The University Of Sheffield.

Structural studies of hemerythrin proteins in the human pathogen *Campylobacter jejuni*

Sami Ali Melebari

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Department of Molecular Biology and Biotechnology
University of Sheffield, UK

November 2017

Abstract

Hemerythrins are oxygen-binding proteins, originally found in the body fluids and tissues of marine invertebrates such as the worms *Sipunculids* and *Priapulida*. In these organisms, a multimeric form of hemerythrin is found in their blood cells and a monomeric myohemerythrin in their muscles. Both, however, are very similar in function and structure and are often described according to the oxidation and ligation states of their iron centre. Hemerythrins differ from haemoglobin and haemocyanin in that they possess non-haem di-iron atom sites that bind one molecule of O_2. Recently, they have been reported in a wide variety of bacteria particularly in microaerophilic and anaerobic bacteria.

Campylobacter jejuni is one species in the family *Campylobacteraceae*, and it is one of the most common causes of gastroenteritis worldwide, but a commensal of chickens and other fowl. *C. jejuni* is a microaerophile and therefore needs a quite low oxygen environment (5-10% O_2) for optimal growth. Throughout its life cycle, however, it is essential for it to be able to survive in the aerobic environment (21% O_2), as well as the oxidative stress response of the human host. There is no clear molecular explanation for the microaerophilic nature of *C. jejuni*. The majority of *Campylobacter species* have been shown to have multiple hemerythrin-like proteins, many of which contain a C-terminal domain of unknown function but with high homology among strains. The analysis of the genome of NCTC11168 in C *jejuni* shows three different genes encoding proteins with iron and oxygen binding hemerythrin-like characteristics: *cj0241c, cj1224* and *cj0045c.* Protein sequences of the Cj0241, Cj1224 and Cj0045 hemerythrins reveal that they all include conserved histidine motifs H, HxxE, HxxxH and HxxxxD, which are important to bind the di-iron centre critical for hemerythrin function. Therefore, the project aimed to resolve the question of the presence of these three genes, determine information on their different roles and the function of some the essential residues in these hemerythrins and their role in oxygen sensing and protection in *Campylobacter jejuni.*

Plasmids for the overexpression of *cj0241c*, *cj1224* and *cj0045c* were obtained from the laboratory of Prof David Kelly. A range of experiments were performed related to function and structure determination. The expression and purification of Cj0241 was very successful, and the structure was determined. The three-dimensional crystal structure was solved at 1.1 Å resolution using Single-wavelength Anomalous Dispersion (SAD) data sets. It showed a four alpha helix bundle with a central di-iron site bridged by an oxygen atom and coordinated by the conserved histidine residues. Analysis revealed a clear tunnel down the middle of the protein along which water and other solvent molecules could travel from the protein exterior to the di-iron site. The Cj1224 and Cj0045 proteins were successfully purified and analyses carried out on them although it was not possible to obtain crystal structures. Chemical, biochemical and biophysical experiments to study the properties of the proteins suggested that the C-terminal extensions in the Cj1224 and Cj0045 proteins might possess zinc binding properties. Attempts to measure the autoxidation rates of the hemerythrins implied a sensor or transport role for the proteins.

Acknowledgement

First and foremost, I would like to express my utmost gratitude and appreciation to Allah Almighty for giving me the health and wellness to complete and present this work.

I would like to sincerely thank my supervisor Dr. John Rafferty for his guidance, support, encouragement, feedback and patience during my PhD study, allowing me to complete my thesis.

I am grateful to Professor David Kelly for his advice and his help whenever I approached him. I am also grateful to Professor David Rice for his assistance, explanations which were instrumental in my understanding of my research. I would also like to thank Dr. Patrick Baker for his valuable guidance during my studies.

I am also deeply grateful Dr. Svetlana Sedelnikova for her assistance and tireless help with manipulation of protein purification protocols and general guidance in the lab. Without her support and patience getting results would not have been possible.

A special thanks to Fiona Rodger for her help with organisation and management of the lab and follow-up on requirements crucial for my research. Many thanks to Dr. Claudine Bisson for her operational assistance.

Also, my great appreciation to Dr. Mark Collins, Dr. Adelina Martin, Mr. Simon Thorpe and Mr. Neil Bramall from the Mass Spectrometry (MS) service laboratory. Also thanks to Prof Mike Williamson and Andrea Hounslow from the NMR service laboratory. Many thanks to my advisors Dr. Jeremy Craven, Dr. Rosie Staniforh, Dr. Jeffrey Green and Dr. Ewald Hettema.

I cannot thank enough Dr. Bilal R. Malik and Dr. Mohammed A. Nassar from the Biomedical Science for their support, guidance and discussion.

On a personal note, I would like to mention and express special thanks to all my colleagues Abdelhamid Elbrghathi, George William, Faizah Almalki, Reem Farsi, Mona Alharbi, Naer Alkaabawi, Israa Alhawani, Jason Wilson, Hayley Owen, Adli Abdaziz and Azura Mohd Noor for their support and encouragement.

I am truly indebted to my wonderful, sisters and brothers for all absolute support, love, sacrifices and care. There are no words which can demonstrate my gratitude towards them.

Very heartfelt thanks to my wife Maryam for her love, sacrifices, support and sharing with me the good and bad times throughout my journey towards a PhD degree.

I am truly indebted to my mother-in-law for her love, prayers, support and always wishing me success.

To Mohammad, Abdullah, Doha, Aseel and Bassam- you have all my love.

Last but not least, my deepest acknowledgement to my government, Kingdom of Saudi Arabia, Ministry of Health for a generous scholarship which made it possible for me to carry out this piece of research. Without this support, it would not have been possible.

I dedicate this research

to my parents.

Table of Contents

Table of Figures

Chapter 1: Introduction

Chapter 2: Material and Method

Chapter 3: Theory of X-ray Crystallography

Chapter 4: Cj0241 Hemerythrin Expression, purification and Crystallisation

Chapter 6: Cj0045 Hemerythrin Expression, purification and Crystallisation

List of tables

Abbreviations

Biological and chemical terms

APS	Ammonium persulphate
IPTG	Isopropyl β-D-1-thiogalactopyranoside
PEG	Poly ethylene glycol
Amp	Ampicillin
ASP	Aspartic acid
CFE	Cell free extract
CYS	Cysteine
PHE	Phenylalanine
GLU	Glutamic acid
PRO	Proline
HIS	Histidine
Kan	Kanamycine
LEU	Leucine
LYS	Lysine
MET	Methionine
PAGE	Polyacrylamide gel electrophoresis
SDS	Sodium dodecyl sulfate
ILE	Isoleucine
TAE buffer	A mixture of Tris base, acetic acid and EDTA
Tris	Tris(hydroxymethyl) aminomethane)
ALA	Alanine
DTT	Dithiothreitol

Crystallographic terms

Å	Angstrom (10^{-10}m)
ASU	Asymmetric unit
a, b, c, α, β, γ	Real space unit cell dimensions and angles
B factor	Crystallographic temperature factor
d_{hkl}	The distance between planes with indices *h k l* in the reciprocal lattice
COOT	Crystallographic Object Oriented Tool
hkl	Reciprocal lattice indices
F	Structure factor
F_{hkl}	Structure factor for a single reflection with indices *hkl*
$\lvert F_{hkl} \rvert$	Structure factor amplitude for the reflection with indices *hkl*
F_{calcs}	Calculated structure factor
F_{obs}	Observed structure factor
$\lvert F_{calcs} \rvert$	Calculated structure factor amplitude
$\lvert F_{obs} \rvert$	Observed structure factor amplitude
f	Atomic scattering factor
I / σI	Signal to noise ratio
MAD	Multi wavelength anomalous dispersion
MR	Molecular replacement

R-factor	Crystallographic refinement R-factor
R_{free}	Free R-factor
R_{merge}	Merging R-factor
R_{pim}	Precision-indicating merging R-factor
SAD	Single-wavelength anomalous dispersion
u,v,w	co-ordinates in Patterson space
V_m	Matthews Coefficient
XDS	X-ray Detector Software
Z	Number of equivalent positions in the unit cell
ρ	Electron density
λ	X-ray wavelength

Miscellaneous terms

°	Degrees
%	Percentage
nl	Nanoliter
µl	Microliter
BLAST	Basic local alignment search tool
°C	Degrees centigrade
CCP4	Collaborative Computational Project No. 4
3D	Three dimensional
h	Hour
kDa	Kilodalton
LB	Luria Bertani media
MW	Molecular weight
NMR	Nuclear magnetic resonance
OD	Optical density
PCR	polymerase chain reaction
PI	Isoelectric point
RMSD	Root-mean-square deviation
rpm	Revolutions per minute
SOC	Super optimal broth with catabolite repression
UV	Ultraviolet
Times gravity	x g
DdH_2O	Distilled Deionized water
dH_2O	Distilled water
nt	Nucleotide
bp	Base pair
PDB	Protein Data Bank

Chapter 1

Introduction

This chapter presents a general introduction to the predicted hemerythrin proteins from *Campylobacter jejuni.* Beginning with a general overview of the bacterial pathogen *Campylobacter jejuni* with its biological importance being the agent that is the cause of Campylobacteriosis infection in humans. Then it hones in on the family of the hemerythrins proteins and toward the end are stated the aims of this project.

1.1 Characteristics of *Campylobacter jejuni*

Campylobacter jejuni (*C. jejuni*) is one species in the *epsilonproteobacterial* family *Campylobacteraceae*, and is viewed as a major cause for gastroenteritis worldwide (Epps *et al.*, 2013). Taxonomically the *epsilonproteobacteria* are defined on the similarity in the percentage of the 16S rRNA gene sequence. It contains the families of *Helicobacteraceae* and *Campylobacteraceae*, which contain among them all the species of *Helicobacter*, *Campylobacter*, *Arcobacter* and *Wolinella* (Nachamkin *et al.*, 2008). The biochemical characteristics are negative for indole, acetone production and methyl red reaction. Most species are positive for oxidase and catalase activity, while there is an absence of hippurate hydrolysis and nitrate reduction (Nachamkin *et al.*, 2008). Food, particularly poultry, is thought to be the main source of campylobacter infections and an important reservoir of *C. jejuni* (Nachamkin *et al.*, 2008). The bacterium is a Gram-negative organism that appears as small, curved (of a spiral or corkscrew appearance) rods about 0.5 to 5 μm long and 0.2 to 0.8 μm wide ; these are motile with a single polar flagellum or bipolar flagella at the ends of the cell and are not spore forming (Debruyne

et al., 2008; Nachamkin *et al.*, 2008). The bacterium is a microaerophilic organism that needs lower concentration levels of oxygen to survive, usually 3-5 % O_2 and around 2-10% CO_2 (Nachamkin *et al.*, 2008). In old cultures, cells might produce coccoid bodies that are regarded as a degenerative shape as compared with the dormant form of the organism (Debruyne *et al.*, 2008). The *C. jejuni* NCTC11168 strain was the first for which genome was sequenced and originated from a human clinical isolate from 1977 (Parkhill *et al.*, 2000). The evaluation of the morphologies of this strain shows that it exhibits as straight rod in shape, perhaps developing from sub-culturing and adopting to lab conditions in contrast to the original isolate which exhibits a spiral shape (Figure 1.1).

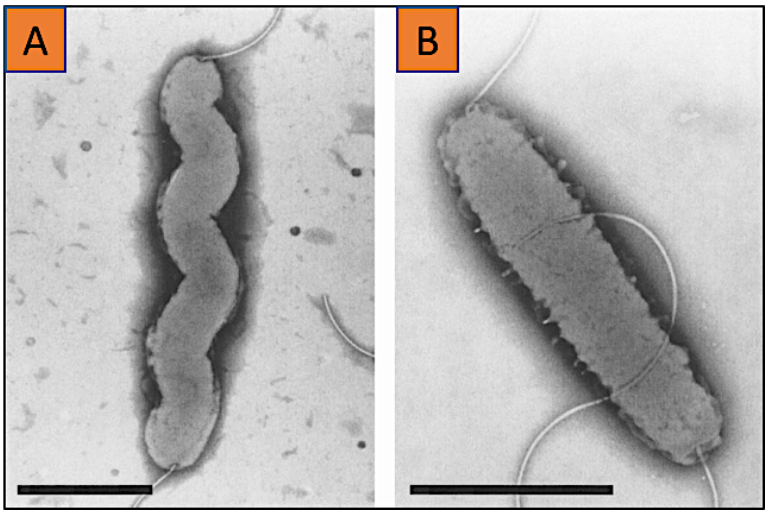


Figure 1.1: Electron microscope images of *C. Jejuni* variant NCTC 11168. Panel (A) The original variant NCTC 11168-O (**O**- Original) appears spiral with a polar flagellum. **Panel (B)** Image of NCTC 11168-GS (**GS**- genome sequence) which exhibits a rod shape. Strain GS was used for genome sequencing. Bars 1µm. Images from (Gaynor *et al.*, 2004).

Campylobacter infection in humans is linked with rare complications and diseases including reactive arthritis, Guillain-Barre syndrome and self-limiting gastroenteritis

(Doorn *et al.*, 2008). A typical case is characterised by fever, diarrhoea, and abdominal cramps.

Campylobacter jejuni has evolved resistance to a variety of antibiotics including penicillians, vancomycin, sulfamethoxazole, trimethoprim, cephalosporins and rifampicin (Epps *et al.*, 2013). Those that have been infected with *Campylobacter* spp., the prognosis is that in most of the cases, a patient will make a good recovery when treated with fluids and electrolytes rather than any specific targeted treatment. But, in extreme cases there is a need to be treated with antibiotics such as fluoroquinolones and macrolides which are administered generally (Silva *et al.*, 2011). Again, the resistance developed by *C. jejuni* to erythromycin and fluoroquinolones may reduce the effective treatment (Alfredson and Korolik, 2007).

1.2 *Campylobacter* taxonomy and metabolism

Theodor Escherich is recognised as the first person to report this bacterium in 1886. He found spiral bacteria in the colons of some children who had died from bacteria that he called 'cholera-infantum' (Butzler, 2004). However, in 1919 the name of the organism was changed to *Vibrio fetus* after it was isolated from aborted cattle fetuses (Smith and Taylor, 1919). In 1931 the name was altered to *Vibrio jejuni* after some researchers found certain bacteria in the jejunum of infected calves (Jones *et al.*, 1931). Later on, in 1963 by Sebald and Véron the name was transferred into the genus *Campylobacter* as a result of their need for microaerophilic conditions, and because of their metabolism and their DNA based composition for growth (On, 2001). *Campylobacter jejuni* strains are commonly unable to catabolise sugar molecules because they lack the phosphofructokinase enzyme (Velayudhan and Kelly, 2002). They get their carbon and energy mostly from amino acids and catabolism of small organic

intermediates (Guccione *et al.*, 2010). However, some studies reveal that certain *C. jejuni* strains have a unique characterised pathway for the metabolism of the L-fucose sugar, that is derived from the host through colonization of the intestine (Stahl *et al.*, 2011). Catabolic substrates are utilised for the tricarboxylic acid cycle (TCA cycle) and gluconeogenesis. *C. jejuni* encodes a full oxidative TCA cycle that changes the carboxylic acids to produce CO_2, ATP and decreasing the equivalents for respiration (Kelly, 2008).

The main feature of *C. jejuni* is that it is oxygen-sensitive microaerophile. This characteristic was obviously unknown during the latter part of the 1970s, when the culture of microaerobic clinical samples arose, this led to the ability to isolate and the understanding that *C. jejuni* was a main human pathogen (Ban Mishu, 2001). The *C. jejuni* microaerophilic nature is considered due to, not having any oxidative stress defence enzymes (Atack and Kelly, 2009), one potential reason could be because of the existence of some vital enzymes in the central metabolism that utilise oxygen-sensitive iron-sulphur clusters. These enzymes exist more commonly in obligate anaerobes, and include pyruvate and 2-oxoglutarate oxidoreductases, which are key TCA cycle enzymes (Kendall *et al.*, 2014).

1.3 Epsilonproteobacteria

The *Epsilonproteobacteria* have various metabolisms in different classes involving organisms having different life styles including free-living and pathogenic bacteria related to other species (Barbara *et al.*, 2006). The pathogens which are linked to the human digestive system include the *Helicobacter and Campylobacter* genera of the family *Helicobacteraceae* and *Campylobacteraceae* (Engberg *et al.*, 2000; Meinersmann *et al.*, 2002). The family of *Campylobacteraceae* currently including *Campylobacter, Arcobacter,* and *Sulfuospirillum* (Lastovica *et al.*, 2014), whilst another order includes

Helicobacter and Wolinella (Takai *et al.*, 2005). *Epsiolonproteobacteria* also plays a significant function in the biochemical cycling of hydrogen, sulfur and nitrogen, flux at water-hydrothermal areas (Takai *et al.*, 2005). An additional function is in producing conditions such as in anoxic water columns, sulfidic sediments and the oxygen-free zones in open ocean and costal environments (Sievert *et al.*, 2008). *Campylobacter* spp. are classified into several serogroups that depend on the protein (H) and lipopolysaccharide (O) antigens. All species of *Campylobacter* are recognised as human gastrointestinal pathogens also as animal pathogens. *C. jejuni*, *C.lari* and *C. coli* are the most common species leading to diarrhoeal disease in industrialised nations, however, *C. fetus* gives severe systematic infections especially to those at higher risk such as the elderly people and immunocompromised (Perez-Perez and Blaser, 1996). Even though *Campylobacter* spp. are motile generally for colonisation, however, *C. gracilis* is known to be non-motile, and in contrast *C. showae* has a multitude of flagella (Silva *et al.*, 2011). Furthermore, oxidase activity has been found in all species except for *C. gracilis* (Vandamme, 2000).

Helicobacter pylori is a common human pathogen that causes active chronic gastritis, peptic, duodenal ulcer disease and gastric cancer (Velázquez, 1999). They are curved into a spiral shape with sheathed multiple flagella. Commonly it is considered that the flagella and spiral morphology are crucial for colonisation of the intestinal mucus layer of the gastrointestinal tracts of humans and a diversity of animals (Eaton *et al.*, 1992;O'Rourke and Bode, 2001). Most *Helicobacter* spp. unlike *Campylobacter* spp., have strong ureolytic activity, particularly those linked with gastric mucosa. Also, oxidization or fermentation of the well-known carbohydrates does not occur (Owen, 1998). Additionally, *Helicobacter* spp have some virulence factors that are produced, damaging the gastric epithelial surface, for example, the CagA protein and vacuolating cytotoxin, which facilitate damage of the gastric mucosa (Luisa and Rainer, 2016).

Arcobacter spp. are similar to *Campylobacter* spp. and are spiral curved, Gram-negative bacteria, with motility driven by unsheathed flagella. However, they are recognised by their ability to grow in both aerobic and microaerobic conditions at temperature lower than 15º C (Lehner *et al.*, 2005). *Arcobacter* includes four species, *Arcobacter cryaerophilus*, *Arcobacter butzleri*, *Arcobacter nitrofigilis* and *Arcobacter skirrowii*. Research has suggested that *Arcobacter*, especially *A. butzleri*, could play a role in human enteric diseases. Also, *A. butzleri* has frequently been seen as a cause of human extraintestinal diseases. However, the potential virulence factors and the mechanisms of pathogenicity of *Arcobacter* spp are little known (Lehner *et al.*, 2005). Moreover, human faeces from stool testing of patients with severe diarrhoea are related to *A. cryaerophilus* and *A. butzleri* (Engberg *et al.*, 2000), and with *A. skirrowii* (Wybo *et al.*, 2004). However linking *Arcobacter* species and extraintestinal invasive diseases is not common, but its importance as a human diarrhoeal factor is still undetermined (Lehner *et al.*, 2005).

Wolinella succinogenes is considered a non-pathogenic, host-associated organism. It is a non-fermenting bacterium, and different from *Campylobacter* and *Helicobacter*, in that it can grow under anaerobic respiration, and in the presence of 2 % oxygen (Claudia *et al.*, 2003)

1.4 Genomic analysis of *Campylobacter jejuni*

The first genome sequence of *C. jejuni* NCTC11168 strain was reported in 2000 (Parkhill *et al.*, 2000). It has a circular chromosome of 1,641,481bp and is expected to encode 1,654 proteins sequences and 54 stable RNA species. Also, it has a dense genome, where 94.3% codes for proteins, and has a 30.6% guanine-cytosine content (GC-content) (Parkhill *et al.*, 2000). One characteristic of the *C. jejuni* genome is the nearly complete

absence of repetitive DNA sequences. There are, however, four repeated sequences within the entire genome; three duplicated or triplicated CDSs and three copies of the ribosomal RNA operon. Furthermore, no sign of any functional insertion sequence (IS) elements, retrons, transposons or prophages in the genome. Another fascinating characteristic of the *C. jejuni* genome is that relatively few of its genes are organised into operons exceptions being ribosomal protein operons and gene clusters that are included in extracellular polysaccharide biosynthesis (EP), lipooligosaccharide biosynthesis (LOS) and flagellar modification. These features give some unique genomic characteristics to C. *jejuni* (Parkhill *et al.*, 2000). Around thirty distinct polyG tracts were recognised which are commonly linked with "contingency" genes in other pathogenic bacteria. They can undergo changes in size by slipped-strand mispairing throughout replication and can affect gene expression because the tract-associated gene will be either transcribed or not depending on the number of G bases (Parkhill *et al.*, 2000). An example is *cj0045c*, a gene encoding for a hemerythrin-like putative iron binding protein which contains a polyG tract (Kendall *et al.*, 2014). Although *C. jejuni* plays a major role as a significant bacterial cause of food-borne disease across the globe, it displays a wide genetic diversity among the *C. jejuni* pathogen strains. Entire genome microarray analyses of 11 *C. jejuni* strains of varied origin discovered close to 30 genetic loci which are either absent or have variations in these strains. The majority of loci are associated with DNA modification, biosynthesis of surface structures and iron acquisition. The remainder are included in metabolism and regulatory processes (Dorrell *et al.*, 2001). This variation will define the unique characteristics of diverse pathogenic strains and the mechanisms that will reveal how the disease comes about.

Other sequenced strains of *C. jejuni* include 81-176 (Hofreuter *et al.*, 2006) RM1221 (Fouts *et al.*, 2005), and 81116 (Pearson *et al.*, 2007). These strains have

generally the same genome structures with some slight differences (Kendall *et al.*, 2014). The 81116 strain has polyG tracts but less than the 11168 strains, and is considered more stable genetically than 11168 (Pearson *et al.*, 2007). Indeed, a number of studies have demonstrated that the genetic instability of NCTC 11168 results in a marked difference between the expression of some genes between the genome sequenced isolate and the original clinical isolate, and an absence in colonisation ability in the lab-cultivated strain (Gaynor *et al.*, 2004). The *C. jejuni* 81-176 is a very infectious strain that displays distinct pathogenic characteristics. The strain was initially found after a United States (U.S.) epidemic of campylobacteriosis that happened as a result of contaminated raw milk (Korlath *et al.*, 1985). Recently, the genome sequence showed some new genetic loci, and a number of deletions, which probably explain the distinct pathogenic characteristics of this strain. Also, several open reading frames (ORFs) were found that are pseudogenes in RM1221 or NCTC 11168 but encode distinct functional proteins in *C. jejuni* 81-176 (Hofreuter *et al.*, 2006).

1.5 Epidemiology and Pathogenicity

Campylobacter jejuni for a long time has been known as a cause of bacterial food-borne infection, and the most predominant bacterial food-borne pathogen in industrial countries. The *Campylobacteriaceae* family contains 16 species and more than 90% of all human Campylobacteriosis infections are caused by *C. jejuni* and *C. coli*. *C. jejuni* can colonise the intestine of a chicken and to a lesser degree those in ducks and turkeys (Dasti *et al.*, 2010). People can get Campylobacteriosis infections due to ingesting polluted drinking water, raw milk, and chicken products (Silva *et al.*, 2011). The contamination of poultry products is the major source of foodborne infections. Epidemiological studies found that between 50-80% of campylobacteriosis infection are attributed to poultry,

especially the consumption and handling of undercooked or raw chicken meat (Kaakoush *et al.*, 2014). Infection by *campylobacter* spp. accounts each year for around 2.5 million cases in the United States and about 340,000 cases in the United Kingdom (Ban Mishu, 2001; Sheppard *et al.*, 2009). This statistic is 13 times the number of cases that accrue combined by *Salmonella* spp., *E. coli* O157:H7 and *Listeria monocytogenes.* In addition, the annual economic problem of *Campylobacter* infection is estimated to be £500 million in the United Kingdom and about $8 billion in the United States (Sheppard *et al.*, 2009).

Campylobacter spp. usually colonise the broiler flocks and continuing to stay within the affected birds till they are slaughtered, without displaying any signs of illness (Silva *et al.*, 2011). Investigation was established in New Zealand to study and analyse the molecular epidemiology of *C. jejuni* over 3-years by using multi-locus sequence typing. The studies found that approximately 60%- 81% of the trade poultry carcasses from the providers were infected with *C. jejuni* (Müllner *et al.*, 2010). The studies showed that many of the carcasses were infected from isolates related to a number of sequence types (ST). Also, the majority of poultry ST have existed in human infection cases, which suggests that poultry is a main source of human infection (Müllner *et al.*, 2010). The intestinal tract of the chicken can harbor a large number of *Campylobacter* spp. This tract, however, through processing, may rupture or leak and the contents are moved to the surface of the carcass. Consequently, *Campylobacter* spp. remains as a liquid film on the chicken skin (Bernaudat *et al.*, 2011). Survival and persistence of *Campylobacter* spp. are promoted by an appropriate micro-environment of the skin. Also, these pathogens are able to persist in the carcass under storage at 4℃ or frozen conditions (Silva *et al.*, 2011). Additionally, horizontal transmission of contamination from the environment to broilers and within the flocks is the most likely source of *Campylobacter* spp. The horizontal transmission occurs quickly once the birds are colonised by *Campylobacter* spp. (Carrillo

et al., 2004). The risk of colonisation and dissemination of *Campylobacter* spp. might be increased with several elements such as the high flock size, wild birds, environmental water supplies, insects, faecal contact, rodents, personnel and other animals (Adkin *et al.*, 2006; Horrocks *et al.*, 2009). Poultry feed does not seem to contribute to *Campylobacter* spp. spread. *Campylobacter* dies quickly in poultry feed and is sensitive to dry conditions, but if water becomes mixed in the feed, it might behave as a transporter for horizontal transmission in a broiler house after the *Campylobacter* spp. colony is established (Silva *et al.*, 2011). Moreover, the possible distribution of *Campylobacter* spp. from breeding flocks by vertical transmission has continued to be discussed. Many researchers have widely studied the vertical transmission within fertile chicken eggs raised in laboratory conditions with no exposure to any external farmyard areas, revealing that they sustained colonisation by *Campylobacter* spp. Furthermore, the carrier rate of *C. jejuni* in the cecum content showed high levels which suggested that the colony was already existent in the fledglings prior to release to the farmyards (Silva *et al.*, 2011). Further, it has been shown that post intracloacal inoculation of one day old broiler fledglings, the *C. jejuni* dissipates very quickly in the lymphoid tissues and stays for lengthy timescale. Additionally, the existence of *Campylobacter* spp. in the chicken egg and hatchery might show possible vertical transmission (Silva *et al.*, 2011).

The prevalence of *Helicobacter* and *Campylobacter* species in broilers has been recognised using species-specific PCR. The PCR studies revealed that *H. pullorum* and *C. consisus* were found in levels of about 31.8 % and 22.6 %, respectively, in chicken samples (Manfreda *et al.*, 2011;Batz *et al.*, 2012). Although *C. jejuni* is the most common cause of bacterial gastroenteritis, it is also thought to be a major causative agent of Guillain Barré syndrome (GBS), where the immune system harms the peripheral nervous system (PNS). It is described by acute or sub-acute symmetrical ascending motor

weakness, areflexia, and mild-to-moderate sensory abnormalities (Nyati and Nyati, 2013; Bae *et al.*, 2014), after infection by *C. jejuni.* Molecular mimicry between the ganglioside epitopes on the human nerves and sialylated lipooligosaccharide on the cell envelope of the pathogen *C.jejuni* leads to a cross-reactive immune response, which results in autoimmune-driven nerve damage (Nachamkin *et al.*, 2002; Nyati and Nyati, 2013).

1.6 Virulence factors of *Campylobacter jejuni*

The pathogen *Campylobacter jejuni* is considered to be the common cause of severe bacterial enteritis of humans in industrialised nations such as the United Kingdom and United States (Blaser *et al.*, 1983; Gaynor *et al.*, 2004). The causes of colonising and infections of the intestinal tracts are not very well understood. Other than the polysaccharide capsule and cytolethal distending toxin (CDT), the *Campylobacter spp.* have a deficiency in the homologues of virulence features, which are common to other pathogens (Guerry, 2007).

1.6.1 Flagella and motility

The movement of the cells and chemotaxis are the most common distinct virulence factors in *C. jejuni.* They play a crucial function in bacterial colonisation of various environments (van Vliet and Ketley, 2001). The motility of *C. jejuni* is driven and mediated by the polar flagella and the accompanying cork-screw movement permits this pathogen to penetrate the mucus layer (Lertsethtakarn *et al.*, 2011).The flagella are an significant factor for the invasion and colonisation of epithelial cells (Newell *et al.*, 1985; Lee *et al.*, 1986). They are composed of flagellin, O-linked glycosylated proteins, encoded by the *flaA* and *flaB* genes (Dasti *et al.*, 2010). The two genes display a high sequence identity of about 95% (Guerry *et al.*, 1990; Nuijten *et al.*, 1990). The expression of *flaA* gene is found at a higher level than the *flaB* gene, and the flagellum of

Campylobacter contains mainly the FlaA protein. The FlaA protein is also reported to be necessary for colonisation (Nachamkin *et al.*, 1993). In addition, a mutation in *flaA* shows a non-motile ability and reduced colonisation of chickens; however, a *flaB* mutant exhibits a slightly declined motile ability (Wassenaar *et al.*, 1994; Hermans *et al.*, 2011). Also, in *C. jejuni* there are a number of genes involved in the flagella system that are regulated by the two flagellin sigma factors σ^{28} and σ^{54} as obtained from microarray data, in strain of NCTC 11168 (Carrillo *et al.*, 2004).

1.6.2 Chemotaxis

The regulation of motility in *C.jejuni* is controlled by a chemotactic signaling system, which permits the bacteria to follow suitable chemical gradients in surroundings and respond to them by flagellar rotation (Lertsethtakarn *et al.*, 2011). From previous studies, mucin and bile, which are secreted by the intestinal tract and gall bladder, affect the chemotaxis of *C. jejuni*, which has a strong attraction to mucin and this was confirmed by flagellar rotation studies (Bhavsar and Kapadnis, 2007). In *C. jejuni* the genome sequence shows it has orthologues of the chemotaxis genes including the *cheA, cheW, cheB, cheY, cheV,* and *cheR*, and these are important for the commensal colonisation of the chick intestine (Hendrixson and Dirita, 2004). The mutants in *cheR* and *cheB* showed a lower capability to colonise the cecum of chickens (Hermans *et al.*, 2011). Furthermore, they may affect the invasion and adhesion of INT 407 cells and the studied mutants were revealed to be able to colonize mice but failed to create disease in animal models (Dasti *et al.*, 2010; Hermans *et al.*, 2011). *C. jejuni* is capable of altering chemotactic motility from environmental signals through Transducer Like Proteins, TLPs, which are trans membrane sensory proteins (Korolik, 2010). The significance of these chemoreceptors in the chemotaxis signaling pathway is recognised through microarray and biochemical

analyses, which showed that the aspartate receptor CcaA of Tlp1 chemoreceptor (Cj506c) was signaling via the CheV rather than the CheW protein in the chemotaxis signaling pathway. This chemotactic signaling pathway could suggest a novel model for *C. jejuni* (Korolik, 2010).

1.6.3 Adhesion and Invasion

Another important aspect of *C. jejuni* pathogenesis is the nature of the interface with intestinal epithelial cells (Bhavsar and Kapadnis, 2007). Unlike *E. coli* and other pathogens, the adhesion of *C. jejuni* to host epithelial cells does not occur by pili or fimbria (Gilbreath *et al.*, 2011; Rubinchik *et al.*, 2012). However, a number of factors have been shown to contribute to the *C. jejuni* adhesion mechanism. For instance, FlaA the flagella filament, has been revealed to be important in adhesion (Yao *et al.*, 1994), but the role of flagella in adherence *in vivo* is less clear (Golden and Acheson, 2002). Furthermore, other adhesins in *C. jejuni* encoded by the genes *peb1*, *jlpA* and *cadF* have been shown as essential for adherence *in vitro* and colonisation *in vivo* (Jin *et al.*, 2001). CadF, an outer membrane protein, is essential for effective invasion and adherence. It was shown to bind specifically with fibronectin, which is situated on chicken epithelial cells (Konkel *et al.*, 1997). CadF and FlpA together can cause the re-arrangement of the actin cytoskeleton via the activation of integrin receptors to assist *C. jejuni* in internalisation. A *cadF* mutant showed significantly reduced colonisation of chicks in contrast to the wild type (Hermans *et al.*, 2011; Heimesaat *et al.*, 2014). The protein PEB1 is an antigenic protein factor displayed on the surface of *C. jejuni* which has a main function in adherence and host colonisation (Müller *et al.*, 2007). A mutation in *peb1A* has been reported to cause 100-fold decrease of *C. jejuni* attachment to HeLa cells , as well as reduced the ability of the pathogen to colonise the intestinal cells in rats (Rubinchik *et al.*, 2012).

1.6.4 Toxin production

The development of infection by *C. jejuni* from diarrhoea to severe bloody diarrhoea and occasionally septicaemia highlights the role of toxins in this disease. However, the only proven toxin recognised is the cytolethal distending toxin (CDT), which is created also by a number of *Campylobacter* spp. including *C. jejuni*, *C. coli*, *C. fetus*, *C. lari*, and *C. upsaliensis* (Dasti *et al.*, 2010).

The creation of CDT by *Campylobacter* was first described by Johnson and Lior in 1988 (Johnson and Lior, 1988), and subsequent research showed several strain difference for toxin activity in *Campylobacter* (van Vliet and Ketley, 2001). This toxin has been shown to encourage cell distension in mammalian cell lines such as Chinese hamster ovary (CHO) cells, HeLa cells and Caco-2 cells in which swelling, elongation and finally cell death are described (Whitehouse *et al.*, 1998). The holotoxin consists of CdtA, CdtB and CdtC proteins. CdtA and CdtC are known to be necessary in *C. jejuni* for binding to the host cell and to activate CdtB to create an active tripartite holotoxin, which exhibits full cellular toxicity (Lara-Tejero and Galan, 2001).

1.6.5 Iron Metabolism

Living organisms require iron which is vital for survival, particularly during electron transport and for roles as a cofactor for many enzymes in redox reactions (Vliet *et al.*, 2002). Moreover, the ferrous ion (Fe^{2+}) which is soluble can easily be transferred into the outer membrane through porins and then transported by a known transporter protein, FeoB, through the cytoplasmic membrane (Hermans *et al.*, 2011). *Campylobacters spp.* have established mechanisms to scavenge enough iron for growth and metabolism. However, iron contributes to various biological processes such as oxygen transport, DNA biosynthesis, gene regulation and the tricarboxylic acid (TCA)

cycle. Moreover, its interaction with proteins in mono or binuclear species, or as a more complex form as a part of haem groups or iron sulfur clusters is crucial (Andrews *et al.*, 2003). Iron also contributes in the creation of reactive oxygen species (ROS) like peroxides (ROOH), super oxide anions, and hydroxyl radicals (OH) and these force pathogens to manage oxidative stress and maintain intracellular iron homeostasis (Storz and Imlayt, 1999; Touati, 2000). In addition, the existence of genes encoding detoxification and multiple iron acquisition systems in the pathogen *Campylobacter spp*. suggest the essential function that iron has in *Campylobacter* virulence and gene regulation (Andrews *et al.*, 2003). However, iron concentration in the cells should be controlled, to prevent growth inhibition or death. So as a result, the protection of iron homeostasis is essential for living organisms which is achieved by regulating iron storage, iron transport and iron metabolism(Andrews *et al.*, 2003). The free iron concentration in a host cell is low and sequestered in host chelators such as haem, lactoferrin (in mucosal surfaces) and transferrin (in serum)(Vliet *et al.*, 2002). *C.jejuni* does not create siderophores, iron carriers secreted in response to iron restriction, but can use the siderophores enterochelin and ferrichrome, which are produced by other bacterial cells (Pickett *et al.*, 1992; Neilands, 1995).

1.6.6 Protein glycosylation

Glycosylation is a method of modifying proteins by adding carbohydrate molecules that is recognised as a significant part of pathogenesis in many bacteria. Glycosylation happens on pili, flagella and adhesion proteins (Power and Jennings, 2003). The glycosylation mechanisms of proteins involving numbers and types of protein, but the overall function of Gram negative bacterial glycosylation is still under research. Glycans can be linked to protein by the hydroxyl groups of serine or threonine residues

to establish an *O*-glycosylation or nitrogen of an asparagine residue to form N-glycosylation (Harald and Christine, 2010). The *O*-glycosylation modification functions in *C.jejuni* have been described to be essential for flagella modification and use Sialic acid-like sugars such as pseudaminic acid derivatives. However, N-glycosylation functions have been recognised to be included in the modification of surface associated proteins (Parkhill *et al.*, 2000).

The first characterisation of the bacterial N- glycosylation pathway in *C. jejuni,* was made when the glycosylation machinery was completely understood (Linton *et al.*, 2005). The major role of glycan modification in *C. jejuni* is unidentified, however, absence of heptasaccharide have led to variation in protein antigenicity (Kelly *et al.*, 2006). N-linked modification enzymes have been encoded by the *pgl* genes in *C. jejuni* (*cj1119c- cj1131c*) (Kelly *et al.*, 2006). However, the inactive *pglG* and *pglI* had little impact on the colonisation in chickens, while, a mutation in *pglD* caused the lack of poultry colonisation (Kelly *et al.*, 2006).

O-linked glycosylation, however, is encoded by *leg* genes or by *pse* genes which encode enzymes for joining the legionaminic acid (LgAm) derivatives (McNally *et al.*, 2006). This system of glycosylation is associated with motility, and also it maybe has a function in invasion and adherence of epithelial cells in chicken (Christine and Brendan, 2005).

1.7 Relationship between O_2 molecules and *C. jejuni*

Campylobacters spp are known as microaerophilic organisms and are thought to be adapted to specific niches that have low oxygen concentrations (Kelly, 2008). Routinely most strains of *C.jejuni* are cultivated in atmospheres of 3 - 10% oxygen and 5-10% carbon dioxide (Kelly, 2008). Also, they have been able to adapt to the difficult

limitation of the oxygen in the gut, perhaps because they depend on a single oxygen-dependent NrdAB-type ribonucleotide reductase (RNR) for DNA synthesis (Sellars *et al.*, 2002). However, during the life cycle of these bacteria they have been found to be able to survive when exposed to a high concentration of oxygen in the external environmental when they have to transfer between hosts (Kendall *et al.*, 2014). The correlation between *C.jejuni* and oxygen levels is not fully understood and therefore is considered to be one of the most important aspects of the biology of this pathogen (Kelly, 2008). The mechanism also for sensing of oxygen is unclear in *C. jejuni.* However, it is obvious from the genome sequencing of these bacteria, that many regulatory systems are absent compared with the aerobic bacteria (Parkhill *et al.*, 2000).

The ability of some strains of *C. jejuni* to grow under high oxygen (15-21%) has been stated in many studies. Adding antioxidants, such as catalase, histidine, superoxide dismutase to the growth medium was found to enhance the survival of *C. jejuni* at an oxygen level between 17-21% (Kaakoush *et al.*, 2007). This suggests that this organism might be more susceptible to free radicals than other aero-tolerant bacteria , but in the presence of antioxidants the resistance to oxygen will be increased (Kaakoush *et al.*, 2007). *C. jejuni* expresses some proteins such as truncated haemoglobin Ctb, protease HtrA, alkyl hydro-peroxidereductase AhpC and ferredoxin FdxA, which are involved in oxygen tolerance (Bury-Moné *et al.*, 2006). It also has proteins that add to the oxygen susceptibility of the bacterium, such as rubredoxin oxidoreductase and L-serine dehydratase (Yamasaki *et al.*, 2004). A cb-type oxidase also plays an important role in *C. jejuni* in that it permits effective energy conservation when oxygen is used as an electron acceptor (Jackson *et al.*, 2007). Recently, Kendall *et al* (2014) have identified protection proteins that possess ligand-binding features typical of hemerythrins which are detailed further in section (1.8.7).

1.8 Hemerythrins

1.8.1 Overview

Hemerythrins are oxygen-binding proteins that were first found in some marine invertebrates species, including the marine worms *priapolida* and *sipunculid*s and lamp shell *brachiopods* (Stenkamp, 1994). These organisms often possess multimeric hemerythrin in their blood cells and a monomeric myohemerythrin in their muscles (French *et al.*, 2008). Both are very similar in function and structure and are often referred to the oxidation and ligation states of the iron centre (Karlsen *et al.*, 2005). Hemerythrin differs from the haemoglobin and haemocyanin, of which the latter appear in all vertebrate and in many invertebrate species, in that it possesses a non-haem di-iron atom site that binds one molecule of oxygen (Karlsen *et al.*, 2005). The hemerythrins have been shown in a wide variety of bacteria (367/2236), a few archaea (21/150) and (4/135) eukaryotic genomes (Alvarez-Carreño *et al.*, 2016). Additionally, over 400 distinct hemerythrin-like sequences in prokaryotic genomes have been detected (French *et al.*, 2008).

1.8.2 Conserved motifs within the hemerythrin sequences

Sequence alignments among different hemerythrins reveal important characteristics. In Figure 1.2 the sequences of hemerythrin and myohemerythrins are aligned for *Phascolopsis gouldii* (phgou_myo/phgou_hem), *Methylococcus capsulatus* (McHr*)* and *Desulfovibrio vulgaris* (Dcrh_Hr). The residues essential for an iron-binding centre can be identified from protein sequences by the existence of the characteristic H.... HxxxE.... HxxxH....HxxxxD motifs. Also, the organisation of these motifs is conserved. This has led to the identification of proteins with the same motifs and organisation as being putative hemerythrins (Holmes *et al.*, 1991; French *et al.*, 2008).

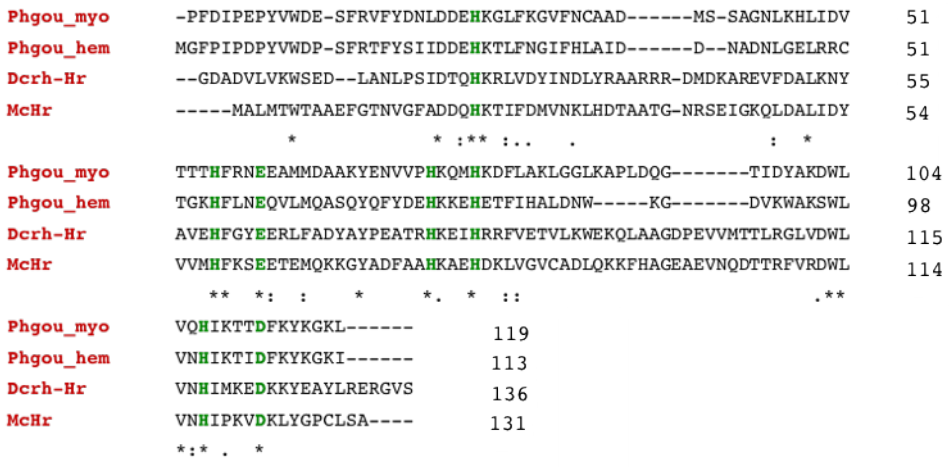


Figure 1.2: Sequence alignment of some hemerythrins using CLUSTALW. Sequences from two eukaryotic hemerythrins from *Phascolopsis gouldii* hemerythrin, *Phascolopsis gouldii* myohemerythrin, and two bacterial hemerythrins from *Methylococcus capsulatus* and *Desulfovibrio vulgaris*. Green highlighted regions indicate motifs conserved amongst hemerythrin proteins. Adapted from French *et al.*, 2008.

1.8.3 Hemerythrin Structure and O_2 Binding Mechanism

The quaternary structure of the hemerythrins consists of four alpha helices in a bundle fold that surrounds the di-iron site (Figure 1.3). The iron ions are usually bound to the protein by carboxylate side chains: one aspartate, one glutamate; and imidazole groups of five histidine residues: three histidine ligands to Fe-1, and the other two ligands to Fe-2 (Hendrickson *et al.*, 1975; Karlsen *et al.*, 2005; French *et al.*, 2008).

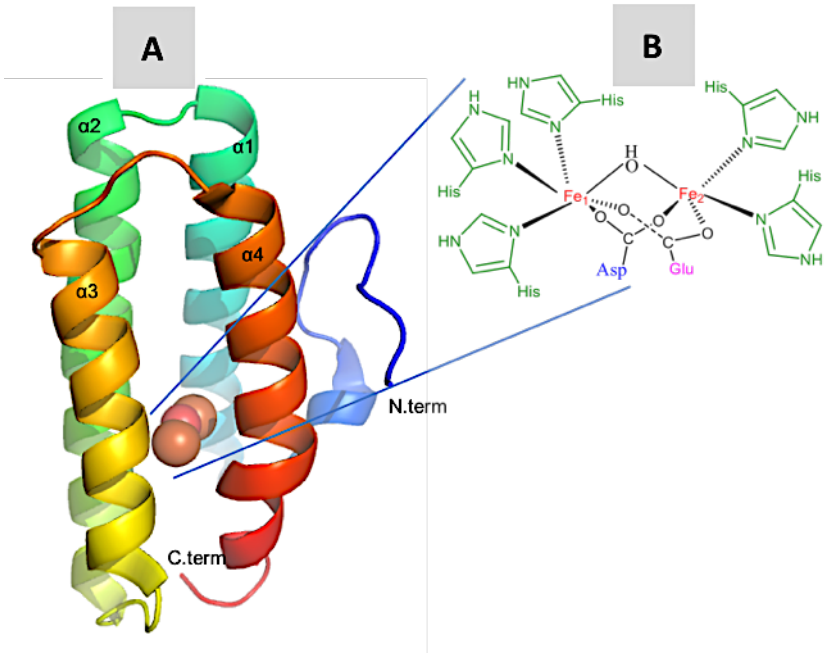


Figure 1.3: Structure of a typical four helix bundle hemerythrin protein with its iron centre. **Panel (A)** shows the molecular model structure of the hemerythrin from marine *Phascolopsis gouldii* worms (PDB: 3AGT). **Panel (B)** shows the classical structure of the iron-binding site of hemerythrin proteins, adapted from (Farmer *et al.*, 2001).

The di-iron centre in the hemerythrin undergoes redox cycling whilst its Fe ions move between Fe^{2+} and Fe^{3+} forms. The hemerythrin iron redox cycling can be divided into three different distinct stages: deoxy-hemerythrin (Fe^{2+}-OH-Fe^{2+}), oxy-hemerythrin (Fe^{3+}-O-Fe^{3+}-OOH) and met-hemerythrin (Fe^{3+}-OH-Fe^{3+}- any other ligand). In the deoxygenated hemerythrin form (deoxy-Hr), the Fe-1 and Fe-2 atoms in their Fe^{2+} oxidation stage are bridged by a hydroxide group (OH^-) (Figure 1.4). During oxygenation, the reaction between O_2 and Fe-2 will produce a single oxygen atom (μ-oxo) bridge, resulting in the creation of a peroxyl group (OOH^-) and an oxidation of both iron atoms to ferric iron forms (Fe^{3+}). This reaction is reversible. After exposure to water, the di-iron

site of oxy-hemerythrin will undergo auto-oxidation to form met-hemerythrin stage, where the two Fe^{3+} atoms are bridged by a single oxygen atom. The reduction of met-hemerythrin will return again to the deoxy-hemerythrin stage (Kao *et al.*, 2008).

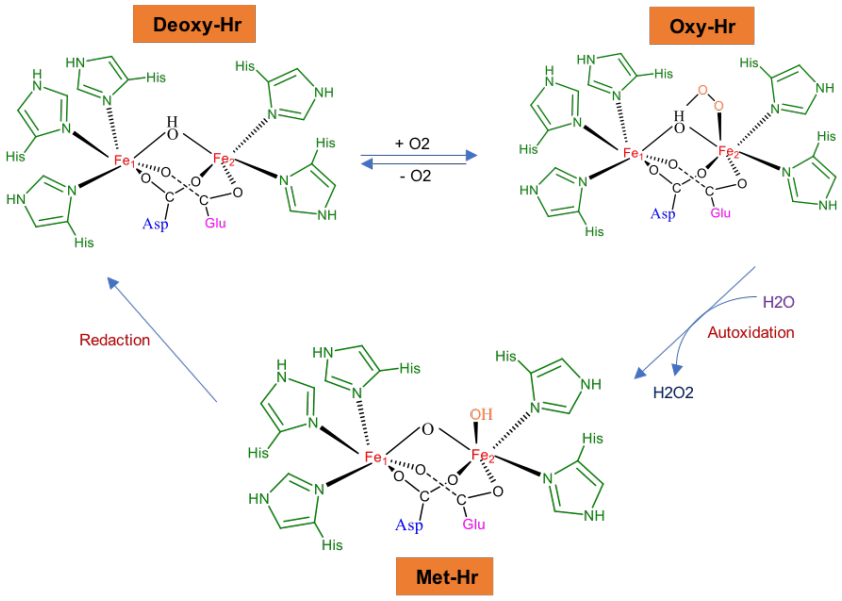


Figure 1.4: The three forms of hemerythrin protein and redox cycling. Adapted from Kao *et al.*, 2008.

1.8.4 Hemerythrins in bacteria

The first report of hemerythrin proteins in bacteria was in *Desulfovibrio vulgaris* and *Methylococcus capsulatus* (Xiong *et al.*, 2000). After that, researchers found during genomic sequencing more than 400 putative hemerythrins in prokaryotes (French *et al.*, 2008). They found also that this protein is more common in anaerobic and microaerobic species, and predominates in motile bacteria (French *et al.*, 2008).

The majority function of the hemerythrin is thought to be related to their oxygen binding properties. This binding of oxygen could be as a storage mechanism; such as in animals, or could be functioning as a delivery system for oxygen such as in *Methylococcus capsulatus* (Karlsen *et al.*, 2005). In addition, binding of oxygen probably acts as a detoxification mechanism, or as a sensory mechanism like DcrH in *Desulfovibrio vulgaris*. On the other hand, it is possible that some hemerythrins may work with iron as a storage or detoxification mechanism, while others work in ways unrelated to the binding of oxygen or metal (French *et al.*, 2008).

Single and multi-domain hemerythrin proteins have been identified in these prokaryotes. However, the single domain is more widely distributed than the multi-domain, which is found in specific groups such as magnetotactic bacteria; *Alphaproteobacteria* or sulphate-reducing bacteria, *Deltaproteobacteria* (French *et al.*, 2008). In the anaerobic bacterium *Desulfovibrio vulgaris*, *DcrH*, it encodes a putative methyl-accepting chemotaxis protein (MCP), which has a hemerythrin domain at its C-terminus. This hemerythrin domain is often called DcrH-Hr and is supposed to contribute in oxygen sensing for aerotaxis (Xiong *et al.*, 2000). When oxygen is bound to the C - terminal Hr domain of DcrH, the N - terminal signalling domain is highly ordered, allowing the transduction of downstream chemotactic signalling pathways.

The methyl-accepting chemotaxis protein (MCP)-hemerythrins are also found in *Rhodospirillum*, *Magnetospirillum*, *Magnetococcus*, and *Shewanella*; in each group of these organisms, the closest relation to the MCP part of the sequence are other MCPs from related organisms, as opposed to other MCP-Hrs, which signifies that in each group the fusion of MCP and hemerythrin took place independently in each species (French *et al.*, 2008). Another example of hemerythrins as oxygen sensors is the enzyme responsible for cyclic di-(3'-5')-guanoside monophosphate (c-di-GMP) synthesis in *Vibrio cholerae*,

diguanylate cyclase. This enzyme, named Vc Bhr-DGC, has a C-terminal diguanylate cyclase (DGC) domain fused to an N-terminal hemerythrin (Schaller *et al.*, 2012). In low oxygen environments, *V. cholerae* form biofilms to aid survival, and this process is induced by increased levels of c-di-GMP (Tischler and Camilli, 2004). The hemerythrin domain acts to regulate the activity of the fused DGC; Vc Bhr-DGC is apparently ~10 times more active as a DGC when the hemerythrin domain is in the diferrous, unbound state than when in the diferric, oxygen bound state (Schaller *et al.*, 2012). Therefore, this hemerythrin ultimately regulates the biofilm formation of *V. cholerae* under low oxygen conditions.

Another example of a single domain hemerythrin in prokaryotic is the McHr from *Methylococcus capsulatus* (Bath). It was found that a hemerythrin-like protein in *Methylococcus capsulatus* (Bath) was stimulated to four times its level when the cells were grown with the particulate methane mono-oxygenase (pMMO) in high concentrations of Cu (Karlsen *et al.*, 2005). The McHr protein in *M.capsulatus* most probably facilitates transport of oxygen to pMMO found in intra-cytoplasmic membranes and thus supports the concentration of dioxygen, particularly with the high level of pMMO produced at 30 μM Cu (Kao *et al.*, 2008).

In strictly anaerobic bacteria, such as *Clostridium sp.,* which are recognised as sensitive to oxygen, it has been found that the hemerythrin proteins might detoxify reactive oxygen species for the organism and this may be followed by a re-reduction generating hydrogen peroxide (French *et al.*, 2008). With aerobic bacteria such as *Burkholderia thailandensis*, *Burkholderia mallei* and *Burkholderia pseudomallei*, the hemerythrin proteins have been encoded between the subunits of ubiquinol oxidase and a putative iron-dependent regulatory protein. It has been suggested that these proteins

play a role in controlling intracellular levels of oxygen or may play a role in respiration under low-oxygen conditions (French *et al.*, 2008).

In contrast to the above point, it is possible that some hemerythrin proteins might have functions relating to their iron-binding nature. For instance, in magnetotactic bacteria such as the *Magnetospirillum magneticum* strain, the hemerythrins bind iron atoms which may be related to the formation of magnetite crystals used in magneto-sensing (French *et al.*, 2008).

1.8.5 Hemerythrins in the *Campylobactereae*

In the *Campylobactereae* the characteristic oxygen-binding of hemerythrin proteins might be involved in the microaerophilic phenotype of these organisms, but their mechanism *in vivo* remains unclear (Kendall *et al.*, 2014). The hemerythrin-like sequences are shown to appear almost in all *Campylobacter species*, which possesses C-terminal Hr domain, was expressed in *E. coli* and purified. The UV-vis absorption and resonance Raman spectra of Cj0241-Hr found a hemerythrin-like iron centre that might play an oxygen-sensing role in this bacterium (Kendall *et al.*, 2014). All of species are obviously related together and have high homology between each strain. In addition, there are a number of single-domain hemerythrins(French *et al.*, 2008). In a number of species, hemerythrins are found to be incomplete, without the C-terminal part HxxxxD motif. However, a study of DNA sequences has shown that the C-terminal part is existing, but it is coded in a different reading frame (French *et al.*, 2008).

The *C. coli*, *C. jejuni* and *C. upsaliensis* hemerythrins appear to be classified into five orthologous groups. However, they differ in term of their polypeptide lengths. Group (1) occurs in all strains and is 240-270 amino acid residues in length. Group (2) has been identified in most strains and is 133 amino acid residues in length, however, this group is

not found in *C. jejuni* RM1221 and *C.coli*. Group (3) has been detected in all strains, except the *C. jejuni subsp. doylei* strain (269.97) and is 199 residues in length. The genes which encode Group (4) and (5) hemerythrins are adjacent, suggesting that they may be hetero-multimers or hetero-dimers, as noticed in some animal hemerythrins (French *et al.*, 2008). *Campylobacter* hemerythrins that were not within any of these five groups, however, were obviously still more similar to the hemerythrins of *Campylobacter* than other hemerythrins from different species. Additionally, most hemerythrin sequences have been found in motile species, raising the prospect that these proteins function as oxygen-sensing aerotaxis mediators (French *et al.*, 2008).

1.8.6 Analysis of *C. jejuni* NCTC11168 Hemerythrins

The analysis of the genome of *C. jejuni* strain NCTC11168 shows three different genes encoding proteins with iron and oxygen binding hemerythrin-like characteristics; *cj0241*, *cj1224* and *cj0045*. There is also a fourth possibility; a pseudogene (*cj0027*) related to these other genes (Kendall *et al.*, 2014). Protein sequences of the Cj0241, Cj1224 and Cj0045 hemerythrins reveal that all include conserved histidine motifs H, HxxE, HxxxH and HxxxxD, which are important to bind the di-iron centre critical for hemerythrin function (Figure 1.5) (French *et al.*, 2008).

The Cj0241 protein is 133 residues in length and has a single domain (Kendall *et al.*, 2014). The *cj0241* gene is located with uncharacterised gene, *cj043c*. However, the two genes *iscS* (*cj0240c*) (iron-sulphur cluster) and *nifU* (*cj0239c*) (nitrogen fixing), which are characterised Fe-S cluster pathways in *Campylobacter jejuni*, are directly located upstream of *cj0241* gene (Figure1.6 A). There is a suggestion that there might be a relationship between these two genes with hemerythrin *cj0241* gene particularly in Fe-S syntheses (Kendall *et al.*, 2014).

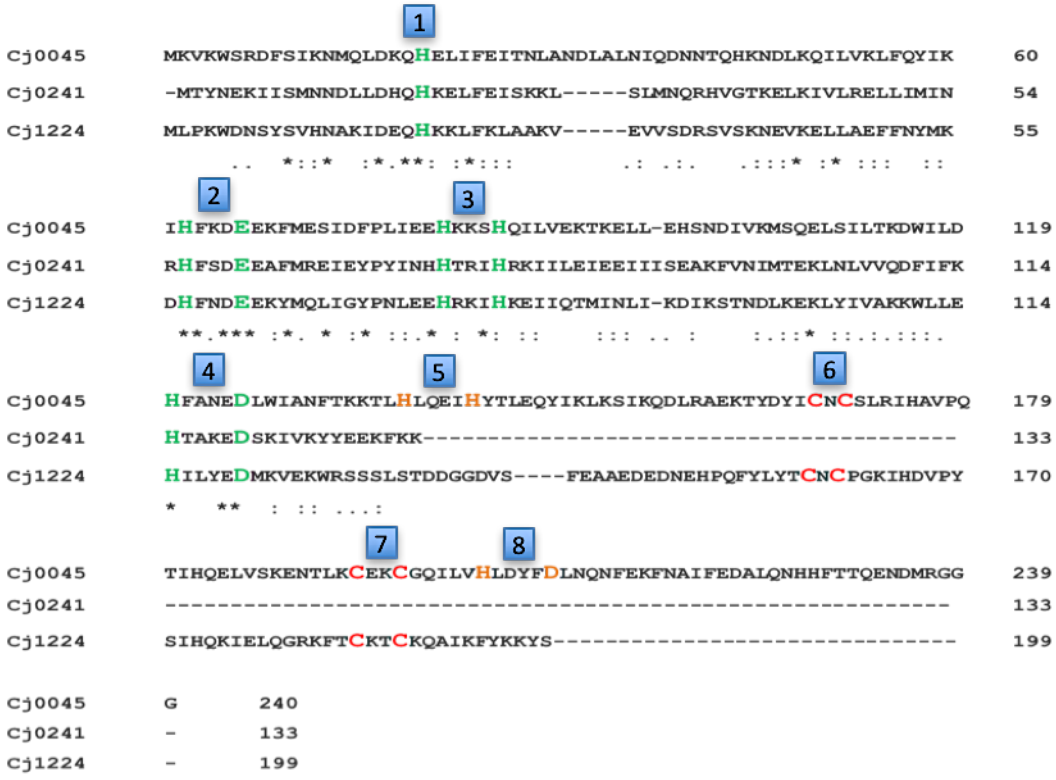


Figure 1.5: Sequence alignment of the three putative hemerythrins (Cj0045, Cj0241 and Cj1224) of *C. jejuni* NCTC 11168. The alignment was carried out using ClustalW. Green highlighted regions (1-4) indicate motifs conserved amongst hemerythrin proteins. Orange highlighted regions in Cj0045 (5, 8) indicate possible additional motifs. Red highlighted regions in Cj10045 and Cj1224 (6, 7) indicate potential metal-binding motifs. * = Position with fully conserved residue. : = Position with residues having strongly similar properties. . = Position with residues having weakly similar properties. Adapted from : Kendall *et al.*, 2014.

Cj1224 is 199 residues in length, with a hemerythrin domain at the N-terminal end. It also has CxC and CxxC motifs in a C-terminal extension of the protein, which have been suggested to be involved in additional redox reactions or metal binding (Kaakoush *et al.*, 2007). The *cj1224* gene is located upstream of an uncharacterised gene,

cj1225 in one direction and downstream of *cj1222* and *cj1223* in the other direction. *cj1222* and *cj1223* encode a histidine kinase and a response regulator respectively which form the well-known DccR/DccS two-component regulator system (Figure 1.6 B). The DccR/DccS regulator system is predicted to regulated transcription of *cj0200c, cj0606, cj1004, cj1356c, cj1626c and cj1723c* genes which are putative periplasmic or membrane proteins (Wösten *et al.*, 2010). These proteins have been shown to be involved in colonisation of the gastrointestinal tract in chicken (Mackichan *et al.*, 2004).

Cj0045 encodes a 240-amino acid protein. The *cj0045* gene is located next to uncharacterised gene function, *cj0044*, and *cj0046* gene which is a pseudogene (putative sodium: sulfate transmembrane transport protein) (Figure1.6 C). It also has CxC and CxxC motifs like Cj1224 with potential redox or metal binding properties, as well as two additional histidine containing motifs that could be potentially be involved in farther iron binding (Kendall *et al.*, 2014).

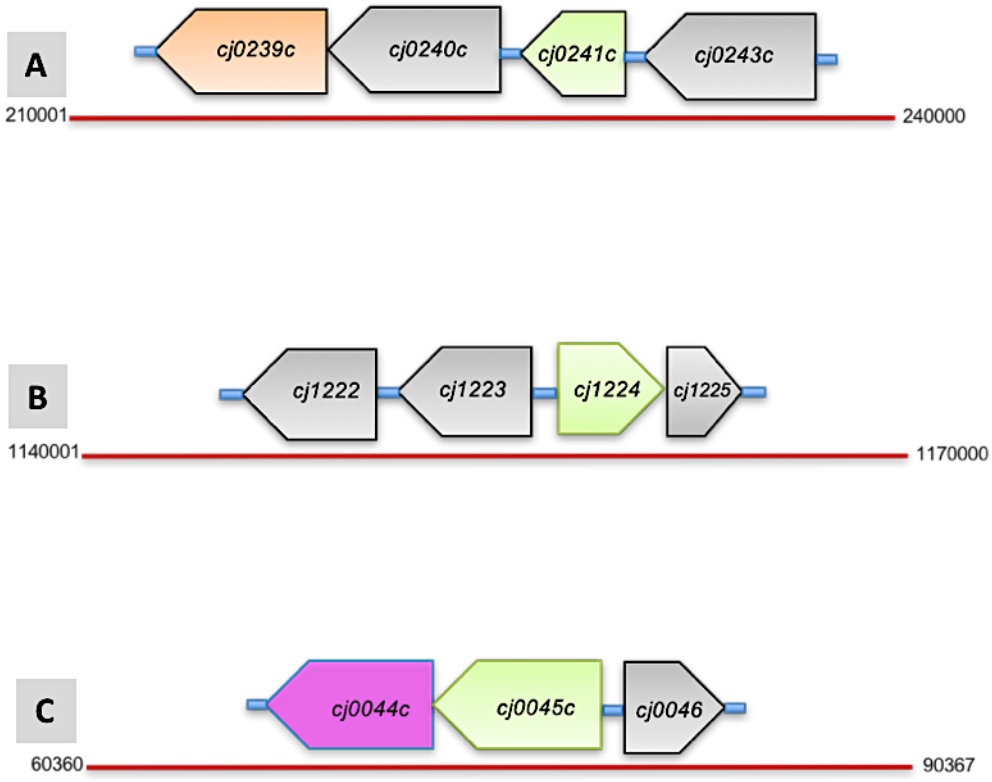


Figure 1.6: The three hemerythrin proteins positions in *C. jejuni* genome. Panel (A) the *cj0241c* gene is directly upstream of the *cj0240c* (iscS) and *cj0239c* (nifU) genes, which may be involved in the formation or repair of [Fe-S] clusters. **Panel (B)** the *cj1224c* gene is located adjacent to *Cj1223c* and *cj1222c* that contain two component systems. **Panel (C)** the *cj0045c* gene locate between uncharacterised gene, *cj0044c,* and pseudogene *cj0046.* The lines and the numbers are indicating the location of the genes within the sequence of the NCTC 11168 genome. Adapted from: Gundogdu et al., 2007.

1.8.7 Functions of *C. jejuni* Hemerythrin Proteins: relationship to Fe-S cluster proteins

The Iron-sulphur (Fe-S) cluster proteins are characterised by the common ability to bind iron-sulphur clusters. They can be used by organisms for a large variety of biochemical functions, such as iron or oxygen-sensing, electron transport, energy metabolism and anaerobic respiration (Py and Barras, 2010). Fe-S clusters can be present in different forms: simple single iron atom with four cysteine residues, and more complex forms which appear in rhombic [2Fe-2S] and cubic [4Fe-4S] formations (Imlay, 2006). Their features result from the ability of iron atoms to cycle between ferrous (Fe^{2+}) and ferric (Fe^{3+}) oxidation states, and this allows the binding of nitrogen or oxygen atoms and the efficient transfer of electrons in organic substrates (Imlay, 2006) .

C.jejuni has metabolic enzymes that use Fe-S clusters as co-factors in the citric acid cycle (TCA). These enzymes are Pyruvate: acceptor oxidoreductase (Por) and 2-oxoglutarate: acceptor oxidoreductase (Oor), which catalyse key intermediate steps for the TCA cycle metabolism in *C. jejuni*. Por catalyse the conversion of Pyruvate to acetyl-CoA and Oor catalysis the conversion of Oxoglutarate to Succinate in the cycle (Figure1.7) (Daucher and Krieg, 1995). Por and Oor are highly sensitive to direct oxygen damage (Kendall *et al.*, 2014). Kendall and co-workers have reported that the deletion either the hemerythrin *cj0241*(codes for *herA*) or *cj1224* (codes for *herB*) genes led to an aerobic growth defect and decreased the activities of the two enzymes, Por and Oor, after exposure to 21% of O_2 (v/v). Under anaerobic growth conditions with slow shaking strains subsequently showed recovery of the activity of Por and Oor to similar levels seen in the wild-type strains. This suggests that one role of hemerythrin is to prevent the Fe-S cluster enzymes, Por and Oor, from being damaged microaerobically or during oxygen transients (Kendall *et al.*, 2014).

Por : pyruvate: acceptor oxidoreductase.
Aan: aconitase.
Icdh: isocitrate dehydrogenase
Oor : 2-oxoglutarate: acceptor oxidoreductase.
Suc : succinyl-CoA synthetase.
Frd : fumarate reductase.
Fum : fumarase.
Mdh : malate dehydrogenase.
Mqo : malate: quinine oxidoreductase.
GltA : citrate synthase.
Nuo : flavodoxin/ferredoxin: quinone oxidoreductase.
Fld : flavodoxin A.
Fd: ferrodoxin.

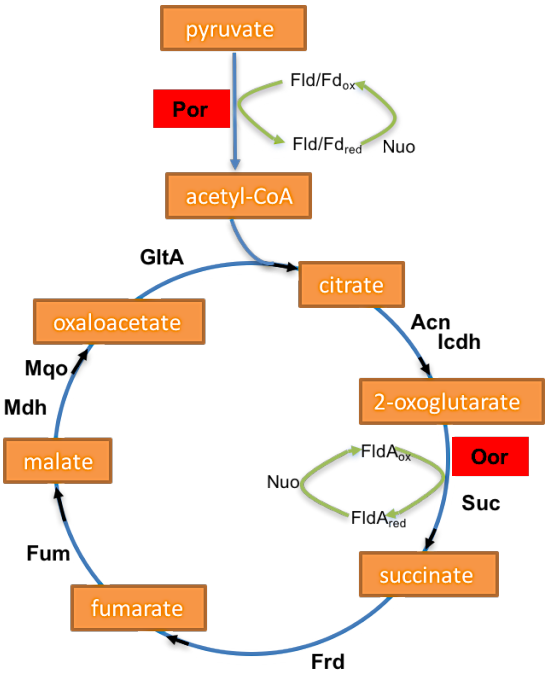


Figure1.7: Schematic illustration the role of Por and Oor enzymes in the Citric acid cycle in *C. jejuni*. *C.jejuni* has metabolic enzymes Pyruvate: acceptor oxidoreductase (Por) and 2-oxoglutarate: acceptor oxidoreductase (Oor) (highlighted in red colour square) that use in Fe-S cluster as co-factors. Adapted and updated from : Kendall *et al.*, 2014 .

The Cj0045 hemerythrin protein is insoluble under routine purification conditions (Kendall *et al.*, 2014), which has thus restricted its further characterisation. The closest homologue of this gene is *fedA* in 81-176 strains (145 residues in length), characterised previously as a **f**lagellar co-**e**xpressed **d**eterminant (fedA). *fedA* however is missing a large segment of 96 amino acids from the C-terminus which in Cj0045 contains the CXC and CXXC domains. Examination of a different reading frame of the intergenic region between *fedA* and the adjacent *cjj81176_0082* gene reveals a short sequence, fedA*

which contains the missing CXC, CXXC and HXXXXD motifs (Figure 1.8). This region is very like the C-terminal part of *cj0045c* gene, involving the existence of a polyG tract that maybe contributes in slipped-strand mispairing, and for this reason the Cj0045 role may be different from the other hemerythrins, Cj0241and Cj1224, and may not be related to enzyme protection (Kendall *et al.*, 2014).

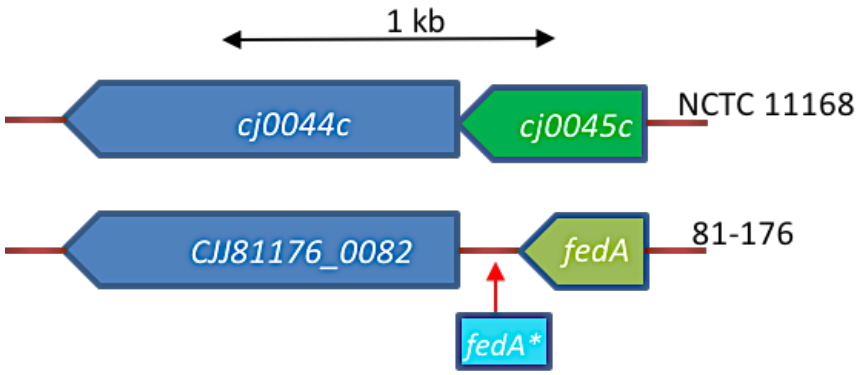


Figure 1.8: A comparison of the homologous open-reading frames in the *C. jejuni* genome of strains NCTC1118 & 81-176. In NCTC 11168, the 723 bp of *cj0045c* is located directly upstream of *cj0044c*, however, in 81-176 there is an intergenic region of 284 bp between the downstream *cjj81176_0082* & *fedA*. This intergenic part is as annotated using a different reading frame as *fedA** and could code for the missing C-terminal extension to the FedA protein. Adapted from :Kendall *et al.*, 2014 .

1.9 Crystal structures of the hemerythrin proteins exhibit a water tunnel in some bacteriohemerythrins

The hemerythrins from marine invertebrates are expected to have a di-oxygen storage role, using the non-heme di-iron centre to carry the O_2 molecule within their vascular system for distribution to tissues. Corresponding to this role, the autoxidation processes of the oxy-Hr form for these invertebrate hemerythrins are very slow, with a half-life ($t_{1/2}$) around 20 h before it is changed to the met-Hr form (Farmer *et al.*, 2001).

The autoxidation half-life of the oxy-form of the hemerythrin-domain in the DcrH sensor protein from the opportunistic pathogen *Desulfovibrio vulgaris* is around 1 minute (Xiong *et al.*, 2000). The X-ray crystal structure of DcrH-Hr reveals a water tunnel that shows significant differences from the animal hemerythrins (Figure 1.9). This tunnel involves mainly hydrophobic residues that permits the quick autoxidation process of the oxy-stage in DcrH-Hr (Isaza *et al.*, 2006). This quick autoxidation process corresponds with the role of this hemerythrin domain as the di-oxygen sensor in the chemotaxis receptor, which could use a redox-dependent conformational alteration to transmit a signal for the presence of oxygen to the adjacent methylation domain (Isaza *et al.*, 2006). The sensory mechanism might include a rotation of the side chain of Leu 115 in the water tunnel to permit the oxygen molecule access to the di-iron site of the structure (Chen *et al.*, 2015). The autoxidation half-life of the McHr hemerythrin from *Methylococcus capsulatus* is less 1-hour, which is longer than that for the hemerythrin-like domain of the sensor protein DcrH, but quicker than that of the hemerythrins in marine invertebrates. The autoxidation half life of L114Y and L114F variants of McHr are increased compared with wild type, consistent with restricting the water tunnel and their behaviour has become more similar to the marine invertebrates. It is argued that because the invertebrates do not have the putative water tunnel, the small size and non-polar nature

of oxygen allow it to reach into the iron centre via "micropores" and the oxy-Hr stage could be stable and not disposed to the autoxidation reaction (Chen *et al.*, 2015). So, a water tunnel is not essential for the hemerythrin proteins that are working only as storage mechanism for di-oxygen as in invertebrate animals.

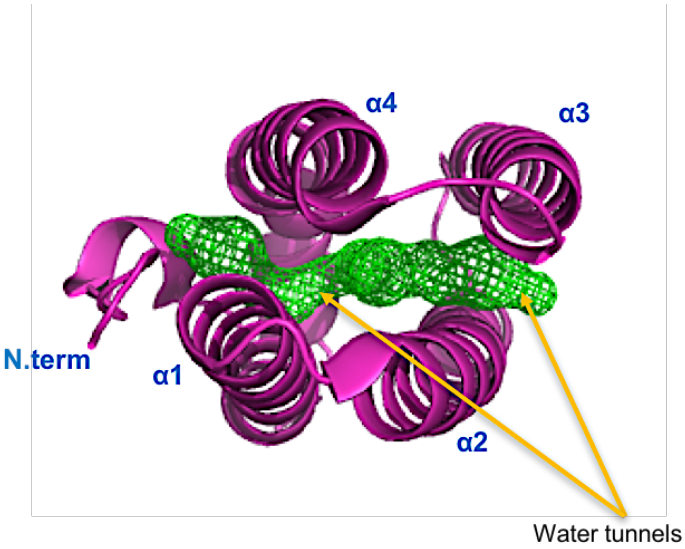


Figure 1.9: An example of the putative water tunnels in Met- DcrH-Hr from *D. vulgaris*. The helices are shown in purple, and water tunnels are highlighted in green mesh. Adapted from: Chen *et al.*, 2015.

1.10 Zinc-Binding Cysteines

Cysteine gives proteins an excellent functional group because of the distinct chemical features of its thiol group (Giles *et al.*, 2003a). It has a wide redox activity and unique metal binding characteristics, which makes it the main catalytic component of enzyme function. In redox activity, cysteine residues with sulfur in many oxidation conditions have been recognised in proteins (Giles and Jacob, 2002). A further characteristic of cysteine is its exceptional metal binding ability. The thiol group ligand

strongly binds with metal ions such as Zn^{2+}, $Fe^{2+/3+}$, Cu^{+} and Cd^{2+} (Giles *et al.*, 2003b). The interaction of cysteine and metals is important for metal properties to be used in biology. Experimentally, metal ion–binding proteins are characterised and recognised using gel electrophoresis, nuclear magnetic resonance spectroscopy, electrophoretic mobility shift assay, mass spectrometry and metal-affinity column chromatography absorbance spectroscopy (Lu *et al.*, 2012).

The *Campylobacter* hemerythrins proteins; Cj1224 and Cj0045, have got as mentioned in section (1.8.6) cysteine motifs within them. These motifs have a prediction from the Phyre^2 service might be a zinc finger within it that will be discussed in chapters 5 and 6. Zinc is the 2nd most popular transition metal after iron ion in living organisms (Andreini *et al.*, 2006). The bond between zinc ions and sulfur play a role in keeping the zinc homeostasis inside the cell (Giles *et al.*, 2003b).

Zinc-binding protein structural motifs are seen in the well-characterised zinc fingers (Figure 1.10). The zinc finger (ZF) classical domains were originally found eukaryotes, until one was recognised in the Ros transcriptional regulator protein in *Agrobacterium tumefaciens*. The ZF of Ros has a Cys_2His_2 motif and binds DNA in a particular mode and folds in a domain bigger than its eukaryotic equivalent which involves 58 amino acids (Chou *et al.*, 1998). Ros homologues have been identified in various bacteria and thus suggesting that the ZF structural motif is extensive also in the prokaryotic kingdom (Malgieri *et al.*, 2015). Zinc fingers usually have a Cys_2His_2 or Cys_4 coordination environment (Andreini *et al.*, 2006). They often stabilise the protein structure for interaction with other types of proteins and biological molecule such as DNA and RNA (Berg and Shi, 1996). Although the functional roles of many zinc finger proteins are not much understood, most are annotated in proteins acting as transcription activators (Razin *et al.*, 2012).

Figure 1.10: A zinc finger (ZF) binding motief in the major groove of the DNA. The zinc finger from protein ZNF217 bound to DNA complex (PDB: 3UK3). Source: Vandevenne *et al.*, 2013.

1.11 Aims of the project

Hemerythrins are oxygen-binding proteins; they differ from haemoglobin and haemocyanin in that they possess non-haem di-iron sites that bind one molecule of O_2. They have been reported in a wide variety of bacteria particularly in microaerophilic and anaerobic bacteria. The majority of *Campylobacter species* have been shown to have multiple hemerythrin-like proteins many of which contain a C-terminal domain of unknown function but with high homology among strains (French *et al.*, 2008). The analysis of the genome of *Campylobacter jejuni* strain NCTC11168 displays three different genes encoding proteins with iron and oxygen binding hemerythrin-like characteristics: *cj0241, cj1224 and cj0045* (Kendall *et al.*, 2014).

The overall aim of the project was to answer the question of why these bacteria have three hemerythrins, and what are the differences between them in terms of function. Other goals include discovering the importance of key residues in these hemerythrins and their role in oxygen sensing and protection in these pathogenic bacteria. A major approach was to try to determine the structures of the three hemerythrins Cj0241, Cj1224 and Cj0045 by using X-ray crystallography and to compare their structures to known models of O_2 binding.

Chapter 2

Materials and Methods

This chapter will describe in general the procedures that were used during this project. It is divided into two parts and the first part discusses the general experiments used with all three *C. jejuni* hemerythrin proteins. The second part discusses specific experiments used with each protein. Also, it includes the chemicals, equipment and reagents used in this study. The study of all three hemerythrins in general was aimed at crystallising the proteins and achieving their 3-D structures using X-ray crystallography.

A study using X-ray crystallography has a requirement for large amounts of soluble and properly folded pure protein. However, it has become clear that achieving this amount of protein from natural sources is often impossible due to low expression yields. There is a discussion of the theory behind the methods used with these three types of hemerythrins. The chapter also explains in detail the laboratory methods used to clone, overexpress and purify the hemerythrin proteins.

2.1 Bacterial strains

The bacterial strains that used in this research are recorded in Table (2.1).

2.2. Storage of strains and plasmids

All strains were kept at 4°C for up to two weeks on solid media. For longer periods, however, strains were saved in the form of glycerol stocks, in 50 % (v/v) sterile glycerol, and kept at -80°C. Plasmids stocks were saved in dH_2O at -20°C.

Table 2.1: The bacterial strains used in this thesis

Bacterial Strain	Source
Cj0241_6xHis Plasmid constructed using the pET21b vector with the *cj0241c* gene inserted in the multiple cloning site between the NdeI/XhoI restriction sites to give a C-terminal hexa-histidine tag.	Prof. Dave Kelly's lab
Cj1224 _6xHis Plasmid constructed using the pET21b vector with the *cj1224c* gene inserted in the multiple cloning site between the NdeI/XhoI restriction sites to give a C-terminal hexa-histidine tag.	Prof. Dave Kelly's lab
6xHis_Cj1224 Plasmid constructed using the pET28b vector with the *cj1224c* gene inserted in the multiple cloning site between the NdeI/XhoI restriction sites to give an N-terminal hexa-histidine tag.	Produced for this thesis
Cj1224_notag Plasmid constructed using the pET21b vector with the *cj1224c* gene inserted in the multiple cloning site between the NdeI/XhoI restriction sites.	Produced for this thesis
Cj0045_6xHis Plasmid constructed using the pET21b vector with the *cj0045c* gene inserted in the multiple cloning site between the NdeI/XhoI restriction sites to give a C-terminal hexa-histidine tag.	Prof. Dave Kelly's lab
6xHis_Cj0045 Plasmid constructed using the pET28b vector with the *cj0045c* gene inserted in the multiple cloning site between the NdeI/XhoI restriction sites to give an N-terminal hexa-histidine tag.	Produced for this thesis

2.3. Growth media and conditions

2.3.1 Luria-Bertani (LB)

The strains of Bacteria were normally allowed to grow in sterile Luria-Bertani medium (Thermo Fisher Scientific, Catalog #: 22700025), according to the protocol mentioned in Sambrook and Russell, 2006, as shown in the appendix. All media were prepared using distilled water (dH_2O). To sterilise the media, an autoclave was used with standard conditions: 121°C, 20 minutes, 15 psi. Solid media were kept at 4°C for up to a month and liquid media were stored at room temperature. The suitable antibiotics were added to LB-agar after cooling the media to around 40-45 °C. The LB agar was then distributed and poured into Petri plates and allowed to solidify.

Solid media were kept at 4°C for up to a month, and liquid media was kept at room temperature.

2.3.2 M9 Minimal medium

M9 minimal media (Thermo Fisher Scientific, Catalog #: A1374401) is a commonly used microbial growth medium, which has only salts, and nitrogen compounds in a buffered form. It is normally complemented with additional carbon and amino acid sources (see in the appendix).

2.3.3 Super Optimal broth with Catabolite repression (SOC medium)

The Super Optimal broth with Catabolite repression (SOC) (Thermo Fisher Scientific Catalog #: B9020S) is a nutrient-rich liquid broth that is supplied as a liquid (see appendix). In this medium, the transformation of competent cells with recombinant plasmid took place.

2.4 Transformation using chemically competent cells

Transformation followed the heat shock method (Hanahan, 1983). Recombinant plasmids were transformed into BL21 (DE3) competent cells. Eppendorf tubes with 50 µl aliquots of competent cells were taken from the freezer at -80°C and defrosted in ice for 3-5 minutes. When defrosted, and prior to incubation in ice for 20 minutes, 2 µl of plasmid DNA preparation (around 100 ng/µl) was added. The mixture was then exposed to a heat shock at 42 °C for 45 seconds and placed back in the ice for 2 minutes. 950 µl of LB media was added and then the cells were incubated at 37°C at 250 rpm for 45 minutes. This was then plated (around 130 µl – 150 µl) on LB agar having the suitable selective antibiotic. The plates were then incubated at 37° C overnight.

2.5 Protein overexpression

An alternative technique can be utilised to express all the three hemerythrin proteins involved in this project, where minor changes lead to optimisation and increase in the expression yield, these will be discussed in the following chapters.

2.5.1 Small-scale overexpression

A 1 ml culture sample that was grown overnight in 5ml LB media was inoculated in 50 ml LB media in a conical flask (250 ml) which contained the appropriate selective antibiotic. This was then incubated at 37˚C and shaken at 250rpm. The cells grew until the approximate optical density (OD_{600}) at 600nm reached 0.6-0.8 before adding 50 µl IPTG to a concentration of 1mM. 1ml samples were then taken as pre-induction samples (P). The first flask was incubated at 37˚C with the second at 25˚C. Then, 1 ml samples were taken after 1, 3 and 5 hours of incubation from each flask. Finally, the flasks were incubated overnight. All the samples were centrifuged at 13000 rpm for 1 minute at 4˚C

in a bench top centrifuge SIGMA 3-16PK, and the cell pellets were kept at -20˚C. The samples were then analysed using SDS-PAGE.

2.5.2 Expression Level and Solubility Test

Protein stability and solubility are important factors when attempting to crystallise proteins (Bolen, 2004). The solubility of the protein was tested by using the BugBuster® reagent (Novagen), this is an industrial product for rapid bacterial cell lysis. By using the manufacturers procedure, the cells were lysed rapidly. A combination of 0.1 ml BugBuster reagent plus 5µl Benzonase were mixed with the cell pellets (from 1ml of culture) in the Eppendorf tube, slowly added with frequent pipetting until the cells were totally suspended. Benzonase is an endonuclease enzyme and has a property that destroyed all forms of RNA and DNA that existed in the sample. The mixture in the tube was then incubated at room temperature for around 10 minutes prior to centrifugation at 17000 rpm, at 4˚C for 10 minutes in the refrigerated centrifuge SIGMA 3-16PK. The insoluble pellets were re-suspended in 0.1 ml ddH_2O, while the soluble supernatant protein was placed in a new tube.

2.5.3 Measurement of protein concentration

The concentration of the protein was calculated by using the Bradford method. The protein molecules bind to the dye of Coomassie Brilliant Blue G-250, producing a changing colour in the protein solution sample. The changing of the colour was measured by using a spectrophotometer and protein concentration was determined by using the Bradford method as shown in the appendix. Reagents used in this method were bought from BioRad.

2.5.4 Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis (SDS-PAGE)

Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis (SDS-PAGE) method was used normally through this project to assess the expression and solubility of the target protein. SDS-PAGE was accomplished according to a standard protocol (Laemmli, 1970). The Gel and buffer materials in this thesis were used, using the 'Manual of Molecular Cloning'. All SDS-PAGE gels described in this thesis were 12% acrylamide concentration.

2.5.4.1 Gel preparation

Gels are generally polymerised between two plates of glass in a gel caster. A 1 mm or 1.5 mm gel plate was used and the gels involved a 12% resolving solution with a 6% stacking gel. The mixture of resolving solution (see appendix) was transferred by a pipette into a gel caster, and was then overlaid with propane solution to permit the gel to set. Once set, the gel was rinsed and the propane solution carefully removed using filter paper. The stacking solution (see appendix) was made and transferred by pipette to the top of resolving gel. A comb was placed at the top to make the sample wells. The comb was then removed after the gel polymerised and was ready for electrophoresis.

2.5.4.2 Sample preparation

The protein concentration in the samples to be analysed on the gel was estimated using the Bradford method. The samples, which contained 10 – 15 µg of total protein, were supplemented with 15 – 20 µl of dH_2O. 2 µl of Reducing Agent (RA) and 5 µl of Sample Loading Buffer (SLB) from Invitrogen (Life technologies) were mixed in and then the solution was boiled for 2-3 minutes at 100˚C.

2.5.4.3 Running condition

After loading the samples in the gel, it was then immersed in a running buffer, which had a composition of 14.4 g /l glycine, 1 g/1 SDS and 3 g /l Tris- HCL with a final pH 8.8. Electrophoresis was conducted at 200V until the tracking dye had moved to the bottom of the gel, which took around 40-45 minutes.

2.5.4.4 Staining and de-staining

To visualise the resolved proteins, the gel was coloured in 0.1%[w/v] Coomassie blue (R250) stain (see appendix) and gently rocked for 20-30 minutes. It was then transferred to a de-staining solution (see appendix), and incubated with gentle rocking until the protein bands were clearly visible. Mark12^{TM} Unstained Standard was used to calibrate the gel.

2.6 Large-scale overexpression

Once successful conditions were found for target protein expression, two culture flasks having 500ml LB media with appropriate antibiotics were inoculated with 10 ml culture samples grown overnight. These cultures were grown at suitable temperature with shaking at 250 rpm until the OD_{600} achieved 0.6. The mixtures were then induced with 1mM IPTG and allowed to grow under the same conditions overnight. The cell-free extract (CFE) was then separated from the cell debris via centrifugation at 14,334 g for 20 minutes at 4˚C using an F500 rotor (Beckman). The cells were then gathered in a 50 ml falcon tubes and harvested by centrifugation again at 5,000 g for 30 minutes to remove any excess media. The supernatant was disposed and the cell pellets kept at -20° C until needed.

2.7 Protein purification

For X-ray crystallography, a large volume of high purity, homogeneous and concentrated of the protein is typically required. During working on the three hemerythrin proteins, Cj0241, Cj1224 and Cj0045, different types of column chromatography were used throughout this project to purify them involving affinity chromatography, ion exchange and gel filtration. Generally, the chromatography was performed on an AKTA Pure System. The purity of the target protein was usually estimated by SDS-PAGE electrophoresis. A Bradford assay was used to measure the protein concentration.

2.7.1 Cell disruption

The initial step in the purification method of all target proteins is cell lysis. This was conducted by the process of sonication that uses high frequency sonic pulses to break down the bacterial cell walls. Cell pellets were defrosted and re-suspended in an appropriate buffer before homogenisation. The cell suspension was distributed between two 20ml vials and placed on ice. Sonication was performed using a Soniprep150 machine and a medium probe was employed at amplitude of 16-microns in 2 cycles of 20 seconds. The cell-free extract (CFE) was separated from the cell debris via centrifugation at 70,000 g for 10minutes at 4˚C using a JA-25.50 rotor (Beckman).

2.7.2 Purification of recombinant proteins by nickel affinity chromatography

The CFE containing soluble His-tagged recombinant protein was injected into a 5 ml HisTrap HP column (GE Healthcare) connected with an AKTA Pure machine. Histidine residues in a tagged protein connect to the nickel ions in the column. The target protein connects to the column with high affinity and high specificity; however, the non-His-tagged proteins will pass through the column. After that, to elute the bound

recombinant His-tagged protein from the column, a gradient of increasing concentration of 0.5 M imidazole was used. The imidazole is the functional group in the histidine side chain (Figure 2.1), and the increase in this imidazole concentration in the buffer will compete with the histidine tag for binding with the nickel column. The purity and amount of protein were analysed by SDS-PAGE.

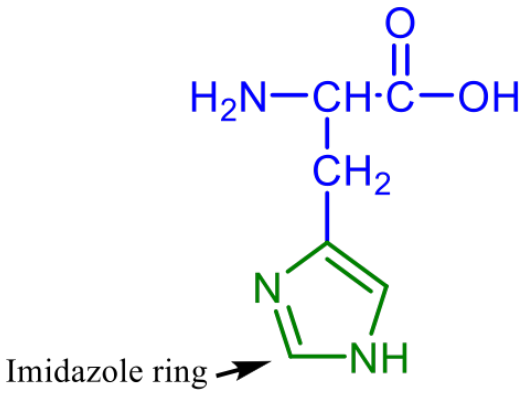


Figure 2.1: Histidine contains an imidazole functional group with the molecular formula $C_6H_9N_3O_2$.

2.7.3 Ion exchange chromatography

In this method, proteins are separated by using ion exchange chromatography based on their general net surface charge. The level of amino acids that are on the surface layer are used to evaluate the net surface charge of a protein, as well as their dependence on the pH (Hames and Hooper, 2005). The pI of a protein is the same as the pH of its condition or solution when there is no overall surface charge, and the overall charge on the protein at this point is neutral. Variations in the pH of the protein solution can lead to changes of the surface charge for example if it is lower than the protein pI, this leads to a

net positive charge and if higher then the net effect will be negative. Generally, ion exchange chromatography can be split into two main types; 1) cation exchange chromatography utilises resins with negative charge, which bind with protein of positive charge, and 2) anion exchange chromatography utilises a resin with positive charge to bind with a negative charged protein. After loading the protein sample onto an ion exchange chromatography column in low ionic strength buffer, oppositely charged proteins to connect with the beads of the column. The concentration of the ionic solution is then increased gradually by adding salts such as sodium chloride, these salts can react within the column competing for the charged groups and allowing the proteins to elute from the column (Figure 2.2a). DEAE sepharose (GE Healthcare) is an example of an anion exchange chromatography material that contains a matrix of cross-linked sepharose beads with DEAE bare on its surface, and it was used to purify Cj1224 and Cj0045 hemerythrin proteins which are detailed further in sections (2.11.10) and (2.12.1) respectively.

2.7.4 Gel filtration

The Gel filtration method, referred to as size exclusion chromatography, is a technique to separate a mixture of proteins according to their mass by allowing them to pass through a column that is filled with porous beads of polyacrylamide or carbohydrate polymer (e.g. agarose or dextran). Large molecules cannot enter the smaller pores in the beads, so they are not free to roam outside of the beads. However, smaller molecules and ions can enter the pores, and therefore will elute after the large molecules have eluted. A Hi-Load Superdex 200, 1.6x60 cm column (GE Healthcare life science) was the gel filtration column that is used in this project, and it has a separation range for molecules with molecular weights between 10000 and 600000 Da (Figure 2.2b). Samples of 1-2ml

were applied on the column and gel filtration was performed in buffer containing 50 mM Tris-HCl pH 8.0 and 0.5 M NaCl at flow rate 1.5ml/min.

The gel filtration chromatogram is also useful to estimate the molecular weight of proteins. This can be done by comparing partition coefficient, Kav, for a protein to a calibration curve of partition coefficients plotted against the log of the molecular weight for a specific gel filtration column (Figure 2.3). The result can also help in estimating the oligomeric state of the protein.

$$Kav = \frac{\text{Eluted volume} - \text{Void volume}}{\text{Total volume} - \text{Void volume}}$$

2.8 Protein dialysis

The dialysis method was used to remove unwanted small molecules such as salts or reducing agents (Figure 2.2c). Protein dialysis was conducted using a 12 mm or 14 mm dialysis tubing of semipermeable membrane (Sigma-Aldrich, UK) with a molecular weight cut-off (MWCO) of 10 kDa.

2.9 Protein concentration

At various stages of the protein preparation it was necessary to increase the protein concentration in solution. This was achieved using Vivaspin centrifugal concentrators (GE Healthcare), which contain two chambers, divided by a polyethersulfone (PES) membrane having pores of a controlled size. The devices are available in different sizes to work with volumes between 50µl to 20ml and a molecular weight cut-off (MWCO) from 3000 to 100000 Dalton. Generally, the centrifugation was done at 4715 rpm in a bench top centrifuge SIGMA 3-16PK at 5-10 °C until the protein sample reached the required concentration as measured using a Bradford assay.

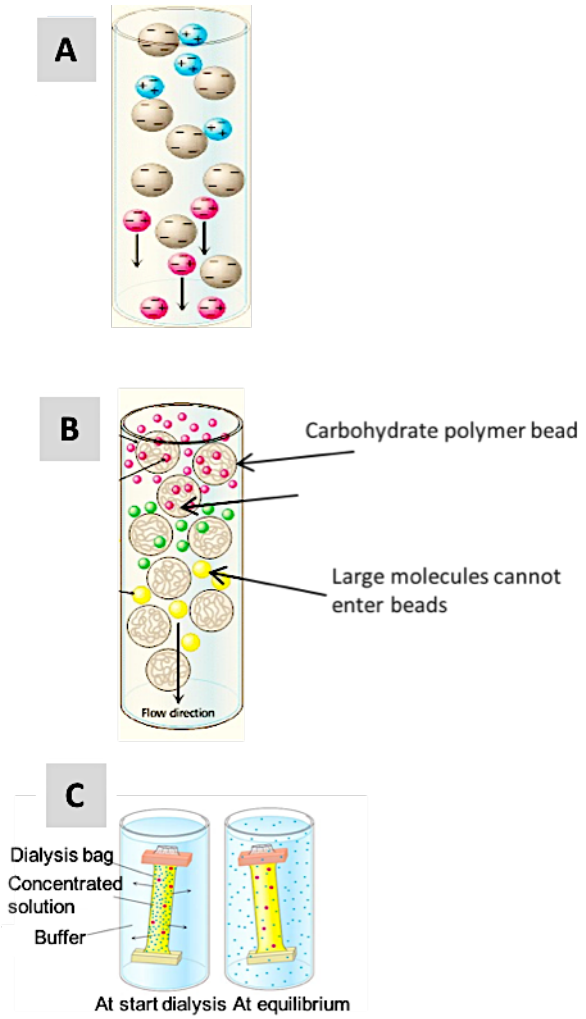


Figure 2.2: Protein purification strategies. Panel (A) Ion-exchange chromatography. This technique splits out proteins of different charge. Those that have positive charge bind to negative charged beads, whilst negative ones flow through. **Panel (B) Gel filtration chromatography**. A small solution of proteins is allowed to flow into a column containing porous beads. In this case larger proteins cannot protrude into the internal volume of the beads, and hence emerge quicker than the smaller ones. **Panel (C) Dialysis**. Protein molecules (red) are contained inside the dialysis bag, however small molecules (blue) diffuse through to the surrounding medium. Adapted from: Stryer, 1995.

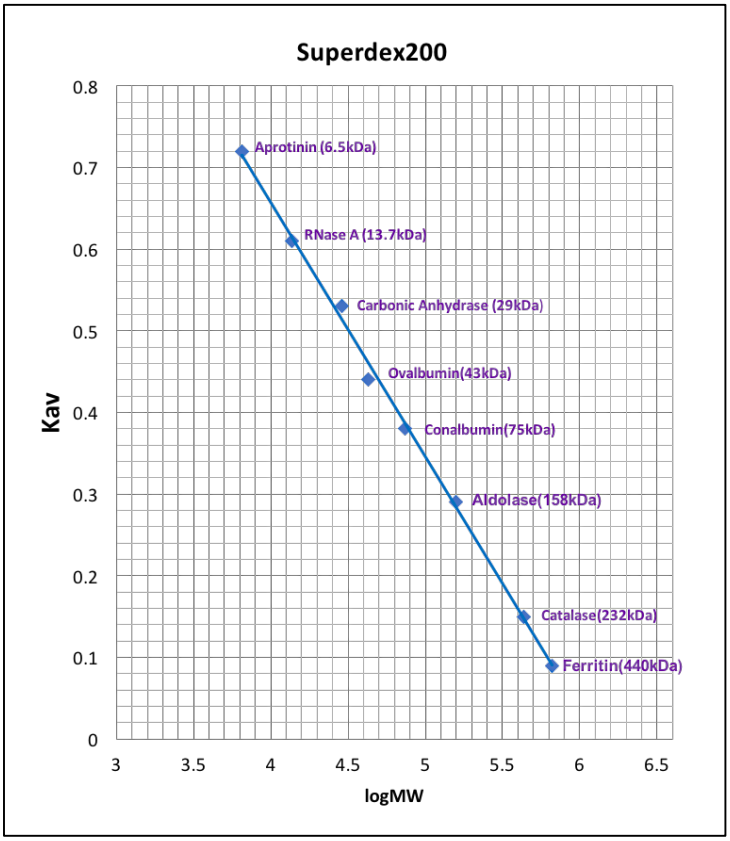


Figure 2.3: Calibration curve for gel filtration Superdex 200 column done with Gel filtration calibration kits (Ge Healthcare).

2.10 Material and Method: Cj0241-Hr

The following section describes in detail all the procedures used to obtain the 3-D structure of Cj0241-Hr protein.

2.10.1 Protein purification

Over expression was performed aerobically in *E. coli* strain BL21 (DE3) transformed with the construct Cj0241_6xHis. The cells were grown in LB media supplemented with ampicillin antibiotic (50 mg ml^{-1}) and the desired protein induced by 1mM IPTG when the culture OD_{600} reached 0.6 at 25°C. The cells were harvested by centrifugation at 9000 rpm for 20 minutes at 4˚C, and stored at –20˚ C. The Cj0241-Hr protein purification was carried out by nickel affinity chromatography. The cell pellet, around 4g, was re-suspended in a total volume of 20 ml of binding buffer (buffer A: 20mM phosphate saline buffer (PBS), 0.5M NaCl, 20mM imidazole at pH 7.4) before disruption by sonication (Soniprep150 machine, 16-microns in 2 cycles of 20 seconds).

The cell-free extract (CFE) was separated from the cells debris via centrifugation at 70000g rpm for 10minutes at 4˚C, using a JA-25.50 rotor. The protein concentration in the CFE was 7.5mg/ml. Therefore, the total protein in the CFE was 150 mg. The CFE containing soluble His-tagged Cj0241-Hr protein was then injected into a 5 ml HisTrap HP column connected to an AKTA purifier system at a flow rate of 5ml/min at room temperature and equilibrated with buffer A. 6ml fractions were collected for unbound material. The column was washed with a column volume (CV) of 5ml of binding buffer A. The recombinant hemerythrin Cj0241 protein was then eluted from the column by 50ml (10CV) gradient of increasing imidazole concentration from 20mM to 500mM using the elution buffer (buffer B: 20mM PBS, 0.5M NaCl, 500 M imidazole at pH 9.2). 3ml fractions were collected during the elution. The elution profile and the progress of

the purification were analysed using SDS-PAGE. Two fractions with the highest purity protein were combined (V=6ml, C=1.6mg/ml) and the total protein was 9.6 mg.

2.10.2 Protein crystallisation

To prepare the purified Cj0241-Hr protein for crystallisation experiments, a sample was concentrated by centrifugation at 4˚C using a VivaSpin column with a 10,000 Dalton MWCO. The concentration process continued until the final concentrate was obtained with a completely clear sample without any sign of precipitation. The purified hemeyrthrin protein sample, at this stage contained 20 mM PBS, 0.5M NaCl and 0.14M imidazole. However, for crystallisation 10mM Tris-HCL pH 8.0 and 0.1M NaCl was required. The protein needed NaCl in order to prevent precipitation and for the protein to stay in solution. So, the sample was concentrated to about 0.75ml using a VivaSpin column. Then a 0.5ml Zeba Spin Desalting Column was used to change the buffer, and about 5ml of 10mM Tris-HCL pH 8.0 plus 0.1M NaCl buffer was used. Therefore, the final crystallisation sample contained 7.5 mg/ml of Cj0241-Hr protein in 10mM Tris-HCL buffer pH 8.0 and 0.1M NaCl.

The initial crystallisation screening trial was set up with a Matrix Hydra II Plus One (Figure 2.4) crystallisation robot (Thermo Fisher Scientific, USA). The concept of this machine is based on the principle of a sitting-drop vapour-diffusion approach. Screening the protein crystals was performed in 96 well plates where each condition tested used 200nl of protein plus 200nl of crystallisation reagent equilibrated against 50ul of the same crystallisation reagent. 3 different commercial crystallisation conditions screens were used; JCSG, PACT and PEG screens (Qiagen). The plate was centrifuged for 2 minutes at a speed 2000 rpm to allow the protein and precipitants to react before incubating at a suitable temperature, usually 17˚C.

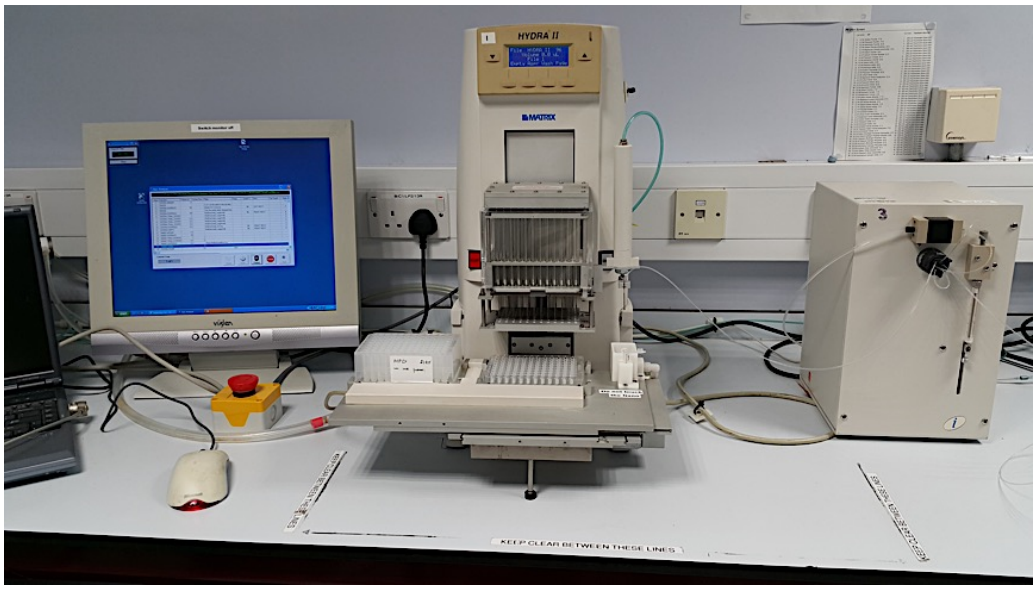


Figure 2.4: The Matrix Hydra II Plus One robot, which was used in the crystallisation experiment.

2.10.3 Optimisation trials

Screen conditions that produced crystals were used to optimise the quality of these crystals for data collection. This optimisation was performed manually in 24-well plates using the hanging drop vapour diffusion technique. Conditions for the optimisation experiment were achieved by using different precipitant concentrations (14 – 24%) and pH (6.5 – 8.0) at 17°C with 7.5 mg/ml protein concentration. Using a ratio of 1:1, a volume of protein and crystal condition were mixed and held on a pre-siliconised cover slip, this was then placed on the reservoir including 300-500 µl of crystal condition and sealed with grease. The plates were incubated at 17˚C, and the growth of crystals checked on a daily basis by inspection under the microscope.

2.10.4 X-ray data collection

X-ray radiation can have harmful effects on the protein crystals due to their sensitive nature. This sensitivity can lead to damage of the protein and hence the quality of the data being collected. Therefore, before the crystals were used for data collection, they were mounted on loops and soaked quickly in a cryo-protectant to stop ice crystal formation. It is the use of low temperature that reduces X-ray radiation damage. Typically, 20 % v/v ethylene glycol was used as cryo-protectant in the crystallisation solution, which contained 0.2M K formate pH7.5 and 20% $^{w}/_{v}$ PEG3350. The crystals were plunged into liquid nitrogen for storage and transported to the synchrotron radiation source at the DLS near Oxford, United Kingdom. The diffraction data were indexed and integrated using XDS and scaling by Xscale from the Xia2 program (Winter, 2010). This gave an mtz file of the indexed reflections with their intensities (I) and their related statistical errors (σI).

2.10.5 Structure Determination for Cj0241

2.10.5.1 Automated Molecular Replacement (MR)

Initially the X-ray diffraction data of the Cj0241 crystal extended to a resolution of 1.7Å. The Cj0241 crystal belongs to space group $P2_1$, with the following unit cell parameters a = 30.9Å, b = 48.4Å, c = 43.7Å, α = 90°, β = 98.7° and γ = 90° **(designated here as Data set 1)**. To get the initial phases for an electron density map, molecular replacement by using the PHASER program from the CCP4 suite was performed with 6 different template models that had been obtained from the Phyre[2] server (**P**rotein **H**omology/analog**Y** **R**ecognition **E**ngine v 2.0). The Phyre[2] is a web - server that facilitates the prediction of the 3D-structure of a protein from its sequence using the techniques and principles of homology modeling through sequence alignment based upon all known structures within the Protein Data Bank (PDB). However, all attempts were unsuccessful. Because it was believed that the Cj0241-Hr structure contained two iron ions, it was hypothesised that phase information could be generated by Single Wavelength Anomalous Diffraction (SAD).

2.10.5.2 Single wavelength Anomalous Diffraction (SAD)

The Cj0241-Hr structure was solved by using the Single-wavelength Anomalous Diffraction (SAD) strategy using data collected on Beamline I04 at DLS. It used the anomalous signal wavelength near the iron edge. This wavelength was chosen from a fluorescence scan across wavelengths around the iron atoms absorption edge. The crystal diffracted to 2.1Å resolution, and the data was indexed using the program Xia2. The asymmetric unit was predicted to contain one copy of Cj0241 with solvent content of approx. 35%. The data was analysed, scaled and merged using Xscale from the Xia2

program (Winter, 2010). The space group $P2_1$ with cell dimensions a = 31.2 Å, b = 48.5 Å, c = 43.6 Å, α = 90°, β = 100.2° and γ = 90° (**designated here as Data set 2).**

The initial set of phases and electron density map were obtained from this data using the AutoSol wizard from the PHENIX suite (Adams *et al.*, 2010). The input included the reflection data file, number of expected Fe anomalous scattering atoms (2) and the protein sequence file. An initial automatic trace of the protein structure was made which was then improved by inspection of the map and model in the COOT program from the CCP4 suite (Emsley and Cowtan, 2004). The model was refined using the phenix. refine option in the PHENIX suite. In the later stages of modeling and refinement, the REFMAC5 program was used with its full coordinate and B-factor refinement plus TLS refinement. The TLS refinement is a mathematical model, which aims to describe the positional displacement of each atom in the molecule by proposing that each atom is part of a rigid body displaced around a mean position (Schomaker and Trueblood, 1968 ; Adams *et al.*, 2010). Water molecules were added in the last refinement cycles.

2.10.6 High resolution structure of Cj0241-Hr

To confirm a likely binding site for an oxygen molecule in the structure, the Cj0241-Hr crystals from a JCSG screen condition (0.2M Ammonium nitrate pH6.3 plus 20% PEG3350) were soaked with sodium azide (NaN_3) before addition of 20% ethylene glycol as cryo-protectant. The NaN_3 concentration was 10mM and the incubation time was 30-60 seconds. They were immediately frozen in liquid nitrogen and then sent to the DLS for data collection. Sodium azide was used in this experiment because azide is known as a stable adduct with the diferric site, and expected to exhibit the same characteristic as an O_2 adduct when attached to one of the iron ions (Isaza *et al.*, 2006). The crystals were noted during this incubation to show a distinct colour change from

yellow to orange. A data set was collected from one crystal which diffracted to 1.1 Å resolution **(designated here as Data set 3)**, and it also belonged to space group $P2_1$, but the cell dimensions were not the same as seen before with Data sets 1 & 2 and it was therefore not isomorphous.

The crystal had unit cell dimensions of a = 28.5 Å, b = 46.2 Å, c = 48.4 Å, α = 90°, β = 93.8° and γ = 90°. The initial set of phases was determined by molecular replacement (MR) with the PHASER program from the CCP4 suite using the model built with Data set 2. This model was then further rebuilt into the map using COOT and refined in REFMAC5 with further rounds of rebuilding and refinement until all the electron density including solvent had been accounted for.

2.10.7 Model Validation

The structure quality and validation were achieved by uploading the coordinate files into the MolProbity software (Chen *et al.*, 2010), which can be downloaded from http://molprobity.biochem.duke.edu/ , a more detailed explanation on the method of validation can be found in section (3.13). Validation was also conducted in COOT before running the final refinement using REFMAC5.

The program PyMol (DeLano, 2002) was used to examine and analyse the model. It was also used to create most of the figures of the protein structure in this thesis. Alternative sources for some figures are being written under each figure.

2.11 Materials and methods: Cj1224-Hr protein

The production and purification of Cj1224 was attempted using the same procedures as in Cj0241. However, a number of additional steps were used to improve the purification of Cj1224-Hr.

2.11.1 Protein purification

Protein purification was done essentially as described in section (2.10.1). In addition, extra ferrous sulphate was added to a final concentration of 0.1 mM into the liquid media when growing the strain for induction of protein expression.

The elution profile and the progress of the purification were analysed using SDS-PAGE. Two 3 ml fractions with the purest protein were combined with a concentration of 1.8mg/ml, and thus the total protein was approx. 11mg.

Also, a number of additional steps were used to improve the purification and promote the crystal of Cj1224-Hr protein to grow:

2.11.1.1 Buffer for purification

Instead of phosphate saline, Tris buffer was also tested during the purification such that buffer A (lysis buffer) contained 50mM Tris-pH 8 plus 0.5 M NaCl, and buffer B (elution buffer) contained buffer A plus 0.5 M Imidazole.

2.11.1.2 Metal Ions

Various metal ions have contributed to the stabilisation and helped solubilise proteins (McPherson, 1990). This may help to encourage the crystallisation of proteins. Crystallisation trials were attempted with the Cj1224-Hr protein using 0.5M and 1M

$ZnCl_2$ after exchanging the protein buffer into 10mM Tris-HCL buffer pH 8 and 0.1M NaCl using Zeba Spin Desalting Columns.

2.11.1.3 Reducing Agents

As previously stated, Cj1224-Hr contains cysteine residues, so to minimise the chance of protein aggregation by potentially unwanted disulphide bond formation that may prevent crystallisation, 100mM Dithiothreitol (DTT) reducing agent was added to the Cj1224-Hr for some sets of trials.

2.11.2 Protein crystallisation

For crystallisation trials, the purified protein Cj01224-Hr was concentrated to a variety of different concentrations: 5 mg/ml, 9 mg/ml, 12 mg/ml, 18 mg/ml and 25mg/ml in 10mM Tris-HCL buffer pH (5,6,7,8) and 0.1M NaCl using a Zeba spin-desalting column. The initial crystallisation screening trial was set up by a Matrix Hydra II Plus One crystallisation robot (Thermo Fisher Scientific, USA) with various commercial screening kits including: PACT, JCSG, PEG, Ammonium sulphate, Proplex, Morpheus and MPD screens (Qiagen). The plates incubated at 17°C and also at 4°C.

2.11.3 Optimisation trials

Optimisation trials were performed manually in 24-well plate using the hanging drop vapour diffusion technique. The experiments were done by changing the concentration of PEG (6000, 8000, 3350) with (12, 14, 16, 18, 20, 22, 24, 26, 28 and 30 % w/v) and varying the pH buffer (5.0, 6.0, 7.0 and 8.0) for small crystals size that was discovered in different condition. A droplet on the cover slip was contained 2µl protein and 2µl of well solution in a hanging drop vapour diffusion method. The cover slip was

then placed on the reservoir including 300-500 µl of crystal condition and sealed with grease.

2.11.4 Data collection

The crystals were mounted in loops straight from their growth conditions, and typically soaked with cryo-protectant solution containing additional 20 % v/v ethylene glycol. However, some crystals were taken directly from the conditions, which contained 2-methyl-2, 4-pentanediol (MPD). The precipitant MPD can itself function as cryo-protectant and thus keeps crystals from cracking due to formation of crystalline ice when doing flash-freezing (Murshudov *et al.*, 1997). The crystals were sent to the DLS near Oxford, United Kingdom, to collect the data as explained for Cj0241, but unfortunately, they were found to be salt.

2.11.5 Estimating Cj1224-Hr protein folding with Nuclear Magnetic Resonance (NMR) spectroscopy

The state of the Cj1224 -Hr protein was evaluated by the Nuclear Magnetic Resonance (NMR) to assess if it is folded. A 1mM solution of Cj1224-Hr purified protein in 20mM sodium phosphate buffer plus 50mM NaCl pH6.5 was sent to the nuclear magnetic resonance (NMR) spectroscopy lab (The University of Sheffield, Firth Court Building, Sheffield, UK). A 1D ^{1}H NMR spectrum was collected.

2.11.6 Cj1224 hemerythrin protein with and without His tags

Because of the previous attempt to get crystals of Cj1224 protein had failed, alternative constructs without any tag or with an N-terminal histidine tag were made to improve the chances of crystallisation.

2.11.6.1 Design of primers

The Cj1224-Hr gene was amplified from the purified plasmid Cj1224 _6xHis, which expressed the C-terminally tagged Cj1224 by using the polymerase chain reaction (PCR). Two types of vector were used in this study: pET21b (+) (Figure 2.5) and pET28b (+) (Figure 2.6). The Cj1224-Hr gene was cloned into pET21b (+) to produce a protein product without a histidine tag (Cj1224_notag). In the second experiment Cj1224-Hr gene was cloned into pET28b (+) to produce a protein product with an N-terminal histidine tag (6xHis_Cj1224).

The same restriction enzymes sites, Nde1 and Xho1 were used with each vector. According to the manufacturer's instructions primers were typically 20–40 nucleotides in length, with 40–60% GC content, and a melting point in the range of 52°C to 60°C. The sequences of the primers used for amplification of the *cj1224* gene are given in Table 2.2.

Table 2.2: Primers used for amplification of *cj1224-Hr* gene

Primers	Sequence
Cj1224-F	CCG CAT ATG CTT CCA AAA TGG GAT AAC
Cj1224-R	GCC CGG CTC GAG TCA AGA ATA TTT TTT GTA

All primers in this study were purchased from Eurofins laboratories. They arrived purified and were re-suspended to 100 µM in dH_2O upon receipt.

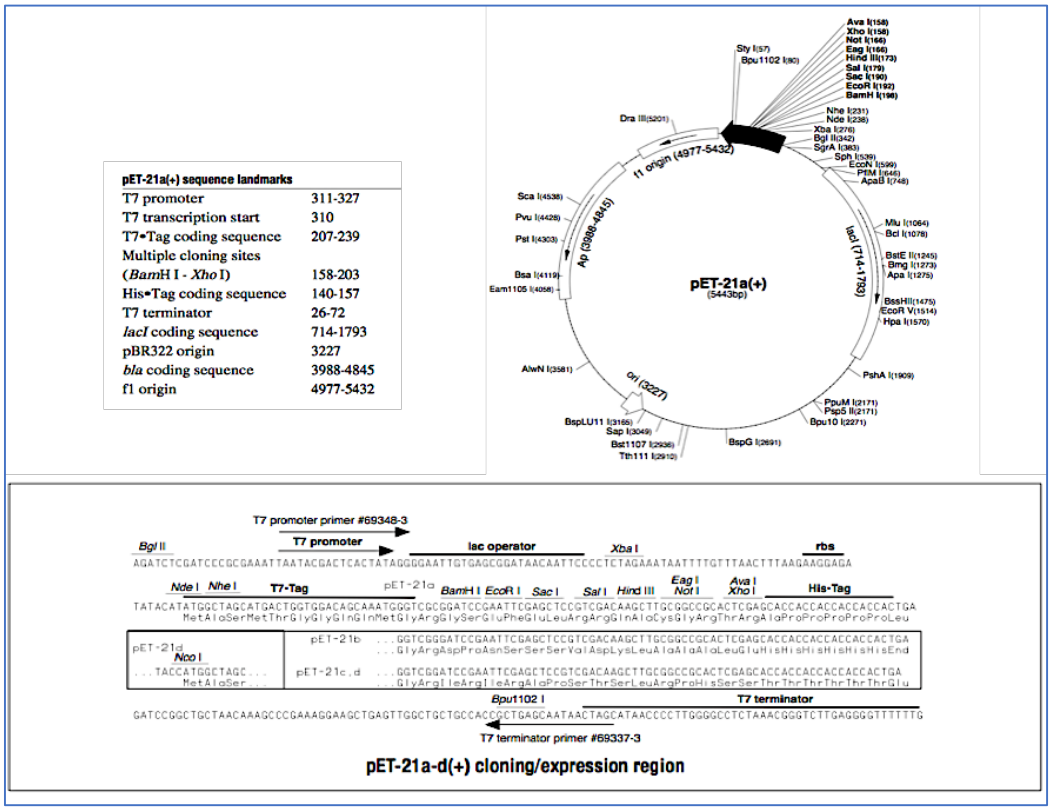


Figure 2.5: pET-21a (+) vector map. pET-21a-d (+) cloning/expression region. The coding region of the Cj1224 gene was cloned between the XhoI1 and NdeI restriction sites of pET-21b (+) vector for overexpression in *E.coli* BL21 (DE3) without a histidine tag. Adapted from: http://www.helmholtz-muenchen.de/fileadmin/PEPF/pET_vectors/pET-21a-d_MAP.pdf

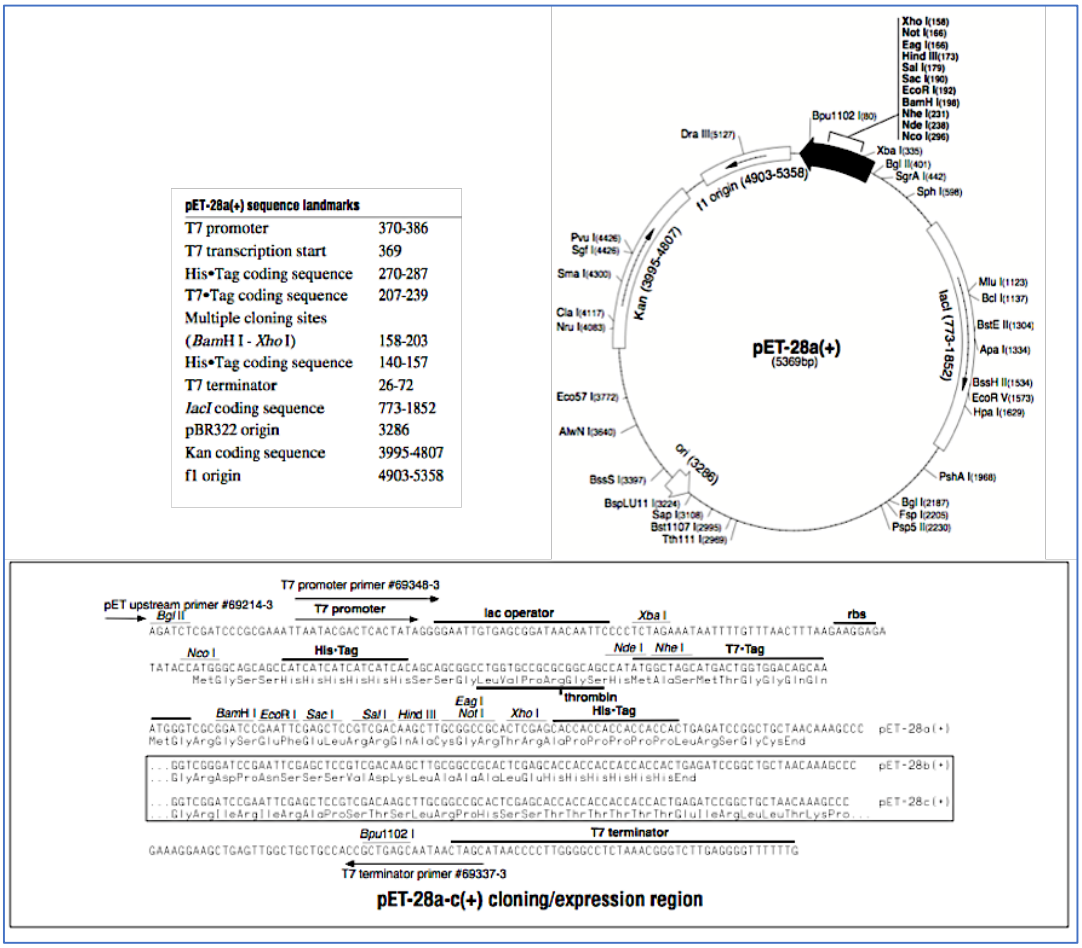


Figure 2. 6: pET-28a (+) vector map. pET-28a-c (+) cloning/expression region. The coding region of the Cj1224 gene was cloned between the XhoI1 and NdeI restriction sites of pET-28b (+) vector for overexpression in *E.coli* BL21 (DE3) with an N-terminal histidine tag. Adapted from: http://www.helmholtz-muenchen.de/fileadmin/PEPF/pET_vectors/pET-28a-c_map.pdf

2.11.6.2 Polymerase Chain Reaction (PCR)

The PCR technique (Bartlett and Stirling, 2003) was used during this study for two different purposes: (1) to amplify the target gene for cloning experiments; and (2) to perform colony PCR to verify the presence of the insert gene in the purified plasmid post-cloning. In both applications, a typical reaction was made up of the following:

Component	50µ Reaction	Final concentration
Q5 High-Fidelity 2X Master Mix	25 µl	1X
10 µM Forward Primer	2.5 µl	0.5 µM
10 µM Reverse Primer	2.5 µl	0.5 µM
Template DNA	2 µl	< 1,000 ng
Nuclease-Free Water	18 µl	

The PCR reaction mixture was set up in a thermal cycling machine; [3]Prime, using the following conditions:

Initial denaturation	98˚C	2:0 min	
Denaturation	98 ˚C	30 sec	
Annealing	60 ˚C	30 sec	
Elongation	72 ˚C	1:30 min	X 25 Cycles
Final Elongation	72 ˚C	7:0 min	
Final hold	4 ˚C		

In order to perform colony PCR, the plasmid DNA was released from colonies by re-suspending them in 30 µl of sterile water and then boiled in a heating block for 10 minutes at 100℃ to lysate the whole cells.

2.11.6.3 Agarose gel electrophoresis

In order to evaluate the DNA fragments agarose gel electrophoresis was applied. The procedure was conducted by loading the sample on a 1% agarose gel, and then

observation of their movement through the gel matrix. The gel matrix was formed by dissolving 0.4 g of agarose into 40 ml of TAE buffer (40 mM Tris pH 8.0, 1 mM EDTA, 20 mM glacial acetic acid). When the agarose completely dissolved and was easy to touch by hand, 5 µl of GelRed stain was added, this led to a fluorescent effect of the solution allowing a visual observation of the DNA migration in UV light. The gel was then placed into a gel rack, left to solidify for around 15-20 minutes, and was then allowed to sit in the TAE buffer. These test samples were mixed with 5 x DNA loading buffer (Bioline, UK). The samples were loaded next to 5 µl of the marker Hyperladder^{TM} I (Bioline, UK) to define the molecular weight of the DNA fragments. The electrophoresis was then run at 80 V for 60 minutes. DNA inserts were visualised using an ultraviolet (UV) light source.

2.11.6.4 Purification of PCR products from the agarose gel

The process used to extract the DNA fragments from the agarose gel used the QIAquick Gel Extraction kit and by following the manufacturer's instructions (see appendix). The resultant DNA purified was then used in the ligation stage.

2.11.6.5 Restriction digest

Before the ligation experiment, both the DNA from the Cj1224-Hr PCR reaction and the vectors (pET21b and pET28b) were cut with restriction enzymes (endonucleases) Xho1 (NEB) and NdeI (NEB) in double digest reactions, to get suitable ends for the cloning experiments (Table 2.3). The DNA fragments and plasmids were incubated in separate tubes with the restriction buffer (Cutsmart) to get optimal reactivity in the restriction digest. The digestion was undertaken for 4 hours at 37˚C in a total volume of 20 µl. The reaction was set up as following:

Table 2.3: Double digests composition in tubes with *cj1224* DNA or the two vectors

	pET21b /pET28b vector	Cj1224-Hr DNA
DNA	10µl	10µl
Nde1	1.0µl	1.0µl
Xhol1	1.0µl	1.0µl
Buffer	2.0µl	2.0µl
Total volume	20µl	20µl

After 2 more hours, an extra 1.0µl of Ndel & Xhol was added into each tube to increase and confirm the cutting process. The digested DNA (PCR products, pET21b and pET28b vectors) were analysed using agarose gel electrophoresis. Following digestion, the enzymes were inactivated by incubation at 65℃ (time dependent on manufacturer's guidelines). The buffer and enzyme were later removed from the sample using the QIAquick® PCR Purification Kit (Qiagen).

2.11.6.6 Ligation experiment

2.11.6.6.1 Phosphatase treatment

Before the ligation experiment, the digested vectors were treated with Shrimp Alkaline Phosphatase (rSAP). The rSAP is useful to remove the 5' phosphate group and reduce the risk of self-ligation of linearised plasmid DNA. Using the manufacturer's instructions, these reactions were completed accordingly. 2µl from Cutsmart buffer were added into the vector tubes (pET21b / pET28b). Then the reaction tube was incubated at 37° C for 30 minutes after addition of 1ul rSAP. The Shrimp Alkaline Phosphatase after that was inactivated for 5 min at 65°C.

2.11.6.6.2 Ligation of PCR product into plasmids

The desired DNA fragments were ligated into the vectors (pET21b and pET28b) by utilising the bacteriophage T_4 ligase (Sambrook and Russell, 2006), this reaction creates phosphodiester bonds between adjacent 5'-phosphate and 3'-OH ends of DNA. It is also able to join both blunt and cohesive ended DNA fragments.

Ligation was achieved in 25µl volume by using a range of estimated molar ratios 1:1, 2:1, 3:1 and 5:1 (DNA insert: plasmid) (Table 2.4). The ligation buffer was mixed to reach a concentration of 1X and then left overnight to incubate at room temperature.

Table 2.4: The ligation reaction composition

Materials	Tube 1 (1:1)	Tube 2 (2:1)	Tube 3 (3:1)	Tube 4 (5:1)
T4 ligation buffer 5X	3µl	3µl	3µl	3µl
Vector DNA	5µl	5µl	5µl	10µl
Insert DNA	5µl	10µl	15µl	5µl
T4 ligase	1µl	1µl	1µl	1µl
H_2O	11µl	6µl	1µl	6µl
Total volume	25µl	25µl	25µl	25µl

2.11.7 Transformation experiment

As stated and undertaken in section (2.4), chapter (2), 50 µl aliquots of competent *E. coli* DH5a cells were removed and used in the transformation process. 1ml of LB media were added to the cells and then incubated at 37 °C at 250 rpm for 40 minutes before plating on LB agar containing appropriate antibiotics; 100 µg /ml ampicillin for Cj1224_notag, and 50 µg /ml kanamycin for 6xHis_Cj1224. The plates were then incubated at 37 °C overnight.

2.11.8 Plasmid purification for screening of possible clones (Plasmid sequencing)

A common process of screening transformed plasmids into *E. coli* is using the QIAprep Spin Miniprep Kit from (QIAGEN, UK), and by applying the standard technique of DNA purification, plasmid was obtained by extracting from the *E. coli.*
By taking one colony of DH5α *E. coli* cells with the plasmid pET21b or pET28b from an LB agar plate and was allowed to inoculate with 5 ml of LB medium supplemented with appropriate antibiotics (100μg/ml ampicillin for pET21b and 50μg/ml Kanamycin for pET28b). The culture was left to incubate overnight at 37° C at 250 rpm rotor shaking speed. The cells were then harvested by centrifugation at 4500 rpm for 20 minutes and the plasmids were purified using a QIAprep Spin MiniPrep Kit according to the manufacturer's instructions (see appendix). Purified plasmids resultants were sent to GATC Biotech Ltd for DNA sequencing using T7F (forward) and T7R (reverse) primers.

2.11.9 Expression and solubility test

2.11.9.1 Small-scale overexpression

The resultant product was transformed into a competent cell of *E. coli* strain BL21 (DE3) (Novagen), and was plated onto appropriate antibiotics. Initially, a small-scale overexpression experiment was undertaken in order to evaluate the expression for the cloned gene and the protein solubility. A colony from the transformed BL21 (DE3) with Cj1224_notag or 6xHis_Cj1224 was taken and used to prepare a 5ml starter culture using LB medium, added to appropriate antibiotics and incubated overnight. One ml from this culture was then used to inoculate two 50ml cultures of LB media having the appropriate antibiotics that would determine the protein solubility. To allow the cells to grow they were left at 37 °C temperature with shaking at 250 rpm until the absorbance at 600 nm

reached about 0.6. Then, one ml of each culture was placed into an Eppendorf tube, and pelleted prior to being stored at -20°C. They were then used on SDS-PAGE as the pre-induction control samples. When the optical density of the cells reached 0.6, IPTG was mixed in, to give a final concentration of 1mM. After 16 hours induction at 25°C or 37°C on a rotary shaker at 220rpm, 1ml from each flask was collected and centrifuged and the pellet then frozen at -20°C for following SDS-PAGE analysis. A solubility test was performed as explained in section (2.5.2 & 2.5.3).

2.11.9.2 Large-scale overexpression

Using two 2L flasks, 500 ml of LB media was added, and at the same time the appropriate antibiotics, and inoculated with the 10ml of starter culture. The flasks were left at 37°C with shaking at 250rpm until the OD_{600nm} reached 0.6, before adding IPTG to a concentration of 1 mM for induction. The cells were harvested by centrifugation in an F500 rotor at 14334g for 20 min at 4°C, and the cells were stored at -20°C before being used in a purification experiment. 1ml pre- and post induction samples were taken during the growth to verify the expression and solubility levels.

2.11.10 Purification of non-tagged Cj1224-Hr protein

Approximately 11 g of cell paste was re-suspended in the lysis buffer (buffer A: 40 ml of 50 mM Tris pH 8.0). Cells were then disrupted by sonication (4x20sec). Cell debris and insoluble protein were removed by centrifugation at 46000g for 25 min and 4°C. The protein content of the Cell Free Extract (CFE) was estimated using a Bio-Rad assay to be approx. 670mg (V=48, C=14mg/ml). The CFE was then applied onto a 20ml DEAE FF column (the pump provides flow rate 5ml/min, fractions 6ml), and protein was eluted by a 200 ml of gradient of NaCl from 0-0.5 M in buffer A. The fractions were analysed by SDS-PAGE. The fraction containing Cj1224-Hr was then concentrated to

total protein 9 mg (3 mg/ml in 3 ml) using a Vivaspin concentrator with a 10 kDa MWCO. This sample was then applied onto a gel filtration column; Hi-Load Superdex 200, 1.6x60 cm column (GE Healthcare life science), equilibrated with buffer A (50 Mm Tris pH8) plus 0.5M NaCl, and the Cj1224-Hr protein was then eluted using the same buffer in 2 ml fractions at flow rate 1.5ml/min. Peak fractions from the chromatogram were analysed by SDS-PAGE. Fractions containing the most pure Cj1224-Hr protein were combined and concentrated using a Vivaspin concentrator. The buffer was then replaced by using a Zeba Desalt Spin Column to 5 mM Tris pH 8.0 and 50 mM NaCl for use in crystallisation trials.

2.11.11 Cj1224-Hr with N-terminal His-tag Purification

In this experiment two types of buffer were used: phosphate saline buffer and Tris buffer. The trials were conducted as previously explained in chapter (2) sections (2.11.1& 2.11.1.1).

2.11.11.1 Crystallisation trial

For crystallisation trials, the purified Cj01224-Hr with N-terminal His-tag and without His-tag were prepared with a variety of different concentrations (7 mg/ml, 12mg/ml, 20mg/ml) with buffer 10mM Tris-HCL pH at a range of pH values (5,6,7,8) and 0.1M NaCl using Zeba spin-desalting column. Different commercial crystallisation conditions screens were used including; JCSG, PACT, PEG, MPD, Ammonium sulphate, Morpheus, Classics and Proplex. Most plates were incubated at 17° C and some at 4° C.

2.11.12 Site directed mutagenesis in Cj1224 hemerythrin

The ICP-MS analysis showed that the protein bound zinc ions in addition to binding iron ions. The C-terminus in Cj1224-Hr has four cysteine residues. Clusters of

cysteines in proteins can display a high affinity toward metal ions including zinc ions. Therefore, to test the effect of removing the cysteine residues, they were changed to alanine by employing a site-directed mutagenesis kit from NEB with reference to the manufacturer's procedure as shown in the appendix. Purified plasmids resultants were then sent to GATC Biotech for sequencing using T7F and T7R primers to verify if the mutations had succeeded.

2.11.13 Analysis of Free Thiol and Disulfide Bond

DTNB (5,5'-dithiobis- [2-nitrobenzoic acid]) or Ellman's reagent is a chemical, which will evaluate quantitatively the concentration or number of Sulfhydryl (-SH) or thiol groups in a protein sample. The experiment was undertaken to assess the concentration of reduced cysteine residues in Cj1224-Hr protein and the status of this amino acid. A blank was used to set the background absorbance (Abs) value using 0.1 ml of DTNB reagent in 0.9 ml of buffer A (0.1M Na phosphate, 1mM EDTA pH 7.3) or buffer B (6M GuHCl, 0.1M Na phosphate, 1mM EDTA pH 7.3) and measured at 412 nm. After that, 20µl of 6.4 mg/ml concentration (0.256 mM) protein solution purified in phosphate buffer was added to the cuvette, mixed and then incubated for 2 - 5 minutes, and the Abs measurement taken at 412 nm. The concentration of TNB in µM was calculated by the Abs reading at 412 nm / 0.0136, where 0.0136 = Micromolar extinction coefficient for TNB at 412 nm. The concentration of TNB in µM is equal to the amount of reduced SH group in the protein.

2.11.14 Microseed Matrix Screening Experiment

This technique was used to enhance and promote protein crystal growth. The benefit of this approach is that it is most appropriate to proteins where little or no nucleation is usually detected. Previous crystals obtained from the related Cj0241-Hr

protein were used as seeds to attempt to grow Cj1224-Hr crystals. The trial was performed by the method of D'Arcy *et al.*, (2007). Crystals of Cj0241 grown from 0.2M potassium formate pH7.3 20% (w/v) PEG 3350 were collected and crushed with a glass probe. 6ul of the reservoir solution used to grow the crystals was added and the crystal suspension was repeatedly drawn up and down in a pipette tip before the mixture was transferred to a Seed Bead tube (Douglas Instruments). The Seed Bead tube was vortexed for 2 minutes with incubation on ice every 30s. The JCSG screen was used to find potential crystallisation conditions by combining 0.5 µl of Cj1224-Hr protein plus 0.35 µl of reservoir and 0.15µl of seed stock. The plate was then incubated at 17°C.

2.11.15 Investigation of the DNA binding properties of Cj1224-Hr

To study the ability of Cj1224-Hr protein to bind with DNA, single and double DNA stranded samples (Single stranded DNA were PCR primers and double stranded DNA was a PCR DNA fragment from *Campylobacter Jejuni*) were examined using an electrophoretic mobility-shift assay (EMSA) on a 1% agarose gel (Table 2.5). The method depends on the principle that the protein-DNA complex migrates more slowly than the free DNA molecule when exposed to polyacrylamide or agarose gel electrophoresis.

A 0.1 mM solution of N-terminal His-tagged Cj1224-Hr and nucleic acid were mixed together. The experiment was done with different DNA: protein ratios and a 1:40 ratio was found to give the most clear result of shifting. The mixtures were analysed by electrophoresis through a 1% agarose gel and examination by UV.

Table 2.5: Single and double DNA stranded samples used in EMSAs with Cj1224-Hr protein

DNA	Oligonucleotide sequence	Length
Single stranded	tgctcctggtaaaattcacgatg	23 nucleotides
Single stranded	cgtcatatgatgaaagttaaatggagtagagattttagcattcg	44 nucleotides
Double stranded	atgctgccgaaatgggataacagctatagcgtgcataacgcgaaaattgatgaacagcataaa aaactgtttaaactggcggcgaaagtggaagtggtgagcgatcgcagcgtgagcaaaaacga agtgaaagaactgctggcggaatttttttaactatatgaaagatcattttaacgatgaagaaaaat atatgcagctgattggctatccgaacctggaagaacatcgcaaaattcataaagaaattattca gaccatgattaacctgattaaagatattaaaagcaccaacgatctgaaagaaaaactgtatatt gtggcgaaaaaatggctgctggaacatattctgtatgaagatatgaaagtggaaaaatggcgc agcagcagcctgagcaccgatgatggcggcgatgtgagctttgaagcggcggaagatgaagat aacgaacatccgcagttttatctgtatacctgcaactgcccgggcaaaattcatgatgtgccgta tagcattcatcagaaaattgaactgcagggccgcaaatttaccgcgaaaacctgcaaacagtgc attaaattttataaaaaatatagctaa	600 base pairs

2.12 Materials and methods: Cj0045-Hr protein

Previous attempts to produce soluble Cj0045-Hr were unsuccessful. A further attempt was made in this study by growing cells *E. coli* BL21 (DE3) competent transformed with the Cj0045_6xHis expression plasmid, at 15 $^{\circ}$C on a rotary shaker at 200 rpm before induction by 1mM IPTG when the culture reached an OD_{600} of 0.6 and then incubation for 16 hrs. Additionally, extra iron ions were provided by addition of ferrous sulphate to a final concentration of 0.1 mM, into the LB liquid media when growing the strain for induction of protein expression.

2.12.1 Purification of Cj0045-Hr protein

The purification was attempted by nickel-affinity chromatography using the same lysis and elution buffers and protocol as used with Cj0241 in chapter (2) section (2.10.1). The sample which was obtained from Cj0045 failed to bind to the Ni column. A fraction of the filtrate was collected and dialysed overnight against 1L 50mM tris pH 8.0 to

remove all the salts. The protein was estimated using a Bio-Rad assay to be approx. 140 mg (V=18ml, C=7.8 mg/ml). The sample was then applied onto a 10 ml DEAE FF column on AKTA pure (Flow rate 5ml/min, fractions 2.5 ml), and protein was eluted by a 100 ml of gradient of increasing NaCl from 0-0.5M in 50Mm Tris pH 8.0 (buffer A). The fractions were analysed by SDS-PAGE, and fractions 13-15 which containing the highest level of the Cj0045 protein were combined and concentrated using a Vivaspin concentrator to total protein 18mg. This sample was then applied onto a gel filtration column 1.6x60cm HiLoad Superdex200 equilibrated with 0.5M NaCl plus 50mM tris pH 8.0 (buffer A), flow rate 1.5ml/min. The Cj0045-Hr protein was eluted using the same buffer in 2 ml fractions. Peak fractions from the chromatogram were analysed by SDS-PAGE. Fractions containing the most pure Cj0045-Hr protein were combined and concentrated using a Vivaspin concentrator. The buffer was then replaced by using a Zeba Desalt Spin Column to 10mM Tris pH 8.0 and 0.1M NaCl for use in crystallisation trials.

2.12.2 Tandem mass spectrometry (MS/MS)

An experiment was conducted to verify the identification of the soluble protein that had been obtained from the purification experiment with Cj0045-Hr. A 10ul sample of Cj0045-Hr protein at concentration 4mg/ml was subjected to chemical modification using 8 μl of 6M urea solution, 2 μl of 0.5M iodoacetamide and 26 μl of 0.1M ammonium bicarbonate. The reaction was left for 20 minutes, and then 4μl of 1M trypsin was added to cut the protein into fragments. The sample was kept on the shaker at 37°C for around 3hrs. The peptide fragments were detected by mass spectrometry, analysis of the sample was assessed by comparing and contrasting the values obtained against the theoretical peptide masses of proteins by a data search using MaxQuant (Cox and Mann, 2008).

The experiment was conducted by Dr. Mark Collins from the Department of Biomedical Science, The University of Sheffield.

2.12.3 Crystallisation trials of Cj0045-Hr

After confirming the protein, the initial crystallisation screening trial was set up by a Matrix Hydra II Plus One crystallisation robot (Thermo Fisher Scientific, USA) at a protein concentration of 11mg/ml with various commercial screening kits including: PACT, JCSG, PEGS, Ammonium sulphate, Proplex, Morpheus and MPD screens. Further initial crystal trials using the JCSG, PACT, MPD, PEG and Classics screens with 11 mg/ml protein were performed which involved additional iron chloride ($FeCl_3$) at a concentration of 25µM. All the plates were incubated at 17°C.

2.12.4 Cj0045 hemerythrin protein with N-terminal His tag

Because of the difficulties in purifying C-terminally His-tagged Cj0045, a new construct was made for the Cj0045-Hr with an N-terminal histidine tag. New primers for Cj0045-Hr were designed according to the manufacturer's recommendations (New England BioLabs) as explained previously in section (2.11.6.1), and a new clone was created. The gene was cloned in a pET28b vector, and the restriction enzymes Nde1 and Xho1 were used. The sequences of the primers that were used for amplification of the cj0045-Hr gene is given in Table 2.6.

Table 2.6: Primers designed for Cj0045-Hr N-terminal

Primers	Sequence
Cj0045-F	CGT CAT ATG AAA GTT AAA TGG AGT AGA GAT TTT AGC ATT CG
CJ0045-R	TTA CTC GAG CTA TCC CCC CCC CCT

The PCR procedure (Bartlett and Stirling, 2003) was used during this study and a typical reaction used the same composition as previously described in section (2.11.6.2). The PCR reaction was carried out in a ^{3}Prime thermal cycling machine, using the same conditions explained in section (2.11.6.2).

The remainder of the steps including the analysis of the DNA for the new clone, purification of the DNA fragments from the agarose gel, the ligation of the DNA into the linearised pET28b vector and the transformation experiments were all undertaken as explained for the production of the N-terminal His-tagged Cj1224-Hr in section (2.11.6).

2.12.4.1 Expression and solubility test

The resultant product was transformed into a competent cell of *E. coli* strain BL21 (DE3) (Novagen), and was plated onto appropriate antibiotics. Initially, a small-scale overexpression experiment was undertaken in order to evaluate the expression for the cloned gene and the protein solubility. A colony from the transformed BL21 (DE3) with 6xHis_Cj0045 was taken and used to prepare a 5ml starter culture using LB medium, with kanamycin 50ug/ml antibiotics and incubated overnight. One ml from this culture was then used to inoculate 50ml of LB media in each of 3 conical flasks also containing kanamycin 50ug/ml antibiotics. To allow the cells to grow they were left at 37°C temperature with shaking at 250 rpm until the absorbance at an OD_{600} nm reached about 0.6. Then, one ml of each culture was placed into an Eppendorf tube, and pelleted prior to being stored at -20° C. They were then used on SDS-PAGE as the pre-induction control samples. Also, IPTG was added to a final concentration of 1mM in each flask to induce expression of the desired gene and production of the protein. After 16 hours induction at 15°C, 25°C or 37°C on a rotary shaker at 150rpm or 250rpm, 1ml from each flask was

collected, centrifuged and the cell pellet then frozen at -20°C for SDS-PAGE analysis. A solubility test was conducted as described previously in section (2.5.2 & 2.5.3).

2.12.4.2 Expression of Cj0045-Hr in M9 minimal medium

In attempt to improve the level of soluble expression, M9 medium was employed. The starter culture was grown at 37°C in LB medium with 50μg/ml kanamycin with shaking at 250 rpm overnight. Cells were harvested by centrifugation at 15000g for 20 min and re-suspended in 5ml of M9 medium. The cells were then divided into two culture flasks containing 50ml of M9 minimal medium with kanamycin (50 μg/ml). They were allowed to incubate for growth at 37° C with shaking at 220 rpm until the OD_{600nm} reached 0.6 prior to adding 1mM IPTG for induction. One flask of culture was incubated at 25°C and the second at 37°C at 220 rpm overnight and the other two flasks were one in 15°C with 150 rpm and the 2nd in 15°C with 200 rpm. A 1ml sample from each flask was collected, centrifuged and the pellet then frozen at -20°C for SDS-PAGE analysis. The cells were defrosted and prepared for analysis of solubility of expression as described previously in section (2.5.2 & 2.5.3).

2.13 Preparation of deoxy-, oxy- and met-forms of the recombinant Cj0241 and Cj1224 hemerythrins

The main aim and objective of this experiment was to determine and evaluate the autoxidation reaction half-life ($t_{1/2}$) of *Campylobacter jejuni* hemerythrin proteins. Firstly, the hemerythrin protein was purified and the buffer was then changed to 50mM Tris-HCl pH (6, 7 or 8). The deoxy-form of the recombinant Cj0241-Hr and Cj1224-Hr was obtained by sodium dithionite ($Na_2S_2O_4$) treatment. The trial procedures were done in an anaerobic environmental chamber (Modular Atmosphere Controlled System, DW

Scientific) at room temperature. The purified recombinant hemerythrins were dialysed against five, ten and twenty equivalents of sodium dithionite with protein concentration at 5 mg/ml. The dialysis was repeated three times each time for 6 hrs. to obtain the deoxy-Hr form. After that, the deoxy-Hr form was dialysed verses degassed dithionite-free Tris-HCl buffer, pH 8.0, for 15-18 hrs to transfer the excess sodium dithionite. Dithionite-free buffer was changed each 5-6 hrs. To get the oxy and met-stages, the deoxy-Hr was openly exposed to air. The deoxyHr and other forms of hemerythrins were confirmed by using a UV–Visible Spectrophotometer, Cary 50 Conc., VARIAN, at wavelengths 250 nm – 600 nm in quartz cells (10 mm).

2.14 The enzyme-based system for depleting the oxygen levels from Cj0241-Hr and Cj1224-Hr

An enzyme-based system to remove oxygen was tested, which involved glucose oxidase (GOx) and catalase from *Aspergillus niger*. An oxygen electrode (Rank Bros) was firstly calibrated in 2ml 50mM Tris pH8 with stirring until the O_2 level steadied and the concentration was recorded as being 100%. Then a grain of sodium dithionite was added to the buffer to reduce the level of oxygen. After the O_2 electrode calibration, a test reaction was set up for Cj0241 and Cj1224 hemerythrin proteins, with & without the oxygen depletion system (glucose/ glucose oxidase/ catalase). A sample at 1 mg/ml of protein was added to 2 ml of Tris buffer, and the oxygen level was recorded by the electrode, then 16 mM glucose, 4 U/mL of glucose oxidase and 20 U/mL of catalase were added. The oxygen level declined to zero, the mixture in the O_2 electrode chamber was then removed to a cuvette, which was degassed with nitrogen gas for 5min, to check the spectrum. The UV spectroscopy analysis was carried out at room temperature and the

spectra collected using a Cary UV-vis spectrophotometer at wavelengths 250 nm – 600 nm.

2.15 Investigation of possible water tunnels in Cj0241-hr structure

CAVER is a software tool generally used for identification and visualisation of channels or tunnels in protein structures (Chovancova *et al.*, 2012). The access channels to the di-iron centre in the Cj0241-Hr protein were investigated using the CAVER program 3.0. The software was used as a plugin in the PyMol graphics program. A default value was taken for the probe radius of 0.9. The program was also used with models of Cj1224-Hr and Cj0045-Hr derived from the Cj0241-Hr structure using the Swiss-modeller program suite.

Chapter 3

Theory of X-Ray Crystallography

X-ray crystallography is a method used to define the three-dimensional structure of biological material in a crystalline state. There are thousands X-ray crystallographic structures placed in the Protein Data Bank (Rhodes, 2006). This chapter will provide a short discussion of the methods and theory of X-ray crystallography, because it is one of the main techniques utilised in this research study.

3.1 The Crystal Features

Many macromolecules including proteins are able to create crystals under certain conditions. In the case of protein crystallisation, each protein molecule can adopt an identical orientation and produce crystals in an orderly three-dimensional array of molecules. The molecules in the protein crystals are held together by non-covalent interactions and form a repeating lattice (Rhodes, 2006). A crystalline lattice involves the same repeating unit, which contains multiple copies of the protein and is known as a unit cell (Figure 3.1A). The unit cell can be described by three dimension lengths a, b and c and three angles α, β and γ between them (Figure 3.1B). The arranging of the unit cells in a crystal or any internal symmetry of the unit cell can be explained as the crystal's space-group. The space-group is expressed by its lattice type, offering the arrangement of lattice points in three dimensions, and the crystal system that determines symmetry operators inside the unit cell (Rhodes, 2006).

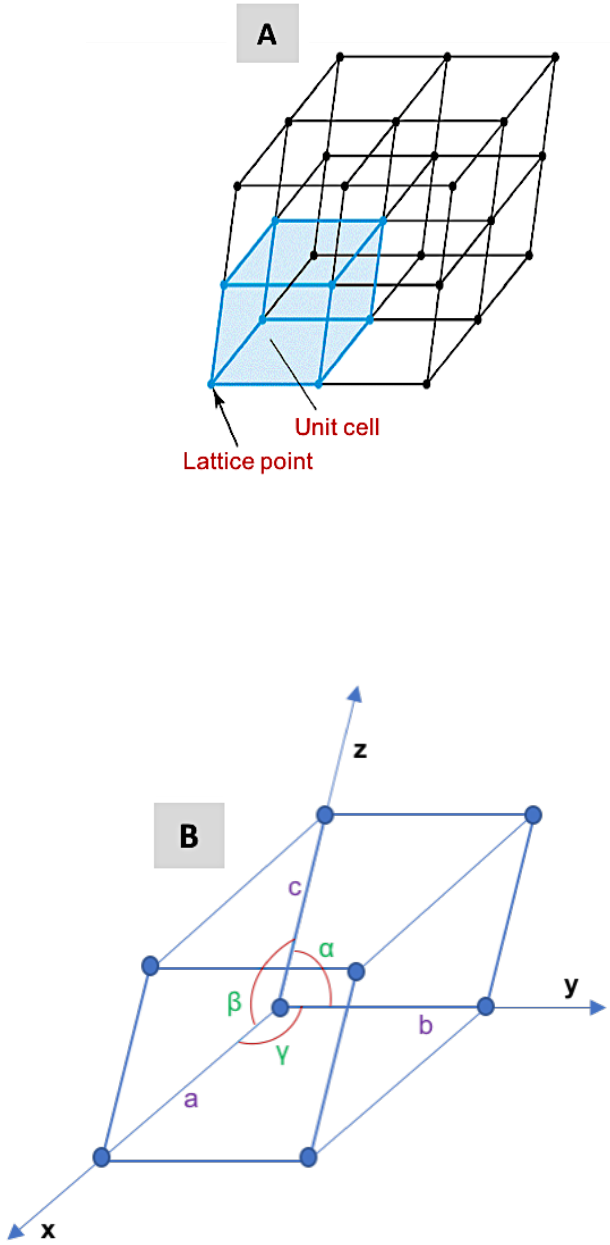


Figure 3.1: **The Unit Cells. Panel (A)** Eight unit cells in a crystalline lattice. **Panel (B)** One unit cell from a crystalline lattice. The location of an atom in the unit cell can be identified by its coordinates x, y and z. Adapted from: Rhodes, 2006 .

There are four potential rotational operation symmetries permitted in three dimensions, creating folding symmetry; 2-fold (180^o), 3-fold (120^o), 4-fold (90^o) and 6-fold (60^o) rotation. In addition, there are eleven potential screw axes, where rotational symmetry is joined with a translation along the rotation axis. These symmetry elements are described by numbers next to a letter defining the lattice type, and provide crystallographic point groups. The four rotational symmetries are labeled 2, 3, 4 and 6 – fold symmetry with the potential screw axes of 2_1, 3_1, 3_2, 4_1, 4_2, 4_3, 6_1, 6_2, 6_3, 6_4, and 6_5 (Blow, 2002)

3.2 Preparation of the Protein Sample for Crystallisation

Correct sample preparation for the crystallisation process is essential in order to obtain the protein crystals. The homogeneity and purity of the protein sample is critical when obtaining crystals that diffract to high resolution (Dessau and Modis, 2011). In crystallisation, a step is required to bring the macromolecule (e.g. nucleic acid, protein, protein-nucleic acid complex or protein-protein complex) to supersaturation state. The sample must hence be concentrated to the highest concentration, usually 2-50 mg/ml, without producing any precipitation or aggregation of the macromolecule.

In most protein crystal growing methods, the target protein is dissolved in an aqueous solution having a precipitant such as polyethylene glycol or ammonium sulfate. This precipitant is concentrated until it is just below that required to precipitate the protein. Subsequently, water is removed slowly by evaporation process to increase together precipitant and protein concentrations, causing precipitation or crystallisation, as the protein becomes insoluble in the crystallisation conditions. Gradual precipitation is more favoured to create larger crystals, however fast precipitation may make numerous small crystals, or an amorphous solid (Rhodes, 2006). In theory, there are commonly three

stages to grow protein crystals: nucleation, post-nucleation or growth and cessation of growth. Once the nucleus has been created, protein particles move from the main solution and start to form a layered crystal structure (Figure 3.2), which can grow after that to macroscopic dimensions (Rhodes, 2006).

Depending on whether the protein is soluble or a membrane protein, several factors can effect protein crystal growth including: temperature, pH, detergents, ions, metal, purity, protein concentration, homogeneity and the existence of many additives and their concentrations. Optimising each of these factors can take a long time; therefore, commercial kits are commonly used to decide initial nucleation and growth conditions (Rhodes, 2006).

3.3 Crystallisation Techniques

There are different techniques that are used to crystallise the target proteins, all of which aim to induce supersaturated solutions of protein macromolecules. The following section describes vapour diffusion as one of these methods.

3.3.1 The Vapour Diffusion Method

Vapour diffusion is carried out by mixing the protein and crystallisation reagent in a typically one-to-one ratio then, placing the mixture in a sealed environment close to a reservoir of precipitants. Because the concentration of the crystallisation mixture is lower than that of the reservoir, its vapour pressure is different and the water or solvent normally diffuses out from the sample into the reservoir. This leads to an increase in the concentration of protein and the precipitants, which causes the protein to come out of the solution if it passes its solubility limit in the conditions.

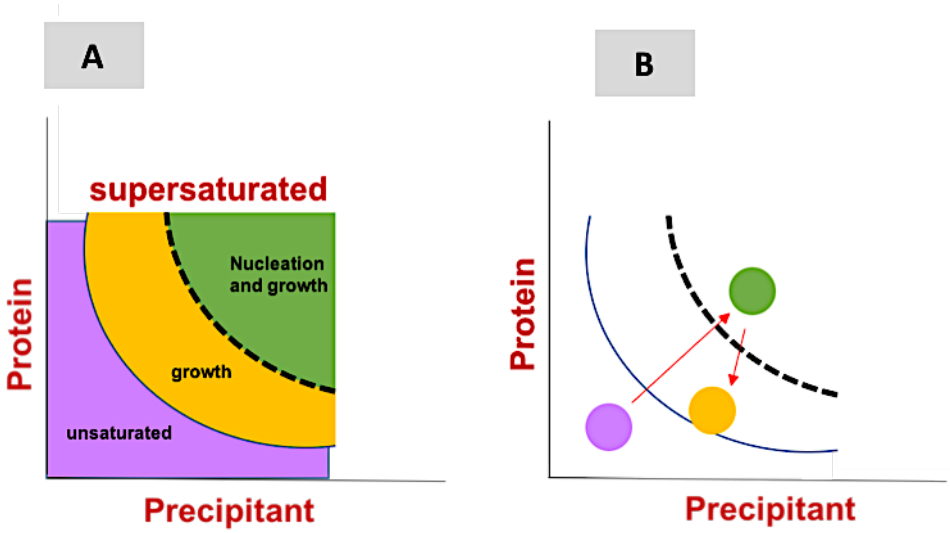


Figure 3.2: Schematic illustration of the protein crystallisation phase diagram. Panel (A): The scheme shows the protein crystallisation phases. The purple zone illustrates the unsaturated period when the solution is not saturated with protein. The green zone illustrates the situations that permit both nucleation and growth. The yellow zone illustrates the situations that help growth only. **Panel (B):** The method to get large crystals; the crystallisation procedure moves from the unsaturated stage to the nucleation and growth stage, then changes quickly to the growth stage until crystal growth quits. Adapted from : Rhodes, 2006 .

This method is used because it permits a slow change to take place in the concentrations of both the protein and the precipitant; this helps in the growth of large and well-formed crystals. The vapour diffusion method can be performed as either a hanging drop or a sitting drop process.

3.3.1.1 Hanging Drop Method

The hanging drop method is performed by placing often 1-2 µl of the purified protein sample on a clean, pre-siliconised cover slip and mixing it in a 1:1 ratio with a crystallisation solution. This is inverted and placed above a small well containing 300-500 µl from the same crystallisation solution (Figure 3.3). After incubation, at room or another chosen stable temperature, this process can produce crystals of a significant size. The coverslip is sealed on top of the reservoir using immersion oil to confirm a secure seal so the equilibration between the hanging drop and reservoir can occur.

3.3.1.2 Sitting Drop Method

Clearly trying to load every crystallisation buffer and mixing it with the protein sample in a ratio of 1:1 for hundreds of crystallisation conditions and placing them one by one on coverslips is time consuming. In the sitting drop technique (Figure 3.4), sample loading can be automated using crystallisation robots. The principles of this technique are like those of hanging drop; the target protein and crystallisation buffer drop is normally placed (often in a 1:1 ratio is applied) on a crystallisation tray next to the reservoir of crystallisation buffer. The container then is airtight sealed with sealing film.

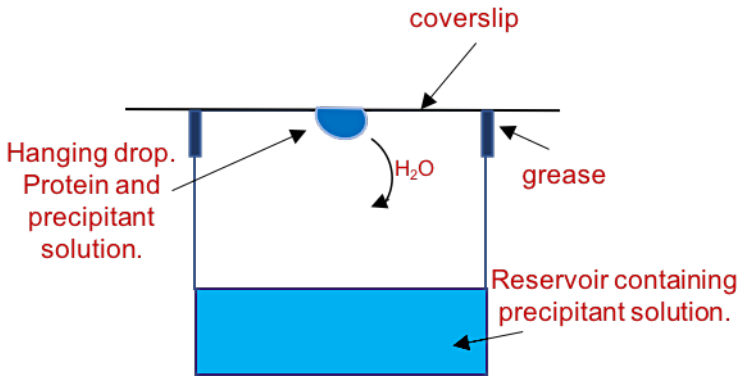


Figure 3.3: The hanging drop method

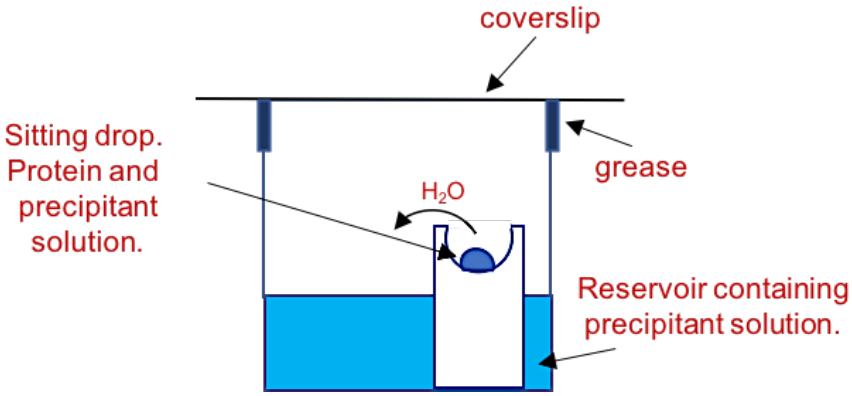


Figure 3.4: The sitting drop vapour diffusion

3.4 Crystal Mounting for Data Collection

During the data collection from single crystals, a crystal is held in the X-ray beam and oriented and rotated in different directions for collection periods ranging from seconds to minutes. Thus, the crystal mount should hold the crystal steadily and not make additional diffraction. It should also be stable during data collection and not damage the crystal. It is practically impossible to ensure all of these conditions at once, but efforts are made to meet them as much as possible (Hasegawa, 2012). In order to reduce the radiation damage to the crystals, a cryo-crystallography technique is used and this also keeps the crystal hydrated. This procedure is usually applied these days; it includes introducing the crystal into liquid nitrogen or nitrogen gas at 100 K using a special loop (Figure 3.5). The main disadvantage of this method is that the liquid nitrogen can give rise to the formation of ice crystals, which can disrupt the protein crystals and cause poor diffraction data. To reduce this situation, a cryo-protectant is introduced into the mother liquor to avoid ice crystal formation. Usually cryo-protectants such as glycerol, ethylene glycol and 2-methyl-2, 4-pentanediol (MPD) are used. It is essential to optimise the freezing period and the type/concentration of the cryo-protectants as both can degrade the protein crystal (Pflugrath, 2004)

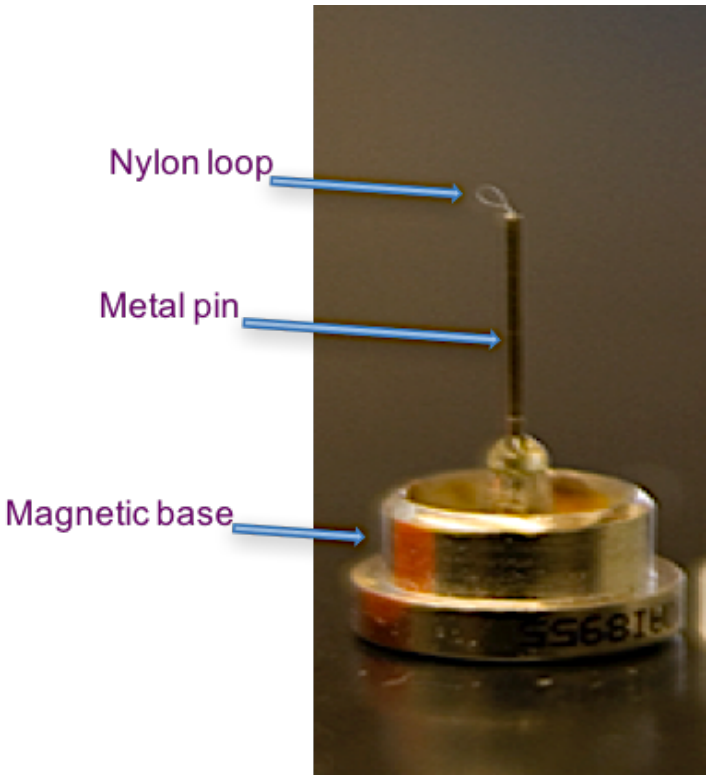


Figure 3.5: A loop used to hold the crystal. The loop is attached to a magnetic base and metal pin. The loops are then fixed on a goniometer head on the X-ray data collection equipment (Pflugrath, 2004).

3.5 X-ray Sources

X-ray crystallography can be described generally as a tool used to identify atomic and macromolecular structures that are in the crystalline state. The crystalline atoms can cause X-rays to diffract into numerous precise directions. Measuring the intensities and angles of the diffracted beams can create a three-dimensional image of the density of electrons within the crystal. However, from this electron density much information can be seen such as the locations of the atoms within the crystal and their disorder as well as chemical bonds and angles.

X-rays are the form of electromagnetic radiation which is the most suitable to explain the atomic structure of protein because they have a wavelength typically between 0.1 – 100 Å which corresponds with the bond distances in the crystals (Rhodes, 2006). There are common sorts of systems that are used for the generation of X-rays such as X-ray tubes, rotating anode sources and synchrotron radiation sources. However, there are some advantages in the use of synchrotron sources which are detailed further in section (3.5.1).

3.5.1 Synchrotron Radiation Source

A synchrotron light source is one the most common powerful tool for provision of X-rays for use in protein crystallography (Figure 3.6). It is made of a particle accelerator that injects electrically charged particles, often electrons, into a circular particle storage ring, which consists of curved and straight parts much like a polygon. In this storage ring, the electrons travel at nearly the speed of light and they are confined into a circular motion using powerful magnetic fields. The electrons are forced to follow a curved path which alters the acceleration of the electrons, and which causes the release of the X-ray radiation at tangents to the curved beamlines. The intensity of the radiation is improved by two systems of accessory devices: wigglers and undulators. Both of these

devices increase the intensity of the emitted radiation by altering the flight path causing local acceleration of the electrons many times during a short distance (Rhodes, 2006). The beamlines are located tangentially around the storage ring providing X-ray beams in which crystals are mounted for data collection. The X-rays are regularly selected by the use of a monochromator (typically a silicon crystal); therefore, only monochromatic X-rays of a particular wavelength are permitted to get to the target protein sample on the beamline. The advantage of using the synchrotron radiation is the intensity of these radiation permitting X-ray data to be collected over much shorter exposure times before the protein crystals are completely damaged. Also, it allows the collection of the data from very small size crystals, and crystals that display weak diffraction patterns when using rotation anode tubes method. In addition, it has the advantage of allowing the user to select the suitable wavelengths, such as for anomalous scattering experiments that are critical in experiments such as Multi-wavelength Anomalous Dispersion (MAD) or Single -wavelength Anomalous Dispersion (SAD) (Rupp, 2010).

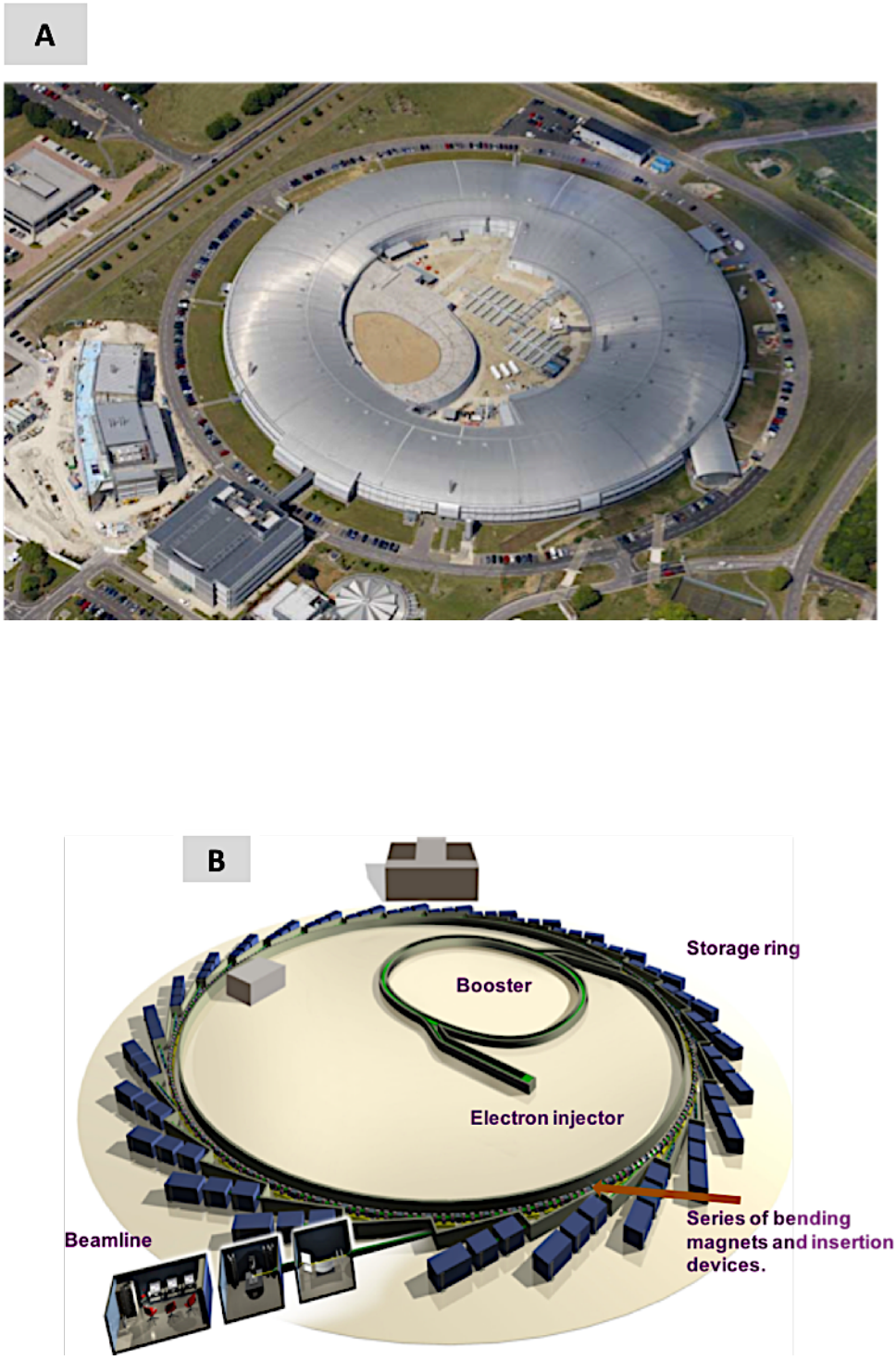


Figure 3.6: The UK's national synchrotron science facility near Oxford. Panel (A) An aerial photograph of the Diamond Light Source (DLS)**. Panel (B)** Diagram of the Diamond Light Source. Taken from the DLS website.

3.6 X-Ray Diffraction

The X-ray beams are electromagnetic radiation that can make a disturbance in the electric field and magnetic field (Figure 3.7). X-ray diffraction happens when the X-ray beam is scattered with no variation in wavelength. Normally the scattered waves are 180^{o} out of phase to the incident wave. When the incident X-ray photon hits a protein crystal, the electrons around the atoms will oscillate with similar frequency to the incident X-ray wave and the emitted X-ray wave can be seen on the detector as diffraction. The diffracted X-rays are characteristic of the crystal and depend on the presence of a certain set of lattice planes (Rupp, 2010). In addition, the emitted waves can undergo two types of interference depending on how the waves overlap with each other: when the troughs or crests of two waves align with each other, their amplitudes add together. This state is recognised as constructive interference (in phase) (Figure 3.8A). However, when the trough of one wave meets the crest of another wave, the subsequent waves will have no amplitude, and therefore be completely cancelled out. This state is known as destructive interference (out of phase) (Figure 3.8B)

3.6.1 Braggs Law

In 1915 William and Lawrence Bragg shared the Physics Nobel Prize for their discovery of the laws, which governed X-ray diffraction from NaCl crystals (Bragg, 1913). The crystal unit cell contains numerous planes, called lattice planes that run in various directions through the lattice points of atoms. An index is given to each set of planes (Miller index). The distance between the planes controls the diffraction angle of the X-rays in line with Bragg's law. Bragg's law explains that when a beam of X-ray wavelength (λ) strikes a set of the lattice planes at an angle (θ), all the atoms in the planes will scatter the X-rays and form or emit reflection waves at the same angle (Blow, 2002).

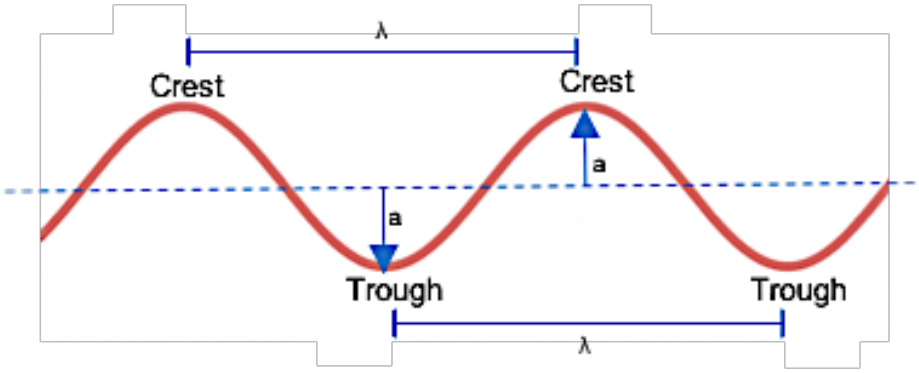


Figure 3.7: The wave displays the wavelength and amplitude. The wavelength λ is the distance from one point to the next equivalent point along on the wave. The amplitude, a, represents the height of the wave crest or trough above its mean level. The picture was adapted from:http://www.bbc.co.uk/bitesize/standard/physics/telecommunications/communication_using_waves/revision/5/

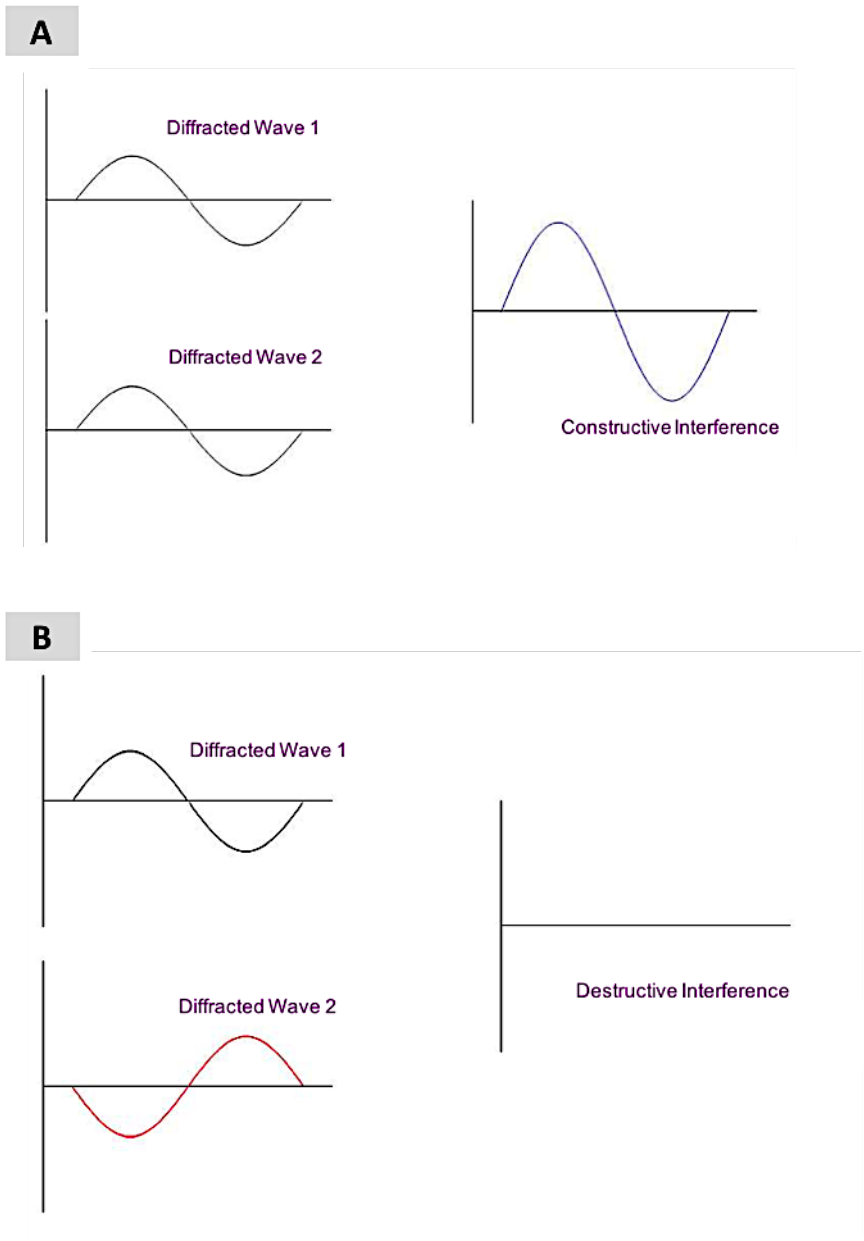


Figure 3.8: Wave Interference. (A) Constructive interference & **(B)** Destructive interference. The picture was adapted from: http://www.asdlib.org/onlineArticles/ecourseware/Bullen_XRD/

In order to understand how Bragg's Law works, consider the two incident X-ray waves as shown in Figure (3.9). The first X-ray beam strikes (atom Z) upon the upper lattice plane with an incident angle of (θ), however, the second X-ray wavelength beam crosses and continues throughout the upper lattice plane and strikes (atom B) the second lattice plane. The second X-ray beam has to travel a further distance AB + BC, so for the two X-ray beams to be in phase, the extra distance between the two planes AB + BC is an integral (n) multiple of the wavelength (λ). This can be expressed as:

$$AB + BC = n\lambda$$

Suppose (d) as the hypotenuse of right angle triangle formed by ABZ (Z is the point in the upper layer from where the beam diffracts), the distance AB will be:

$$AB = d \sin\theta$$

and because AB = BC, and AB+BC= 2AB, then $n\lambda = 2AB$. Combining the two terms, $AB = d \sin\theta$ and $n\lambda = 2AB$, will yield Bragg's law:

$$n\lambda = 2d\sin\theta$$

Hence, the X-ray diffraction is detected when the distance of the crystal layers (d), the incident angle (θ) and the incident X-ray wavelength (λ) are satisfied as stated in Bragg's Law. If $n\lambda = 2d\sin\theta$ according to Bragg's Law is not satisfied, then a blank page will appear showing a reflective image that is a blank space.

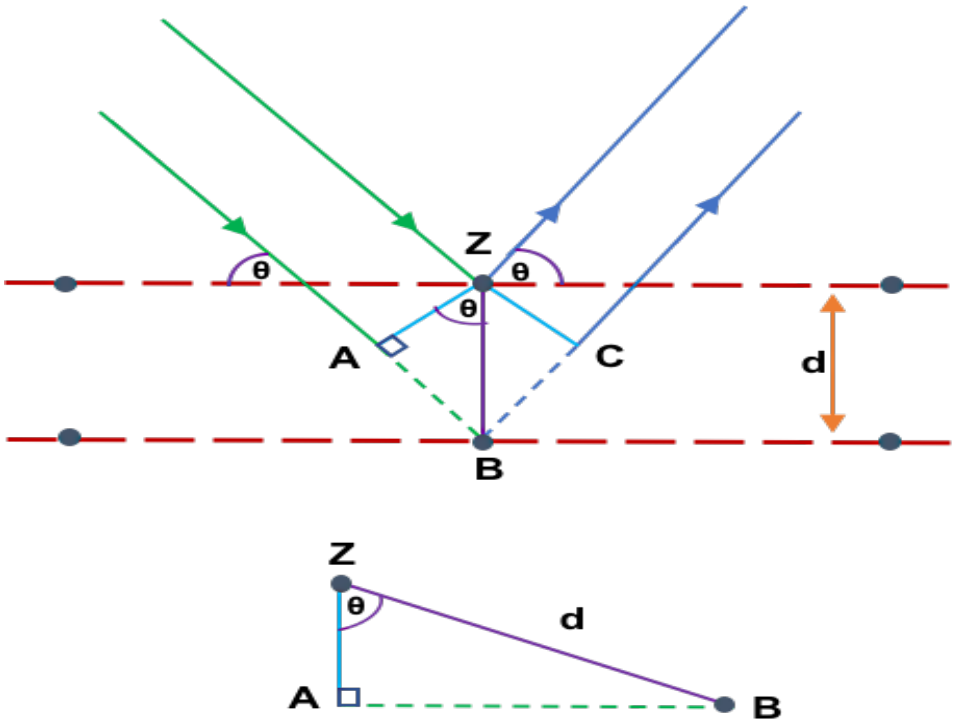


Figure 3.9: A schematic illustration of Bragg`s law using the reflection geometry & trigonometry. The two incident X-ray beams are scattered from points Z & B and lead to diffraction of the beam. The lower beam moves an additional length (AB + BC). d is the distance between Z & B (interplanar spacing), and is the hypotenuse of the right-angled triangle ABZ. For the wave in the upper and lower beam to remain in phase and interfere constructively, the additional distance must be an integral number (n) of wavelengths (λ). This only happens if $n\lambda$ = AB + BC where AB = BC and AB = $d\sin\theta$. The result is that diffraction is only observed when $n\lambda = 2d\sin\theta$ (Bragg's Law).

3.7 Data Collection

Data collection is the most important step in X-ray crystallography. There have been a number of commonly used X-ray detectors for collecting diffraction data:

3.7.1 Image Plate Detectors

The image plate detector has a very extensive capacity to record the reflections of widely variable intensity. It contains a layer of plastic that is treated with a thin coating of a phosphor material, typically crystalline such as BaFBr fixed with Europium (Eu^{2+}). Once the X-rays hit the plate they interact with the Eu^{2+} thus leading to excitation and loss of an electron creating Eu^{3+}. Following X-ray exposure, data are read from the image plate using a fine red laser beam, which can cause Eu^{3+} to return back to stable Eu^{2+} (Rhodes, 2006). The resulting emission of a blue light at a wavelength of 390 nm after that can be detected by a photomultiplier, which reads the diffraction patterns and converts them to a digital signal. Image plates can be deleted by exposure to bright visible light and reprocessed indefinitely. The image plate's detectors are low sensitivity and it can take a long time to collect the data. Therefore, they are not ideally suited to synchrotron experiments, which demand very fast readout time.

3.7.2 Charged-Coupled Device (CCD) Detectors

In this detector, the image is formed on a fluorescent screen connected to a CCD by fiber-optic cables (Figure 3.10). The crystal-diffracted X-rays strike the phosphor screen and are changed into photons of visible light, which move towards a pixel on a CCD chip. Every pixel in the array after that can accumulate electrons from absorption events that are read out to a computer, and the data are recorded (Rhodes, 2006).

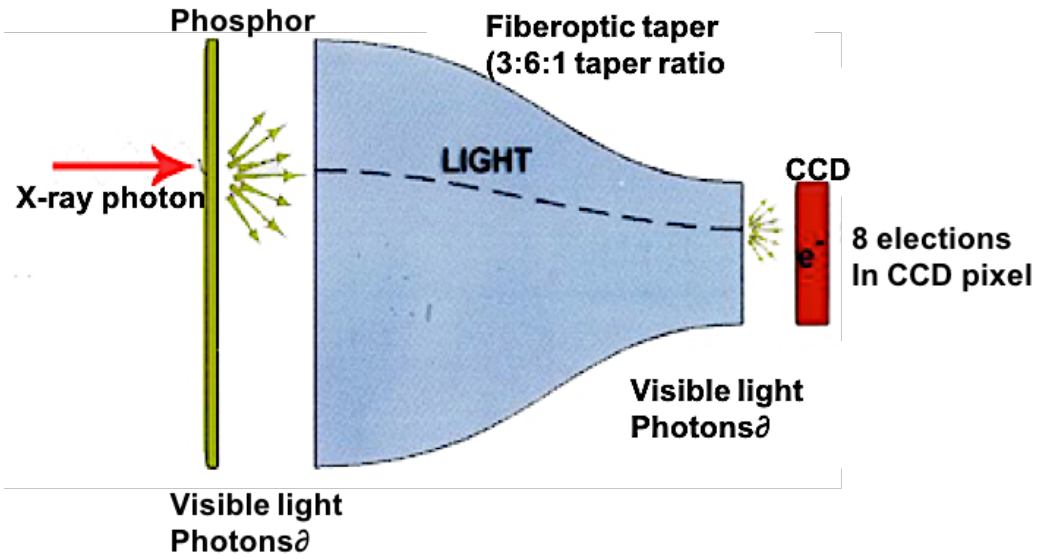


Figure 3.10: Diagram showing Charge Coupled Device (CCD). Adapted from: http://www.xtal.iqfr.csic.es/Cristalografia/parte_06-en.html

The CCD detectors have many advantages that include rapid read-out times, reducing the data collection time, which therefore can reduce the effect on the crystal from radiation damage.

3.7.3 Pilatus Detectors

The Pilatus detector counts a single photon in a pixel array where the incoming photons are directly measured, which permits the data to be collected continuously. It is available on many beamlines at synchrotrons. The Pilatus system has advantages over CCD detectors and image plate detectors. The system has a very large dynamic range and low noise, reducing the time of data collection via a very fast signal readout.

3.8 Data Processing

When the diffraction data has been collected from the protein crystal, it needs to be processed by different programs to get the information on the crystal contents, involving the measurement of intensities and position of reflections. These programs include steps for data indexing and integration, and scaling and merging data.

3.8.1 Data Indexing and Integration

After collecting the diffraction data of the crystal protein, they should be indexed and integrated from all the images to yield a complete dataset. This is performed automatically using the Xia2 systems at the Diamond Light Source to achieve X-ray diffraction data processing by indexing and integrating the data using the programs XDS and scaling by Xscale (Winter, 2010). During the process of indexing, the crystal space group and unit cell dimensions of the crystal are determined. This processing can be done by measuring the distance between reflection spots from one or different multiple images. Ideally two images 90° apart are normally used as a good starting point. When the

parameters of the crystal and detector have been assessed to high accuracy, the integration of the dataset can be obtained. The crystal parameter information requires defining the distance between the crystal and detector, the relationship between the X-ray beam and crystal i.e., what are vertical and horizontal axis directions, and also the angle at which the crystal lattice setting is relative to the X-ray beam. The data are then integrated by measuring all reflections from all images and converting them into a standard mtz file, which is used in the following data processing.

3.8.2 Scaling Data and Merging

The mtz file output from XDS is used as input for XScale, which combines all reflections from different images and operates a scaling factor to every individual reflection. This scale is used to adjust for any changes that came from the data collection process such as radiation damage and the resulting statistical analysis is used as an analytical tool to study data quality and to review if the data collection was acceptable.

The XScale program creates an output log file with measurements or statistics that explain the quality of the data. For instance, completeness measures how much of the possible data has actually been recorded. The dataset should be as near as possible to 100% complete. High multiplicity, the repetition of the same reflection, can give further accurate data and also is necessary for use of the weakest anomalous signal. R_{merge} (or R_{sym}) is used for the calculation of an agreement between the independent measurements of symmetry correlating to reflections in a data set as shown by the equation:

$$R = \frac{\Sigma_{hkl}\,\Sigma_j |I_{hkl,j} - < I_{hkl} >|}{\Sigma_{hkl}\Sigma_j\, I_{hkl,j}}$$

Where $\langle I_{hkl} \rangle$ is the average of a symmetry or Friedel mate that is related to the observations of a unique reflection. This discrepancy of R_{merge} can be solved by the

introduction of $R_{p.i.m.}$. This R Factor is a precision-indicating merging factor which highlights the average measurement of the full data set, the description of $R_{p.i.m}$ can be seen in the equation below:

$$R_{pim} = \frac{\sum_{hkl}\sqrt{\frac{1}{n-1}}\sum_{j=1}^{n}|I_{hkl,j} - \langle I_{hkl,j}\rangle|}{\sum_{hkl}\sum_{j}I_{hkl,J}}$$

Another indicator that can be determined is the redundancy-independent merging R value $R_{r.i.m.}$ (or R_{meas}) This R factor or value measures the accuracy of each measurement and is independent on how often the measurement of the reflection has been taken.

$$R_{meas} = \frac{\sum_{hkl}\sqrt{\frac{n}{n-1}}\sum_{j=1}^{n}|I_{hkl,j} - \langle I_{hkl,j}\rangle|}{\sum_{hkl}\sum_{j}I_{hkl,j}}$$

Furthermore, I/σI is another indicator of the quality data that is providing information on the signal to noise ratio of the data. The values for a typical resolution cutoff are commonly argued with many crystallographers preferring no cutoff and others when I/σI is lower than 2.0 (Evans, 2006).

3.9 Phase Problem and its Solution in X-ray Crystallography

There are a number of processing package programs that can be used in X-ray crystallographic work to analyse crystal diffraction data. The phase problem in physics is linked to the problem of losing information regarding the phase of an X-ray wave that can happen during determination of a structure from diffraction data (Winn *et al.*, 2011). However, the biggest challenge in X-ray crystallography is to determine phases that are needed together with the intensities of each reflection in order to produce an electron density map, and hence obtain a model of the protein's structure (Taylor, 2010). Hence, in X-ray crystallography, there are various ways to find the lost phases such as the molecular replacement (MR) experiments, multiple-wavelength anomalous diffraction

(MAD) and single-wavelength anomalous diffraction (SAD). Phasing was attempted using a SAD technique for the work with Cj0241-Hr protein, in collaboration with my supervisor Dr. John Rafferty, University of Sheffield. The process and techniques of these programs used are explained in chapter (4).

Experimental phasing by SAD method was conducted using the AutoSol wizard from the Phenix suite (Adams *et al.*, 2010). Molecular replacement, however, was conducted using the PHASER program (McCoy *et al.*, 2007) from the CCP4 suite (Winn *et al.*, 2011). Initial structural model building and refinement was done with the Autobuild program using phenix. refine in the PHENIX suite.

3.9.1 The Patterson Function (Patterson Map)

The Patterson function developed by Arthur Lindo Patterson in 1935 at Massachusetts Institute of Technology is used to solve the phase problem in X-ray crystallography by defining atomic locations by using the intensities (amplitudes squared) as coefficients instead of structure factors (amplitudes and phases). It can calculate the diffraction pattern with a lack of phase information (Glusker, 1984). The Patterson function P (uvw) is a Fourier transform (FT) of the reflection intensities assuming all phase angles are set to 0.

$$P(uvw) = \frac{1}{V}\sum_{hkl}\left|F(hkl)\right|^2 \cos[2\pi(hu + kv + lw)]$$

where *u*, *v* and *w* are related to Patterson cell coordinates of a point on the Patterson map (Drenth, 2007). The peaks in the Patterson map are interatomic distance within the same molecule and between different molecules in the crystal. The Patterson function can be used to find small numbers of heavy atoms bound to a protein. It is useful in phasing procedures such as MR to help find the orientation and location of a model using rotation

and translation functions, and in SAD, MIR and MAD for finding heavy atom locations in crystals derivatized with the heavy atoms.

3.9.2 Single-Wavelength Anomalous Dispersion (SAD)

Single-wavelength anomalous dispersion (SAD) was used to solve the protein structure in this research study. The reason for choosing this technique is because SAD uses a single dataset at a single suitable wavelength. One advantage of this system is reducing the time of the X-ray beam shining onto the crystal, therefore, reducing potential radiation damage to the crystal during collecting the data (Rhodes, 2006; Rupp, 2010).

The heavy atom substructure can be derived from the SAD data (Rupp, 2010). Most atoms, such as iron atoms, show anomalous dispersion when the X-ray wavelength is close to the atom's absorption edge. This characteristic of these elements is considered as highly important when solving the phase in X-ray crystallography. The hemerythrin proteins contain two iron atoms, thus, the X-ray wavelength can be set at the synchrotron to the specific absorption energy of iron to collect anomalous datasets. However, the X-ray wavelength used in protein crystallography is not close to the absorption edge of the majority of atoms that compose the proteins such as oxygen, nitrogen and carbon; these atoms do not participate in the anomalous scattering.

3.9.2.1 Anomalous Scattering

Anomalous scattering happens when the incident X-ray photon of the right wavelength is absorbed by heavy atoms in the crystal, creating a transition between inner shell electrons (orbitals) to a higher energy level. When the excited electron returns to the inner shell, as the energy is lost from the atom, a new X-ray photon is emitted but with a different amplitude and phase to the incident X-ray. This phenomenon is known as the anomalous scattering. At the wavelength when the photon has sufficient energy to release

the electron out of its orbital, there is a fast change in the absorption factor, usually represented as the atom's absorption edge. As mentioned above, the majority of atoms that compose proteins such as oxygen, nitrogen and carbon do not contribute to the anomalous scattering, but do normal scattering, because the X-ray wavelength used is not near the absorption edge of these atoms. However, proteins can have components with atoms that have the behavior of anomalous scattering such as Se, Hg, Fe…etc., with the normal scattering from the other atoms (Figure 3.11). Initially the presence heavy atoms in the protein crystal are confirmed experimentally by an X-ray fluorescence scan at the synchrotron before beginning to collect the data. This permits selection of the wavelengths at which anomalous scattering will be best to collect the datasets. Usually wavelengths at the peak of the absorbance edge and the inflection point are used to determine the structure (Figure 3.12). In order to combine the amplitude and phase of the anomalous scattered wave, the atomic scattering factors are taken into consideration:

$$f = f_{\circ} + f' + \mathbf{i}f''$$

Where $f_{\circ}$ is the normal scattering factor that is depends on the Bragg angle, f' and f'' are dependent on wavelength. f' is a heavy atom dispersion term that adjusts the normal scattering factor, whereas the f'' is the absorption term that is 90° advanced in phase

3.9.2.2 Friedel's Law

In Friedel's law the pair of reflections $|F_{h,k,l}|$ and $|F_{-h,-k,-l}|$ have the same structure amplitudes and their phases have opposite sign (Blow, 2002). However, this law breaks down in the presence of an anomalous scatterer, so that $|F_{h,k,l}| \neq |F_{-h,-k,-l}|$. The anomalous component of the scattered wave from the heavy atom is usually at 90° to the advancing non-anomalously scattered wave (Figure 3.13) and this creates the differences in $|F_{h,k,l}|$ and $|F_{-h,-k,-l}|$. The anomalous differences can be used to find the anomalous scatterers.

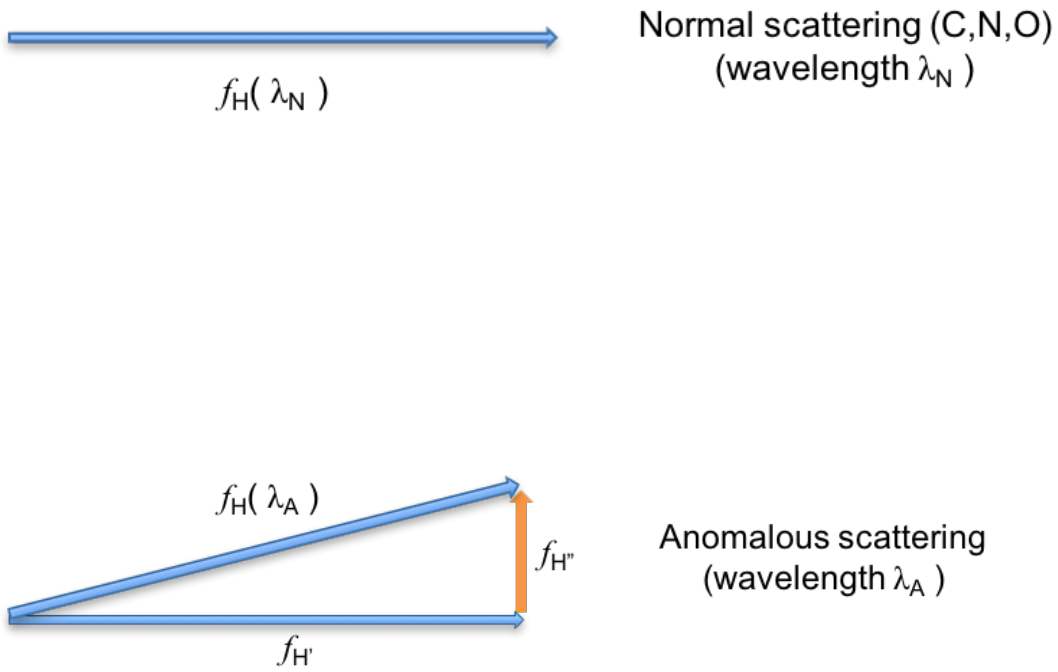


Figure 3.11: A heavy atom scattering factor f_H at two different wavelengths. At wavelength λ_N it scatters normally. At wavelength λ_A there is an anomalous scattering component $f'_{H''}$ for which the phase is 90° in advance of the normal component, and also the normal component of the scattering is reduced to $f_{H'}$. Taken from Blow, 2002

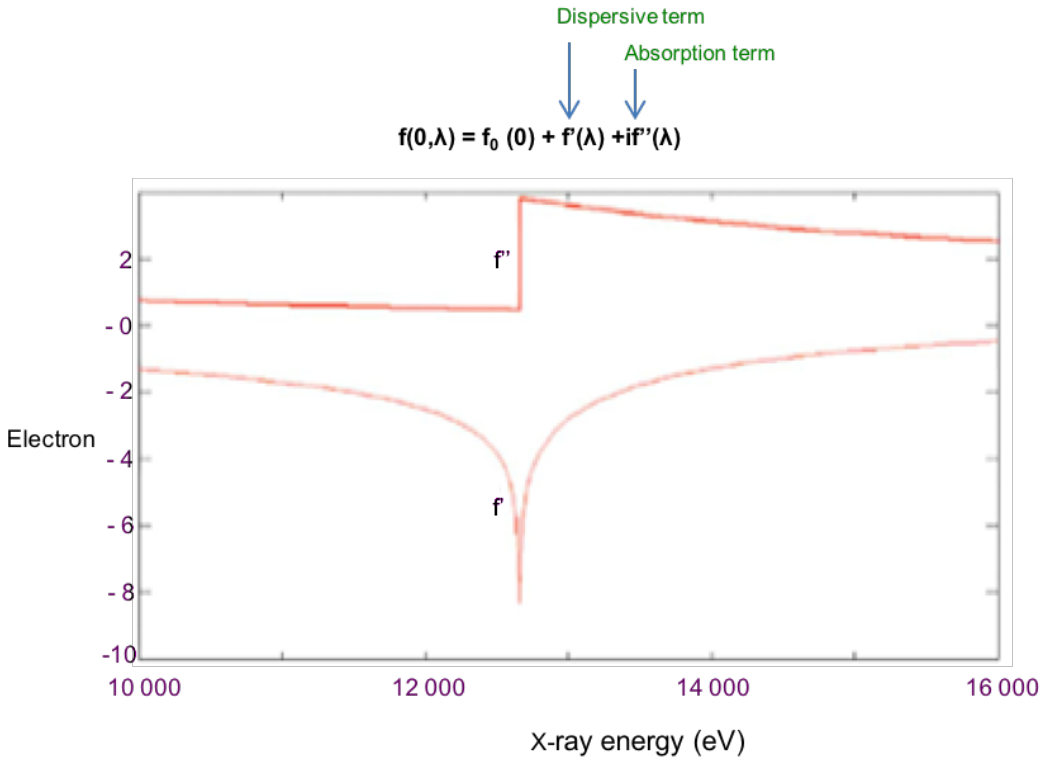


Figure 3.12: Variation in anomalous scattering signal against incident X-ray energy near the *K* edge of the heavy atom (Selenium). Adapted from Taylor, 2010.

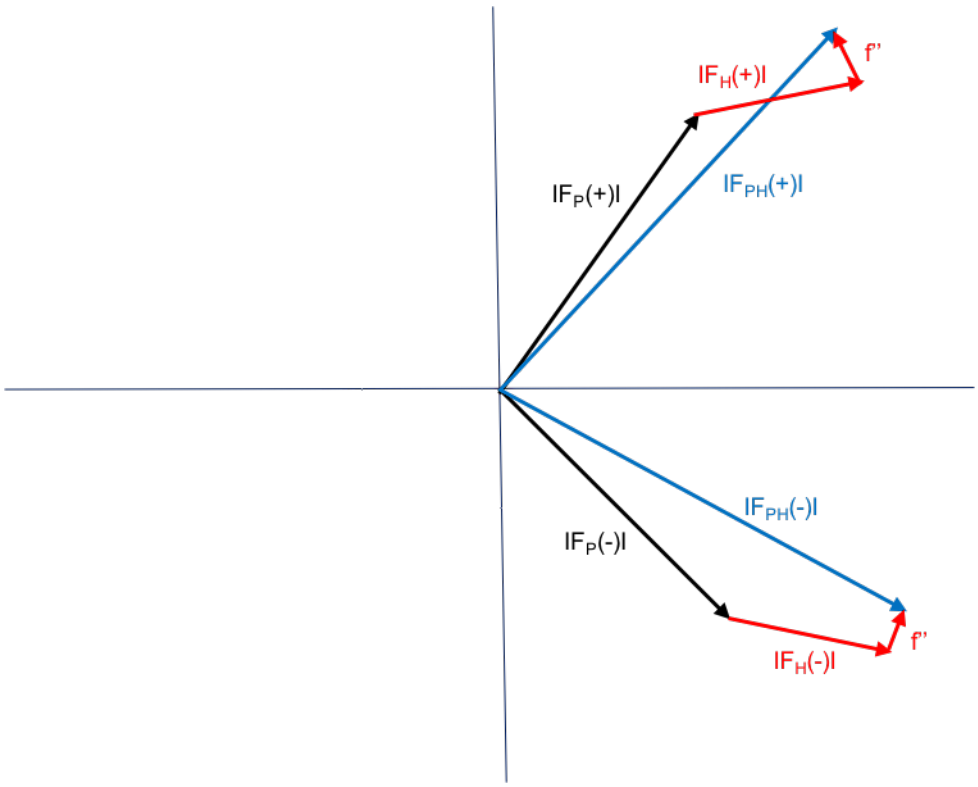


Figure 3.13: Breakdown of Friedel's law when an anomalous scatterer is present. $f = f_\circ + f' + i f''$. $|F_{h,k,l}| \neq |F_{-h,-k,-l}|$ or $|F_{PH}(+)| \neq |F_{PH}(-)|$. $\Delta F^{+/-} = |F_{PH}(+)| - |F_{PH}(-)|$ is the Bijvoet difference. Adapted from (Taylor, 2010).

3.9.2.3 Phasing

During anomalous scattering, the difference in scattering intensity (I_{hkl} and $I_{-h, -k, -l}$) (proportional to the amplitudes ($|F_{hkl}|$ and $|F_{-h, -k, -l}|$)) is called a Bijvoet difference. The Bijvoet differences can be used to determine anomalous scattering due to the heavy atom in F_{PH}. The position of the heavy atoms in the unit cell can be determined by using a Patterson map. The peaks obtained from the Patterson maps can be used to calculate the amplitudes and phases of the atomic structure factors of the heavy atoms. From these phases and amplitudes, phases for the original reflection including the protein part can then be derived computationally using X-ray crystallography software.

In the SAD technique, there are two final choices for the electron density map from the first set of phase estimates and the small differences in the resulting maps can be strengthened to select the correct phase by applying techniques such as solvent flattening (Taylor, 2010).

3.9.2.4 Solvent Flattening

The solvent flattening method is one of the most effective ways to improve the electron density map. In a protein crystal, the solvent content is ranged from approx. 27% to 65% with an average of about 43% (Matthews, 1968; Chruszcz *et al.*, 2008). These solvent areas compared to the protein areas are very disordered. So, as a result of this absence of order in the solvent area, the electron density map displays fewer features and has a lower mean density value (Kantardjieff and Rupp, 2003). The solvent flattening method sets a mean electron density value for the solvent region while the protein regions are left unaltered. Subsequently, the electron density distribution including the solvent regions is used to calculate the structure factors. In addition, the new phase angles can be

used to improve the electron density of the map. By flattening the solvent areas, the phases are more improved and the protein areas of the density map are improved too. This procedure can be repeated many times to obtain significant results in the quality and also improve the map density in the protein region (Rupp, 2010)

3.9.3 Molecular Replacement (MR)

Another method in X-ray crystallography to determine the phase solution is by using molecular replacement (MR). In this technique, there is a need for the existence of a previously solved protein structure that has some structural homology to the unknown protein that the diffraction data has derived from (McCoy *et al.*, 2007). The phase information obtained for the target structure can be used to create electron density maps. This method requires the model protein to be placed into a similar orientation as the unknown protein in the unit cell, and normally this rotation happens by using the Patterson maps of the known and unknown proteins. The right placing within the unit cell then needs a translation stage of the rotated model.

3.9.3.1 Rotation Function

The rotation function compares the Patterson map of the unknown structure with the map of the known homologue structure model in different orientations. The Patterson maps are oriented and overlapped in stages within three-dimensional space. In every orientation of the model, the program calculates the Patterson functions of the model of the known protein structure and the target protein using its structure factor amplitudes from the experimental data. Then it compares them and checks the agreement of the two, until the maximum similar orientation is found. This procedure is assessed by a correlation coefficient and is called the Rotation Function. The following stage is the translation function, that is necessary to define the exact location within the unit cell.

3.9.3.2 Translation Function

The translational function is normally a procedure of arranging the oriented protein to be located in the asymmetric unit. The rotated model needs to be translated into the right coordinates, x, y, z, to determine the accurate location within the unit cell. Generally, this procedure is used if there is a single symmetry copy or more than one molecule in the asymmetrical unit, which can take a long time to process as every copy will need to be oriented. However, molecular replacement is considered the fastest technique for defining initial phase for the reflections.

3.10 Electron Density Calculation

Once the phase of the X-ray scattering has been defined, it is possible to calculate an electron density map to create an atomic model of the protein structure. The relationship between the structure factor (F) from reflection *hkl* and the electron density (ρ) at cell coordinates xyz and volume V of the unit cell can be identified by the following equation as seen below, but the prime objective in this protein crystallography is to determine and calculate the electron density (ρ) at all the positions of x, y, z in the cell rather than a diffraction pattern.

$$\mathbf{F}(hkl) = V \int_{x=0}^{1} \int_{y=0}^{1} \int_{z=0}^{1} \rho(xyz) \exp[2\pi i(hx + ky + lz)] dx dy dz$$

This calculation permits the diffraction pattern of a known structure to be estimated or can be calculated if the electron density cloud is known; therefore, the Fourier transform of **F** (*hkl*) allows the electron density for every fractional coordinate x, y and z in the unit cell to be calculated by the following equation;

$$\rho(x\ y\ z) = \frac{1}{V} \sum_{hkl} |F(hkl)| \exp[-2\pi i(hx + ky + lz) + i\alpha(hkl)]$$

This calculation can produce a map of electron density that can be interpreted in terms of atomic locations permitting a model of the protein structure to be built. When a model structure has been constructed, it is then refined versus the data.

3.11 Model Rebuilding and Refinement

Once the initial phases of the target protein have been calculated, and an appropriate model is obtained, the model needs to be refined to improve it by iterative rounds of re-building and refinement. The goal of this procedure is to build the structure model that best describes the experimental data. The COOT program was used to examine and visualise the electron density map during the re-building of the model (Emsley and Cowtan, 2004). The REFMAC5 programs in the CCP4 suite and phenix.refine option in the PHENIX suite (Adams *et al.*, 2010) were used to refine the structures of the models in this project. By refinement and rebuilding of the old structure several times, the agreement increased between the new model and the electron density map. The refinement includes re-location of the side-chains, adding molecules or atoms corresponding to un-built parts, where there is definitive density in the map, changing residues, adding or deleting water molecules (Emsley and Cowtan, 2004). The new model after refinement can be assessed by comparing the value of the R-factor and the free R-factor that can measure the agreement between the collected data and obtained model.

3.12 Evaluating the Refinement

During the refinement and re-building process, it is important to monitor the improvement of the agreement between the model and the diffraction data. This agreement can be analysed by evaluating the value of an R factor. The R factor (R_{work}) is a measurement tool of the quality of the model desired. It is described as the ratio of the

sum of differences between observed $|F_{obs}|$ and calculated structure $|F_{calc}|$ amplitudes to the sum of observed structure amplitudes and it is described by the following equation:

$$R = \frac{\sum ||F_{obs}| - |F_{calc}||}{\sum |F_{obs}|}$$

A decreasing value of the R-factor during refinement describes how the model is corresponding well with the diffraction data. Its values rely on the order in the crystal, the resolution, B-factors, the diffraction data etc. The second factor used for validation is a free R factor (R_{free}) which is defined by using a small random subset of the diffraction data (typically 5%) that is omitted from the refinement procedure to avoid the overfitting (bias) of the model during refinement (Terwilliger *et al.*, 2008). The R_{work} and R_{free} should be close to each other, and a value of more than 5% difference shows that the model has been over fitted to the data and become biased. The free R factor is expressed as the following equation:

$$R_{free} = \frac{\sum_{hkl \sqsubset T} ||F_{obc}| - |F_{calc}||}{\sum_{hkl \sqsubset T} |F_{obc}|}$$

Where $hkl \sqsubset T$ describes all reflections from the free R set. At the end of any refinement with a R factor and R_{free} that cannot be improved, the model properties should be stereochemically and conformationally sensible, and must be validated before placing it in the databanks.

3.12.1 The B-factor

The B-factor, also known as the "temperature factor" or by the more precise term as the atomic displacement parameter (ADP), is a measure of how the atom vibrates or oscillates around its mean position in a particular model. The B-factor values usually vary from 1 Å to 100 Å. i.e. low-resolution data often gives models with high B-factors because of the variations in locations of atoms distributed in the copies of the protein

through the crystal lattice. Atomic vibration is recognised in that the side chain atoms in a protein molecule have more freedom of movement than those in the main chain. Higher temperature means higher vibrational energy of the atoms, therefore, more movement from the mean equilibrium location. This vibration can disturb the data diffraction, so it is appropriate to assign a B-factor or "temperature factor" to each atom (Rupp, 2010). During the refinement, atoms in the model structure which have more movement or freedom can be seen by comparing the B-factor values and therefore some insight can be found into the dynamics of these atoms in the model (Rhodes, 2006).

3.12.3 The Occupancy factor

Another parameter involved in refinement is the occupancy of each atom in the structure model. The occupancy of most atoms in the protein structure have a full occupancy set at 1.0, except when the atom has two or more likely locations such as when the side chain occur has two conformations. If the two conformations have equal frequency, the atoms involved have occupancies of 0.5 each in their two likely locations. For multiple conformations, the occupancy of the protein atoms must sum to 1.0. For non-protein atoms such as metal cofactors or solvent, other values for the occupancy are possible and can be correlated with the temperature factor of the atom and the surrounding atoms (Rhodes, 2006).

3.13 Model Validation

The validation of a structure model must be performed during the refinement and building process. It depends on the experimental guide provided by using the electron density map and chemical knowledge of bond angles, bond lengths and other chemical properties of a protein molecule. The R_{work} and R_{free} factors can avoid general errors and

over fitting but lack some information about local problems such as mis-built loops or residue side chain errors.

A Ramachandran plot or a [φ, ψ] plot is a good indicator to visualise energetically the permitted regions for the conformation of the backbone dihedral angles phi (ϕ) and psi (ψ) of the amino acids in the protein structure. The values of phi (ϕ) and psi (ψ) are very good signs of the model quality. They have been seen as clusters in the allowed regions and only a small number of residues (typically <5%) should lie in disallowed regions. The MolProbity program also accomplishes a number of validation procedures involving analysis of rotamer distributions, hydrogen bonding, atom clashes and van der Waals forces in the interfaces between the components of a structure (Chen *et al.*, 2010).

Chapter 4

Cj0241 Hemerythrin Expression, Purification and Crystallisation

This chapter will explain the work on Cj0241-Hr protein as presented in chapter (1), section (1.8.6). The chapter will discuss the experiments that were done to find crystals of Cj0241, solve the structure, refine it and do some functional studies.

4.1 Confirmation of the sequence of the insert *cj0241c* gene in the pET-21b plasmid

The *cj0241c* gene that had been inserted into pET-21b plasmid was validated by DNA sequence analysis after extracting the DNA construct using a Miniprep kit (Qiagen), and sending it to GATC Biotech. The sequence was analysed using the Basic local alignment search tool (BLAST) (Altschul *et al.*, 1990) at the National Centre for Biotechnology Information http://blast.ncbi.nlm.nih.gov/Blast.cgi, which showed 100% identity with the sequence of Cj0241 from the universal protein resource (UniProt) database of protein sequence as presented in chapter (1) section (1.8.6) in Figure (1.5).

4.2 Testing the expression of the hemerythrin Cj0241 protein

Transformation and expression were performed aerobically in *E.coli* strain BL21 (DE3) as described previously in chapter (2), section (2.4). The experiment was performed to define the best temperature and time following induction for overproduction of Cj0241. Two temperatures were tested (37°C and 25°C) and the time period of the expression was analysed using an SDS gel as shown in Figure (4.1).

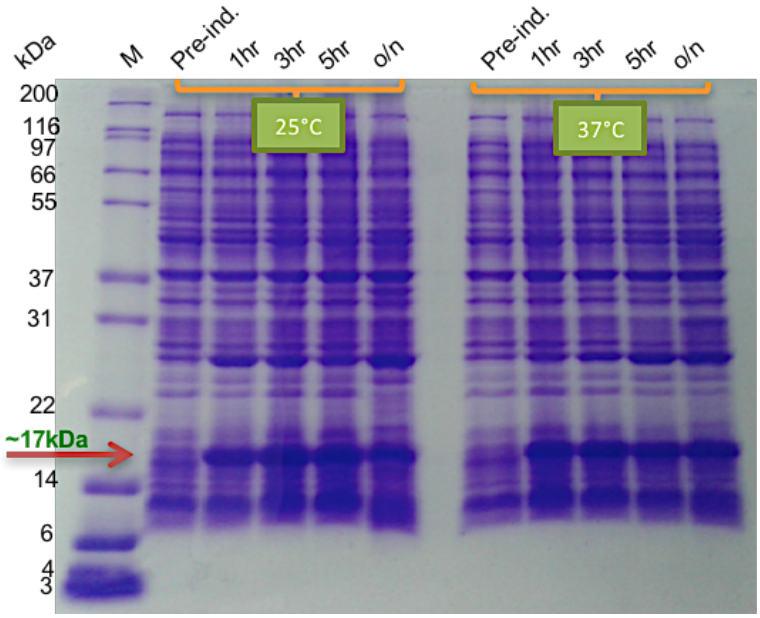


Figure 4.1: The overexpression of Cj0241-Hr protein. Samples were collected before induction then at 1, 3, 5 and 16 (overnight) hours post- induction. A band designated by an arrow, at approx. 17 kDa in all induced lanes displays that the Cj0241-Hr protein expression was positively induced. M= The Mark12™ Unstained Standard, numbers show dimensions of molecular weight of sample protein in kDa.

From an observation of the gel, it is clear that the Cj0241-Hr protein (indicated by the arrow- ~17 kDa) reached overexpression during the time periods. Also, the desired protein has to be soluble and this was also checked at 37°C and 25°C using small scale overexpression as described in chapter (2) section (2.5.1). Both temperatures encouraged the bacteria to produce soluble protein (Figure 4.2).

4.3 Purification of His Tag- Cj0241 hemerythrin

The Cj0241-Hr protein was purified using a nickel affinity column chromatography technique as explained earlier in chapter (2) section (2.10.1). Two peaks were observed from the chromatogram (Figure 4.3A). The elution profiles of fractions from 13 to 17 were checked for protein concentration using the Bradford method (section 2.5.3), and the progress of purification was analysed using 12% SDS-PAGE as presented in Figure (4.3B). From the gel, we can see that the band of Cj0241-Hr protein appeared at molecular weight of approx. 17 kDa. Lanes 1 and 2, which are cell debris and cell free extract (CFE) respectively showing the high level of hemerythrin expression. Lanes 3, 4 & 5 indicate that all the hemerythrin Cj0241 was bound to the Ni column. The separation of the hemerythrin from contamination and protein was not very good but fractions 15 and 16 corresponding to lanes 10 and 11 respectively were the purest fractions. These two fractions were combined with total protein at 9.6 mg, and used to prepare a crystallisation sample as described in chapter (2) section (2.10.2).

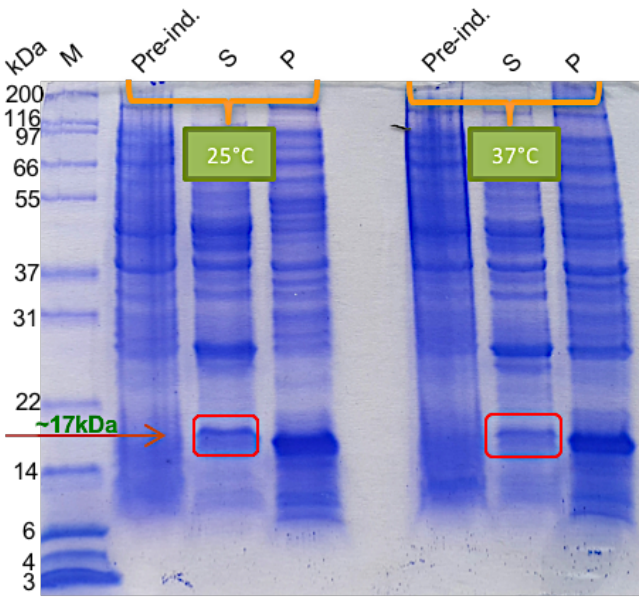


Figure 4.2: Cj0241-Hr protein solubility was checked at 25°C and 37°C. Both temperatures permitted the bacteria to produce soluble protein, indicated with an arrow. S = soluble protein, P= insoluble protein and M= The Mark12™ Unstained Standard, numbers show dimensions of molecular weight of sample protein in kDa.

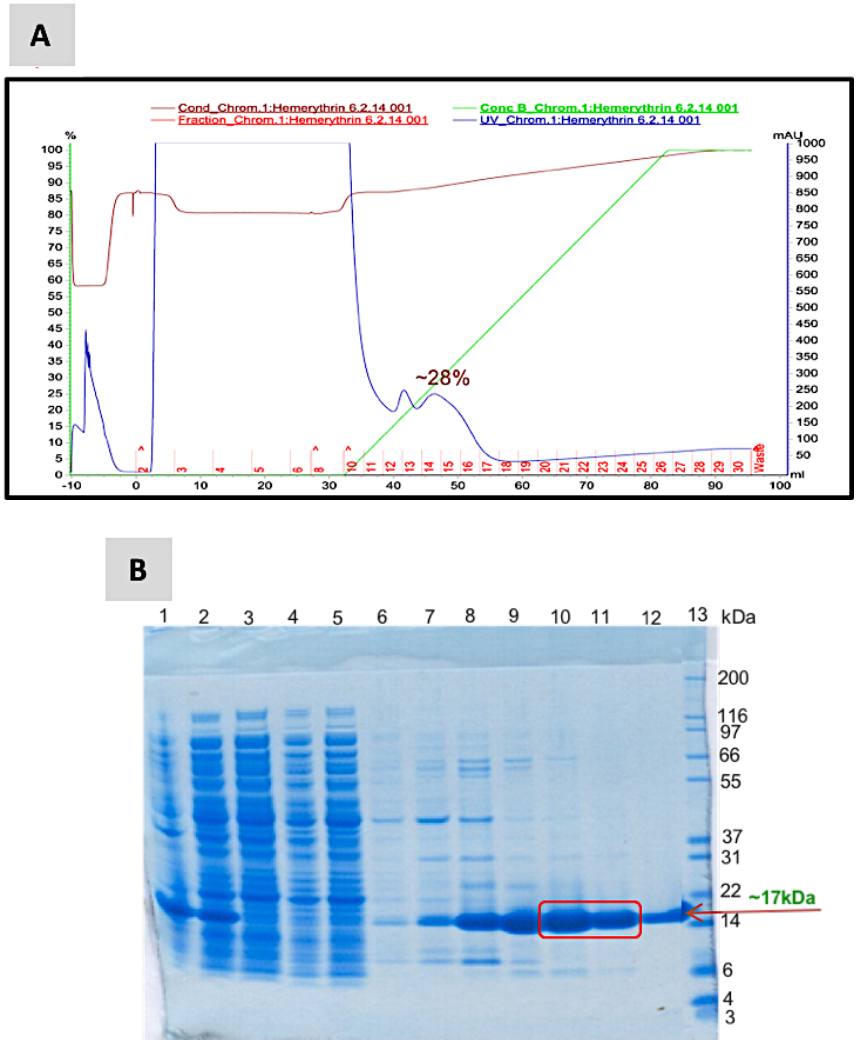


Figure 4.3: Analysis of the elution of Cj0241-Hr from a Ni-affinity column. Panel (A) The chromatogram the purification of Cj0241-Hr via Nickel affinity column (HisTrap HP 5ml from GE Healthcare™). Following sample application and column wash, an increasing gradient (0-70%) of 0.5M Imidazole was applied. Cj0241-Hr was eluted at ~28 %, which corresponds to 0.14M Imidazole. **Panel (B)** SDS-PAGE analysis illustrating the purity of the Cj0241-Hr protein at various steps of the purification process. A band for Cj0241-Hr protein appeared which had molecular weight ~17kDa, as indicated by an arrow. Lanes 1 and 2 are cell debris and CFE, respectively, and show the high level of hemerythrin expression. Lanes 3 -5 corresponds to washing fractions (2,5&6) and they indicate that all the hemerythrin Cj0241 was bound to the Ni column. The hemerythrin can be seen in lanes 6-12 corresponding to fractions (12-17) respectively. Fractions 15 &16 (red square) the purest protein was combined together to prepare the crystallisation sample. Lane 13 the Mark12™ (M) Unstained Standard (Thermo Fisher Scientific).

4.4 Gel filtration analysis of oligomeric state of Cj0241

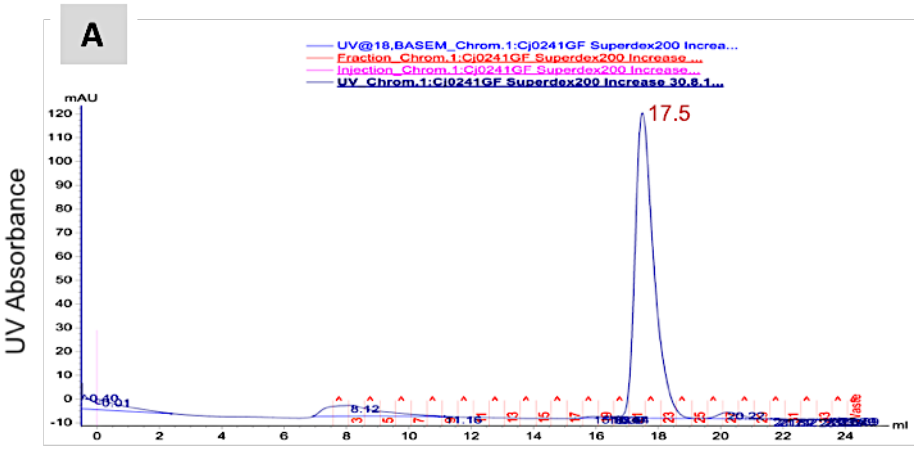
A sample of Cj0241-Hr protein was used in an analytical gel filtration experiment to examine the size of the protein in solution. 1.5 mg of Cj0241 protein purified using His Trap HP column was applied onto a Superdex 200 Increase column, equilibrated in buffer 0.5M NaCl, 50Mm Tris-HCL pH8.0. The column was then run at a flow rate of 0.5ml/min. 0.5ml fractions were collected. A single peak was eluted at 17.5 ml. This corresponds to a K_{av} of 0.60, which is equivalent to a 16 kDa molecular weight, this is in line or close to the expected monomeric mass of Cj0241-Hr protein (Figure 4.4).

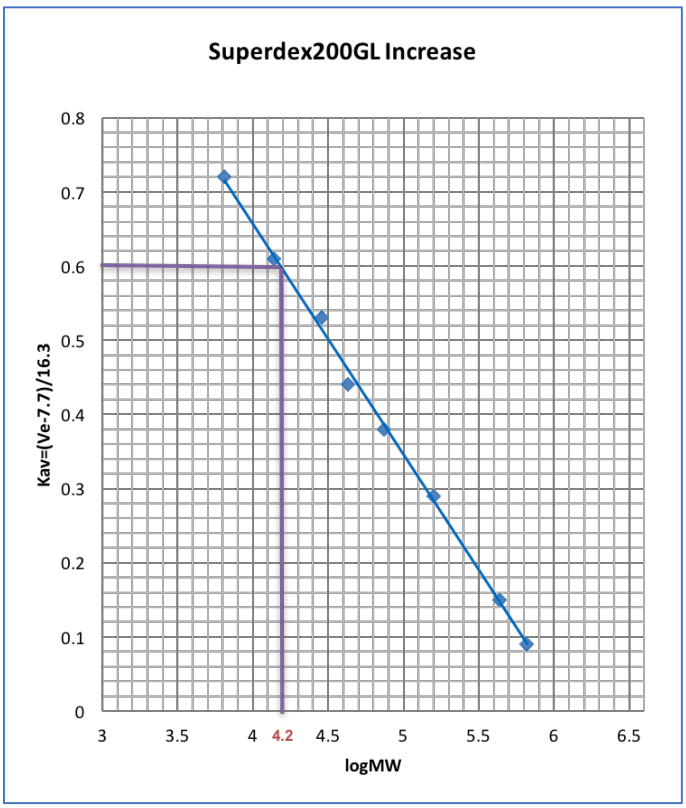
4.5 Sample preparation of Cj0241-Hr for crystallisation trials

For crystallisation trials, the purified protein Cj0241-Hr was concentrated to 7.5 mg/ml in 10mM Tris-HCL buffer pH 8.0 and 0.1M NaCl using a Zeba spin- desalting column. The initial crystallisation screening trial was set up by a Matrix Hydra II Plus One robot as mentioned in chapter (2) section (2.10.2). The purified Cj0241-Hr has a greenish- yellow colour, which may be coming from the iron ions expected to be bound by the protein; however, more details about protein characteristics will be mentioned in chapter (4) section (4.13.6). In addition, the molecular weight of the purified protein was checked by mass spectrometry. The result exhibited the molecular weight of this protein as 17123.16 Da (Figure 4.5) as expected from the intact native protein sequence.

4.6 Initial crystallisation using commercial screens

Initial crystallisation experiments were carried out using various screening kits including the PACT screen, JCSG screen and PEGS screen (Qiagen), with a Hydra automated crystallisation robot, and all trays were kept in the incubation room at 17ºC.





B

Use the calibration plot for Hi-Load superdex 200 column chart to calculate the protein MW:
K_{av} = 0.60
LogMW = 4.2
MW = 16 kDa

Figure 4.4: Gel filtration purification of Cj0241. Panel (A) The elution profile from Superdex200 Increase gel filtration column of Cj0241.**Panel (B)** The calibration plot for Hi-Load Superdex 200 Increase column. Calibration was obtained using standard proteins (Gel healcare low and high molecular weight kits).

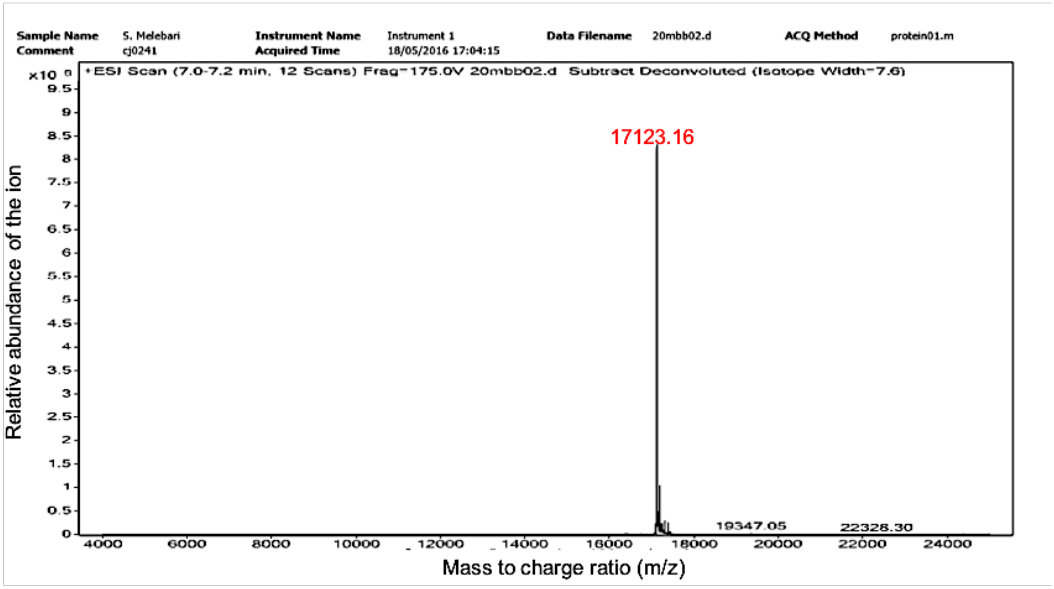


Figure 4.5: Cj0241-Hr Mass Spectrometry analysis. The results obtained from Mass Spectrometry verified the molecular weight of Cj0241-Hr protein, and agreed with a predicted molecular weight of 17122.86 Da as calculated from the native sequence. The result corresponds to the Cj0241 sequence, minus an N-terminal methionine, plus 2 linker residues and a 6-histidine tag.

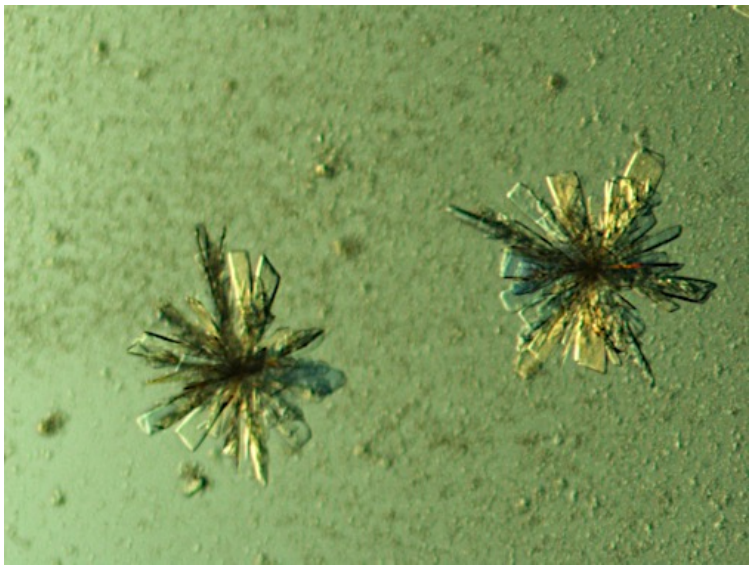


Figure 4.6: Initial crystallisation hits of Cj0241-Hr protein crystals. The crystals showed at condition A10 in the JCSG screen with 0.2M potassium formate pH7.3 and 20% (w/v) PEG 3350 precipitant.

After 3 days, thin needle-shape crystal clusters were found (Figure 4.6) in the JCSG screen in well condition A10, which contained 0.2M potassium formate pH7.3 and 20% (w/v) PEG 3350 precipitant.

4.7 Optimisation from initial screen

The initial crystals that were grown in well A10, the conditions of the JCSG screen were optimised to produce large crystals by varying the concentration of PEG, potassium formate and buffer pH as described in chapter (2) section (2.10.3). Crystals were produced that were suitable for X-ray data collection, grown at 0.2 M K Formate pH 7.5 & 8 in 20%(w/v) PEG 3350.

4.8 Experimental Structure determination of Cj0241-Hr protein

For X-ray diffraction measurements, the crystals were picked from their conditions using nylon loops and quickly soaked into a cryo-protected solution and sent to the Diamond Light Source (DLS) in Oxford as mentioned in chapter (2) section (2.10.4).

4.8.1 X-ray data collection and processing

As explained above, the Cj0241-Hr crystal data were collected at the DLS synchrotron in Oxford. The X-ray diffraction data set of the best Cj0241 crystal diffracted to 1.8Å resolution (Figure 4.7). Data sets were collected using an exposure time of 0.150 seconds with 0.15° oscillation and wavelength 0.92000 Å at beamline I04 for a total of 1200 images. The processing of the X-ray diffraction data sets was done using the auto-indexing and collection method of the xia2 system in the 3dii mode (Winter, 2010). This Cj0241-Hr crystal belonged to the monoclinic crystal system (Figure 4.8). It is recognised

by unequal length between the three axes a, b and c, and the two interaxial angles α and γ equal to 90°, while the β angle does not need to be 90° ($a \neq b \neq c$, $\alpha = \gamma = 90°$, $\beta \neq 90°$).

The space group was $P2_1$ with a 2-fold screw axis along b (Figure 4.9), with unit cell dimensions a = 30.9 Å, b = 48.4 Å, c = 43.7 Å, α = 90°, β = 98.7° and γ = 90°. The consequence of processing and data set collection provided the total R_{merge} 0.044 (0.625 for the outer shell), the overall I/σI was 19.3 (2.7 for the outer shell) and the completeness was overall 97.1%. In general, these results show sensible indicators for the data quality (Weiss, 2001; Wlodawer *et al.*, 2008). The processing statistics of Cj0241-Hr are exhibited in Table (4.1).

4.8.2 Matthews Coefficient (Vm) calculation

The Matthews Coefficient calculations (Matthews, 1968) are performed to estimate the number of molecules in the asymmetric unit because it is useful to know the number of protein copies in the cell unit to do molecular replacement. The calculation value depends on the molecular weight of Cj0241-Hr protein and the unit cell volume. Calculation of Vm using the Matthews program from the CCP4 suite suggested that there was one molecule copy in the asymmetric unit cell with a Vm of 1.89 $Å^3$/Da and a solvent content of 34.81%.

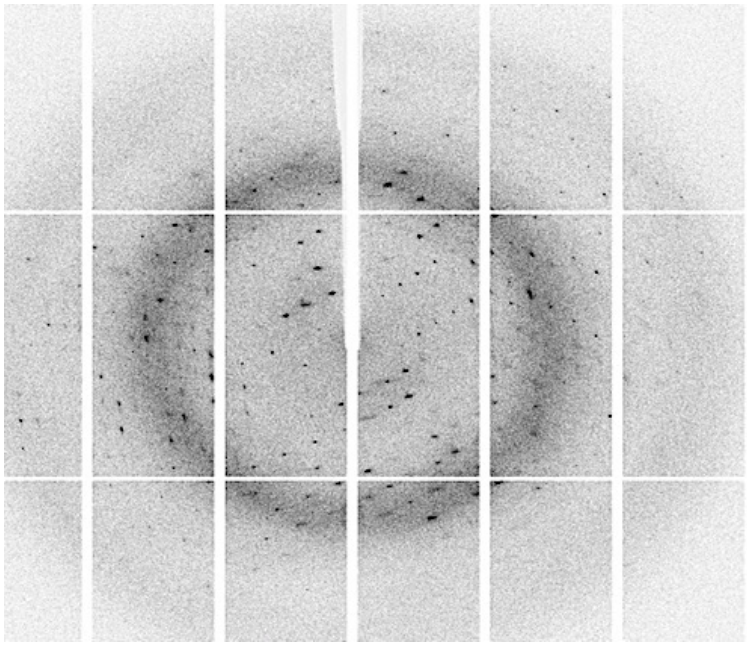


Figure 4.7: X-ray diffraction image of Cj0241-Hr protein crystal. X-ray diffraction pattern of a Cj0241-Hr crystal that had been soaked in 20 % v/v ethylene glycol in the buffer contained 0.2M K. format pH7.5 and 20% PEG3350, collected to 1.8 Å resolution using beamline I04 at Diamond light source. The exposure time was 0.150 s, the wavelength 0.92000 Å, and the oscillation range per image was 0.15°.

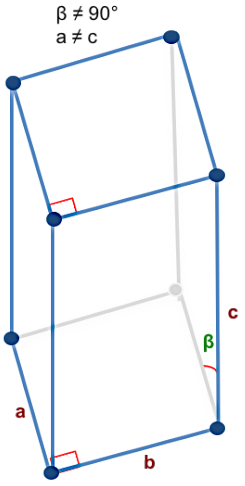


Figure 4.8: **Standard unit cell.** The Cj0241-Hr protein crystal shows according to the unit cell dimensions it is monoclinic type, which is shaped like a rectangular prism with a parallelogram as its base. Adapted from Blow, 2010.

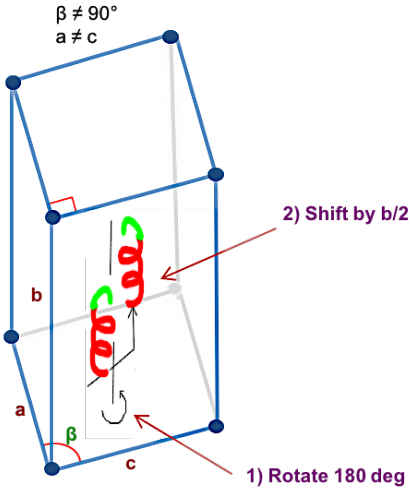


Figure 4.9: Packing in a $P2_1$ unit cell. The 2-fold screw axis contains a rotation of 180° around an axis parallel to b, and followed by a translation of ½ b in the b-direction. Adapted from Blow, 2010.

Table 4.1: The data collection statistics for Cj0241-Hr crystal (**1st dataset**).

Data collection	Overall	Low	High
Low resolution limit (Å)	43.16	43.16	1.81
High resolution limit (Å)	1.76	7.87	1.76
$*R_{merge}$	0.044	0.026	0.625
$**R_{meas}$	0.048	0.028	0.677
$***R_{pim}$	0.018	0.011	0.257
Total observations	82838	981	6250
Total unique observations	12352	148	925
Mean (I)/σ(I)	19.3	62.8	2.7
Completeness	97.1	94.9	97.5
Multiplicity	6.7	6.6	6.8
Unit-cell parameters			
a,b,c (Å)	30.86	48.42	43.66
α, β, γ (o)	90.00	98.68	90.00
Space group	$P2_1$		
Beamline at DLS	I04		
Crystal system	Monoclinic		
Wavelength (Å)	0.92000		

$$*R_{merge} = \frac{\sum_{hkl}\sum_j |I_{hkl,j} - \langle I_{hkl,j}\rangle|}{\sum_{hkl}\sum_j I_{hkl,j}}$$

$$**R_{meas} = \frac{\sum_{hkl}\sqrt{\frac{n}{n-1}}\sum_{j=1}^{n} |I_{hkl,j} - \langle I_{hkl,j}\rangle|}{\sum_{hkl}\sum_j I_{hkl,j}}$$

$$***\mathrm{R}_{pim} = \frac{\sum_{hkl}\sqrt{\frac{1}{n-1}}\sum_{j=1}^{n} |I_{hkl,j} - \langle I_{hkl,j}\rangle|}{\sum_{hkl}\sum_j I_{hkl,J}}$$

4.8.3 Predicted structure of Cj0241-Hr protein

Initially a predicted three-dimensional (3D) structure of Cj0241Hr from the protein sequence was done using the Phyre^2 server to make a model for use in molecular replacement with the data collected from the crystal as explained in chapter (2), section (2.10.5.1). Six template models were obtained from this program based on hemerythrin structures in the protein data bank. These models suggested that the structure of Cj0241-Hr has primarily four alpha helices in a bundle (Figure 4.10).

4.8.4 Molecular replacement (MR) and model building of Cj024-Hr protein

Success of molecular replacement (MR) is based on the availability of a good model, which could be from the same or another species or even the same protein. The models found from Phyre^2 had sequence similarity less than 25% in all cases. Attempts at molecular replacement (MR) using the automated search option in the PHASER program (McCoy *et al.*, 2007) from the CCP4 suite were unsuccessful. However, because the Cj0241-Hr structure was believed to contain two iron ions and these atoms have the anomalous scattering behaviour, phase information was pursued by Single Wavelength Anomalous Diffraction phasing (SAD) as mentioned in chapter (3), section (3.9) to contribute to solving the structure of this protein.

4.8.5 Single wavelength Anomalous Diffraction phasing (SAD)

Several cryo-protected crystals of Cj0241-Hr protein were sent to the DLS for the SAD phasing experiments (Figure 4.11). To decide the correct wavelength for a SAD experiment prior to collecting data, the fluorescence scan was performed around the theoretical Fe^{2+} absorption edge as explained in chapter (3) section (3.9.2).

The peak energy and atomic scattering factors *values* (*f' and f''*) were determined from the fluorescence scan, and the scan showed a good signal for the presence of Fe^{2+} (Figure 4.12) at 7128 eV. The SAD data were collected at the peak wavelength. The crystal data collected on the DLS beamline I04 had a 2.1 Å resolution limit. A total of 900 images were gathered with an oscillation angle of 0.2°, and the exposure period was 1.0 second per image at wavelength 1.73940 Å. The data was processed using xia2/3daii. The crystal belongs also to the monoclinic system and the space group was $P2_1$, with unit-cell parameters a= 31.23 Å, b= 48.51 Å, c= 43.60 Å, a = 90.00°, β = 100.21°, γ = 90. 00°. The data scaling statistics for this structure are shown in Table 4.2.

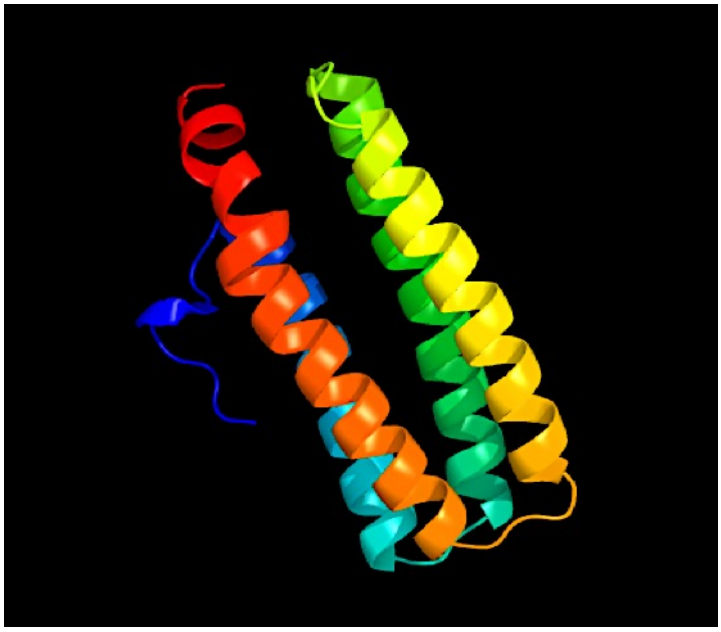


Figure 4.10: **The structure for Cj0241Hr predicted by the Phyre2 server using a model based on PDB entry 2awy.** The model demonstrates that the predicted structure of Cj0241-Hr contains four alpha helices in a bundle. A total of 127 residues (95% of Cj0241 sequence) have been modelled with 100.0% confidence by the single highest scoring template 2awy (PDB code) which had a 24% sequence identity.

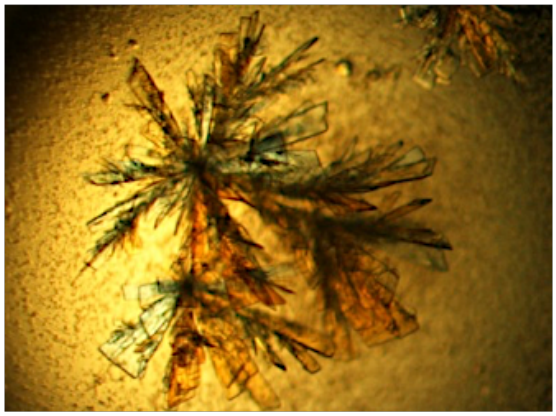


Figure 4.11: Crystal condition for Cj0241-Hr crystal used with SAD technique (2nd dataset). Crystals obtained from at 0.2 M K Formate pH 7.5 in 20%(w/v) PEG 3350, were soaked approx. 30-60 seconds in 20% ethylene glycol as cryo-protectant before sent to the DLS.

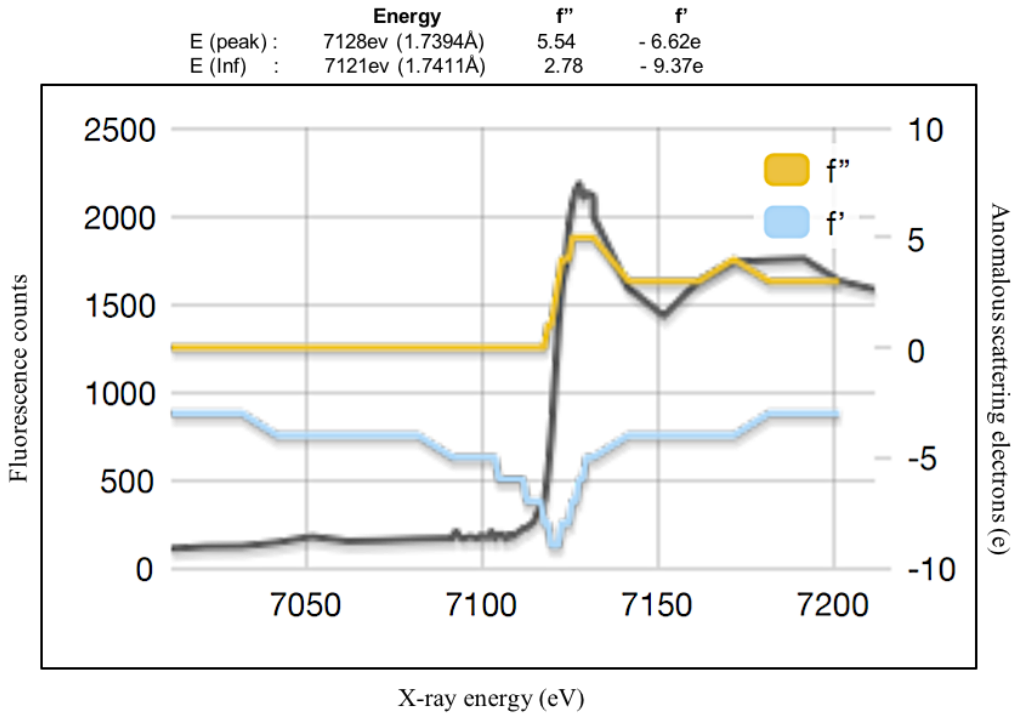


Figure 4.12: The output of a fluorescence Fe edge scan from Cj0241-Hr. The fluorescence was undertaken to recognise the iron atoms absorption edge before adjusting the X-ray wavelength to 1.73940 Å for data gathering. The scan result shows the X-ray absorption spectrum (black curve) and the dispersive scattering (f' – blue curve) and anomalous scattering (f'' – yellow curve) components. The combinations of the two different components make the black curve features and they can be calculated from it.

Table 4.2: The data collection statistics for Cj0241-Hr crystal from SAD processing **(2nd dataset).**

Data collection	Overall	Low	High
Low resolution limit (Å)	32.14	32.14	2.13
High resolution limit (Å)	2.07	9.26	2.07
*R_{merge}	0.048	0.042	0.513
**R_{meas}	0.057	0.050	0.714
***R_{pim}	0.031	0.027	0.495
Total observations	37382	548	441
Total unique observations	6746	96	162
Mean (I)/σ(I)	18.0	34.3	2.3
Completeness	85.5	97.1	28.7
Multiplicity	5.5	5.7	2.7
Unit-cell parameters			
a,b,c (Å)	31.23	48.51	43.60
α, β, γ (o)	90.00	100.21	90.00
Space group	$P2_1$		
Beamline at DLS	I04		
Crystal system	Monoclinic		
Wavelength (Å)	1.73940		

$$*R_{merge} = \frac{\Sigma_{hkl}\Sigma_j|I_{hkl,j} - \langle I_{hkl,j}\rangle|}{\Sigma_{hkl}\Sigma_j I_{hkl,j}}$$

$$**R_{meas} = \frac{\Sigma_{hkl}\sqrt{\frac{n}{n-1}}\Sigma_{j=1}^{n}|I_{hkl,j} - \langle I_{hkl,j}\rangle|}{\Sigma_{hkl}\Sigma_j I_{hkl,j}}$$

$$***R_{pim} = \frac{\Sigma_{hkl}\sqrt{\frac{1}{n-1}}\Sigma_{j=1}^{n}|I_{hkl,j} - \langle I_{hkl,j}\rangle|}{\Sigma_{hkl}\Sigma_j I_{hkl,J}}$$

4.9 Experimental Phasing

To locate the Fe atoms and explain the experimental data obtained from the SAD technique and then calculate phases from this solution, the PHENIX program was used to perform phase improvement, interpretation of the electron density map and initial automated model building. The structure was constructed using AutoSol within the PHENIX program suite (Adams *et al.*, 2010).

4.9.1 Experimental Phasing and Building with PHENIX (Python-based Hierarchical Environment for Integrated Xtallography)

PHENIX is used for calculating experimental phases with SAD data in the AutoSol wizard. The AutoSol started by defining the position of the heavy atoms in the unit cell and estimated the initial phases for the protein. However, AutoSol identified 5 sites, which was unexpected as Cj0241-Hr only had 2 expected Fe positions. Examination of these sites (Figure 4.13A) indicated that sites 1 & 2 had high occupancy of 0.72 & 0.61 respectively, compared with the other three sites which had low occupancy of 0.19, 0.31, 0.42 respectively, and represent an incomplete refinement of the heavy atoms positions which most probably only involve to iron site. A figure of merit (FOM) estimated for these sites was 0.44, which led onto making the first electron density map (Figure 4.13B) designed using only the heavy atom locations.

Having recognised the heavy atom locations and estimated a set of experimental phases, the AutoSol was then utilised to create a model into the density map. This model structure is guided by previous information of the main sequence. AutoSol output showed the Cj0241 model contains four alpha helices. The model building with 104 of the 133 expected resides with 87 assigned sequence and a further 17 positions built as poly-ALA backbone. The electron density map covered most of the structure.

A

Fe site	Occupancy	B-factor	X, Y, Z location
Fe-1	0.72	30.39	3.998, 0.009, 12.183
Fe-2	0.61	32.29	1.686, 45.606, 11.201
Fe-3	0.19	43.45	- 1.152, - 25.972, -11.817
Fe-4	0.31	43.87	- 4.231, - 24.337, - 11.185
Fe-5	0.42	48.57	- 1.713, - 25.845, - 10.879

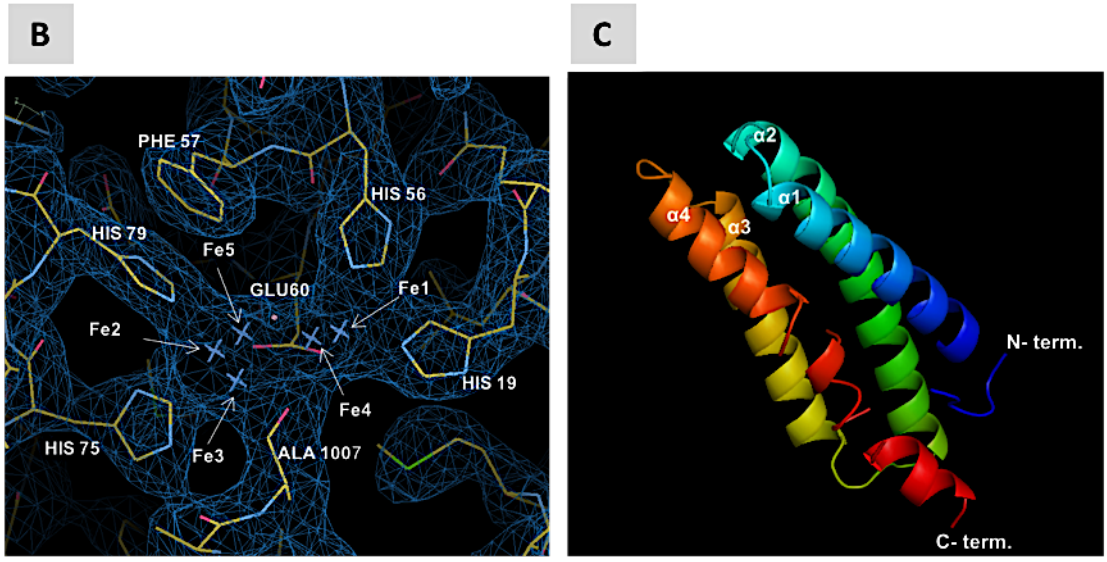


Figure 4.13: Finding heavy atom places and creating an initial electron density map of Cj0241-Hr. The Fe of Cj0241 encodes 2 iron ions sites, but the AutoSol recognises 5. **Chart (A)** facts the temperature factor, occupancy, and real space places for each possible site of heavy atoms. **Panel (B)** displays the electron density map estimated from only the experimental heavy atom phases. **Panel (C)** shows the overall structure of Cj0241-Hr after calculated the phasing by AutoSol.

In addition, some regions within the map appeared without any residues. These regions were completely rebuilt again using the COOT program (Emsley and Cowtan, 2004), and the initial R-factor and R_{free} factor were 0.28, 0.33 respectively (Figure 4.13C).

4.9.2 Model refinement

The Cj0241 model was improved using iterative cycles of rebuilding and refinement using the PHENIX Refine option and COOT program, until no more changes to either the geometry or the R_{factor} (R_{free} and R_{work}) were observed. Every round of refinement was achieved by 3-10 cycles of maximum probability restrained refinement, permitting restrained individual isotropic B-factors. The isotropic B-factors suppose that the thermal oscillation of an atom in a molecule can happen equally in each direction around a mean location. In each cycle, modifications to the atomic model were created when missing atoms were introduced, errors in model conformation were corrected, and solvent molecules were added.

Hemerythrins normally contain two iron ions in the active site and their location was clear during refinement of the Cj0241 structure and corresponded to the locations of peaks 1 & 2 found by PHENIX. Difference electron density maps were calculated (F_o-F_c), and a positive difference density feature was seen between the two iron atoms. An oxygen atom was put in and refined as a μ-oxo bridge. The model of the structure was also improved with water molecules for which there was clear evidence in the map as shown in COOT with a sigma level of the electron density of at least 2 in the 2Fo-Fc map or 3 in Fo-Fc map. In addition, the water molecules were placed into different density features between 2.6 – 3.2 Å (Wlodawer *et al.*, 2008) from the model when the correct type of protein atoms were taken into consideration. Furthermore, other solvents that existed in the cryo-protectant and crystallisation buffers such as the polyethylene glycol

and glycerol were added when they matched with the electron density map. Additionally, the residues of some amino acids that were not fitted automatically with the map were repaired and located to be within the electron density. As a result of these refinements, the model fit to the electron density map permitted the construction of 119 residues of the 133 expected residues with R-factor and R_{free} factor were 0.26, 0.31 respectively (Figure 4.14).

4.9.3 Building Cj0241-Hr structure with higher resolution data (model 1)

Even though the electron density map was improved relative to the initial map, the structure still was not complete and there was a lot of negative electron density in different places around the map, particularly around the active site. In addition, it was not possible to produce more improvement in the refinement values. So, to get a better model for Cj0241-Hr, the structure was refined against the earlier 1.8 Å higher resolution data (1^{st} dataset) with edited FreeR flags. The result of this allowed more residues to be built and improved quality in the maps; this structure is referred to as **"model 1"**. More refinements were done as above followed by manual rebuilding within the new part of the improved map. The new map allowed to build 121 residues of the 133 expected residues, with not much decrease in the refinement value of R-factor and R_{free} factor which were 0.25, 0.30 respectively. However, there were still the problems with the negative electron density maps in different places around the structure of Cj0241-Hr that were found.

4.9.4 Reduction of model bias by omit- map technique

Reduction in the bias of the maps from the Cj024-Hr imperfect protein model was investigated using the omit map technique via the OMIT option within the PHENIX suite.

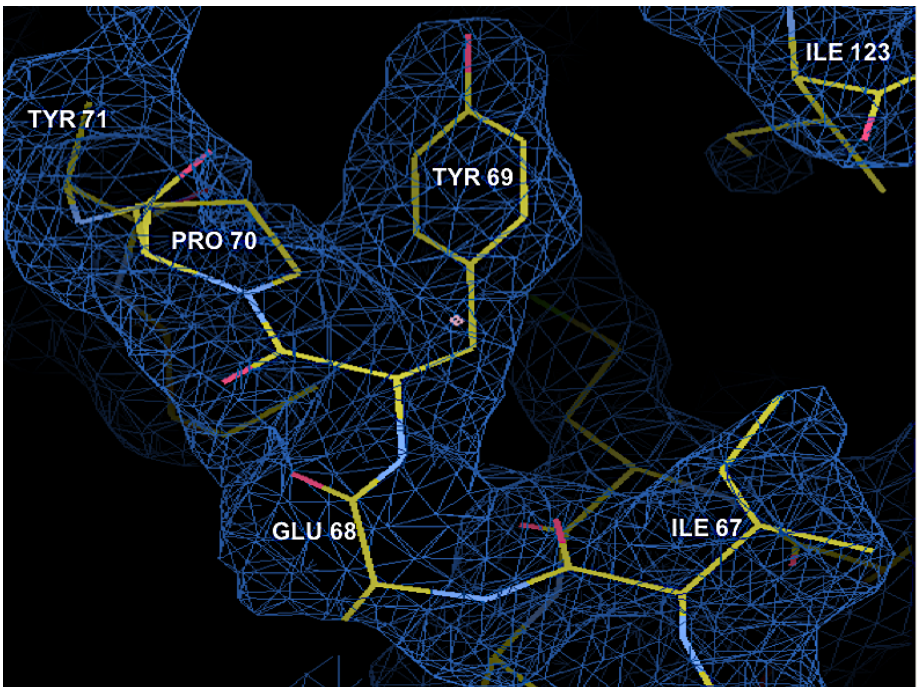


Figure 4.14: A typical part of the 2Fo-Fc and Fo-Fc electron density maps calculated using the peak wavelength data (2nd dataset) and the refined phases from the model. It shows the fit of the model to the map, which permitted to construct 119 residues of the 133 expected residues with R-factor and R_{free} factor were 0.26, 0.31 respectively.

The goal of this method was to make an improved map to which a better model could be fitted. Model bias could be decreased by refinement of the partial structures, which were built by omission of spherical regions around selected parts of the model. In addition, during this technique, the method of simulated annealing (often known a molecular dynamic refinement) was used in refinement. Each section of 10 residues from the structure was omitted in turn and a short-simulated annealing refinement was made to reduce model bias from the missing region. The resultant fit of the model to the map with the same previous 121 residues of the 133 expected residues and the R_{free} and R_{work} were now 0.30 and 0.24 respectively. However, there were still the problems with the negative electron density maps in different places around the structure of Cj0241-Hr that were found.

4.9.5 Negative electron density peaks around the two iron ions

The hemerythrin proteins as mentioned in chapter (1) section (1.8.3) can be classified according to their oxidation condition into three different stages: deoxy-hemerythrin (Fe^{2+}-OH-Fe^{2+}), oxy- hemerythrin (Fe^{3+}-O-Fe^{3+}-OOH) and met-hemerythrin (Fe^{3+}-OH-Fe^{3+}) (Kao *et al.*, 2008). The current structure of Cj0241-Hr is expected to be in the met-form. This stage is considered normal for finding the hemerythrin protein. The Cj0241 structure would not contain an oxygen molecule because the iron is expected to undergo auto-oxidation under aerobic conditions during the purification and crystallisation. Although a good model of the Cj0241-Hr structure was achieved, the refinement of the Fe atoms was not fully stable and showed massive negative electron density features around the location of two iron ions (Figure 4.15). To solve this unlikely electron density map, further adjustment of the occupancy to 0.5 relieved some of the problem of this negative density features (Figure 4.16).

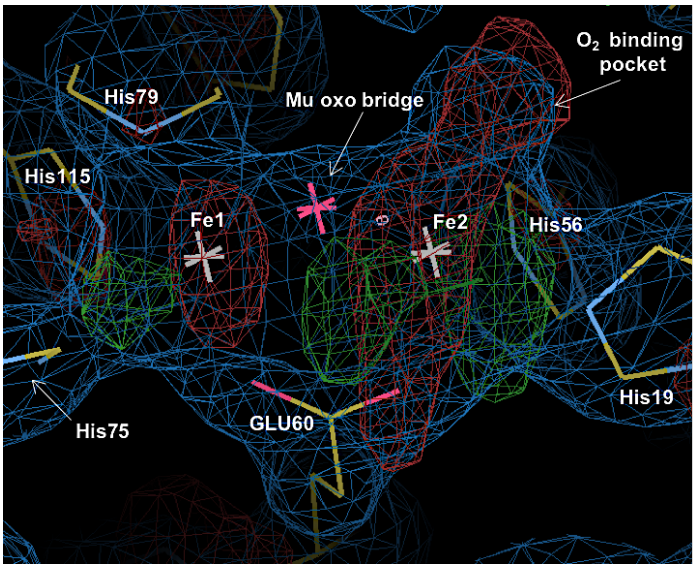


Figure 4.15 : The active site of Cj0241-Hr strucure shows massive negative map density. The active site of Cj0241-Hr strucure shows massive negative map density Fo-Fc electron density map features around the two iron ions.

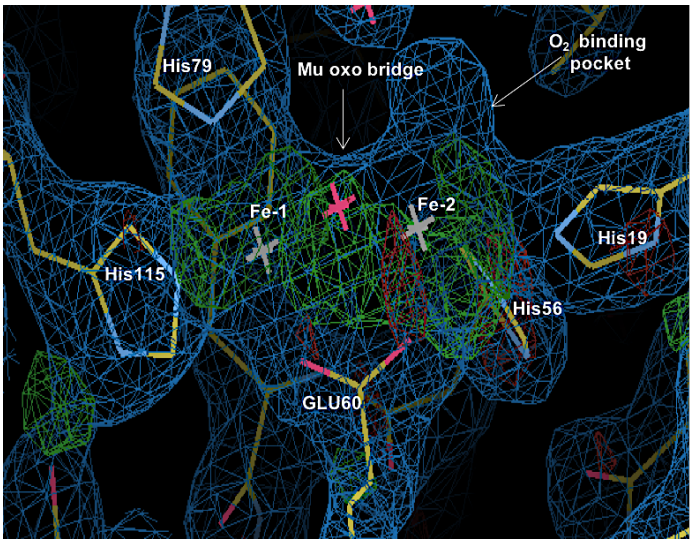


Figure 4.16: The active site of Cj0241-Hr structure shows massive of positive map density. The active site of Cj0241-Hr strucure shows the negative density map partially disappears after decreasing the occupancy of Fe ions from 1.0 to 0.5. However, the map still needs more refinement because there is a lot of positive electron density map features which appeared.

However, a lot of positive electron density map features appeared, which implies that the occupancy of Fe change from 1.0 to 0.5 was too large.

4.9.6 B-factor values

During the refinement, the B-factors were not stable throughout the structure. They showed unusual distributions where for example, the CE1 atom of the side chain phenyl ring of Phe57 had a very elevated B-factor compared with the other atoms in the ring, which did not make any sense (Figure 4.17). Incorrect positions of atoms in the Cj0241-Hr structure will tend to cause B-factors to be higher or lower than expected. To solve this problem, deletion, rebuilding and refinement of the recovered sensible values in some residues but not in all cases. Therefore, because of the problems with the refinement of the Fe-O-Fe site and a general instability noted in some side chain atom B-factors, refinement was instated using the REFMAC5 program (Murshudov *et al.*, 1997).

4.9.7 Refinement Cj0241-Hr structure with REFMAC5

REFMAC5 refines atomics model by regulating the model parameters to find the model that greatest explains the experimental data. The last refined coordinate file from the PHENIX server was input into REFMAC5 and used in full coordinates, TLS and restrained isotropic B-factor refinement. The electron density map around the two iron ions became clear. Also, the structure showed a clear electron density feature in the possible site for the O_2 binding pocket, which was modeled as a water molecule after the O_2 is released, in agreement with the electron density shape and appropriate occupancy for a water molecule in refinement (Figure 4.18). Furthermore, the B-factors of the atoms in each amino acid in the pdb file exhibited more consistent values. TLS refers to the translation-libration-screw rotation of a group of atoms, either an entire molecule or a domain (Korostelev and Noller, 2007).

A

```
ATOM    392  N   PHE A  57      16.627  18.847  30.457  1.00 36.81
ATOM    393  CA  PHE A  57      17.881  18.080  30.481  1.00 36.88
ATOM    394  C   PHE A  57      17.942  17.094  29.330  1.00 54.32
ATOM    395  O   PHE A  57      19.003  16.823  28.790  1.00 40.32
ATOM    396  CB  PHE A  57      18.063  17.297  31.785  1.00 34.69
ATOM    397  CG  PHE A  57      18.294  18.161  33.002  1.00 43.24
ATOM    398  CD1 PHE A  57      18.534  19.518  32.888  1.00 45.46
ATOM    399  CD2 PHE A  57      18.281  17.591  34.275  1.00 52.59
ATOM    400  CE1 PHE A  57      18.743  20.307  34.038  1.00104.28
ATOM    401  CE2 PHE A  57      18.496  18.362  35.423  1.00 51.06
ATOM    402  CZ  PHE A  57      18.720  19.719  35.315  1.00 43.44
```

B

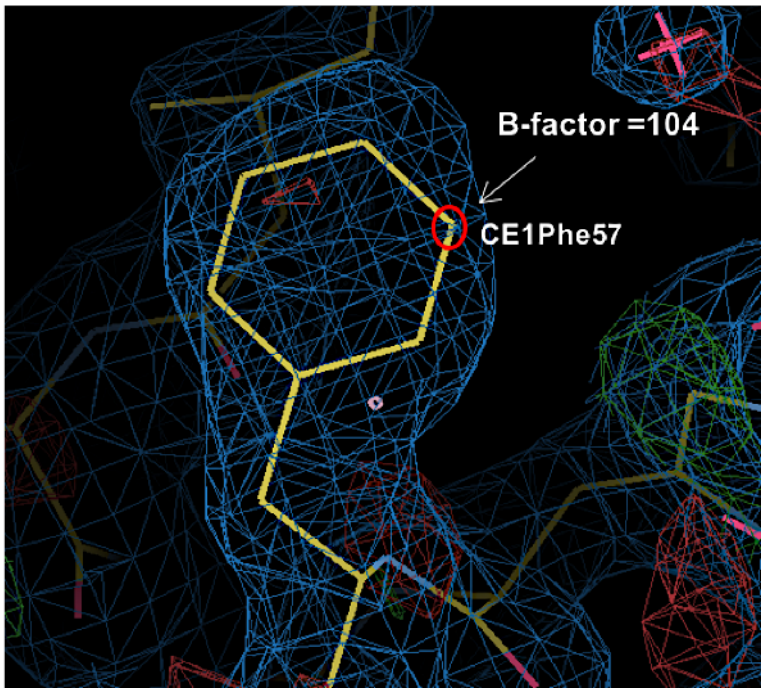


Figure 4.17: Unstable B-factor refinement of the Cj0241Hr model. The CE1 atom in the sidechain of Phe 57 has an unusually high value. **(A)** The values from the coordinates of Phe57 refined with PHENIX. The elevated B–factor is circled. **(B)** The fit of the Phe57 sidechain to the 2Fo-Fc electron density map shows good coverage for all the atoms.

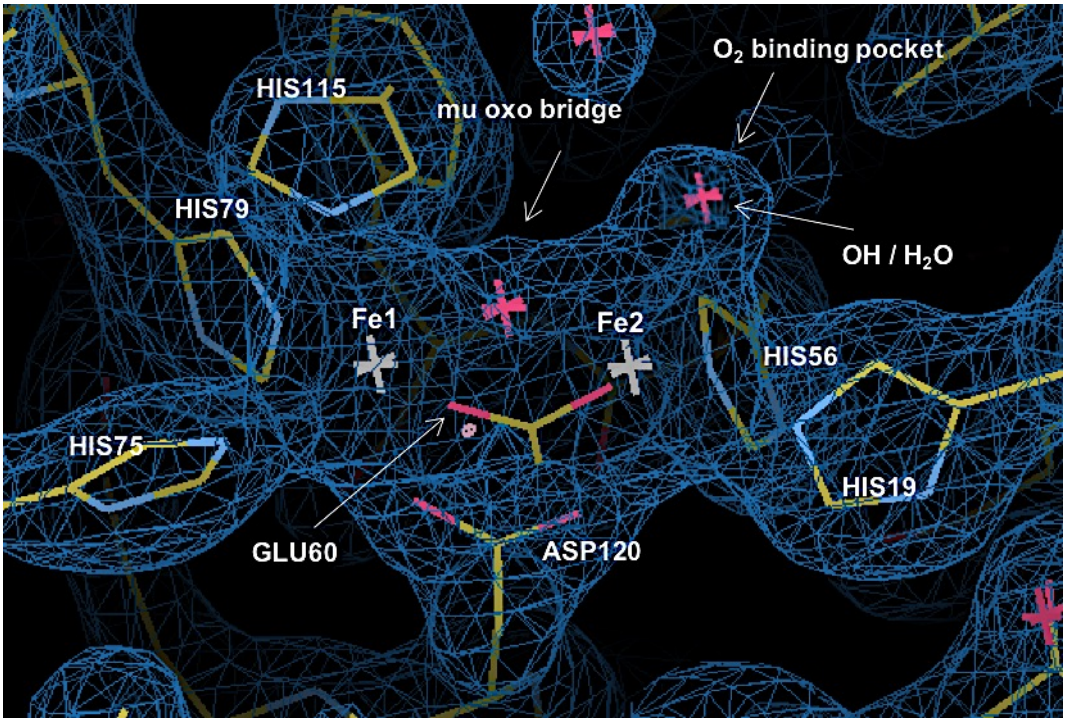


Figure 4.18: The active site of the Cj0241-Hr structure from FEFMAC5. The negative and positive electron density map around the two iron ions was not present after refinement by REFMAC5 with using TLS refinement.

Using TLS refinement helped to improve the model and resulted in a reduction in R-factor (R_{free} and R_{work}) when compared with an identical refinement of the same model. Furthermore, a general improvement in the fit of the residues with an apparent reduction in noise peak was observed (Figure 4.19). The refinement, rebuilding and assessment using REFMAC5 and COOT were carried on until there was no structure improvement possible and no more water molecules could be added. The refinement statistics of Cj0241-Hr presented a reduction in R-factor and R_{free} to 0.21 and 0.27, respectively.

4.9.8 Cj0241 Model Quality Validation

The crystal structure of Cj0241-Hr protein shows that it has contained one molecule within the asymmetric unit as expected from Matthews coefficient estimates. There was no electron density for the initial seven amino acids at the N-terminus and the final five residues at the C-terminus of the molecule plus six-histidine tag and the leucine and glutamate the two linker residues between the natural C- terminus and the tag. The structure building was validated by examining it with the structure analysis tools within COOT (Emsley and Cowtan, 2004) and the MOLPROBITY server (Chen *et al.*, 2010) to check the rotamer , clash score and Ramachandran plots. Both programs showed that all side chain and main chain parameters appeared within the predictable range for the resolution of the data, and all residues dropped within the permitted areas of the Ramachandran plot (Figure 4.20).

4.9.9 The Cj0241 hemerythrin model 1

The final model of Cj0241-Hr structure (model 1) was composed of 121 residues, monomeric protein, formed into four α helices in a bundle and a di-iron centre with a μ oxo-bridging ligand. The di-iron centre includes an additional hydroxyl group or water

A

```
ATOM    384  N   PHE A  57      16.587  18.802  30.519  1.00 29.34
ATOM    385  CA  PHE A  57      17.845  18.040  30.484  1.00 26.04
ATOM    386  CB  PHE A  57      18.015  17.269  31.797  1.00 25.27
ATOM    387  CG  PHE A  57      18.306  18.131  32.994  1.00 25.63
ATOM    388  CD1 PHE A  57      18.599  19.455  32.876  1.00 24.15
ATOM    389  CE1 PHE A  57      18.856  20.235  34.022  1.00 28.52
ATOM    390  CZ  PHE A  57      18.887  19.653  35.305  1.00 25.76
ATOM    391  CE2 PHE A  57      18.606  18.303  35.428  1.00 28.94
ATOM    392  CD2 PHE A  57      18.276  17.565  34.293  1.00 27.88
ATOM    393  C   PHE A  57      17.887  17.052  29.317  1.00 27.96
ATOM    394  O   PHE A  57      18.936  16.836  28.715  1.00 25.12
```

B

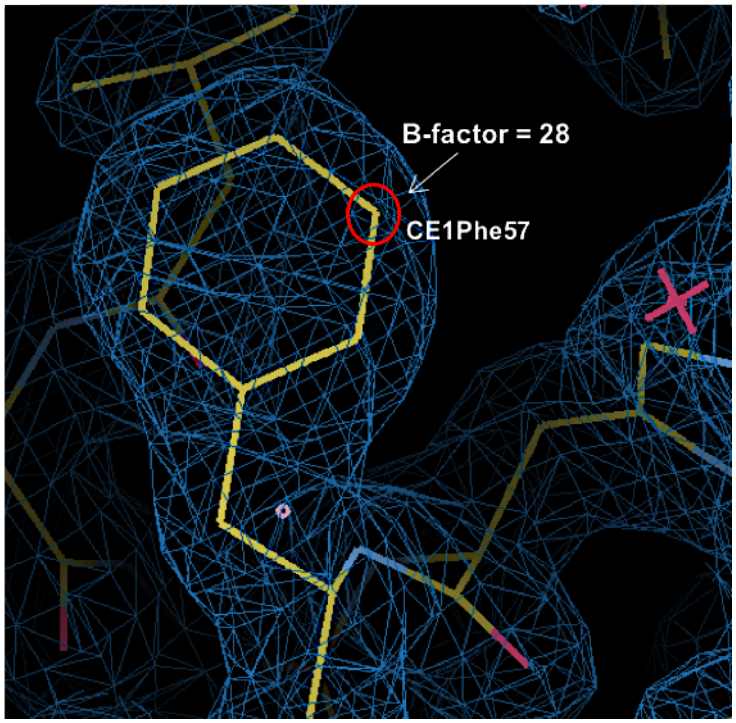


Figure 4.19: Stable B-factor refinement of the Cj0241Hr model from REFMAC5. Stable B-factor refinement and sensible & consistent values for the B-factors of the atoms. The CE1 atom in the side chain of Phe 57 has a low value compared with the value from PHNEX. **(A)** The values from the coordinates of Phe57 refined with REFMAC5. **(B)** The fit of the Phe57 side chain to the 2Fo-Fc electron density map shows good coverage for all the atoms.

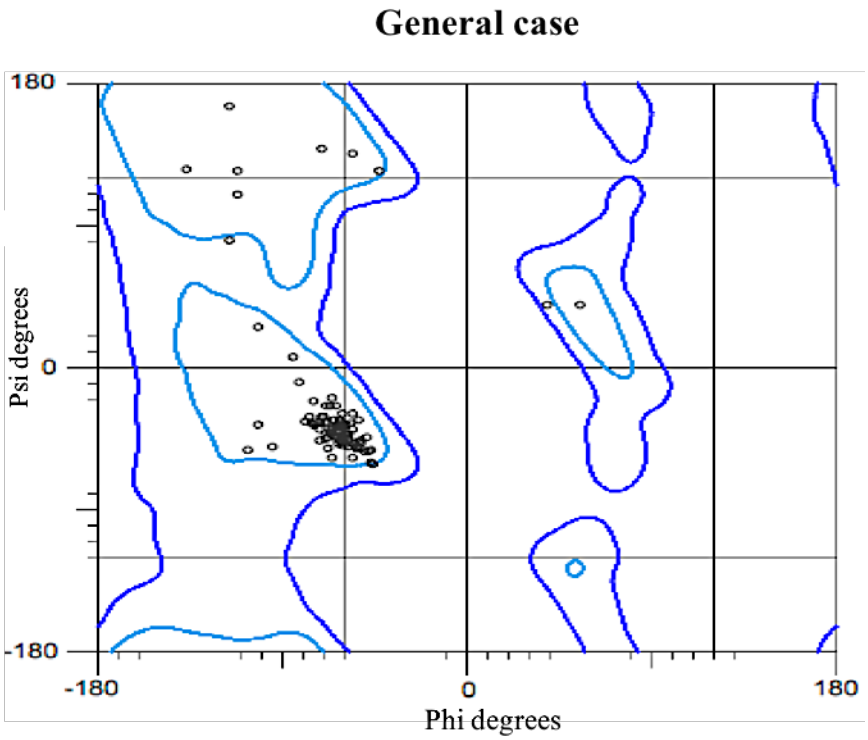


Figure 4.20: The Ramachandran plot for the final model of Cj0241-Hr (Model 1). The Ramachandran analysis of the Cj0241-Hr model shows 95.80% of 121 residues have been shown in favorable locations with no amino acids occupying the prohibited region disallowed because of steric hindrance. The figure was produced using Molprobity servers http://molprobity.biochem.duke.edu

molecule (OH^-/H_2O) attached to one of the iron atoms in what is believed to be the O_2 binding pocket. The di-iron centre was also coordinated by two types of functional group: the carboxylate groups in the side chains of ASP120 and GLU60 which act as bridging ligands to both iron atoms FE1 and FE2; the imidazole groups of HIS19, HIS56 and HIS75 which act as ligands to Fe-1 and HIS79 and HIS115 which act as ligands to Fe-2 (Figure 4.21). There are also 56 water molecules and 4 ethylene glycol molecules from cryo-protectant built it into the model. The final refinement statistics for the Cj241-Hr structure are written in Table 4.3, and it is called the **"model 1".**

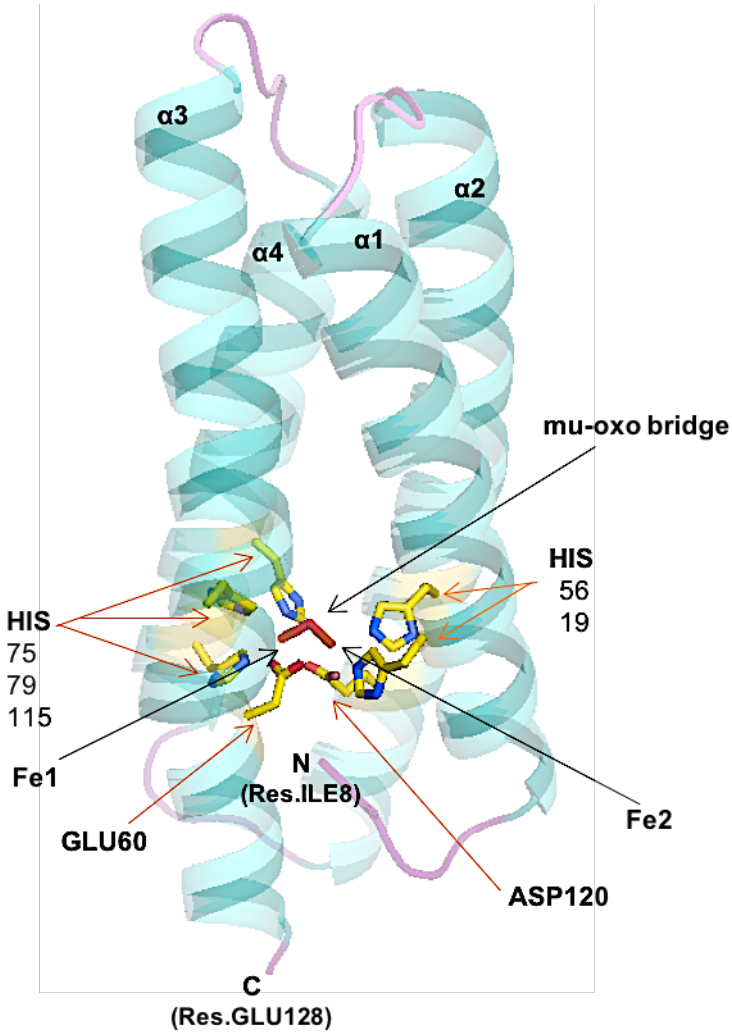


Figure 4.21: A carton representation of the Cj0241-Hr structure. The final model of Cj0241 protein consists of four parallel alpha helices labeled 1-4, and a μ-oxo bridged di-iron centre (orange) coordinated by the side chains of ASP 120 & GLU 60, which bridge the two iron irons. The HIS19, HIS56 & HIS75 coordinate to Fe1, and HIS79 & HIS115 coordinate Fe2.

Table 4.3: Data Refinement statistics for Cj0241-Hr structure model 1.

Data Refinement Statistics	Cj0241-Hr model 1
Resolution (Å)	2.1
Protein molecules in asymmetric unite	1
Number of residues	121
Number of H_2O molecules	56
Number of atoms	1001
Number of reflections	11737
Average B values (Å^2)	
Whole chain	40
Main chain	37
Side chain	42
R.M.S Deviation:	
Bond lengths (Å)	0.016
Bond angles (°)	1.569
Ramachandran plot:	
Favoured regions (%)	95.80
Outliers (%)	0.00
R_{work}	0.21
R_{free}	0.27
Active centre distances (Å)	
Fe1-His75	2.44
Fe1-His79	2.29
Fe1-His115	2.42
Fe1-Glu60	2.07
Fe1-Asp120	2.08
Fe1-O	1.92
Fe2-His19	2.42
Fe2-His56	2.27
Fe2-Glu60	2.11
Fe2-Asp120	2.01
Fe2-O	1.95
Fe1-Fe2	3.39
Fe1-O-Fe2(deg)	28.50
Fe2-H_2O/OH	2.07

4.10 High resolution structure of Cj0241-Hr (dataset 3)

The first structure was believed to be a met-hemerythrin on the basis that the oxygen has been released from the structure during the redox cycling. To get some insight into the active site of the Cj0241-Hr structure and the predicted binding between Fe2 and an O_2 molecule in oxygen binding pocket, attempts were made to soak Cj0241-Hr crystals (Figure 4.22) in sodium azide (NaN_3) as explained previously in chapter (2) section (2.10.6). The crystal data was collected on the DLS at the beamline I04 to a resolution limit of 1.1 Å. A total number of 1900 images were collected with an oscillation angle of 0.10° per image, and the exposure time was 0.1 second per image at wavelength 0.9749 Å. The diffraction data was integrated, indexed and scaled using the Xia2 software package. From the dataset in Table (4.4), the results show that the completeness is reasonable (98.7%) with a R_{merge} 0.035, R_{pim} (0.029) and R_{meas} of 0.048, these are good indicators, and added to the multiplicity which was greater than 2. The mean (I) / σ (I) value was close to 2 at the outer shell. In general, these results show sensible indicators for the data quality (Weiss, 2001; Wlodawer *et al.*, 2008). The crystal belongs also to the monoclinic system and the space group was $P2_1$, with unit-cell parameters a = 28.45 Å, b = 46.21 Å, c = 48.38 Å, a = 90.00°, β = 93.00°, γ = 90.00°, which was not isomorphous to the previous data set. The data scaling statistics for this structure are shown in Table (4.4).

4.10.1 Solution and refinement of the data

So, because the crystal was not isomorphous to the previous form, the structure model was solved by the molecular replacement technique using the PHASER program from CCP4 with the first model of Cj0241Hr as a search model.

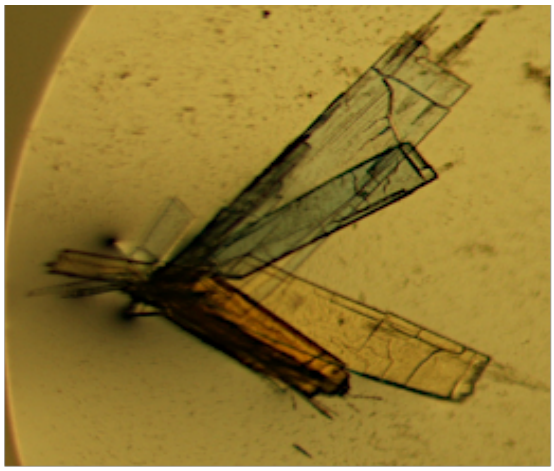


Figure 4.22: Crystal condition for the Cj0241-Hr with NaN_3 . Crystals obtained from 0.2M Ammonium nitrate pH6.3 and 20% PEG3350, were soaked for around 30-60 seconds in 10mM mM NaN_3 before addition of 20% ethylene glycol as cryo-protectant and then sent to the DLS.

Table 4.4: The data collection statistics for Cj0241-Hr crystal (**3rd dataset**).

Data collection	Overall	Low	High
Low resolution limit (Å)	14.19	14.19	1.12
High resolution limit (Å)	1.09	4.87	1.09
$*R_{merge}$	0.035	0.028	0.524
$**R_{meas}$	0.054	0.043	0.727
$***R_{pim}$	0.029	0.023	0.413
Total observations	169521	1809	9534
Total unique	51518	574	3554
Mean (I)/σ(I)	12.8	34.1	1.9
Completeness	98.7	93.4	92.7
Multiplicity	3.3	3.2	2.7
Unit-cell parameters			
a,b,c (Å)	28.45	46.21	48.38
α, β, γ (°)	90.00	93.83	90.00
Space group	$P2_1$		
Beamline at DLS	I04		
Crystal system	Monoclinic		
Wavelength (Å)	0.97949		
Number of molecules in the asymmetric unit	1		

$$*R_{merge} = \frac{\sum_{hkl}\sum_j |I_{hkl,j} - \langle I_{hkl,j}\rangle|}{\sum_{hkl}\sum_j I_{hkl,j}}$$

$$**R_{meas} = \frac{\sum_{hkl}\sqrt{\frac{n}{n-1}}\sum_{j=1}^{n} |I_{hkl,j} - \langle I_{hkl,j}\rangle|}{\sum_{hkl}\sum_j I_{hkl,j}}$$

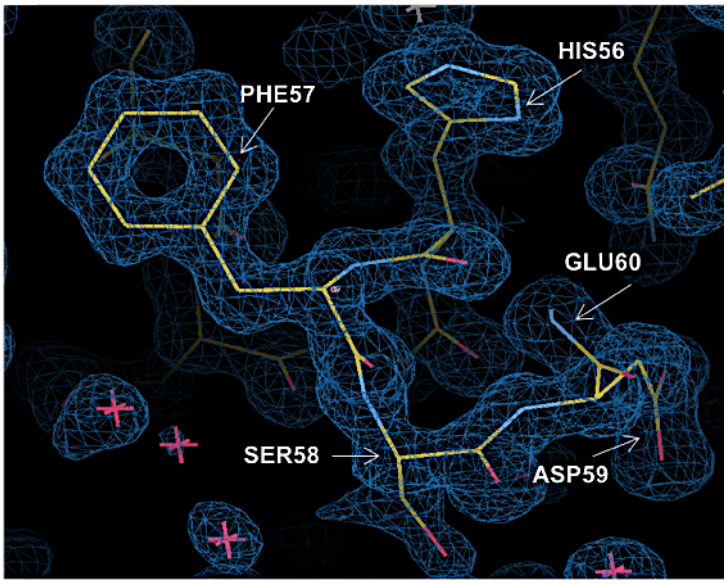
$$***R_{pim} = \frac{\sum_{hkl}\sqrt{\frac{1}{n-1}}\sum_{j=1}^{n} |I_{hkl,j} - \langle I_{hkl,j}\rangle|}{\sum_{hkl}\sum_j I_{hkl,J}}$$

In addition, the Fe-O-Fe unit was removed from the search model before performing the molecular replacement calculations. The COOT program was used to inspect the results and a $2F_O$-F_c map was estimated from these data in REFMAC.

The Fe-O-Fe unit was positioned manually once again into the observed density in the map. Initially individual isotropic B-factors were refined, and eventually anisotropic refinement of individual B-factors was achieved, and the R-factor declined to 0.21 and R_{free} dropped to 0.23. Unfortunately, the azide ligand to Fe-2 was not observed in the oxygen-binding pocket as had been hoped and seen with other hemerythrin structures (Isaza *et al.*, 2006). However, a much higher data set to a resolution 1.1 Å was obtained from this experiment and permitted a more detailed examination of the structure.

The overall fitting of the first model and new data (3rd data set) was good with clear electron density map around nearly all the residues of the structure (Figure 4.23), and this structure is referred to as **"model 2"**. Furthermore, the $2F_O$-F_C electron density map in the region consistent to an exogenous ligand pocket to Fe2 was formed as containing a water/hydroxide molecule as explained previously in section (4.9.7) (Figure 4.24). Attempts to position an N^-_3 did not give a good fit to the density and resulted in large negative density peaks after refinement (Figure 4.25). No other exogenous electron density was seen in the binding pocket and the refinement by REFMAC5 showed a decrease in both the R-factor to 0.17 and R_{free} to 0.19. Future options, however, to obtain a structure with azide bound to Cj0241-Hr might include increasing the time period to soak the crystals, incubating the crystals with higher concentrations of sodium azide and co-crystallisation with azide.

A



B

ATOM	456	N	PHE	A	57	-19.458	4.912	58.974	1.00 9.86
ATOM	457	CA	PHE	A	57	-18.797	5.554	60.109	1.00 10.07
ATOM	458	CB	PHE	A	57	-18.983	4.770	61.427	1.00 11.10
ATOM	459	CG	PHE	A	57	-20.355	4.818	61.985	1.00 10.21
ATOM	460	CD1	PHE	A	57	-21.376	5.608	61.515	1.00 9.92
ATOM	461	CE1	PHE	A	57	-22.614	5.590	62.107	1.00 10.90
ATOM	462	CZ	PHE	A	57	-22.833	4.831	63.232	1.00 11.81
ATOM	463	CE2	PHE	A	57	-21.831	4.059	63.710	1.00 12.15
ATOM	464	CD2	PHE	A	57	-20.606	4.047	63.122	1.00 12.44
ATOM	465	C	PHE	A	57	-17.319	5.805	59.810	1.00 11.05
ATOM	466	O	PHE	A	57	-16.754	6.839	60.156	1.00 11.29

Figure 4.23: The electron density map from the crystals soaked with azide. Panel (A) Molecular replacement using the new data from the azide soaked crystal (3rd data set) gave a good electron density map around nearly all residues of the structure of Cj0241-Hr. **Panel (B)** The B-factors were lower and stable when comparing with the structure from the low-resolution data set. Compare the results with those in Figures 17 & 19.

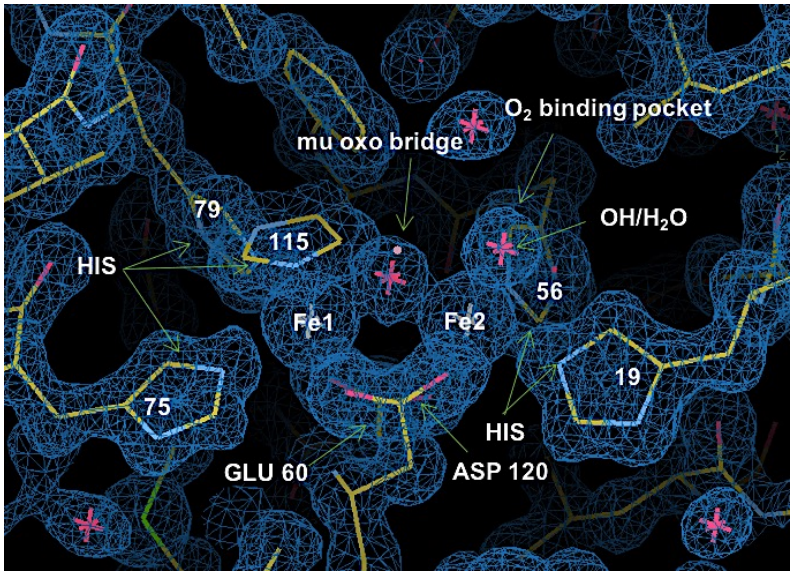


Figure 4.24: The active site of Cj0241 and the oxygen-binding pocket site. The $2F_O$-F_C electron density map of the di-iron site and exogenous ligand pocket of Cj0241 at 1.1 Å resolution revealed either water or a hydroxo ligand occupies an exogenous site on Fe2.

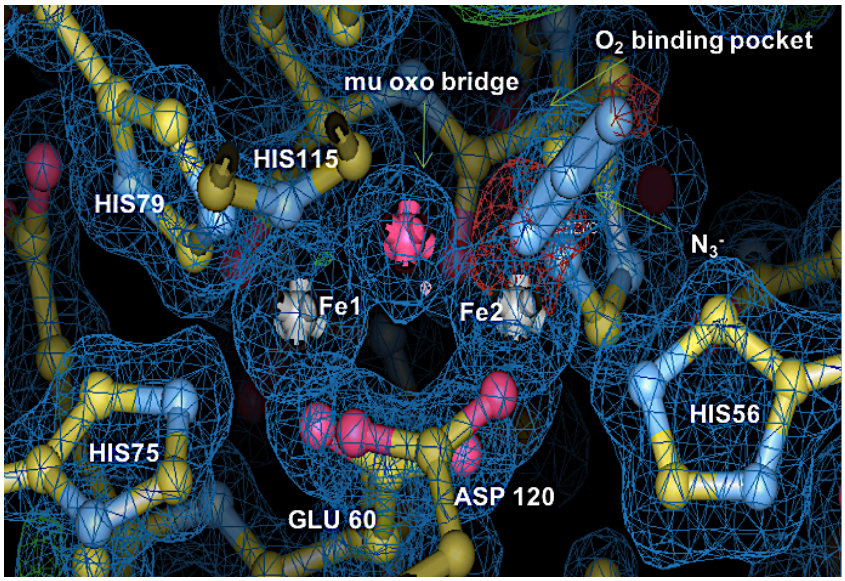


Figure 4.25: A carton representation the active site of Cj024-Hr with Azide form. Azide ion structure is too large to fit with the electron density map and refinement did not work.

4.10.2 Additional residues in high-resolution map

The electron density map calculated from the high-resolution data revealed new residues in the structure, which were not found in the first model. The first model of Cj0241-Hr structure missed 7 residues in the N-terminal and also 6 residues in C-terminal. However, the new data collection revealed electron density for these residues and gave a structure from the residue 2 to residue 133 (Figure 4.26). In addition, Cj0241-Hr construct has an artificial six histidine tag at the C-terminus, and the new data clearly exhibited these six histidines following the leucine and glutamate the two linker residues between the natural C- terminus and the tag (Figure 4.27).

4.11 Disorder Structure

Although the new collected data gave a high resolution, 1.1 Å, the new structure had weaker electron density in the map in the loop between helices 3 & 4 for residues Glu93, Ala94, Lys95, Phe96, Val97, Asn98 and Ile99. This portion of the electron density map in the lower resolution model (model 1) was clear and continuous around all these residues (Figure 4.28). All attempts at refinement and improvement of the model in this region were unsuccessful. These differences could be associated with packing interactions or arrangement in the two structures.

4.12 Cj0241 Validation and final model (model 2)

The crystal structure of Cj0241-Hr protein showed that it has one molecule in the asymmetric unit as predictable from the Matthews coefficient calculation numbers. In addition, the high resolution structure was validated by examining it with the structure analysis tools within COOT (Emsley and Cowtan, 2004) and Molprobity server (Chen *et al.*, 2010).

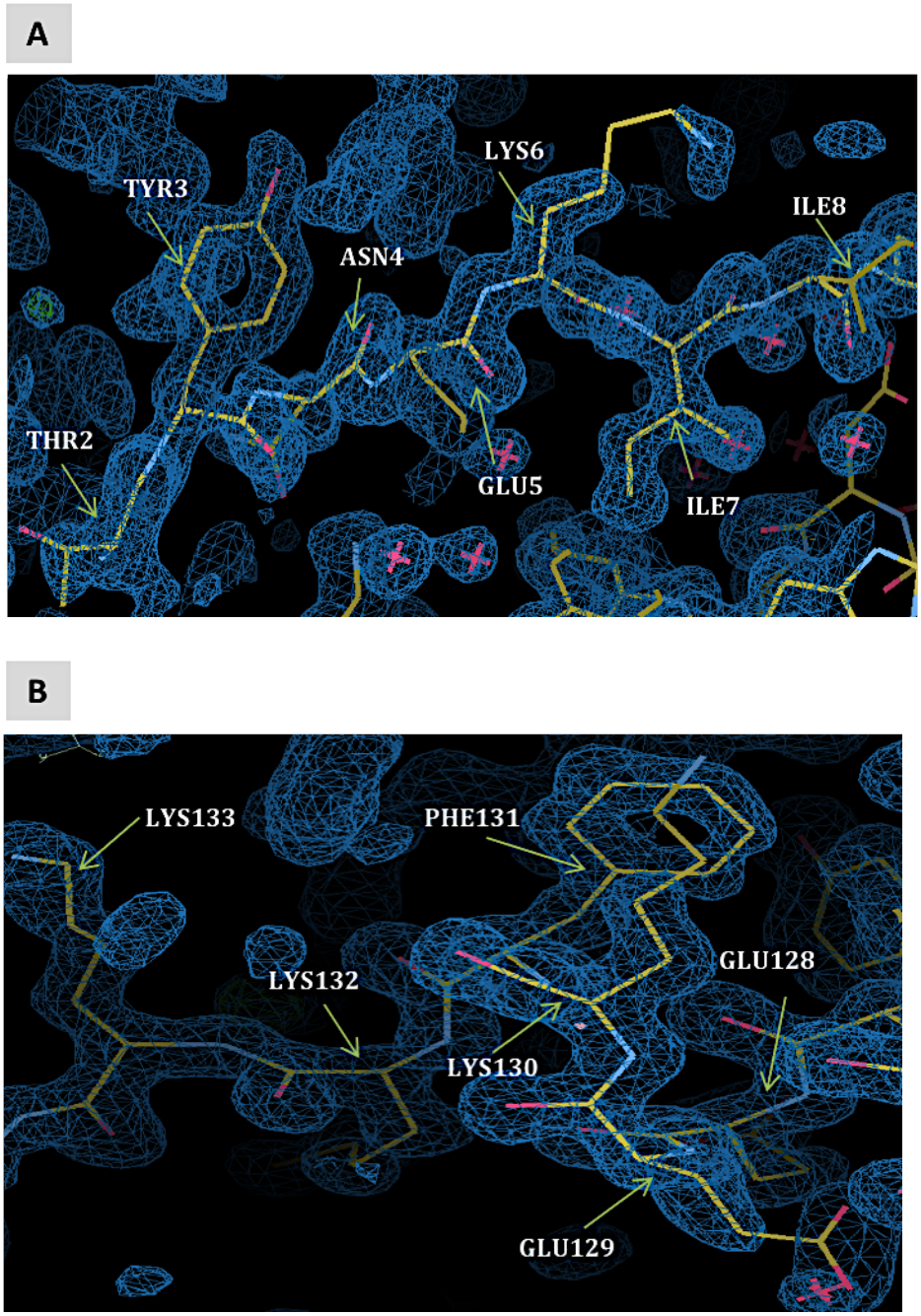


Figure 4.26: The electron density maps from the crystals were soaked with azide improved to build new residues. The new data allowed the building of new residues in the structure either in the N-terminal or C-terminal regions that were not found in the first model. **Panel (A)** N-terminal region shows new residues built from 2 to 7. **Panel (B)** C-terminal region shows new residues built from 129 to 133.

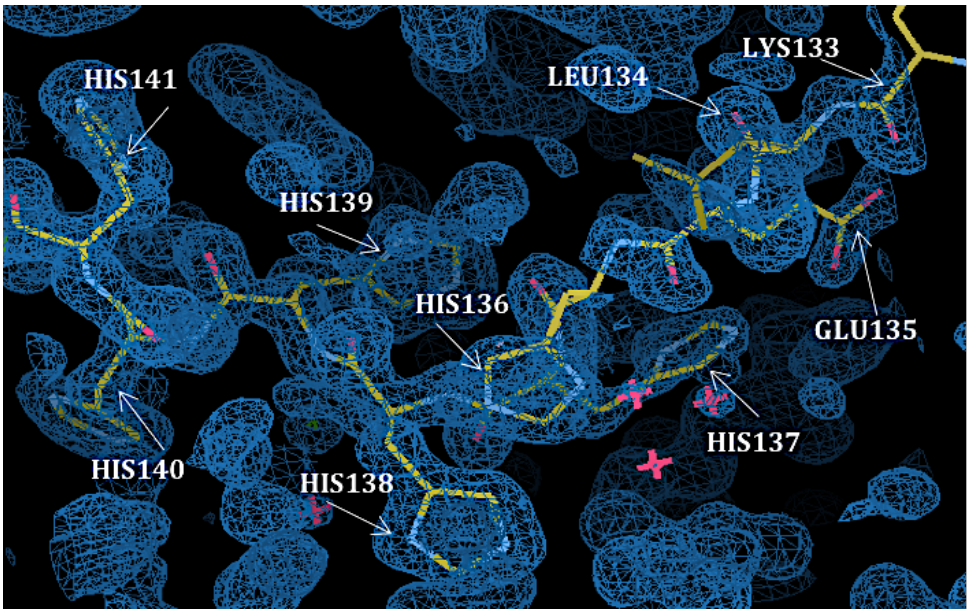


Figure 4.27: Evidence for purification tag. The map from the high-resolution data revealed a six-histidine affinity tag following the linker residues Leu134 and Glu135.

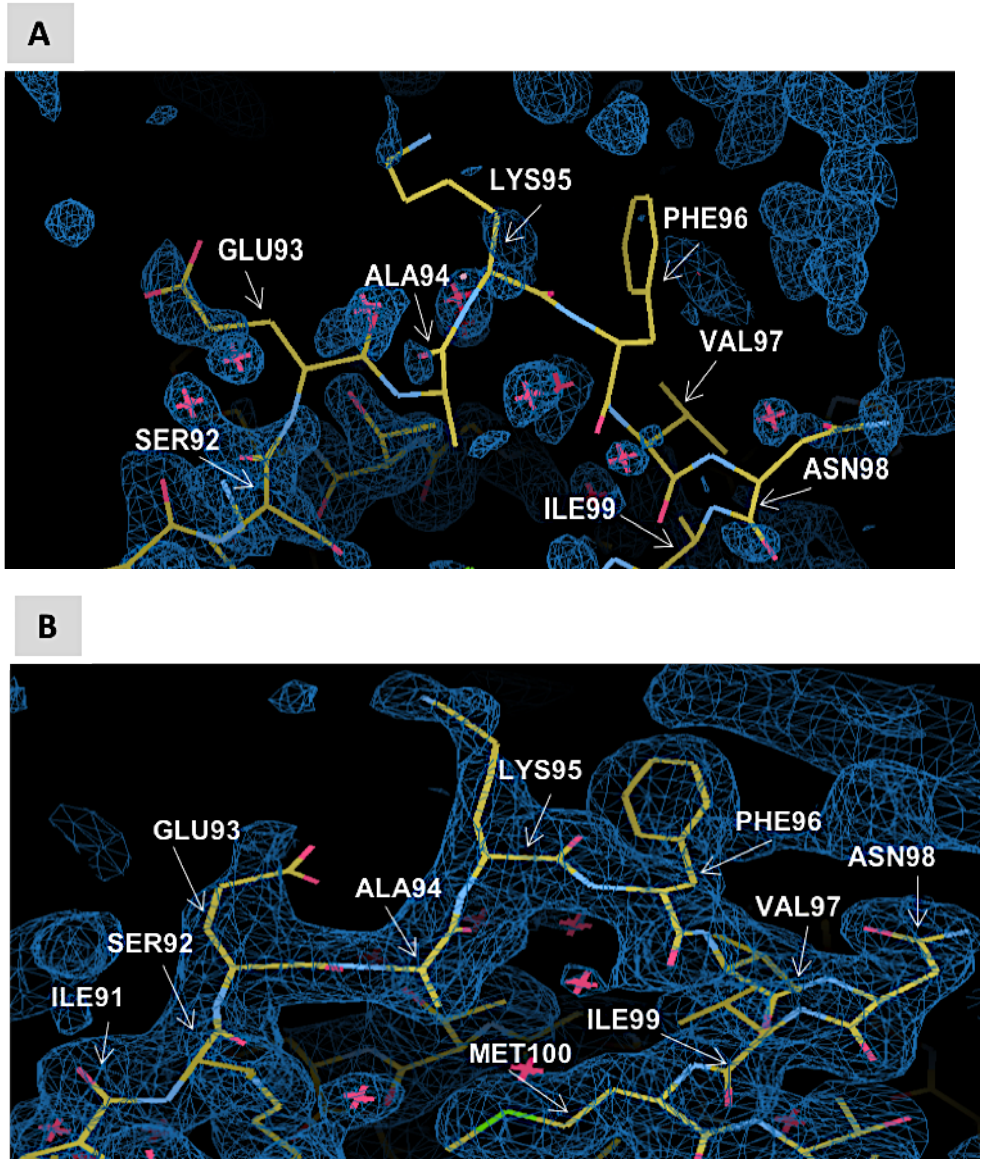


Figure 4.28: The electron density map for the loop linking α-helices 3 & 4. (A) Cj0241-Hr structure from the high-resolution data exhibited a weak electron density map between residues 93 - 99, which represents a disordering in the loop. **(B)** The map for the same portion of the structure from the low-resolution data showed residues covered with the map.

Both programs showed that all side chain and main chain parameters appeared within the predictable range for the resolution of the data, and all residues dropped within the permitted areas of the Ramachandran plot, which showed that the structure in general is of a good quality (Figure 4.29).

The final model of Cj024-Hr (Figure 4.30) from the high-resolution dataset is composed of a four-alpha helix bundle. The molecule contains α1: 2–36, α2: 40–66, α3: 71–90 and α4: 100– 133 as explained previously on model 1. There was an electron density map covering all these residues of the structure including the 2 linkers and six histidine tagged. However, a part of the structure between 93-99 residues looked disorderedly most likely because of the crystal packing condition. The μ-oxo-bridging ligand of the di-iron centre is coordinated with two types of groups as described before in chapter (4) section (4.9.9). The result of refinement Cj0241-Hr model 2 provided R_{work} of 0.17 and R_{free} of 0.19. The root means square deviation (RMSD) from the ideal for bond lengths is 0.015 Å and for bond angles is 1.731°, which lie within acceptable range (Wlodawer *et al.*, 2008). In addition, the overall average B-factor was 16.6 $Å^2$ for the protein (more details in section (4.13.2)). The structure was also supported with 184 water molecules and 3 ethylene glycol molecules from cryo-protectant built it into the model. The water molecules support is clear evidence in the map as shown in COOT with a sigma level on the electron density of 2 in 2Fo-Fc map or 3 in Fo-Fc map. The final refinement statistics for Cj241-Hr structure from the high-resolution data (model 2) are presented in Table (4.5). The table also, contains the measurement of the active site of the protein structure of Cj0241.

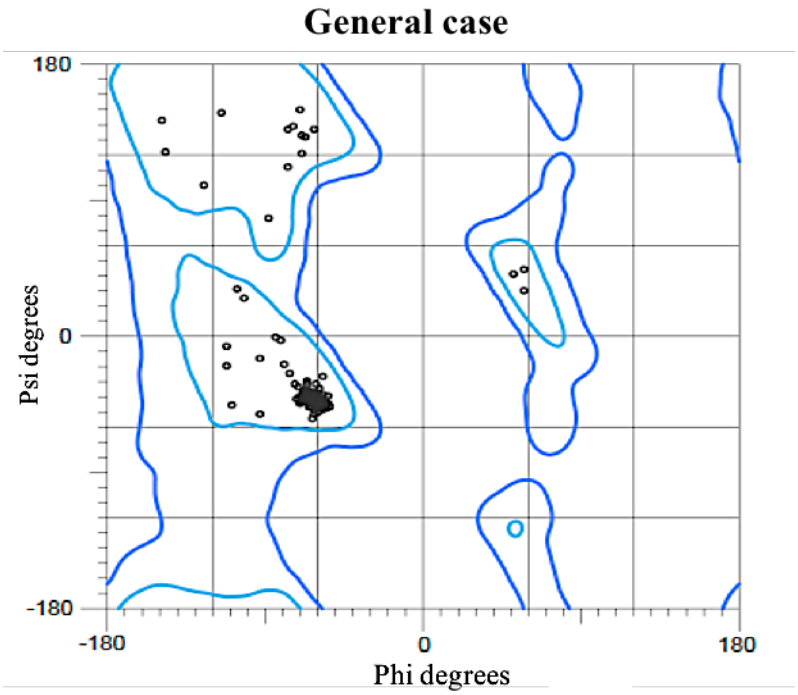


Figure 4.29: Ramachandran plot for the final model of Cj0241-Hr. Ramachandran analysis of the Cj0241-Hr model show 98.45% of 140 residues (132 from Cj0241 sequence plus 2 linker residues plus a 6-histidine tag) have been modeled in favorable locations with no residues occupying the prohibited region defined by steric hindrance. The figure was produced using the Molprobity server http://molprobity.biochem.duke.edu .

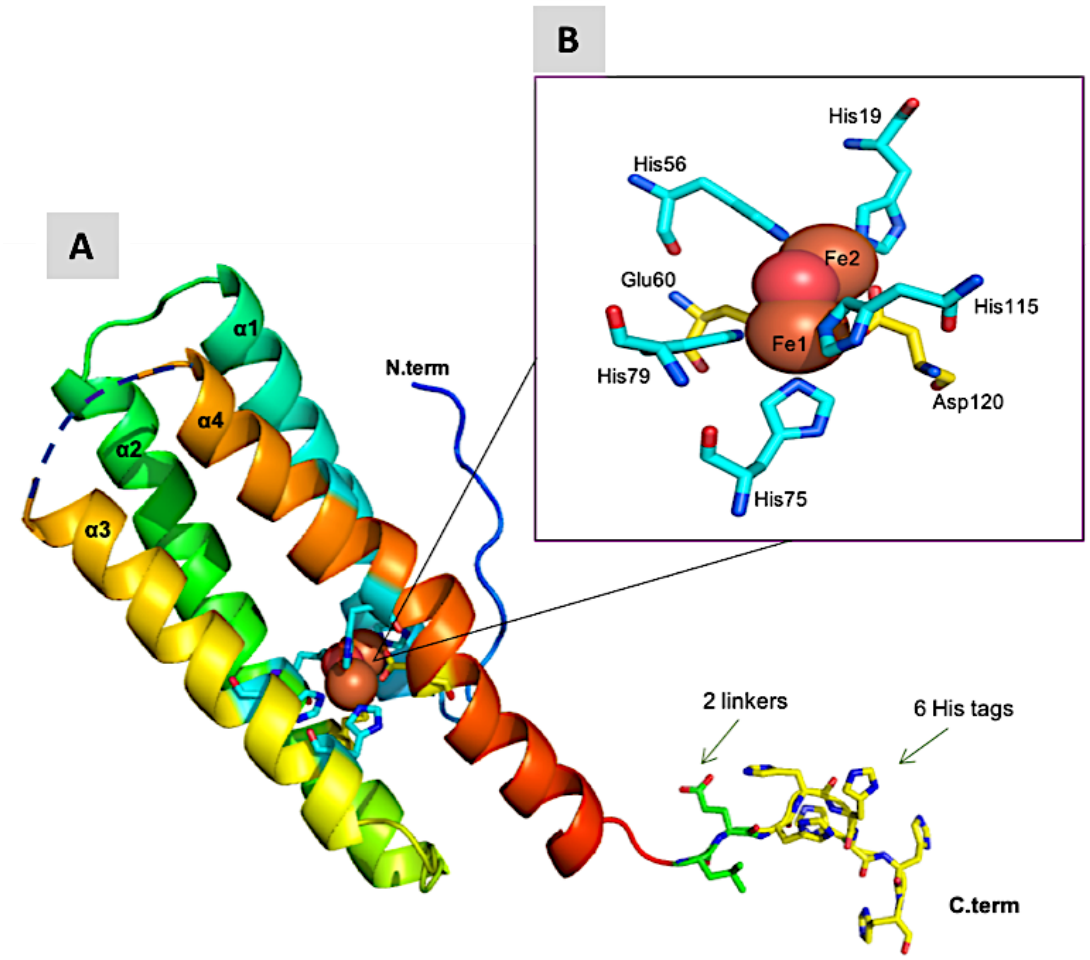


Figure 4.30: A cartoon representation of the CJ0241-Hr structure with helices represented as coils. Panel (A) The structure consists of four parallel alpha helices (rainbow coloured N- to C-terminus), and a μ-oxo bridging ligand of the di-iron centre (orange). The two iron ions are coordinated by the carboxylate groups from the side chains of Asp120 & Glu60 (yellow carbon atoms) and the imidazole groups of five histidine residues (light blue carbon atoms) where three histidines (His19, His56 & His75) ligand to Fe1 and the other two (His79 & His115) ligand to Fe2. The dotted line indicates that is where the disordered loop occurs. The N-terminal residues and the poly-histidine tag at the C-terminus can be seen extending away from the helical bundle where they make additional contacts with other molecules in the crystal lattice. **Panel (B)** The di-iron centre is shown in close-up on the right of the figure.

Table 4.5: Refinement statistics for Cj0241-Hr structure modelled from the high-resolution dataset (model 2).

Data Refinement Statistics	Cj0241-Hr model 2
Resolution (Å)	1.1
Protein molecules in asymmetric unite	1
Number of residues	132 plus 2 linkers plus 6 histidine tagged
Number of H_2O molecules	184
Number of atoms	1234
Number of reflections	49079
Average B values (Å^2)	
Whole chain	16
Main chain	14
Side chain	18
R.M.S Deviation:	
Bond lengths (Å)	0.015
Bond angles (°)	1.731
Ramachandran plot:	
Favoured regions (%)	98.45
Outliers (%)	0.00
R_{work}	0.17
R_{free}	0.19
Active centre distances (Å) :	
Fe1-His75	2.17
Fe1-His79	2.17
Fe1-His115	2.16
Fe1-Glu60	2.09
Fe1-Asp120	2.08
Fe1-O	1.39
Fe2-His19	2.17
Fe2-His56	2.17
Fe2-Glu60	2.17
Fe2-Asp120	2.03
Fe2-O	1.92
Fe1-Fe2	3.40
Fe2-H_2O/OH	2.06

4.13 Cj0241-Hr protein 3D - structural analysis

4.13.1 Distortions in Cj0241-Hr protein helices

Alpha helices in protein structure are not always found in perfect helical conformation, and they often show structural distortions (Geetha, 1996). Given the shape of the 4 alpha helices in Cj0241-Hr structure, it is very clear from looking at Fig 30A that helix 4 is curved or distorted in the region between residues His115 and Asp120 (both of which coordinate the Fe atoms), so that the helix does not run straight like helices 1 and 3. However, this feature is conserved with some other hemerythrin structures such as McHr from *Methylococcus capsulatus* [Bath] and DcrH-Hr from *Desulfovibrio vulgaris*. Helix 2 is also slightly curved, but not the same as in the other species such as McHr and DcrH-Hr. This feature is possibly to ensure that His56 and Glu60 can make best contacts with the iron atoms. The bends in the helices that are close to the active site have a significant functional role to play in their interaction with the adjacent atoms (Geetha, 1996). The distortions are mostly found due to side chain packing or the interaction between the solvent and the side chains of the helices because the exposed C=O and N-H groups tend to form H-bonds to solvent (Geetha, 1996). This gives increase to a bend in the helix axis. Another possibility for distortion is induced by certain specific amino acids such as the presence of Proline (Pro) residues, which induce the distortions in the way of the helix axis. This is due to the proline being unable to shape a regular alpha-helix because of steric hindrance occurring from its cyclic side chain that also blocks the main chain N atom and avoids it making a hydrogen bond (Barlow and Thornton, 1988). However, this last possibility is not found with Cj0241-Hr structure, but does appear with the McHr (Chen *et al.*, 2015) structure due to the existence of a Pro within the bend part (Figure 4.31).

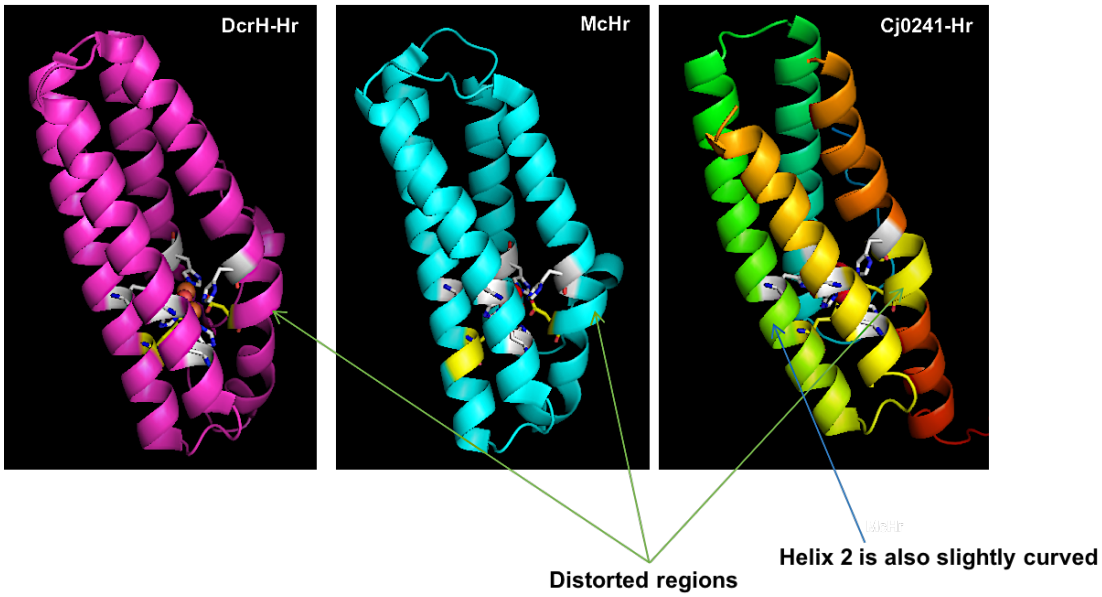


Figure 4.31: Distortions in Cj0241-Hr protein helices. Helix 4 is curved or distorted in the region between residues His115 and Asp120, so that the helix does not run straight like helices 1 and 3. This feature is also found with some other structures such as Dcrh-Hr and McHr. Furthermore, helix 2 is also slightly curved compared with other species.

4.13.2 B-value distribution

The thermal motion is an oscillation of an atom around its mean location because of the temperature and is described by its B-factor. Crystal disorder implies the atom does not occupy the same location in each unit cell or in each asymmetric unit (Rhodes, 2006). The atomic scattering factor is usually disturbed by the difference in atomic location caused by crystal disorder and/or the real motion of temperature variations. An increased value of the B-factor indicates more uncertainty in the locations of atoms. Temperature can produce atomic fluctuations/mobility that disturb the B-factors, which are also called "Debye- Waller factors" or "temperature-factors". From the table (4.5), it is clear that the main chain of Cj0241Hr has a lower distribution of B factor compared with the side chains due to the atoms in side-chains have more freedom and flexible movement than the atoms within the main chain.

Atoms with the low B-factors show a well-ordered portion of the structure. Atoms with high B-factors correspond to a portion of the structure that indicates it has a higher level of flexibility and mobility of individual atoms and side chains (Cesarini *et al.*, 2012). However, it is suggested that the main contributor to the B-values derives from the crystal disorder rather than the thermal motion of the atoms (Rhodes, 2006)

The B-factors of the atoms in the loop between helices 3&4 (which has seven missing residues) are significantly higher (flexibility) than the overall B-factors of the other parts of the structure. Therefore, this could be due to the amino acid residues at the end of the loop has a low number of contacts with other amino acids (Radivojac *et al.*, 2004). In addition, the B factor distributions within the 4 helices have been shown to be lower when comparing with the other parts. In general, the "**Cj0241-Hr model 2"** structure has lower B factors comparing with "**Cj0241-Hr model 1"** structure as shown in table (4.3), which has higher flexibility in several residues of the structure as can be

seen in Figure (4.32). In addition, in model 2 has the His136, which is one of the group of His-tag at the end of the C-terminus, the electron density map in this part is somewhat weaker. Therefore, even when this residue was refined it had also high B-values.

4.13.3 Electrostatic surface potential of Cj0241

The electrostatic surface potential of Cj0241 was calculated using Pymol (Figure 33). Analysis of the charge distribution on the surface shows that the Cj0241 has a mixed distribution of negatively and positively charged surface residues, as well as neutral residues.

Positively charged arginine and lysine residues are equally distributed on the surface and are balanced by the negative charges from glutamates and aspartate residues.

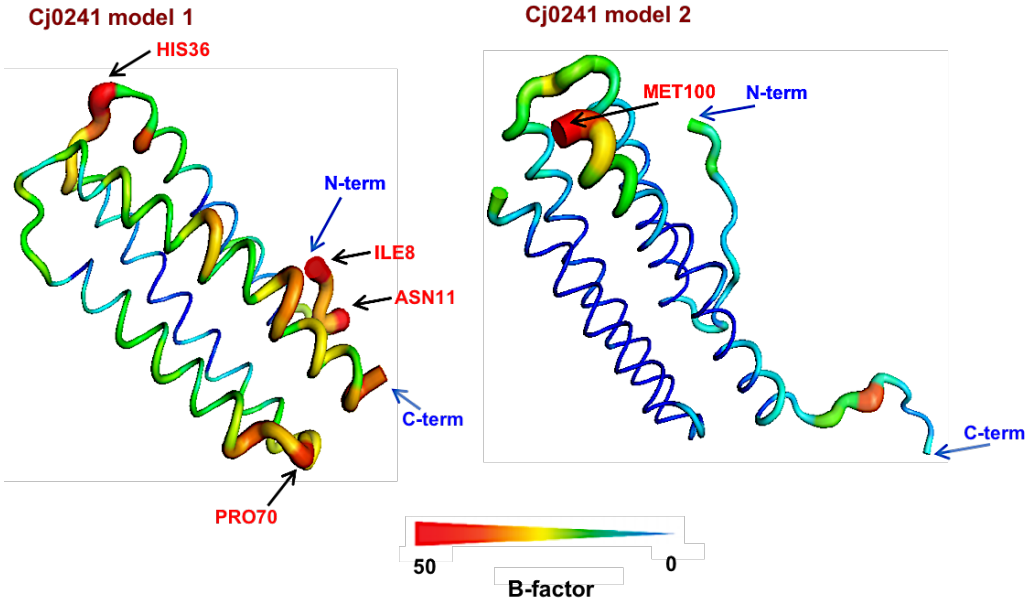


Figure 4.32: Location of the amino acids with the highest B-factor on the Cj0241-Hr structures model 1 & model 2. By using " B -factors putty" option in the Pymol program (DeLano & Lam, 2005), the regions with the highest B-factors and therefore probably the most flexible sites are exhibited in orange-red, while the more rigid ones are displayed in green-blue (Cesarini *et al.*, 2012).

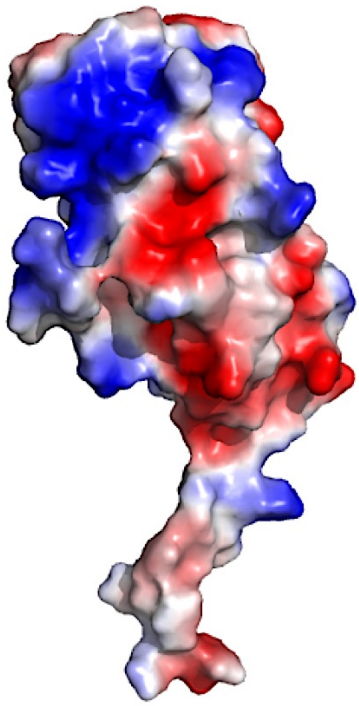


Figure 4.33: The electrostatic potential map of the positively and negatively charged residues on the surface of Cj0241-Hr protein. The electrostatic potential maps in two different orientations. Negatively charged regions (from aspartate/glutamate residues) are shown in red and positively charged regions (from lysine/arginine residues) are shown in blue. Diagram created using Pymol.

4.13.4 The contribution of water molecules to Cj0241-Hr protein structure

Protein hydration is very essential for three dimensional structure, conformation and activity of proteins (Tompa *et al.*, 2015). The Cj0241-Hr structure from the high-resolution data is surrounded by 184 water molecules. Their distribution around the Cj0241Hr protein model is shown in Figure (4.34). Water molecules were added as described in section (4.9.2) using COOT program. Using COOT, a dummy water molecule is rolled around the surface of a protein which helps analyse the hydration shells around the protein structure. In this higher resolution structure, they were also seen inside the protein molecule structure (Figure 4.35 C). Two water molecules were found down in the middle of the structure of the molecule. One of them is adjacent to the di-iron centre and the other one is further down a channel in the middle of the molecule. This supports the argument that the Cj024-Hr protein has a water tunnel to give access to the di-iron site which is detailed further in section (4.13.5).

The variations of the protein surface are controlled by the network of H_2O molecules (Hospital *et al.*, 2017). Some of these water molecules contribute to the structure and stability of the protein by bridging through hydrogen bonding (Patel *et al.*, 2014). These linking water molecules should be considered while understanding the structure of the hemerythrin molecule. Some of these water molecules interact directly with the protein surface are dynamically restricted and are ordered while additional water molecules are hydrogen bonded to these and stay dynamically active (Zhong *et al.*, 2011).

In the Cj0241-Hr model 1 structure, most of the water molecules seem to have direct contacts with the protein structure. However, in the model 2 of Cj0241-Hr structure can be seen more water molecules evenly spread all over the surface and show there are clear signs of a different grouping/bunching together of water molecules.

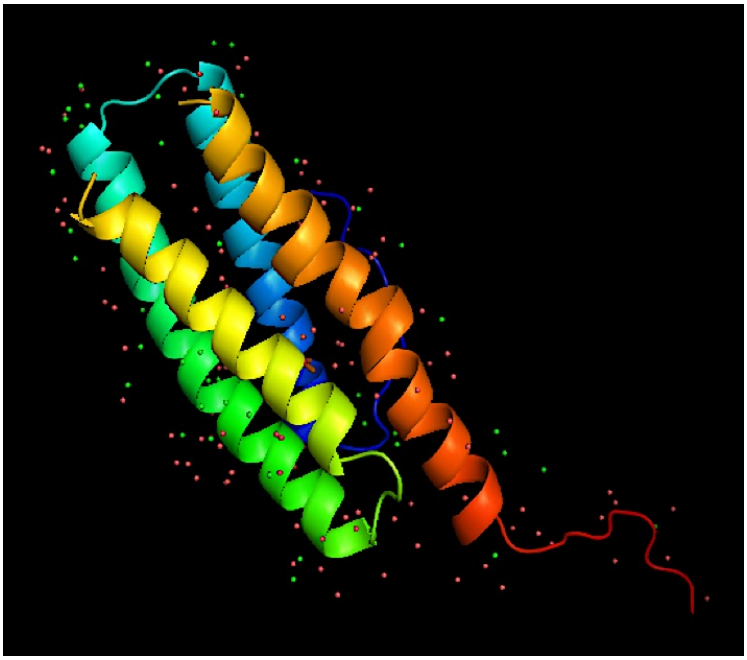


Figure 4.34: The water molecules surrounding the Cj0241-Hr protein structure. COOT analysis of the water molecules around the surface of the Cj0241-Hr protein showing those directly associated with the surface (first hydration shell, red colour), and those which are not interacting directly (second hydration shell, green colour).

In Cj0241-Hr model 2 there are 72% of the water molecules in direct contact with the structure and this is called "th**e first hydration shell**". However, there are another 28% of the water molecules that do not interact directly with the protein surface, but just with other water molecules and together these are called the "**second hydration shell**".

4.13.5 The Cj0241 met-Hr crystal structure exhibits the presence of the putative water tunnel in Cj0241-Hr

Analysis of the crystal structure of Cj0241-Hr using the CAVER program (Chovancova *et al.*, 2012) shows possible access routes via two tunnels to the di-iron centre that might be linked to its autoxidation. The tunnels are oriented approximately in parallel to the long axis of the α-helix bundle. They have separate exits onto the surface and, overlap near to the di-iron centre and provide the most likely route for oxygen molecules to reach the metals and for the hydrogen peroxide products of the reaction to leave. In Figure (4.35 A) the tunnel 1 (pink mesh) is lined via the side chains of the residue SER26, LEU29, SER30, ASN33, ASN105 and VAL108.

The Tunnel 2 (blue mesh) is lined by LEU46, LEU49, ILE86, ILE89, SER92, MET100, THR101, GLU102 and VAL107. However, the two tunnels overlap together and the region in common is lined by HIS19, HIS56, HIS79, HIS115, ILE53, PHE57, ILE82, PHE111 and ILE112, and is terminated by the residues HIS19, HIS56, HIS79, HIS115 and THR116 at the iron centre in the core of Cj0241-Hr protein (Figure 4.35 B). In Figure (4.35 C) a water molecule can be seen within the tunnel and this supports the argument that the water tunnels in Cj0241 are the entrance of to reach the di-iron centre. There is a difference between the two tunnels for length, where the blue tunnel is approx. 20Å long and the pink tunnel is approx. 17Å long, and the tunnels finish approx. 4Å away

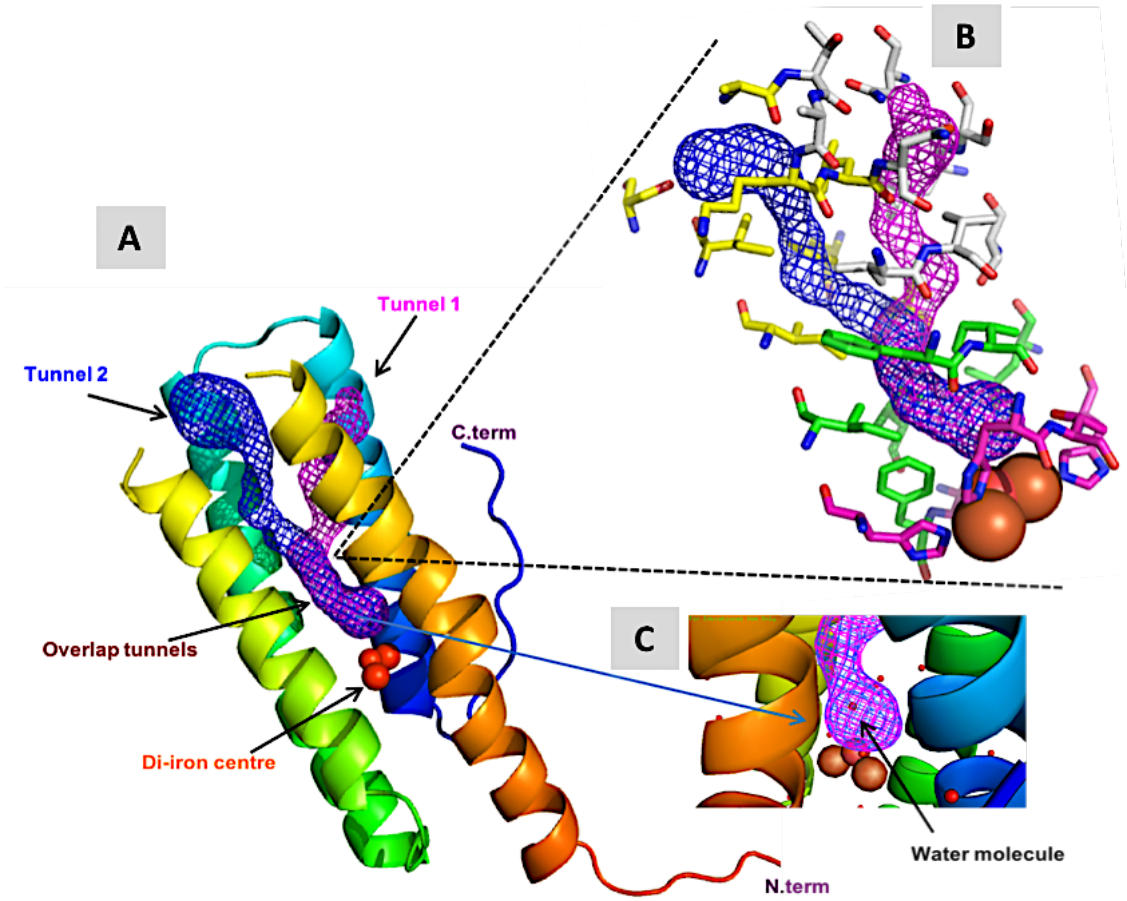


Figure 4.35: Cj0241Hr structure reveals two water tunnels. Panel (A) The crystal structure of the Cj0241-Hr showing a four-helix bundle, a non-heme di-iron centre (red) and the analysis by CAVER 3.0 showed possible water tunnels (deep blue & pink mesh) oriented parallel with the long axis of the four–helix bundle. **Panel (B)** the amino acids around the water tunnel; yellow colour amino acids around tunnel 1, white around tunnel 2, green around the overlap of the two tunnels and pink colour amino acids terminate the water tunnel. **Panel (C)** the water molecule within the water tunnel supports the argument that the water tunnel plays a role in the transfer of solutes to and from the di-iron centre of Cj0241-Hr protein.

from the iron centre. Also, the channels narrow in the middle and return to their widest approx. 2Å diameter near the di-iron centre core.

4.13.6 UV visible spectrophotometric analysis

A UV-visible spectrophotometric analysis of the Cj0241Hr hemerythrin proteins in general showed two major peaks at approx. 336 nm and 375 nm and a minor peak at 500 nm (Figure 4.36). These features are related to the di-iron centre absorbance, and characteristic of the met form of the hemerythrins in *C.jejuni* (Kendall *et al.*, 2014). During redox cycling the reduction of the stage met protein via auto-oxidation results in the deoxy shape of the protein, whereas a hydroxide bridges the di-iron binuclear centre. Re-oxidation of the hemerythrin protein by oxygen to give the hydroperoxyl group can be followed by an encouragement of the 500 nm characteristic that is diagnostic of the oxy stage of hemerythrins (Kao *et al.*, 2008).

Measurement of the autoxidation rate might reveal some idea about the function of the hemerythrins in *C. jejuni*, because different rates are proposed for storage, transporter or oxygen sensing hemerythrins. However, measurement of the UV-visible spectra for the deoxy and oxy hemerythrin forms proved impossible. These spectra are important stages for estimating the autoxidation rate by measuring the oxy-hemerythrin half-life ($t_{1/2}$). It was not possible to monitor the change because the protein was unstable under reducing conditions and always precipitated. The experiment was conducted (as described in chapter (2) section 2.13) using various concentrations of sodium dithionite ($Na_2S_2O_4$) and also sodium borohydride ($NaBH_4$) to produce the deoxy form. The hemerythrin protein was very sensitive and readily aggregated and precipitated. Concentrations of sodium dithionite as low as approx. 0.5mM gave rise to precipitation of Cj0241, but the lower concentrations tested did not show any signs of reducing the

enzyme to the deoxy form (Figure 4.37). Furthermore, an enzyme based system for depleting the oxygen level in solution of Cj0241-Hr protein to create a non-reducing environment (as described in chapter (2) section 2.14) was used to find the de-oxy form. However, UV spectrum analysis showed that the resulting sample still contained only the met form of the hemerythrin protein. The experiment was repeated three times but the results remained the same.

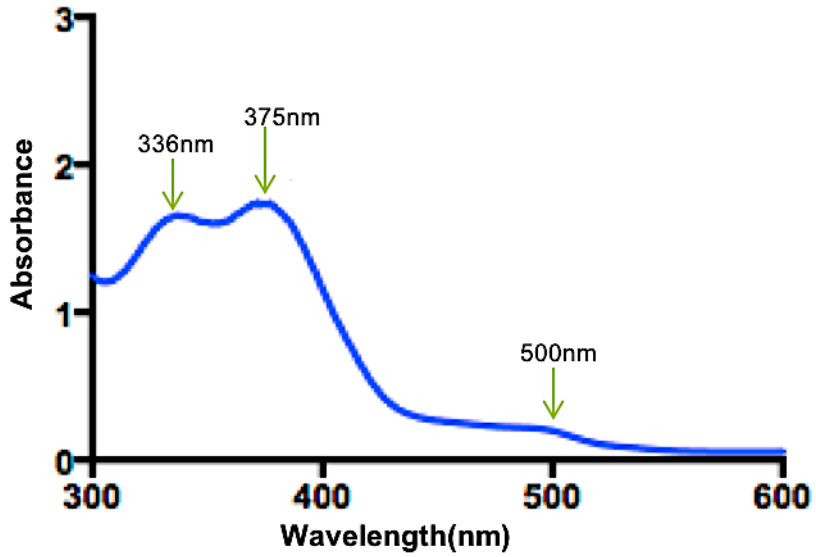


Figure 4.36: Visible spectrum of Cj0241 met-Hr. The spectrum exhibited two major absorption bands at 336 nm and 375 nm and a weaker characteristic band at approx. 500 nm, which are explained by the ligand-to-metal charge transfer (LMCT) transitions from the μ-oxo bridge to Fe(III) within the di-iron entre. These spectral characteristics are diagnostic of the met stage of hemerythrins (Kao *et al.*, 2008).

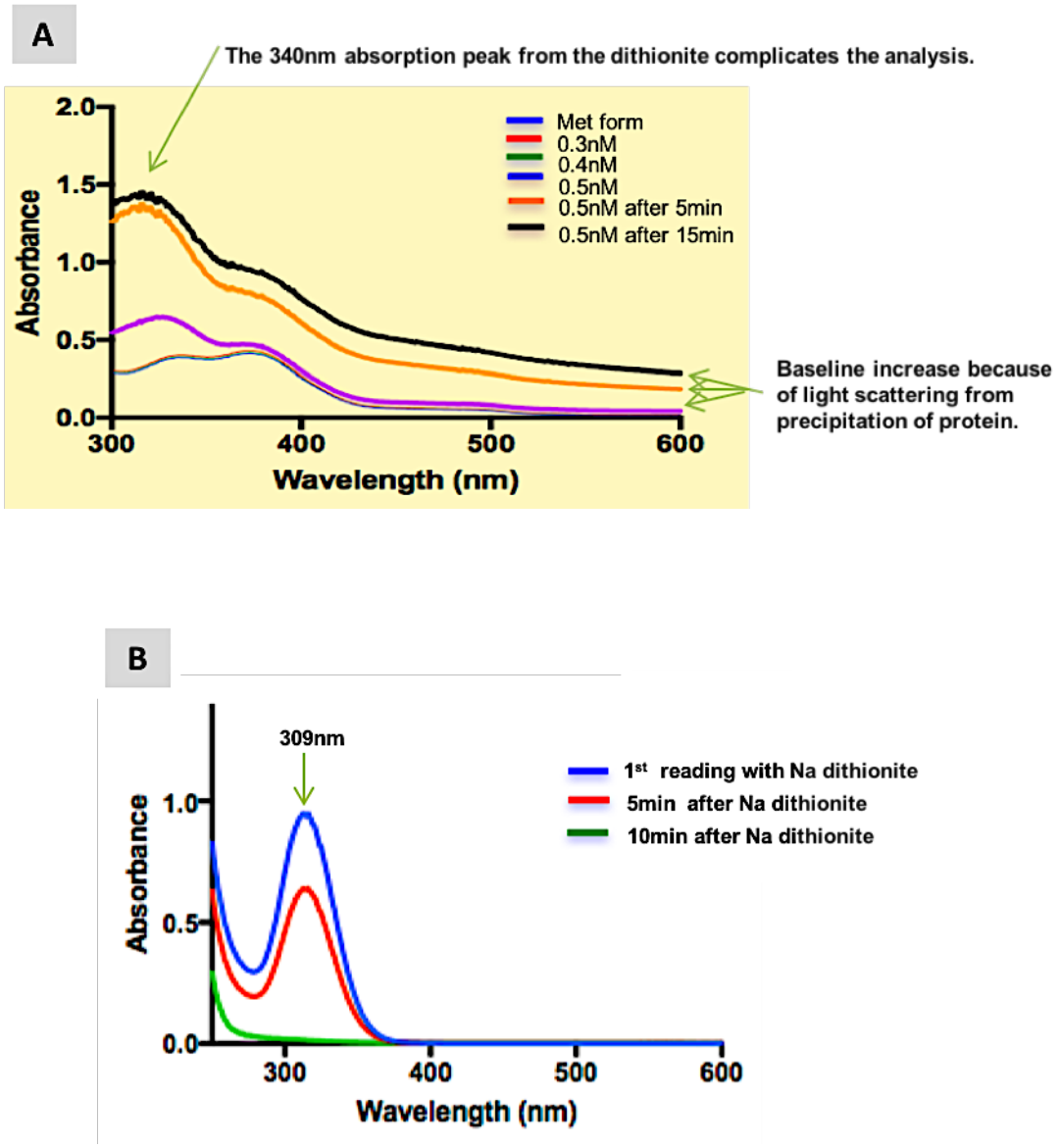


Figure 4.37: Visible spectrum of Cj0241 met-Hr. Panel (A) Concentrations of 0.3mM & 0.4mM of sodium dithionite showed no difference in the spectra from the original shape of the met form without dithionite. However, upon adding 0.5mM dithionite, the Cj0241-Hr protein started to aggregate. **Panel (B)** A reference absorbance spectrum for 5mM dithionite solution showed a peak at approx. 309nm, which disappeared after 10 to 12 min as the dithionite, decayed.

4.13.7 Crystal structure of the met-Cj0241Hr comparison with the other hemerythrin species

The crystal structure of the met-Cj0241-Hr shows a similar di-iron metal centre and μ-oxo bridge that has been seen in the structures of hemerythrins from *Methylococcus capsulatus* [Bath] (McHr), from *Desulfovibrio vulgaris* (DcrH-Hr) and from the invertebrate *Phascolopsis gouldii* (Isaza *et al.*, 2006; Kao *et al.*, 2008; Chen *et al.*, 2015). A comparison of the di-iron sites of the hemerythrins in these species (Figure 4.38) shows very similar geometries and link distances of the two iron (Fe) atoms and oxygen atoms, as well as the surrounding residues. The iron–ligand bond distances of the μ-oxo bridged di-iron centre in Cj0241-Hr agree with those in previous met-Hr structures.

Fe1 is surrounded with six-coordinating ligands in all cases: His75, His79, His115, Asp120 and Glu60 (residue numbers from Cj0241Hr). Likewise, Fe2 is associated with a hydroxide and four ligands His19, His56, Asp120 and Glu60. The Fe1–Fe2 distance (3.4 Å) in met-Cj0241-Hr is similar to that in DcrH- Hr (3.4 Å) and McHr 3.5 Å (Chen *et al.*, 2015), however , it is longer than those created in met-Hrs from invertebrates (3.2 Å) (Farmer *et al.*, 2001).

Similar water tunnels were recognised in the crystal structure of DcrH-Hr and McHr, however, the structures of the invertebrate animals show no water tunnels (Farmer *et al.*, 2001; Isaza *et al.*, 2006). The autoxidation rate $t_{1/2}$ differs between these species. In McHr it is ~ 50 min, but in the case of DcrH-Hr it is less than 1 min (Chen *et al.*, 2015). Many attempts were made to measure the $t_{1/2}$ for Cj0241-Hr as explained previously, but a completely reliable value was not possible. However, it is possible that the autoxidation of the oxy-Hr from Cj0241-Hr is similar to that of other bacterial hemerythrins from different species because of the existence of the water tunnel in the structure compared with the marine invertebrates that did not show this tunnel and have long autoxidation

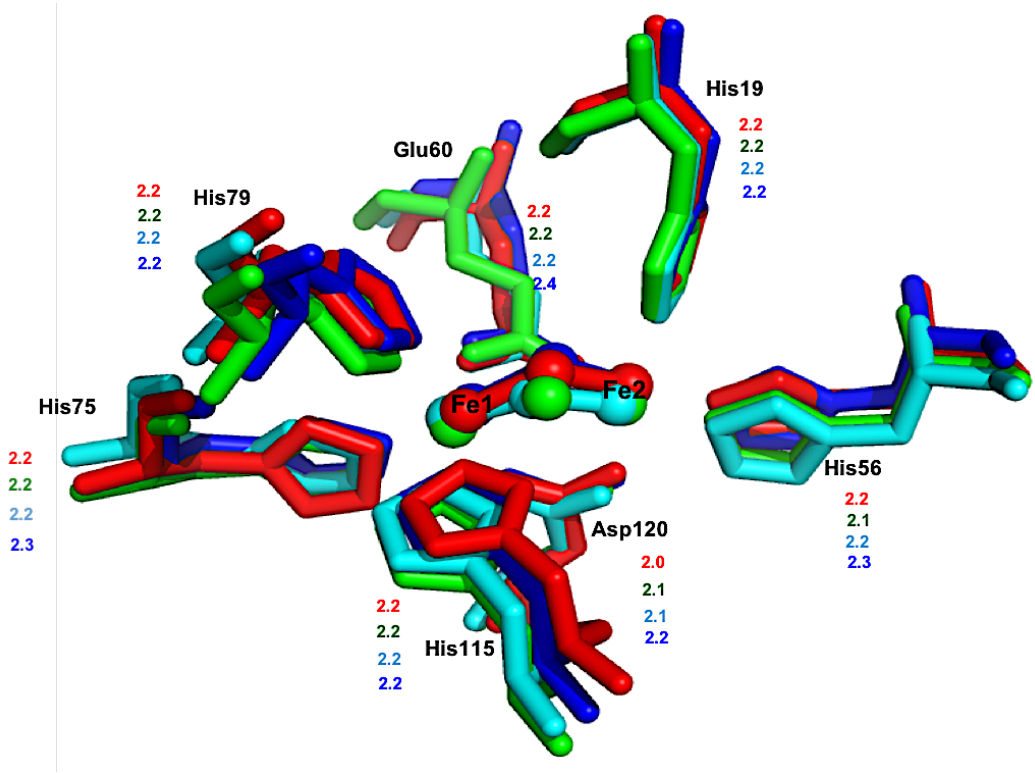


Figure 4.38: The iron–ligand bond distances of the μ-oxo bridged di-iron centre in Cj0241-Hr agree with those in previous met-Hr structures. Superimposed structures of the di-iron site in met-Cj0241-Hr (red), met-DcrH-Hr from *D. vulgaris* (PDB: 2AWY; cyan), met-McHr from *Methylococcus capsulatus* (PDB: 4XPX; green) and met-Hr from *P. gouldii* (PDB: 1I4Y; blue). The coordination link distances are shown for the Fe atoms and their related amino acid residue ligands.

half-life times. Moreover, the length dimension of the water tunnel between the protein surface and the metal centre is quite similar to the other bacteria. It is longer and more like the McHr tunnel (approx. 13 Å in length) than the tunnel in DcrH-Hr (approx. 11 Å in length).

From previous studies Chen and co-workers have found the Leu114 in McHr plays the function role of a water regulator in the water tunnel of McHr (Chen *et al.*, 2015). They mutated the L114 to A, F and Y, and found a 12-20-fold rise in the autoxidation process of the L114Y and L114F in the mutants comparing with the wild type of McHr. Additionally, the same results were found with mutations of the conserved residue Leu104 in myoHr (Xiong *et al.*, 2000) , Leu115 in DcrH-Hr (Isaza *et al.*, 2006) and Leu99 in Hr (Farmer *et al.*, 2001). Seemingly, the hydrophobicity of the leucine residue in these organisms perhaps plays a regulatory role in the autoxidation processes (Farmer *et al.*, 2001), and control the di-oxygen binding features in these organisms (Figure 4.39). A sequence alignment of Cj0241, DcrH-Hr, Hr, McHr, and myoHr (Figure 4.40), possibly suggests that the hydrophobic Ile112 in Cj0241 has the same role acting as a regulatory gate for the entrance of H_2O molecules to reach and contact the di-iron site of Cj0241.

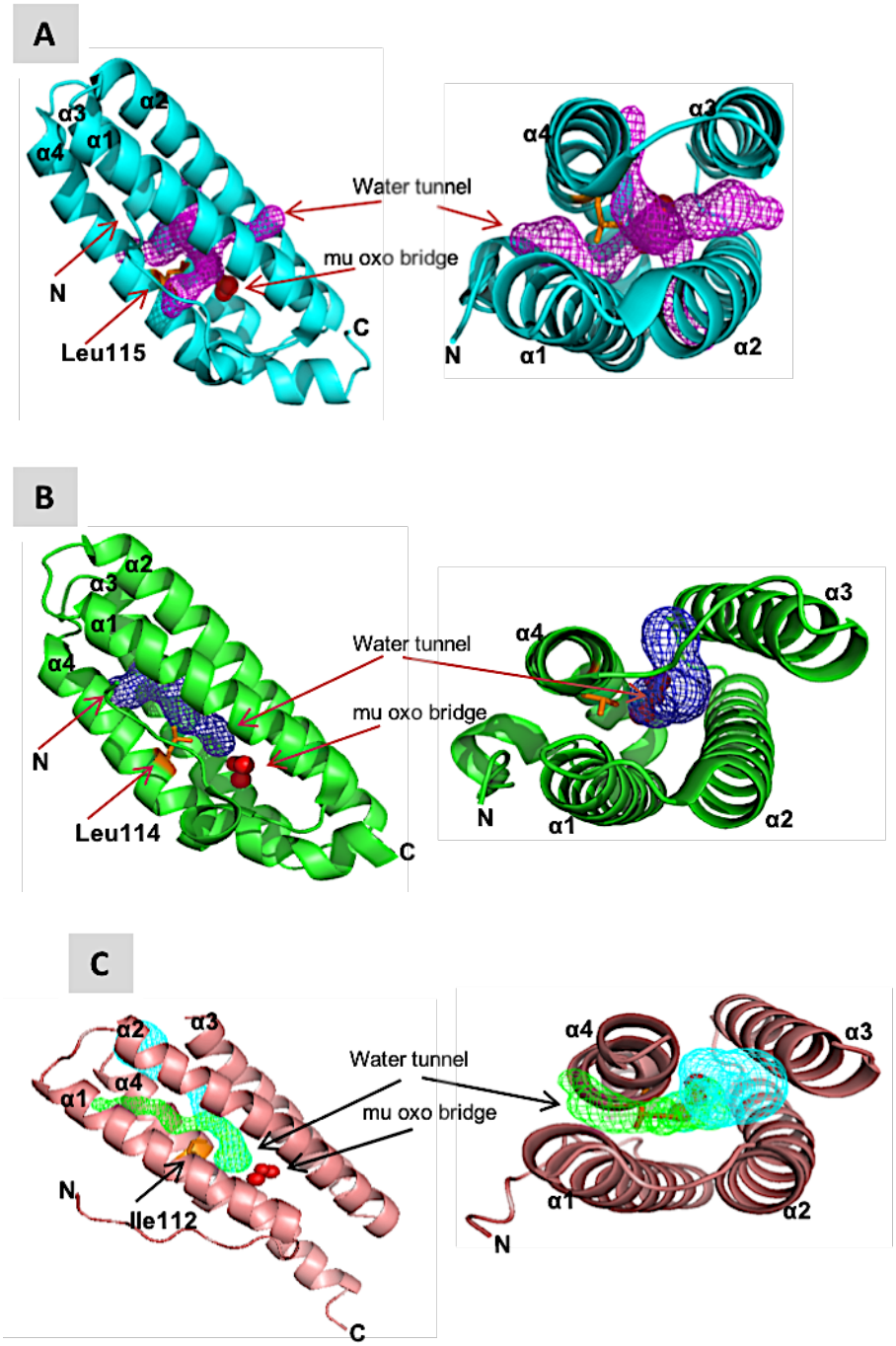


Figure 4.39: Top and side views comparison of the putative water tunnel in DcrH-Hr, McHr and Cj0241Hr. (A) Met- DcrH-Hr from *D. vulgaris* (PDB: 3AGT; purple): the water tunnels are highlighted in pink mesh. **(B)** Met-McHr from *M. capsulatus* (Bath): the water tunnel is shown in light blue mesh. **(C)** Cj0241 from *Campylobacter jejuni* (model 2): the water tunnels are highlighted in green and cyan mesh. Also, the structures show the Leu15 and Leu14 in DcrH-Hr and McHr, respectively. They are hypothesised possibly to have roles to limit the passage of H_2O molecules to the di-iron centre (Xiong *et al.*, 2000 ; Chen *et al.*, 2015). In addition, Ile112 from Cj0241 is shown which perhaps has the same function.

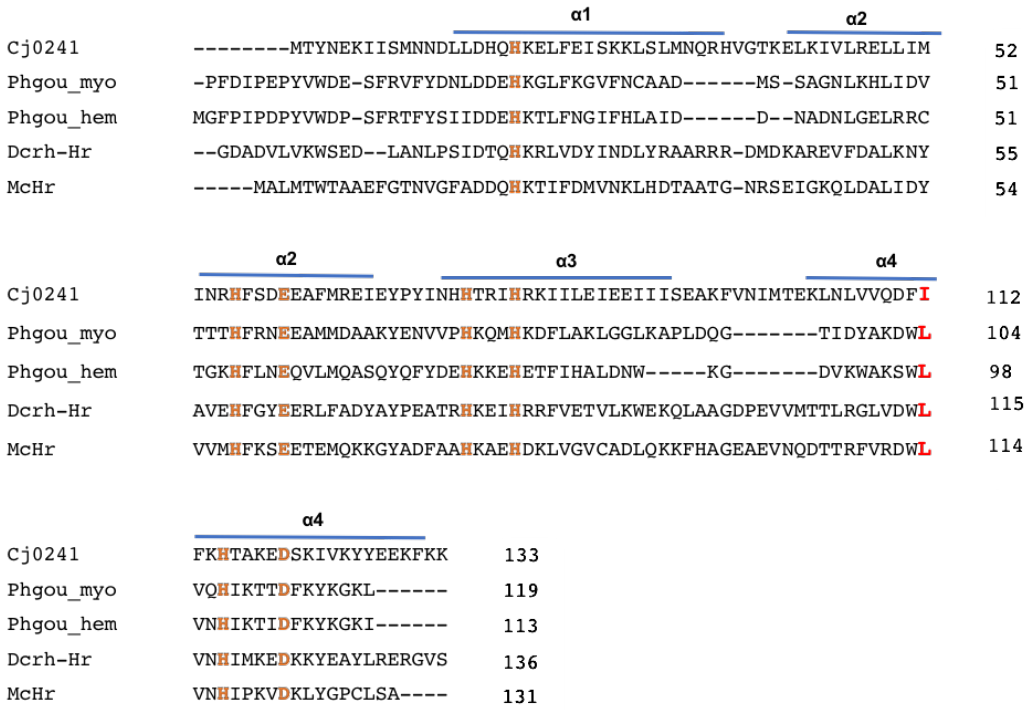


Figure 4.40: Sequence alignment of Cj0241 with various hemerythrins. The alignments of the hemerythrins from C.jejuni (Cj0241-Hr), *P. gouldii* myohemerythrin (Phgou_myo) (NCBI: gi253340), *P. gouldii* hemerythrin (Phgou_hem) (NCBI: gi6694943), the hemerythrin- like domain in *D. vulgaris* (DcrH) (NCBI: gi6685084) and *M. capsulatus* (Bath) (McHr) (NCBI: gi81682690), were achieved using ClustalW2 (Larkin *et al.*, 2007). The four alpha helices in the bundle of Cj0241-Hr structure are given by α1, α2, α3, and α4 at the top of the first sequence. The hydrophobic residue Ile112 in Cj0241-Hr may play a water valve role as seen in the studies that were conducted on Leu98, Leu104, Leu114 and Leu115) of the Phgou_myo, Phgou_hem, DcrH and McHr, respectively (highlighted in **red** colour). The amino acid residues supporting the coordination of di-iron core are highlighted in orange colour.

4.13.8 Effects of crystal packing on protein loops

The crystal packing effects on protein loop structures have been examined from previous studies by (I) a comparison of the loops in proteins that were crystallised in alternative packing orders, as well as (II) hypothetical prediction of loops in both with and without the inclusion of the crystal situation. Results showed that in a marginal of cases, loop geometries are reliant on crystal packing effects (Rapp and Pollack, 2005).

The Cj0241-Hr structure from the high resolution 3rd dataset (model 2) showed a weaker electron density map in the loop between helices 3 & 4 (Glu93, Ala94, Lys95, Phe96, Val97, Asn98 and Ile99). In contrast, this part in the lower resolution model from the 2nd dataset (model 1) had a strong density map, which covered around all these residues. Although many refinements were conducted to optimise this portion, improvement of the model compared with the first built structure from the 2nd dataset was not successful. These differences could be perhaps associated with the crystal packing interactions or arrangement in the two structures. As previously mentioned, although the two structures have the same space group $P2_1$, the crystals were not isomorphous and the packing arrangement was quite diverse between the two crystals.

Perhaps changes in loops or side chain conformation are likely to be related to a combination of inter-molecular (interactions with ligands or surrounding proteins in the crystal) forces, and intra-molecular (the local environment of the loops or side-chain in the protein)(Jacobson *et al.*, 2002).

In the loop of the low-resolution structure there is some contact with residues from molecules around it (Figure 4.41 A), and this possibly helps to stabilise the conformation the loop in the structure. Whereas, in the high-resolution structure, there is some space but the residues nearby are not close enough to make contacts, or keep that loop conformation (Figure 4.41 B). The same hypotheses of crystal packing can also be given

for the C- & N- termini in the high-resolution structure model 2. The C-terminal tag of one molecule and the N-terminal section of another adjacent molecules interact directly with each other in the high-resolution form, stabilising both N- & C- terminal residues in the high-resolution structure (Figure 4.42). Also, this interaction would certainly help to stabilise the crystal lattice overall between the molecules and gives better packing in the crystal, therefore leading to higher resolution data.

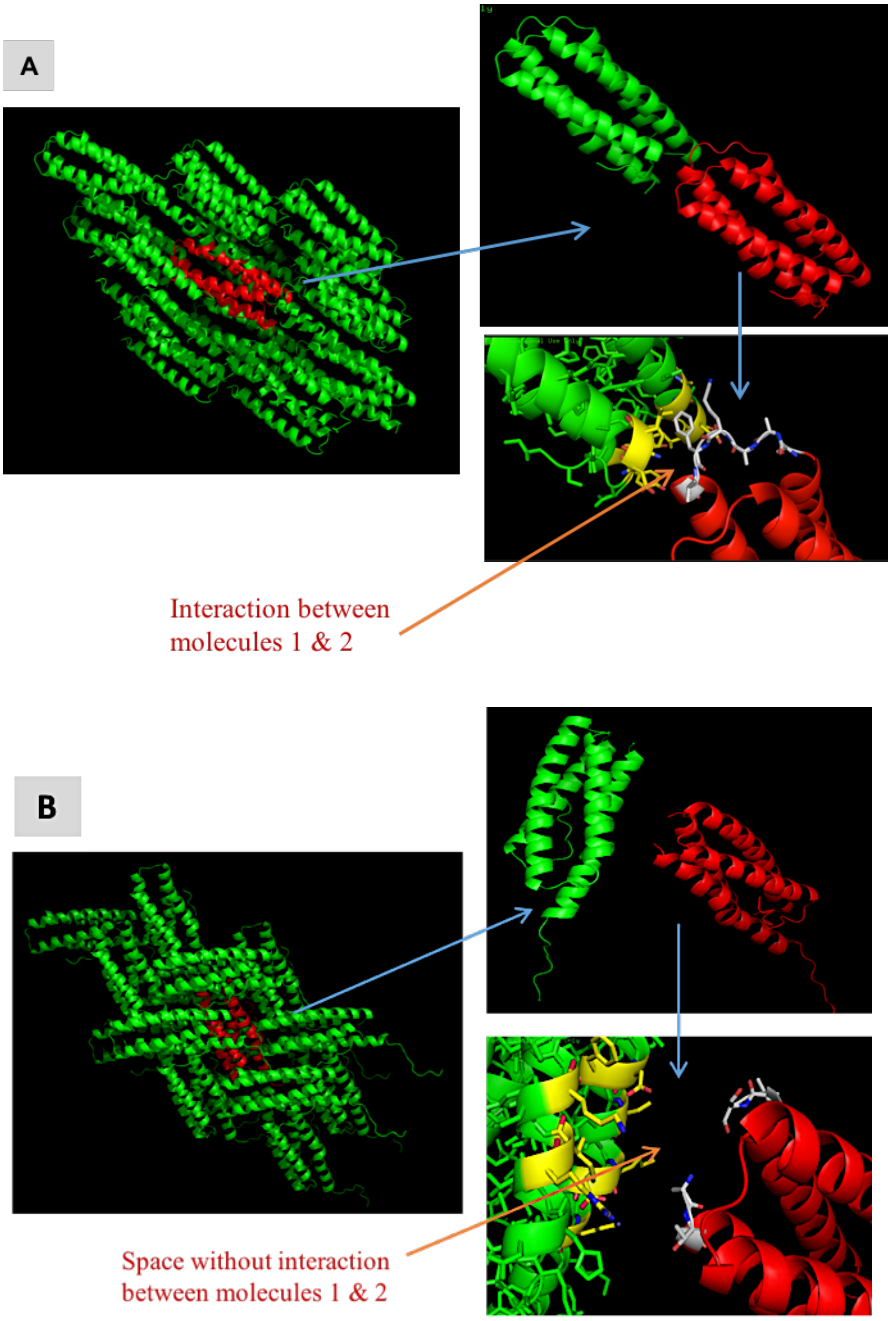
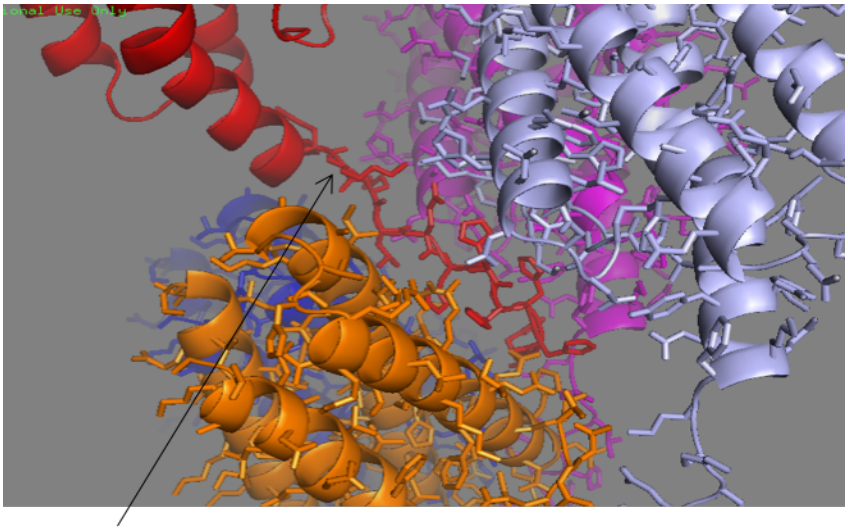


Figure 4.41: The crystal packing of Cj0241 in model 1& model 2 crystal forms. The crystals from the low resolution 2nd data set and the high resolution 3rd data set both belonged to space group $P2_1$ but had different unit cell dimensions and different crystal packing arrangements that have contributed to both the difference in resolution of the data and the stabilisation of certain surface loops on the proteins. **Panel (A)** the loop between helices α3 & α4 in the low-resolution structure interacts with the neighbouring protein by crystal contact. Phenylalanine packing into a nice hydrophobic pocket and making a favorable packing interaction. **Panel (B)** The same loop in the high-resolution structure does not interact with the neighbours in the structure and therefore the loop is not stabilised.



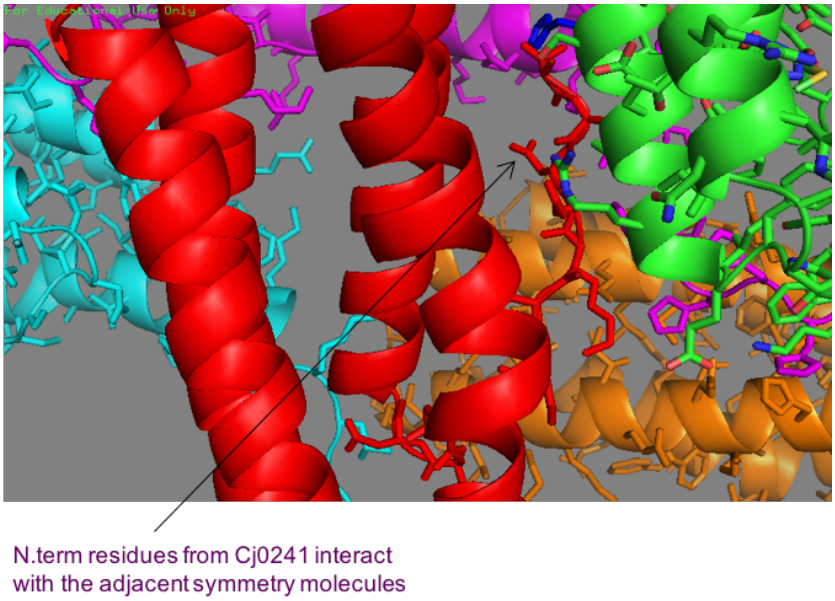


Figure 4.42: Cj0241-Hr model 2 crystal packing of C & N termini. The crystal packing of Cj0241 model 2 shows the C- & N- terminal residues are interacting with the adjacent molecules.

Chapter 5

Cj1224 Hemerythrin Expression, Purification and Crystallisation

This chapter will describe and discuss the results of the work on Cj1224-Hr protein. The methods from transformation to crystallisation stages were carried out with this type of hemerythrin. However, no crystals were discovered from this protein. Consequently, numbers of attempts were made to avoid the factors that perhaps influence the ability to form crystals protein. Also, the study includes using ICP-MS analysis, site-directed mutagenesis experiments and prediction structure of Cj1224-Hr using some computational programs.

5.1 Confirmation of the sequence of the inserted *cj1224c* gene in the pET-21b plasmid

The *cj1224c* gene that had been inserted into pET-21b plasmid was validated by DNA sequence analysis after extracting the DNA construct using a Miniprep kit (Qiagen) as described previously in chapter (2), section (2.11.8). An analysis using the tool of Basic local alignment search (BLAST) (Altschul *et al.*, 1990) at the National Centre for Biotechnology Information http://blast.ncbi.nlm.nih.gov/Blast.cgi, showed 100% identity with the sequence of Cj1224 from the universal protein resource (UniProt) database of protein sequence as presented in Figure (1.5) in chapter (1), section (1.8.6).

5.2 Testing the expression of the hemerythrin Cj1224 protein

Transformation and expression were performed aerobically in *E. coli* strain BL21 (DE3) as described previously in chapter (2), sections (2.4) and (2.5). Experiments were performed to define the best temperature and time following induction for overproduction of Cj1224. Two temperatures were tested (37 °C and 25 °C), and the time period of the expression was analysed using SDS-PAGE as described in Figure (5.1). From an observation of the gel, it is clear that the Cj1224-Hr protein (indicated by the arrow- ~25 kDa) achieved overexpression during the time-courses. In addition, it is noticeable the expression of soluble protein Cj1224-Hr appeared in both temperatures (Figure 5.2).

5.3 Purification of His Tag- Cj1224 hemerythrin

To purify the Cj1224-Hr protein, nickel affinity chromatography was carried out on samples after the induction of the protein expression from a large-scale culture, as described in the material and method chapter (2), section (2.10.1). From the chromatogram (Figure 5.3A) a peak was observed and fractions 15 to 20 were checked for protein concentration using the Bradford method as described previously in chapter (2) section (2.5.3). The progress of purification was analysed using12% SDS-PAGE as presented in Figure (5.3B). From the gel, a band of protein appeared which had molecular weight of approximately 25kDa corresponding to the Cj1224-Hr protein. Lanes 2 and 3 show the high level of hemerythrin expression and lanes 4 to 6 indicate that all the Cj1224-Hr was bound to the Ni column. Fractions 18 & 19 were joint to give a 6ml sample at 3.5mg/ml, and used to prepare a crystallisation sample.

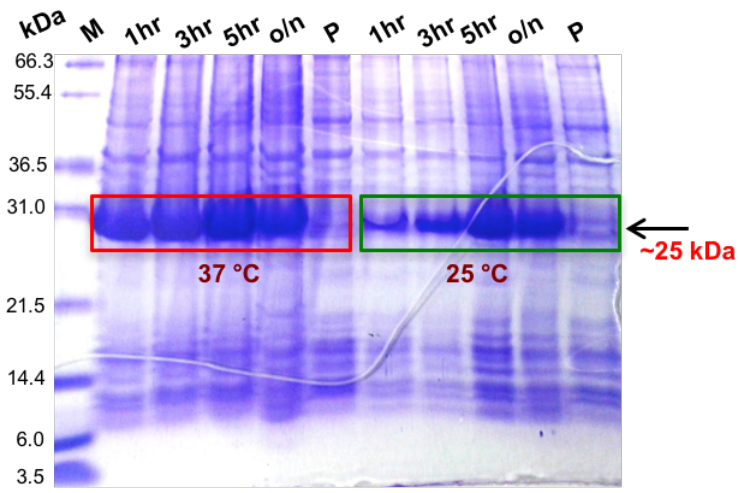


Figure 5.1: The overexpression of Cj1224-Hr protein. Cells were cultivated to 0.6 (OD_{600}) and induced using 1mM of IPTG prior to incubation with 250 rpm shaking at either 25˚C or 37˚C. Samples were collected before induction then at 1, 3, 5 and 16 hours over night (o/n) post-induction. A band indicated by an arrow at about 25 kDa in all induced lanes indicates that the Cj1224-Hr protein expression was positively induced. M= The Mark12™ Unstained Standard, numbers show dimensions of molecular weights of sample protein in kDa.

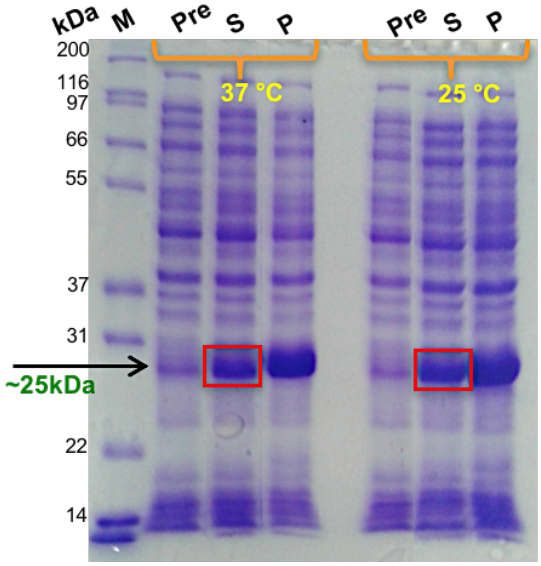


Figure 5.2: Cj1224-Hr protein solubility was checked at 25°C and 37°C. Both temperatures permitted the bacteria to produce soluble protein, arrow indicator. Pre = pre-induction, S = soluble protein, P= insoluble protein and M= The Mark12™ Unstained Standard, numbers show dimensions of molecular weight of sample protein in kDa.

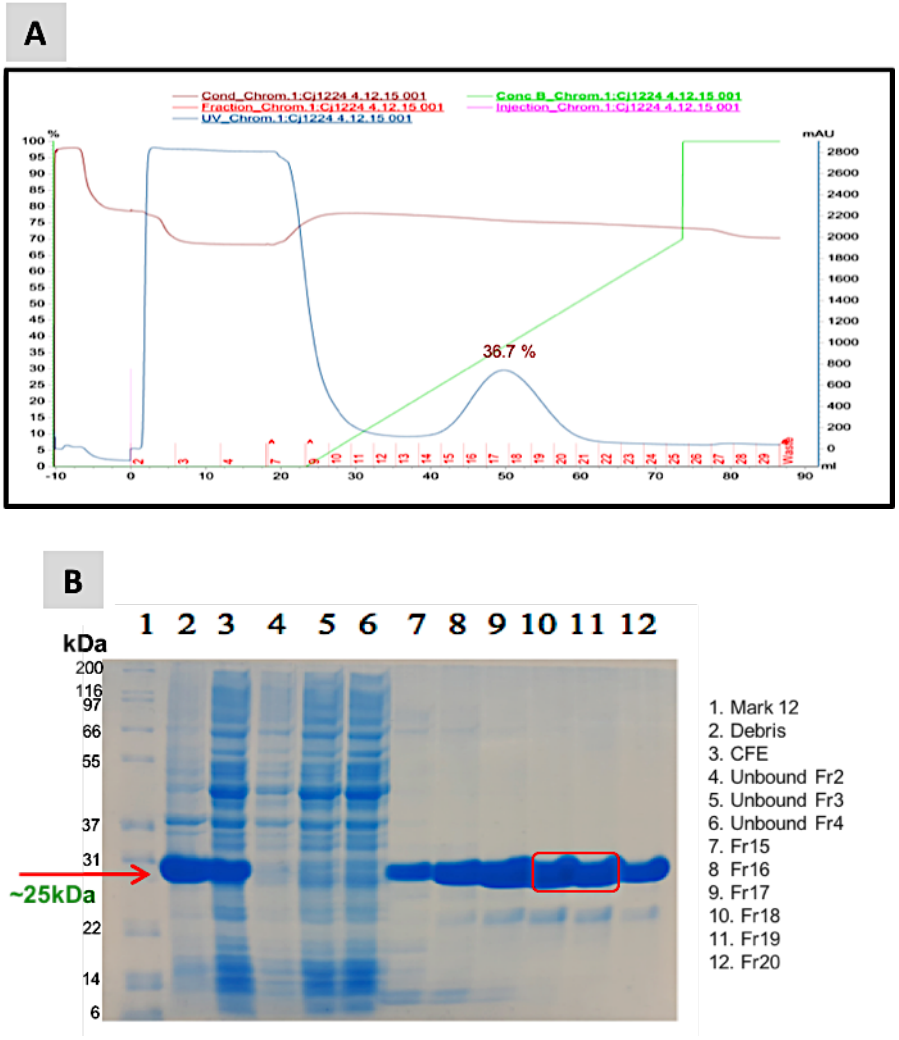


Figure 5.3: The chromatogram obtained via Nickel affinity column for Cj1224-Hr protein and SDS-PAGE analysis. Panel (A) The chromatogram for the purification of Cj1224-Hr via Nickel affinity column (HisTrap HP 5ml from GE Healthcare™). Following sample application and column wash, an increasing gradient (0-70%) of 0.5M Imidazole was applied. Cj1224-Hr was eluted at ~36.7 %, which corresponds to 0.18M Imidazole. **Panel (B)** An SDS-PAGE analysis illustrating the purity of the Cj1224-Hr protein at various steps of the purification process. Lane 1 The Mark12™ (M) Unstained Standard (Thermo Fisher Scientific), numbers show dimensions of molecular weight of sample protein in kDa. Lane 2 and 3 are the cell debris and cell free extract respectively, and lanes 4- 6 from washing fractions. Lanes 7-12 correspond to the peak in the chromatogram and contained the most Cj1224-Hr protein. Lanes 10 & 11 (red square) were combined together to prepare the crystallisation sample.

The purified protein was coloured like Cj0241-Hr protein, which may be coming from iron ions that are bound to the Cj1224-Hr protein (Figure 5.4). In addition, the molecular weight of the purified protein was checked by Mass spectrometry (Department of Chemistry, The University of Sheffield), and the result exhibited the molecular weight of this protein to be 24697.29 Da as expect from ProtParam tool services at approx. 24697.15 Da (Figure 5.5).

5.4 Initial crystallisation using commercial screens

The purified hemerythrin Cj1224 protein was concentrated for crystallisation purposes as described in chapter (2) section (2.11.2). The sample was then transferred into buffer 10mM Tris-HCL pH 8.0 + 0.1M NaCl and used at 7.5 mg/ml protein concentration. Initial crystallisation experiments were carried out using various screening kits: PACT, JCSG, PEGS, Ammonium sulphate, Proplex, Morpheus and MPD commercial screens (Qiagen), with a Hydra II e-drop automated pipetting system (Matrix) as described in chapter (2) section (2.10.2). All the trays were kept in the incubation room at 17º C. After around 48hrs, some cluster crystals were observed in several screening plates, but unfortunately no successful protein crystals were produced from these screen process. Also, the screens were carried out using metal ions and reducing agents (DTT), as described in chapter (2) section (2.11.1.2 and 2.11.1.3), but again the screening showed only salt crystals. Therefore, some analysis was conducted to find the reasons why this protein has not crystallised as explained in the next pages.

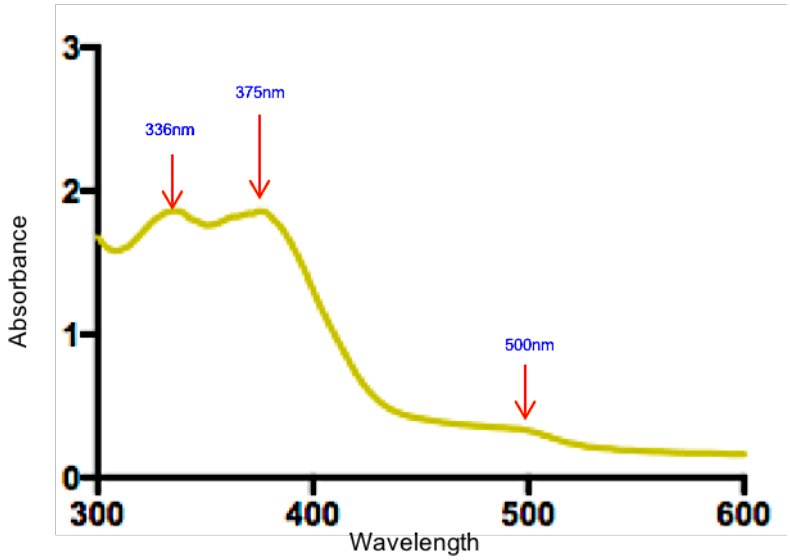


Figure 5.4: Visible spectrum of Cj1224-Hr protein. The UV-spectrum confirmed the protein sample contains iron ions from its characteristic μ-oxo bridge to Fe (III) and -OOH to Fe (III) ligand-to-metal charge transfer (LMCT) bands at 336 nm 375 nm and 500nm.

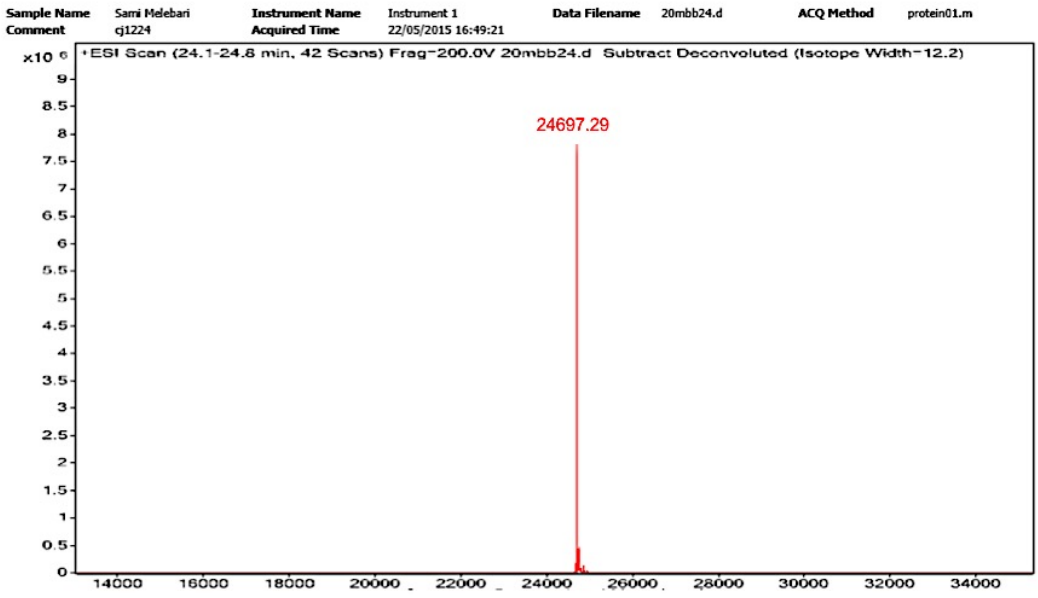


Figure 5.5: Cj1224-Hr protein Mass Spectrometry analysis. The results obtained from mass spectrometry analysis verified the molecular weight of Cj1224-Hr, and agreed with the ProtParam tool predicted molecular weight of 24697.15 Da. The result included the Cj1224 sequence, plus 2 linker residues and a 6 hisidine residue tag.

5.5 Analysis of protein folding using Nuclear Magnetic Resonance (NMR) spectroscopy

The state of Cj1224-Hr protein was evaluated by Nuclear Magnetic Resonance (NMR) spectroscopy to find out whether it had folded, which would cause a problem in crystallising the protein. The sample was sent to the NMR lab as described in chapter (2) section (2.11.5), and the result obtained showed that the protein was folded (Creighton, 1992). From the chart (Figure 5.6)., the peaks in the NMR spectrum were sharp, narrow, and with a dispersion over a large range of chemical shifts; below 0.5ppm and above 8.5 ppm.

5.6 Analysis of Free Thiol group and Disulfide Bond

Disulfide bonds have significant roles in the solubility and stability of proteins. In other proteins, the free cysteine residue is part of the catalytic activity site. The existence of a disulfide bond in proteins and peptides has been shown to effect the stability of a protein (Maulik *et al.*, 2009). So, it is important to study the characteristics of cysteine residue thiol groups within Cj1224-Hr protein because their effect on things like degradation, aggregation or conformational state. The existence of a free thiol group in the cysteine residue in a protein can be quantitatively assessed by different techniques. The initial reagent commonly used for approximating the number of thiol groups was called Ellman's reagent (5,5'-dithiobis-(2-nitrobenzoic acid) or DTNB). The compound 5-thio-2-nitrobenzoic acid (TNB) is released in the reaction between thiol group(s) and DTNB (Figure 5.7), which absorbs at 412 nm and pH 8.0 (Maulik *et al.*, 2009). Therefore, the rate of this reaction is dependent on several factors: 1) the reaction pH, 2) the pKa of the sulfhydryl and 3) steric and electrostatic effects.

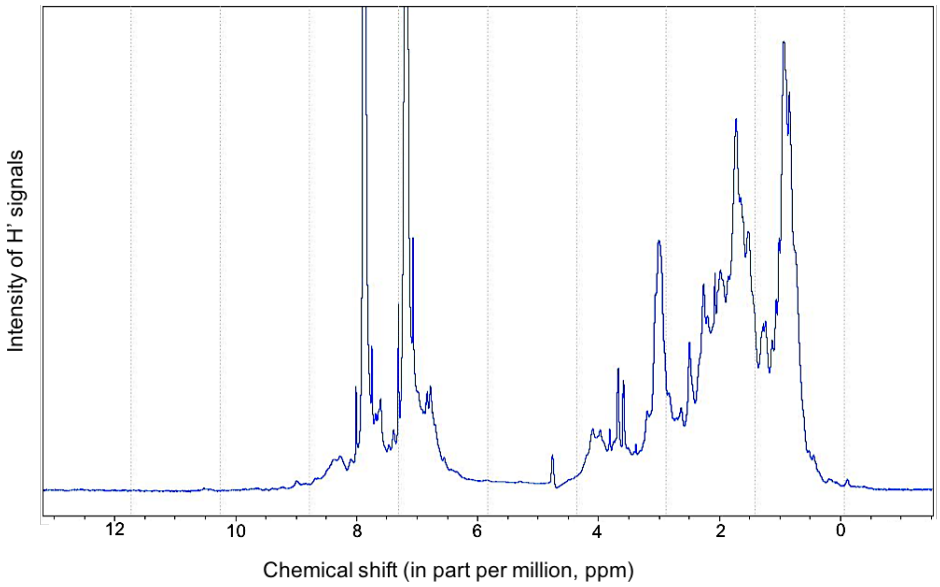


Figure 5.6: ^{1}H-NMR spectrum of Cj1224-Hr protein. The peaks showed dispersion below 0.5 ppm and above 8.5 ppm of chemical shift, and make sign that the Cj1224 is folded protein.

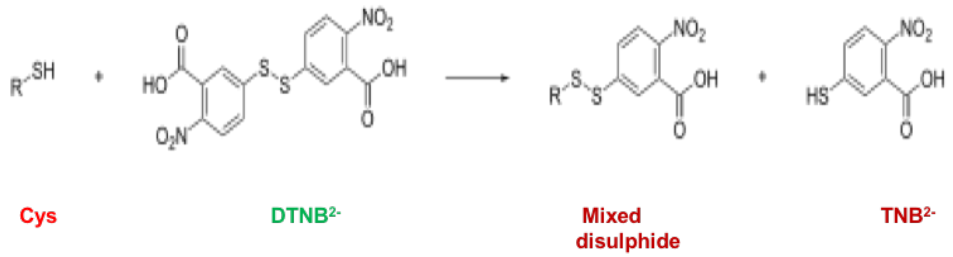


Figure 5.7: Reaction of DTNB with a thiol (R-SH)

The protein concentration used in the assay was 0.256 mM, and the free cysteine content of Cj1224-Hr assessed in two buffer systems as described in chapter (2) section (2.11.13). Buffer A was that used in the purification of Cj1224-Hr and contains 0.1M Na phosphate and 1mM EDTA pH 7.3. However, buffer B was that used to denature the protein and contains 6M GuHCl, 0.1M Na phosphate, and 1mM EDTA pH 7.3.

The results obtained from treatment of Cj1224-Hr with Ellman's reagent in buffer A (Table 5.1) illustrated one pair of cysteine residues could have remained as a bridge, and the other two cysteine residues were free but only one was accessible. Alternatively, three cysteine residues could have been buried and not accessible to DTNB. The DNTB molecule is quite big, it could possibly be able to interact only with the cysteine thiol groups, which are found on the surface of the molecule and free to the solution. Therefore, it was possible that not all the molecules of Cj1224-Hr had an S-S bridge in existence and a proportion of the molecules had this S-S bridge broken for some reason.

Table 5.1: Treatment of Cj1224-Hr with Ellman's reagent in buffer A

Pr. amount (µl)	Abs 412nm	TNB uM	Mol thiol group/ Mol protein
20	0.067	4.9	1
40	0.147	11	1.1
60	0.270	20	1.3

Treatment of Cj1224-Hr with Ellman's reagent in buffer B (Table 5.2) revealed more than 2 cysteine residues available. However, in a small protein like Cj1224, when it is unfolded all cysteine residues should be accessible unless they are in a disulphide bridge.

Table 5.2: Treatment of Cj1224-Hr with Ellman's reagent in buffer B

Pr. amount (µl)	Abs 412nm	TNB uM	Mol thiol group/ Mol protein
20	0.191	14	2.7
40	0.348	26	2.5
60	0.523	38	2.4

Taken together, the results from buffer A and buffer B can be most simply interpreted by a Cj1224 protein in solution where no disulphide bridges are formed but only one of the cysteine residues is accessible to DTNB in the folded state.

5.7 Microseed Matrix screening experiment

A microseed-matrix technique has been found to induce the nucleation event in standard crystallisation screens (Luft and Detitta, 1999). Using this method, the previous crystals from the related Cj0241-Hr were used as seeds and introduced into new drops of Cj1224-Hr protein as explained in chapter (2) section (2.11.14). Unfortunately, the result of this trial was salt crystals too. However, one crystal did provide diffraction data, but it had unit cell parameters the same values as those seen in Cj0241-Hr. It was concluded that this was the result of either poor seed preparation or bad handling that carried over Cj0241 protein solution and thus highlighted the difficulties with the approach.

5.8 Cj1224 hemerythrin protein with and without His-tags

A polyhistidine-tag has proved increasingly popular during the past decade as a purification method for recombinant protein (Carson *et al.*, 2007). Cloning vectors

incorporate commonly six repeated histidines and a linker, sometimes with a protease-cleavage site, to the C- or N-terminus of the interest protein. These tags enable to bind of the expressed desired protein to a nickel-affinity column. The tag can then be removed by a protease either on the column or after the elution process. The common understanding is that these histidine tags have no influence on the function and structure of the protein (Chant *et al.*, 2005). However, in crystallography, the tag maybe plays an important role. They may increase the aggregation or alter the solubility of the purified protein (Carson *et al.*, 2007). Also, crystallographers think these tags could probably delay or prevent crystallisation because they might interfere possibly with crystal formation by blocking lattice contacts (Carson *et al.*, 2007). Additionally, there is some argument that the position of the histidine tag as C-terminal or N-terminal may also influence the crystallisation of the native protein (Carson *et al.*, 2007). Thus, new primers were designed for *cj1224* gene to put the tag at the N-terminal end of the protein and also to take the tag off completely.

5.8.1 Cloning of Cj1224-Hr

5.8.1.1 Amplification of the cj1224-Hr gene

The initial Cj1224 expression plasmid with a C-terminal His-tag was used as the DNA source. The *cj1224* gene was amplified by PCR as described in chapter (2) section (2.11.6.2) using the Cj1224-Hr forward and reverse primers and Q5® High-Fidelity 2X Master reaction mixture. To analyse the PCR products, they were run on a 1% agarose gel, which showed a band of the predicted size of approx. 600 bp (Figure 5.8). A Gel Extraction Kit (Qiagen) was used to purify the PCR products as explained in chapter (2) section (2.11.6.4).

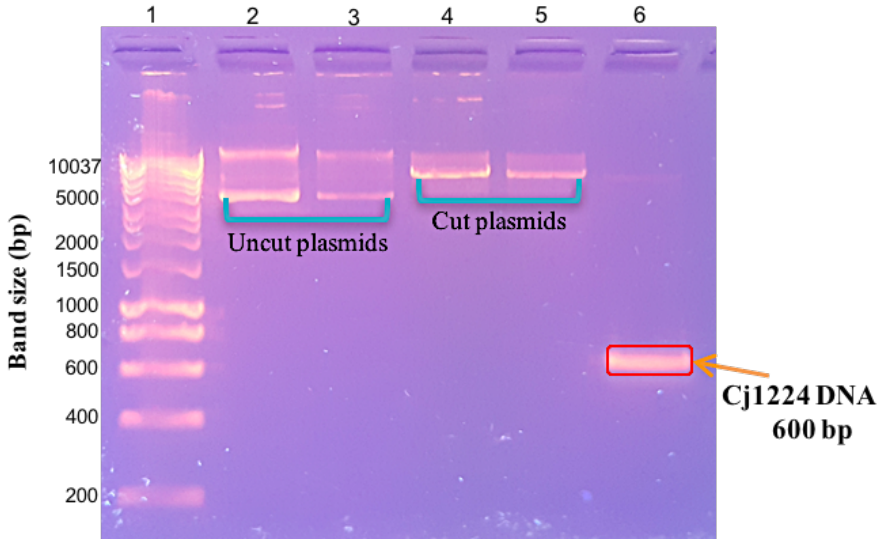


Figure 5.8: 1% DNA agarose gel of *cj1224-Hr* gene amplification and digested plasmids pET21b & pET28b. Lane (1) is DNA hyper ladder^{TM} 1kb molecular weight marker (BIOLINE). Lane (6) is Cj1224-Hr PCR product band and the arrow sign shows the expected size of 600 bp. Lanes 2 & 3 are pET21b & pET28b respectively before being cut using restriction enzymes. Lanes 4 & 5 are pET21b & pET28b respectively after being cut using the restriction enzymes and show at expected sizes of 5443 bp & 5369 bp respectively.

5.8.1.2 Restriction enzymes digest

The PCR product of *cj1224-Hr* gene was then digested using NdeI and XhoI restriction enzymes prior to insertion into a pET28b or a pET21b plasmid that was cut with the same enzymes as previously stated in chapter (2) section (2.11.6.5). In both plasmids, a stop codon was inserted to prevent the C-terminal tag inclusion. The digested samples were then analysed with a 1% agarose gel, and the bands showed at the predicted sizes (Figure 5.8). The digested vectors after that were treated with Shrimp Alkaline Phosphatase (rSAP) before setting up the ligation reaction as explained in chapter (2) section (2.11.6.6.1). The ligation mixes were transformed into competent DH5α *E. coli* cells (Novagen). As plasmids (pET21b & pET28b) carry an antibiotic resistance gene (ampicillin for pET21b or kanamycin for pET28b), the subsequent transformants were spread on LB-agar plates supported with the suitable antibiotic (100μg/ml ampicillin pET21b or 50μg/ml kanamycin for pET28b), and the plates were incubated at 37°C overnight. Next day some colonies were observed on the agar plates, and therefore screened for successful cloning.

5.8.1.3 Confirmation of the existence cj1224-Hr gene in the pET21b and pET28b plasmids

The presence of the *cj1224* gene in the pET21b and pET28b plasmids was checked and confirmed by DNA sequence analysis and both sequences showed 100% match with that in the NCBI database, see sequence in chapter (1) section (1.8.6). From the plated clones, a number of colonies were taken and the possible constructs were then extracted using a technique of miniprep kit (Qiagen) as described in chapter (2) section (2.11.8). The resultant miniprep samples were sent with T7 forward and reverse primers to GATC Biotech to check the sequence.

5.8.1.4 Expression and solubility investigation

The confirmed construct was transformed into a competent cell of *E. coli* strain BL21 (DE3) (Novagen), and plated on LB agar media supplemented with appropriate antibiotics. The experiment was conducted as explained in chapter (2) section (2.11.9). Two types of colonies from transformed BL21 (DE3) were used: Cj1224_notag and 6xHis_Cj1224, to check the expression and solubility at two different temperatures (25°C & 37 °C) with 1mM concentration of IPTG. The protein concentration of the samples to be analysed on the gel was estimated using the Bradford method as explained in chapter (2) section (2.5.3). The majority of Cj1224-Hr protein in each case was over-expressed in the insoluble fractions as shown by SDS-PAGE analysis, however, luckily there was still some small amount of protein in the soluble fractions at both temperatures. The molecular weight of the expressed protein was the same as expected at approx. 25 kDa (Figure 5.9). The large-scale culture growth of Cj1224-Hr protein with overexpression was conducted as explained in chapter (2) section (2.11.9.2) at 25 °C with shaking at 220rpm.

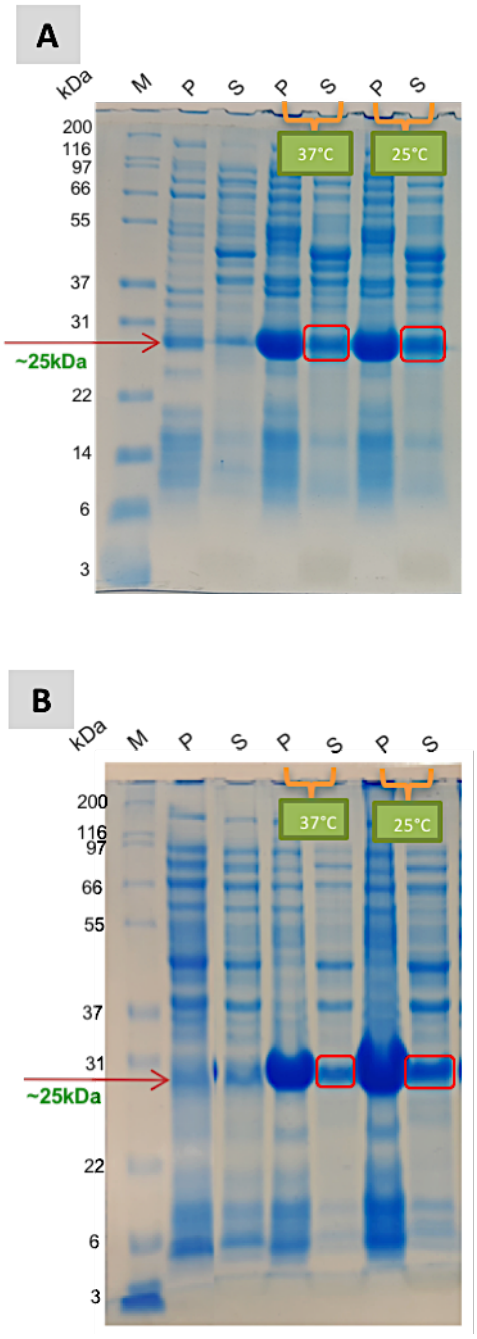


Figure 5.9: SDS PAGE analysis of the expression of Cj1224-Hr at 25° C & 37 ° C with 1mM IPTG. Panel (A) CJ1224-Hr without His-tag. **Panel (B)** shows Cj1224-Hr with N-terminal His-tag. The experiment showed that there were small amounts of soluble protein at 25° C & 37 ° C. M = molecular weight marker, P = insoluble fraction, S = soluble fraction.

5.8.1.5 Purification of non-tagged Cj1224

A one litre culture of Cj1224-Hr without His-tag produced 11g of cell paste, which was lysed and purified as described in chapter (2) section (2.11.10). The protein was applied on a 20ml DEAE-FF column and the fractions were analysed by SDS-PAGE (Figure 5.10). Fractions 11 to 14 contained the most pure Cj1224-Hr, and were combined and concentrated with total protein of approx. 9mg. This sample was then applied on a Superdex200 gel filtration column. The Cj1224-Hr was eluted from the column at approx. 80ml, which corresponding to an apparent MW of 31.6 kDa, indicating its monomeric state in solution (Figure 5.11A & C). Also, the results were analysed by SDS-PAGE and fractions 23 & 24 contained the most pure Cj1224-Hr protein (Figure 5.11B) were combined with a total protein yield of 2.8 mg. This sample was then concentrated to 19mg/ml and the buffer was then exchanged by using Zeba Desalt Spin Columns to 5mM Tris pH 8.0 and 50mM NaCl for use in crystallisation trials.

5.8.1.6 Crystallisation trials of Cj1224-Hr

The sample of untagged Cj1224-Hr protein at 11mg/ml concentration was screened. Different commercial crystallisation screens were used including the JCSG, PACT, PEG, MPD, Ammonium sulphate and Proplex suites (Qiagen / Molecular Dimensions). However, unfortunately all results for these trials were salt crystals.

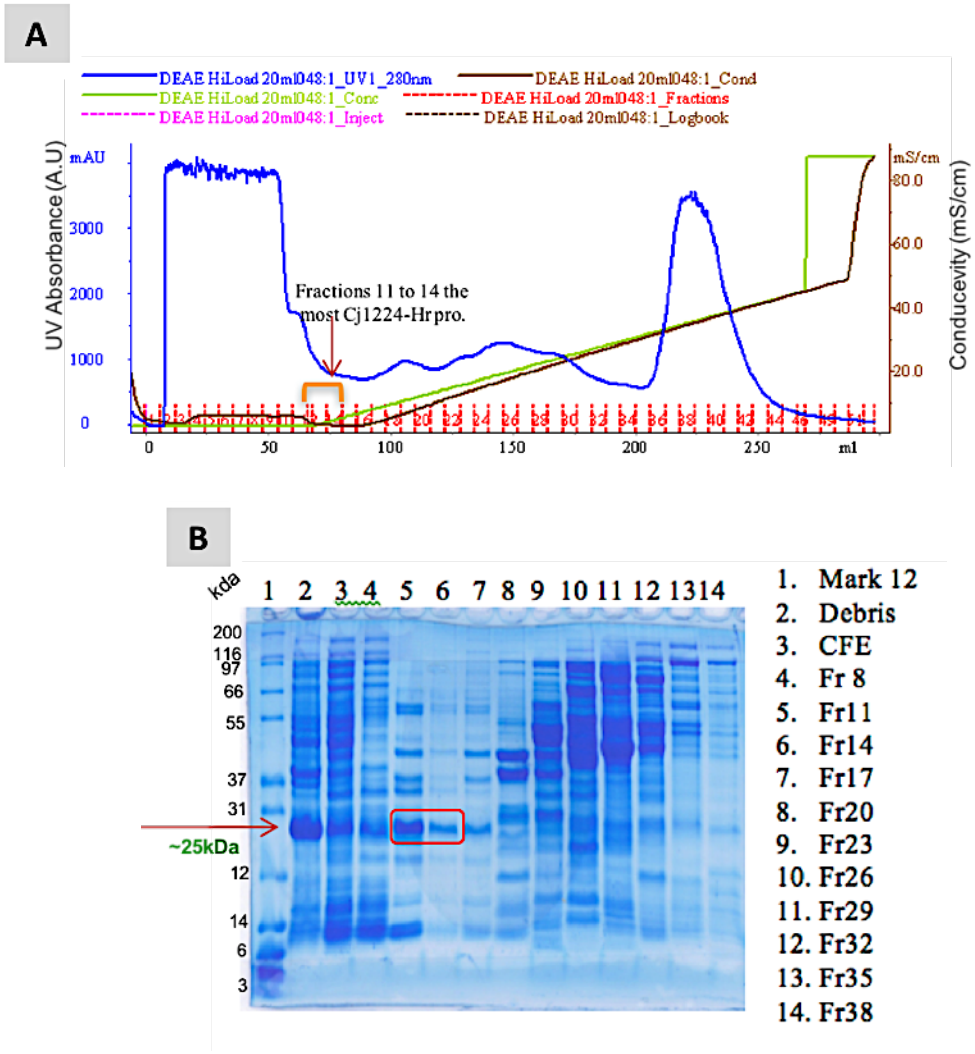


Figure 5.10: Ion exchange purification of untagged Cj1224-Hr on a DEAE-FF column. Panel (A) the chromatogram from the DEAE-FF column represents the fractions after applying the sample on ion exchange chromatography. Cj1224-Hr was eluted from the column by 200ml of gradient of 0-0.5M NaCl in 50mM Tris pH8.0 at flow rate 5ml/min. **Panel (B)** 12% SDS-PAGE analysis of the fractions from the DEAE-FF column. The Cj1224-Hr protein is shown (the arrow sign) with size ~25 kDa. Fractions 11-14 were combined (red box) with total protein of approx. 13mg.

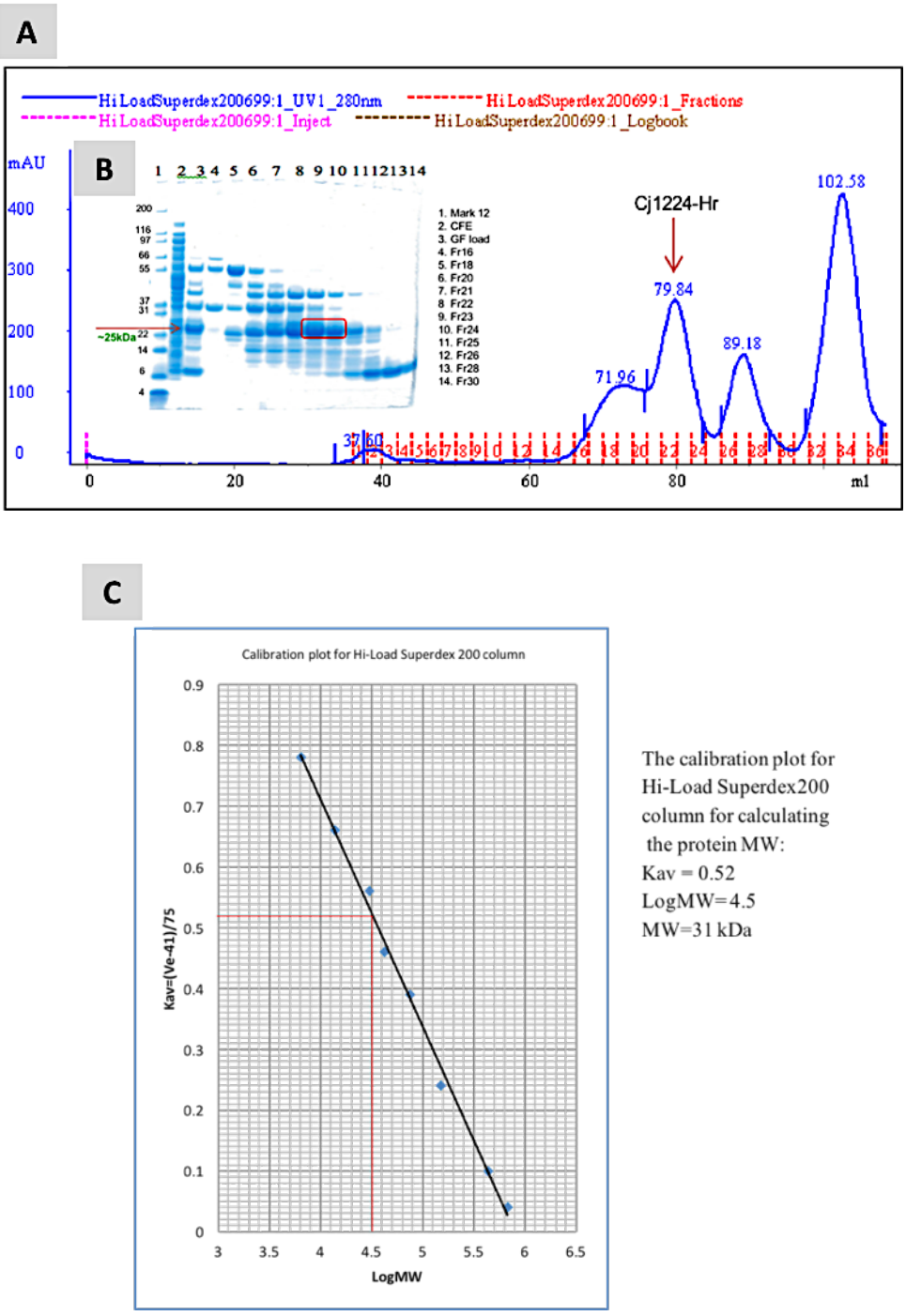


Figure 5.11: Gel filtration purification of untagged Cj1224-Hr. Panel (A) the chromatogram of gel filtration on a 1.6x60cm Hiload Superdex200 column loading and elution of Cj1224 in a peak at 79.84 ml. **Panel (B)** 12% SDS-PAGE (Nupage 4-12% BT protein gel) analysis of purification progress of Cj1224-Hr preparation. Fractions 23 and 24 (red square box) were combined with total protein was yield 2. 8mg.The Cj1224-Hr protein is shown (the arrow sign) with size ~25 kDa. **Panel (C)** The calibration plot for Hi-Load Superdex200 column. The calibration plot for Hi-Load Superdex 200 showed that the Cj1224-Hr corresponded to an apparent MW of 31 kDa, indicating its monomeric state in solution.

5.8.1.7 Purification of Cj1224-Hr with N-terminal His-tag

In this experiment two types of buffer systems were used; phosphate buffer and Tris buffer. The trials were attempted using the same protocol as explained in chapter (2) section (2.11.1 & 2.11.1.1). The progress of the purification was checked by SDS-PAGE analysis (Figure 5.12).

5.8.1.8 Crystallisation trials

For crystallisation trials, the N-terminal tagged Cj01224-Hr purified from either phosphate buffer or Tris buffer was buffer exchanged and concentrated to different concentrations; 5 mg/ml, 12 mg/ml, 18 mg/ml and 25mg/ml in 10mM Tris-HCl buffer pH (5,6,7,8) and 0.1M NaCl using Zeba spin-desalting column. Also, different commercial crystallisation condition screens were used including JCSG, PACT, PEG, MPD, Ammonium sulphate, Morpheus, Classics and Proplex (Qiagen/Molecular Dimensions). The plates were incubated at 17 $^{\circ}$C and also some plates at 4 $^{\circ}$C. Unfortunately, again no protein crystals were obtained from these trials.

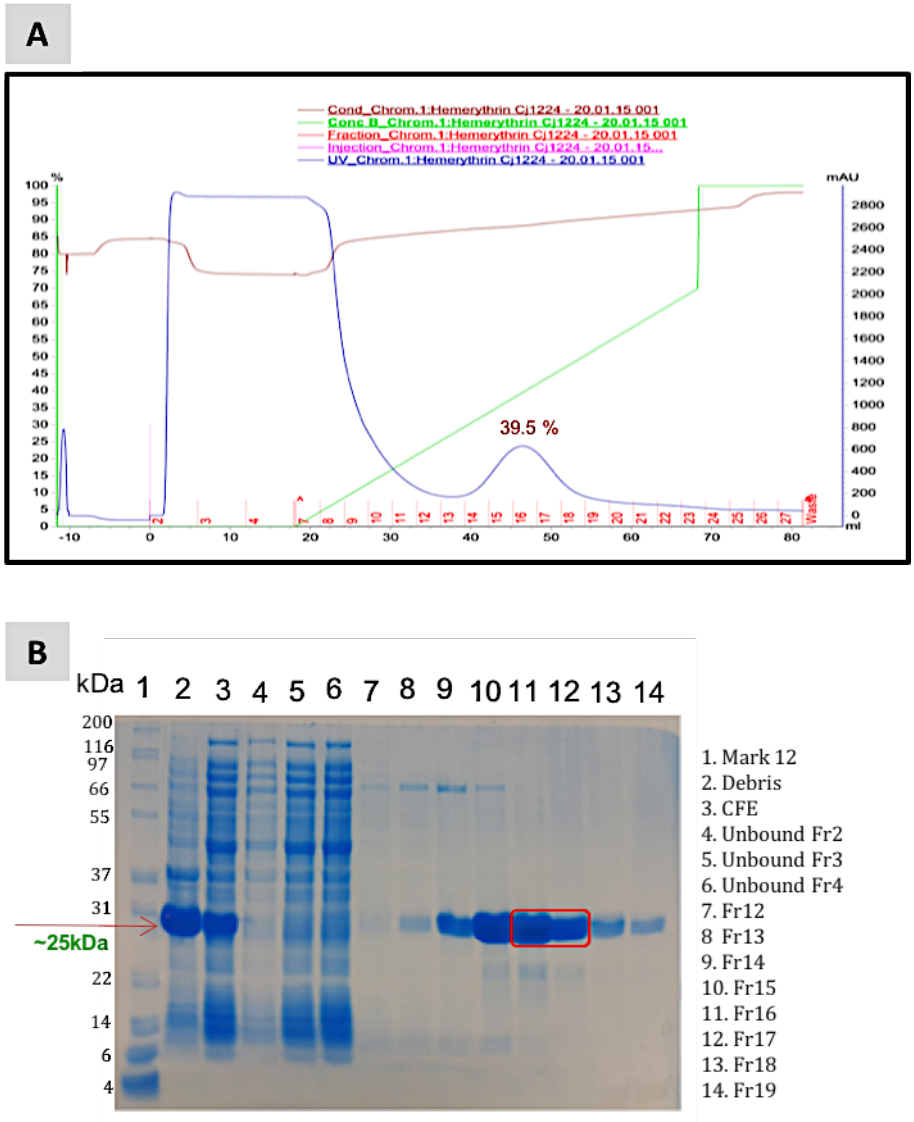


Figure 5.12: Purification of Cj1224-Hr protein with N-terminal His-tag using Tris buffer. Panel (A) The chromatogram of the purification of Cj1224-Hr via Nickel affinity column (HisTrap HP 5ml from GE Healthcare™). Following sample application and a column wash, an increasing gradient (0-70%) of 0.5M Imidazole was applied (green line). Cj1224-Hr was eluted at ~39.5 %, which corresponds to 0.20M Imidazole. **Panel (B)** 12% SDS-PAGE analysis of the fractions from the column. The Cj1224-Hr protein is shown (the arrow sign) with size ~25 kDa. Lane 1: Protein molecular weight marker (Mark 12^{TM}) lane 2: cell debris, lane 3: cell free extract, lanes 4-6: from washing fractions, and lanes 7-14: from elution fractions. Fractions 16 & 17were combined and then concentrated for crystallisation trials.

5.9 Inductively Coupled Plasma Mass Spectrometry (ICP-MS) analysis

The presence of metal ions bound to Cj1224-Hr was examined using inductively coupled plasma atomic emission-mass spectrometry analysis (ICP-MS). The analysis clearly demonstrated the presence of iron ions (Table 5.3). Also, the analysis showed a significant peak of zinc ions with Cj1224-Hr protein, which suggests this protein also could be binding with zinc ions as well as with the iron ions.

Table 5.3: ICP-MS result shows the presence of zinc ions beside iron ions in Cj1224-Hr Protein.

Sample	Pro.conc.	Fe	S	Zn	Ni
Buffer	0.00	<0.01	<0.01	<0.01	<0.01
Cj1224- N.term His-tag	0.08	0.08	0.55	0.05	0.00
Cj1224- C.term His-tag	0.08	0.08	0.58	0.05	0.00

*All results unit = mM

The structure of Cj1224-Hr (with C or N terminal tag) is expected to contain 2 iron ions and 9 sulfur atoms for each protein molecule. Also, the structure may contain 1 Zn atom, as there is possibly a zinc finger created with 4 Cys residues around a zinc ion. Therefore, to see how many Zn ions are associated with each molecule, the molar ratio of the protein to the metal was estimated. According to ICP-MS analysis results, the ratio of Fe: S should be approximately 1:5 with 2 iron atoms and 9 sulphurs. This is confirmed in the results of the ICP analysis but this also revealed the possible presence of one zinc ion.

This finding of zinc ions with Cj1224-Hr protein may open some important questions about the possible role of zinc ions in the function of the Cj1224-Hr, the possible interaction of zinc ions with the cysteine SH- groups and their potential influence on crystallisation as discussed below.

5.10 The Cj1224-Hr have possible zinc-binding site

The Zinc ion is one of the most common metal ions found to bind proteins in living organisms. It is the second most abundant ion found within cells behind iron ions when considering biologically related transition metals (Vallee and Auld, 1990). In proteins, a zinc ion is normally liganded by the amino acid side chains of glutamic acid, aspartic acid, histidine and cysteine (Auld, 2001). Cysteine residues are recognised to play vital functions in proteins, involving binding to many metal ions. It can particularly exhibit high affinity toward zinc ions (Zn^{2+}), and this subsequent Zn^{2+}- cysteine complex is an important of protein structure for stabilization, regulation and catalysis (Pace and Weerapana, 2014). Changes in zinc concentrations in cells can alter protein expression quantities and changeable binding of Zn to proteins plays a role in cell signalling (Halloran, 1993) . The Cj1224-Hr sequence contains four cysteine residues at the C-terminus. The existence of these cysteine residues may reflect a function in binding metal ions. The possible properties linked to the nature of the metal binding in Cj1224 could also mean this protein has a function in oxygen sensing in a similar way to other studied hemerythrins (Xiong *et al.*, 2000; Kao *et al.*, 2008; Schaller *et al.*, 2012). In addition, Kendall mentioned that binding free iron with CXC and CXXC motifs in the presence of oxygen molecules perhaps minimised the harm by the Fenton reaction (Kendall *et al.*, 2014). However, this study on Cj1224 using ICP-MS has provided some indication of possible zinc ion binding for these residues.

5.10.1 Cj1224-Hr structure has a predicted zinc finger

The structure prediction from the Swiss-Model server has suggested that the Cj1224-Hr has a 4-alpha helix bundle, and it is 32% identical with Cj0241-Hr (Figure 5.13).

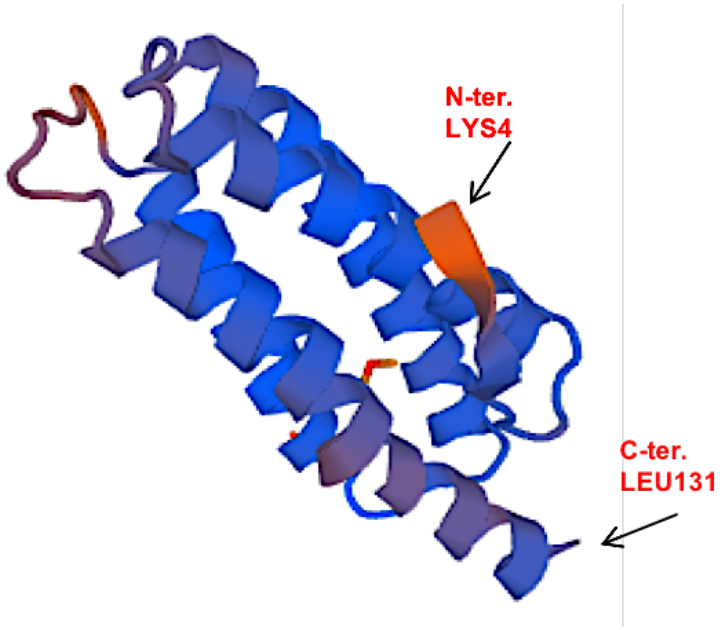


Figure 5.13: Predicted structure of Cj1224-Hr based on Cj0241-hr structure from Swiss model. Swiss-Model server suggests the Cj1224-Hr contains four alpha helices, consistent with a 32% sequence identity.

In addition, the C-terminal part of the structure, which contains the cysteine residues, has been suggested by the Phyre[2] server to be a zinc finger although the models have only low sequence identity and give low confidence values (Figure 5.14A). This is consistent with the structure of Cj1224-Hr protein binding with zinc ions as found from ICP-MS analysis. Zinc fingers usually have a Cys_2His_2 or Cys_4 motif in numerous proteins (Figure 5.14B), which binds zinc to stabilise the structure (Andreini *et al.*, 2006). Even though the functional roles of many zinc finger proteins are not much understood, many acts as transcription activators, regulatory subunits, or have catalytic functions (Razin *et al.*, 2012). Additionally, they may act as protein: protein interaction domains. For instance, cysteines in one of the two Cys_2-Cys_2 type of zinc fingers in Fep1 protein, an iron–responsive GATA factor, can also work as ligands for a different Fe-S cluster in aerobically purified Fep1 to replace its normal Fe-S cluster (Antimo *et al.*, 2016). Thus a zinc finger motif could have dual roles that are controlled by the O_2 concentration. As mentioned previously, there is evidence that the hemerythrins in *Campylobacter jejuni* make direct interaction with Por and Oor enzyme proteins, which contain important Fe-S clusters in the citric acid cycle (Kendall *et al.*, 2014) and which might possibly support the hypothesis of a zinc finger in Cj1224-Hr with a role to provide an alternative Fe-S cluster when under oxygen stress.

Interestingly, a number of other residues identified as being conserved in the alignment (Tyr158, Ile186, Val192) are packed together on the structure and adjacent to the zinc finger and its associated cysteines (Figure 5.14C). The observation of this apparent clustering adds weight to the suggestion that this protein contains a zinc finger domain but clearly only a full structure determination will confirm this proposal.

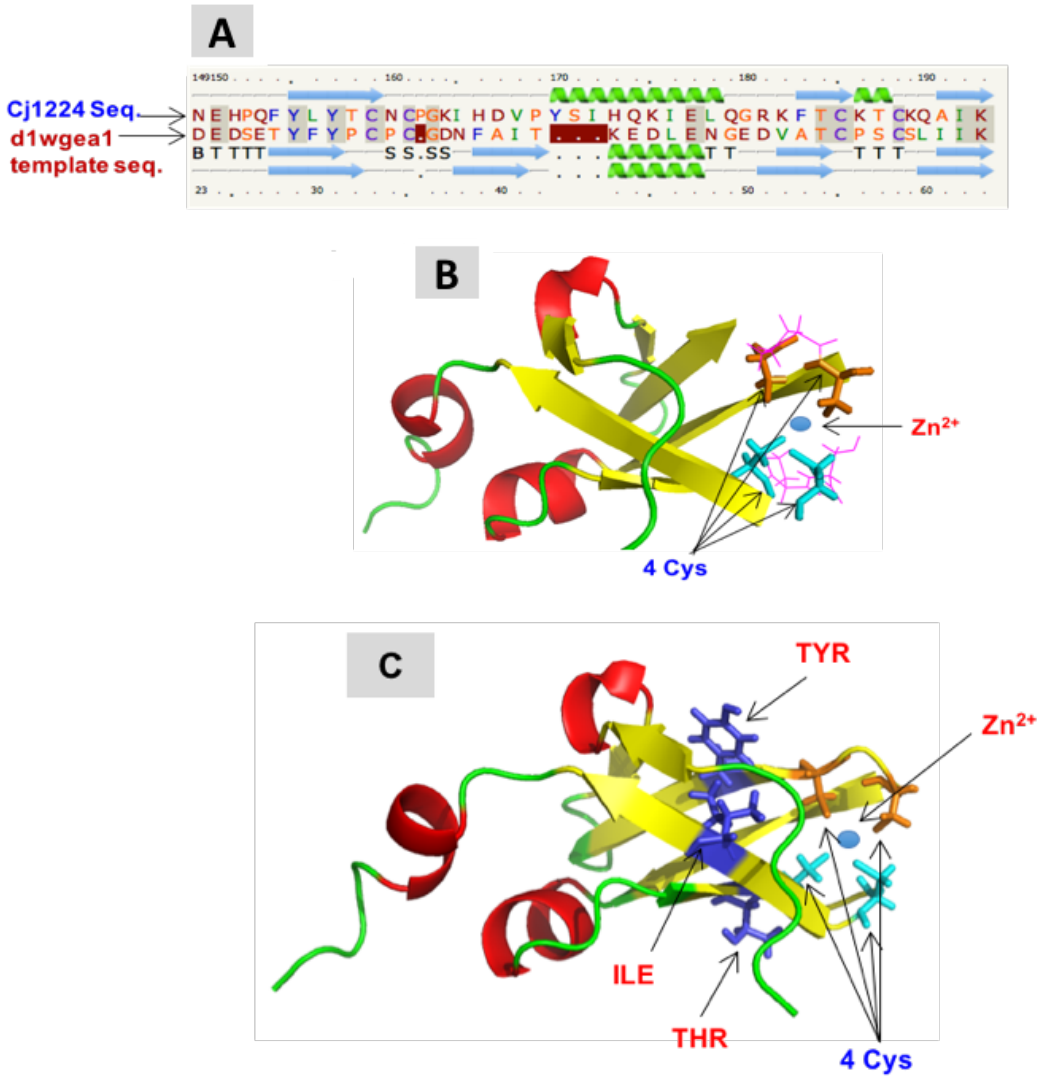


Figure 5.14: Possible zinc finger structure formed by the C-terminal region of Cj1224-Hr. An alignment sequence of Cj1224-Hr with mouse DESR1 protein (PDB entry 1wge) and a cartoon drawing showing the possibility of a zinc ion in Cj1224-Hr structure binding with the 4-cysteine amino acids. **Panel (A)** shows the sequence alignment of the C-terminal part of Cj1224 and the 1wge which has 29% identical sequence. **Panel (B)** shows the predicted structure of the C-terminal region in Cj1224 as modelled by the Phyre2 server. It shows the 4 Cys residues binding with a zinc ion. **Panel (C)** show a number of residues identified as being conserved in the alignment (Tyr158, Ile186, Val192) are packed together on the structure and adjacent to the zinc finger and its associated cysteine.

5.10.2 Cj1224-Hr protein possibly is capable of binding to DNA fragments

The zinc finger domains are normally found in multiple groups within a single protein, and can be stabilised by interaction with other proteins and biomolecules such as RNA and DNA (Pace and Weerapana, 2014). These interactions are perhaps involved in a range of biological processes such as DNA repair, DNA replication, transcriptional regulation and DNA packing (Ofran *et al.*, 2007). To study the ability of Cj1224-Hr to bind with single and double stranded DNA, the electrophoretic mobility-shift assay (EMSA) was used on a 1% agarose gel as described in chapter (2) section (2.11.15). In Figure (5.15) DNA probes (single / double stranded) without protein migrated as a single band. However, upon addition of protein, this band migrated more slowly through the gel and supports the idea that the Cj1224-Hr protein possibly is capable of binding to the DNA fragments. This further supports the claim of Cj1224 forming a zinc finger domain(Si *et al.*, 2015). The zinc finger from the predicted structure of Cj1224 is located at the C-terminus. Thus, there are a number of possible functions of the zinc finger in Cj1224-Hr: acting as a binding domain with DNA fragments perhaps for the sensing of oxygen levels; acting as a protein interaction domain, such as interacting with Por and Oor as suggested by Kendall (Kendall *et al.*, 2014) and thus perhaps providing an alternative Fe-S cluster as seen with Fep1; a totally different and as yet uncharacterised Function (Antimo *et al.*, 2016). In addition, from the model of its structure, Cj1224-Hr is predicted to contain six solvent tunnels from the outside to its di-iron centre (Figure 5.16), which are different from those in Cj0241, using the prediction tool CAVER 3.0 with probe radius 0.8.

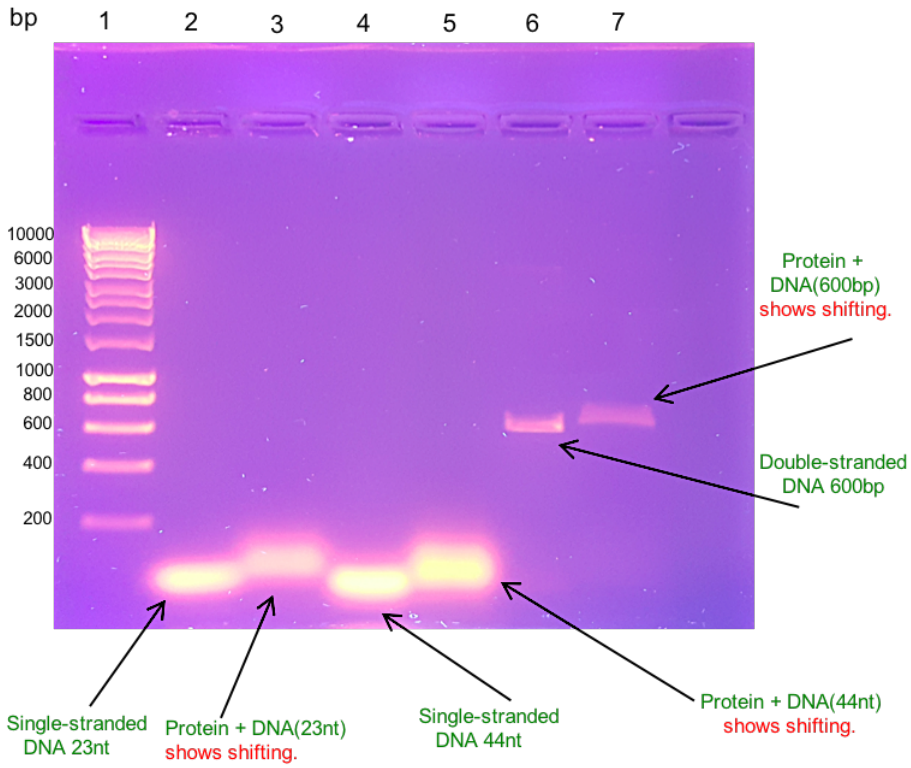


Figure 5.15: Single and double stranded DNA electrophoretic mobility-shift assay (EMSA) using Cj1224. The non-specific binding of Cj1224-Hr to DNA was tested and analysed on a 1% agarose gel. Lane 1 is DNA hyper ladder™ 1kb molecular weight marker (BIOLINE). Lanes 2 & 4 DNA single stranded without Cj1224-Hr protein (23 & 44 bp respectively). Lanes 3 & 5 DNA single stranded (23 & 44 bp respectively) with Cj1224-Hr protein, which exhibited shifted band positions. Lane 6 DNA double stranded (600bp) without Cj1224-Hr protein. Lane 7 DNA double stranded (600bp) with Cj1224-Hr protein, which exhibited a shifted band position.

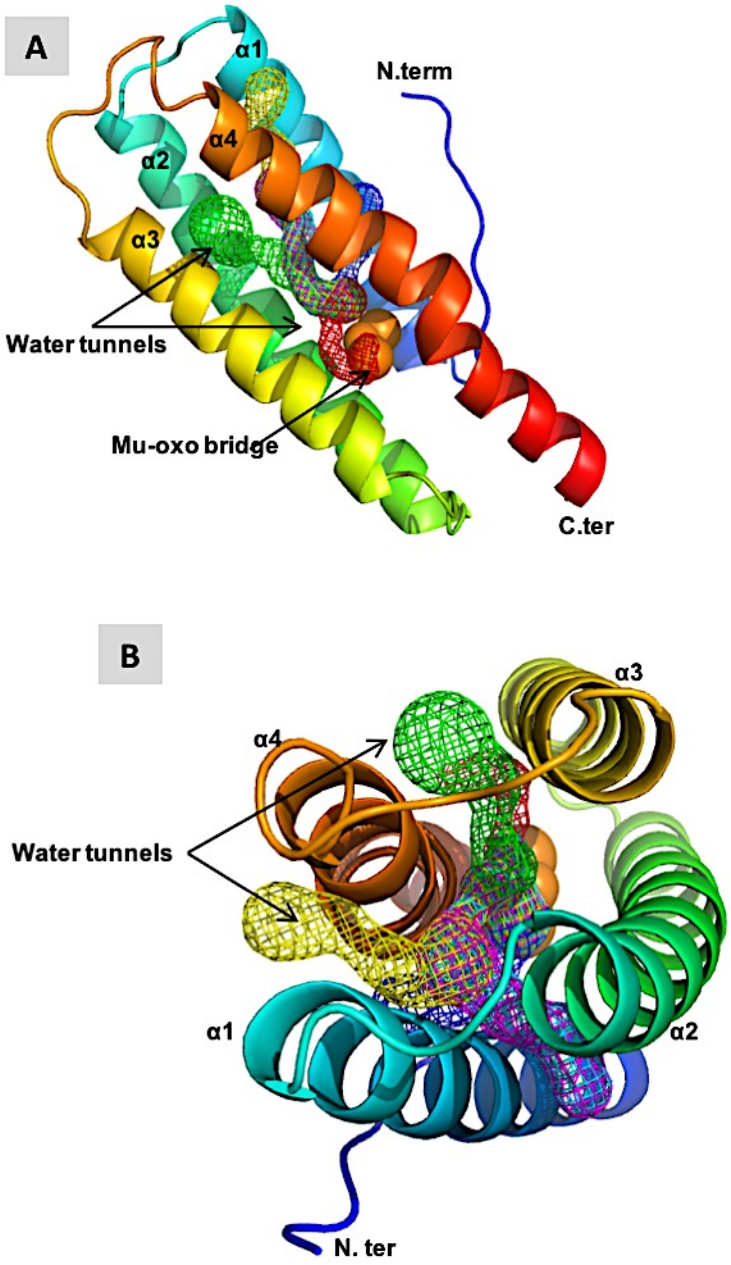


Figure 5.16: A Prediction of the solvent tunnels in a Cj1224-Hr protein structure model. The tunnels were calculated using the CAVER 3.0 with standard probe radius 0.8 which is plugin to PyMol. **Panel (A)** shows a side view of Cj1224-Hr structure which comprises a four- helix bundle (rainbow colour scheme), a non-heme di-iron centre (orange spheres) with six solvent tunnels (shown in different coloured mesh). **Panel (B)** shows a top view of the structure with solvent tunnels. The tunnels were found oriented parallel and perpendicular with the long axis of the four-helix bundle.

5.11 Investigation of Zinc binding to Cj1224-Hr

To confirm that the cysteine residues possibly bind with zinc ions, cysteine residues of Cj1224-Hr with an N-terminal His-tag were changed to alanine. The 4 cysteine residues were divided into two groups: first group contains Cys179 & Cy181; the second group contains Cys205 &Cys208 (Figure 5.17). Single and double mutations were performed with both groups using the site-directed mutagenesis procedure from NEB as described in chapter (2) section (2.11.12). The sequencing confirmed the mutation in each group as explained in chapter (5) section (5.1). Also, the constructs were transformed into *E. coli* and small-scale expression testing was performed at different temperatures (16°C, 25°C, 37°C), to check the solubility of Cj1224-Hr after the mutations were created as explained previously in section (2.5.1). The purification was done using the protocol as described in chapter (2) section (2.10.1). The SDS PAGE analysis (Figure 5.18A) from the double mutations (Cys179Ala & Cys181Ala) reveal that this new protein was insoluble. This could be because the protein unfolded as a result of the mutations. Also, the zinc ions might be acting as a centre around which the C-terminal part folded as previously shows in Figure (5.14b). Removing the 2 cysteines and potentially therefore no zinc binding, led the C-terminal part to cause the protein to unfold and resulting in the protein aggregating then precipitating.

Single mutation was done also with Cys179 and the SDS PAGE analysis showed a little amount of soluble protein from the single mutation at 25°C at the expected molecular weight of approx. 25kDa (Figure 5.18B).

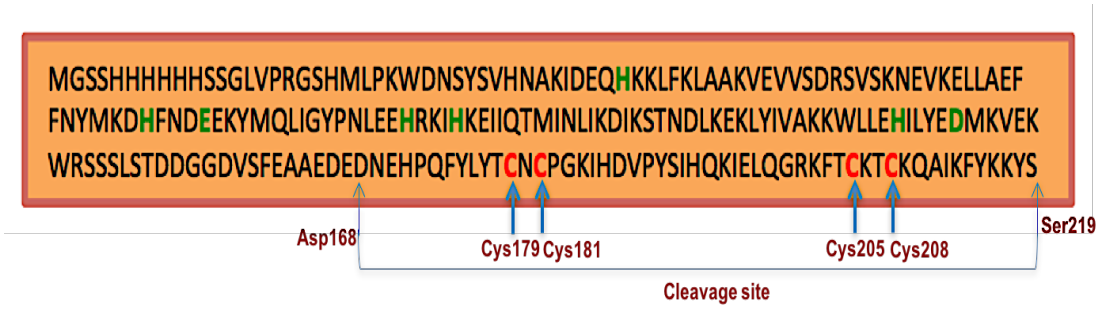


Figure 5.17: Sequence of Cj1224-Hr with N-terminal His-tag. Green highlighted region indicates the histidines involved in Fe binding. The sequence also contains 4 cysteine residues (red highlighted regions) which are divided into two groups; first group is Cys179 & Cys181 and the second group is Cys205 & Cys208. Single and double mutations were performed with these two groups. In addition, the blue arrow indicates the portion of the sequence from ASP168 to SER219, which was not seen in an initial MS/MS analysis.

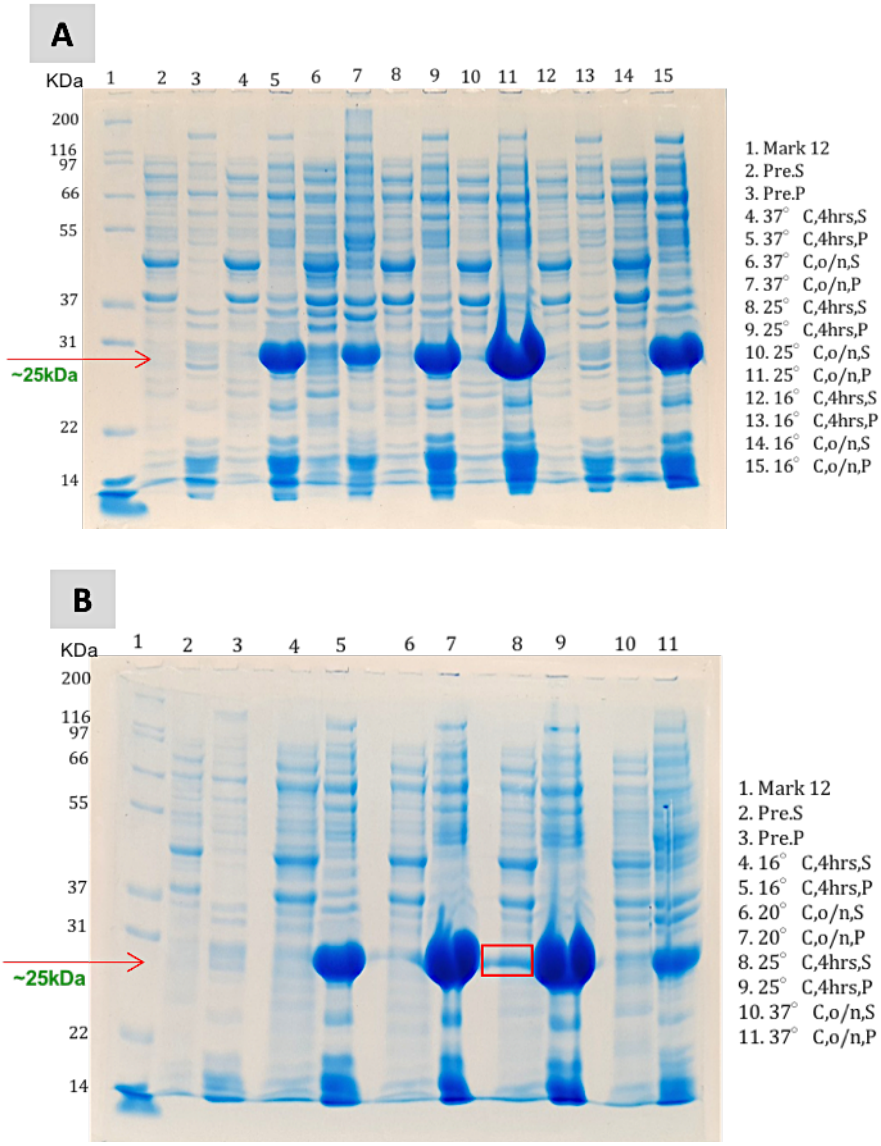


Figure 5.18: SDS-PAGE analysis of small-scale expression and solubility of Cj1224-Hr mutated variants at different temperatures. Panel (A) shows the SDS-PAGE of Cj1224-Hr testing for expression and solubility with double mutated variant (Cys179Ala & Cys181Ala), which displayed no soluble protein. **Panel (B)** shows the SDS-PAGE of Cj1224-Hr testing for expression and solubility with single mutation (Cys179Ala), which displayed a little amount of soluble protein (red square) at 25°C, 220rpm shaking.

However, the SDS PAGE analysis of the chromatogram from the purification exhibited a band at less than the expected molecular weight of the Cj1224-Hr mutated variant protein (Figure 5.19B). Nevertheless, some of the purified Cys179Ala variant protein was sent to ICP-MS to check for the presence of metal ions, especially zinc ions, and some was used in crystallisation experiments with various screening kits. Additionally, some was sent for intact protein MS analysis to check the molecular weight. Unfortunately, the crystallisation trial did not display any crystal forms. The MS result showed the molecular weight to be 20919.83 Da, as expected from SDS-PAGE but shorter than full length Cj1224. Further, in contrast an MS/MS sequencing analysis of digested protein confirmed the protein identity with a 90.9% sequence coverage, including all the fragments of Cj1224-Hr protein and the mutation site (Figure 5.20).

One possibility to explain these differences in molecular weights is that the Cj1224-Hr was proteolysed by endogenous proteases from *E.coli* perhaps into two fragments: 5 kDa & 20 kDa. In SDS -PAGE and intact MS analysis only 20 kDa fragments would be detected, however, in MS/MS analysis after protein digestion, all peptides from both fragments would be detected. However, this does imply that the small C-terminal domain remains associated with the large N-terminal helix bundle after proteolysis and through the purification on the Ni column. The small fragment has perhaps been lost from the bottom of the gel in Figure (5.19).

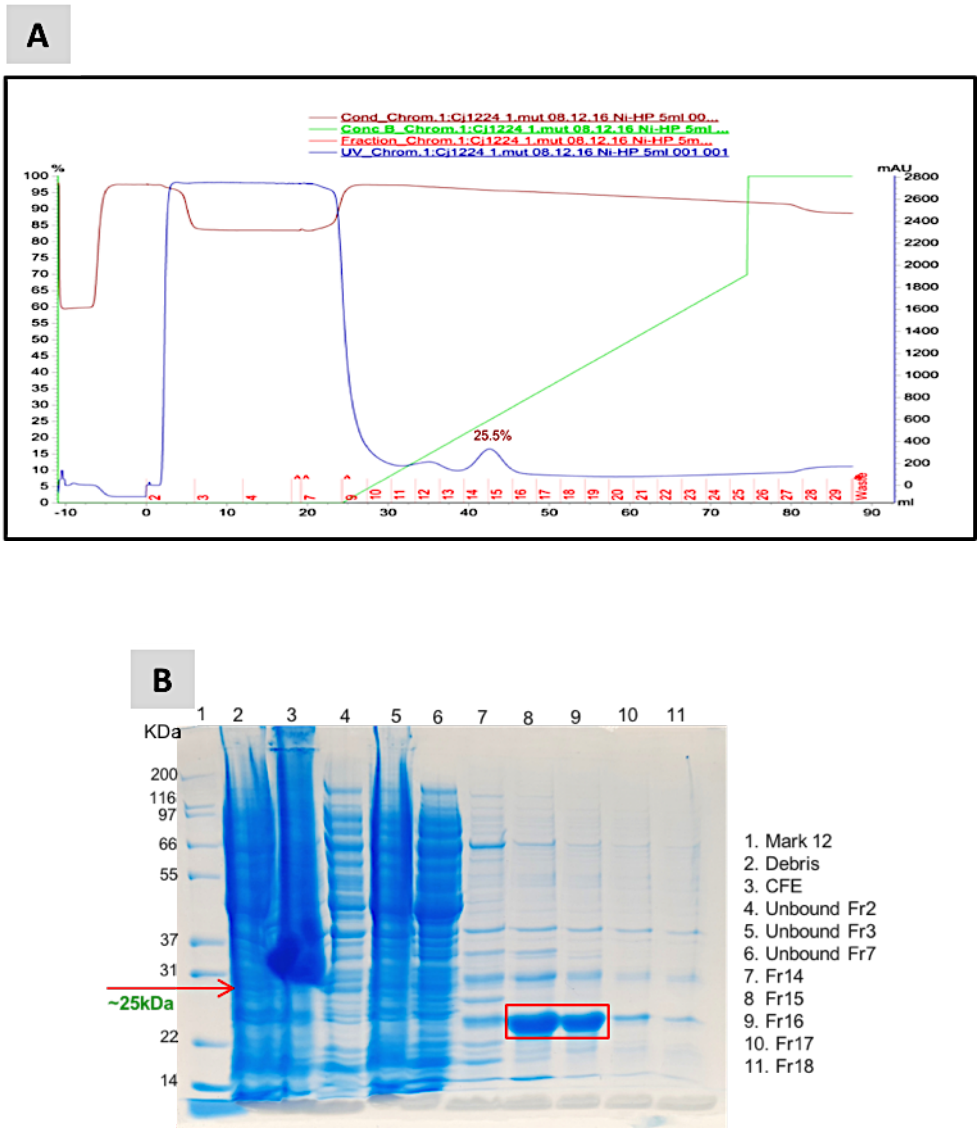


Figure 5.19: The purification of N-terminal His-tag Cj1224-HR single (Cys179Ala) mutated variant. Panel (A) the chromatogram of the purification of Cj1224-Hr via Nickel affinity column (HisTrap HP 5ml from GE Healthcare™). Following sample application and column wash, an increasing gradient (0-70%) of 0.5M Imidazole was applied (green line). Cj1224-Hr was eluted at ~25.5 %, which corresponds to 0.13M Imidazole. **Panel (B)** shows an SDS-PAGE analysis of the eluted fractions. Lanes 8 & 9 (red square) show the thick bands less than the expected molecular weight of Cj1224-Hr as seen in the CFE in lane 3.

1	Sequence
2	MLPKWDNSYSVHNAKIDE
3	GSSHHHHHHSSGLVPRGSHMLPKWDNSYSVHNAKIDE
4	LPKWDNSYSVHNAKIDE
5	QHKKLFKLAAKVE
6	QHKKLFKLAAKVEVVSDRSVSKNE
7	VVSDRSVSKNE
8	FFNYMKDHFNDE
9	FFNYMKDHFNDEE
10	FFNYMKDHFNDEEKYMQLIGYPNLE
11	EKYMQLIGYPNLE
12	EKYMQLIGYPNLEE
13	KYMQLIGYPNLE
14	KYMQLIGYPNLEE
15	IIQTMINLIKDIKSTNDLKE
16	KLYIVAKKWLLE
17	DMKVEKWRSSSLSTDDGGDVSFE
18	KWRSSSLSTDDGGDVSFE
19	KWRSSSLSTDDGGDVSFEAAEDE
20	DNEHPQFYLYT**ANC**PGKIHDVPYSIHQKIE
21	HPQFYLYT**ANC**PGKIHDVPYSIHQKIE
22	LQGRKFTCKTCKQAIKFYKKYS

Figure 5.20: MS/MS analysis for N-terminal His-tagged Cj1224 Cys179Ala variant. The MS/MS analysis showed all the fragments including those containing the mutation site for Cys179Ala (CNC → ANC), lines 20&21. This region of the protein was detected when performing the digestion in solution using GluC digestion buffer where the cleavage occurs between Glu and Asp residues.

The ICP-MS result showed there were still iron ions in the protein sample, but no zinc ions comparing with the wild type (Table 5.4). Regardless of the mismatch between the numbers of atoms within the protein molecule, there is no zinc ion associated with the protein molecule. Consequently, this may be further possible evidence that the wild type of Cj1224-Hr is binding zinc ions by its cysteine residues besides binding with iron ions. However, as can be seen from table (5.4) the ratio of iron to sulphur is approximately 1:1 unlike earlier experiment where a ratio of 1:9 is seen (see table 5.3). this suggests the experiment describe above its to be repeated as it may be unreliable.

Table 5.4: ICP-MS results show no zinc ions with Cj1224-Hr Cys179Ala variant

Sample	Pro.conc.	Fe	S	Zn	Mn	Ni
Buffer	0.00	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
Cj1224-N.terminal His-tag with 1_mutation	0.07	0.06	0.06	0.00	0.00	0.00

All results unit = mM

Another round of mutagenesis was performed in which the other pair of cysteine residues Cys205 & Cys208 (second group) were subjected to the same procedures that have been done with the first group of cysteine residues Cys179 & Cy181. The experiments included the site-directed mutagenesis procedure from NEB, sequencing analysis, transformation, checking the solubility and purification trials. From the purification experiments in both single and double mutations, the SDS-PAGE showed the same results as above. However, in the SDS-PAGE analysis of the single mutation (Cys205Ala) as can be seen in Figure (5.21) there were two bands; the thin band was displayed at a molecular weight of approx. 25kDa and the thick band at approx. 20kDa. These two bands were cut directly from the SDS-PAGE gel and analysed by the MS/MS method. The results showed that the thin band at 25 kDa came from a contaminating protein of *E.coli*.

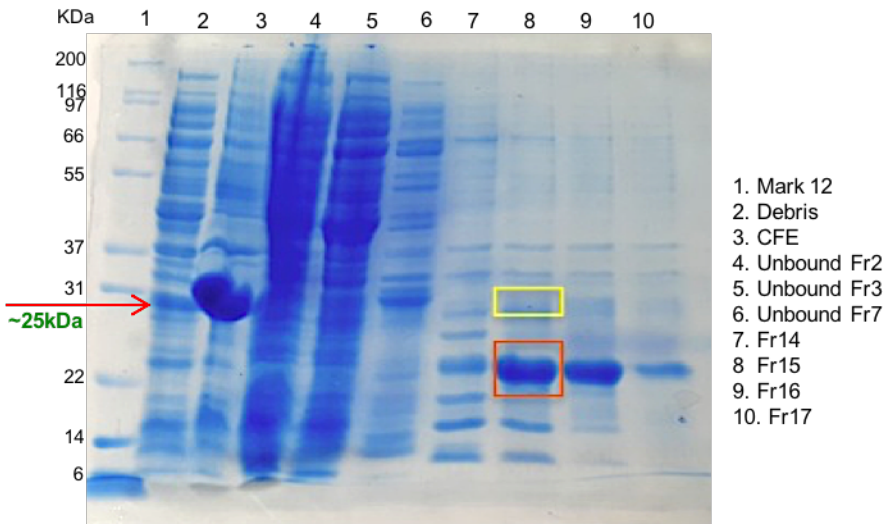


Figure 5.21: SDS-PAGE of the eluted fractions from the purification of N-terminal His-tagged Cj1224-Hr variant (Cys205Ala). Two bands are seen; a thin band at approx. 25kDa (yellow square) and a thick band at approx. 21kDa (red square). Both bands (thin & thick) were analysed by MS/MS. The thick band was confirmed to be Cj1224-Hr, however, the thin band came from an *E. coli* contaminant. The red arrow is the expected molecular weight of Cj1224-Hr as seen in the CFE (lane 3).

The thick band with a molecular weight of approx. 20 kDa belonged to the N-terminal sequence of Cj1224-Hr protein, but the fragment with the mutation at the C-terminus was not detected in the sample. This is in contrast with the results from the previous experiment where the mutation site was found by MS/MS. In this second analysis, the digestion was carried out in the gel rather than on a sample in solution. The gel would separate the fragments generated by endogenous *E. coli* proteases. The molecular weight of the thick band would correspond to about the N-terminal 170 residues and thus excluding the mutation site.

5.12 UV visible spectrophotometric analysis of Cj1224-Hr

The UV-visible spectrophotometric analysis of hemerythrin proteins in general show two major peaks at approx. 336 nm and 375 nm and a weaker peak at 500 nm, like Cj0241 showed previously in Figure (5.5). These features are related to the di-iron centre absorbance, and thus characteristic of the met form of the hemerythrins in *C. jejuni* (Kendall *et al.*, 2014). An attempt was made to measure the autoxidation rate of Cj1224 hemerythrin like Cj0241-Hr, which might give some information about the hemerythrin function such as to whether it is linked with the storage, transport or sensing of oxygen. All attempts failed to give clear results because the Cj1224 tends to precipitate. The reason behind this could be most likely associated with the sensitivity of the Cj1224 hemerythrin to the sodium dithionite (or sodium borohydride) like the Cj0241. Furthermore, an enzyme based system for depleting the oxygen level in solution with Cj1224-Hr protein was also used as explained in chapter (2) section (2.14), but still the UV spectrum showed only the met form instead of the de-oxy form. However, because of the potential sensitivity of Cj1224-Hr to sodium dithionite, a serial dilution of sodium dithionite was made to find which concentration might be possibly starting to affect the

hemerythrin protein. The experiment was conducted as explained in chapter (2) section (2.13). Therefore, concentrations of sodium dithionite as low as approx. 0.5mM gave the Cj1224-Hr spectra that showed more convincing loss of the peaks for the de-oxy form than Cj0241-Hr. The concentration and hence the effect of the sodium dithionite was followed by its peak in the spectrum at 309 nm, as showed previously in chapter (4) section (4.13.6), and the spectra were expected to belong to the deoxy-hemerythrin stage. At the point when the dithionite had apparently all gone, the cuvette lid was opened and the peaks in the spectral curve gradually increased for 2 minutes at the points expected for the oxy-hemerythrin stage, after which the hemerythrin was expected to return to the met-form by auto-oxidation. The curve returned to the met-form and remained for around 6 minutes. After that the protein started to aggregate and precipitate (Figure 5.22).

The aim of this experiment was to measure the exact rate of the autoxidation, but this was very challenging to conduct because the problem of the aggregation and precipitation of the hemerythrins proteins. However, although the exact rate could not be measured, it appeared that the rate occurred in minutes rather than seconds, as expected for a sensor protein as with the domain in DcHr (Chen *et al.*, 2015), or hours like the marine invertebrate hemerythrins (Farmer *et al.*, 2001).

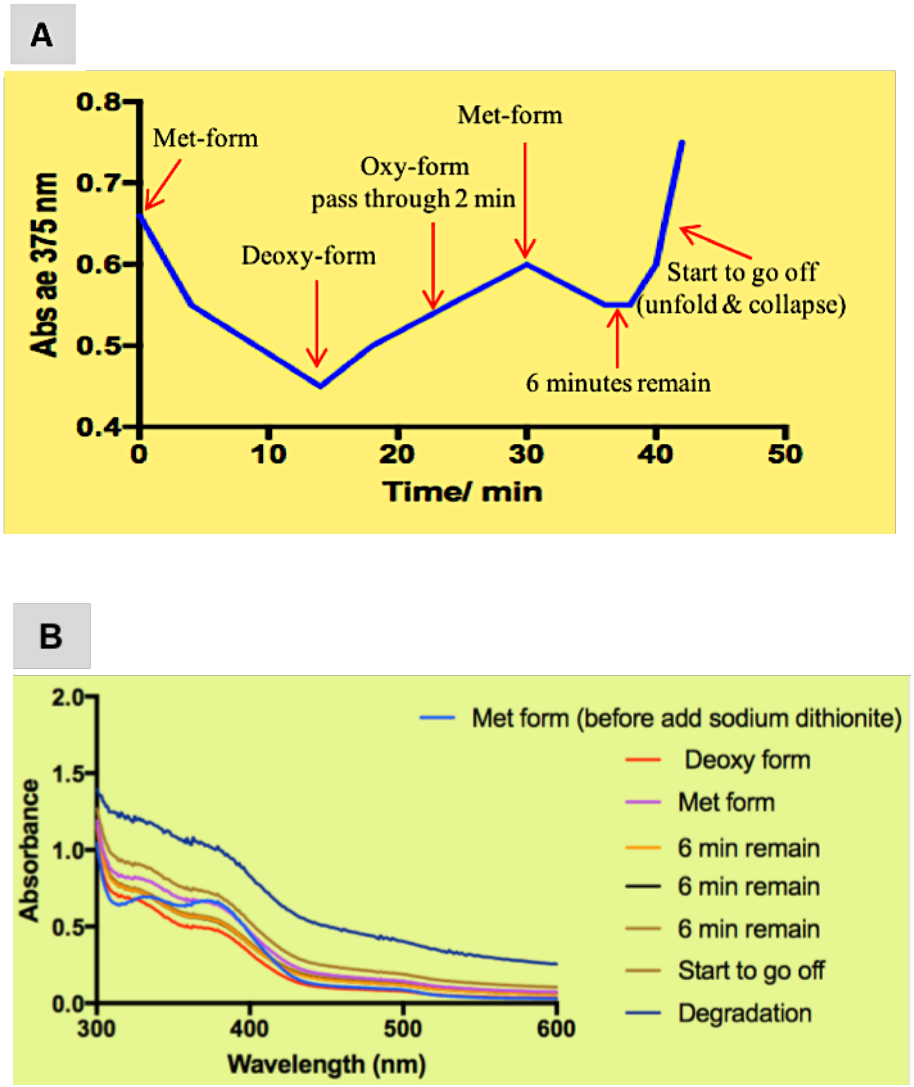


Figure 5.22: Cj1224-Hr UV-vis Spectroscopy analysis. Panel (A) peak kinetics at 375 nm shows the interconversion of the 3 forms of hemerythrin. **Panel (B)** The full UV-visible spectra.

Chapter 6

Cj0045 Hemerythrin Expression, Purification and Crystallisation

This chapter will describe the work done with Cj0045-Hr protein. More information is found in chapter (1), section (1.8.6 / 1.8.7). Similar methods were used for the two hemerythrins Cj0241 and Cj1224. Cj0045-Hr is larger when compared with the two other hemerythrins Cj0241 and Cj1224. It has 240 residues in length and around 30kDa molecular weight. Also, it has additional metal binding motifs like Cj1224, as well as two extra motifs characteristic of di-iron centre binding hemerythrins (Figure1.5). From previous studies, there have been several attempts to get an amount of soluble protein *in vitro* from this type of hemerythrin, but all were not successful (Kendall *et al.*, 2014). However, a further attempt was made to get soluble protein of Cj0045-Hr, and this is described in more details in this chapter.

6.1 Confirming insertion of *cj0045* gene in the pET-21b plasmid

The *cj0045* gene that had been inserted into pET-21b plasmid was validated by DNA sequence analysis after extraction of the DNA construct using a Miniprep kit (See appendix). A small amount of the DNA (5µl with 5 µl T7F or T7R) was sent to GATC Biotech for sequencing. This was followed by an analysis using the Basic local alignment search tool (BLAST) (Altschul *et al.*, 1990) at the National Centre for Biotechnology Information http://blast.ncbi.nlm.nih.gov/Blast.cgi. It showed 100% identity with the sequence of Cj0045, as shown in chapter (1) section (1.8.6).

6.2 Testing the expression of recombinant His Tag- Cj0045-Hr protein

Cj0045-Hr protein as mentioned previously in chapter (1) section (1.8.7) was produced as insoluble protein under routine purification conditions; 25℃ and 37℃ with 1mM IPTG in a shaker incubator rotation of 250 rpm. The attempts have included various auto-induction media. Also, other overexpression strains such as Origami B from *E.coli* were tested, and the LB media was supplemented with iron ions to support these bacteria. In addition attempts have been made to use the technique of protein denaturation by urea and then refolding, but this way was not successful and it was decided that this particular method was unsuitable for this particular protein (Kendall *et al.*, 2014). In the studies described here, variations in the IPTG concentrations, growth temperatures and the rotation frequency have been tested to obtain soluble protein (Figure 6.1). Interestingly, some soluble protein was found for cultures grown at 15 °C with 1mM IPTG induction at 150 rpm and 200 rpm speed of rotation (Figure 6.2). Consequently, large-scale overexpression cultures were grown under these conditions to produce soluble Cj0045-Hr protein for the following experiments.

6.3 Purification of His-tagged Cj0045 hemerythrin

Approximately 5.6 g of cell paste was used to produce the Cj0045, and the experiment was conducted using the same approach taken with Cj0241 and Cj1224, chapter (2) section (2.10.1). The protein sample concentration was measured in the Cell Free Extract (CFE) using a Bio-Rad protein assay, and it was approx. 13 mg/ml in the supernatant fraction. Therefore, the total protein in the CFE was 226 mg. This sample was then loaded onto a 5 ml Nickel affinity chromatography column at room temperature and equilibrated with lysis buffer A. The column was washed with the lysis buffer to discard any unbound protein.

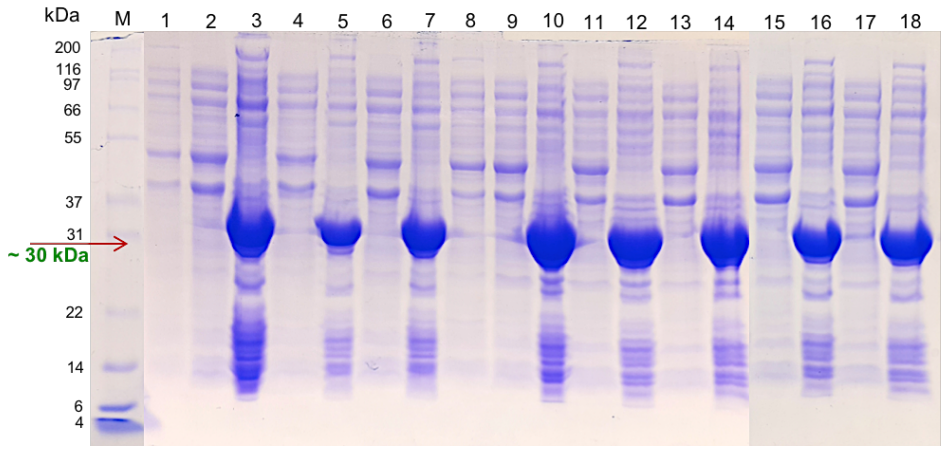


Figure 6.1: SDS-PAGE gel analysis of the expression of recombinant His-tagged Cj0045-Hr protein at different temperatures and concentrations of IPTG, at 250rpm shaker incubator speed. The gel shows all Cj0045-Hr protein in the insoluble fractions. Lane 1: molecular weight marker and red arrow indicates approximate 30 kDa MW region of Cj0045-Hr protein. Lane 2: soluble pre-induction sample. Lanes 3,4: soluble and insoluble, 17°C, 1mM IPTG. Lanes 5,6: soluble and insoluble, 17°C, 0.5mM IPTG. Lanes 7,8: soluble and insoluble, 25°C, 1mM IPTG. Lanes 10,11: soluble and insoluble, 25°C, 0.5mM IPTG. Lanes 12,13: soluble and insoluble, 30°C, 1mM IPTG. Lanes 14,15: soluble and insoluble, 30°C, 0.5mM IPTG. Lanes 16,17: soluble and insoluble, 37°C, 1mM IPTG. Lanes 18,19: soluble and insoluble, 37°C, 0.5mM IPTG. All growth was carried out using LB. Lanes 3-18 show post-induction samples at the stated conditions.

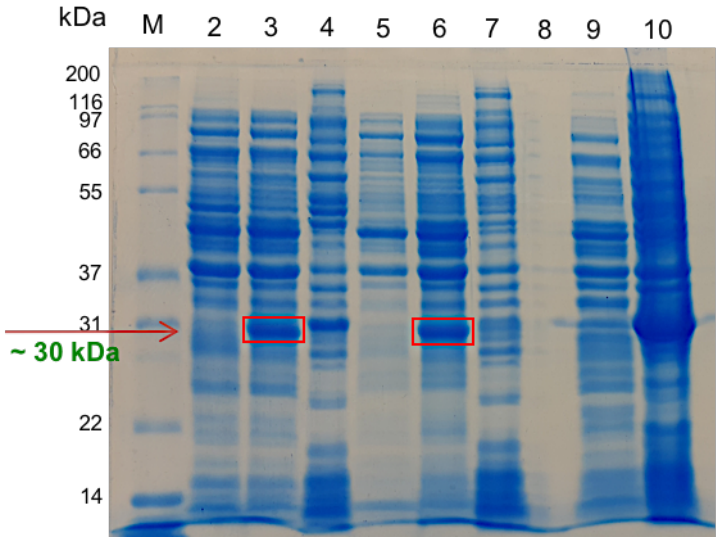


Figure 6.2: SDS-PAGE analysis of growth at 15°C. The Cj0045-Hr appeared in soluble fractions (red square) at 15 °C at incubator rotation speed 150 rpm & 200 rpm with 1mM IPTG induction. Lane 1: molecular weight marker, and red arrow indicates approx. 30 kDa MW region of Cj0045-Hr protein. Lanes 2,5,8 : soluble pre-induction samples. Lanes 3,4 : soluble & insoluble culture, 15°C, 1mM IPTG (150rpm). Lanes 6,7 : soluble & insoluble culture, 15°C, 1mM IPTG (200 rpm). Lanes 9,10 : soluble & insoluble culture, 37°C, 1mM IPTG (250rpm). The rotation speeds of the incubator are given in parentheses.

The recombinant hemerythrin Cj0045 protein was then eluted from the column using an increasing gradient of buffer B (elution buffer) containing 50 mM Tris-pH 8.0, 0.5M NaCl and 500 M imidazole. From the chromatogram, unexpectedly, as can be seen from Figure (6.3A), the recombinant Cj0045-Hr protein was not eluted in the gradient from the nickel column. Therefore, fractions were analysed by SDS-PAGE to check for the desired protein by molecular weight. Surprisingly, the majority of the Cj0045-Hr protein was released in the flow-through fractions 4-8 (Figure 6.3B). This means the recombinant Cj0045-Hr protein did not bind with the nickel column, even though it contained a six-histidine tag at the C-terminal end. It was suspected that the Tris buffer was possibly unsuitable. Consequently, a second attempt was made to purify the Cj0045-Hr using a different buffer; phosphate saline buffer instead of Tris buffer was used. However, this second attempt was still unsuccessful.

There are several possible explanations for this situation. It might be that the His-tag is combined into the structure of the protein and inaccessible to the Ni column, or it may be the tag is cleaved off during sample preparation. A possible reason for an inaccessible tag is that the His-tag could be interacting with metal ions within the structure and not binding to the column. Therefore, fractions 4 - 8 that contained Cj0045-Hr protein were combined, and the purification repeated but this time using an ion-exchange method, followed by gel filtration as explained previously in chapter (2) section (2.12.1). The Cj0045 protein sample was applied onto a Superdex 200 gel filtration column, 1.6x60 cm (GE healthcare life science), in 0.5M NaCl 50mM Tris pH 8, and Cj0045-Hr protein was eluted using the same buffer. Fractions were analysed by 12% SDS-PAGE to check for the presence of the desired protein by molecular weight.

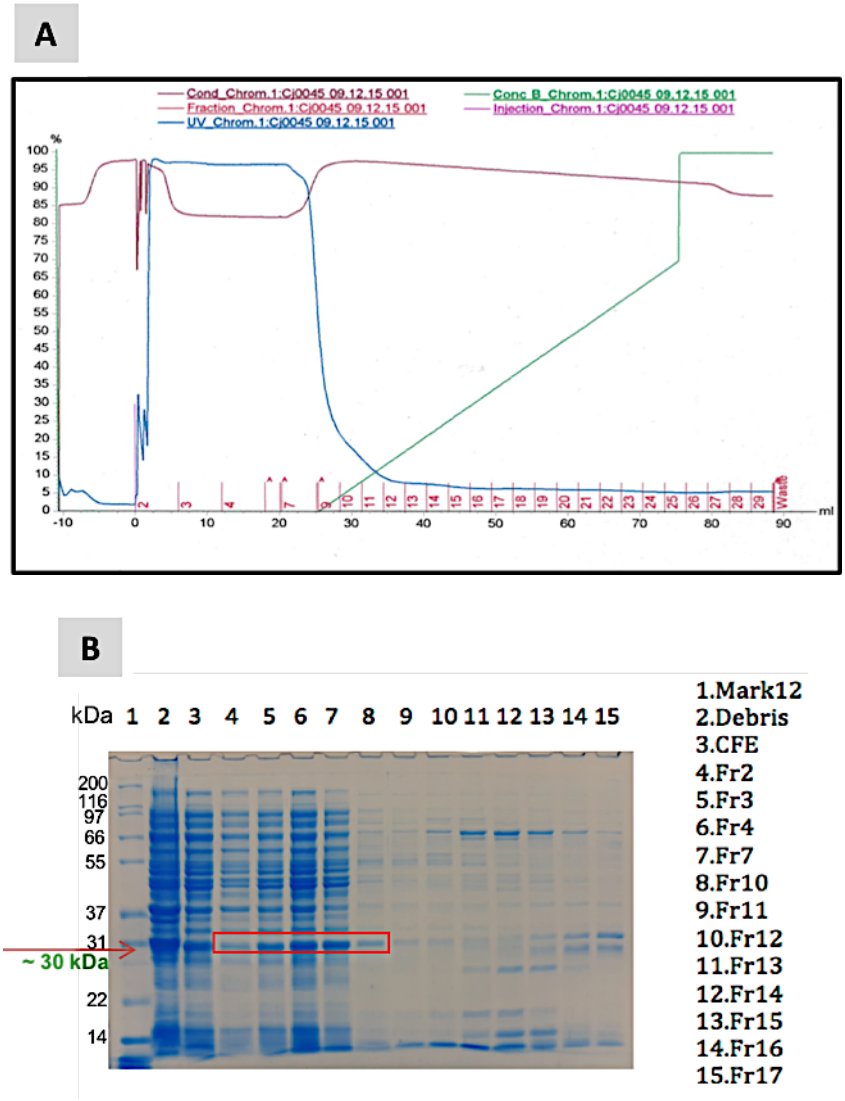


Figure 6.3: The Chromatogram obtained via Nickel affinity column for Cj0045-Hr protein, and SDS-PAGE. Panel (A) The chromatogram of the purification of Cj0045-Hr via Nickel affinity column (HisTrap HP 5ml from GE Healthcare™) shows Cj0045-Hr protein did not elute in the increasing imidazole gradient. **Panel (B)** 12% SDS-PAGE analysis shows that most of the Cj0045-Hr protein did not bind and came in the flow-through fractions 4-8 (red square).

Cj0045-Hr protein was eluted from the column at approx. 85ml, corresponding to an apparent MW of 21 kDa, indicating its monomeric state in solution (Figure 6.4). Peak fractions 24 and 25, which appeared to contain the Cj0045-Hr (Figure 6.5), were combined with total protein approx. 4 mg. This was concentrated to 11 mg/ml by using a Vivaspin concentrator with a 10 kDa MWCO. The buffer was then exchanged by using Zeba Desalt Spin Column to 10mM Tris pH 8.0 and 0.1M NaCl and used for crystallisation trials. A sample of purified protein was sent for mass spectrometry to check the molecular weight.

6.4 Mass spectrometry analysis

The molecular weight of Cj0045-Hr protein was checked by Mass spectrometry. The analysis found a different value to what was found from the database (Figure 6.6). The molecular weight results of Cj0045-Hr using ProParam tool is 29.450 kDa, however, the MS result shows it was 28.919 kDa. This means 531Da of the protein was missing from the Cj0045-Hr sequence. It could be possible that the protein had lost some histidine tags from the sequence; its mass was 28.627 kDa without His-tag, and this could be one of the reasons for the lack of binding to the Ni-column. Therefore, further investigation was required to be sure the purified protein was Cj0045-Hr.

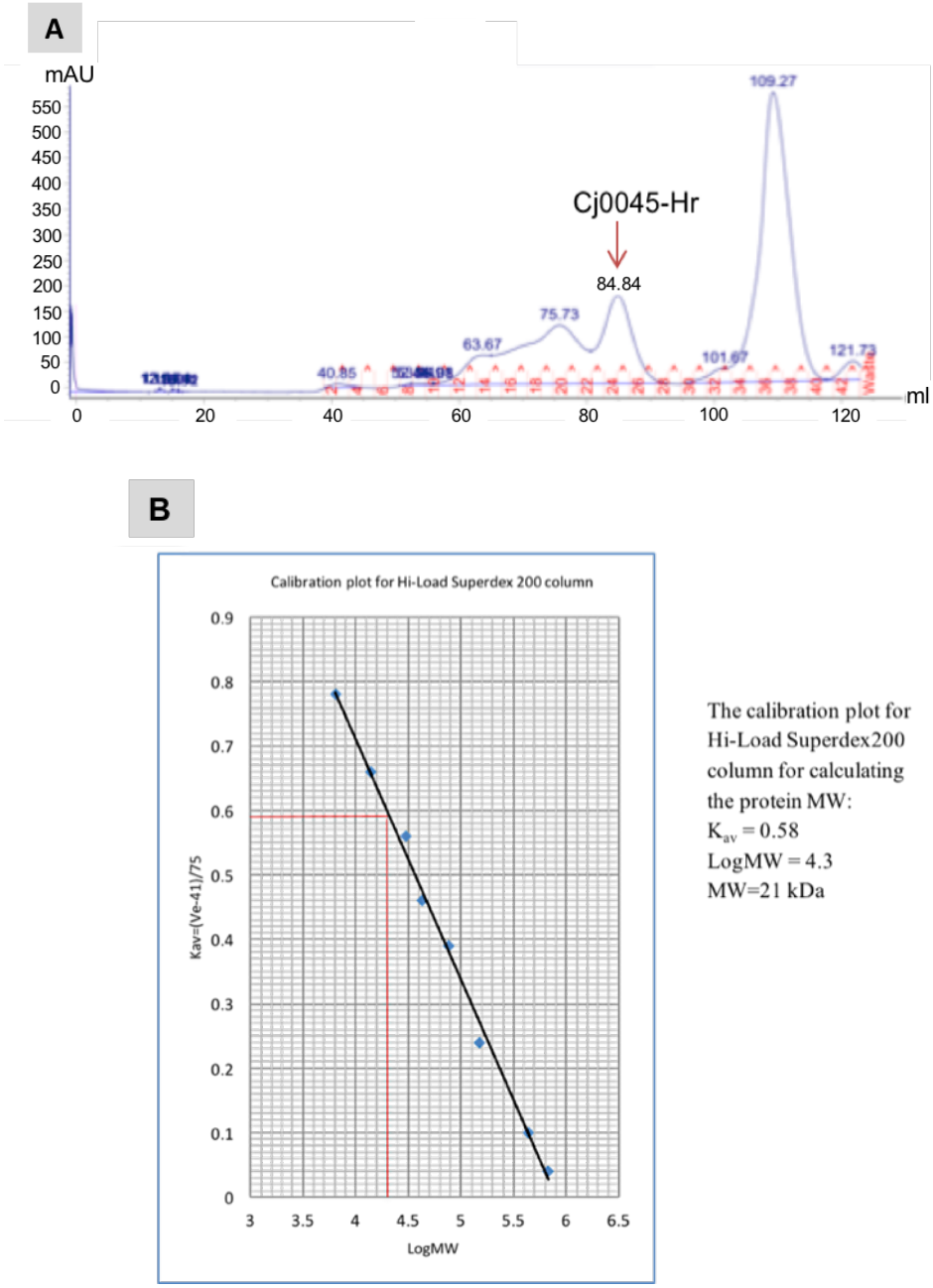


Figure 6.4: Gel filtration purification of Cj0045-Hr protein. Panel (A) The chromatogram of the gel filtration on a 1.6x60cm Hiload Superdex200 column loading and elution of Cj0045 in a peak at 84.84 ml. **Panel (B)** The calibration plot for Hi-Load Superdex 200 column chart. The calibration plot for Hi-Load Superdex 200 showed that the Cj0045-Hr corresponded to an apparent MW 21 kDa, indicating its monomeric state in solution.

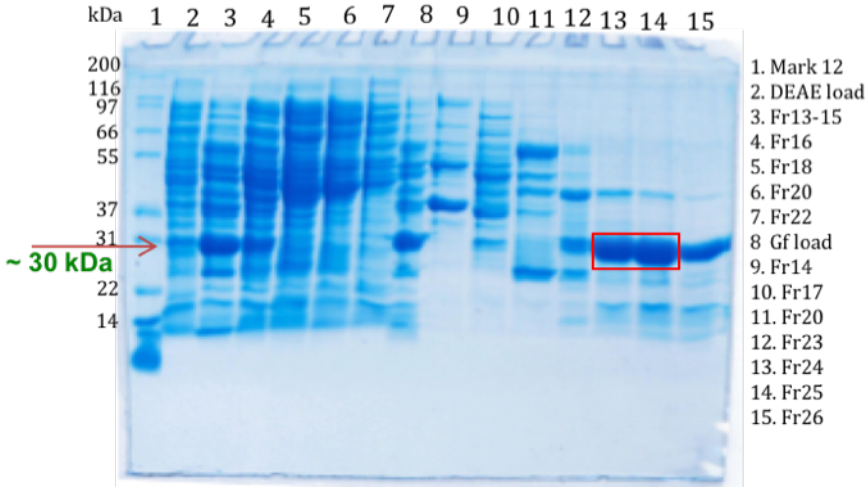


Figure 6.5: 12% SDS-PAGE analysis of purification progress of Cj0045-Hr preparation. Most Cj0045 purified protein appear in fractions 22-26. Fractions 24 & 25 (red square box), lanes 13 & 14, were combined with total protein of 4mg. The Cj0045-Hr protein is shown (the arrow sign) with size ~30 kDa.

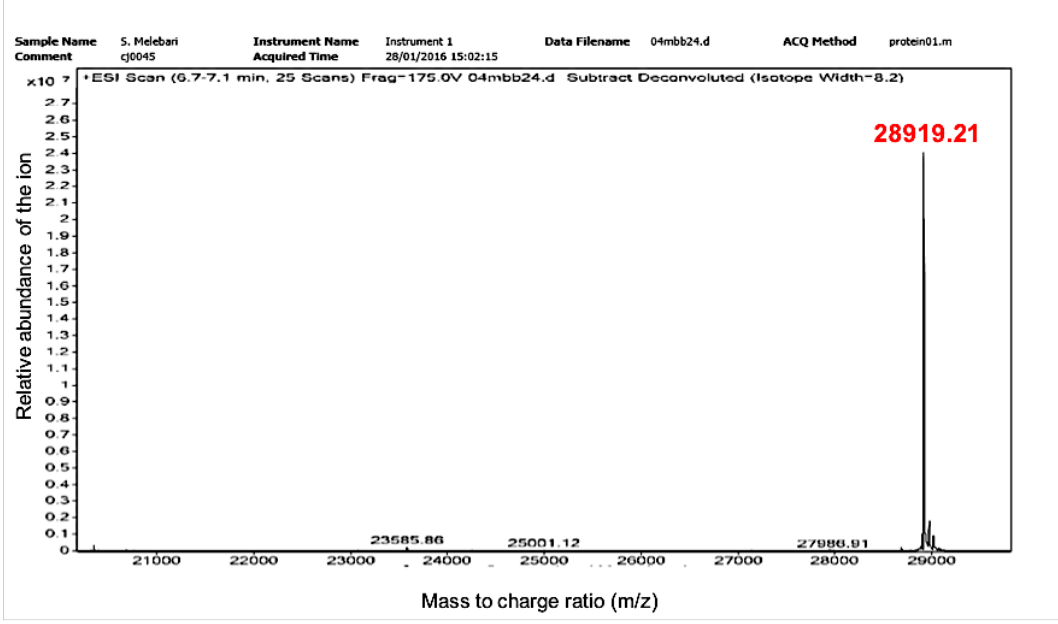


Figure 6.6: Cj0045-Hr protein Mass Spectrometry analysis. The Mass spectrum shows the molecular weight of Cj0045-Hr was 28.919 kDa, which is slightly different from protein database results (29.45 kDa).

6.5 MS/MS analysis for Cj0045-Hr protein

To confirm the identity of the purified protein sample, a small amount of this protein was sent for MS/MS sequencing to check by analysis of peptide fragments as explained in chapter (2) section (2.12.2). The digested protein was then analysed and the result showed 31 fragments of Cj0045-Hr protein were detected after comparing them with the protein sequence database (Figure 6.7). Consequently, this result confirmed the purified was Cj0045-Hr protein. Interestingly, the six-Histidine tags were found, but some residues were not detected. In theory, however, it should have identified these peptides given their size, but other physiochemical parameters might have precluded their identification: 1) they may not fragment very well or 2) they could have unknown post-translational modifications.

6.6 Crystallisation trials of Cj0045-Hr

After confirming the identity of the protein, the initial crystallisation screening trial was set up by a Matrix Hydra II Plus One crystallisation robot (Thermo Fisher Scientific, USA) with various screening kits as explained in chapter (2) section (2.12.3). The protein concentration was 11mg/ml as explained in section (6.4) and all the plates were incubated at 17^{o} C. After a few days, some cluster crystals were observed in several screening plates, but unfortunately they were found to be salt. Further effort was made to encourage Cj0045-H protein to produce crystals; the screen was carried out using metal ions ($FeCl_3$) and reducing agent (DTT) as described in chapter (2) section (2.12.3), but again the screening showed only salt crystals.

A

No.	Sequence
1	DEEKFMESIDFPLIEEHKK
2	DWILDHFANEDLWIANFTK
3	DWILDHFANEDLWIANFTKK
4	ELLEHSNDIVK
5	FMESIDFPLIEEHK
6	FMESIDFPLIEEHKK
7	FNAIFEDALQNHHFTTQENDMR
8	FNAIFEDALQNHHFTTQENDMRGGGLEHHHHHH
9	GGGLEHHHHHH
10	IHAVPQTIHQELVSK
11	IHAVPQTIHQELVSKENTLK
12	IHFKDEEK
13	IHFKDEEKFMESIDFPLIEEHK
14	IHFKDEEKFMESIDFPLIEEHKK
15	KSHQILVEK
16	KTLHLQEIHYTLEQYIK
17	KTLHLQEIHYTLEQYIKLK
18	LKSIKQDLRAEK
19	MSQELSILTK
20	NDLKQILVK
21	NMQLDKQHELIFEITNLANDLALNIQDNNTQHK
22	NMQLDKQHELIFEITNLANDLALNIQDNNTQHKNDLK
23	QHELIFEITNLANDLALNIQDNNTQHK
24	QHELIFEITNLANDLALNIQDNNTQHKNDLK
25	SHQILVEK
26	SHQILVEKTK
27	SHQILVEKTKELLEHSNDIVK
28	SIKQDLRAEK
29	TKELLEHSNDIVK
30	TLHLQEIHYTLEQYIK
31	TLHLQEIHYTLEQYIKLK

B

MKVKWSRDFSIKNMQLDKQHELIFEITNLANDLALNIQDNNTQHKNDLKQILVKLFQYIKIHF
KDEEKFMESIDFPLIEEHKKSHQILVEKTKELLEHSNDIVKMSQELSILTKDWILDHFANEDLW
IANFTKKTLHLQEIHYTLEQYIKLKSIKQDLRAEKTYDYICNCSLRIHAVPQTIHQELVSKENT
LKCEKCGQILVHLDYFDLNQNFEKFNAIFEDALQNHHFTTQENDMRGGGLEHHHHHH

Figure 6.7: MS/MS analysis for Cj0045 with C-terminal His-tag. Panel (A) The MS/MS analysis showed 31 fragments including the His-tag sequence (yellow highlight). These fragments were detected when performing the digestion in solution using Trypsin, which cleaves at the carboxylic side of arginine (R) and lysine (K) residues. **Panel (B)** The analysis program did not identify certain peptides (green highlight). Trypsin cleavage sites are shown in red highlight.

6.7 Cj0045-Hr protein with N-terminus His-tags

In crystallography as mentioned previously, there is some debate regarding the position of the His-tag, and whether perhaps the positioning of the tag at C-terminal or N-terminal end of the protein maybe influences the ability to crystallise the native protein (Carson *et al.*, 2007). So, an attempt was made to obtain protein crystals from Cj0045 by changing the position of the His-tag from C-terminal to N-terminal. New primers for *cj0045* gene were designed as explained in chapter (2), section (2.12.4) and new clones created followed by growth and then purification of them. The trial was done using pET28b as a vector (Figure 2.6).

6.7.1 Cloning of Cj0045-Hr

6.7.1.1 Amplification of the cj0045-Hr gene

The truncated *cj0045-Hr* gene, which had been previously inserted into a pET-21b vector between the Ndel and Xhol restriction enzyme sites, was amplified by PCR as described in chapter (2) section (2.11.6.2), using the Cj0045-Hr forward and reverse primers and Q5® High-Fidelity 2X Master reaction mixture. To analyse the PCR products, they were run on a 1% agarose gel (see section, 2.11.6.3). The agarose gel showed a band with an expected size of approx. 720 bp (Figure 6.8), and a Gel Extraction Kit after that was utilised for purification of the PCR products (see appendix).

6.7.1.2 Restriction enzyme digest and ligation of Cj0045 into pET28b

The PCR product of *cj0045-Hr* gene was digested with the restriction enzymes Xol1 (NEB) and NdeI (NEB) prior to insert into pET28b vector that was cut with the same enzymes, as previously stated in chapter (2) section (2.12.4).

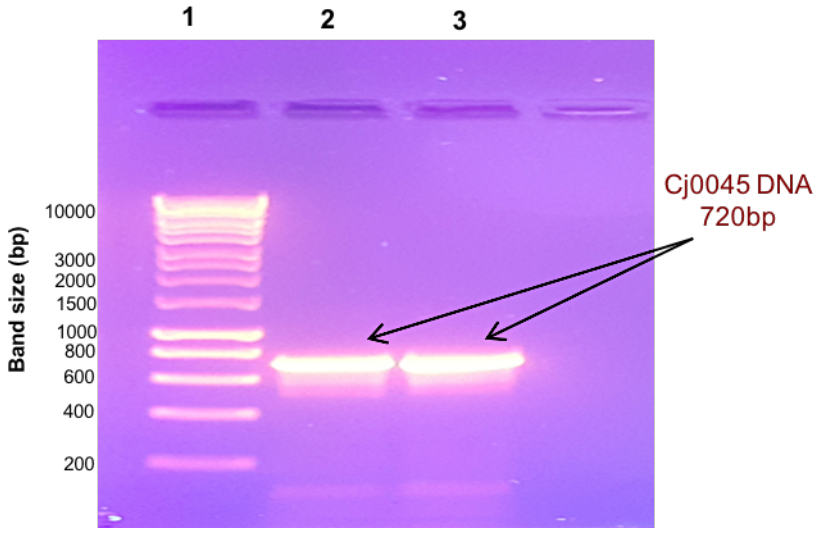


Figure 6.8: 1% DNA agarose gel of *cj0045-Hr* gene amplification. Lane (1) is DNA hyper ladder 1 molecular weight marker. Lanes 2 & 3 PCR product, and the arrows indicate Cj0045-Hr PCR product band at the expected size of 720 bp.

The DNA fragments and plasmids were incubated in separate tubes with restriction buffer (Cutsmart) to get a best efficiency in the restriction digest. The digestion was accomplished for 4 hours at 37℃ in a total volume of 20 µl. The reaction was set up as explained in chapter (2), section (2.11.6.6). Following digestion, the enzymes were inactivated by incubation at 65℃ for 20 minutes according to the manufacturer's guidelines. The buffer and enzyme were later removed from the sample using the QIAquick® PCR Purification Kit (Qiagen). The digested DNA (PCR products and pET28b plasmid) was analysed using 1% agarose gel electrophoresis (Figure 6.9). Before the ligation experiment, the digested vectors were treated with Shrimp Alkaline Phosphatase (rSAP) as explained in section (2.11.6.6.1). The desired DNA fragments were ligated after that into the linearised vector (pET28b) by using the bacteriophage T_4 ligase (Sambrook and Russell, 2006) , as explained in section (2.11.6.6.2). The ligation mixes were transformed into competent DH5α *E. coli* cells (Novagen). As vector pET28b carries a kanamycin resistance gene, the subsequent transformant was spread on LB-agar plates supported with kanamycin 50µg/ml, and the plates were incubated overnight at 37°C. Next day some colonies were observed on the plates, and consequently screened for effective cloning.

6.7.1.3 Confirmation the existence of cj0045-Hr gene in pET28b plasmids

Correct insertion of the *cj0045* gene in the pET28b vector was checked. Some colonies from the plated clones were picked and the potential constructs were obtained using a Miniprep kit (Qiagen) (see appendix). The subsequent miniprep samples were sent with standard T7 forward and reverse primers to GATC Biotech to check the sequence and DNA sequence analysis confirmed the correct product was made.

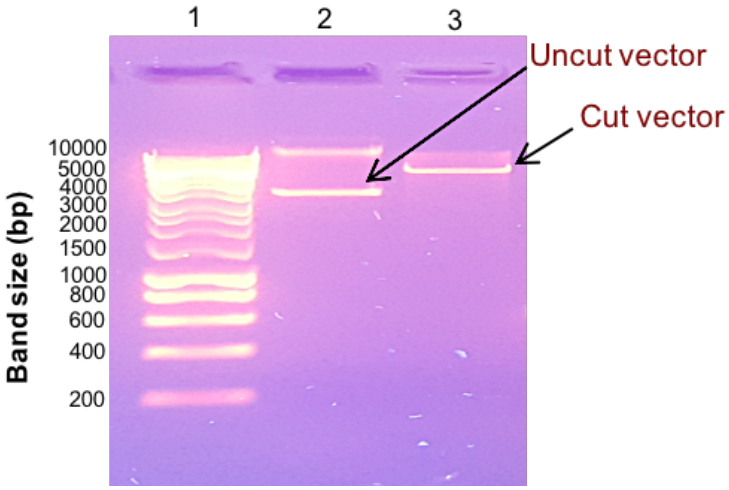


Figure 6.9: 1% DNA agarose gel of digest vector pET28b. Lane (1) is DNA hyper ladder^{TM} 1kb molecular weight marker (BIOLINE). Lane 2 pET28b before cutting using restriction enzymes digest. Lane 3 pET28b after cutting using restriction enzymes digest, and it shows a band at expected size of approx. 5000 bp.

6.7.2 Expression and solubility test

6.7.2.1 Small-scale overexpression

The construct was transformed into a competent *E. coli* strain BL21 (DE3) (Novagen), and plated with appropriate antibiotics onto LB-agar. Firstly, a small-scale overexpression experiment was prepared to check the expression level of the cloned gene and the possibility of protein solubility at 15 °C, 25 °C and 37 °C, as previously explained in chapter (2) section (2.12.4.1). In each condition, Cj0045-Hr protein was over-expressed, but was in the insoluble fractions as seen by SDS-PAGE analysis (Figure 6.10). The molecular weight of the expressed protein was the same as expected at approx. 30 kDa.

6.7.2.2 M9 minimal medium cultures for expression of Cj0045

In order to improve the expression of soluble Cj0045-Hr, attempts were made to grow cultures using M9 minimal media instead of LB media as mentioned in section (2.3.2)**.** The experiment was carried out at 15 °C, 25 °C and 37 °C as explained in chapter (2) section (2.12.4.2) and the results analysed by SDS-PAGE (Figure 6.11). Unfortunately, the same results were obtained without any soluble protein expression.

6.8 Structure prediction of Cj0045-Hr protein

Based on the Cj0241 structure, the Swiss-Model server suggested that the Cj0045-Hr has a bundle of four alpha helices, consistent with its 30% sequence identity with Cj0241-Hr (Figure 6.12). In addition, the C. terminal region of the sequence contains four cysteine residues motifs like Cj1224-Hr. For this region, the Phyre[2] server provides models with low sequence identity and confidence values but also suggests a model which contains a zinc finger structure (Figure 6.13).

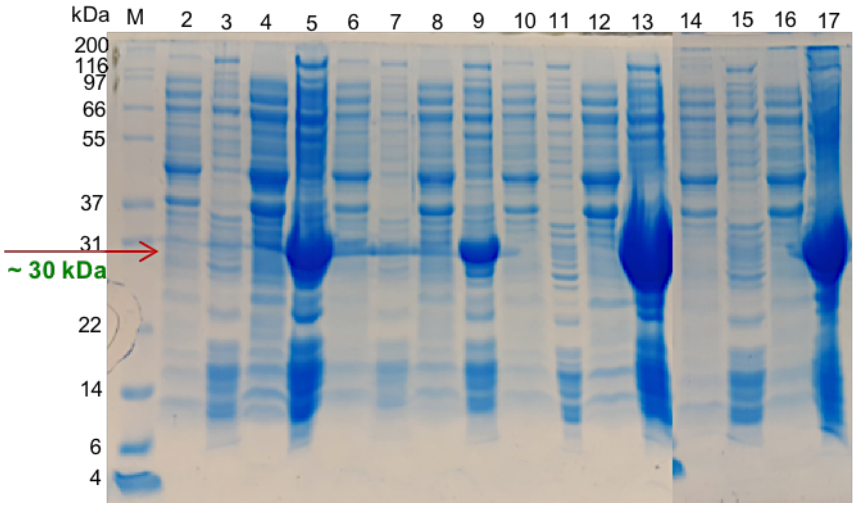


Figure 6.10: SDS-PAGE gel analysis of the expression of recombinant N-terminal His-tagged Cj0045-Hr protein under different temperatures with 1mM IPTG induction. The gel shows all of the Cj0045-Hr appeared in insoluble fractions. Lanes show the results from pre- and post-induction, soluble and insoluble cultures at various temperatures as indicated. Lane 1: molecular weight marker and red arrow indicates approximate 30kDa Mw region of Cj0045-Hr protein. Lanes 2,3: soluble and insoluble pre-induction, 15 °C (200 rpm); lanes 4,5: soluble and insoluble induced culture, 15°C (200 rpm); lanes 6,7: soluble and insoluble pre-induction, 15°C (150 rpm); lanes 8,9: soluble and insoluble induced culture, 15°C (150 rpm); lanes 10,11: soluble and insoluble pre-induction, 25°C (220 rpm); lanes 12,13: soluble and insoluble induced, 25°C (220 rpm); lanes 14,15: soluble and insoluble pre-, 37°C (220 rpm); lanes 16,17: soluble and insoluble induced, 37°C (220 rpm). The rotation speeds of the incubator shaker are given in parentheses.

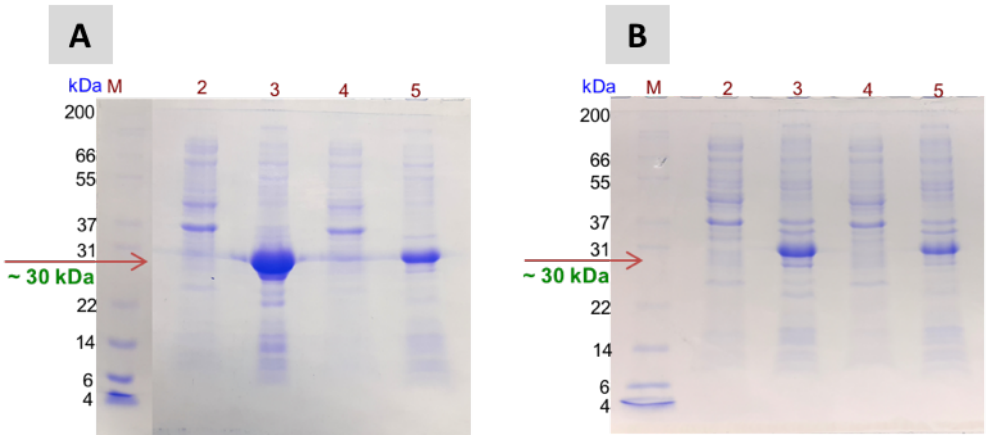


Figure 6.11: SDS-PAGE gel analysis of the expression of recombinant N-terminal His-tag-Cj0045-Hr protein using M9 minimal medium. The cultures were grown under different temperatures with 1mM IPTG induction. The gel shows all the Cj0045-Hr in the insoluble fractions. Lanes show the results from pre- and post-induction, soluble and insoluble cultures at various temperatures as indicated. **Panel (A)** Lane 1: molecular weight marker and red arrow indicates approximate 30kDa MW region of Cj0045-Hr protein. Lanes 2,3: soluble and insoluble, 15 °C (150 rpm); lanes 4,5: soluble and insoluble induced, 15°C (200 rpm). **Panel (B)** Lane 1: molecular weight marker and red arrow indicates approximate 30kDa MW region of Cj0045-Hr protein. Lanes 2,3: soluble and insoluble, 25°C (220 rpm); lanes 4,5: soluble and insoluble induced, 37°C (220 rpm). The rotation speeds of the incubator shaker are given in parentheses.

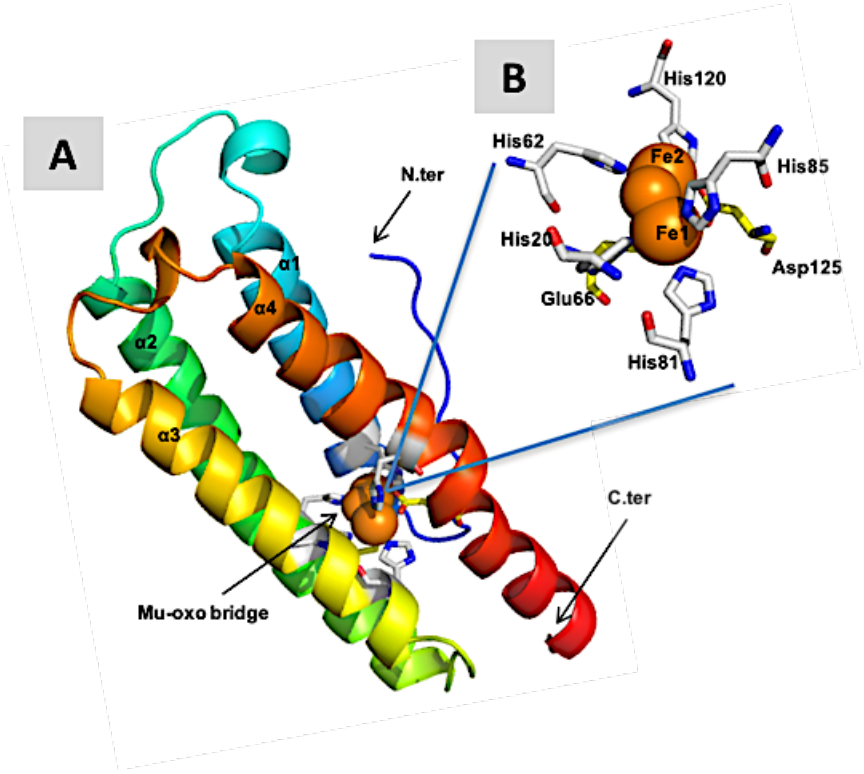


Figure 6.12: Predicted structure of Cj0045-Hr based on the Cj0241-hr structure from Swiss-model. Panel (A) Swiss-Model server suggests the Cj0045-Hr is contains four alpha helices, consistent with a 30% identity with Cj0241-Hr. **Panel (B)** The location of the expected di-iron centre is also shown by displaying the expected amino acid residue ligands in stick form.

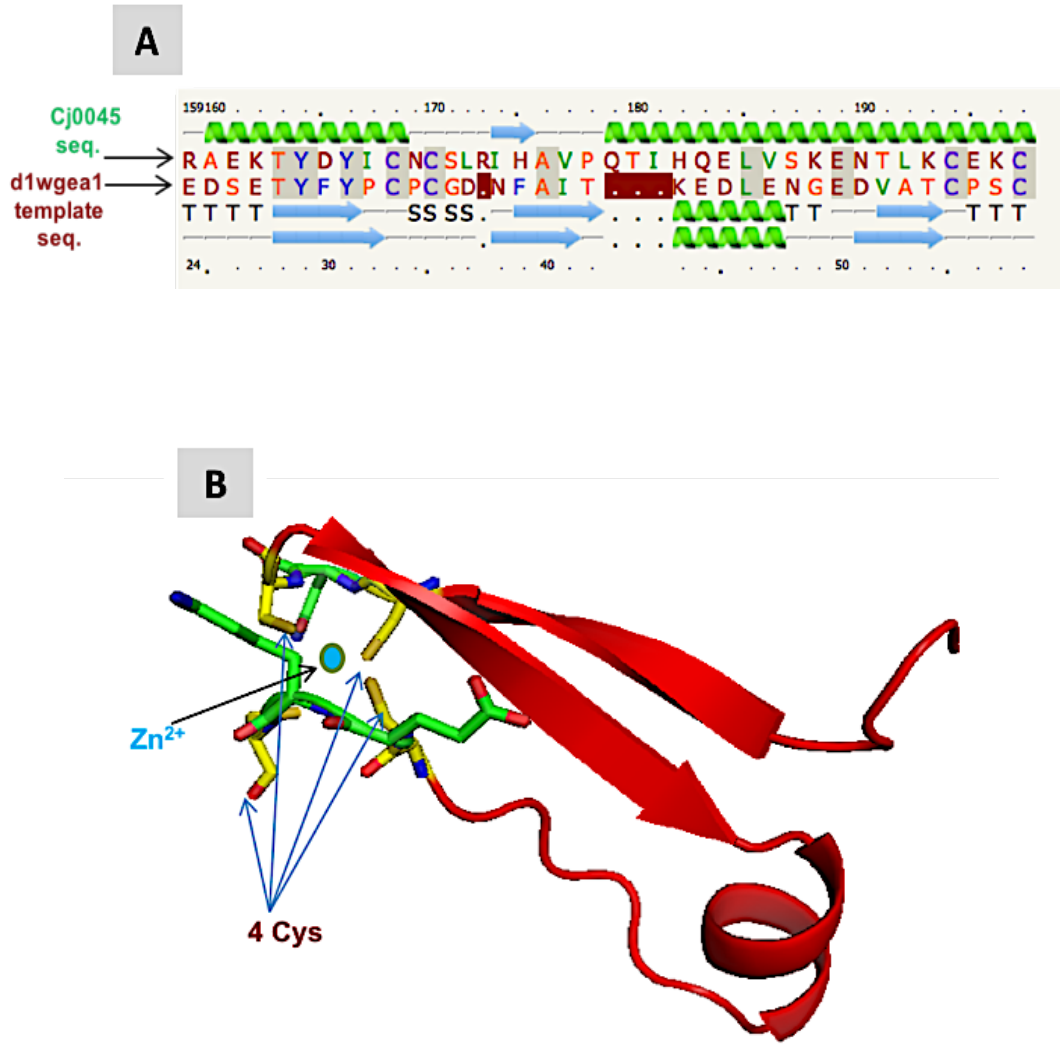


Figure 6.13: The results of modelling the C-terminal region of Cj0045 with the Phyre2 server. Panel (A) The sequence alignment of the C-terminal part of Cj0045 and its template model from Phyre2 (from PDB entry 1wge). It shows approx. 25 % sequence identity. **Panel (B)** The predicted structure of the C-terminal part of Cj0045 shows the 4 Cys residues potentially binding a zinc ion.

This could suggest that the Cj10045-Hr protein structure perhaps is also binding additional zinc ions, as suggested with Cj1224-Hr. Zinc fingers usually have a Cys2 His 2 or Cys 4 in motif in many proteins (Si *et al.*, 2015). Moreover, analysis of the Cj0045 model for potential tunnels using the CAVER program (Chovancova *et al.*, 2012) revealed six possible water tunnels (using the same default parameter values in the program as done with the analysis of the original Cj0241 structure) (Figure 6.14). Even though these tunnels were based on a model derived from the Cj0241 structure, they are noticeably different in shape from those seen in Cj0241. They all provide access to the di-iron core and might be linked to its solute accessibility and the autoxidation process.

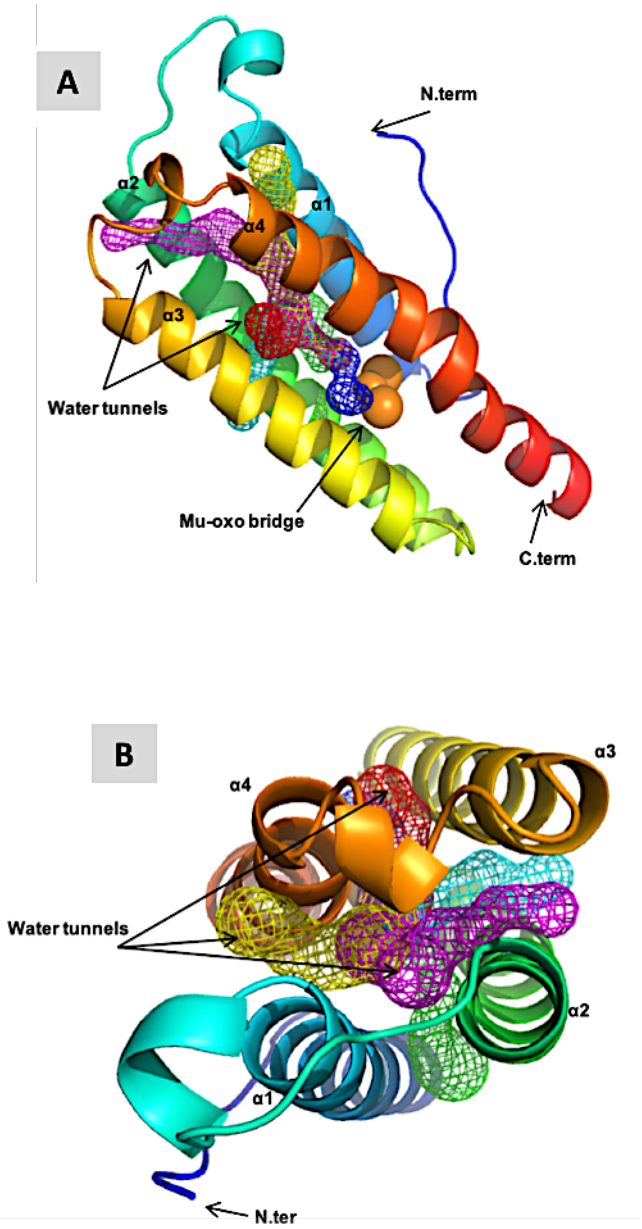


Figure 6.14: A prediction of the solvent tunnels in a possible model of the Cj0045-Hr protein structure. The tunnels were calculated using the CAVER 3.0 with standard probe radius 0.8, which is plugin to PyMol. **Panel (A)** shows a side view of Cj0045-Hr model structure, which comprises a four- helix in bundle (rainbow colours, N- to C- terminus), a non-heme di-iron centre (orange) with six predicted water tunnels (different coloured mesh). **Panel (B)** shows a top view of the same model structure with predicted water tunnels. The water tunnels were found oriented either mainly parallel with or perpendicular to the long of the four-helix bundle.

Chapter 7

Summary and future work

7.1 Summary

Campylobacteriosis has been defined as a developing foodborne disease and the main cause of gastroenteritis in humans (Kwan *et al.*, 2008). It is estimated to be the most common causative agent of foodborne diseases, followed by non-typhoidal *Salmonella, Shigella spp*. and *Escherichia coli* 0157:H7 (Mead *et al.*, 1999). *C. jejuni* infection is one of the most generally recognised bacterial causes of acute gastroenteritis worldwide (Houf and Stephan, 2007) . The pathogenesis is spreading through the oral–fecal route, and can colonise the intestinal mucosa of most warm-blooded animals. This colonisation is commensal, and in commercial broiler chickens *C. jejuni* is found at its highest levels in the mucosal crypts in the small intestine (Newell and Fearnley, 2003). In developing countries, the *Campylobacter spp.* is a significant cause of childhood morbidity from diarrheal illness. Even though the campylobacteriosis is usually a self-limiting disease in healthy people, a severe gastrointestinal disease can happen in immunocompromised people. Furthermore, post infection problems from *Campylobacter* such as reactive arthritis and Guillain-Barré syndrome can occur in both immunocompetent and immunocompromised people(Adzitey and Nurul, 2011)

Campylobacter is considered a microaerophilic bacterium that has optimal growth at usually 3-5 % O_2 and approx. 2-10% CO_2. However, it is able to grow in some aerobic conditions. The complete lack of ability for growth under 21% (v/v) O_2 is not because of an absence of oxidative stress defences (Lombard *et al.*, 2000). *C. jejuni* has well-characterised enzymes that can protect cells from molecular O_2 and work as defences

such as catalase, superoxide dismutase (SOD), alkyl-hydroperoxide reductase (AhpC) and desulforuberythrin (Rrc)(Parkhill *et al.*, 2000) . *C. jejuni* microaerophily is also not because of an inability to express these defences, because it has been shown in chemostat culture with high levels of O_2 that a number of these proteins such as AhpC, Tpx and SodB are expressed when compared to similar cultures which have been grown in low and optimal O_2 situations (Kendall *et al.*, 2014). However, two key metabolic enzymes have been identified in *C. jejuni* which are particularly oxygen sensitive. The enzymes pyruvate: acceptor oxydoreductase (Por) and 2-oxoglutarate: acceptor oxidoreductase (Oor) possess iron-sulfur clusters that are damaged by microaerobic conditions or during high O_2 transitions in the C. jejuni lifecycle (Kendall *et al.*, 2014). These two enzymes are normally only found in anaerobic bacteria and replace the usual enzymes (pyruvate dehydrogenase and 2-oxoglutarate dehydrogenase) found in aerobic bacteria. This would suggest they might be playing a key function in the microaerophilic nature of the organism. It is possible that reactive oxygen species and molecular oxygen are damaging the iron–sulfur clusters of these enzymes, resulting in inactive enzymes and the production of free ferrous iron that can cause the generation of further oxidative radicals (Imlay, 2006). There is evidence in *C. jejuni* that mutants of the genes *cj0241* and *cj1224* encoding hemerythrin proteins are more sensitive to oxygen and have reduced Por and Oor activity (Kendall *et al.*, 2014). This suggest that the single-domain O_2-binding hemerythrins could also be included in the mechanisms for prevention of oxygen harm in microaerophilic bacteria, and perhaps reflects an evolutionary adaptation to the presence of free molecular oxygen (Alvarez-Carreño *et al.*, 2016).

These hemerythrin-like sequences appear in almost all Campylobacter species with high homology between each strain with unknown functions (French *et al.*, 2008). The analysis of the genome of NCTC11168 in *C. jejuni* shows three different genes

encoding proteins with iron and oxygen binding hemerythrin-like characteristics; *cj0241, cj1224* and *cj0045*. Protein sequences of the Cj0241, Cj1224 and Cj0045 hemerythrins reveal that all include conserved histidine motifs H, HxxE, HxxxH and HxxxxD, which are important to bind the di-iron centre critical for hemerythrin function (Kendall *et al.*, 2014). In addition, there are differences between them that include the molecular weight and possible metal binding motifs in C-terminal extensions.

The *cj0241c* gene has an oxygen transport classification, and it encodes a di-iron centre hemerythrin protein and is located directly upstream of the iron-sulfur centre biosynthesis genes *iscS* and *nifU* (Kendall *et al.*, 2014). The function of Cj0241 in vivo is unclear. However, a new hypothesis based upon a recovery of activity analysis suggests a role for Cj0241 in the protection of damaged Fe-S clusters in Por or Oor, possibly by isolating molecular O_2 to keep a balance of intracellular O_2 helpful to Fe-S cluster repair or synthesis (Kendall *et al.*, 2014).

This thesis presents the first structure of a Cj0241-Hr protein from *C. jejuni*. Cj0241 has been expressed successfully and purified to homogeneity using a standard purification protocol by Ni-affinity column chromatography. Purified Cj0241 protein possesses UV spectral features at 336, 375 and ~500 nm corresponding with characteristic features of previously studied hemerythrins (Kao *et al.*, 2008; Kendall *et al.*, 2014). Cj0241-Hr protein was successfully crystallised using PEG3350 as a precipitant by the sitting-drop vapour-diffusion method. A full diffraction dataset was collected using a synchrotron facility to 1.1 Å resolution from a single crystal of hemerythrin Cj0241. The structure of Cj0241 was solved using SAD data from the two iron atoms in Cj0241. Cj0241 has a typical α-helical bundle architecture. It consists of four parallel α helices, and an active site of di-iron at the core of the structure that is able to bind one oxygen molecule. The di-iron ions are both Fe (II) in the deoxy-hemerythrin, and Fe (III) in the

oxy-hemerythrin, as well as Fe (III) in a met-hemerythrin (met-Hr), which can also exist upon autoxidation via water. The iron ions are connected to the protein backbone by the carboxylate side chains of aspartate (120) and glutamate (60), and the imidazole groups of five histidine residues: three histidines (19, 56 & 75) are bound to Fe-1, and the other two (79 & 115) bound to Fe-2.

The structure Cj0241-Hr was found in the met-Hr form. The oxygen binding pocket site was modeled as a water molecule after the O_2 is released, in agreement with the electron density shape and appropriate occupancy for a water molecule in refinement.

The structure of Cj0241-Hr also revealed two potential water tunnels to the di-iron core. These tunnels conserved the approximate shape and geometry of other tunnels found in hemerythrins in species such as *Desulfovibrio vulgaris* and *Methylococcus capsulatus*. The tunnels are suggested to link with a fast autoxidation rate during transition of the water and oxygen molecules. It was not possible to get a $t_{1/2}$ value for this process because the Cj0241hemrythrin protein was very sensitive to the reducing agents sodium dithionite ($Na_2S_2O_4$) and sodium borohydride ($NaBH_4$) and easily aggregated and precipitated. However, a comparison of the non-heme di-iron site geometry, and the dimensions of the tunnels in the Cj0241-Hr and the other species might imply a role for the Cj0241-Hr as a transporter or sensor protein to the oxygen molecules.

The second hemerythrin studied here was Cj1224, and it may have a similar function as Cj0241-Hr (Kendall *et al.*, 2014). It has spectral features like Cj0241, but unlike Cj0241, it has two possible additional metal binding motifs (CXC and CXXC) in its C-terminal region, which may have a functional role. Many attempts were made to purify and crystallise this protein, but all attempts to crystallise the protein and determine its structure were unsuccessful. Interestingly the metal binding properties of Cj1224-Hr

were analysed using ICP-MS technique, and demonstrated the presence of a significant peak for zinc ions as well as the bound iron ions. In addition, the Phyre2 server provided a model, which suggested the C-terminal region of Cj1224 could be a zinc finger and consistent with the ICP-MS analysis. To check the ability of this protein for DNA interaction, an electrophoretic mobility-shift assay (EMSA) was performed and showed some evidence of band shifts. This was again consistent with the hypothesis that the Cj1224-Hr structure has a zinc-finger in the C-terminal region because these structural features are known for their DNA and protein interaction roles.

The other *C. jejuni* hemerythrin, Cj0045, studied here has previously been suggested to play a different but unknown role to Cj0241 and Cj1224(Kendall *et al.*, 2014). Previous studies with Cj0045-Hr also had yielded only insoluble protein(Kendall *et al.*, 2014), however, in this study some soluble Cj0045-Hr protein was obtained by varying the growth conditions. This was verified by mass spectrometry and MS/MS analysis. Many attempts were conducted to get protein crystals, but they were all unsuccessful. The C-terminal region was also analysed using the Phyre2 server, which again suggested that it might be a zinc finger.

7.2 Future work

In spite of some new results being discovered through the work in this project, many aspects concerning the mechanism of the hemerythrins in *C. jejuni* still remain unanswered and need extensive biochemical, physical and structural studies to achieve the full understanding of the importance of these proteins. Understanding the mechanisms by which pathogens respond to oxidative stress, might reveal anti-microbial targets and shed indications on pathogenic mechanisms. Options for future work include:

1. Finding more reliable reducing system alternatives to sodium dithionite and sodium borohydride, that do not damage the protein and enable the autoxidation rate for wild type (WT) protein to be determined *in vitro*.
2. Testing the hypothesis that the residue Ile112 acts as a regulatory gate in the channel for entry of H_2O molecules to reach the di-iron site and controls the di-oxygen binding features in Cj0241-Hr. Mutagenesis studies (site-direct mutagenesis) with residue Ile112 could be used to probe this, and Ile112 could be mutated to Alanine. The autoxidation rate of the variant could then be checked and compared with the wild type.
3. Mutation experiments on residues that are considered to play important functional role within the hemerythrin Cj0241 structure. For example, the composition of residues around the active site, those clearly bind iron ions. An investigation could be made into how critical each one is to the system and how many are really needed, although the conservation of the motifs and the coordination geometry of the iron centres would suggest that the structure might need all them. These residues include His19, His56, His75, His79, His115, Glu60 and Asp120, which can be mutated to Alanine. Changing the Glu60 and Asp120 residues for Asp/Gln or Glu/Asn, respectively, could be done to look at their importance in bridging the iron ions. The mutant protein could be crystallised and checked to find out whether it is folded up the same way, or see if the mutation has made any massive change in the structure. Furthermore, to check the function affect (*in vivo*) a complementation assay with a *C. jejuni cj0241* null mutant using either wild type or mutated versions of *cj0241* on a vector. An assay for Por activity can then be checked as previously described (Kendall *et al.*, 2014). All family of different mutants of *cj0241c* can be investigated in the same way. Also, site- directed

mutagenesis trial of Asp residue to Glu residue (or vice versa) in Cj0241-Hr might also give some evidence if the hemerythrin can accommodate the two iron ions, or does the protein still work as to their role function *in vivo* using a complementation assay. The experiment can be performed also using the crystallography structure technique to check the active site shape, and the ICP-MS analysis to estimate the iron ions.

4. The interaction between Cj0241-Hr and the other proteins in *Campylobacter* might be recognised through purification of the protein on HisTrap HP columns, into which the cell free extract (CFE) of *C. jejuni* is injected. Elimination of unbound CFE by washing the column could possibly leave some proteins in the CFE binding with Cj0241 and able to be purified for further analysis. This is a possibility because there was a hypothesis that Cj0241-Hr protein and Por and Oor can interact together (Parrish *et al.*, 2007). The experiment could confirm the interaction seen previously using the yeast two-hybrid system screens.
5. In addition, there is a possibility that Cj0241 could form a multimeric protein (octameric or trimeric) with one of the other hemerythin, because we know that some hemerythrins from eukaryotes do form multimeric hemerythrins (French *et al.*, 2008). A possible function with this interaction is that the Cj0241 acts with Cj1224 to carry out their role. The Cj0241 possibly binds with Por (Parrish *et al.*, 2007) and if it binds with Cj1224, there is a possibility that Cj1224 might offer a functional part in its C-terminal extension, which is not present in Cj0241. The extension in Cj1224 possibly contains a zinc finger motif from our studies. The zinc finger in the yeast Fep1 protein was suggested to provide an alternative Fe-S cluster or provide cysteines as alternative cluster ligands (Antimo *et al.*, 2016).

Also, the interaction could be checked *in vivo* using the yeast two-hybrid system screens.

6. The Cj1224 hemerythrin in *C. jejuni* may also have direct interaction with Por and Oor enzyme proteins, which might agree with the hypothesis of a zinc finger. A yeast two-hybrid system screens could be used to verify this.
7. Cj0045 could be tested to see whether it can form DNA-protein or RNA-protein or protein-protein interactions. The experiments could be performed as explained for Cj1224-Hr with a variety of different sources of DNA and RNA using electrophoretic mobility-shift assay (EMSA) and the yeast two-hybrid system screens.
8. Because the *cj1224* and *cj0045* from the strain NCTC 11168 have not crystallised, it may be valuable to find another *cj1224* and *cj0045* from a different strain of *Campylobacter* that has 70%-80% sequence identity, which might suggest surface residues have different behaviour in these proteins and might help for crystallisation experiments and better handling.

References

Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.w., Kapral, G.J., Grosse-kunstleve, R.W.*, et al.* (2010). PHENIX : a comprehensive Python-based system for macromolecular structure solution. Acta Crystallographica Section D *66*, 213-221.

Adkin, A., Hartnett, E., Jordan, L., Newell, D., and Davison, H. (2006). Use of a systematic review to assist the development of *Campylobacter* control strategies in broilers. Journal of Applied Microbiology *100*, 306-315.

Adzitey, F., and Nurul, H. (2011). *Campylobacter* in poultry: Incidences and possible control measures. Research Journal of Microbiology *6*, 182-192.

Alfredson, D.A., and Korolik, V. (2007). Antibiotic resistance and resistance mechanisms in *Campylobacter jejuni* and *Campylobacter coli* (Oxford, UK), pp. 123-132.

Altschul, S.F., Gish, W., Miller, W., Meyers, E.W., and Lipman, D.J. (1990). Basic Local Alignment Search Tool. Journal of Molecular Biology *215*, 403-410.

Alvarez-Carreño, C., Becerra, A., Lazcano, A., and Nikolaidis, N. (2016). Molecular Evolution of the Oxygen- Binding Hemerythrin Domain. PLoS ONE *11*.

Andreini, C., Banci, L., Bertini, I., and Rosato, A. (2006). Counting the zinc- proteins encoded in the human genome. Journal of proteome research *5*, 196.

Andrews, S.C., Robinson, A.K., and Rodríguez-Quiñones, F. (2003). Bacterial iron homeostasis (Oxford, UK), pp. 215-237.

Antimo, C., Barry, D.H., Adriana, E.M., Rossella, M., Alessandra, G., Andrea, B., Giulietta, S., Giovanni, M., and Maria Carmela Bonaccorsi Di, P. (2016). *Pichia pastoris* Fep1 is a [2Fe- 2S] protein with a Zn finger that displays an unusual oxygen-dependent role in cluster binding. Scientific Reports *6*.

Atack, J.M., and Kelly, D.J. (2009). Oxidative stress in *Campylobacter jejuni*: responses, resistance and regulation. Future microbiology *4*, 677-690.

Auld, D. (2001). Zinc coordination sphere in biochemical zinc sites. Biometals *14*, 271-313.

Bae, J.S., Yuki, N., Kuwabara, S., Kim, J.K., Vucic, S., Lin, C.S., and Kiernan, M.C. (2014). Guillain–Barré syndrome in Asia. J Neurol Neurosurg Psychiatry *85*, 907.

Ban Mishu, A. (2001). *Campylobacter jejuni* Infections: Update on Emerging Issues and Trends. Clinical Infectious Diseases *32*, 1201-1206.

Barbara, J.C., Annette Summers, E., Megan, L.P., and Ken, T. (2006). The versatile ε-proteobacteria: key players in sulphidic habitats. Nature Reviews Microbiology *4*, 458.

Barlow, D.J., and Thornton, J.M. (1988). Helix geometry in proteins. Journal of Molecular Biology *201*, 601-619.

Bartlett, J.M., and Stirling, D. (2003). A short history of the polymerase chain reaction. PCR protocols, 3-6.

Batz, M.B., Hoffmann, S., and Morris Jr, J.G. (2012). Ranking the disease burden of 14 pathogens in food sources in the United States using attribution data from outbreak investigations and expert elicitation. Journal of food protection *75*, 1278-1291.

Berg, J.M., and Shi, Y. (1996). The Galvanization of Biology: A Growing Appreciation for the Roles of Zinc. Science *271*, 1081-1085.

Bernaudat, F., Frelet-Barrand, A., Pochon, N., Dementin, S., Hivin, P., Boutigny, S., Rioux, J.-B., Salvi, D., Seigneurin-Berny, D., Richaud, P.*, et al.* (2011). Heterologous Expression of Membrane Proteins: Choosing the Appropriate Host. PLoS ONE *6*.

Bhavsar, S., and Kapadnis, B. (2007). Virulence factors of *Campylobacter*. The Internet Journal of Microbiology *3*.

Blaser, M.J., Taylor, D.N., and Feldman, R.A. (1983). Epidemiology of *Campylobacter jejuni* infections. Epidemiologic reviews *5*, 157.

Blow, D.M. (2002). Outline of crystallography for biologists (Oxford: Oxford : Oxford University Press, 2002).

Bolen, D.W. (2004). Effects of naturally occurring osmolytes on protein stability and solubility: issues important in protein crystallisation. Methods *34*, 312-322.

Bragg, W.H., and Bragg, W.L. (1913). The Reflection of X- rays by Crystals. Proceedings of the Royal Society of London Series A, Containing Papers of a Mathematical and Physical Character (1905-1934) *88*, 428-438.

Bury-Moné, S., Kaakoush, N.O., Asencio, C., Mégraud, F., Thibonnier, M., De Reuse, H., and Mendz, G.L. (2006). Is *Helicobacter pylori* a True Microaerophile? Helicobacter 11, 296-303.

Butzler, J.P. (2004). *Campylobacter* , from obscurity to celebrity (Oxford, UK), pp. 868-876.

Carrillo, C.D., Taboada, E., Nash, J.H.E., Lanthier, P., Kelly, J., Lau, P.C., Verhulp, R., Mykytczuk, O., Sy, J., Findlay, W.A.*, et al.* (2004). Genome- wide Expression Analyses of *Campylobacter jejuni* NCTC11168 Reveals Coordinate Regulation of Motility and Virulence by flhA. Journal of Biological Chemistry *279*, 20327-20338.

Carson, M., Johnson, D.H., McDonald, H., Brouillette, C., and Delucas, L.J. (2007). His-tag impact on structure. Acta Crystallographica Section D *63*, 295-301.

Cesarini, S., Bofill, C., Pastor, F.I.J., Reetz, M.T., and Diaz, P. (2012). A thermostable variant of *P. aeruginosa* cold-adapted LipC obtained by rational design and saturation mutagenesis. Process Biochemistry *47*, 2064-2071.

Chant, A., Kraemer-Pecore, C.M., Watkin, R., and Kneale, G.G. (2005). Attachment of a histidine tag to the minimal zinc finger protein of the *Aspergillus nidulans* gene regulatory protein AreA causes a conformational change at the DNA- binding site. Protein Expression and Purification *39*, 152-159.

Chen, K.H.C., Chuankhayan, P., Wu, H.-H., Chen, C.-J., Fukuda, M., Yu, S.S.F., and Chan, S.I. (2015). The bacteriohemerythrin from *Methylococcus capsulatus* (Bath): Crystal structures reveal that Leu114 regulates a water tunnel. Journal of Inorganic Biochemistry *150*, 81-89.

Chen, V.B., Arendall, W.B., Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010). MolProbity : all- atom structure validation for macromolecular crystallography. Acta Crystallographica Section D *66*, 12-21.

Chou, A.Y., Archdeacon, J., and Kado, C.I. (1998). Agrobacterium Transcriptional Regulator Ros is a Prokaryotic Zinc Finger Protein that Regulates the Plant Oncogene ipt. Proceedings of the National Academy of Sciences of the United States of America *95*, 5293-5298.

Chovancova, E., Pavelka, A., Benes, P., Strnad, O., Brezovsky, J., Kozlikova, B., Gora, A., Sustr, V., Klvana, M., Medek, P.*, et al.* (2012). CAVER 3.0: A Tool for the Analysis of Transport Pathways in Dynamic Protein Structures (Tool for Analysis of Tunnel and Channel Dynamics). *8*, e1002708.

Christine, M.S., and Brendan, W.W. (2005). Protein glycosylation in bacterial mucosal pathogens. Nature Reviews Microbiology *3*, 225.

Chruszcz, M., Potrzebowski, W., Zimmerman, M.D., Grabowski, M., Zheng, H., Lasota, P., and Minor, W. (2008). Analysis of solvent content and oligomeric states in protein crystals - Does symmetry matter? Protein Science *17*, 623-632.

Claudia, B., Mark, E., Guenter, R., Jörg, S., Christa, L., Oliver, K., Ramkumar, N., Roland, G., Andrea, R., Heike, K.*, et al.* (2003). Complete genome sequence and analysis of *Wolinella succinogenes*. Proceedings of the National Academy of Sciences of the United States of America *100*, 11690.

Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nature Biotechnology *26*, 1367-1372.

Creighton, T.E. (1992). Protein folding (WH Freeman and Company).

D'Arcy, A., Villard, F., and Marsh, M. (2007). An automated microseed matrix-screening method for protein crystallisation. Acta Crystallographica Section D-Biological Crystallography *63*, 550-554.

Dasti, J.I., Tareen, A.M., Lugert, R., Zautner, A.E., and Groß, U. (2010). *Campylobacter jejuni*: A brief overview on pathogenicity- associated factors and disease- mediating mechanisms. International Journal of Medical Microbiology *300*, 205-211.

Daucher, J.A., and Krieg, N.R. (1995). Pyruvate: ferredoxin oxidoreductase in *Campylobacter species*. Canadian journal of microbiology *41*, 198-201.

Debruyne, L., Gevers, D., and Vandamme, P. (2008). Taxonomy of the family *Campylobacteraceae*. In *Campylobacter*, Third Edition (American Society of Microbiology), pp. 3-25.

DeLano, W.L. (2002). The PyMOL user's manual. DeLano Scientific, San Carlos, CA *452*.

Dessau, M.A., and Modis, Y. (2011). Protein crystallisation for X- ray crystallography. Journal of Visualised Experiments.

Dorrell, N., Mangan, J.A., Laing, K.G., Hinds, J., Linton, D., Al-Ghusein, H., Barrell, B.G., Parkhill, J., Stoker, N.G., Karlyshev, A.V.*, et al.* (2001). Whole genome comparison of *Campylobacter jejuni* human isolates using a low- cost microarray reveals extensive genetic diversity. Genome research *11*, 1706.

Drenth, J. (2007). Principles of protein X-ray crystallography (Springer Science & Business Media).

Eaton, K., Morgan, D., and Krakowka, S. (1992). Motilty as a factor in the colonization of gnotobiotic piglets by *Helicobacter pylori*. J Med Microbiol *37*, 123-127.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics.

Engberg, J., On, S., Harrington, C., and Gerner-Smidt, P. (2000). Prevalence of *Campylobacter*, *Arcobacter*, *Helicobacter*, and *Sutterella spp*. in human fecal samples as estimated by a reevaluation of isolation methods for Campylobacters. Journal of Clinical Microbiology *38*, 286-291.

Epps, S., Harvey, R., Hume, M., Phillips, T., Anderson, R., and Nisbet, D. (2013). Foodborne *Campylobacter*: Infections, Metabolism, Pathogenesis and Reservoirs. International Journal of Environmental Research and Public Health *10*, 6292-6292.

Evans, P. (2006). Scaling and assessment of data quality. Acta Crystallographica Section D: Biological Crystallography *62*, 72-82.

Farmer, C.S., Kurtz Jr, D.M., Liu, Z.-J., Wang, B.C., Rose, J., Ai, J., and Sanders-Loehr, J. (2001). The crystal structures of *Phascolopsis gouldii* wild type and L98Y methemerythrins: structural and functional alterations of the O_2 binding pocket. J Biol Inorg Chem *6*, 418-429.

Fouts, D.E., Mongodin, E.F., Mandrell, R.E., Miller, W.G., Rasko, D.A., Ravel, J., Brinkac, L.M., DeBoy, R.T., Parker, C.T., Daugherty, S.C.*, et al.* (2005). Major Structural Differences and Novel Potential Virulence Mechanisms from the Genomes of

Multiple *Campylobacter* Species (Sequencing of Four *Campylobacter* Species). PLoS Biology *3*, e15.

French, C.E., Bell, J.M.L., and Ward, F.B. (2008). Diversity and distribution of hemerythrin- like proteins in prokaryotes (Oxford, UK), pp. 131-145.

Gaynor, E., Cawthraw, S., Manning, G., and Mackichan, J. (2004). The genome-sequenced variant of *Campylobacter jejuni* NCTC 11168 and the original clonal clinical isolate differ markedly in colonization, gene expression, and virulence-associated phenotypes. Journal of bacteriology *186*, 503-517.

Geetha, V. (1996). Distortions in protein helices. International Journal of Biological Macromolecules *19*, 81-89.

Gilbreath, J., Cody, W., Merrell, D., and Hendrixson, D. (2011). Change Is Good: Variations in Common Biological Mechanisms in the Epsilonproteobacterial Genera *Campylobacter* and *Helicobacter*. In Microbiol Mol Biol Rev, pp. 84-+.

Giles, G.I., and Jacob, C. (2002). Reactive sulfur species: An emerging concept in oxidative stress. Biological Chemistry *383*, 375-388.

Giles, N.M., Giles, G.I., and Jacob, C. (2003a). Multiple roles of cysteine in biocatalysis. Biochemical and Biophysical Research Communications *300*, 1-4.

Giles, N.M., Watts, A.B., Giles, G.I., Fry, F.H., Littlechild, J.A., and Jacob, C. (2003b). Metal and Redox Modulation of Cysteine Protein Function. Chemistry & Biology *10*, 677-693.

Glusker, J.P. (1984). The Patterson function. Trends in Biochemical Sciences *9*, 328-330.

Golden, N., and Acheson, D. (2002). Identification of motility and auto agglutination *Campylobacter jejuni* mutants by random transposon mutagenesis. Infect Immun *70*, 1761-1771.

Guccione, E., Hitchcock, A., Hall, S.J., Mulholland, F., Shearer, N., Van Vliet, A.H.M., and Kelly, D.J. (2010). Reduction of fumarate, mesaconate and crotonate by Mfr, a novel oxygen- regulated periplasmic reductase in *Campylobacter jejuni*. Environmental Microbiology *12*, 576-591.

Guerry, P. (2007). *Campylobacter* flagella: not just for motility. Trends in Microbiology *15*, 456-461.

Guerry, P., Logan, S.M., Thornton, S., and Trust, T.J. (1990). Genomic organization and expression of *Campylobacter* flagellin genes. Journal of bacteriology *172*, 1853-1860.

Halloran, T.V. (1993). Transition Metals in Control of Gene Expression. Science *261*, 715-725.

Hames, D., and Hooper, N. (2005). Biochemistry, 3 th. New York: Taylor dan Francis.

Harald, N., and Christine, M.S. (2010). Protein glycosylation in bacteria: sweeter than ever. Nature Reviews Microbiology *8*, 765.

Hasegawa, K. (2012). Introduction to single crystal X-ray analysis. The Rigaku Journal *28*, 14-18.

Heimesaat, M.M., Lugert, R., Fischer, A., Alutis, M., Kühl, A.A., Zautner, A.E., Tareen, A.M., Göbel, U.B., and Bereswill, S. (2014). Impact of *Campylobacter jejuni* cj0268c knockout mutation on intestinal colonization, translocation, and induction of immunopathology in gnotobiotic IL-10 deficient mice. PloS one *9*, e90148.

Hendrickson, W.A., Klippenstein, G.L., and Ward, K.B. (1975). Tertiary structure of myohemerythrin at low resolution. Proceedings of the National Academy of Sciences of the United States of America *72*, 2160.

Hendrixson, D.R., and Dirita, V.J. (2004). Identification of *Campylobacter jejuni* genes involved in commensal colonization of the chick gastrointestinal tract. Molecular Microbiology *52*, 471-484.

Hermans, D., Van Deun, K., Martel, A., Van Immerseel, F., Messens, W., Heyndrickx, M., Haesebrouck, F., and Pasmans, F. (2011). Colonization factors of *Campylobacter jejuni* in the chicken gut. Veterinary research.

Hofreuter, D., Tsai, J., Watson, R., Novik, V., Altman, B., Benitez, M., Clark, C., Perbost, C., Jarvie, T., Du, L.*, et al.* (2006). Unique Features of a Highly Pathogenic *Campylobacter jejuni* Strain. Infection and Immunity *74*, 4694-4707.

Holmes, M.A., Le Trong, I., Turley, S., Sieker, L.C., and Stenkamp, R.E. (1991). Structures of deoxy and oxy hemerythrin at 2.0 Å resolution. Journal of Molecular Biology *218*, 583-593.

Horrocks, S.M., Anderson, R.C., Nisbet, D.J., and Ricke, S.C. (2009). Incidence and ecology of *Campylobacter jejuni* and *coli* in animals. Anaerobe *15*, 18-25.

Hospital, A., Candotti, M., Gelpí, J.L., and Orozco, M. (2017). The Multiple Roles of Waters in Protein Solvation. Journal of Physical Chemistry B *121*, 3636-3643.

Houf, K., and Stephan, R. (2007). Isolation and characterization of the emerging foodborn pathogen Arcobacter from human stool. Journal of Microbiological Methods *68*, 408-413.

Imlay, J. (2006). Iron- sulphur clusters and the problem with oxygen. Molecular Microbiology *59*, 1073-1082.

Isaza, C.E., Silaghi-Dumitrescu, R., Iyer, R., Kurtz, D.M., and Chan, M. (2006). Structural Basis for O sub(2) Sensing by the Hemerythrin- like Domain of a Bacterial Chemotaxis Protein: Substrate Tunnel and Fluxional N Terminus. Biochemistry (Washington) *45*, 9023-9031.

Jackson, R.J., Lee, L.J., Gidley, M.D., Wainwright, L.M., Lightfoot, J., Poole, R.K., Elvers, K.T., and Park, S.F. (2007). Oxygen reactivity of both respiratory oxidases in *Campylobacter jejuni*: The cydAB genes encode a cyanide- resistant, low-affinity oxidase that is not of the cytochrome bd type. Journal of bacteriology *189*, 1604-1615.

Jacobson, M.P., Friesner, R.A., Xiang, Z., and Honig, B. (2002). On the Role of the Crystal Environment in Determining Protein Side-chain Conformations. Journal of Molecular Biology *320*, 597-608.

Jin, S., Joe, A., Lynett, J., and Hani, E. (2001). JlpA, a novel surface- exposed lipoprotein specific to *Campylobacter jejuni*, mediates adherence to host epithelial cells. Molecular Microbiology *39*, 1225.

Johnson, W.M., and Lior, H. (1988). A new heat- labile cytolethal distending toxin (CLDT) produced by *Campylobacter* spp. Microbial Pathogenesis *4*, 115-126.

Jones, F.S., Orcutt, M., and Little, R.B. (1931). Vibrios (*Vibrio jejuni*, n. sp.) associated with intestinal disorders of cows and calves. The Journal of experimental medicine *53*, 853.

Kaakoush, N.O., Miller, W.G., De Reuse, H., and Mendz, G.L. (2007). Oxygen requirement and tolerance of *Campylobacter jejuni*. Research in Microbiology *158*, 644-650.

Kaakoush, N.O., Sodhi, N., Cox, J.M., Mitchell, H.M., Chenu, J.W., and Riordan, S.M. (2014). The interplay between *Campylobacter* and *Helicobacter* species and other gastrointestinal microbiota of commercial broiler chickens. Gut Pathogens *6*.

Kabsch, W. (2010a). Integration, scaling, space- group assignment and post- refinement. Acta Crystallographica Section D *66*, 133-144.

Kabsch, W. (2010b). XDS. Acta Crystallographica Section D *66*, 125-132.

Kantardjieff, K.A., and Rupp, B. (2003). Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein–nucleic acid complex crystals. Protein Science *12*, 1865-1871.

Kao, W.-C., Wang, V.C.C., Huang, Y.-C., Yu, S.S.F., Chang, T.-C., and Chan, S.I. (2008). Isolation, purification and characterization of hemerythrin from *Methylococcus capsulatus* (Bath). Journal of Inorganic Biochemistry *102*, 1607-1614.

Karlsen, O.A., Ramsevik, L., Bruseth, L.J., Larsen, Ø., Brenner, A., Berven, F.S., Jensen, H.B., and Lillehaug, J.R. (2005). Characterisation of a prokaryotic haemerythrin from the methanotrophic bacterium *Methylococcus capsulatus* (Bath. FEBS Journal *272*, 2428-2440.

Kelly, D.J. (2008). Complexity and Versatility in the Physiology and Metabolism of *Campylobacter jejuni*. *Campylobacter*, 3rd Edition, 41-61.

Kelly, J., Jarrell, H., Millar, L., and Tessier, L. (2006). Biosynthesis of the N- Linked Glycan in *Campylobacter jejuni* and Addition onto Protein through Block Transfer. Journal of bacteriology *188*, 2427-2434.

Kendall, J.J., Barrero-Tobon, A.M., Hendrixson, D.R., and Kelly, D.J. (2014). Hemerythrins in the microaerophilic bacterium *Campylobacter jejuni* help protect key iron–sulphur cluster enzymes from oxidative damage. Environmental Microbiology *16*, 1105-1121.

Konkel, M.E., Garvis, S.G., Tipton, S.L., Anderson, J.D.E., and Cieplak, J.W. (1997). Identification and molecular cloning of a gene encoding a fibronectin- binding protein (CadF) from *Campylobacter jejuni.* Molecular Microbiology *24*, 953-963.

Konkel, M.E., Monteville, M., Rivera-Amill, V., and Joens, L. (2001). The Pathogenesis of *Campylobacter jejuni*- Mediated Enteritis. Current Issues in Intestinal Microbiology *2*, 55-71.

Korlath, J.A., Osterholm, M.T., Judy, L.A., Forfang, J.C., and Robinson, R.A. (1985). A Point- Source Outbreak of Campylobacteriosis Associated with Consumption of Raw Milk. The Journal of Infectious Diseases *152*, 592-596.

Korolik, V. (2010). Aspartate chemosensory receptor signalling in *Campylobacter jejuni*. Virulence *1*, 414.

Korostelev, A., and Noller, H.F. (2007). Analysis of Structural Dynamics in the Ribosome by TLS Crystallographic Refinement. Journal of Molecular Biology *373*, 1058-1070.

Kwan, P.S.L., Birtles, A., Bolton, F.J., French, N.P., Robinson, S.E., Newbold, L.S., Upton, M., and Fox, A.J. (2008). Longitudinal study of the molecular epidemiology of *Campylobacter jejuni* in cattle on dairy farms. Applied and environmental microbiology *74*, 3626.

Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. nature *227*, 680-685.

Lara-Tejero, M., and Galan, J. (2001). CdtA, CdtB, and CdtC Form a Tripartite Complex That Is Required for Cytolethal Distending Toxin Activity. Infection and Immunity *69*, 4358-4365.

Larkin, M.A., Blackshields, G., Brown, N., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I., Wilm, A., Lopez, R.*, et al.* (2007). Clustal W and clustal X version 2.0. Bioinformatics *23*, 2947-2948.

Lastovica, A.J., On, S.L., and Zhang, L. (2014). The Family *Campylobacteraceae*. In the Prokaryotes (Springer), pp. 307-335.

Lee, A., Rourke, J.L., Barrington, P.J., and Trust, T.J. (1986). Mucus colonization as a determinant of pathogenicity in intestinal infection by *Campylobacter jejuni*: A mouse cecal model. Infection and Immunity *51*, 536-546.

Lehner, A., Tasara, T., and Stephan, R. (2005). Relevant aspects of Arcobacter spp. as potential foodborne pathogen. International journal of food microbiology *102*, 127-135.

Lertsethtakarn, P., Ottemann, K.M., and Hendrixson, D.R. (2011). Motility and Chemotaxis in *Campylobacter* and *Helicobacter*. Annu Rev Microbiol *65*, 389-410.

Linton, D., Dorrell, N., Hitchen, P.G., Amber, S., Karlyshev, A.V., Morris, H.R., Dell, A., Valvano, M.A., Aebi, M., and Wren, B.W. (2005). Functional analysis of the *Campylobacter jejuni* N- linked protein glycosylation pathway. Molecular Microbiology *55*, 1695-1703.

Lombard, M., Touati, D., Fontecave, M., and Nivière, V. (2000). Superoxide reductase as a unique defense system against superoxide stress in the microaerophile Treponema pallidum. Journal of Biological Chemistry *275*, 27021-27026.

Lu, C.-H., Lin, Y.-F., Lin, J.-J., Yu, C.-S., and Vertessy, B.G. (2012). Prediction of Metal Ion– Binding Sites in Proteins Using the Fragment Transformation Method. PLoS ONE *7*.

Luft, J.R., and Detitta, G.T. (1999). A method to produce microseed stock for use in the crystallisation of biological macromolecules. Acta Crystallographica Section D *55*, 988-993.

Luisa, F.J.-S., and Rainer, H. (2016). The CagA toxin of *Helicobacter pylori*: abundant production but relatively low amount translocated. Scientific Reports *6*.

Mackichan, J.K., Gaynor, E.C., Chang, C., Cawthraw, S., Newell, D.G., Miller, J.F., and Falkow, S. (2004). The *Campylobacter jejuni* dccRS two- component system is required for optimal in vivo colonization but is dispensable for in vitro growth. Molecular Microbiology *54*, 1269-1286.

Malgieri, G., Palmieri, M., Russo, L., Fattorusso, R., Pedone, P.V., and Isernia, C. (2015). The prokaryotic zinc- finger: structure, function and comparison with the eukaryotic counterpart, pp. 4480-4496.

Manfreda, G., Parisi, A., Lucchi, A., Zanoni, R.G., and De Cesare, A. (2011). Prevalence of *Helicobacter pullorum* in conventional, organic, and free- range broilers and typing of isolates. Applied and environmental microbiology *77*, 479.

Matthews, B.W. (1968). Solvent content of protein crystals. Journal of Molecular Biology *33*, 491-497.

Maulik, V., Jennifer, S., and Teruna, J. (2009). The role of thiols and disulfides in protein chemical and physical stability. Curr Protein Pept Sci *10*, 614-625.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. Journal of Applied Crystallography *40*, 658-674.

McNally, D., Hui, J., Aubry, A., Mui, K., Guerry, P., Brisson, J.-R., Logan, S., and Soo, E. (2006). Functional Characterization of the Flagellar Glycosylation Locus in

Campylobacter jejuni 81-176 Using a Focused Metabolomics Approach. Journal of Biological Chemistry *281*, 18489-18498.

McPherson, A. (1990). Current approaches to macromolecular crystallisation (Oxford, UK), pp. 1-23.

Mead, P.S., Slutsker, L., Dietz, V., McCaig, L., Bresee, J., Shapiro, C., Griffin, P., and Tauxe, R. (1999). Food- related illness and death in the United States. In Emerg Infect Dis, pp. 607-625.

Meinersmann, R., Patton, C., Evins, G., Kaye, W., and Fields, P. (2002). Genetic diversity and relationships of *Campylobacter* species and subspecies. International journal of systematic and evolutionary microbiology *52*, 1789-1797.

Müller, A., León-Kempis, M.D.R., Dodson, E., Wilson, K.S., Wilkinson, A.J., and Kelly, D.J. (2007). A Bacterial Virulence Factor with a Dual Role as an Adhesin and a Solute-binding Protein: The Crystal Structure at 1.5 Å Resolution of the PEB1a Protein from the Food-borne Human Pathogen *Campylobacter jejuni*. Journal of Molecular Biology *372*, 160-171.

Müllner, P., Collins-Emerson, J.M., Midwinter, A.C., Carter, P., Spencer, S.E.F., van Der Logt, P., Hathaway, S., and French, N.P. (2010). Molecular epidemiology of *Campylobacter jejuni* in a geographically isolated country with a uniquely structured poultry industry. Applied and environmental microbiology *76*, 2145.

Murshudov, G., Vagin, A.A., and Dodson, E. (1997). Refinement of macromolecular structures by the maximum- likelihood method. Acta Crystallographica Section D-Biological Crystallography *53*, 240-255.

Nachamkin, I., Liu, J., Li, M., Ung, H., Moran, A.P., Prendergast, M.M., and Sheikh, K. (2002). *Campylobacter jejuni* from patients with Guillain-Barré syndrome preferentially expresses a GD(1a)-like epitope. Infection and immunity *70*, 5299.

Nachamkin, I., Szymanski, C.M., and Blaser, M.J. (2008). *Campylobacter* (ASM Press).

Nachamkin, I., Yang, X., and Stern, N. (1993). Role of *Campylobacter- jejuni* flagella as colonization factors for 3- day- old chicks - analysis with flagellar mutants. Applied And Environmental Microbiology *59*, 1269-1273.

Neilands, J.B. (1995). Siderophores: structure and function of microbial iron transport compounds. The Journal of biological chemistry *270*, 26723.

Newell, D.G., and Fearnley, C. (2003). Sources of *Campylobacter* colonization in broiler chickens. In Appl Environ Microbiol, pp. 4343-4351.

Newell, D.G., McBride, H., and Dolby, J.M. (1985). Investigations on the role of flagella in the colonization of infant mice with *Campylobacter jejuni* and attachment of *Campylobacter jejuni* to human epithelial cell lines. J Hyg *95*, 217-227.

Nuijten, P., Vanasten, F., Gaastra, W., and Vanderzeijst, B. (1990). Structural and functional-analysis of 2 *Campylobacter-jejuni* flagellin genes. Journal Of Biological Chemistry *265*, 17798-17804.

Nyati, K.K., and Nyati, R. (2013). Role of Infection in the Pathogenesis of Guillain-Barré Syndrome: An Update. BioMed Research International *2013*.

O'Rourke, J., and Bode, G. (2001). Morphology and ultrastructure.

Ofran, Y., Mysore, V., and Rost, B. (2007). Prediction of DNA- binding residues from sequence. Bioinformatics *23*, i347-i353.

On, S. (2001). Taxonomy of *Campylobacter*, *Arcobacter*, *Helicobacter* and related bacteria: current status, future prospects and immediate concerns. J Appl Microbiol *90*, 1S-15S.

Owen, R. (1998). *Helicobacter* - species classification and identification. Br Med Bull *54*, 17-30.

Pace, N., and Weerapana, E. (2014). Zinc- Binding Cysteines: Diverse Functions and Structural Motifs (Basel: MDPI AG), pp. 419-434.

Parkhill, J., Wren, B.W., Mungall, K., Ketley, J.M., Churcher, C., Basham, D., Chillingworth, T., Davies, R.M., Feltwell, T., Holroyd, S.*, et al.* (2000). The genome sequence of the food- borne pathogen *Campylobacter jejuni* reveals hypervariable sequences. Nature *403*, 665.

Parrish, J.R., Yu, J., Liu, G., Hines, J.A., Chan, J.E., Mangiola, B.A., Zhang, H., Pacifico, S., Fotouhi, F., DiRita, V.J.*, et al.* (2007). A proteome- wide protein interaction map for *Campylobacter jejuni*. Genome Biology *8*, R130-R130.

Patel, H., Grüning, B.A., Günther, S., and Merfort, I. (2014). PyWATER: a PyMOL plug-in to find conserved water molecules in proteins by clustering. Bioinformatics *30*, 2978-2980.

Pearson, B.M., Gaskin, D.J.H., Segers, R.P.A.M., Wells, J.M., Nuijten, P.J.M., and van Vliet, A.H.M. (2007). The complete genome sequence of *Campylobacter jejuni* strain 81116 (NCTC11828). Journal of bacteriology *189*, 8402.

Perez-Perez, G.I., and Blaser, M.J. (1996). *Campylobacter* and *helicobacter.*

Pflugrath, J.W. (2004). Macromolecular cryocrystallography— methods for cooling and mounting protein crystals at cryogenic temperatures. Methods *34*, 415-423.

Pickett, C., Auffenberg, T., Pesci, E., Sheen, V., and Jusuf, S. (1992). Iron acquisition and hemolysin production by *Campylobacter jejuni*. Infection and Immunity *60*, 3872-3877.

Power, P.M., and Jennings, M.P. (2003). The genetics of glycosylation in Gram- negative bacteria. FEMS Microbiology Letters *218*, 211-222.

Py, B., and Barras, F. (2010). Building Fe- S proteins: bacterial strategies. In Nat Rev Microbiol, pp. 436-446.

Radivojac, P., Obradovic, Z., Smith, D., Zhu, G., Vucetic, S., Brown, C., Lawson, J., and Dunker, A. (2004). Protein flexibility and intrinsic disorder. *13*, 71-80.

Rapp, C.S., and Pollack, R.M. (2005). Crystal packing effects on protein loops. Proteins: Structure, Function, and Bioinformatics *60*, 103-109.

Razin, S., Borunova, V., Maksimenko, O., and Kantidze, O. (2012). Cys2His2 zinc finger protein family: Classification, functions, and major members. Biochemistry Moscow *77*, 217-226.

Rhodes, G. (2006). Crystallography made crystal clear [electronic resource] : a guide for users of macromolecular models, 3rd ed. edn (Amsterdam ; Boston: Amsterdam ; Boston : Elsevier/Academic Press, c2006).

Rubinchik, S., Seddon, A., and Karlyshev, A.V. (2012). Molecular mechanisms and biological role of *Campylobacter jejuni* attachment to host cells. European journal of microbiology & immunology *2*, 32.

Rupp, B. (2010). Biomolecular crystallography : principles, practice, and application to structural biology (New York: New York : Garland, c2010).

Sambrook, J., and Russell, D.W.S. (2006). The condensed protocols from molecular cloning: a laboratory manual.

Schaller, R.A., Ali, S.K., Klose, K.E., and Kurtz, D.M. (2012). A bacterial hemerythrin domain regulates the activity of a Vibrio cholerae diguanylate cyclase. Biochemistry *51*, 8563.

Schomaker, V., and Trueblood, K.N. (1968). On the rigid- body motion of molecules in crystals. Acta Crystallographica Section B *24*, 63-76.

Sellars, M.J., Hall, S.J., and Kelly, D.J. (2002). Growth of *Campylobacter jejuni* supported by respiration of fumarate, nitrate, nitrite, trimethylamine-N-oxide, or dimethyl sulfoxide requires oxygen. Journal of bacteriology *184*, 4187-4196.

Sheppard, S.K., Dallas, J.F., Strachan, N.J.C., Macrae, M., McCarthy, N.D., Wilson, D.J., Gormley, F.J., Falush, D., Ogden, I.D., Maiden, M.C.J.*, et al.* (2009). *Campylobacter* genotyping to determine the source of human infection. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America *48*, 1072.

Si, J., Zhao, R., and Wu, R. (2015). An Overview of the Prediction of Protein DNA-Binding Sites (Basel: MDPI AG), pp. 5194-5215.

Sievert, S.M., Scott, K.A., Klotz, M.G., Chain, P.S.G., Hauser, L.J., Hemp, J., Huegler, M., Land, M., Lapidus, A., Larimer, F.W.*, et al.* (2008). Genome of the

Epsilonproteobacterial Chemolithoautotroph *Sulfurimonas denitrificans*. Applied and Environmental Microbiology *74*, 1145-1156.

Silva, J., Leite, D., Fernandes, M., Mena, C., Gibbs, P.A., and Teixeira, P. (2011). *Campylobacter* spp. as a foodborne pathogen: a review. In Front Microbiol.

Smith, T., and Taylor, M.S. (1919). Some morphological and biological characters of the spirilla (*Vibrio fetus*, n. sp.) associated with disease of the fetal membranes in cattle. The Journal of experimental medicine *30*, 299.

Stahl, M., Friis, L.M., Nothaft, H., Liu, X., Li, J., Szymanski, C.M., and Stintzi, A. (2011). L-fucose utilization provides *Campylobacter jejuni* with a competitive advantage. Proceedings of the National Academy of Sciences of the United States of America *108*, 7194.

Stenkamp, R. (1994). Dioxygen and hemerythrin. In Chem Rev, pp. 715-726.

Storz, G., and Imlayt, J.A. (1999). Oxidative stress. Current Opinion in Microbiology *2*, 188-194.

Stryer, L. (1995). Biochemistry, 1995. Newyork, NY: WH Freeman and Co, Fourth Google Scholar.

Takai, K., Hirayama, H., Nakagawa, T., Suzuki, Y., Nealson, K.H., and Horikoshi, K. (2005). *Lebetimonas acidiphila* gen. nov., sp. nov., a novel thermophilic, acidophilic, hydrogen-oxidizing chemolithoautotroph within the *Epsilonproteobacteria*, isolated from a deep-sea hydrothermal fumarole in the Mariana Arc. International journal of systematic and evolutionary microbiology *55*, 183.

Taylor, G.L. (2010). Introduction to phasing. Acta Crystallographica Section D *66*, 325-338.

Terwilliger, T.C., Grosse-Kunstleve, R.W., Afonine, P.V., Moriarty, N.W., Adams, P.D., Read, R.J., Zwart, P.H., and Hung, L.W. (2008). Iterative- build OMIT maps: map improvement by iterative model building and refinement without model bias. Acta Crystallographica Section D *64*, 515-524.

Tischler, A., and Camilli, A. (2004). Cyclic diguanylate (c- di- GMP) regulates *Vibrio cholerae* biofilm formation. Molecular Microbiology *53*, 857.

Tompa, K., Bokor, M., Verebélyi, T., and Tompa, P. (2015). Water rotation barriers on protein molecular surfaces. Chemical Physics *448*, 15-25.

Touati, D. (2000). Iron and Oxidative Stress in Bacteria. Archives of Biochemistry and Biophysics *373*, 1-6.

Vallee, B., and Auld, D. (1990). Zinc coordination, function, and structure of zinc enzymes and other proteins. Biochemistry (Washington) *29*, 5647-5660.

van Doorn, P.A., Ruts, L., and Jacobs, B.C. (2008). Clinical features, pathogenesis, and treatment of Guillain-Barré syndrome. Lancet Neurology *7*, 939-950.

van Vliet, A., and Ketley, J. (2001). Pathogenesis of enteric *Campylobacter* infection. Journal Of Applied Microbiology *90*, 45S-56S.

Vandamme, P. (2000). Taxonomy of the family *Campylobacteraceae*. In *Campylobacter* (ASM Press), pp. 3-26.

Vandevenne, M., Jacques, D.A., Artuz, C., Nguyen, C.D., Kwan, A.H.Y., Segal, D.J., Matthews, J.M., Crossley, M., Guss, J.M., and Mackay, J.P. (2013). New insights into DNA recognition by zinc fingers revealed by structural analysis of the oncoprotein ZNF217. The Journal of biological chemistry *288*, 10616.

Velayudhan, J., and Kelly, D. (2002). Analysis of gluconeogenic and anaplerotic enzymes in *Campylobacter jejuni*: an essential role for phosphoenolpyruvate carboxykinase. Microbiology-(UK) *148*, 685-694.

Velázquez, M., and Feirtag, J.M. (1999). *Helicobacter pylori*: characteristics, pathogenicity, detection methods and mode of transmission implicating foods and water. International journal of food microbiology *53*, 95-104.

Vliet, A.H.M., Ketley, J.M., Park, S.F., and Penn, C.W. (2002). The role of iron in *Campylobacter* gene regulation, metabolism and oxidative stress defense (Oxford, UK), pp. 173-186.

Wassenaar, T.M., Bleumink-Pluym, N.M.C., Newell, D.G., Nuijten, P.J.M., and Van der Zeijst, B.A.M. (1994). Differential flagellin expression in a *flaA flaB*$^{+}$ mutant of *Campylobacter jejuni*. Infection and Immunity *62*, 3901-3906.

Whitehouse, C., Balbo, P., Pesci, E., Cottle, D., Mirabito, P., and Pickett, C. (1998). *Campylobacter jejuni* cytolethal distending toxin causes a G sub(2)-phase cell cycle block. Infection and Immunity *66*, 1934-1940.

Weiss, M.S. (2001). Global indicators of X-ray data quality. Journal of Applied Crystallography *34*, 130-135.

Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G.W., McCoy, A.*, et al.* (2011). Overview of the CCP4 suite and current developments. Acta Crystallographica Section D *67*, 235-242.

Winter, G. (2010). xia2 : an expert system for macromolecular crystallography data reduction. Journal of Applied Crystallography *43*, 186-190.

Wlodawer, A., Minor, W., Dauter, Z., and Jaskolski, M. (2008). Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures (Oxford, UK), pp. 1-21.

Wösten, M.M.S.M., van Dijk, L., Parker, C.T., Guilhabert, M.R., van Der Meer-Janssen, Y.P.M., Wagenaar, J.A., and van Putten, J.P.M. (2010). Growth phase- dependent

activation of the DccRS regulon of *Campylobacter jejuni*. Journal of bacteriology *192*, 2729.

Wybo, I., Breynaert, J., Lauwers, S., Lindenburg, F., and Houf, K. (2004). Isolation of *Arcobacter skirrowii* from a patient with chronic diarrhea. Journal of clinical microbiology *42*, 1851.

Xiong, J., Kurtz, D.M., Ai, J., and Sanders-Loehr, J. (2000). A hemerythrin- like domain in a bacterial chemotaxis protein. Biochemistry *39*, 5117.

Yamasaki, M., Igimi, S., Katayama, Y., Yamamoto, S., and Amano, F. (2004). Identification of an oxidative stress- sensitive protein from *Campylobacter jejuni* , homologous to rubredoxin oxidoreductase/rubrerythrin. FEMS Microbiology Letters *235*, 57-63.

Yao, R., Burr, D.H., Doig, P., Trust, T.J., Niu, H., and Guerry, P. (1994). Isolation of motile and non- motile insertional mutants of *Campylobacter jejuni* : the role of motility in adherence and invasion of eukaryotic cells. Molecular Microbiology *14*, 883-893.

Zhong, D., Pal, S.K., and Zewail, A.H. (2011). Biological water: A critique. Chemical Physics Letters *503*, 1-11.

Appendix

1. Luria-Bertani (LB)

Luria Broth (LB) and Luria Agar (LA) media preparation.

Ingredients	LB medium	LA agar
Tryptone	10 g	10 g
Yeast extract	5 g	5 g
NaCl	10 g	10 g
Bacteriological agar	----------	15 g
H_2O	1 L	1 L

2. M9 Minimal medium

Ingredients	Stock concentration	Volume added
5x M9 salts*	-	200 ml
$MgSO_4$	1 M	2 ml
Glucose	20%w/v	20 ml
$CaCl_2$	1M	0.1 ml
Sterile dH_2O	-	Up to 980 ml

*5x M9 salts have been added to dH_2O to a final volume of 1L:

$Na_2HPO_4.7H_2O$	64 g
KH2PO4	15 g
NaCl	2.5 g
NH4Cl	5.0 g

3. Super Optimal broth with Catabolite repression (SOC medium)

Ingredients	Amount for 1L
Tryptone	20 g
Yeast extract	5 g
Glucose	3.6 g
NaCl	0.5 g
1M KCL	2.5 ml
1M $MgCl_2$	10ml
1M $MgSO_4$	10ml
ddH_2O	To 1000 ml

All media were sterilized by autoclaving, and desired antibiotics were added.

4. 12% Resolving solution

Materials	Amounts	Note
30% acrylamide	2.5 ml	
1M Tris pH 8.8	2.35 ml	Hydroxymethyl aminomethane
Milli Q (water)	1.28 ml	
10% SDS	62.5 µl	Sodium Dodecyl Sulfate
TEMED	6.25 µl	Tetramethylethylenediamine
10% APS	62.5 µl	Ammonium persulphate

5. Stacking solution

Materials	Amounts	Note
30% acrylamide	0.75 ml	
1M Tris pH 6.8	0.47 ml	Hydroxymethyl aminomethane
Milli Q (water)	2.46 ml	
10% SDS	37.5 ul	Sodium Dodecyl Sulfate
TEMED	3.75 ul	Tetramethylethylenediamine
10% APS	37.5 ul	Ammonium persulphate

6. Coomassie blue stain

Materials	Amounts
Coomassie blue	0.1%(w/v)
Methanol	1 vol
Acetic acid	1 vol
H_2O	1 vol

7. De- staining solution

Materials	Amounts
Acetic acid	10% (v/v)
Methanol	10% (v/v)

8. Bio-Rad protein assay

A protein solution of concentration 1-10ug was placed into a cuvette (1.6ml volume), to this mixture 0.8ml of milliQ water and 0.2ml of pure Bio-Rad Dye Reagent was added. The cuvette was sealed using parafilm and mixed with gentle shaking. A measurement of the OD959 should be taken at this point, the value should be in the range 0.1 and 0.7. If the reading is higher, the concentration of the protein should be diluted or a smaller volume be used and if lower, then an increase in the solution is needed with a decrease in the volume of water to ensure that the volume of the test is within 1ml. In order to calculate the protein concentration, the following formula should be used:

OD959 x 15 / Volume of protein = protein concentration (mg/ml)

9. Purification of PCR products from the agarose gel (Gel Extraction)

The target DNA fragment was sliced from the gel by using a scalpel and transferred into an Eppendorf tube 1.5 ml. The gel slice was weighed and three volumes of buffer QG (Guanidium Thiochloride) being added to each volume of agarose (e.g. 300ul to 100mg agarose) for gel solubilisation and DNA binding to the column's membrane. The gel tube was then placed into the incubation at 50°C for 10 minutes or until completely dissolved. One gel volume of isopropanol was mixed into the tube and the contents were transferred to a QIAquick spin column and centrifuged for 1 minute at 13000 rpm to bind the DNA. The flow-through was discarded, then 500 µl of buffer QG was added and centrifuged for 1 minute. 750 µl of PE washing buffer (with ethanol) was used to centrifuge for 1 minute to remove the salts and carbohydrate. The flow-through was discarded again, and more centrifugation for 1 minute was done to take out the residual buffer. The QIAquick spin column was then placed into a clean 1.5 ml Eppendorf tube. To elute the DNA, 30-40 µl of EB buffer (10 mM Tris-HCl, pH 8.5) or sterile water was added to the centre of the

QIAquick membrane. The column was allowed to stand for 3-5 minutes, and then centrifuged for 1 minute.

10. Plasmid purification

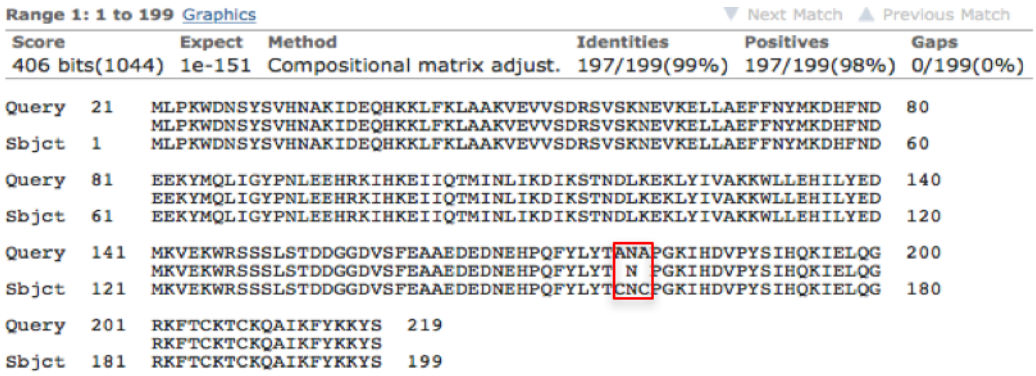
A small-scale plasmid preparation was performed according to the normal protocol of plasmid DNA purification with the QIAprep Spin Miniprep Kit from Qiagen as the following. One colony of DH5α *E. coli* cells with the plasmid pET21b or pET28b were taken from an LB agar plate and used with 5 ml of LB medium supplemented with appropriate antibiotics, 100μg/ml ampicillin for pET21b and 50μg/ml Kanamycin for pET28b. The culture was left to grow overnight at 37° C at 250 rpm. The cells were then harvested by centrifugation at 4500 rpm for 20 minutes. Cell pellets were re-suspended in 250 μl buffer P1 (50mM Tris-HCl, pH 8.0, 10mM EDTA and 100 μg/ml of RNAase A) which contains RNase to avoid RNA contamination of the purified plasmid. The resultant solution was moved into a 1.5 ml Eppendorf tube, and the cells were then subjected to alkaline lysis by adding 250 μl from buffer P2 (200mM NaOH and 1% SDS), and the tube contents were mixed by gentle inversion. Then 350μl of buffer N3 (3M Potassium acetate, pH5.5) was used to neutralize the lysis reaction with the tube being gently shaken a few times to mix all the contents, and then centrifuged at 13000 rpm for 10 minutes. The supernatant was moved to a QIAprep spin column and then centrifuged at 13000 rpm for 1 minute. The flow-through was discarded and the column washed by adding 750 μl of buffer PE, containing ethanol allowing the DNA to precipitate in the column leading to removal of the salts, and centrifuged at 13000 rpm for 1.0 minute. After removing the flow through a second centrifugation at 13000 rpm took place to remove any residual buffer PE. After that, the QIAprep column was transferred to a new 1.5 ml Eppendorf tube and the DNA was eluted by adding 30-40 μl of EB buffer (10 mM Tris-

HCl, pH 8.5) or sterile water, and incubated at room temperature for 2-5 minutes. The tube was then centrifuged at 13000 rpm for one minute to collect the pure plasmid DNA. All plasmid DNA was stored at −20˚C after checking the DNA concentration.

11. Site directed mutagenesis

The method of this experiment includes a PCR amplification of a template plasmid containing the gene using mutagenic primers. The primers were designed according to the NEB online primer design software, NEBaseChanger^{TM}, to calculate the optimal annealing temperature (Ta). PCR reactions of a total volume of 25 µl were set up for each mutation containing 1 µl DNA template (1-25 ng final amount), 1 µl (0.5 µM final conc.) of forward and reverse primers, 12.5µl Master Mix and 9µl of Nuclease-free water. The reaction mixes were subjected to an initial denaturation at 98^{o} C for 30 seconds before 25 cycles of denaturation at 98^{o} C for 10 seconds, annealing at 50-72^{o} C for 30 seconds and elongation at 72^{o} C for 3 minutes before a final extension at 72^{o} C for 2 minutes. 3 µl of the PCR product was taken for analysis by agarose gel electrophoresis. A 1 µl sample of this product was added to 5 µl of a unique Kinase-Ligase-DpnI (KLD) reaction buffer. A 1µl sample of KLD enzyme was added plus 3µl of nuclease free water and it was incubated for 5 minutes at room temperature. 5 µl of the KLD reaction was taken for transformation into 50 µl of chemically competent cells using the same protocol as for DH5α cells. Transformants were selected by plating on LB agar containing the appropriate antibiotics and grown overnight at 37^{o} C before being picked into LB media containing the appropriate antibiotics and grown again overnight at 37^{o} C at 250 rpm. The plasmids were recovered using a QIAprep Spin MiniPrep Kit (QIAGEN). Purified plasmids were then sent to GATC Biotech for sequencing using T7F and T7R primers to verify the mutations had succeeded.

12. Sequence alignment of two mutations in cysteine residues Cys 179/Cys181:



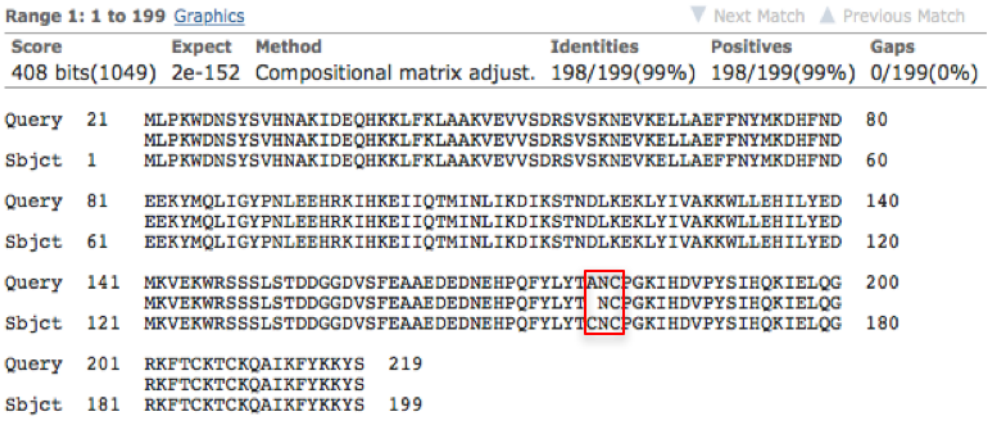
Range 1: 1 to 199 Graphics Next Match Previous Match

Score	Expect	Method	Identities	Positives	Gaps
406 bits(1044)	1e-151	Compositional matrix adjust.	197/199(99%)	197/199(98%)	0/199(0%)

```
Query  21   MLPKWDNSYSVHNAKIDEQHKKLFKLAAKVEVVSDRSVSKNEVKELLAEFFNYMKDHFND  80
            MLPKWDNSYSVHNAKIDEQHKKLFKLAAKVEVVSDRSVSKNEVKELLAEFFNYMKDHFND
Sbjct  1    MLPKWDNSYSVHNAKIDEQHKKLFKLAAKVEVVSDRSVSKNEVKELLAEFFNYMKDHFND  60

Query  81   EEKYMQLIGYPNLEEHRKIHKEIIQTMINLIKDIKSTNDLKEKLYIVAKKWLLEHILYED  140
            EEKYMQLIGYPNLEEHRKIHKEIIQTMINLIKDIKSTNDLKEKLYIVAKKWLLEHILYED
Sbjct  61   EEKYMQLIGYPNLEEHRKIHKEIIQTMINLIKDIKSTNDLKEKLYIVAKKWLLEHILYED  120

Query  141  MKVEKWRSSSLSTDDGGDVSFEAAEDEDNEHPQFYLYTANAPGKIHDVPYSIHQKIELQG  200
            MKVEKWRSSSLSTDDGGDVSFEAAEDEDNEHPQFYLYT N PGKIHDVPYSIHQKIELQG
Sbjct  121  MKVEKWRSSSLSTDDGGDVSFEAAEDEDNEHPQFYLYTCNCPGKIHDVPYSIHQKIELQG  180

Query  201  RKFTCKTCKQAIKFYKKYS  219
            RKFTCKTCKQAIKFYKKYS
Sbjct  181  RKFTCKTCKQAIKFYKKYS  199
```

Query: The sequence translated from sequencing.
Subject: The Cj1224 sequence from the database.

13. Sequence alignment of one mutation in Cys 179 residue:



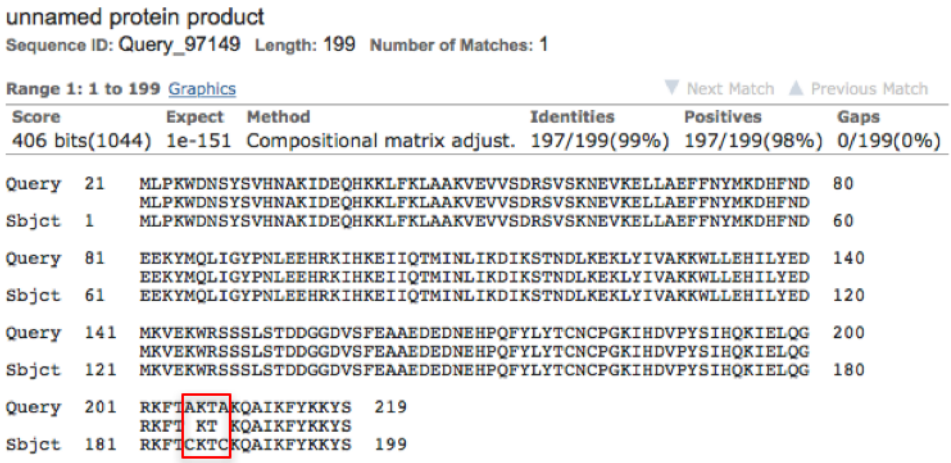
Range 1: 1 to 199 Graphics Next Match Previous Match

Score	Expect	Method	Identities	Positives	Gaps
408 bits(1049)	2e-152	Compositional matrix adjust.	198/199(99%)	198/199(99%)	0/199(0%)

```
Query  21   MLPKWDNSYSVHNAKIDEQHKKLFKLAAKVEVVSDRSVSKNEVKELLAEFFNYMKDHFND  80
            MLPKWDNSYSVHNAKIDEQHKKLFKLAAKVEVVSDRSVSKNEVKELLAEFFNYMKDHFND
Sbjct  1    MLPKWDNSYSVHNAKIDEQHKKLFKLAAKVEVVSDRSVSKNEVKELLAEFFNYMKDHFND  60

Query  81   EEKYMQLIGYPNLEEHRKIHKEIIQTMINLIKDIKSTNDLKEKLYIVAKKWLLEHILYED  140
            EEKYMQLIGYPNLEEHRKIHKEIIQTMINLIKDIKSTNDLKEKLYIVAKKWLLEHILYED
Sbjct  61   EEKYMQLIGYPNLEEHRKIHKEIIQTMINLIKDIKSTNDLKEKLYIVAKKWLLEHILYED  120

Query  141  MKVEKWRSSSLSTDDGGDVSFEAAEDEDNEHPQFYLYTANCPGKIHDVPYSIHQKIELQG  200
            MKVEKWRSSSLSTDDGGDVSFEAAEDEDNEHPQFYLYT NCPGKIHDVPYSIHQKIELQG
Sbjct  121  MKVEKWRSSSLSTDDGGDVSFEAAEDEDNEHPQFYLYTCNCPGKIHDVPYSIHQKIELQG  180

Query  201  RKFTCKTCKQAIKFYKKYS  219
            RKFTCKTCKQAIKFYKKYS
Sbjct  181  RKFTCKTCKQAIKFYKKYS  199
```

Query: The sequence translated from sequencing.
Subject: The Cj1224 sequence from the database.

14. Sequence alignment of two mutations in Cys 205/Cys 208 residues:



unnamed protein product
Sequence ID: Query_97149 Length: 199 Number of Matches: 1

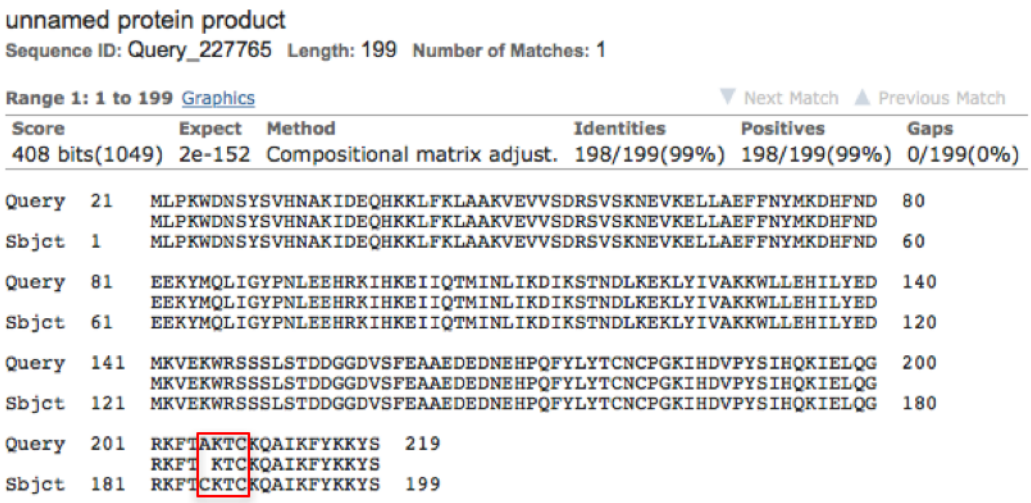
Range 1: 1 to 199 Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
406 bits(1044)	1e-151	Compositional matrix adjust.	197/199(99%)	197/199(98%)	0/199(0%)

```
Query  21   MLPKWDNSYSVHNAKIDEQHKKLFKLAAKVEVVSDRSVSKNEVKELLAEFFNYMKDHFND  80
            MLPKWDNSYSVHNAKIDEQHKKLFKLAAKVEVVSDRSVSKNEVKELLAEFFNYMKDHFND
Sbjct  1    MLPKWDNSYSVHNAKIDEQHKKLFKLAAKVEVVSDRSVSKNEVKELLAEFFNYMKDHFND  60

Query  81   EEKYMQLIGYPNLEEHRKIHKEIIQTMINLIKDIKSTNDLKEKLYIVAKKWLLEHILYED  140
            EEKYMQLIGYPNLEEHRKIHKEIIQTMINLIKDIKSTNDLKEKLYIVAKKWLLEHILYED
Sbjct  61   EEKYMQLIGYPNLEEHRKIHKEIIQTMINLIKDIKSTNDLKEKLYIVAKKWLLEHILYED  120

Query  141  MKVEKWRSSSLSTDDGGDVSFEAAEDEDNEHPQFYLYTCNCPGKIHDVPYSIHQKIELQG  200
            MKVEKWRSSSLSTDDGGDVSFEAAEDEDNEHPQFYLYTCNCPGKIHDVPYSIHQKIELQG
Sbjct  121  MKVEKWRSSSLSTDDGGDVSFEAAEDEDNEHPQFYLYTCNCPGKIHDVPYSIHQKIELQG  180

Query  201  RKFTAKTAKQAIKFYKKYS  219
            RKFT KT KQAIKFYKKYS
Sbjct  181  RKFTCKTCKQAIKFYKKYS  199
```

Query: The sequence translated from sequencing.
Subject: The Cj1224 sequence from the database.

15. Sequence alignment of one mutation in Cys 205 residue:



unnamed protein product
Sequence ID: Query_227765 Length: 199 Number of Matches: 1

Range 1: 1 to 199 Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
408 bits(1049)	2e-152	Compositional matrix adjust.	198/199(99%)	198/199(99%)	0/199(0%)

```
Query  21   MLPKWDNSYSVHNAKIDEQHKKLFKLAAKVEVVSDRSVSKNEVKELLAEFFNYMKDHFND  80
            MLPKWDNSYSVHNAKIDEQHKKLFKLAAKVEVVSDRSVSKNEVKELLAEFFNYMKDHFND
Sbjct  1    MLPKWDNSYSVHNAKIDEQHKKLFKLAAKVEVVSDRSVSKNEVKELLAEFFNYMKDHFND  60

Query  81   EEKYMQLIGYPNLEEHRKIHKEIIQTMINLIKDIKSTNDLKEKLYIVAKKWLLEHILYED  140
            EEKYMQLIGYPNLEEHRKIHKEIIQTMINLIKDIKSTNDLKEKLYIVAKKWLLEHILYED
Sbjct  61   EEKYMQLIGYPNLEEHRKIHKEIIQTMINLIKDIKSTNDLKEKLYIVAKKWLLEHILYED  120

Query  141  MKVEKWRSSSLSTDDGGDVSFEAAEDEDNEHPQFYLYTCNCPGKIHDVPYSIHQKIELQG  200
            MKVEKWRSSSLSTDDGGDVSFEAAEDEDNEHPQFYLYTCNCPGKIHDVPYSIHQKIELQG
Sbjct  121  MKVEKWRSSSLSTDDGGDVSFEAAEDEDNEHPQFYLYTCNCPGKIHDVPYSIHQKIELQG  180

Query  201  RKFTAKTCKQAIKFYKKYS  219
            RKFT KTCKQAIKFYKKYS
Sbjct  181  RKFTCKTCKQAIKFYKKYS  199
```

Query: The sequence translated from sequencing.
Subject: The Cj1224 sequence from the database.