five coordinate with a His120 residue located in its fifth coordination site, while the sixth coordination site is vacant. It is covalently bound to the bundles via two disulphide bonds creating a CXXCH motif. The relatively low temperature factor showed that the protein is tightly organised and well ordered.

The crystal structure of L16Y solved 0.98 Å. A clear electron density peak equivalent to the phenol ring at leucine 16 position was apparent confirming the success of the mutation. The phenol ring of tyrosine stacked parallel to the haem group within the Van der Waals contact distance. The iron of the haem group found to be five coordinate with his 120 in the fifth coordination site while the sixth coordination site was vacant. The overall structure of the mutant protein was similar to the recombinant protein.

The solvent channel to the haem site which is seen in type I cytochromes such as in *C. vinosum* is blocked by tryptophan 56 in AXCP. In order to open this channel in AXCP, tryptophan 56 was mutated to glycine. At the same time leucine 16 was mutated to phenylalanine therefore the double mutant L16F/W56G was generated. The crystal structure of this mutant was resolved at 0.96 Å resolution. The phenyl ring of phenylalanine 16 positioned at the same way as for the phenol ring of tyrosine of L16Y mutant. Regarding the channel, replacing tryptophan by glycine, did not open the channel.

Leucine 16 was replaced by glycine generating L16G. The crystal structure of this mutant was solved at 1.18 Å resolution. Unexpectedly, the iron haem was six coordinate, with a water molecule in its sixth coordination position and His 120 in its fifth coordination site. This is the first reported six coordinate crystal structure of cytochrome c′ in its relaxing state. Replacing leucine by glycine has also increased the planarity of the haem group..

6.13 References:

---

[1] Hasnain S. S., Wakatsuki S. (2008) "Biophysical methods: structure, function and dynamics studies of macromolecular assemblies using electron, lasers, neutrons and x-rays" *Current Opinion in Structural Biology* **18,** 577-580.

[2] Sugahara M., Yusuhiko A., Morikawa Y., Kageyama Y., and Kunishima N. (2008) "Nucleant-mediated protein crystallization with the application of microporous synthetic zeolites" *Acta Crystallographica* **D64**, 686-695.

[3] Saridakis E., and Chayen N. E. (2009) "Towards a 'universal' nucleant for protein crystallisation" *Trends in Biotechnology* **27,** 99-106.

[4] G. M. Sheldrick (1990) "Phase annealing in SHELX-90: a direct method for large structures" *Acta Crystallographica* A**46,** 467-473.

[5] Dauter Z., Lamzin V., and Wilson K. S. (1997) "The benefits of atomic resolution" *Current Opinion in Structural Biology* **7**, 681-688.

[6] Strange R. W., Ellis M., and Hasnain S. S. (2005) "Atomic resolution crystallography and XAFS" *Coordination Chemistry Reviews* **249**, 197-208.

[7] Cusanovich M. A., Tedro S. M. and Kamen M.D. (1970) "*Psedomonas denitrificans* cytochrome cc` " Archives of biochemistry and biophysics, **141**, 557-570.

[8] Dauter Z. (2003) "Protein structures at atomic resolution" *Methods in Enzymology* **368**, 288-337.

[9] Broennumann Ch. and *et al* (2006) "The Pilatus 1M detector" *Journal of Synchrotron Radiation* **13**, 120-130.

[10] Leslie A. G. W. (1992) "Recent changes to the MOSFLM package for processing film and image plate data" *joint CCP4/ESF-EACMB Newsletter on protein crystallography*, No. **26**

[11] Evans P. (1993) "Data reduction: data collection and processing" *Proceedings of the CCP4 study weekend. Data collection and processing*, 114–123.

[12] CCP4 (1994) "The CCP4 Suite: programs for protein crystallography" *Acta Crystallographica*, **D50**, 760-763.

[13] French G. S., and Wilson K. S. (1978) "On the treatment of negative intensities" *Acta Crystallographica* A**34,** 517-525.

[14] Wilson A. J. C. (1949) "The probability distribution of x-ray intensities" *Acta Crystallographica* **2,** 318-321.

[15] Vagin, A. A. and Teplyakov, A. (1997) "MOLREP: an automated program for molecular replacement" *Journal of Applied Crystallography* **30**, 1022-1025.

[16] Lawson D. M., Stevenson C. E. M., Andrew C. R., and Eady R. R. (2000) "Unprecedented proximal binding of nitric oxide to heme: Implication for guanylate cyclase" *The EMBO Journal* **19,** 5661-5671.

[17] Murshudov G. N, Vagin A. A., and Dodson E. J. (1997) "Refinement of macromolecular structures by the maximum-likelihood method" *Acta Crystallographica* **D53**, 240-255.

---

[18] Emsley P., and Cowtan K. (2004) "Coot: model –building tools for molecular graphics" *Acta Crystallographica* D**60**, 2126-2132

[19] Lamzin V. S., and Wilson K. S. (1993) "Automated refinement of protein models" *Acta Crystallographica* D**49**, 129-147.

[20] Laskowski R. A., MacArthur M. W., Moss D. S., and Thornton J. M (1993) "PROCHECK: a program to check the stereochemical quality of protein structures" *Journal of Applied Crystallography* **26**, 283-291.

[21] Davis I. W., Leaver-Fay A., Chen V. B., Block J. M., Kapral G. J., Wang X., Murray L. W., Arendall III W. B., Snoeyink J., Richardson J. S., Richardson D. C. (2007) "MolProbity: all-atom contacts and structure validation for proteins and nucleic acids" *Nucleic acid research* **35**, W375-W383.

[22] Ramachandran G. N., Ramakrishnan C., and Sasisekhran V (1963) "Stereochemistry of polypeptide chain configurations" *Journal of molecular biology* **7**, 955.

[23] Dobbs A. J., Anderson B. F., Faber H. R., and Baker E. N. (1996) "Three dimensional structure of cytochrome c′ from two Alcaligenes species and the implications for four-helix bundle structures" *Acta Crystallographica* D**52**, 356-368.

[24] Barbieri S. (2007) "Mutagenesis, spectroscopy, and crystallography of recombinant cytochrome c′ from *A. xylosoxidans*" *PhD thesis,* University of Wales, Bangor.

[25] DeLano, W.L. (2002) "The PyMOL Molecular Graphics System" DeLano Scientific, Palo Alto, CA, USA. **http://www.pymol.org**

[26] Pauling L., Corey R. B., Branson H. R. (1951) "The structure of proteins: two hydrogen-bonded helical conformations of the polypeptide chain" *Proceedings of the National Academy of Science in Washington* **37:** 205-211.

[27] Allen J. W. A., Daltrop O., Stevens J. M., and Ferguson S. J (2003) "C-type cytochromes: diverse structures and biogenesis systems pose evolutionary problems" Philosophical Transactions of the Royal Society, B **358,** 255-266.

[28] Berman H. M., Westbrook J., Feng Z., Gilliland G., Bhat T. N., Weissig H., Shindyalov I. N., Bourne P. E. (2000) "The protein data bank" *Nucleic acid research* **28**, 235-242.

[29] Beitlich T., Kühnel K., Schulze-Briese C., Shoeman R. L., and Schlichting I. (2007) "Cryoradiolytic reduction of crystalline heme proteins: analysis by UV-Vis spectroscopy and x-ray crystallography" *Journal of Synchrotron Radiation* **14**, 11-23.

[30] Yano J., Kern J., Irrgang KD., Latimer M., Bergmann U., Glatzel P., Pushkar Y., Biesiadka J., Loll B., Sauer K., Messinger J., Zouni A., Yachandra V. K. (2005) "X-ray damage to the Mn_4Ca complex in single crystals of photosystem II: A case study for metalloprotein crystallography" *Proceedings of the National Academy of Science* **23**, 12047-12052.

[31] Krissinel E. and Henrick K. (2004) "Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions" *Acta Crystallographica* **D**60, 2256-2268.