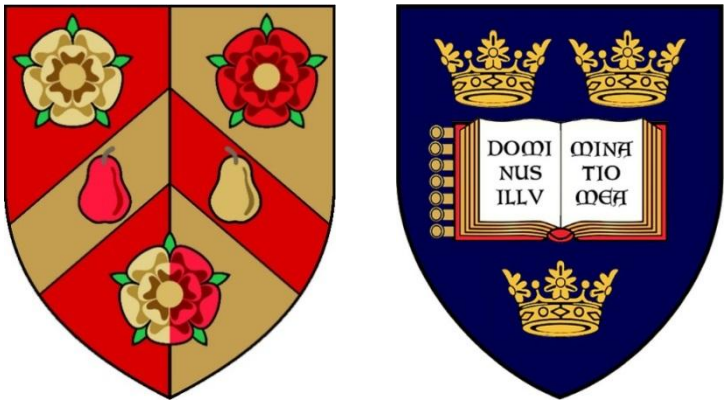
Substrate Recognition by Holocytochrome *c* Synthase in Cytochrome *c* Biogenesis System III



Yulin Zhang

Wolfson College
Department of Biochemistry
University of Oxford
Trinity Term 2015

A thesis submitted to the board of the Faculty of Medical Sciences of the University of Oxford in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Some of the work presented in this thesis has been published:

Zhang, Y., Stevens, J.M. and Ferguson, S.J. "Substrate recognition of holocytochrome *c* synthase: N-terminal region and CXXCH motif of mitochondrial cytochrome *c*." ***FEBS Lett.*** 2014. 588, 3367-74.

Stevens, J.M., **Zhang, Y.**, Muthuvel, G., Sam, K.A., Allen, J.W. and Ferguson, S.J. "The mitochondrial cytochrome *c* N-terminal region is critical for maturation by holocytochrome *c* synthase." ***FEBS. Lett.*** 2011. 585, 1891-6.

Substrate Recognition by Holocytochrome *c* Synthase in Cytochrome *c* Biogenesis System III

Yulin Zhang, Wolfson College, University of Oxford
A thesis submitted for the degree of Doctor of Philosophy, Trinity Term 2014

C-type cytochromes are ubiquitous proteins with crucial functions in organisms, which include electron transfer and apoptotic signalling. In eukaryotic organisms, mitochondrial cytochrome *c* is located in the intermembrane space, and it is a component of the electron transport chain; it is responsible for transferring electrons from Complex III to Complex IV. The regulated release of cytochrome *c* from mitochondria results in the activation of a signal transduction pathway leading to controlled cell death, or apoptosis. In mitochondrial *c*-type cytochromes, the heme is bound to both cysteines of a CXXCH motif located near the N-terminus. The covalent heme attachment in *c*-type cytochromes, the final step in its biosynthesis, is achieved by different cytochrome *c* biogenesis systems in different organisms. Out of these systems, System III, found in many eukaryotes, has a single component - holocytochrome *c* synthase (HCCS) which is the enzyme responsible for the catalysis of heme binding to cytochrome *c*. HCCS recognises apocytochrome *c* as a substrate upon the import of the apocytochrome from the cytosolic space to the mitochondrial intermembrane space. The requirements of amino acid sequence for HCCS recognition had remained an intriguing question, despite the relatively long period since the discovery of the enzyme. Thus, HCCS in System III and its substrate recognition is the subject of this thesis.

This thesis describes the experiments showing that the N-terminal region of the mitochondrial cytochrome *c* protein is important for substrate recognition, as well as further characterisation of this sequence by mutagenesis. Out of several highly conserved residues in the N-terminus, a phenylalanine residue in the N-terminus is identified to be critical for heme attachment by HCCS. The role of this phenylalanine residue in the interaction between the two proteins was probed by substituting it with a range of residues. Furthermore, the importance of the spacing between the key phenylalanine residue and the CXXCH motif was investigated. A single-cysteine variant of the mitochondrial cytochrome *c* with a single bond to the heme is produced by HCCS, but heme attachment only occurs if histidine is present as an axial ligand to the heme iron. Replacement of the histidine with other potential iron-ligating residues abolished heme attachment. These results bring insight into the critical features in amino acid sequence of cytochrome *c* for the substrate recognition specificity of HCCS.

Sequence analysis on the N-terminal region of mitochondrial cytochromes *c* in a variety of organisms reveals evolutionary implications for cytochrome *c* biogenesis systems. It also attempts to explain the reason for negative results in previous chapters for the analysis of the N-terminal region of cytochrome *c*. An improved method for human HCCS production is also described in this thesis, for the exploitation of purification and characterisation in future studies of HCCS.

Acknowledgements

It has been a long and challenging journey. Reaching the final bend and looking back, this thesis could not have been accomplished without the help of many people. In my humble acknowledgement, I would like to take this opportunity to sincerely thank all the individuals who had offered priceless help towards the completion of my project and this thesis.

First of all, I would like to thank my principal supervisor Prof. Stuart Ferguson for accepting me as a D. Phil student, for his foresighted guidance and his precious support, and for his unsurpassed knowledge and wise advice provided throughout my time in this group. Also, I owe a huge debt of gratitude to my immediate supervisor Dr Julie Stevens, for all the motivation, patience and inspiration she gave, as well as countless discussions throughout the four years.

My sincere gratitude is extended to Dr Despoina Mavridou, Dr Shilpa Bali and Dr Lorena Álvarez who took the time out of their busy schedules to provide considerable help for my project and for their invaluable feedback and support throughout the whole four-year period, as well as proofreading of my thesis. I am deeply appreciative of Dr James Allen for his generosity and support in the first two years of my D. Phil course.

Special acknowledgment is also extended to Dr Felicity Alcock, Dr Mark Kail, Dr Michael Lukey, Dr Martin Krehenbrink, Dr Ciarán Kelly, Paraskevi Kritsiligkou, Dr Matthew Young, Dr Marta Wojnowska, Dr Nick Green, Dr Fernanda Rodriguez, Dr Richard Gillams, Dr Sebastian Busch, Karl Brune and all the other members of Ferguson, Berks, McLain and Bayer labs, for their company, help and encouragement.

I am very grateful to my new friends, Shevket and Andrew, for their valuable opinions and discussions regarding my thesis. I also thank Rachel, Catherine, Jo, Cendie, Sam, Matt, Alex, Albert, Drew and all the other companions in the lab I worked with and supervised in the past four years, for the assistance received and unforgettable time spent together. Thanks are also due to my old friends from Imperial College London (Biochemistry year 2010/2011) who came to Oxford: Jacob, Johnny, Joe, Janice, Jerry, Ben, Pirada, Liuqing and Xiao, and I wish them all the best for the future.

I gratefully acknowledge Wolfson College for the financial support for covering the cost of my conference attendance. I also thank all my friends from Malvern, in Oxford and in London for the joy we had together, especially the Sharmans, Lukas, Dan, Dominic, Hao, Weiwei, Yiqun, Aaron and Liz. Finally, I want to thank my family for their continuous support to my aspirations, and I dedicate this thesis to them.

Abbreviations

A	ampere
aa	amino acid
ABC	ATP-binding cassette
ADP	adenosine diphosphate
AP	alkaline phosphatase
ATP	adenosine triphosphate
Bcl	B-cell lymphoma protein
BLAST	basic local alignment search tool
BLASTp	basic local alignment search tool for protein sequences
BSA	bovine serum albumin
^{o}C	degree Celcius
c	centi (10^{-2})
CCB	cofactor assembly on complex C subunit B
CCHL	cytochrome *c* heme lyase
Ccm	cytochrome *c* maturation
CoQ	coenzyme Q
Da	Dalton
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
Dsb	disufide bond oxidoreductase
EDTA	ethylenediaminetetraacetic acid
ESI	electrospray ionisation
FPLC	fast-performance liquid chromatography
g	gram
g	gravity (9.81m s^{-1})
GST	glutathione-S-transferase
HCCS	holocytochrome *c* synthase
HCC_1S	holocytochrome c_1 synthase
IMS	intermembrane space
IPTG	isopropyl-1-thio-b-D-galactopyranoside
k	kilo (10^{3})
l	litre
LB	Luria-Bertani
m	metre; milli (10^{-3})
M	molar
min	minute
MOPS	3-propanesulfonic acid
MLS	microphthalmia with linear skin defects syndrome
NADH	nicotinamide adenine dinucleotide
n	nano (10^{-9})
Nrf	nitrite reductase, formate dependent
OD_{550}	optical density at 550nm
PAGE	polyacrylamide gel electrophoresis
PDB	Protein Data Bank

PCR	polymerase chain reaction
s	second
SDS	sodium dodecyl sulfate
SOB	super optimal broth
TAE	tris-acetate-EDTA
TMBZ	3,3',5,5'-tetramethylbenzidine
TOF	time of flight
TOM	translocase of the mitochondrial outer membrane
Tris	Tris-(hydroxymethyl) aminomethane
Tween-20	polyoxyethylene (20) sorbitan monolaurate
U	unit
UniProt	universal protein resource
UV	ultraviolet
V	volt
v	volume
w	weight
WT	wild-type
$\Delta\Psi$	membrane potential
µ	micro (10^{-6})

Contents

List of figures

List of tables

Chapter 1

Introduction

1.1 Energy transfer in biochemistry

Energy transduction in all organisms is essential for their survival. In the sense of different ways of harvesting energy in this process, organisms are classified into autotrophs and organotrophs based on their ways of utilising energy for activities [1]. The greatest difference between the autotrophs and organotrophs is that organotrophs universally utilise reduced organic compounds for energy, in contrast with the carbon fixation of CO_2 to reduced organic compounds by autotrophs. Therefore, the metabolic diversity in the biosphere can be categorised in terms of organotrophy or autotrophy, and the nature of the electron donor and acceptor [1]. Most complex life forms such as fungi and animals normally require respiration involving oxygen [2]. The oxidative phosphorylation chain, or respiratory chain, transfers electrons from various substrates to oxygen as a terminal electron acceptor, generating a proton electrochemical gradient that drives adenosine triphosphate (ATP) synthase [3], as shown in Figure 1.1. The oxidative phosphorylation reaction results in the formation of ATP by phosphorylation of adenosine diphosphate (ADP). The ATP molecule has been called the ‘molecular unit of currency’ of intracellular energy transfer [4].

A more modern and technically more accurate view is that protons moving down their electrochemical gradient represent the fundamental unit of energy currency. Oxidative phosphorylation occurs in association with the inner mitochondrial membrane of eukaryotes and the cytoplasmic boundary membrane of bacteria and archaea. In photosynthetic organisms, electron transfer chains are again involved in generating a proton electrochemical

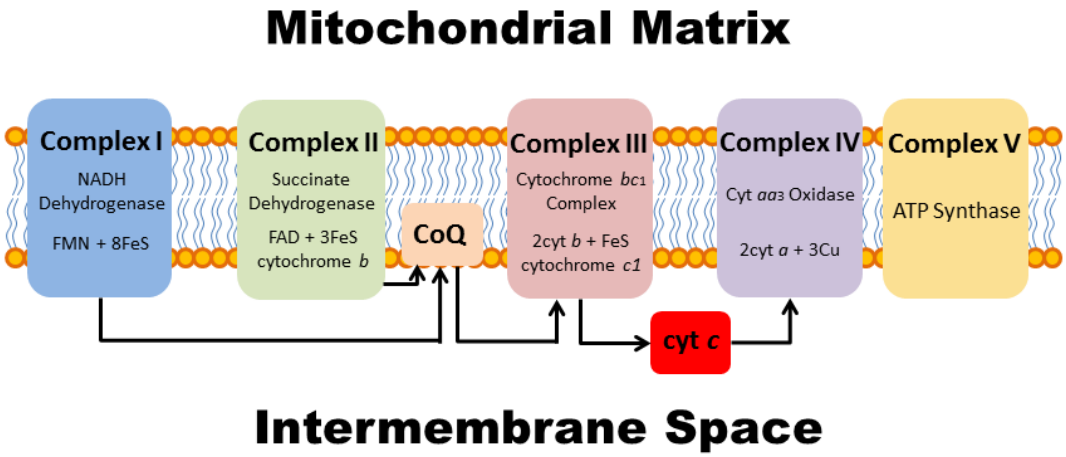


Figure 1.1 Schematic diagram of the mitochondrial respiratory chain. Its components - complexes I, II, III, IV and V are illustrated from left to right embedded in the phospholipid bilayer of the mitochondrial inner membrane. Cytochrome *c* and coenzyme Q are also indicated. The reaction centres and electron carriers are labelled. The flow of electrons between the components is indicated by arrows. Adapted from Nicholls and Ferguson [5].

gradient, in this case using the energy absorbed as sunlight to drive this process rather than the energy released in respiration.

Photosynthetic organisms are not the only autotrophs. There are bacterial species that can grow on inorganic matter and can use respiration both to generate the ATP and the reductant by reversed electron transfer that are needed to fix carbon dioxide. Respiratory or electron transfer chains are not required in organisms or cells which rely on fermentation or an analogous process, for example species of obligate anaerobic bacteria.

1.2 The cytochrome protein family

Energy transduction in organisms usually requires interactions between proteins as part of the electron transfer process. One of the important classes of proteins involved in energy transfer are cytochromes, which are an important heme-associated protein family found in most organisms; they have a variety of functions [6]. Cytochromes have the heme prosthetic group bound to the polypeptide chain, by either non-covalent interactions or covalent bonds. One of the major roles which cytochromes perform is electron carriers, as constituents of electron transport chains [6] as a vital link within energy transduction pathways.

Cytochromes that are members of electron transfer chains are categorised according to the different types of hemes. Cytochrome *c* contains heme *c*, cytochrome *b* contains heme *b*, cytochrome *d* contains heme *d* and cytochrome *a* contains heme *a* [7]. The details for these hemes will be discussed below. Cytochrome *f* is a cytochrome containing heme *c* but with an atypical protein fold for a cytochrome [7]. Not all cytochromes are involved in electron transfer. For example, the well-known cytochrome P450, so called because of its optical spectrum under certain conditions, is a family of proteins that catalyse hydroxylation of substrates and are involved in processes such as synthesis of antibiotics in prokaryotes, synthesis of hormones and other chemicals in plants, and particularly the drug metabolism and steroidogenesis in mammalians [8]. Cytochrome P450 contains non-covalently bound *b*-type heme, and so it can be defined as a member of *b*-type cytochrome family. Other proteins that contain *b*-type heme include, catalase, haemoglobin, myoglobin [9] and nitric oxide synthase.

1.2.1 Heme prosthetic group (heme *b*)

Heme is a prosthetic group found in a number of protein families in the vast majority of living organisms [10,11] with diverse functions including electron transfer, oxygen transport and storage, catalysis, gas sensing, and gene regulation [7]. It is a tetrapyrrole molecule liganding a central iron ion, and it is also called iron protoporphyrin IX (see Figure 1.2). The term *b*-type heme appears to come from the original characterisation of cytochromes and molecules, such as hemoglobin and myoglobin. These proteins are usually described as containing heme rather than *b*-type heme, even though hemoglobin contains the same non-covalently bound molecule as a *b*-type cytochrome. Other molecules with a tetrapyrrole structure, such as chlorophyll, also play important roles in the metabolism in photosynthetic organisms [12].

Heme is an essential component of cytochrome proteins, where it has functions including electron transport, energy generation, and chemical transformation [13]. The heme molecule contains methyl groups in positions 1, 3, 5 and 8; vinyl groups in position 2 and 4; propionic acid groups in positions 6 and 7 [13] (Figure 1.2). The ferrous heme in cytochromes has three characteristic absorption peaks (α, β, and γ) in UV-visible spectra [6], providing a useful method for examining the nature of heme binding, and its oxidation state, within proteins. This ‘standard’ or *b*-type heme is the most ubiquitous type of heme in the natural world. Heme *b* is attached to cytochromes *b* via coordinations of the heme iron via amino acid side chains [14], as well as other non-covalent interactions between the heme and the protein. Examples of *b*-heme include hemes b_{562} and b_{566}, associated with

two cytochromes *b* of the Complex III (Figure 1.1) [15,16]; these are named according to their wavelength

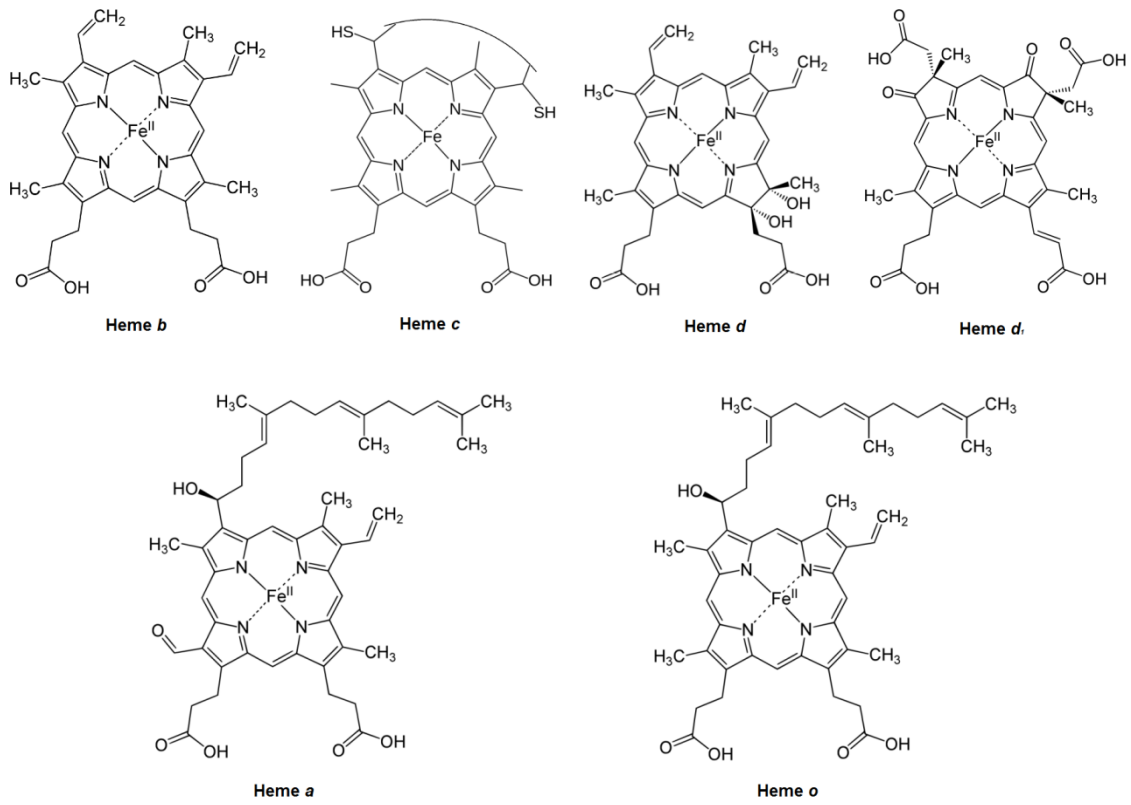


Figure 1.2 The structural formulae of heme *a*, heme *b*, heme *c*, heme *d*, heme d_1 and heme *o*. The polypeptide chain is represented by the arc. Adapted from Bowman and Bren [7], Newton [17] and Caughey *et al.* [18].

absorption properties.

Heme biosynthesis in mitochondria is a complex multi-step pathway with 11 enzymes involved, and the heme precursor is imported and exported across the mitochondrial inner membrane several times [12]. The first step of heme biosynthesis is the production of 5-

aminolevulinic acid by 5-aminolevulinate synthase, which is followed by the condensation of two molecules of aminolevulinic acid to yield one molecule of pyrrole porphobilinogen, a reaction that is catalysed by porphobilinogen synthase. Four molecules of pyrrole porphobilinogen are linked by porphobilinogen deaminase to produce a linear oligomeric pre-uroporphyrinogen that is closed to a ring tetrapyrrole intermediate uroporphyrinogen III by uroporphyrinogen III synthase. This cyclic molecule is then sequentially modified by uroporphyrinogen III decarboxylase, oxygen-dependent coproporphyrinogen III oxidase, coproporphyrinogen III dehydrogenase and protoporphyrinogen IX oxidase to form the protoporphyrin IX. Ferrochelatase then inserts the iron ion into this ring to complete the final step [12]. This final step takes place in the matrix side of the inner mitochondrial membrane. How heme is transported to other locations in the cell (for example, some cytochrome P450 molecules associated with the endoplasmic reticulum), is unknown. Heme synthesis in bacteria is more diverse than the mitochondrial pathway [12], but as the work described in this thesis is concerned with mitochondrial heme protein assembly, this topic of bacterial heme synthesis is not discussed here.

1.2.2 Heme *a*, heme *d* and heme *o*

Different modifications on heme *b* result in other forms of heme with specific functions. Heme *a* differs from the 'standard' *b*-type heme in respect of two modifications to the tetrapyrrole ring. These are the hydroxyfarnesylation at the vinyl 2 position and the formylation at the methyl 8 position [7]. This *a*-type heme is only found in many cytochrome oxidases of respiratory chains. Seemingly there is no explanation for this advantage of the modification of the tetrapyrrole ring at this moment. One kind of

cytochrome oxidase is often called cytochrome aa_3, but there is only one type of *a*-type heme present [5]. The subscript 3 denotes that one of the hemes is in a different environment than the other.

O-type heme is a species that is formed on the pathway to *a*-type from *b*-type heme. An enzyme called heme *o* synthase catalyses a farnesyl addition to vinyl group 2 on heme *b* to produce heme *o*, and heme *a* synthase catalyses the modification from heme *o* to heme *a* [19]. *Escherichia coli*, for example, lacks an enzyme that catalyses the conversion of *o*-type heme to *a*-type heme, and hence it can produce a cytochrome *bo* oxidase molecule that contains one molecule of the *b*-type heme and one of the *o*-type.

D-type heme is found in the bacterial cytochrome *bd* oxidases and is different from *b*-type heme with the modification on its propionic acid side chain [17]. Finally, there is the d_1 heme that is found in one type of bacterial cytochrome cd_1 nitrate reductase. It is not a standard heme, because the tetrapyrrole ring is partially saturated and modified in other ways [7,20].

1.2.3 *C*-type cytochromes and 'heme *c*'

The important characteristic for heme *c* is the covalent bonds formed between the two heme vinyl groups and the protein to which it is attached. Typically, two cysteine residues in a highly conserved CXXCH motif form thioether bonds to the heme group [7]. Thus there is really no such separate molecule as a *c*-type heme, because it cannot exist as a free species. A *c*-type cytochrome actually contains *b*-type heme (s) covalently attached to the

polypeptide chain. The covalent heme attachment in *c*-type cytochromes has been conserved in organisms for a very long time [19].

The *c*-type cytochrome family is diverse, and Ambler [21] had categorised most of its members into four classes. Class 1 of cytochrome *c* is the largest group in the family, and it includes small (8-12 kDa) and hydrophilic globular proteins usually with one heme bound. Examples of Class 1 are shown in Figure 1.3. Class 2 cytochromes *c* are also monoheme, and they have a four-helix bundle fold. Class 3 are multiheme *c*-type cytochromes with a low reduction potential, and the heme axial ligation is typically bis-His. Class 4 members are multiheme proteins with a high molecular mass (>40kDa) and a reaction centre constituted by these heme groups. The hemes in Class 4 cytochrome *c* usually have His-Met or bis-His axial ligation [7]. Examples of *c*-type cytochromes from Class 2, 3 and 4 are shown in Figure 1.4.

It is an important question as to why this type of covalent heme binding exists. The exact reason for this phenomenon is still debated. However, there have been a number of suggestions made in recent years. As has been discussed before, Class 3 and Class 4 *c*-type cytochromes consists of many members with multiple hemes and a high heme:polypeptide chain ratio [7,22]. It has been proposed [22-24] that the covalent tethering of the heme in cytochromes *c* enables a more compact packing of hemes, and a smaller number of amino acids residues per heme than would be the case if separate heme pockets, as found for example in hemoglobin, were to be used. Since this suggestion was made, there have been many more of this kind of multiheme *c*-type cytochrome structures reported [22]. One of the most notable new proposals is the suggestion of a heme wire across the bacterial outer

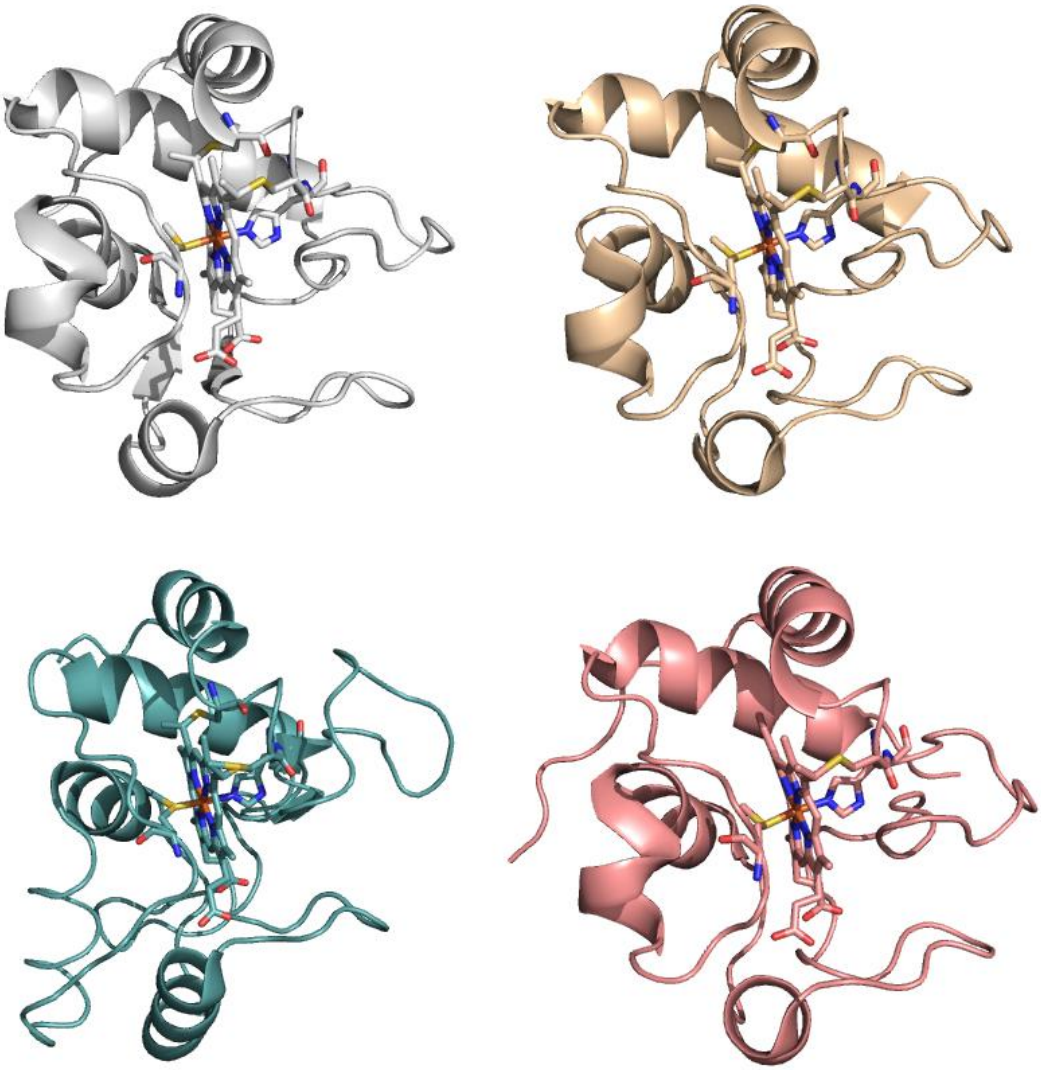


Figure 1.3 3-D structures of monoheme *c*-type cytochromes in different organisms: fungus *Sacchromyces cerevisiae* iso-1-cytochrome *c* (Protein Data Bank code: 1YCC) in grey; mammal *Equus caballus* heart cytochrome *c* (Protein Data Bank code: 1HRC) in light orange; bacteria *Paracoccus denitrificans* cytochrome c_{550} (Protein Data Bank code: 155C) in teal; trypanosomatid protozoan *Crithidia fasciculata* cytochrome *c* (Protein Data Bank code: 2YK3) in flesh tint. Note the model for *P. denitrificans* cytochrome c_{550} has a suspected alpha helix comparable to the other three models in the lower left corner in the diagram (between residues P81-T89), and the failure for recognition of this helix might be due to the limitations of the software used to produce the figure.

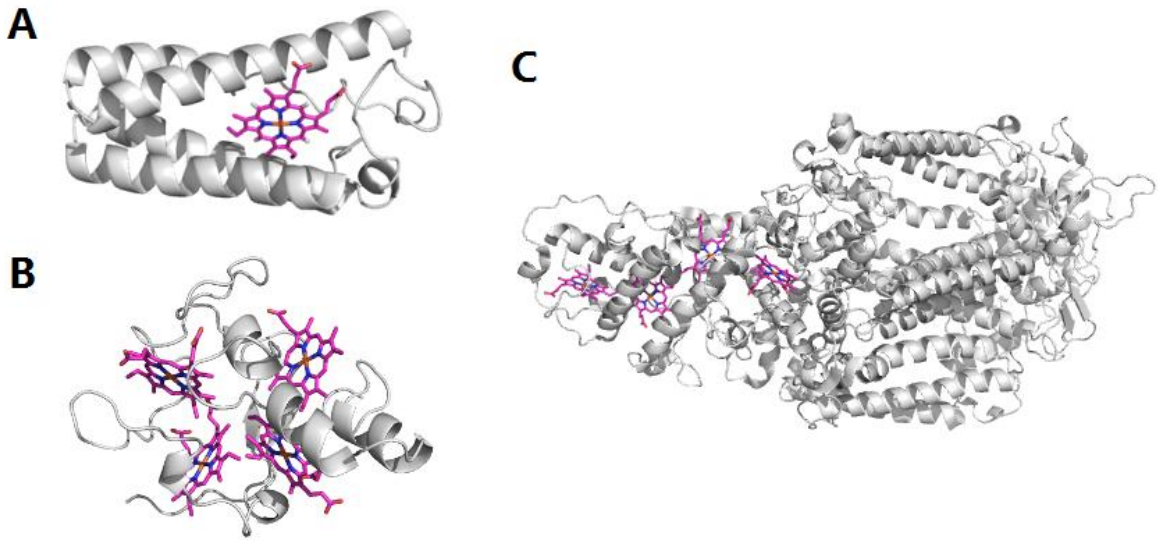


Figure 1.4 3-D structures of examples of *c*-type cytochromes from classes 2, 3 and 4: (A) Class 2: *Rhodopseudomonas capsulatus* cytochrome c' (PDB entry 1CPQ); (B) Class 3: *Desulfovibrio vulgaris Hildenborough* cytochrome c_3 (PDB entry 1CTH); (C) Class 4: Photosynthetic complex centre of *Rhodopseudomonas viridis* (PDB entry 1PRC). The heme prosthetic groups are highlighted in magenta colour.

membrane that would allow electron conduction to/from the periplasm to/from the exterior of the cell [25]. The advantage of high heme density in protein cannot be extrapolated as the reason for existence of the monoheme mitochondrial cytochrome *c*, and the bacterial equivalent cytochrome *c* (termed c_2) and c_1 (both with only one heme binding pocket) though. It was suggested that the covalent binding of the heme in *c*-type cytochromes was an evolutionary remnant [23]. Cytochromes *c* and c_1 first arose in the periplasm of Gram-negative bacteria, and there could be a risk of losing the valuable iron-containing heme, if non-covalently attached to the protein, from the microorganism to the exterior media. Hence, it was proposed that covalent binding between the heme and the polypeptide chain evolved in order to prevent heme loss [23]. Recently, there have been arguments against

this view. First, it was later discovered that Gram-negative bacteria and mitochondria in many eukaryotes use different cytochrome *c* biogenesis systems [26]. Any redundancy of covalent attachment in mitochondria would not be able to explain the need for the emergence of a separate System III for mitochondria *c*-type cytochrome biogenesis after its evolution from Gram-negative bacteria. Second, since this hypothesis was proposed, it has emerged that there are in fact *b*-type cytochromes in the periplasm (one example is cytochrome b_{562} in *E. coli* [27]). This suggests the loss of heme from *b*-type cytochromes in the periplasm is not significant enough to provide a selection pressure for the evolution of the covalent binding in *c*-type cytochromes.

Another approach to investigate the possible advantage of *c*-type cytochrome would be if an analogue of mitochondrial cytochrome *c* with heme non-covalently bound could function in the mitochondrial respiratory chain. This would examine the significance of the covalent linkage in the biological function of cytochromes *c*. This specific type of experiment has not been carried out, but two thermophilic bacterial monoheme *c*-type cytochromes, with structural similarity to mitochondrial cytochrome *c*, have been converted to non-covalently heme binding *b*-type cytochromes, namely *Hydrogenobacter thermophilus* c_{552} and *Aquifex aeolicus* cytochrome c_{555} [28,29]. In both cases, the proteins retained the native overall fold and adopted the non-covalent binding with the heme to become essentially a *b*-type cytochrome, although the stability of the resultant proteins is decreased. For the *H. thermophilus* protein, the redox potential for the *b*-type variant is lower than that of the wild type (170mV rather than 245mV) [28]; but according to another study, the redox potential values of variants with a single cysteine bond are similar to wild type (225mV and 250mV to 245mV) [30]. Thus, it is not clear that the presence or absence

of the thioether bonds has much direct effect on the redox potential of the cytochrome *c*. The latter depends on a number of complex factors essentially depending upon the environment of heme within the protein. Given that redox potential depends on the environment of the heme, a reasonable explanation is that the redox potential in the *c*-type cytochrome is adjusted by the covalent bonds, as a consequence of the precise environment of the heme [31]. Therefore the covalent bonds enable its function within the mitochondrial respiratory chain.

The improved stability of cytochrome *c* to cytochrome *b* is also a potential advantage [31]. The conversion of the *H. thermophilus* cytochrome c_{552} to a *b*-type cytochrome is accompanied by a decrease in stability of the protein, according to titrations of protein unfolding under temperature or pH change and as a function of guanidine hydrochloride concentration, as judged by observation of decrease in the α-helical proportion of the protein according to the UV CD spectra. [28]. Similarly, conversion of *E. coli* cytochrome b_{562} into a *c*-type cytochrome should be correlated with an increase in stability [27]. If the loss of stability of *H. thermophilus* cytochrome c_{552} when it is converted to a *b*-type cytochrome can be extrapolated to mitochondrial cytochrome *c*, then a *b*-type cytochrome with the same fold as mitochondrial cytochrome *c* might be insufficiently stable. However, stability of proteins depends on many factors: the stability of *Pseudomonas aeruginosa* cytochrome c_{551} can also be enhanced by several mutations based on the amino acid sequence of the hyperthermophilic *A. aeolicus* c_{551} [32]. An *A. aeolicus* cytochrome c_{551} without thioether bonds can form a stable structure, which is virtually same as a *b*-type cytochrome [29]. Currently, it is not clear why mitochondrial cytochrome *c* and c_1 could not be replaced by *b*-type cytochromes of similar stability and redox potential. The fact

that this has not happened naturally suggests that there is an alternative reason why such a molecule could not be achieved by natural selection.

1.3 Cytochrome *c* biogenesis systems

Cytochrome *c* biogenesis is the process of heme attachment to the apocytochrome *c* to form the heme-attached cytochrome *c* or holocytochrome *c*. This process also leads to the folding of apocytochromes *c* which almost always are either partially folded or unfolded [22]. This process is always catalysed by biogenesis proteins of varying complexity. It has long been known that there are different cytochrome *c* biogenesis enzyme systems utilised by different organisms and organelles [26]. Four major systems have been defined from studies of cytochrome *c* synthesis in a variety of organisms, and they have been named Systems I, II, III and IV [33-35]. System I is utilised by the majority of Gram-negative bacteria, archaea, the mitochondria in plants, red algae and some other organisms; System II is found in most Gram-positive bacteria, β-, δ-, and ε-proteobacteria, cyanobacteria and chloroplasts of plants [36]; System III is used by some protozoa and the mitochondria of animals and fungi and System IV in oxygenic photosynthetic organisms [37,38]. The major components of System I, II and III are shown in Figure 1.5. There are also the as yet unidentified cytochrome *c* biogenesis Systems V and VI found in euglenoid species in the kingdom *Exavata*, and *Bacillus* and heliobacteria species respectively [35]. The common feature of Systems I, II and III is the presence of a conserved heme-binding motif on the cytochrome, but the maturation systems themselves display a surprising level of diversity

in their number of protein components, structures and subcellular locations. In cytochrome b_6f complex in

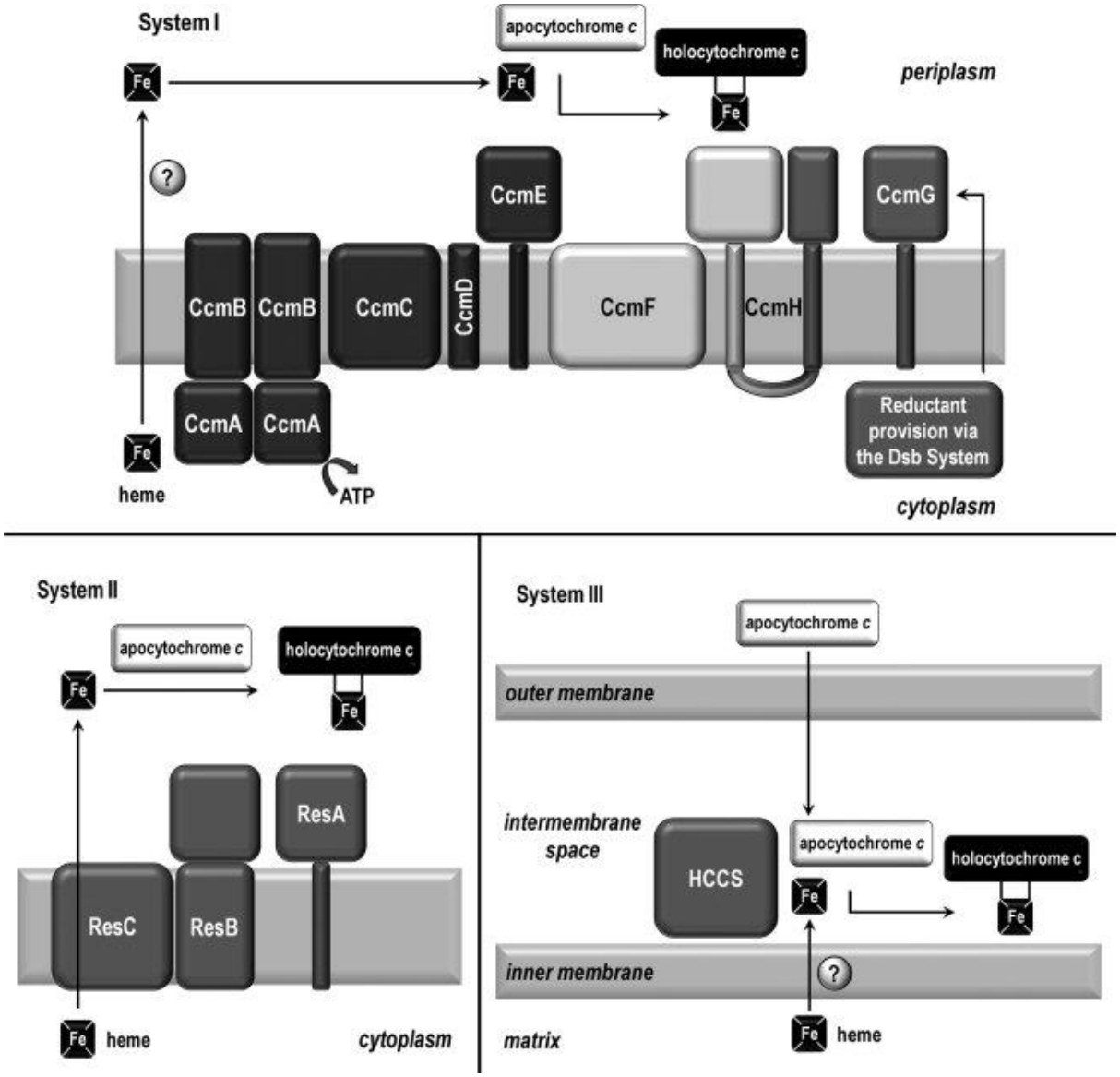


Figure 1.5 The three major cytochrome *c* maturation Systems I, II and III, based on *Escherichia coli*, *Bacillus subtilis* and *Saccharomyces cerevisiae* respectively. Adapted from Mavridou *et al.* [39].

photosynthetic species, there is also a unique heme called heme c_i which is covalently bound to a cysteine on cytochrome b_6 via one thioether bond whose formation is catalysed by System IV. The hypothetical System V catalyses the attachment of heme to a single cysteine in the euglenoid cytochromes *c* (such as the trypanosome cytochrome *c*) [35]. Systems V and VI are unique because the euglenoid genomes lack any of the holocytochrome *c* synthetic proteins from Systems I-IV [40], and genomes of Firmicutes lack any of these proteins too [41].

1.3.1 System I

The alternative name for System I is the cytochrome *c* maturation system (Ccm System). In *E. coli* the system contains eight components, *CcmABCDEFGH*, which are expressed as a single operon. In some organisms, the two domains of CcmH (N-terminal and C-terminal) exist as two separate polypeptides known as CcmH and CcmI [42]. The other two components in this system are DsbA and DsbD, proteins which are responsible for the disulphide interchange reactions of the apocytochrome *c*. The oxidising environment of the bacterial periplasm leads to oxidation of the cysteine residues from the heme-binding CXXCH motif to bind together to form an internal disulphide bridge, which must be prevented. DsbA and DsbD are necessary to carry out the delicate control mechanism of the oxidation and reduction of these cysteines [42]. Most of the components of System I are integral membrane proteins or membrane anchored proteins [42], as shown in Figure 1.5 (based on *E. coli*).

The CcmA, CcmB and CcmC components seem to form a putative ATP-binding cassette (ABC) transporter which is involved in the heme transfer process to the protein CcmE. CcmA is the cytoplasmic ATP hydrolysis subunit and CcmB and CcmC are transmembrane components which may pass heme to the CcmE protein [42,43]. The nature and mechanism of the heme transport in System I is debated; there is currently no evidence to demonstrate how heme is transported across the membrane to the periplasm. It is thought that a conformational change, brought about by ATP hydrolysis by CcmA, is transferred to CcmC, which allows heme to be transferred to CcmE [44]. CcmD is a small membrane protein that facilitates the interaction of other components of the system [42]. CcmE is a heme chaperone and is the pivotal component of the system [45]. It was discovered that CcmE can bind heme covalently via a histidine residue [46] and deliver the heme to the apocytochrome *c* [42]. The heme is released from CcmE with the aid of CcmC, CcmD and CcmF [47,48]. CcmG is a periplasmic thioredoxin, and N-terminal CcmH acts as a thiol-disulfide oxidoreductase [49]. They are thought to be involved in the reduction of the cysteines in the CXXCH motif of apocytochromes [42]. CcmG is in turn reduced by the DsbD protein. DsbD is a unique transmembrane protein that is able to initiate a thiol-disulfide cascade to transfer reductant from cytoplasmic therodoxin to the periplasmic side [42] and is a vital component of cytochrome *c* biogenesis in System I. C-terminal CcmH or CcmI is a tetratricopeptide repeat containing protein that is thought to interact with apocytochrome to facilitate productive interactions [50]. The interactions between CcmE, CcmF and CcmH are important for the heme attachment reaction [42]. It is proposed that these components are able to form protein supercomplexes during their function [39,44,51]. A variant system called System I* exists in some sulphate-reducing bacteria that lacks N-terminal CcmH or CcmG, and has cysteine instead of histidine in the

heme binding motif in CcmE [52]. This System I* can function heterologously in *E. coli*, whereas the typical System I cannot function without the CcmH or CcmG components [42]. The bacterial nitrite reductase NrfA is also a *c*-type cytochrome utilising a Nrf operon (Nrf*EFG*) as part of its biogenesis which is functionally similar to System I [42,53]. The NrfE, NrfF and NrfG are orthologues of CcmF, N-terminal CcmH and CcmI/C-terminal CcmH [42]. It is significant that several different maturation proteins are needed simply because the substrate has a heme-binding motif with a different heme ligand, i.e. CXXCK instead of CXXCH. The substrate specificity of System I is otherwise very broad, and it will attach heme to a large array of proteins that feature a CXXCH motif. [54,55]

1.3.2 System II

System II, which is also termed as the cytochrome *c* synthesis system (Ccs System), has fewer components than System I, with only three or four components (ResA/CcsX, ResB/CcsB/Ccs1, ResC/CcsA and CcdA or DsbD) [56]; different components or orthologues exist in different organisms [57], and the cartoon representing System II shown in Figure 1.5 is based on *Bacillus subtilis*. The disulphide reductant component is either CcdA or DsbD in different organisms, and the disulphide bond formation could be catalysed by Bdb proteins BdbD and BdbC [33,37]. CcsA/ResC and CcsB/ResB form a heterodimeric complex System II CCS (CcsBA/ResBC, or CCS_{II}), and this module exists as a fusion protein in some ε-proteobacteria species [57]. It has been proposed that this CCS_{II} complex catalyses heme transport. It has substrate recognition for apocytochrome, and catalyses the thioether bond formation for heme attachment. System II proteins are evolutionarily linked to CcmC and CcmF of System I [57]. Two highly conserved histidine

pairs within CCS_{II} were found to be essential for the maturation of cytochrome *c* [58]. One of the isoforms of CcsBA-type protein in the ε-proteobacteria *Wolinella succinogenes*, NrfI, is designated to attach heme to the nitrate reductase NrfA [57].

Organisms that contain Systems I and II use the ubiquitous secretory (Sec) pathway to direct the apocytochrome to the periplasm. A signal peptide sequence on the apocytochrome is cleaved by a peptidase upon import into the periplasm [59].

Almost all of the plants use System I for cytochrome *c* biogenesis in mitochondria, and they use System II to produce cytochrome *c* (including cytochrome *f*) in chloroplasts (situated in the thylakoids) [37], in addition to System IV, resulting in three different cytochrome *c* biogenesis systems coexisting in a single plant cell.

1.3.3 System V and System VI

The organisms in the phylum *Euglenozoa* have mitochondrial cytochrome *c* with only one thioether bond formed with the heme, and although it has a very similar structure and chemical properties with other cytochromes *c* in System III [35,60], there is a lack of any of the functional proteins from System I-IV in these organisms, such as species in the genera *Trypanosoma* and *Leishmania*, from genomic evidence [40,41]. This system is termed as System V; however, no clear functional proteins of this system have been identified so far [38]. System VI is the term for a speculated cytochrome biogenesis system

in firmicutes which includes some *Bacillus* species; in this system, a cytochrome *b* subunit of the cytochrome *bc* complexes (similar to the cytochrome b_6f complex in thylakoid) has a heme c_i covalently bound [35,61]. The components of System IV (which will be discussed in Section 1.3.4) which are responsible for the heme c_i attachment in oxygenic photosynthetic species are missing in *B. subtilis* and other firmicutes, hence the putative System VI is defined [62].

1.3.4 System III and System IV

System III is a single component system with the holocytochrome *c* synthase being the only component, with cytochrome *c* as its substrate, shown in Figure 1.5. The system is much simpler than the other systems, which, as described before, have multiple components present with different specific roles in cytochrome *c* assembly [37]. In some eukaryotic organisms such as *Saccharomyces cerevisiae* (baker's yeast), there are two HCCS homologues which are responsible for binding heme to cytochrome *c* or cytochrome c_1, and they are termed HCCS and HCC_1S respectively according to their specific substrates [63,64]. It is also reported that there is another component in the fungal System III, Cyc2p, as a NAD(P)H-dependent heme reductase [65-67]. It may have additional functions in other biochemical processes in mitochondria [35]. Its mutation leads to a significant drop in cytochrome *c* maturation in yeast [68], but it is only present in fungi, algae and protozoa but not in other organisms using System III [38]. In animals and some other organisms such as *Amoebozoa*, only one homologue of HCCS exists, and its function is assumed to catalyse the heme attachment to both cytochrome *c* and cytochrome c_1 [38]. Despite its simplicity, the substrate recognition of apocytochorome *c*,

was still a mystery and remained unresolved at the outset of the work discussed in this thesis. Further discussion of the progress on this system is included in the introductions of the result chapters (mainly Sections 3.1, 4.1 and 5.1).

System IV is an unusual cytochrome *c* biogenesis system. It is specifically designated for the heme attachment of heme c_i on cytochrome b_6 of the cytochrome b_6f complex (a homologue of the cytochrome bc_1 complex in the electron transport chain) [44]. Heme c_i is only attached to the cytochrome b_6 by a single thioether bond. System IV has at least four components (CCB1, CCB2, CCB3 and CCB4) which are all integrated into thylakoid membranes, and the active sites are on the negative side of the membranes [62]. It is suggested that these four proteins are conserved among all organisms performing oxygenic photosynthesis [62]. A model of CCB mediated heme attachment to cytochrome b_6 has been proposed, and this model involves formation of transient complexes of CCB1-b_6, CCB3-b_6 and CCB2-CCB4-CCB3-b_6 [62].

1.4 Biological functions of mitochondrial cytochrome *c*

Similar to other cytochromes, *c*-type cytochromes in mitochondria are involved in biological functions vital for the survival of organisms, such as electron transfer [6]. Mitochondrial cytochrome *c* is involved in cellular respiration and programmed cell death or apoptosis [69]. The mitochondrial respiratory chain is a critical system existing in the mitochondria in respiring eukaryotic cells. Electron transfer, coupled with proton pumping, results in the energy transduction to yield ATP [5]. There are roughly 20 discrete electron

carriers in this system. Most of these electron carriers are embedded within protein supercomplexes with many subunits, and these complexes are termed Complex I (NADH dehydrogenase), II (Succinate dehydrogenase), III (Cytochrome bc_1 complex), IV (Cytochrome aa_3 oxidase) and V (ATP synthase), as shown in Figure 1.1 [5]. There are several cytochrome respiratory chain components which act as electron carriers, and one of these components, cytochrome c_1, is a subunit of Complex III in the system [70]. The covalent or non-covalent binding of heme is a necessary step in the formation of these complexes with cytochrome components. [71]. Complex III catalyses the electron transfer from another prosthetic group in the mitochondrial inner membrane - Coenzyme Q - to cytochrome *c* (Figure 1.1) [72]. Complex IV – another component of the electron transport chain – is termed cytochrome *c* oxidase because it receives electrons from ferric cytochrome *c*. Unlike Coenzyme Q, cytochrome *c* is located in the intermembrane space and is a free protein disassociating to and from the membrane in the respiratory chain. Hence, it acts as a mobile electron carrier between Complex III and Complex IV, with the heme changing between ferrous and ferric states [7]. Cytochrome *c* is also a crucial signalling factor in the initiation of apoptosis and other related signalling pathways [73,74]. The biological apoptotic function of mitochondrial cytochrome *c* will be further discussed below.

The cytochrome c_1 component of Complex III is a protein with a globular domain with structural similarity to mitochondrial cytochrome *c* [35], although the overall sequence similarity is only 13.7% (proteins in *S. cerevisiae*). Cytochrome c_1 is anchored into the membrane and thus to Complex III by a C-terminal transmembrane helix.

1.5 Holocytochrome *c* synthase

Holocytochrome *c* synthase (HCCS), also termed originally as cytochrome *c* heme lyase (CCHL), is the enzyme responsible for the catalysis of heme attachment for cytochrome *c* in System III [33]. This holo- prefix means heme-bound, whereas the apo- prefix means heme-free. It was discovered that the *cyc3* and *cyc2* genes encode HCCS and HCC_1S proteins respectively [63,68,75]. One example, *S. cerevisiae* HCCS, is a 31 kDa single-chain protein. These proteins are not found related to other with protein families or been found to share close homology with other proteins. HCCS is a membrane associated protein on the mitochondrial inner membrane, but does not show the features of a typical membrane anchor or transmembrane domains. It does not traverse the intermembrane space [76]. HCC_1S is a homologue of HCCS, and the yeast HCC_1S has 26% identity and 38.2% similarity to the yeast HCCS protein [35]. In humans, other animals and some other eukaryotic organisms, there is only one copy of HCCS with no isoforms of this protein present [35]. HCCS recognises the apocytochrome imported into the mitochondrial intermembrane space and subsequently catalyses the formation of thioether bonds between heme and the polypeptide chain of cytochrome *c* [35].

HCCS had not been purified at the outset of the present work and since the discovery of the enzyme [63] the direct analysis of its properties has been limited. Therefore, the substrate recognition of HCCS enzymes has been an intriguing topic with very little site-directed mutagenesis having been performed on HCCS except in some disease related positions [35].

C-type cytochrome maturation by HCCS and HCC_1S has been said to be NADH and flavin dependent [77,78] in yeast, but as mentioned before, there is no evidence for this in all organisms. HCC_1S is not able to mature cytochrome *c*, but HCCS is able to mature cytochrome c_1 at a lower level than HCC_1S [79].

It was reported that HCCS would not attach heme to structurally related bacterial *c*-type cytochromes and that one or more CP motifs (a cysteine-proline sequence has often been considered as a heme binding motif in several proteins) in the HCCS sequence were found to be important for HCCS function [80]. The importance of the CP motifs has been questioned recently [81].

A potential breakthrough in understanding HCCS came when it was shown that it was active in the cytoplasm of *E. coli* [82]. Assuming that the *E. coli* cytoplasm did not contain a fortuitous protein partner of HCCS, this finding shows that HCCS can function alone, at least in the reducing environment of the *E. coli* cytoplasm. However, it should be noted that the environment in the mitochondrial intermembrane space is more oxidising and thus a system analogous to Cyc2p in yeast and other fungi (Section 1.3.4) may be needed in all mitochondria containing organisms [67]. One example is a recently discovered thiol-disulfide oxidoreductase protein AtCCMH in the plant *Arabidopsis thaliana* [83].

1.5.1 Import of HCCS and HCC_1S into mitochondria

HCCS (and HCC_1S) are synthesised in the cytosol and imported into mitochondria. The machinery for the import of HCCS and HCC_1S includes Tom20 and Tom22 proteins of the translocase of the outer membrane (TOM) complex in the mitochondrial outer membrane [76,84]. It does not require ATP hydrolysis or a membrane potential, and the import is not typical in that sense [75]. The folding of the HCCS/ HCC_1S plays a negligible role in the import too, as proven by the successful import of a 25 kDa folded fragment of HCC_1S [85]. The mitochondrial import signals for HCCS and HCC_1S have been identified as the two stretches of highly conserved amino acid residue sequence in the third quarter of the HCCS; they are hydrophilic and are not proteolytically cleavable [75,76].

1.6 Mitochondrial cytochrome *c*

As described above, *c*-type cytochromes are essential components in the mitochondria. In some organisms, mitochondrial cytochrome *c* exists in two isoforms. One example is the yeast species *S. cerevisiae*, where there are two isoforms, cytochrome *c* iso-1 and cytochrome *c* iso-2, encoded by *cyc1* and *cyc7* genes respectively. It has been proposed the expression levels of these isoforms are related to growth conditions [86]. Some other organisms, such as *Caenorhabditis elegans*, *Drosophila melanogaster*, and even *Mus musculus*, also have this dual-cytochrome *c* system [87]. Humans have different copies of the cytochrome *c* gene, but only one gene is expressed [88]. Apocytochromes *c*, regardless

of which organism they belong to, are usually unfolded, but sometimes may have transient structures in the process (helical structures may present in the intermediate species) [89,90]; they only fold into a structure upon the heme attachment to the polypeptide chain [35,37].

1.6.1 Structure of apocytochrome *c*

It is widely believed that the apoform of mitochondrial cytochrome *c* is relatively unstructured. Certain observations with circular dichroism show that the protein from horse heart the characteristic fold of the holoprotein is absent from the apoprotein [90].

There may be some structure associated with horse heart apocytochrome *c*, as Daltrop and Ferguson [90] showed that heme would readily bind non-covalently, implying that a nascent heme binding site may be present in the apoprotein. The work with horse heart cytochrome *c* is generally supposed to be capable of extrapolation to other mitochondrial cytochromes *c*. However, there is little direct evidence for this. In the case of bacterial cytochromes *c*, much less is known about the apoproteins. In the case of *Paracoccus denitrificans* cytochrome c_{550}, Daltrop and Ferguson [90] showed that the apoprotein lacked the structural features of the holoform as judged by circular dichroism spectroscopy, and the fluorescent spectra result also led to the same conclusion. There are one or two exceptional *c*-type cytochromes from thermophiles which do retain varying extents of structure in the apoforms [29,32].

There is scarce evidence about the structural state of other apocytchromes *c*, whether the example is cytochrome c_1 or the many diverse multiheme cytochromes found in bacteria. It is assumed that all these proteins have very different structures, if indeed they have structure at all, in the apo relative to the holo states.

The work in this thesis is concerned with the recognition of apocytochrome *c* substrates by the HCCS enzyme. It is not known whether there are any elements of local structure in the apocytochromes *c*. If there are such features, and if they are important for recognition of the apocytochrome by HCCS, then any amino acid substitution in the cytochrome could affect such structure and thus recognition by HCCS. Equally, HCCS may recognise individual amino acids in a completely unstructured part of the apocytochrome. These two extreme possibilities cannot be distinguished in the work described in this thesis.

1.6.2 Heme-binding motif CXXCH and its variations

The amino acid sequence of the cytochrome *c* family contains a highly conserved heme-binding motif, which is usually defined by the amino acid sequence of CXXCH [69]. There are known variants of this motif existing in many organisms [7].

The 2-vinyl and 4-vinyl groups of heme *c* form thioether bonds with the thiol-group of the cysteine residues under reducing conditions [91] to bind the heme covalently to the protein [7]. The histidine residue acts as an axial ligand to the heme and there is a methionine that

acts as another axial ligand in mitochondrial and many, but not all *c*-type cytochromes [22,35]. The X residues represent various residues in different organisms.

1.6.2.1 AXXCH/FXXCH motif

The cytochrome *c* proteins in the protozoan phylum *Euglenozoa* exhibit a heme binding motif with sequence AXXCH or FXXCH in cytochrome *c* and cytochrome c_1 respectively [92,93]. Under these circumstances the attachment of heme requires only one thioether bond. The cytochrome *c* biogenesis system associated with these organisms is classified as System V, as mentioned above [35]. Only one thioether bond through the 4-vinyl group on the heme is formed [35]. The HCCS enzyme of System III is not found as the biogenesis system of these organisms [38]. There are other less common variations of the motif [7]. A 3-D representation of the CXXCH and AXXCH motifs is shown in Figure 1.6.

1.6.3 Import of apocytochrome *c* into mitochondria

The import of mitochondrial cytochrome *c* across the mitochondrial outer membrane involves Tom40 and Tom22 of the TOM complex [84]. One of the features of this import is the lack of a cleavable signal sequence, unlike the import of *c*-type cytochromes into bacterial periplasm using System I [94,95]. The import also does not involve ATP hydrolysis and the membrane potential as sources of energy input [96,97]. The transport of unfolded apocytochrome into the intermembrane space is reversible as the apocytochrome is able to diffuse out through the outermembrane [98]. The apocytochrome *c* heme

attachment by HCCS happens after the import and prevents the fully folded holoprotein from moving back to the cytoplasm [35].

1.6.4 Import of apocytochrome c_1 into mitochondria

Mitochondrial import of apocytochrome c_1 is via a different pathway, and is more complicated than the process of apocytochrome *c* import [99]. The mechanism includes the cleavage of a signal peptide by matrix processing peptidase upon its insertion into the mitochondrial matrix [78]. There are two signal sequences in cytochrome c_1, one is at the N-terminal region and the other one is internal [100]. The internal signal sequence is cleaved after heme attachment by HCC_1S and the hydrophilic domain of the holocytochrome folds [100].

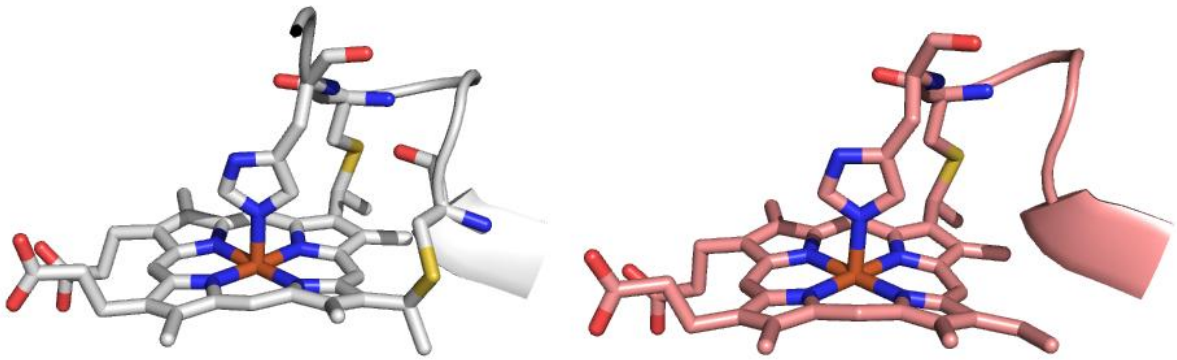


Figure 1.6 3-D structure of the heme binding motifs CXXCH of *Saccharomyces cerevisiae* iso-1-cytochrome *c* (Protein Data Bank code: 1YCC) in grey and AXXCH *Crithidia fasciculata* cytochrome *c* (Protein Data Bank code: 2YK3) in flesh tint with heme covalently bound.

1.7 HCCS, Cytochrome *c* and Health

The HCCS and the product cytochrome *c* have direct implications in human health. The functions of cytochrome *c* and its role in the initiation and regulatory process of apoptosis and other processes [74] makes it important for various aspects of human health.

1.7.1 Cytochrome *c* and Apoptosis

Programmed cell death or apoptosis is a vital process for the regulation of development, the removal of damaged cells, the immune response and the control of cell homeostasis in multicellular organisms [73,74]. The apoptosis mechanism is triggered by stimuli such as DNA damage, stress or protein misfolding [74]. It is regulated differently in different organisms. In mammals, the mitochondrion is the organelle which initiates the apoptotic process [73]. B-cell lymphoma protein (Bcl)-2 protein family members regulate the release of cytochrome *c* from the mitochondria [73], and the release of the cytochrome *c* starts the signalling pathway to activate many members of the caspase family of proteases, such as caspase-9 and executioner caspases, to form a protein oligomer which is termed apoptosome for signalling and to cleave a number of targets and initiate the apoptosis process [74,101]. The release of cytochrome *c* from mitochondria triggers calcium release from the endoplasmic reticulum which causes an amplified release of cytochrome *c* from all the mitochondria in the cell to facilitate apoptosis [101,102].

It has also been suggested that cytochrome *c* release might have other non-apoptosis signalling functions, such as the spermatid individualization process in *D. melanogaster*, as well as platelet formation and monocyte differentiation [74]. The understanding of the role of cytochrome *c* in these physiological processes is still limited at this stage.

1.7.2 Microphthalmia with linear skin defects syndrome (MLS)

The mutation and malfunction of the *HCCS* gene is related to the disease of microphthalmia with linear skin defects syndrome. Only defects in the *HCCS* gene would result in the MLS syndrome [35]. In humans, the gene for HCCS is on the X-chromosome. The variant R197X impairs the import of HCCS into the intermembrane space [103]. Disease variants R217C, R197X [103] and E159K [104] were shown to be incapable of maturing cytochrome *c*. The deletion of exons of the *HCCS* gene can also lead to MLS symptoms [103].

The impairment of import of HCCS in humans has a profound effect on the production of cytochrome *c*, which in turn probably affects the apoptosis-necrosis balance of the patients and causes excess necrosis of various tissues [103]. It has also been proposed that HCCS protein itself may play a role in the initiation of apoptosis [105].

In mammals, because the HCCS gene is on the X-chromosome of which males only retain one copy, the defect results in death of male patients in the embryonic stage [106]. As

females have two copies of X chromosome, the affected individual would be able to survive with one normal copy of the HCCS allele [107]. Affecting the development of eye, brain and skin, MLS is known to cause various symptoms including microphthalmia, sclerocornea, agenesis of the corpus callosum, costovertebral defects, seizures, mental retardation, cardiac abnormalities, diaphragmatic hernia and linear skin lesions of the head and neck [107]. The extent and severity of these symptoms in different patients are varied [35]. Studies of HCCS therefore have the potential to give some insight into these diseases, as well as providing an understanding of the role of the protein in the basic and essential function of producing mitochondrial cytochrome *c*.

1.8 Aims and structure of this thesis

Since its discovery [63] and the subsequent identification and classification of different cytochrome *c* biogenesis systems [26], HCCS (and HCC_1S) has been an enigmatic enzyme with little characterisation of its mechanism of heme attachment or sufficient insight into its evolutionary origin. A few studies concerning its substrate recognition had been published, including by Wang et al. [108] and others [40,109]. The understanding of its mechanism and the elucidation of the model of HCCS activity has developed relatively slowly when compared to other systems [33,37,91]. The distribution of different cytochrome *c* biogenesis systems in mitochondria, as well as the phylogenetic distribution of HCCS and its isoforms, was investigated [38]; a phylogenetic tree for mitochondrial *c*-type cytochrome maturation was constructed.

After the first successful co-expression of mitochondrial cytochrome *c* and HCCS in *E. coli* [82], Sanders *et al.* [110] and Rumbley *et al.* [111] optimised the condition for the overexpression as well as heme maturation of mitochondrial cytochrome *c* with *S. cerevisiae* HCCS in *E. coli*. Feasible co-expression of mitochondrial cytochrome *c* and HCCS in *E. coli* enables the analysis of heme maturation activity of HCCS by physical methods. The work in this thesis aimed to utilise a combination of chemical and spectroscopic methods to identify and analyse the key characteristics present in mitochondrial cytochrome *c* from different organisms for the successful substrate recognition of HCCS, with the aid of bioinformatics tools. Another aim of this thesis is the development and optimisation of procedures for the expression of the HCCS protein itself, as the high-level expression of HCCS in an *E. coli* host could serve as the basis for experiments on purification and mutagenesis studies on HCCS, as well as the investigation of protein-protein interactions between HCCS and mitochondrial cytochrome *c*.

In this thesis, Chapters 3 and 4 involve the identification and further characterisation of the critical features for the HCCS substrate recognition, such as the N-terminal region sequence of mitochondrial cytochromes *c* and the highly conserved N-terminal phenylalanine residue within it, as well as the difference in residue spacing in the N-terminal region of bacterial and mitochondrial cytochromes *c*. Chapter 5 provides further insights in the role of cysteines and histidines in the heme-binding motif CXXCH in mitochondrial cytochrome *c*. Chapter 6 consists of approaches to sequence analysis of various selected organisms for a perspective to compare with the analysis in previous chapters, and Chapter 7 describes the progress in the optimisation of *Homo Sapiens* HCCS expression in *E. coli* for purification to be carried out in future work.

Chapter 2

Material and methods

2.1 Bacterial strains and plasmids

2.1.1 Index of bacterial strains

Table 2.1 lists all the bacterial strains used in this project.

Table 2.1 Bacterial Strains used in the study

Organism	Characterisitics	Reference
Escherichia coli BL21(DE3)	F *ompT hsdSB* (r_B^- m_B^-) *dcm gal* λ(DE3)	Stratagene
Escherichia coli DH5α	φ80d*lacZ*ΔM15 *recA*1 *endA*1 *gyrA*96, thi-1 *hsdR*17(r_K^- m_K^-) *supE*44 *relA*1, *deoR*, Δ(*lacZYA-argF*)U169 *phoA*	[112]
Escherichia coli EC06	F′ Δ(ara-leu)7697 [*araD*139]$_{B/r}$ Δ(*codB-lacI*)3 *galK*16 *galE*15 λ$^-$ e14$^-$ *mcrA*0 *relA*1 *rpsL*150(Str^R) *spoT*1 *mcrB*1 *hsdR2*(r$^-$m$^+$) Δ*ccmABCDEFGH*	[113]
Escherichia coli XL2-Blue Ultracompetent Cells	*recA*1 *endA*1 *gyrA*96 *thi*-1 *hsdR*17 *supE*44 *rel*A1 *lac* [F′ *proAB lacI*^{q}ZΔ*M*15 *Tn*10 (Tet^r) Amy Cm^r]	Agilent Technologies

2.1.2 Index of plasmids

Table 2.2 lists all the plasmids used in this project.

Table 2.2 Plasmids used in the study

Name	Description	Source/Reference
pET3a	Cloning vector, Amp^R	New England Lab
pLysS	Cloning vector, Cm^R	Novagen
pACcyc3	*S. cerevisiae cyc3*, Cm^R	[110]
pEC86	*E. coli ccmABCDEFGH*, Cm^R	[114]
pOScyc1	*S. cerevisiae cyc1*, Amp^R	[110]
pKPD1	*P. denitrificans cycA*, Amp^R	[115]
pWT	*Equus caballus CYCS* + *S. cerevisiae cyc3*, Amp^R	[111]
pCYTc-b	*A. thaliana CYTc-b*	Dr Bonnard

pAt	*A. thaliana CYTc-b*, Amp^R	This work
pTrp	*T. brucei Tb927.8.5120* (cytochrome *c*) with a CXXCH motif modification, Amp^R	[116]
pKSH028	*S. cerevisiae cyc1* N-terminus up to CXXCHX + *P. denitrificans* c_{550} C-terminus, Amp^R	[117]
pKSH029	Recombinant *S. cerevisiae cyc1* N-terminus up to CXXCH + *T. brucei c* C-terminus, Amp^R	Dr K. Sam (unpublished work)
pYZ01	*S. cerevisiae cyc1* N-terminus + *P. denitrificans* c_{550} C-terminus from CXXCH, Amp^R	This work
pYZ02	pOScyc1 with G6A mutation, Amp^R	This work
pYZ03	pOScyc1 with K10A mutation, Amp^R	This work
pYZ04	pOScyc1 with G11A mutation, Amp^R	This work
pYZ05	pOScyc1 with F15A mutation, Amp^R	This work
pYZ06	pOScyc1 with R18A mutation, Amp^R	This work
pYZ07	pWT with C15A mutation on cytochrome, Amp^R	This work
pYZ08	pWT with C18A mutation on cytochrome, Amp^R	This work
pYZ09	pWT with H19K mutation on cytochrome, Amp^R	This work
pYZ10	pWT with H19M mutation on cytochrome, Amp^R	This work
pYZ11	pWT with H19R mutation, Amp^R	This work
pYZ12	pOScyc1 with F15Y mutation, Amp^R	This work
pYZ13	pOScyc1 with F15W mutation, Amp^R	This work
pYZ14	pOScyc1 with F15I mutation, Amp^R	This work
pYZ15	pOScyc1 with F15E mutation, Amp^R	This work
pYZ16	pOScyc1 with T17 deletion, Amp^R	This work
pYZ17	pOScyc1 with K16 deletion, Amp^R	This work
pYZ18	pOScyc1 with G29 truncation, Amp^R	This work
pYZ19	pOScyc1 with H45 truncation, Amp^R	This work
pYZ20	pOScyc1 with K60 truncation, Amp^R	This work
pOScyc1peri	pOScyc1 with an N-terminal *P. aeruginosa* $c_{551/552}$ periplasmic targeting sequence, Amp^R	This work
pYZ21	pOScyc1peri with F15A mutation, Amp^R	This work
pYZ22	pOScyc1peri with F15Y mutation, Amp^R	This work
pYZ23	pOScyc1peri with F15W mutation, Amp^R	This work
pYZ24	pOScyc1peri with F15I mutation, Amp^R	This work
pYZ25	pOScyc1peri with F15E mutation, Amp^R	This work
pYZ26	pOScyc1peri with T17 deletion, Amp^R	This work
pYZ27	pOScyc1peri with K16 deletion, Amp^R	This work
pYZ28	pWT with K6A mutation on cytochrome, Amp^R	This work
pYZ29	pWT with K9A mutation on cytochrome, Amp^R	This work
pYZ30	pWT with G24 truncation on cytochrome, Amp^R	This work
pYZ31	pWT with T29 truncation on cytochrome, Amp^R	This work
pYZ32	pWT with K40 truncation on cytochrome, Amp^R	This work
pYZ33	pWT with N55 truncation on cytochrome, Amp^R	This work
pYZ34	pWT with N71 truncation on cytochrome, Amp^R	This work
pYZ35	pWT with CXXCH-CNAAGSCH motif substitution on cytochrome, Amp^R	This work
pYZ36	pWT with CXXCH-CNAAGSQCH motif substitution on cytochrome, Amp^R	This work

pYZ37	pKPD1 with N-terminal signal sequence truncation, Amp^R	This work
pYZ38	pYZ37 with a Strep-tag (MAWSHPQFEK) insertion, Amp^R	This work
pYZ39	pYZ38 with T15 insertion, Amp^R	This work
pYZ40	pYZ38 with A15 insertion, Amp^R	This work
pYZ41	pYZ38 with K15 insertion, Amp^R	This work
pYZ42	pAt with a Strep-tag (MAWSHPQFEK) insertion , Amp^R	This work
pYZ43	pYZ42 with P8K and P9A mutations, Amp^R	This work
pYZ44	pAt with an N-terminal region (ASFDEAPP) deletion, Amp^R	This work
pYZ45	pYZ45 with a Strep tag (MAWSHPQFEK) insertion, Amp^R	This work
pAtperi	pAt with an N-terminal *P. aeruginosa* $c_{551/552}$ periplasmic targeting sequence, Amp^R	This work
pYZ46	pAtperi with an N-terminal region (ASFDEAPP) deletion, Amp^R	This work
pYZ47	pTrp with with a Strep-tag (MAWSHPQFEK) insertion, Amp^R	This work
pYZ48	pYZ47 with an N-terminal region (PPKERAALPP) deletion, Amp^R	This work
pYZ49	pKSH029 with L21A mutation, Amp^R	This work
pSccyc1	*S. cerevisiae CYT1*, Amp^R	Genscript®
pHCC1S	*S. cerevisiae CYT2*, Amp^R	Genscript®
pHsHCCS	*H. sapiens HCCS*, Amp^R	Genscript®
pHsHCCS /StrepII	pHsHCCS with a Strep-tag insertion	This work

2.1.3 Growth of bacteria on solid and liquid media

The liquid media for bacterial growth is Luria-Bertani (LB) media whose composition is shown in Table 2.3. For the solid media, 1.5% (w/v) agar (Melford) was added and mixed to liquid media before autoclaving.

Table 2.3 Composition of Luria-Bertani (LB) media

Component	**Quantity (g L^{-1})**
Tryptone	10
Yeast extract	10
NaCl	5

2.1.4 Antibiotics

The antibiotics were added at appropriate quantities to the growth media for the prevention of contamination and the antibiotic screening of the colonies. The antibiotics used in this project are listed in Table 2.4, with their concentrations in aqueous (ampicillin) and ethanol (chloramphenicol) solutions.

Table 2.4 Antibiotic concentrations used in the study

Antibiotic	Concentration (μg ml^{-1})
Ampicillin	100
Chloramphenicol	33

2.1.5 Index of oligonucleotide primers

Table 2.5 lists all the oligonucleotide primers used in this project.

Table 2.5 Oligonucleotide primers used in the study

Name	Function	Sequence
XX_F	Yeast-bacterial Chimera CXXCH mutation	5′-GACTAGATGTAAAGCCTGCCACACCATCC-3′
XX_R	Yeast-bacterial Chimera CXXCH mutation	5′- GGATGGTGTGGCAGGCTTTACATCTAGTC-3′
T26M_F	Yeast-bacterial Chimera T26M mutation	5′-GTAAAGCCTGCCACATGATCCAGGCGCC-3′
T26M_R	Yeast-bacterial Chimera T26M mutation	5′-GGCGCCTGGATCATGTGGCAGGCTTTAC-3′

L21A_F	Yeast-trypanosome Chimera L21 mutation	5′-GCCTTTTGTACCGGTGTGGCATTGCGCACATCTAGTCTTGAAAAGTGTAGCACC-3′
L21A_R	Yeast-trypanosome Chimera L21 mutation	5′-GGTGCTACACTTTTCAAGACTAGATGTGCGCAATGCCACACCGGTACAAAAGGC-3′
G24X_F	*E. caballus* cytochrome *c* G24 truncation	5′-CACACTGTCGAAAAATAAGGTAAGAACAAGACCGGTCC-3
G24X_R	*E. caballus* cytochrome *c* G24 truncation	5′-GGACCGGTCTTCTTCTTACCTTATTTTTCCACAGTGTG-3
T29X_F	*E. caballus* cytochrome *c* T29 truncation	5′-GGTGGTAAGAACAAGTAAGGTCCAAACTTGAACGG-3
T29X_R	*E. caballus* cytochrome *c* T29 truncation	5′-CCGTTCAAGTTTGGACCTTACTTGTTCTTACCACC-3
K40X_F	*E. caballus* cytochrome *c* K40 truncation	5′-GAACGGTTTGTTCGGTCGTTAAACCGGTCAAGCTCCAGG-3
K40X_R	*E. caballus* cytochrome *c* K40 truncation	5′-CCTGGAGCTTGACCGGTTTAACGACCGAACAAACCGTTC-3
N55X_F	*E. caballus* cytochrome *c* N55 truncation	5′-CTTACACTGACGCTAACAAGTAAAAGGGTATCACCTGGAAG-3
N55X_R	*E. caballus* cytochrome *c* N55 truncation	5′-CTTCCAGGTGATACCCTTTTACTTGTTAGCGTCAGTGTAAG-3
N71X_F	*E. caballus* cytochrome *c* N71 truncation	5′-CTTTGATGGAATACTTGGAATAACCAAAGAAGTACATTCCTGG-3
N71X_R	*E. caballus* cytochrome *c* N71 truncation	5′-CCAGGAATGTACTTCTTTGGTTATTCCAAGTATTCCATCAAAG-3
K6A_F	*E. caballus* cytochrome *c* K5 mutation	5′-GGGTGACGTTGAAGCGGGTAAGAAGATTTTC-3
K6A_R	*E. caballus* cytochrome *c* K5 mutation	5′-GAAAATCTTCTTACCCGCTTCAACGTCACCC-3
K9A_F	*E. caballus* cytochrome *c* K8 mutation	5′-CGTTGAAAGGGTAAGGCGATTTTCGTTCAAAAG-3
K9A_R	*E. caballus* cytochrome *c* K8	5′-CTTTTGAACGAAAATCGCCTTACCCTTTCAACG-3

	mutation	
G29X_F	*S. cerevisiae* cytochrome *c* G29 truncation	5′-CAATGCCACACCGTGGAAAAGTAAGGCCCACATAAGGTTGGTCC-3
G29X_R	*S. cerevisiae* cytochrome *c* G29 truncation	5′-GGACCAACCTTATGTGGGCCTTACTTTTCCACGGTGTGGCATTG-3
H45X_F	*S. cerevisiae* cytochrome *c* H45 truncation	5′-GGTATCTTTGGCAGATAATCTGGTCACTAACCCAAAG-3
H45X_R	*S. cerevisiae* cytochrome *c* H45 truncation	5′-CTTTGGGTTAGTGACCAGATTATCTGCCAAAGATACC-3
K60X_F	*S. cerevisiae* cytochrome *c* K60 truncation	5′-CGTACACAGATGCCAATATCTAAAAAAACGTGTTGTGGGACG-3
K60X_R	*S. cerevisiae* cytochrome *c* K60 truncation	5′-CGTCCCACAACACGTTTTTTTAGATATTGGCATCTGTGTACG-3
Sccytcperi1_F	*S. cerevisiae* cytochrome *c* periplasmic target sequence insertion	5′-GAAGGAGATATACATATGAAACCGTACGCACTGCTTTCGCTGCTCGCGCTTCTTGGTCTCATCCTCAATTTGAG-3
Sccytcperi1_R	*S. cerevisiae* cytochrome *c* periplasmic target sequence insertion	5′-CTCAAATTGAGGATGAGACCAAGAAGCGCGAGCAGCGAAAGCAGTGCGTACGGTTTCATATGTATATCTCCTTC-3
Sccytcperi2_F	*S. cerevisiae* cytochrome *c* periplasmic target sequence insertion	5′-ACCGTACGCACTGCTTTCGCTGCTCGCCACCGGCACCCTGCTCGCCCAGGGCGCCTGGGCCGAAGACGCTTCTTGGTCTCACTTTCAATTTGAG-3
Sccytcperi2_R	*S. cerevisiae* cytochrome *c* periplasmic target sequence insertion	5′-CTCAAATTGAAAGTGAGACCAAGAAGCGTCTTCGGCCCAGGCGCCCTGGGCGAGCAGGGTGCCGGTGGCGAGCAGCGAAAGCAGTGCGTACGGT-3
G6A_F	*S. cerevisiae* cytochrome *c* G6 mutation	5′-GACTGAATTCAAGGCCGCTTCTGCTAAGAAAGGTGC-3
G6A_R	*S. cerevisiae* cytochrome *c* G6 mutation	5′-GCACCTTTCTTAGCAGAAGCGGCCTTGAATTCAGTC-3
K10A_F	*S. cerevisiae* cytochrome *c* K10 mutation	5′-CCGGTTCTGCTAAGGCTGGTGCTACACTTTTCAAG-3
K10A_R	*S. cerevisiae* cytochrome *c* K10 mutation	5′-CTTGAAAAGTGTAGCACCAGCCTTAGCAGAACCGG-3
G11A_F	*S. cerevisiae* cytochrome *c* G11	5′-GGTTCTGCTAAGAAAGCTGCTACACTTTTCAAG-3

	mutation	
G11A_R	*S. cerevisiae* cytochrome *c* G11 mutation	5′-CTTGAAAAGTGTAGCAGCTTTCTTAGCAGAACC-3
F15A_F	*S. cerevisiae* cytochrome *c* F15 mutation	5′-GAAAGGTGCTACACTTGCCAAGACTAGATGCATC-3
F15A_R	*S. cerevisiae* cytochrome *c* F15 mutation	5′-GATGCATCTAGTCTTGGCAAGTGTAGCACCTTTC-3
R18A_F	*S. cerevisiae* cytochrome *c* R18 mutation	5′-CACTTTTCAAGACTGCTTGTCTACAATGCCAC-3
R18A_R	*S. cerevisiae* cytochrome *c* R18 mutation	5′-GTGGCATTGTAGACAAGCAGTCTTGAAAAGTG-3
F15Y_F	*S. cerevisiae* cytochrome *c* F15 mutation	5′-GAAAGGTGCTACACTTTACAAGACTAGATGTCTAC-3
F15Y_R	*S. cerevisiae* cytochrome *c* F15 mutation	5′-GTAGACATCTAGTCTTGTAAAGTGTAGCACCTTTC-3
F15W_F	*S. cerevisiae* cytochrome *c* F15 mutation	5′-GAAAGGTGCTACACTTTGGAAGACTAGATGTCTAC-3
F15W_R	*S. cerevisiae* cytochrome *c* F15 mutation	5′-GTAGACATCTAGTCTTCCAAAGTGTAGCACCTTTC-3
F15I_F	*S. cerevisiae* cytochrome *c* F15 mutation	5′-GAAAGGTGCTACACTTATCAAGACTAGATGTCTAC-3
F15I_R	*S. cerevisiae* cytochrome *c* F15 mutation	5′-GTAGACATCTAGTCTTGATAAGTGTAGCACCTTTC-3
F15E_F	*S. cerevisiae* cytochrome *c* F15 mutation	5′-GAAAGGTGCTACACTTGAAAAGACTAGATGTCTAC-3
F15E_R	*S. cerevisiae* cytochrome *c* F15 mutation	5′-GTAGACATCTAGTCTTTTCAAGTGTAGCACCTTTC-3
C15A_F	*E. caballus* cytochrome *c* C15 mutation	5′-GATTTTCGTTCAAAAGGCCGCTCAATGTCACACTG-3
C15A_R	*E. caballus* cytochrome *c* C15 mutation	5′-CAGTGTGACATTGAGCGGCCTTTTGAACGAAAATC-3
C18A_F	*E. caballus* cytochrome *c* C18	5′-CAAAAGTGTGCTCAAGCCCACACTGTCGAAAAAG-3

	mutation	
C18A_R	*E. caballus* cytochrome *c* C18 Mutation	5′-CTTTTTCGACAGTGTGGGCTTGAGCACACTTTTG-3
H19K_F	*E. caballus* cytochrome *c* H19 mutation	5′-CAAAAGTGTGCTCAATGTAAGACTGTCGAAAAAGGTGG-3
H19K_R	*E. caballus* cytochrome *c* H19 mutation	5′-CCACCTTTTTCGACAGTCTTACATTGAGCACACTTTTG-3
H19M_F	*E. caballus* cytochrome *c* H19 mutation	5′-CAAAAGTGTGCTCAATGTATGACTGTCGAAAAAGGTGG-3
H19M_R	*E. caballus* cytochrome *c* H19 mutation	5′-CCACCTTTTTCGACAGTCATACATTGAGCACACTTTTG-3
H19R_F	*E. caballus* cytochrome *c* H19 mutation	5′-CAAAAGTGTGCTCAATGTCGCACTGTCGAAAAAGGTGG-3
H19R_R	*E. caballus* cytochrome *c* H19 mutation	5′-CCACCTTTTTCGACAGTGCGACATTGAGCACACTTTTG-3
CX5CH_F	*E. caballus* cytochrome *c* CXXCH insertion	5′-GATTTTCGTTCAAAAGTGTAACGCCGCCGGCAGCTGTCACACTGTCGAAAAAGG-3
CX5CH_R	*E. caballus* cytochrome *c* CXXCH Insertion	5′-CCTTTTTCGACAGTGTGACAGCTGCCGGCGGCGTTACACTTTTGAACGAAAATC-3
CX6CH_F	*E. caballus* cytochrome *c* CXXCH insertion	5′-GATTTTCGTTCAAAAGTGTAACGCCGCCGGCAGCCAGTGTCACACTGTCGAAAAAGG-3
CX6CH_R	*E. caballus* cytochrome *c* CXXCH insertion	5′-CCTTTTTCGACAGTGTGACACTGGCTGCCGGCGGCGTTACACTTTTGAACGAAAATC-3
C550trunc_F	Deletion of the signal sequence of *P. denitrificans* c_{550}	5′-CAGGATGGCGACGCCGCCAAAGGC-3
C550trunc_R	Deletion of the signal sequence of *P. denitrificans* c_{550}	5′-CATCGCGTTTCCTCTTGGGTATCG-3
C550insNT_F	*P. denitrificans* c_{550} T15 insertion	5′-GGCGAGAAAGAATTCAACACTAAGTGCAAGGCTTGCCAC-3
C550insNT_R	*P. denitrificans* c_{550} T15 insertion	5′-GTGGCAAGCCTTGCACTTAGTGTTGAATTCTTTCTCGCC-3

C550Strep_F	*P. denitrificans* c_{550} Strep-tag insertion	5′-CAAGAGGAAACGCGATGGCGTGGAGCCACCCGCAGTTCGAAAAAGCGCAGGATGGCGACGCC-3
C550Strep_R	*P. denitrificans* c_{550} Strep-tag insertion	5′-GGCGTCGCCATCCTGCGCTTTTTCGAACTGCGGGTGGCTCCACGCCATCGCGTTTCCTCTTG-3
C550T15A_F	*P. denitrificans* c_{550} T15 mutation	5′-GGCGAGAAAGAATTCAACGCTAAGTGCAAGGCTTGCCAC-3
C550T15A_F	*P. denitrificans* c_{550} T15 mutation	5′-GGCGAGAAAGAATTCAACGCTAAGTGCAAGGCTTGCCAC-3
C550T15X_F	*P. denitrificans* c_{550} T15 deletion	5′-GGCGAGAAAGAATTCAACAAGTGCAAGGCTTGCCAC-3
C550T15X_R	*P. denitrificans* c_{550} T15 deletion	5′-GTGGCAAGCCTTGCACTTGTTGAATTCTTTCTCGCC-3
C550insKNK_F	*P. denitrificans* c_{550} N15 insertion	5′-GGCGAGAAAGAATTCAAGAACAAGTGCAAGGCTTGCCAC-3
C550insKNK_R	*P. denitrificans* c_{550} N15 insertion	5′-GTGGCAAGCCTTGCACTTGTTCTTGAATTCTTTCTCGCC-3
c1solclone_F	Cloning of *S. cerevisiae* cytochrome c_1 gene	5′-AAAAAACATATGACCGCAGCTGAACACGG-3
c1solclone_R	Cloning of *S. cerevisiae* cytochrome c_1 gene	5′-TTTTGGATCCCTATTCGGCACACCAGTTTAAAAAGG-3
Hcc1sclonin g_F	Cloning of *S. cerevisiae* HCC_1S gene	5′-AAAAAACATATGGATGTCTTCAGACCAACAGG-3
Hcc1sclonin g_R	Cloning of *S. cerevisiae* HCC_1S gene	5′-TTTTGGATCCTCAAAAGGTTTGGCCAGGTGCAAG-3
Atcytc_F	Cloning of *A. thaliana* cytochrome *c* gene	5′-AAAAAACATATGGCGTCGTTTGATGAAGCACC-3
Atcytc_R	Cloning of *A. thaliana* cytochrome *c* gene	5′-TTTTTTGGATCCCTACTTAGGCGCAGTAGATTCCTTC-3
AtStrep_F	*A. thaliana* cytochrome *c* Strep-tag insertion	5′-CTTCCTTGTTGCTTCTATGGCGTGGAGCCACCCGCAGTTCGAAAAATCAATGGCGTCG-3
AtStrep_R	*A. thaliana* cytochrome *c*	5′-CTTCCTTGTTGCTTCTATGGCGTGGAGCCACCCGCAGT

	Strep-tag insertion	TCGAAAAATCAATGGCGTCG-3
P8KP9A_F	*A. thaliana* cytochrome *c* P8 and P9 mutations	5′-TTCACCGGCTTTGGGGTTTCCAGCTTTTGCTTCGTCGAACGACGCCAT-3
P8KP9A_R	*A. thaliana* cytochrome *c* P8 and P9 mutations	5′-ATGGCGTCGTTCGACGAAGCAAAAGCTGGAAACCCCAAAGCCGGTGAA-3
AtdelInterm_F	*A. thaliana* cytochrome *c* N-terminal deletion	5′-TTCCTTGTTGCTTCTTCAATGGGAAACCCCAAAGCCGGTGAA-3
AtdelInterm_R	*A. thaliana* cytochrome *c* N-terminal deletion	5′-TTCACCGGCTTTGGGGTTTCCCATTGAAGAAGCAACAAGGAA-3
AtdelStrep_F	N-terminal deleted *A. thaliana c* Strep-insertion	5′-CTTTTCACCGGCTTTGGGGTTTCCCATTGATTTTTCGAACTGCGGGTGGCTCCACGCCATTGAAGAAGCAACAAGGAAGAAATA-3
AtdelStrep_R	N-terminal deleted *A. thaliana c* Strep-insertion	5′-TATTTCTTCCTTGTTGCTTCTTCAATGGCGTGGAGCCACCCGCAGTTCGAAAAATCAATGGGAAACCCCAAAGCCGGTGAAAAG-3
Atperi1_F	*A. thaliana* cytochrome *c* periplasmic target sequence insertion	5′-TTCCTTGTTGCTTCTTCAATGAAACCGTACGCACTGCTTTCGCTGCTCGCCACCGGGCGTCGTTCGACGAAGCACCT-3
Atperi1_R	*A. thaliana* cytochrome *c* periplasmic target sequence insertion	5′-AGGTGCTTCGTCGAACGACGCCCGGTGGCGAGCAGCGAAAGCAGTGCGTACGGTTTCATTGAAGAAGCAACAAGGAA-3
Atperi2_F	*A. thaliana* cytochrome *c* periplasmic target sequence insertion	5′-GCTTTCGCTGCTCGCCACCGGCACCCTGCTCGCCCAGGGCGCCTGGGCCGAAGACGCGTCGTTCGACGAAGCACCT-3
Atperi2_F	*A. thaliana* cytochrome *c* periplasmic target sequence insertion	5′-AGGTGCTTCGTCGAACGACGCGTCTTCGGCCCAGGCGCCCTGGGCGAGCAGGGTGCCGGTGGCGAGCAGCGAAAGC-3
AtperidelInterm_F	Periplasmically targeted *A. thaliana* cytochrome *c* N-terminal deletion	5′-GCCCAGGGCGCCTGGGCCGAAGACGGAAACCCCAAAGCCGGTGAA-3
AtperidelInterm_R	Periplasmically targeted *A. thaliana* cytochrome *c* N-terminal deletion	5′-TTCACCGGCTTTGGGGTTTCCGTCTTCGGCCCAGGCGCCCTGGGC-3
TrpStrep_F	*T. brucei* cytochrome *c*	5′-AAGTGCTGCACGCTCCTTTGGTGGCATTGATTTTTCGA

	(CXXCH) Strep-tag insertion	ACTGCGGGTGGCTCCACGCCATGCGAATTCTGTTTCCTGTGTG-3
TrpStrep_R	*T. brucei* cytochrome *c* (CXXCH) Strep-tag insertion	5′-AAGTGCTGCACGCTCCTTTGGTGGCATTGATTTTTCGAACTGCGGGTGGCTCCACGCCATGCGAATTCTGTTTCCTGTGTG-3
TrpdelInterm_F	N-terminal deleted *T. brucei c* (CXXCH) Strep-tag insertion	5′-CTCCCCACGCGCTGCGTCACCCATGCGAATTCTGTTTCCTGT-3
TrpdelInterm_R	N-terminal deleted *T. brucei c* (CXXCH) Strep-tag insertion	5′-ACAGGAAACAGAATTCGCATGGGTGACGCAGCGCGTGGGGAG-3
TrpdelStrep_F	N-terminal deleted *T. brucei c* Strep-tag insertion	5′-AAGTGCTGCACGCTCCTTTGGTGGCATTGATTTTTCGAACTGCGGGTGGCTCCACGCCATGCGAATTCTGTTTCCTGTGTG-3
TrpdelStrep_R	N-terminal deleted *T. brucei c* Strep-tag insertion	5′-CACACAGGAAACAGAATTCGCATGGCGTGGAGCCACCCGCAGTTCGAAAAATCAATGCCACCAAAGGAGCGTGCAGCACTT-3
pKK223-3_F	Sequencing primer for pKPD1 and pTrp	5′-CGGTTCTGGCAAATATTCTG-3

2.2 Molecular biology techniques

2.2.1 Purification of plasmid DNA

Plasmid DNA was purified from *E. coli* bacterial cells using the QIAprep® Spin Miniprep Kit (Qiagen) according to the instructions of the manufacturer.

2.2.2 Digestion of DNA using restriction enzymes

For cloning purposes, purified DNA was digested with appropriate restriction enzymes, e.g. *Nde*I, *BamH*I, *Dpn*I (New England Biolabs), according to the manufacturer's instructions.

2.2.3 Agarose gel electrophoresis and DNA extraction

Agarose gel electrophoresis was used to separate DNA molecules based on size. The DNA samples were mixed with 5 × DNA loading buffer (Bioline) or 6 × DNA loading buffer (Novagen) with an appropriate ratio by volume. Agarose gels (40mM Tris-acetate pH 7.5, 1mM EDTA, 1% w/v agarose, 0.2 μg ml^{-1} ethidium bromide) were dissolved, melted, poured and set in plastic tanks. The samples were loaded into the wells and the electrophoresis was run with TAE buffer (40mM Tris-acetate pH 8.0, 1mM EDTA) under a current of 100mA for an appropriate length of time. The Hyperladder I (Bioline) was used as the marker. The DNA on the agarose gels was visualised under Ultraviolet (UV) light from the transilluminator of a Gel Doc 2000 gel documentation system (Bio-Rad) as the DNA-binding ethidium bromide is fluorescent under UV light.

Plasmid or linear DNA was extracted from cut agarose gel bands using the Gel extraction Kit (Qiagen) according to the instructions of the manufacturer.

2.2.4 Ligation of DNA

In gene cloning, the ligation of a DNA fragment with a vector was achieved by linking the strands of DNA together. The plasmid and the insert were digested with appropriate restriction enzymes to produce complementary overhang 'sticky ends'. These linear DNA molecules were then separately purified by agarose gel electrophoresis followed by gel extraction. Dephosphorylation of the digested DNA was required to prevent self-ligation. An appropriate volume of 10 × phosphatase reaction buffer (New England Biolabs) and 1 µl phosphatase was mixed with the vector or insert, and then the mixtures were incubated at 37 ^{o}C for 20 min. The mixtures were then put on an 80 ^{o}C heat block to inactivate the enzyme.

The volume of the ligation reaction mixture was always 10 µl, with 1µl 10 × T4 DNA ligase Reaction Buffer (Sigma) and 1 µl T4 DNA ligase (Sigma). The molar ratios of vector to insert DNA were varied, such as 1:2, 1:4 and 1:8. The total DNA concentration was always in the range of 5-25 ng μl^{-1}. The ligation mixtures were incubated at room temperature overnight.

2.2.5 Commercial gene synthesis

Under some circumstances, genes were externally synthesised by and incorporated into the semi-digested vectors (provided by the lab) by Genescript® service.

2.2.6 Preparation of competent *E. coli* cells

E. coli strains BL21 (DE3), EC06 and DH5α were used as competent cells for transformations. The preparation of stocks of competent cells involves an adapted version of the technique described in Hannan's improved method [112]. A frozen aliquot from the stored stock of cell or a single colony of the chosen strain was picked for the inoculation of 5 ml super optimal broth (SOB) media (Table 2.6). The resulting culture underwent overnight growth at 37 ^{o}C with vigorous shaking. 2 ml of this culture was then transferred to 200 ml SOB media for cell growth under the same conditions until the OD_{550} reached around 0.45. This culture was then incubated on ice for 30 min, and then centrifuged down at 5000 × *g* for 15 min at 4 ^{o}C. The supernatant was then discarded, whereas the pellet was gently resuspended in 66 ml ice-cold buffer RF1 (Table 2.7). The cell suspension was then incubated on ice for a further 60 min followed by centrifugation as above. The supernatant was again discarded, and then the pellet was gently resuspended in 16 ml RF2 buffer (Table 2.8). This final suspension was then split into 220 µl aliquots in Eppendorf tubes. These aliquots were cryo-frozen in liquid nitrogen or dry ice immediately, and then stored in a -80^{o}C freezer.

Table 2.6 Composition of SOB media

Component	Mass
Bacto Tryptone (Peptone)	5 g
Yeast Extract	1.25 g
NaCl	0.146 g
KCl	0.046 g
$MgCl_2.4H_2O$	0.508 g
$MgSO_4.4H_2O$	0.61 g
Add distilled water to 250 ml, and autoclave	

Table 2.7 Composition of RF1 buffer

Component	Mass
RbCl	2.4 g
$MnCl_2.4H_2O$	1.98 g
CH_3COOK	0.589 g
$CaCl_2.2H_2O$	0.3 g
Glycerol	30 g
Add distilled water to 200 ml, adjust pH to 5.8 with 0.2 M acetic acid, and filter sterilise	

Table 2.8 Composition of RF2 buffer

Component	Mass
MOPS	0.209 g
RbCl	0.12 g
$CaCl_2.2H_2O$	1.1 g
Glycerol	15 g
Add distilled water to 100 ml, adjust pH to 6.8 with 1 M NaOH, and filter sterilise	

2.2.7 Plasmid transformation of competent *E. coli* cells

E. coli strain BL21 (DE3) was used for routine gene expression and protein production, and *E. coli* strain DH5α was used for routine molecular biology. Trials for transformations were carried out with XL2-Blue ultracompetent cells (Agilent Technologies) when the ligated product could not be successfully transformed with DH5α.

Frozen 220 µl aliquots of competent *E. coli* cells in an Eppendorf tube were placed on ice. After the suspension had completely thawed, an appropriate amount of the relevant plasmid DNA (typically 50-250 ng) was added to the aliquot. The suspension was then incubated on ice for 30 min before being incubated in a 42 ^{o}C water bath for 30 s, then was incubated on ice again for 2 min. For transformations with chloramphenicol, the recovery step is required: 1ml LB media was added to the suspension, and the mixture was incubated at 37 ^{o}C for 1 hour before the cells pelleted and 1 ml supernatant removed. Either way, 220 µl of suspension was plated onto LB agar supplemented with appropriate antibiotics. LB agar plates were then incubated at 37 ^{o}C overnight.

2.2.8 DNA sequencing

All the DNA sequencing was done externally by Geneservice of Source Biosciences® using Genome Analyzer IIx and HiSeq 2000 high throughput sequencing systems (Illumina). Commercial T7 and M13 stock primers, as well as lab stock primer pKPD223-3, were used in the sequencing.

2.3 Polymerase chain reaction (PCR) techniques

All PCR reactions were performed in a final volume of 50 µl in 0.5 ml Eppendorf tubes. The temperature change was produced by a Techne® TC-512 Thermal cycler. For all the PCR reactions for Quikchange® site-directed mutagenesis and cloning, the KOD Hot Start Polymerase Kit (Novagen) was used to prepare the mixture. This system utilises a recombinant *Thermococcus kodakaraensis* KOD1 DNA polymerase that forms a premixed complex with two monoclonal antibodies that inhibit the DNA polymerase and 3'→5' exonuclease activities at ambient temperatures [118]. It has the features of high fidelity, fast extension speed, and it effectively inhibits primer degradation and reduces non-specific amplification.

The standard reaction setup is shown in Table 2.9. In following sections, specific conditions for x, y and z are applied to this setup. Usually one negative control (without the addition of KOD polymerase) was also made in the set up.

Table 2.9 Standard reaction setup conditions for KOD Hot Start Polymerase

Component	Volume	Concentration
Template DNA solution	x µl	see **Table 2.11**

Forward (5') primer stock solution	y µl	see **Table 2.10** and **Table 2.11**
Reverse (3') primer stock solution	y µl	
25mM $MgSO_4$	4 µl	1.5 mM
dNTP stock solution	1 µl	0.2 mM
DMSO (PCR grade)	z µl	see **Table 2.10**
10 × KOD Hot Start Polymerase Buffer	5 µl	1X
KOD Hot Start Polymerase	1 µl	0.02 U/µl
Water (Molecular biology grade)	to 50 µl	

The cycling condition for the PCR reactions is: (1) Initialisation step: 95 ^{o}C for 2 min; (2) Denaturation step: 95 ^{o}C for 20 s; (3) Annealing step: Lowest primer T_m for 10 s; (4) Extension step: 70 ^{o}C for 10 s kb^{-1}; and (5) Final extension step: 70 ^{o}C for 7 min. The temperature at the annealing step could be increased to minimise non-specific binding when appropriate.

2.3.1 PCR amplification

The amplification of the insert was achieved by PCR reactions. 1 µl of template DNA solution was used, and the primer and DMSO concentrations are varied in different trials as shown in Table 2.10:

Table 2.10 Conditions for Standard reaction setup for PCR amplification

Component	Volume (µl)	Concentration
Forward (5') primer stock solution (10 µM)	2.0, 4.0	0.4 µM, 0.8 µM
Reverse (3') primer stock solution (10 µM)	2.0, 4.0	0.4 µM, 0.8 µM
DMSO (PCR grade, 50%)	0.0, 5.0	0, 5%

2.3.2 Site-directed mutagenesis technique of plasmid DNA

PCR reactions for mutagenesis were performed on the plasmid DNA template following a method based on the guidance of the QuikChange II Site-Directed Mutagenesis Kit (Stratagene). The method involved three steps: (1) PCR synthesis of mutant strands with mutagenic primers; (2) digestion of the template DNA plasmid by the restriction enzyme *Dpn*I (New England Biolabs, Section 2.2.2); (3) transformation of competent *E. coli* cells with the mutant DNA plasmid. Plasmid DNA was then purified from transformed cells on agar plates using the QIAprep® Spin Miniprep Kit (Section 2.2.1) and sequenced (Section 2.2.8) to confirm the correct modification to the sequence in the plasmid. The following varied concentrations were used combinatorially in reactions.

Table 2.11 Conditions for Standard reaction setup for PCR mutagenesis

Component	Volume used (µl)	Concentration
Template DNA solution	1.0, 0.1	0.2 ng/µl, 0.02 ng/µl
Forward (5') primer stock solution (10 µM)	1.0, 0.1	0.2 µM, 0.02 µM
Reverse (3') primer stock solution (10 µM)	1.0, 0.1	0.2 µM, 0.02 µM

2.4 Protein production, extraction and purification

2.4.1 Protein production

5 ml LB media with 1:1000 appropriate antibiotic supplements was inoculated with a single colony of *E. coli* strain BL21 (DE3) or DH5α and incubated aerobically at 37 ^{o}C with vigorous shaking for 16 hours. This 5 ml culture was then used to inoculate (1) 20 ml LB media in a sterile falcon tube; (2) 125 ml LB media in an autoclaved 250 conical flask; or (3) 500 ml LB media in an autoclaved 2 l conical flask. (1) (2) were under a partially anaerobic growth condition, and (3) was under an aerobic condition. When appropriate, 1 mM isopropyl-1-thio-b-D-galactopyranoside (IPTG) was used to induce expression of the proteins when OD_{600} of the culture suspension reached around 0.6. These cultures were incubated at 30 ^{o}C with shaking at 200 rpm for 20 h before harvesting. At least five replicate cultures were analysed for each quantitative experiment.

For the 20ml culture, 1 ml of suspension was transferred to an Eppendorf tube and centrifuged at 5000 × *g*. The supernatant was subsequently discarded, and the cell pellet was mixed with an appropriate volume of Bugbuster Protein Extraction Reagent (Novagen, Section 2.4.2) to lyse the cell. The whole-cell extract was analysed by UV-vis spectroscopy (Section 2.5.1) or biochemical methods (Section 2.6). Alternatively, the periplasm was separated by following the procedure described in Section 2.4.2 and the spheroplasts were lysed by Bugbuster Reagent (Novagen).

For the 125 ml culture or 500 ml culture, the cells were pelleted by centrifugation at 5000 × *g*. The periplasm was then separated from the spheroplasts (Section 2.4.2), and if necessary, the spheroplasts were lysed by sonication (Section 2.4.2). The periplasm extract and cytoplasm extract could then be subjected to spectroscopic analysis (Section 2.5) or

other analyses (Section 2.6). The protein in the extract was purified by FPLC (Section 2.4.3) in some cases.

2.4.1.1 Mitochondrial holocytochrome *c* expression systems in *E. coli*

Pollock *et al.* [82] found that the gene encoding mitochondrial cytochrome *c* could be co-expressed with the gene for the HCCS enzyme in the cytoplasm of *E. coli*. Sanders and Lill [110] and Rumbley et al. [111] developed procedures to overexpress mitochondrial cytochrome *c* with HCCS in *E. coli*, with the two genes on either separate plasmids or on a single plasmid. In both systems, the protein production was optimised to high levels. The two-plasmid system involved cytochrome *c* and HCCS both from *S. cerevisiae* (baker's yeast), whereas the single-plasmid system involved *E. caballus* (horse) cytochrome *c* and *S. cerevisiae* HCCS. Both of the cytochromes *c* had been well characterised. The *S. cerevisiae* cytochrome *c* gene with an added periplasmic targeting sequence (Chapter 3, Section 3.2.3.5) could also be co-expressed with a pEC86 plasmid containing the Ccm system. The plasmids containing cytochrome *c* genes were ampicillin resistant, and the plasmids containing HCCS or Ccm system genes were chloramphenicol resistant.

In this work, different approaches to the analysis of mitochondrial cytochrome *c* were used and both of the single-plasmid and the two-plasmid systems were used in different experiments. For some work that will be presented in Chapter 5, the single plasmid system was selected. This was because the horse cytochrome has a slightly shorter and therefore more tractable N-terminal region than the yeast cytochrome *c*. However, the horse

cytochrome *c* in the single-plasmid system has drawbacks, such as the lack of tags on the cytochrome and HCCS, and the wild-type gene expression were found to have a fluctuating level. Moreover, even though yeast HCCS matures both yeast and horse cytochromes *c*, it is probable that the two-plasmid system has the advantage that both genes are from the same organism (*S. cerevisiae*) which might simplify interpretation of interactions between the two proteins. Hence, the two-plasmid system was chosen for most of the characterisation works, mainly discussed in Chapter 4. Both plasmid systems were able to provide comparable levels of overexpression of holocytochrome *c*.

2.4.2 Cell lysis techniques

The periplasm of harvested cells was separated to produce spheroplasts with the addition of lysozyme, essentially following the procedure as described by Allen *et al.* [119]. Small volumes of *E. coli* cell lysis was achieved by adding appropriate volumes of Bugbuster® Protein Extraction Reagent (Novagen) to the pellet of harvested cells in Eppendorf tubes according to manufacturer's instructions. For larger pellets of cells, the cytoplasmic extraction was prepared by sonication of the spheroplasts with an Ultra Cell® sonicator (Sonics & Materials) followed centrifugation at 16000 × *g* for 35 min at 4 °C.

2.4.3 Fast protein liquid chromatography (FPLC)

An XK26 column (Amersham Biosciences/GE Healthcare) was packed with DEAE Sepharose Fast Flow resin (Amersham Biosciences/GE Healthcare) and set up according

to the manufacturer's instructions. The column was equilibrated with the washing buffer (50 mM Tris-HCl, 1 M NaCl, pH 7.4). The cytoplasmic extract of the relevant protein was then loaded on the column, followed by the washing of several column volumes by the washing buffer. The protein was then eluted with a 300 ml linear salt gradient (0~500 mM NaCl in 50 mM Tris-HCl, pH 8) at a flow rate of 2 ml min^{-1}, collecting 5 ml fractions. The results showing the FPLC-purified protein are shown in Chapter 3, Section 3.2.1.2.

2.5 Spectroscopic techniques

2.5.1 Electrospray ionisation mass spectroscopy

Electrospray ionisation mass spectroscopy (ESI-MS) was used to detect the molecular weight of the relevant species and to confirm its presence. The mass spectra data were collected by an internal service using a Micromass LCT Time-of-flight (TOF) mass spectrometer (Waters).

2.5.2 UV-vis absorption spectroscopy

UV-vis absorption spectroscopy was used to record the light absorption values of a suspension at a set range of wavelengths in order to characterise the properties of the heme group (within the protein). Before the analysis, appropriate amount of sodium dithionite

were added to the samples to fully reduce the heme groups in the sample. The spectra were collected using a UV–visible Lambda 2 spectrophotometer (Perkin Elmer) in quartz cuvettes with a pathlength of 1 cm and a scan speed of 240nm min^{-1}. Single readings at a particular wavelength were recorded by a Smart Spec® Plus spectrophotometer (Bio-Rad) in same cuvettes as above.

2.5.2.1 Pyridine hemochrome spectra

In the pyridine hemochrome assay, pyridine was used to form a hemochrome, or a ferroporphyrin complex with two axial nitrogenous bases [120]. The UV-vis spectra of this hemochrome provide information of the type of iron porphyrin present and any modifications on it, such as covalent binding [121] – the heme group is attached to one or two thioether bonds. It does not provide information about whether the protein is correctly folded. A 19% (v/v) pyridine solution in 0.15 M NaOH was used to determine pyridine hemochrome spectra of the samples. The sample was then analysed according to the method above.

2.6 Biochemical analytical techniques

2.6.1 Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)

Protein gel electrophoresis was carried out to separate the proteins in a suspension based on molecular weight determined by the different distances of migration. Samples were mixed with a suitable volume of 3 × SDS loading buffer (New England Biolabs) before loading by syringe. For SDS-PAGE, pre-cast 10% Bis-Tris gels NuPAGE Gel (Invitrogen/Novex) were run with TBE running buffer (Invitrogen/Novex) at 200V for 35 min. Pre-stained protein markers (SeeBlue Plus 2, Invitrogen) were used for the scale of migration distance. The polyacrylamide gel was analysed under an epi-light of a Gel Doc 2000 gel documentation system (Bio-Rad).

2.6.2 Coomassie Brilliant Blue staining of polyacrylamide gels

Coomassie Brilliant Blue staining was carried out in order to visualise all the protein content within the sample loaded on the polyacrylamide gel. It was performed using the SimplyBlue® SafeStain (Invitrogen). The polyacrylamide gel was washed in distilled water for 40 s in a microwave oven and immersed in the stain for 10 min before destaining with distilled water.

2.6.3 Heme-staining of polyacrylamide gels

Heme-staining [122] was performed on SDS-PAGE gels to detect the presence of covalently bound heme in proteins, including *c*-type cytochromes. The SDS-PAGE gels

with migrated proteins were washed in distilled water for 30 s in a microwave oven three times, followed by 20 min incubation in 70 ml of 0.05 mM Sodium acetate solution (pH 2.5) with shaking. 30 ml of 1.25 mM solution of 3,3',5,5'-tetramethylbenzidine (TMBZ, Sigma) in methanol was then mixed with the solution for another 20 min incubation with constant shaking, then followed by the addition of 300 µl of 30% hydrogen peroxide solution. After appropriate time of staining the solution was decanted and replaced by distilled water. The detection occurs by the formation of a green-blue insoluble compound from the oxidation of TMBZ catalysed by the heme.

2.6.4 Western blotting

Western blotting technique was used to determine the level of expression of proteins in cell extracts immunologically. In this project, N-terminal Strep-tag with a sequence of MASWSHPQFEKIEGR or MAWSHPQFEK was added to the sequence of some wild-type proteins and occasionally their variants in order to probe the level of protein production. The SDS-PAGE gel, the Hybond-C Extra nitrocellulose membrane (Amersham Scientifics) and six pieces of blotting paper were equilibrated by immersing in the Transfer buffer (39 mM Glycine, 47 mM Tris-HCl, 0.0375 (w/v) SDS and 20% (v/v) methanol) for 15 min. A Trans-Blot Transfer Cell (Bio-Rad) was then used to transfer the protein from the polyacrylamide gel to the membrane for 1 hour 20 min at 20V, 100mA, with a sandwich of gel on the membrane with 3 pieces blotting paper on either side. The membrane was then immersed in TBA buffer (50mM Tris-HCl, 150mM NaCl, 0.1% (v/v) Tween-20 (Polyoxyethylene (20) sorbitan monolaurate), pH 7.5) with 3% w/v Bovine Serum Albumin (BSA, Sigma-Aldrich) with constant shaking for 1 hour for blocking. The

antibody used in the following 1 hour incubation (with shaking) was either a Strep-Tactin® Alkaline Phosphatase (AP) conjugated antibody (IBA), a Mouse Monoclonal Anti-polyHistidine Alkaline Phosphatase (AP) antibody (Sigma) or a Mouse Monoclonal Anti-Glutathione-S-transferase (GST) antibody (Sigma) with a concentration of 3.5 µl $(10\ ml)^{-1}$ (for the Strep-Tactin antibody) or 3 µl $(10\ ml)^{-1}$ (for the other two antibodies) in the blocking solution for the appropriate tag for the relevent protein. After that, three 5-minute washes were carried out using TBS buffer. The membrane was developed by having the membrane immersed in the solution of SIGMAFAST BCIP/NBT tablet (10 ml $(1\ tablet)^{-1}$ distilled water) for an appropriate time before quenching by distilled water.

2.6.5 Densitometry

The images of polyacrylamide gels and Western blots were captured from the GeneSnap (Syngene) on a Gbox EF Gel Documentation System (Syngene) and analysed by GeneTools (Syngene) software to produce densitometry measurements.

2.7 Bioinformatics techniques

2.7.1 BLAST

Basic Local Alignment Search Tool (BLAST) is a popular bioinformatics tool to search for protein/ nucleotide sequences with high similarity to the query sequence, with a library created to enable fast comparison and alignment of homologous sequences (http://blast.ncbi.nlm.nih.gov/Blast.cgi) [123]. Protein sequences from the database at the UniProt website (http://www.uniprot.org/) were analysed by BLASTp for the homologues of different query sequences. The results of the BLAST analysis are shown in Chpater 6, and the cut-off date for this study is 29/09/2014.

2.7.2 ClustalW

ClustalW is a widely used multiple sequence alignment tool for biological sequence alignment (http://www.ebi.ac.uk/Tools/msa/clustalw2/) [124]. Highly conserved residues in a protein family can be identified through multiple sequence alignment. ClustalW creates a phylogenetic tree to visualise the taxonomy of the multiple sequences.

Chapter 3

The location of specificity determinants in apocytochrome *c* for HCCS

3.1 Introduction

The cytochrome *c* biogenesis System III (Figure 3.1) has only one or two components – HCCS and in some cases HCC_1S. Its function is heme attachment to the apocytochrome *c* or apocytochrome c_1, in the biogenesis of mitochondrial cytochrome c/c_1. HCCS is located in the intermembrane space in the mitochondria. For the biogenesis to take place, the polypeptide chain of the cytochrome *c* has to be imported from the cytosol. The import of apocytochrome *c* into mitochondria is usually through the Tom40 and Tom22 components of a TOM complex as described in Chapter 1, Section 1.6.2.

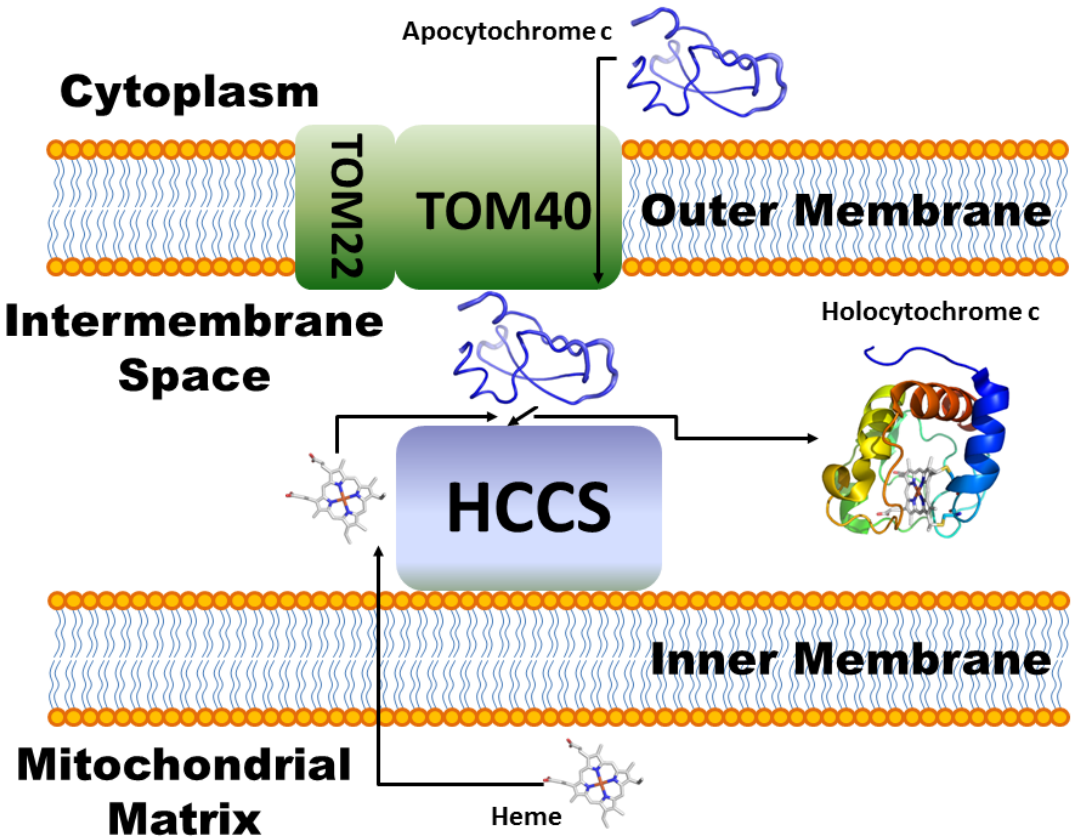


Figure 3.1 The representation of cytochrome *c* biogenesis System III.

Unlike the Ccm System components whose substrate recognition relies almost solely on the CXXCH motif [119,125], HCCS of System III has been shown not to be able to mature bacterial cytochromes, despite the presence of the CXXCH motif [110], such as the cytochrome c_{550} from *Paracoccus denitrificans* whose fold is very similar to the mitochondrial cytochrome *c*, as well as to other bacterial c_2-type cytochromes [126,127] (Figure 3.2). Although a number of residues of mitochondrial cytochrome *c* are highly conserved, many were found to play no vital roles in the maturation process by *S. cerevisiae* HCCS [128]. A previous study has reported some investigations on the regions of cytochrome *c* sequence required for the import and recognition by HCCS [108]. The N-terminal sequence was also examined in random mutagenesis to create a library for two highly conserved residues to assess their roles in structural integrity and constraints [129], but that work did not consider the possible influence of the specificity of HCCS on the results.

A key issue concerning HCCS is what feature it recognises in the apocytochrome *c* substrate. As explained in Chapter 1, there have been indications that the N-terminal region of the substrate apocytochorme *c* is a candidate. Given that the *P. denitrificans* cytochrome c_{550} is not recognised by HCCS, this chapter investigates the possibility that a chimeric protein with the N-terminus of a *S. cerevisiae* mitochondrial cytochrome *c*, but the C-terminus of the *P. denitrificans* cytochrome c_{550} would be recognised by HCCS. Similarly, the *T. brucei* mitochondrial cytochrome *c* is poorly matured by HCCS, even with its AXXCH motif modified to a CXXCH motif. Hence, the level of expression of a chimeric cytochrome *c* with a *S. cerevisiae* cytochrome *c* N-terminal region and a trypanosome cytochrome *c* C-terminal region was also explored.

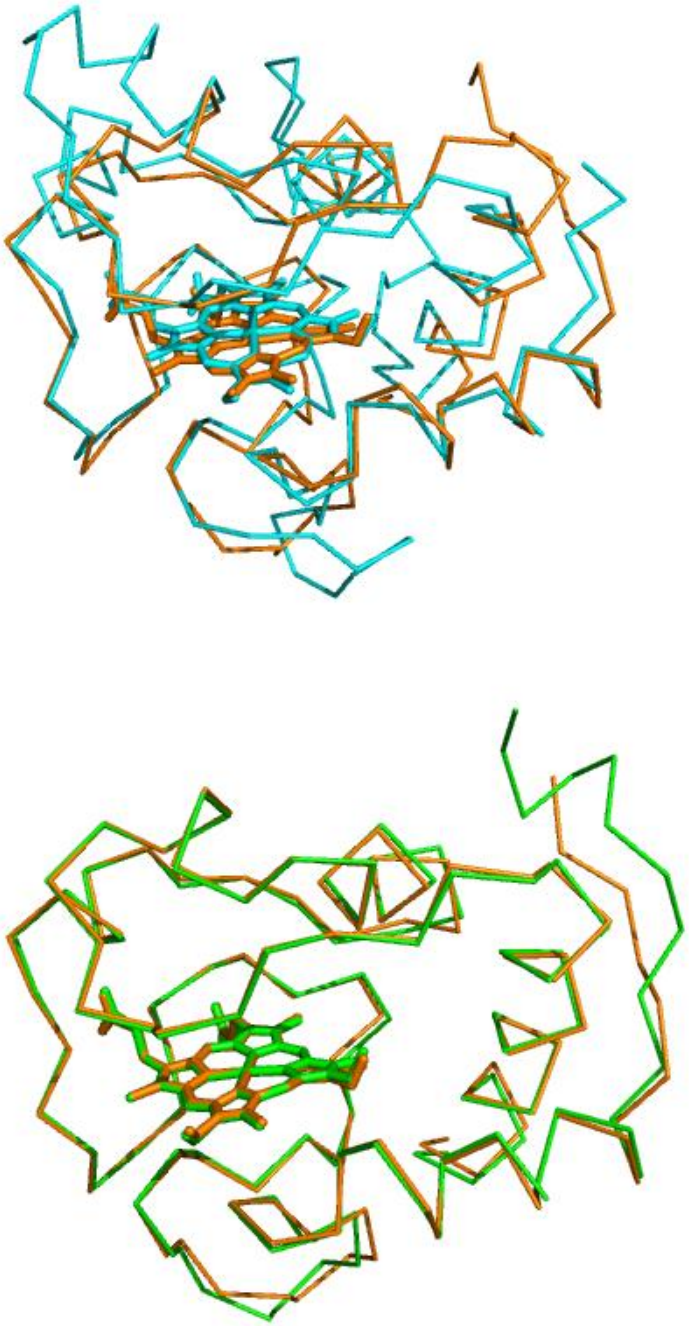


Figure 3.2 The structure alignment of *P. denitrificans* cytochrome c_{550} (cyan) and *S. cerevisiae* cytochrome *c* (orange), as well as the alignment of *S. cerevisiae* cytochrome *c* (orange) and *O. sativa* (rice) mitochondrial cytochrome *c* (green). Note the similar folds of these proteins. The heme group of each cytochrome is depicted. PDB entries are 155C, 1YCC and 1CCR respectively.

Plants utilise the System I/Ccm system for cytochrome *c* biogenesis in mitochondria [50]; however, the amino acid sequences of their mitochondrial cytochromes *c* (UniProt database, from [130,131]) are fairly similar to the animal and yeast ones shown in Figure 3.3. This is intriguing in the sense that a similarly folded cytochrome *c* with a sequence with a 72% similarity (*A. thaliana* to horse, Emboss pairwise alignment) utilises a different cytochrome *c* biogenesis system. Therefore, it is worth exploring whether there are sequence differences that prevent plant cytochromes *c* from being recognised by System III. Similarly, the putative System V is utilised by the trypanosome cytochrome *c*, which

```
Yeast numbering          1         10         20         30         40
P. denitrificans ---------- QDGDAAKGEK EFN-KCKACH MIQAPDGTDI IKGGKTGPNL
T. brucei        -MPPKERAAL PPGDAVRGEK LFKGRAAQCH TGTKGG---- --SNGVGPNL
S. cerevisiae    ------MTEF KAGSAKKGAT LFKTRCLQCH TVEKGG---- --PHKVGPNL
A. thaliana      ---MASFDEA PPGNPKAGEK IFRTKCAQCH TVEKGA---- --GHKQGPNL
A. thaliana      ---MASFDEA PPGNAKAGEK IFRTKCAQCH TVEAGA---- --GHKQGPNL
E. caballus      ---------- -MGDVEKGKK IFVQKCAQCH TVEKGG---- --KHKTGPNL
H. sapiens       ---------- -MGDVEKGKK IFIMKCSQCH TVEKGG---- --KHKTGPNL

                        50         60         70         80         90
P. denitrificans YGVVGRKIAS EEGFKYGEGI LEVAEKNPDL TWTEADLIEY VTDPKPWLVK
T. brucei        YGIVGRKSGT VEGFTYSK-- ---ANQDSGV MWTPQVLDVY LENPKKFMPG
S. cerevisiae    HGIFGRHSGQ AEGYSYTD-- ---ANIKKNV LWDENNMSEY LTNPKKYIPG
A. thaliana      NGLFGRQSGT TPGYSYSA-- ---ANKSMAV NWEEKTLYDY LLNPKKYIPG
A. thaliana      NGLFGRQSGT TAGYSYSA-- ---ANKNKAV EWEEKALYDY LLNPKKYIPG
E. caballus      HGLFGRKTGQ APGFSYTD-- ---ANKNKGI TWKEETLMEY LENPKKYIPG
H. sapiens       HGLFGRKTGQ APGYSYTA-- ---ANKNKGI IWGEDTLMEY LENPKKYIPG

                       100        110        120        130
P. denitrificans MTDDKGAKTK MTFKMGKNQA DVVAFLAQNS PDAGGDGEAA AEGESN
T. brucei        TKMSFAGLKK PQERADLIAY LETLKD---- ---------- ------
S. cerevisiae    TKMAFGG--- --LKKEKDRN DLITYLKKAC E--------- ------
A. thaliana      TKMVFPGLKK PQDRADLIAY LKEGTA---- ---------- ------
A. thaliana      TKMVFPGLKK PQDRADLIAY LKESTAPK-- ---------- ------
E. caballus      TKMIFAG--- --IKKKTERE DLIAYLKKAT NE-------- ------
H. sapiens       TKMIFVG--- --IKKKEERA DLIAYLKKAT NE-------- ------
```

Figure 3.3 Multiple sequence alignment of *Paracoccus denitrificans* cytochrome c_{550} (Signalling sequence removed), *Trypanosoma brucei* cytochrome *c*, *Saccharomyces cerevisiae* iso-1 cytochrome *c* and, two isoforms of *Arabidopsis thaliana* cytochrome *c*, *Equus caballus* cytochrome *c* and *Homo sapiens* cytochrome *c*. The cysteines of the CXXCH heme-binding motif are highlighted in grey. The conserved TKM sequence in which the M is an axial ligand to the heme iron is underlined.

also has a sequence similarity with horse heart cytochrome *c*, and therefore the same question of sequence compatibility with System III arises.

As an essential component of the cytochrome bc_1 complex (System III), cytochrome c_1 has a hydrophilic globular domain and a membrane anchor (shown in Figure 3.4) [35]. Relatively limited study has been done on the substrate recognition of HCCS/HCC_1S in different organisms with respect to the cytochrome c_1 maturation; examples include [78] and [132]. As it has already been mentioned in Section 1.6.4, cytochrome c_1 has a unique mechanism of import into mitochondria, as shown in Figure 3.5. Cytochrome c_1 is produced in the cytosol by ribosomes from a nuclear gene, and has two independent signal sequences together with transmembrane helices on either end of the protein [100]. The N-terminal signal sequence and transmembrane helix are first inserted and anchored into the mitochondrial inner membrane, and this N-terminal signal sequence is cleaved before the internal signal sequence and helix anchor are also inserted into the membrane. However, it is possible that this process is accomplished by either the conservative import model or the stop-transfer model [35]. Conservative import model is also called the re-export model. In this model, the precursor protein is transported from cytosol towards the mitochondrial matrix directly through both the outer membrane and inner membrane. Upon arrival, the N-terminal signal sequence is cleaved and the rest of the polypeptide chain then binds with a factor called hsp60, and through ATP hydrolysis the sequence is then recognised by translocators to locate them to the inner membrane. Finally, the polypeptide is cleaved again and becomes the mature protein [133]. In the stop-transfer model, the polypeptide chain is imported through the two membranes and the sorting sequence recognised by the translocator on the inner membrane. The signal peptide on the N-terminus is cleaved and

the rest of the protein is only imported into the intermembrane space through the outer membrane. It is then cleaved again to produce the mature protein [133]. HCC_1S possibly then binds and interacts with the apocytochrome c_1 with its initial sequence cleaved, and then catalyses the heme attachment. The N-terminal anchor is also cleaved by peptidase after the covalent binding of heme, and the holocytochrome c_1 can fold into the correct conformation and assemble into Complex III [35,100].

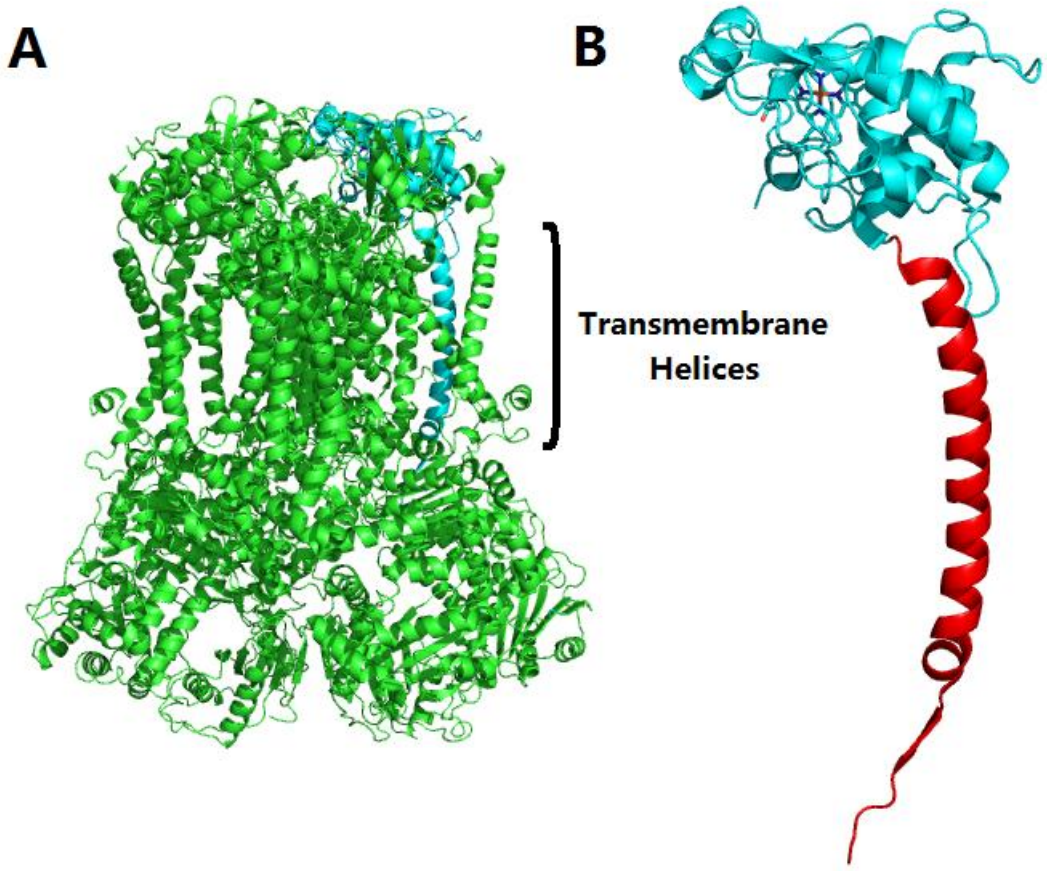


Figure 3.4 (A) Cytochrome bc_1 complex/Complex III of *S. cerevisiae* (PDB entry 1KYO). The cytochrome c_1 (Chain D) is highlighted in cyan colour. **(B)** The modified cytochrome c_1 for expression. The C-terminal transmembrane helix (red) is removed in the sequence to optimise the expression of protein.

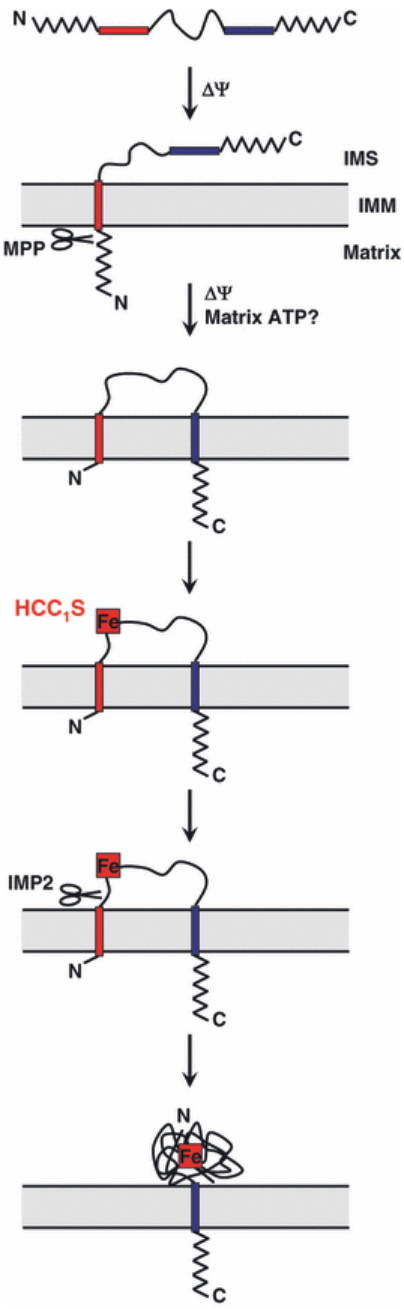


Figure 3.5 Schematic diagram for the mechanism of the import of cytochrome c_1 into the mitochondrial inner membrane. IMS, IMM, MPP, IMP2 and ΔΨ stand for intermembrane space, inner mitochondrial membrane, matrix processing peptidase, inner membrane peptidase 2 and membrane potential. The N-terminal transmembrane helix is coloured red, and the C-terminal transmembrane helix is coloured blue. The heme group is depicted in a red box with 'Fe' written in it. The HCC_1S enzyme is coloured red. Adapted from Allen [35].

In this chapter, different types of experiments were designed to investigate different aspects of the substrate recognition specificity of HCCS, for example the chimeric cytochrome *c* experiments and the analysis of HCCS specificity towards plant and trypanosome cytochromes *c*.

3.2 Results

Both of UV-vis spectroscopy and Heme-staining methods could be used for the detection of holocytochrome *c* in this thesis. The sensitivity of both methods is compared in Figure 3.6, which is adapted from Goddard et al. [115]. It was deduced that the sensitivity of UV-vis spectroscopy is slightly higher than that of Heme-staining, but the thresholds of both of them are very low [115].

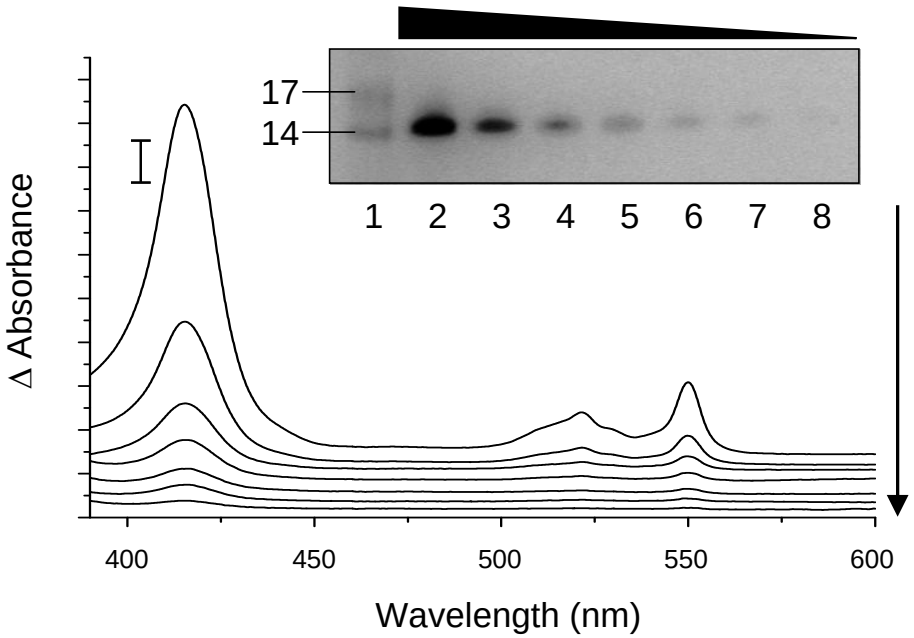


Figure 3.6 Comparison of sensitivity levels of heme-staining and UV-vis spectroscopic techniques. *P. denitrificans* cytochrome c_{550} was expressed, matured and detected by both methods in a sequentially diluted fashion. It is adapted from the supplementary data from Goddard et al [115]. The molecular markers (kDa) are shown in Lane 1.

Table 3.1 is a summary table for the results for cytochromes *c* maturation of the wild-type and variants which mentioned in this chapter for comparison.

3.2.1 Chimeric cytochrome *c*

3.2.1.1 Co-production of *P. denitrificans* cytochrome c_{550} with *S. cerevisiae* HCCS

It has been reported previously (see Chapter 1) that HCCS is not able to mature a bacterial cytochrome *c*, namely *P. denitrificans* cytochrome c_{550}. This statement is confirmed by the absence of heme attachment of *P. denitrificans* cytochrome c_{550} (with signal peptide truncated) when co-produced with *S. cerevisiae* HCCS in the *E. coli* cytoplasm (Figure 3.8 (B)).

3.2.1.2 Construction and production of the yeast-bacteria chimeric cytochrome *c*

A chimeric protein comprising the N-terminal 18 amino acids of iso-1 cytochrome *c* (*CYC1*) from *S. cerevisiae* followed by the C-terminal sequence of cytochrome c_{550} from *P. denitrificans* (the CXXCH heme-binding motif and residues C-terminal to it) was constructed in two steps: (1) substitution of the N-terminus (including the CXXCH motif) of *P. denitrificans* cytochrome c_{550} by the sequence of *S. cerevisiae* (the initial gene construct was completed by Dr K. Sam from Ferguson lab previously) (2) reversion of the CXXCH motif to the *P. denitrificans* c_{550} sequence. The comparison of the N-terminal region to the wild-type cytochromes and other variants is represented in Figure 3.3, the N-terminal region sequence is shown in Figure 3.6 (C) and the final sequence is shown in Figure 3.8 (A) (i).

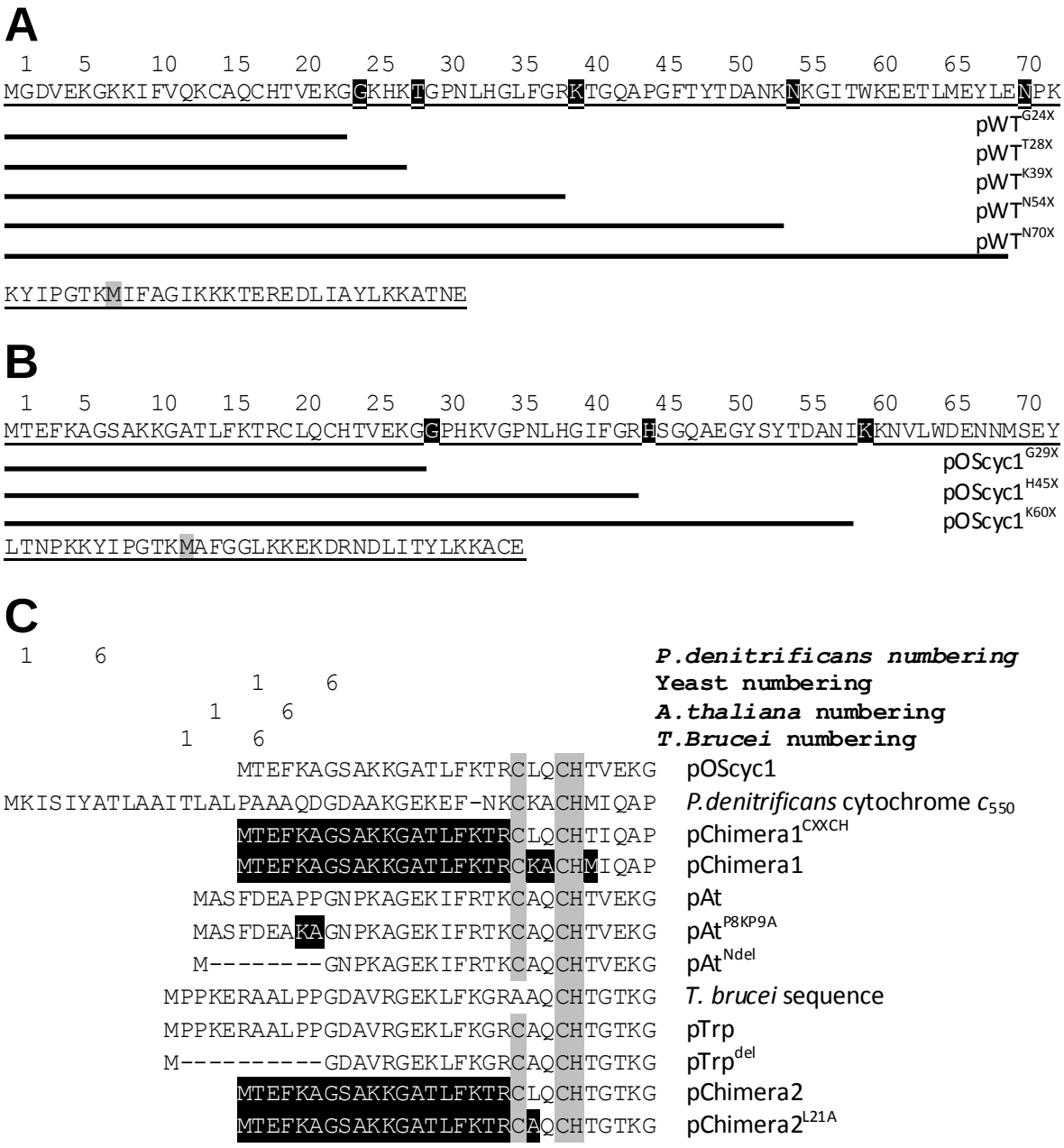


Figure 3.7 (A) Amino acid sequence of *E. caballus* cytochrome *c* and representations of the truncated N-terminal fragments of it with C-termini at residues G24, T29, K40, N55 and N71 (all coloured in black boxes). The heme axial ligand residue M86 is depicted in a grey box. **(B)** Amino acid sequence of *S. cerevisiae* cytochrome *c* iso-1 and representations of the truncated N-terminal fragments of it with C-termini at residues G29, H45 and K60 (all coloured in black boxes). The heme axial ligand residue M86 is depicted in a grey box. **(C)** N-terminal sequences in plasmids of *S. cerevisiae* cytochrome *c* iso-1(pOScyc1), *A. thaliana* cytochrome *c* (pAt), *T. brucei* cytochrome *c* (pTrp) and their variants (including chimeric cytochrome *c*) with *P.denitrificans* cytochrome c_{550} for

comparison. The cysteines and histidines in the CXXCH motif are depicted in grey boxes, and the mutations are depicted in black boxes.

Name of the Plasmid(s)	Results for Maturation
pWT^{G24X}	Failure
pWT^{T28X}	Failure
pWT^{K39X}	Failure
pWT^{N54X}	Failure
pWT^{N70X}	Failure
$pOScyc1^{G29X}$ + pACcyc3	Failure
$pOScyc1^{H45X}$ + pACcyc3	Success
$pOScyc1^{K60X}$ + pACcyc3	Success
$pChimera1^{CXXCH}$ + pACcyc3	Success
pChimera1+ pACcyc3	Success
pAt + pACcyc3	Failure
pAt^{P8KP9A} + pACcyc3	Failure
pAt^{Ndel} + pACcyc3	Failure
pAtperi + pEC86	Moderate Success
pTrp + pACcyc3	Failure
$pTrp^{del}$ + pACcyc3	Failure
pChimera2 + pACcyc3	Success
$pChimera2^{L21A}$ + pACcyc3	Success
pOScyc1+ pACcyc3	Success
pOScyc1+ pHCC1S	Failure
pSccyc1+ pACcyc3	Failure
pSccyc1+ pHCC1S	Failure

Table 3.1 List of results for maturation of wild-type cytochromes *c* and the variants made in Chapter 3.

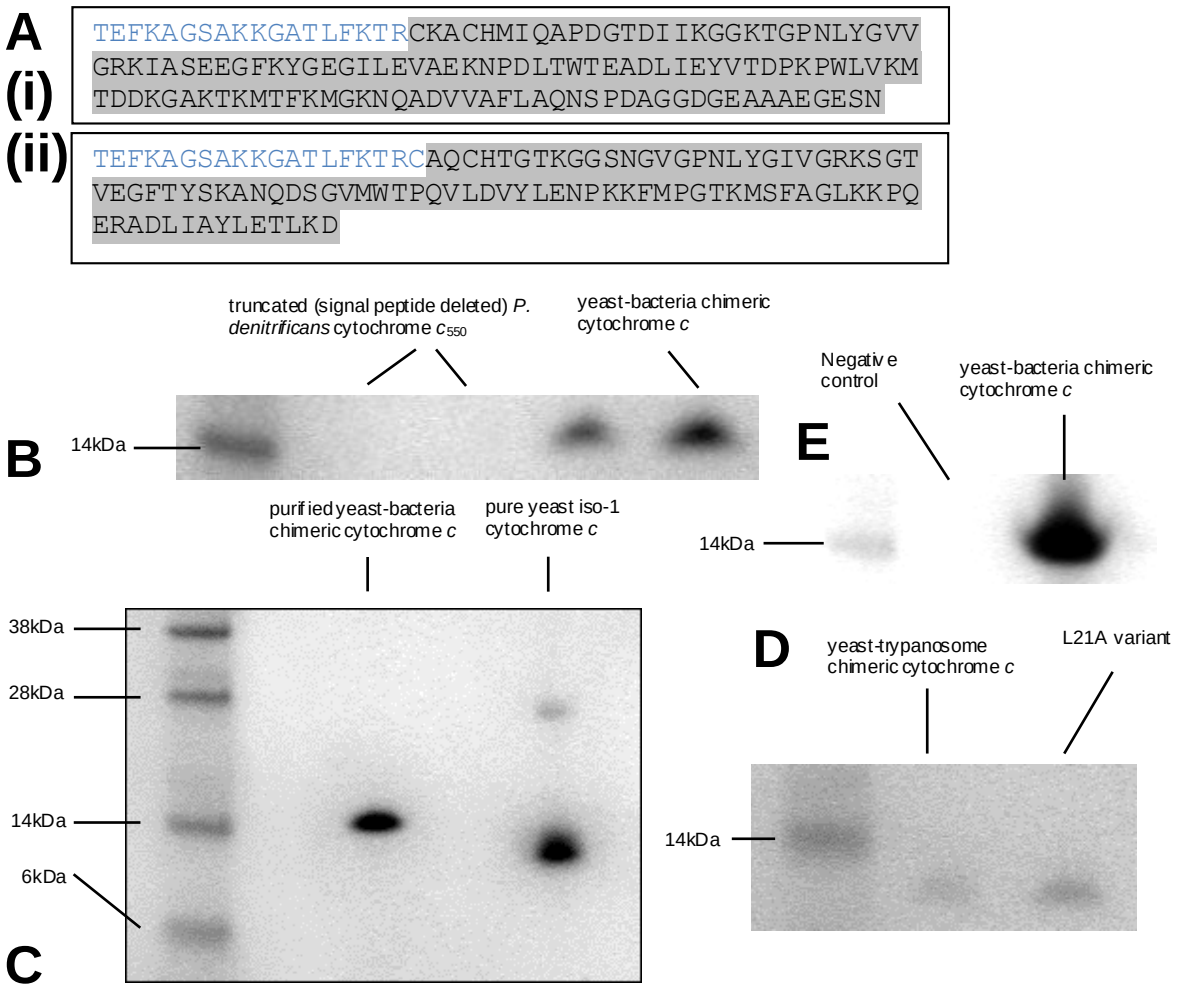


Figure 3.8 (A) The amino acid sequences of two chimeric cytochromes *c*. The sequences in blue are the *S. cerevisiae* iso-1 cytochrome *c* N-terminal sequences, and that highlighted in grey are (i) the sequence of the cytochrome c_{550} from *P. denitrificans* (excluding the amino acids N-terminal to CKACH) and (ii) the sequence of the cytochrome *c* from *T. brucei* (excluding the amino acids N-terminal to CAQCH). **(B)** Heme-stained SDS-PAGE analysis of cytoplasmic extracts of the yeast-bacteria chimeric cytochrome *c* and the N-terminal truncated (signal peptide deleted) *P. denitrificans* cytochrome c_{550}. Equal volumes of cytoplasmic extracts - diluted so as to contain equal amounts of material from wet cell mass - were loaded on each lane and the data were normalised according to wet cell mass. **(C)** The yeast-bacterial cytochrome *c* chimera (shown in A) analysed by SDS–PAGE stained for covalently attached heme. Pure yeast iso-1 cytochrome *c* (mass = 12 588 Da, Sigma) was used as a control for comparison. **(D)** The yeast-trypanosome cytochrome *c* chimeras analysed by SDS–PAGE stained for covalently attached heme. Equal volumes of cytoplasmic extracts - diluted so as to contain equal amounts of material from wet cell mass - were loaded on each lane and the data were normalised according to wet cell mass. The molecular markers are labelled as shown in the figures. **(E)** Negative control (sample

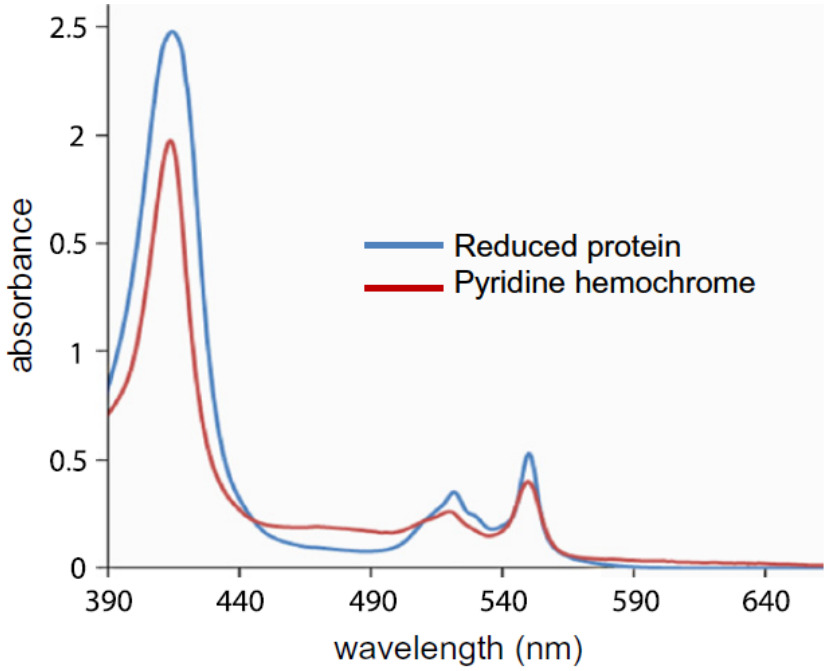
without the HCCS containing plasmid) and the yeast-bacterial chimeric cytochrome *c* analysed by SDS–PAGE heme-staining technique.

3.2.1.3 UV-vis spectroscopy and Heme-stained SDS-PAGE analysis of the yeast-bacteria chimeric cytochrome *c*

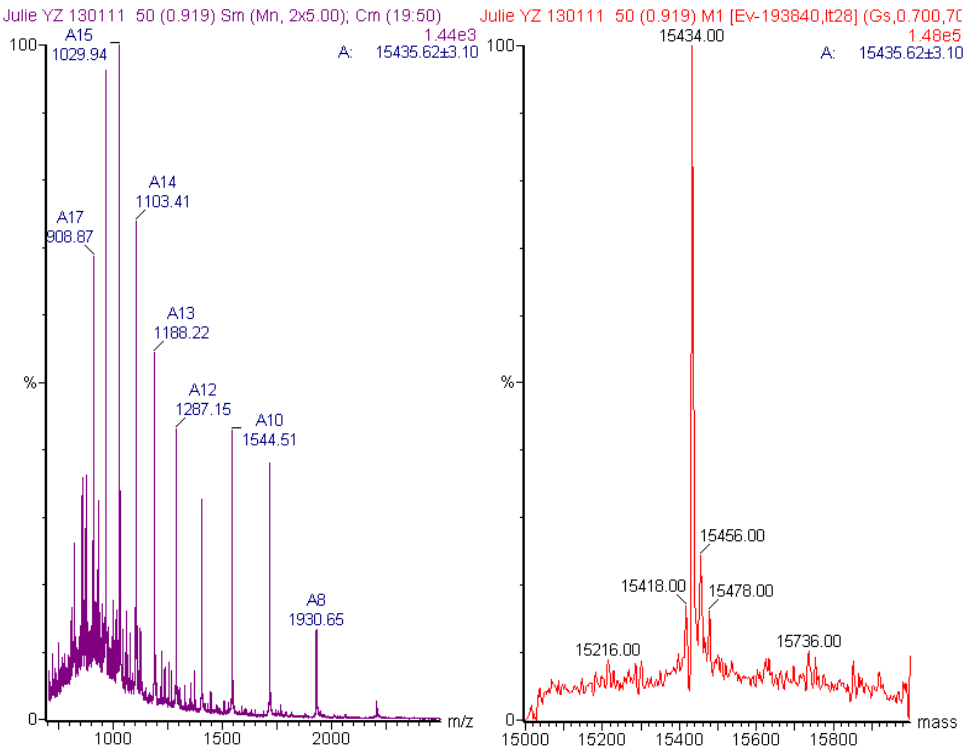
The gene for the chimera was co-expressed in the *E. coli* cytoplasm with *S. cerevisiae* HCCS from a separate plasmid (pACcyc3). The periplasmic fraction was removed from the harvested cells to avoid interference by endogenous cytochromes in that compartment. The cytoplasmic extract had a distinct red colour and was fractionated by anion-exchange chromatography. The red fractions with the purified protein were then analysed by electrospray (ES) mass spectrometry, which showed the predominant protein species to have a mass of 15,435.6 Da (±3.1 Da). The mass spectra result in Figure 3.9 (B) indicates that the predicted mass of the chimeric protein (which has its N-terminal methionine cleaved, as would be expected in the *E. coli* cytoplasm) is 14,816 + 616.5 Da for heme, resulting in a total mass of 15,432.5 Da. This difference falls within experimental error of the value observed. Thus, the presence of the successful expression and maturation of this chimeric cytochrome *c* could be confirmed. The cytoplasmic extract was analysed by UV–vis spectroscopy and SDS–PAGE followed by heme-staining (Figure 3.8 (C) and Figure 3.9 (A)).

The UV-vis spectra of the reduced protein and the pyridine hemochrome were indistinguishable from wild-type *P. denitrificans* cytochrome c_{550}. The amount of chimeric

protein obtained was comparable to the *S. cerevisiae* cytochrome *c* itself produced from a same type of plasmid in the presence of plasmid-encoded *S.cerevisiae* HCCS. A negative



A



B

Figure 3.9 (A) UV-vis spectroscopic analysis of the purified chimeric yeast-bacterial cytochrome. The spectrum on reduction with disodium dithionite is shown in blue and that following the addition of 0.15 M NaOH and 19% (vol/vol) pyridine shown in red. In the spectrum of the reduced protein the wavelength maxima are 415, 520 and 550 nm. The α-band is at 550 nm in the pyridine hemochrome spectrum. **(B)** The mass spectroscopic data of the purified chimeric yeast-bacterial cytochrome.

control (sample grown containing the plasmid with chimeric cytochrome *c*, but not the plasmid with HCCS) is shown in Figure 3.8 (E).

3.2.1.4 Construction and expression of the yeast-trypanosome chimeric cytochrome *c*

The N-terminal amino acid stretch of iso-1 cytochrome *c* (*CYC1*) from *S. cerevisiae* (18 residues) is followed by the C-terminal sequence of cytochrome *c* from *T. brucei* (the AXXCH heme-binding motif and residues C-terminal to it). Similar to the yeast-bacteria chimeric protein, the yeast-trypanosome chimeric cytochrome *c* was also constructed in two steps: (1) substitution of the N-terminus (including the AXXCH motif) of *T. brucei* cytochrome *c* to the sequence of *S. cerevisiae* (this initial work was also completed by Sam, K. from Ferguson lab previously) (2) reversion of the AXXCH motif to *T. brucei* c_{550} sequence by the substitution L21A. The N-terminal region sequence is shown in Figure 3.7 (C) and the final sequence is shown in Figure 3.8 (A) (ii).

3.2.1.5 Heme-stained SDS-PAGE analysis on the yeast- trypanosome chimeric cytochrome *c*

The co-expressions of the initial chimeric protein and the L21A variant *with S. cerevisiae* HCCS were successful. The heme-stained SDS-PAGE analysis result is shown in Figure 3.8 (D). The bands on the gel demonstrate the successful expression and maturation of both proteins in *E. coli*. as the band has an expected molecular weight ~12.3 kDa. No UV-vis spectrum was collected from this variant.

3.2.2 Truncated mitochondrial cytochrome *c*

It has been shown previously by Veloso *et al.* [134] that a stable N-terminal fragment of horse heart cytochrome *c* could be chemically synthesised and has heme attached to it by liver mitochondrial extracts. If such a fragment can be formed by HCCS, then the importance of the N-terminal sequence of mitochondrial cytochrome *c* could be further demonstrated. The approach involved the truncation of the C-terminus of mitochondrial cytochrome *c* which gives fragments with the N-terminal region including the CXXCH motif. This strategy was performed on both horse heart cytochrome *c* and *S. cerevisiae* cytochrome *c* to produce eight variants in total.

3.2.2.1 Truncated *E. caballus* mitochondrial cytochrome *c*

A stop codon was inserted at five different positions in the gene of *E. caballus* cytochrome *c* to produce five variants with different lengths of N-terminal region: G24X, T29X, K40X, N55X and N71X. X marks stand for the stop codon positions. These truncation

variants are represented in Figure 3.7 (A). The genes for the truncated variants and the wild-type protein were co-expressed with *S. cerevisiae* HCCS in *E. coli*. Figure 3.10 (A) shows the results of heme-stained SDS–PAGE analysis of cytoplasmic extracts of these variants. The heme-staining analysis indicates that none of these variants had a detectable maturation of holocytochrome *c*, since no bands can be observed.

3.2.2.2 Truncated *S. cerevisiae* mitochondrial cytochrome *c*

A stop codon was introduced in different positions in the gene of *S. cerevisiae* iso-1 cytochrome *c* (CYC1) to remove different lengths of the C-terminal region, producing three variants: G29X, H45X and K60X. The X indicates the position of the inserted stop codon, as indicated on the sequence alignment in Figure 3.7 (B). The genes for the truncated variants and the wild-type protein were co-expressed with *S. cerevisiae* HCCS in *E. coli*; Figure 3.10 (B) shows the results of heme-stained SDS–PAGE analysis of cytoplasmic extracts. Heme-stained bands were observed with predicted molecular masses (~6.5 kDa and ~8 kDa) for H45X and K60X variants, but no reproducible product for the shortest variant, G29X, was detected. The low levels of maturation for these heme-containing truncated cytochromes *c* (illustrated in Figure 3.10 (C)) can be ascribed to relative instability of these molecules. The marks/bands at the very low molecular weight may represent degraded fragments of cytochrome *c* with heme attached, as well as free heme dissociated from these heme fragments. A negative control is included in Figure 3.11, and as there is no band observed on the blank negative, it can be deducted that at least some of the components of the stained material on the gel are from degraded cytochromes.

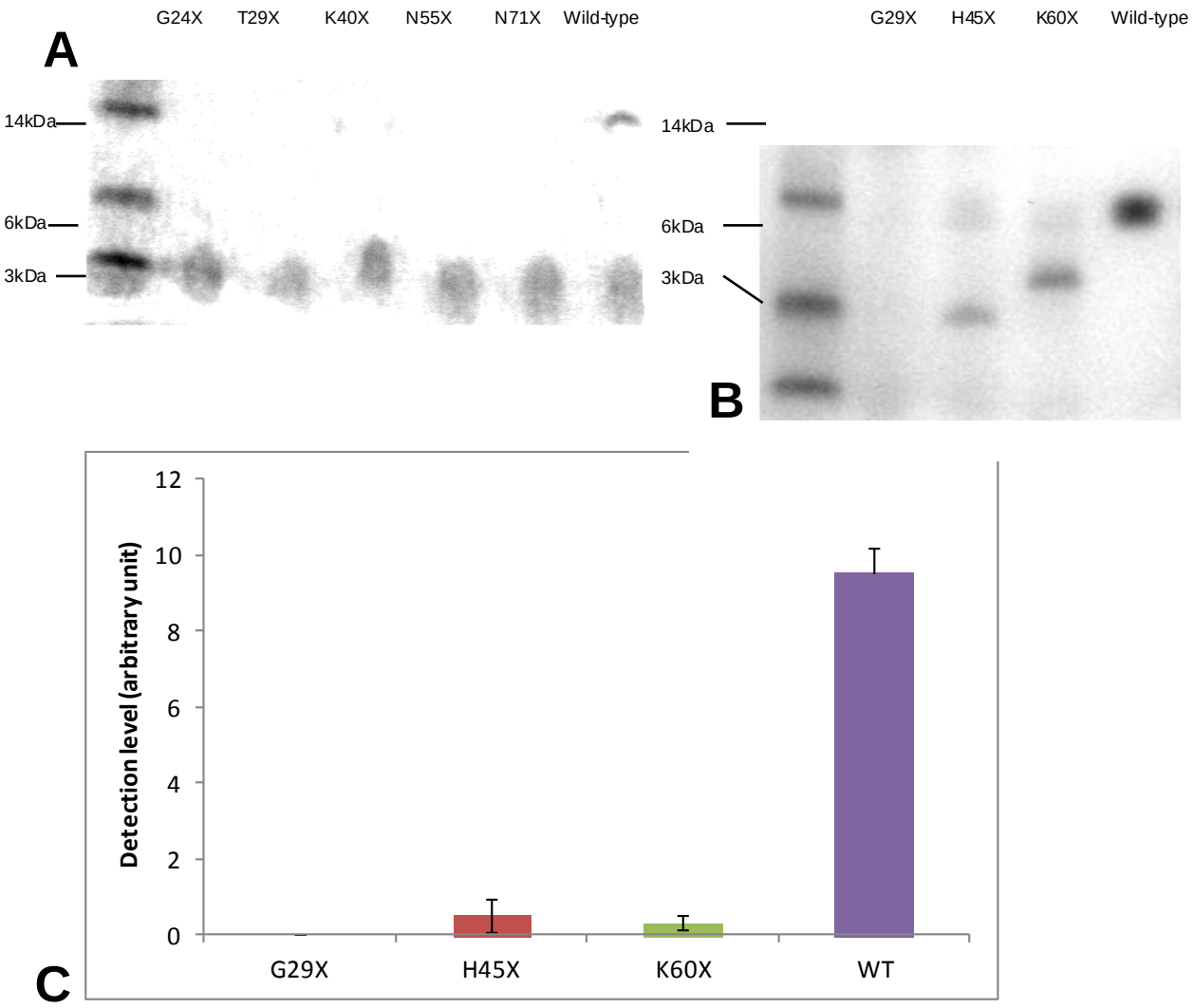


Figure 3.10 (A) Heme-stained SDS-PAGE analysis of cytoplasmic extracts of wild-type and C-terminal truncation variants of *E. caballus* cytochrome *c*. Equal volumes of cytoplasmic extraction (diluted so as to contain equal amounts of material from wet cell mass) were loaded on each lane and normalised according to wet cell mass. The bands in the bottom of the gel in each lane are stains for free hemes (or from the degradation of cytochromes). The molecular markers are labelled as shown in the figures. **(B)** Heme-stained SDS-PAGE analysis of cytoplasmic extracts of *E. coli* containing wild-type and C-terminal truncation variants of *S. cerevisiae* cytochrome *c*. Equal volumes of cytoplasmic extraction (diluted so as to contain equal amounts of material from wet cell mass) were loaded on each lane and normalised according to wet cell mass. Note that the wild-type cytochrome *c* was 100-fold diluted. The molecular markers are labelled as shown in the

figures. **(C)** Average densitometry recordings for G29X, H45X, K60X variants and the wild-type *S. cerevisiae* cytochrome *c* (six repeats were made on all variants and wild-type).

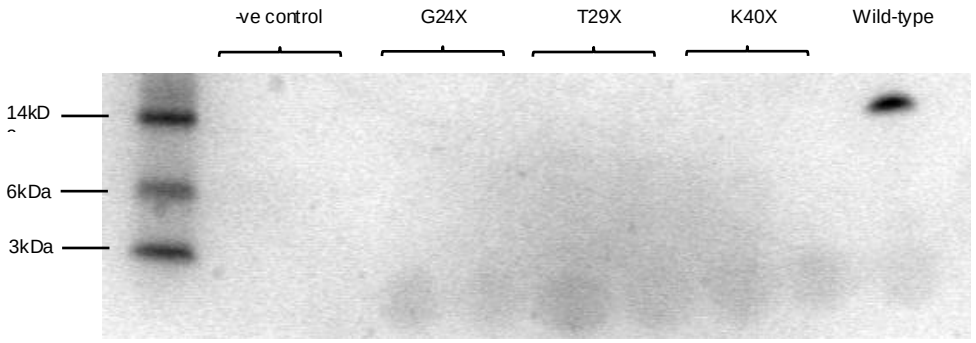


Figure 3.11 Heme-stained SDS-PAGE analyses of cytoplasmic extracts of wild-type and C-terminal truncation variants of *E. caballus* cytochrome *c*. Equal volumes of cytoplasmic extraction (diluted so as to contain equal amounts of material from wet cell mass) were loaded on each lane and normalised according to wet cell mass. The bands in the bottom of the gel in each lane are stains for free hemes (or from the degradation of cytochromes).

3.2.3 *A. thaliana* cytochrome *c* variants

Although the amino acid sequence length and the expected folds for the *S. cerevisiae* cytochrome *c* and the *A. thaliana* mitochondrial cytochrome *c* are similar (since the mitochondrial cytochrome *c* of another plant, *Oryza sativa* (rice), has a very similar fold to that shown in the alignment of cartoons in Figure 3.2), and the sequence homology is high, the *A. thaliana* cytochrome *c* differs from the *S. cerevisiae* cytochrome *c* in some ways and one of the major differences is the N-terminal region. However, it should be noted that the

apocytochromes *c* are generally without clearly defined secondary structure [89,135], whereas the holocytochromes *c* are folded and have stable structures, as mentioned in Chapter 1.

In view of the demonstrated importance of the N-terminus of apocytochrome *c* for HCCS recognition, it is notable that the N-terminal region of *A. thaliana* and indeed all plant cytochromes *c* is longer than for yeast and mammalian cytochromes *c*. *A. thaliana* cytochrome *c* is processed by System I, and it is possible that the apocytochrome *c*, despite many sequence similarities with yeast cytochrome *c* to the immediate N-terminal side of the CXXCH motif, could not be a substrate for an HCCS enzyme.

One of the distinct features from the two isoforms of *A. thaliana* mitochondrial cytochrome *c* sequences is the sharing of a proline repeat in the N-terminal region which is shared among many other plant cytochromes. Thus, attempts were made to determine whether the wild-type or variants with plant-specific *A. thaliana* cytochrome *c* features removed could be processed by HCCS.

3.2.3.1 The cloning of *A. thaliana* cytochrome *c*

Two isoforms (*CYTc-a* and *CYTc-b*) exist in the genome of *A. thaliana* (Genebank: AY062853, AY045944) and the latter was chosen for the attempts of cloning and expression. The *CYTc-b* gene on the plasmid was PCR amplified (Chapter 2, Section 2.3). The DNA fragments containing *CYTc-b* gene were then purified by agarose gel

electrophoresis followed by gel extraction (Chapter 2, Section 2.2.3). The DNA fragment containing *CYTc-b* gene was then ligated (Chapter 2, Section 2.2.4) into cloning vector pET3a to produce a plasmid pAt. The plasmids containing cDNA clones of *A. thaliana* cytochrome *c* were a gift kindly provided by Dr Géraldine Bonnard of University of Strasbourg.

3.2.3.2 The co-expression of the *A. thaliana* cytochrome *c* gene with the *S. cerevisiae* *HCCS* gene

The cloned *A. thaliana* cytochrome *c* sequence on the plasmid pAt with an N-terminal Strep-tag was co-expressed with *S. cerevisiae* HCCS in *E. coli*. The heme-stained SDS-PAGE analysis and Western blot for a cytoplasmic extraction is shown in Figure 3.12 (A) (B). The *A. thaliana* cytochrome *c* was not produced nor matured, as no trace of apocytochrome *c* could be found on a Western blot using anti-Strep antibodies; or otherwise, the *A. thaliana* cytochrome *c* was degraded readily when produced. The N-terminal Strep-tag was not considered to be problematic for the expression and/or maturation, as it will be shown in Chapter 4 that such a tag did not interfere with HCCS action on tagged *S. cerevisiae* apocytochrome *c*.

3.2.3.3 Lysine and Alanine substitutions of the N-terminal Prolines of *A. thaliana* cytochrome *c*

The conserved proline repeats in the N-terminal region of *A. thaliana* cytochrome *c* were replaced with a lysine and an alanine in order to modify this feature to have an N-terminal

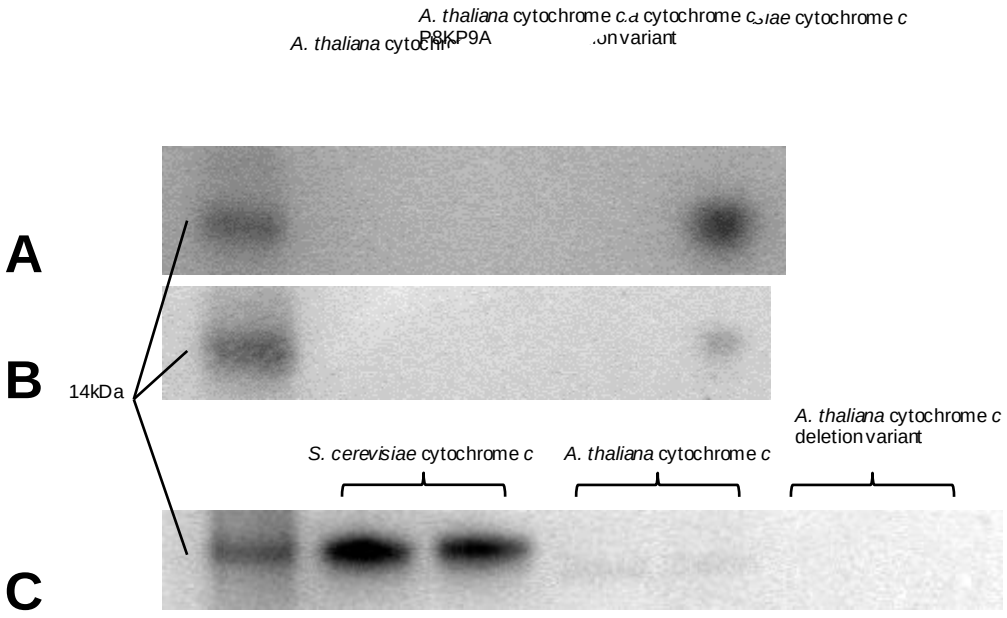


Figure 3.12 (A) Heme-stained SDS-PAGE analysis of cytoplasmic extracts of *E. coli* for the co-expression and maturation trial of *A. thaliana* mitochondrial cytochrome *c* (and its variants) with *S. cerevisiae* HCCS gene, and *S. cerevisiae* cytochrome *c* as a control. Equal volumes of cytoplasmic extracts - diluted so as to contain equal amounts of material from wet cell mass - were loaded on each lane. The molecular markers are labelled as shown in the figures. **(B)** Western Blot of cytoplasmic extracts of *E. coli* for the co-expression trial of *A. thaliana* mitochondrial cytochrome *c* gene (and its variants) with *S. cerevisiae* HCCS gene, and *S. cerevisiae* cytochrome *c* as a control. An anti-Strep antibody (IBA) was used in the blotting. The volumes of cytoplasmic extractions loaded in each lane are equal (diluted so as to contain equal amounts of material from wet cell mass). **(C)** Heme-stained SDS-PAGE analysis of periplasmic extracts of *E. coli* containing wild-type and N-terminal deletion variants of periplasmically targeted *A. thaliana* cytochrome *c* and periplasmically targeted *S. cerevisiae* cytochrome *c* as a control. Duplicate extracts for each protein are shown.

region more similar to the *S. cerevisiae* cytochrome *c*, so that the *S. cerevisiae* HCCS possibly might recognise and mature the cytochrome. The N-terminal region sequences of

these variants are shown in Figure 3.7 (C), and the results of heme-staining analysis on SDS-PAGE and Western blot are shown in Figure 3.12 (A) (B); the lysine and an alanine double substitution variants were not detectably produced as holoproteins and no heme-stained bands could be seen on the gel.

3.2.3.4 N-terminal deletion variant of *A. thaliana* cytochrome *c*

The eight-residue N-terminal element, including the prolines (ASFDEAPP), was deleted from the wild type *A. thaliana* cytochrome *c* to make the deletion variant Atdel, as shown in Figure 3.7 (C). As already mentioned, the N-terminal sequence of the *A. thaliana* cytochrome *c* is longer than the N-terminus of *S. cerevisiae* cytochrome *c*, therefore this deletion variant resembles the *S. cerevisiae* cytochrome *c* to an even greater extent than the substitution variant. No detectable amounts of holoprotein or even apoprotein were observed from the results in Figure 3.12 (A) (B).

3.2.3.5 Periplasmic targeted variants of *A. thaliana* cytochrome *c*

In contrast to System III, System I has a broader specificity, and usually attaches heme to any proteins with a CXXCH motif [55]. Thus the Ccm apparatus, or System I of *E. coli* that functions in the periplasm, essentially recognises only the CXXCH motif and thus *A. thaliana* mitochondrial holocytochrome *c* and its variants might be formed, whereas the HCCS system is seemingly unable to process *A. thaliana* cytochrome *c* and/or variants to

form holoproteins. To test the compatibility of the cloned *A. thaliana* cytochrome *c* with System I in the periplasm of *E. coli*, a periplasmic targeting sequence from *P. aeruginosa* cytochrome $c_{551/552}$ (MKPYALLSLLATGTLLAQGAWAED) was inserted at the N-terminus of the *A. thaliana* cytochrome *c* in two steps due to the length of this signal sequence. The result is shown in Figure 3.12 (C), with very faint bands detected for *A. thaliana* cytochrome *c* on the gel of heme-stained SDS-PAGE analysis, but no detection of the periplasmically targeted deletion variant of *A. thaliana* cytochrome *c* was found.

3.2.4 *T. brucei* cytochrome *c* deletion variant

As explained in Chapter 1, the cytochromes *c* of trypanosomatid organisms are distinct because they have an AXXCH rather than CXXCH sequence in the N-terminal region. Following alteration of AXXCH to CXXCH it has been shown that the protein is processed by System I [40]. However, previous attempts to obtain significant expression and maturation of cytochrome *c* using HCCS of the CXXCH variant protein have had little success [116] even though the N-terminal sequence of the protein is similar to that of *S. cerevisiae* and animal cytochromes *c*, as shown in Figure 3.3. However, interestingly the N-terminal region is longer than in yeasts and animals and has similarities with plant cytochromes *c*, with similar proline repeat elements (Figure 3.7 (C)). Consequently, the region coding for the most extreme N-terminus in the *T. brucei* cytochrome *c* sequence was deleted, as shown in the alignment of the N-terminal sequences of the variants in the plasmid in Figure 3.7 (C). The maturation by HCCS of this variant was tested. No detectable heme attachment could be seen, and no apoprotein formation could be observed

on the Western blot using antibody against the N-terminal Strep-tag (Figure 3.13). Another possibility is the instability that was introduced by the deletion of the N-terminus of the sequence led to the degradation of the polypeptide chain before the step of heme attachment took place.

3.2.5 Cytochrome c_1 and HCC_1S

The importance of the N-terminal region of mitochondrial cytochrome *c* for recognition by HCCS (this Chapter) has been shown using expression of the two proteins in *E. coli* for the maturation of the cytochrome *c*. As explained in Chapter 1, and further discussed in Chapter 6, the second type of mitochondrial cytochrome *c* - cytochrome c_1, is in some organisms processed by a specific synthase HCC_1S. Clearly, it is now necessary to show if it is sequence to the N-terminal side of the CXXCH motif that is also recognised in this case. One way to do this would be to express cytochrome c_1 and HCC_1S in *E. coli*. However, as has been stated in the Introduction to this chapter in Section 3.1, cytochrome c_1 is a membrane anchored protein with a complex N-terminal sequence, part of which is required for signalling and targeting the apoprotein to its final location in the mitochondrion. Hence this experiment is not straightforward. However, the structure of cytochrome c_1 shows that the bulk of the protein, including the immediate N-terminal region adopts a globular fold (Figure 3.4). Therefore it ought to be possible to truncate the gene in such a way that a hydrophilic and soluble version of cytochrome c_1 is made. An analogous production of a soluble cytochrome c_y from the bacterium *Rhodobacter capsulatus* with the membrane anchor helix removed could be taken as a precedent for this

type of experiment [136]. The sequence in *S. cerevisiae* cytochrome c_1 to be removed is shown in Figure 3.14 (A). An inserted N-terminal Strep-tag on the cytochrome c_1 gene and

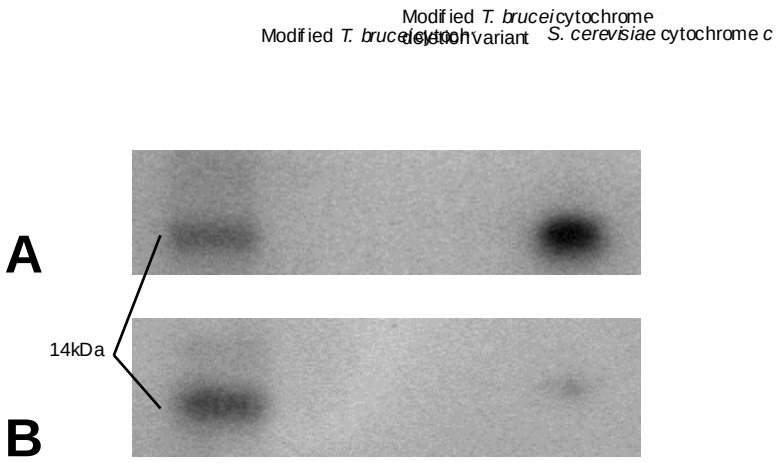


Figure 3.13 (A) Heme-stained SDS-PAGE analysis of cytoplasmic extracts of *E. coli* containing for the co-expression trial of modified *T. brucei* mitochondrial cytochrome *c* gene (and its variants) with *S. cerevisiae* HCCS gene, and *S. cerevisiae* cytochrome *c* as a control. Equal volumes of cytoplasmic extracts - diluted so as to contain equal amounts of material from wet cell mass - were loaded on each lane. The molecular markers are labelled as shown in the figures. **(B)** Western Blot of cytoplasmic extracts of *E. coli* containing for the co-expression trial of modified *T. brucei* mitochondrial cytochrome *c* gene (and its variants) with *S. cerevisiae* HCCS gene, and *S. cerevisiae* cytochrome *c* as a control. Anti-Strep antibody (IBA) was used in the blotting. The volumes of cytoplasmic extractions loaded in each lane are equal (diluted so as to contain equal amounts of material from wet cell mass).

a polyHistidine tag added on the HCC_1S gene enable Western blot analysis to be carried out.

Attempts to clone both a soluble version of cytochrome c_1 and HCC_1S, were unsuccessful in the sense that even though the PCR amplified gene constructs cut from plasmids pJAc1 and pJAhcc1s (from Dr J. W. Allen, Ferguson lab) and vectors were produced, the ligation

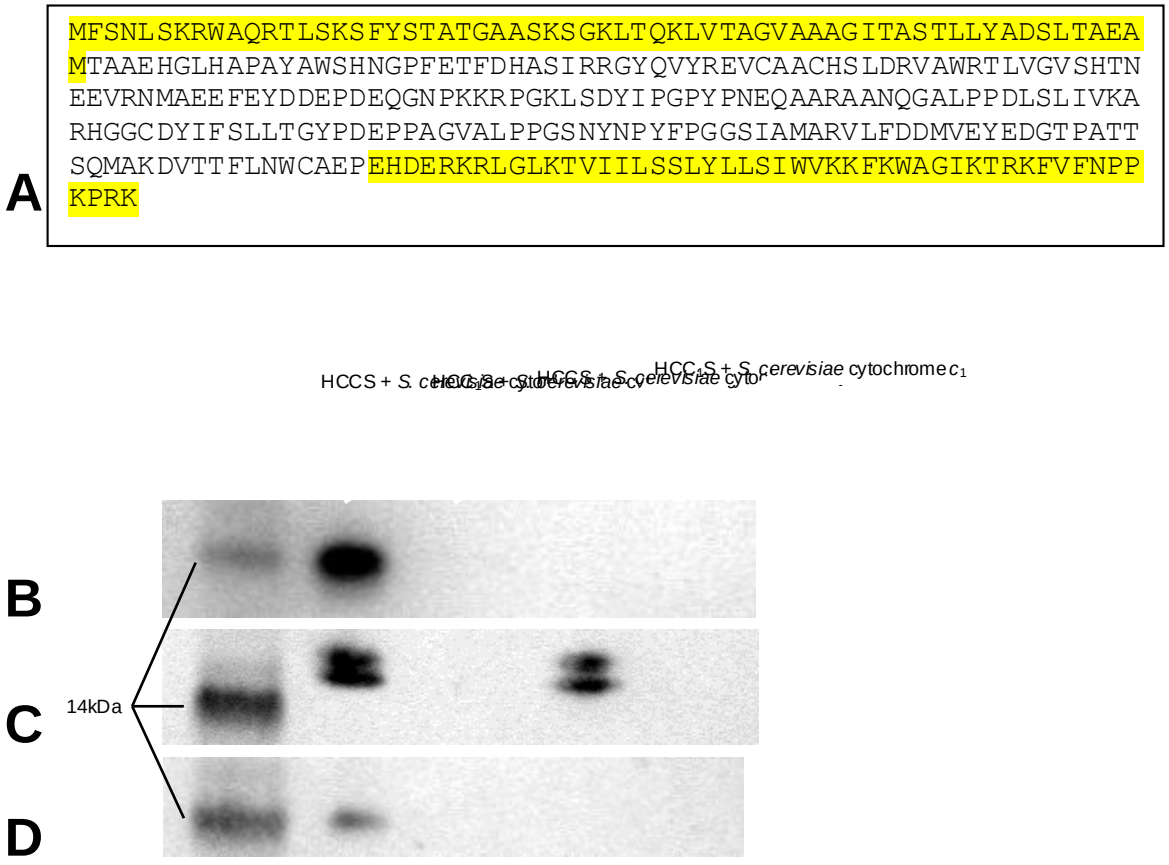


Figure 3.14 (A) The amino acid sequence of the modified *S. cerevisiae* cytochrome c_1 with the initial signal sequence and the transmembrane anchor removed (highlighted in yellow). **(B)** Heme-stained SDS-PAGE analysis of cytoplasmic extracts of cross expression of *S. cerevisiae* cytochrome c/c_1 genes and HCCS/HCC_1S genes. Equal volumes of cytoplasmic extracts - diluted so as to contain equal amounts of material from wet cell mass - were loaded on each lane. The molecular markers are labelled as shown in the figures. **(C)** Western Blot of cytoplasmic extracts of cross expression of *S. cerevisiae* cytochrome c/c_1 genes and HCCS/HCC_1S genes. An anti-polyHistidine antibody (Sigma) was used in the blotting. The volumes of cytoplasmic extractions loaded in each lane are equal (diluted so as to contain equal amounts of material from wet cell mass). The molecular markers are labelled as shown in the figures. **(D)** Western Blot of cytoplasmic extracts of cross expression of *S. cerevisiae* cytochrome c/c_1 genes and HCCS/HCC_1S genes. An anti-Strep antibody (IBA) was used in the blotting. The volumes of cytoplasmic

extractions loaded in each lane are equal (diluted so as to contain equal amounts of material from wet cell mass). Lane M shows the 14 kDa molecular marker. Lane 1 shows HCCS/yeast cytochrome *c* combination, lane 2 the HCC_1S/yeast cytochrome *c* combination, lane 3 the HCCS/yeast cytochrome c_1 combination, and lane 4 shows the HCC_1S/yeast cytochrome c_1 combination.

product, i.e. plasmids which contained the desired DNA sequences could not be obtained, and attempts of transformation of the ligation products tried in XL-2 Blue ultracompetent cells did not yield any colonies. Rather than spend too long in seeking explanations for these failures, plasmids containing chemically synthesised DNA sequences (sequences of genes plus tags) were purchased (Chapter 2, Section 2.2.5) and attempts at co-expression from the two different plasmids were tried in *E. coli* strain BL21 (DE3). Unfortunately no maturation of cytochrome c_1, or holocytochrome could be detected using either heme-staining, which would probe for the holoprotein, or with the Western blotting using anti Strep-tag antibody which would probe for the total amount of protein (Figure 3.14 (B) (D)). Similarly, no detectable production of HCC_1S was observed (Figure 3.14 (C)) as judged by Western blotting using antibody against the polyHistidine tag.

3.3 Discussion

3.3.1 Chimeric cytochromes

The *P. denitrificans* cytochrome c_{550} and mitochondrial cytochromes *c* share a common fold as shown in Figure 3.2 [126,127]. The work in this chapter provides persuasive

evidence that the part of the apocytochrome *c* that is recognised is the N-terminus. The fusion of the *S. cerevisiae* N-terminus to the *P. denitrificans* cytochrome c_{550} from the CXXCH motif onwards enables HCCS to produce a chimeric cytochrome *c*. Spectroscopic criteria, particularly the position of the α-band in the pyridine hemochrome spectrum at 550 nm are important. This indicates that the two vinyl groups of the heme are saturated, consistent with correct attachment to the heme via two thioether bonds, as observed in typical *c*-type cytochromes [63]. The importance of the XX residues within the CXXCH motif is limited, because as the heme-binding cysteines and histidine are largely conserved throughout *c*-type cytochromes in different organisms, the XX are different among them. For example, the XX residues are KA for the *P. denitrificans* protein and LQ for the *S. cerevisiae* protein. Yet the maturation apparently proceeds efficiently with KA as judged at least qualitatively by the similar amounts of chimeric protein and wild-type *S. cerevisiae* protein produced by the system. The level of residue conservation of XX in mitochondrial cytochromes *c* will be further discussed in Chapter 6.

The overall result is consistent with the older finding of Veloso *et al.* [134] who showed that heme was covalently attached by *S. cerevisiae* mitochondria to a synthesised peptide corresponding to the 26 N-terminal residues, including the CXXCH motif of horse heart cytochrome *c*. However, this older work localised the required recognition only to an extended version of N-terminal region including the CXXCH motif, and hence it was not possible to fully characterise the sequence that was produced with heme attached to the peptide. Nye and Scarpulla [137] presented evidence of heme attachment to a C-terminal fusion of yeast cytochrome *c*, to chloramphenicol acetyl transferase. Having acquired a globular fold the latter was retained on the cytosolic side of the outer mitochondrial

membrane with at least part of the cytochrome in the intermembrane space [137]; the cytochrome thus spanned the outer membrane with only the N-terminus accessible to the HCCS. This observation supports the evidence presented in this work that it is the N-terminal region that is critical for recognition by HCCS. It is also notable that HCCS effectively processes variants of *S. cerevisiae* cytochrome *c* in which the heme–iron axial methionine ligand is replaced by several residues, showing that HCCS activity does not require this C-terminal axial ligand [138]. Previously, it has been speculated that it is possible that the HCCS enzyme acts as a type of chaperone (similar to CcmE in System I), stabilising the cytochrome's overall structure to optimise the proximity of heme vinyl groups to the critical cysteines [139]. The results in this work evidently argue against this interpretation and support a model in which HCCS interacts with the N-terminus of unfolded apocytochrome and that acquisition of the final fold follows the heme attachment event. The extent of unfolding of the apocytochrome is unclear because previous work has shown that even mitochondrial apocytochrome *c* is able to acquire sufficient tertiary structure to bind heme non-covalently [90]. Thermophilic *c*-type cytochromes which were modified to contain non-covalent attachment with the heme, i.e. *b*-type cytochrome, also preserved the overall fold [28,29].

The same conclusion about the importance of the N-terminus can be demonstrated from experiments in which the N-terminal region of the trypanosome cytochrome *c* was replaced by the N-terminus from *S cerevisiae*. *T. brucei* cytochrome *c* is known to have little or no detectable maturation when processed by System III [40]. In this case, a mutation to generate the variation L21A resulted in slightly higher levels of cytochrome *c* synthesis than the chimera cytochrome *c* with the XX residues being LQ (same as the *S.*

cerevisiae sequence). In fact when substituting the LQ sequence in the *S. cerevisiae* CXXCH sequence by AQ, the resulting trypanosome sequence improved holocytochrome production again indicates that the XX is not important for HCCS function. One of the likely explanations for the increase in yield might be that the CAQCH sequence helps to stabilise the chimeric apoprotein, thus providing more substrate for HCCS.

During the course of the work presented in this thesis, several other studies were published which came to the same conclusion as to the importance of the N-terminal region for HCCS recognition [140]. One of these involved fusing the N-terminus of the *S. cerevisiae* cytochrome *c* to a protein domain known as PDZ, and the resultant protein had heme attached [141]. Other work [142,143] involved specific mutagenesis of the N-terminal region and that will be discussed in Chapter 4 in which residue-specific characterisation of the N-terminus will be presented.

3.3.2 Truncated mitochondrial cytochrome *c*

The *E. caballus* cytochrome *c* in the single-plasmid expression system by System I was not tagged, whereas the *S. cerevisiae* cytochrome *c* in the two-plasmid expression system has a Strep-tag on its N-terminus. Western blot was not carried out on these variants as the positive result of truncated *S. cerevisiae* cytochrome *c* in 3.2.2.2 already proves the point that truncated holocytochrome can be produced, the Western blot is not entirely necessary to be performed to demonstrate this.

Braun and Thöny-Meyer [54] reported successful heme attachment on synthesised short-length peptides. Detectable amounts of two truncated forms of *S. cerevisiae* cytochrome *c* were observed, but the shortest one (G29X) and the *E. caballus* truncation variants were not detectable. The shortest truncated fragment, G29X, was only expressed once in 20+ repeats (data not shown). The absence of reproducible heme attachment of the G29X variant is probably due to internal peptide degradation of this short polypeptide chain. The failure of maturation of the horse heart cytochrome *c* truncation variants is probably due to some subtle aspects relating to limitation of cross species co-expression. Another explanation is the higher proteolytic susceptibility to of truncated horse heart cytochrome in *E. coli*, leading to the degradation of the fragments, as shown in the staind bands probably representing the heme-attaching fragments in Figure 3.11. Although the fold and sequences are very similar, *E. caballus* cytochrome *c* differs from *S. cerevisiae* cytochrome *c* in some positions in the amino acid sequences.

The levels of holocytochrome *c* for the two longer fragments are significantly lower than the wild-type, probably due to the reduced stability, and thus the degradation of these N-terminal fragments of cytochrome *c*. The heme attachment to the truncated protein is consistent with the recent demonstration that heme was attached to an N-terminal fragment fused to a PDZ domain [141]. The truncation experiments also provide further evidence to the suggestion that the methionine residue M83, an axial ligand to the heme iron in holocytochrome *c*, is not required for HCCS activity. It has also been reported earlier that the M83A protein is readily produced by HCCS [138,144], certainly to a greater extent than the truncations.

3.3.3 *A. thaliana* and *T. brucei* variants

There seemed to be no particular reason as to why the two-plasmid system using *S. cerevisiae* HCCS and a gene for *A. thaliana* cytochrome *c* would not produce a holocytochrome *c* product, especially since the N-terminal region of the plant cytochrome *c* has some sequence similarity with the sequence stretch known to be important for HCCS action (which will be further discussed in Chapter 4). The failure to detect any product led to the hypothesis that the N-terminal extension in the plant protein compared with that *of S. cerevisiae* might be preventing maturation or expression, but attempts to delete and/or mutate this also resulted in failure to detect holoprotein. There was also a failure to detect the apoprotein under all circumstances. When the plant protein was directed to the periplasm and post-translationally modified by the Ccm system to allow attachment of heme then a small amount of product was observed. This observation shows that the plant gene could be transcribed and translated by *E. coli*, but seemingly not very efficiently.

This failure is possibly due to the failure of the optimised system to adapt to the specific and delicate production and maturation requirements for *A. thaliana* cytochrome *c*, as the periplasmically targeted *A. thaliana* cytochrome *c* was only produced in very low level in *E. coli* by the Ccm System. It is possible that the codon usage of plants differs significantly from that of *E. coli*. It is possible to have different triplet codons for the genetic code for one particular amino acid. For example, six different triplets codes for leucine: UUA,

UUG, CUU, CUC, CUA and CUG. In different organisms, the use of the degenerate genetic code, even for the same amino acid sequence, may differ. The compositions of available tRNA species may also differ in cells in different organisms. The data on the Kazusa DNA Research Institute website (http://www.kazusa.or.jp/codon/) shows the codon usage of *E. coli* has significant differences from that of *A. thaliana*, for several codons. During the finishing stage of my project, Moreno-Beltran *et al.* [145] reported the successful expression and maturation of *A. thaliana* cytochrome *c* with the cloning of the DNA sequence into a pET28a vector.

Similar failures for unknown reasons may surround the failure to produce and mature full length *T. brucei* cytochrome *c* with the HCCS, as even though the native AXXCH motif was changed to CXXCH. This result can be contrasted with the successful production of the chimeric protein containing the *S. cerevisiae* N-terminus and the *T. brucei* protein from the CXXCH motif to the C-terminus, discussed in Section 3.3.1. This observation argues against that the codon usage being a problem. Overall, it must be concluded that there is a feature in the *T. brucei* N-terminus that is incompatible with the action of HCCS. This subject is discussed further in Chapter 6.

3.3.4 Cytochrome c_1 and HCC_1S

The reason for the failure to detect any production of the water soluble recombinant cytochrome c_1 protein (in even the apo-form) is unknown. Very rapid and easy proteolysis and degradation could be a reason. With only negative results to hand, we cannot be

certain that the HCC_1S was expressed. As the expressions all involve genes chemically synthesised, there ought not to have been any problem with codon usage, as that was taken into account in the sequence design. There is a precedent for this membrane anchor removal for a membrane bound *c*-type cytochrome in *R. capsulatus*, as discussed in Section 3.2.4, but the latter is a bacterial protein instead of a mitochondrial protein, and there might be differences in folding and processing that are hindered in this case. As far as the cytochrome c_1 polypeptide is concerned, any future work might involve exploring a number of trials with different truncations, as the one chosen here might have been sub-optimal despite the structure based choice.

3.4 Perspectives

Despite the importance of mitochondrial cytochrome *c* in respiration and apoptosis, knowledge of how it is formed by a posttranslational modification reaction catalysed by holocytochrome *c* synthase has been lacking. In this chapter, the demonstration in this work that the yeast-bacterial and trypanosome-bacterial chimeric cytochromes *c*, as well as N-terminal truncated fragments of cytochrome *c* are recognised by HCCS, allows us to conclude that the major, and possibly only, recognition features between the apocytochrome and the synthase reside in the cytochrome N-terminal region, a conclusion that is in good agreement with a variety of earlier data in the literature [137] and [138]. The failure of expression and/or maturation of the *A. thaliana* cytochrome *c* and

recombinant *T. brucei* cytochrome *c* could be attributed to different factors, and it is possible that different strategies and/or expression systems applied would make a difference.

In the following chapter, the examination of the significance of specific residues and the spacing difference between bacterial protein and eukaryotic protein in the N-terminal region of mitochondrial cytochrome *c* in HCCS substrate recognition will be presented.

Chapter 4

N-terminal region of mitochondrial cytochrome *c* in the substrate recognition of HCCS

4.1 Introduction

In the Results and Discussion sections in Chapter 3, it has been established that it is the N-terminal region (excluding the X residues of the CXXCH motif) of apocytochrome *c* that is critical for the processing by HCCS. This information might be able to explain the findings of several previous studies of mutagenesis on the biosynthesis of holocytochrome *c* [108,128]. However, the identification of the importance of the N-terminal region enabled further characterisation and analysis of this part of the amino acid sequence. This chapter aims to investigate the importance of specific residues in the N-terminal region of mitochondrial cytochrome *c* for recognition of apocytochrome *c* by HCCS.

Allen [35] noted the existence of several residue numbering systems in the studies of mitochondrial cytochromes *c*. Given the removal of the initial methionine in the protein sequence in mitochondrial cytochromes *c*, sometimes the counting starts from the residue next to this methionine. In some other cases, counting starts from this methionine. For the consistency with previously published work and convenience, this thesis uses two numbering systems: yeast (*S. cerevisiae*) numbering and horse (*E. caballus*) numbering, with horse numbering using the methionine as residue 1, whereas the yeast numbering using from the residue next to this methionine as residue 1.

4.2 Results

4.2.1 Conserved residues in N-terminal region of cytochrome *c*

The N-terminal region of mitochondrial cytochromes *c* is a stretch of amino acid residues with many highly conserved residues. Figure 4.1 shows a sequence alignment of the N-terminal region of mitochondrial cytochrome *c* from 15 different organisms performed by

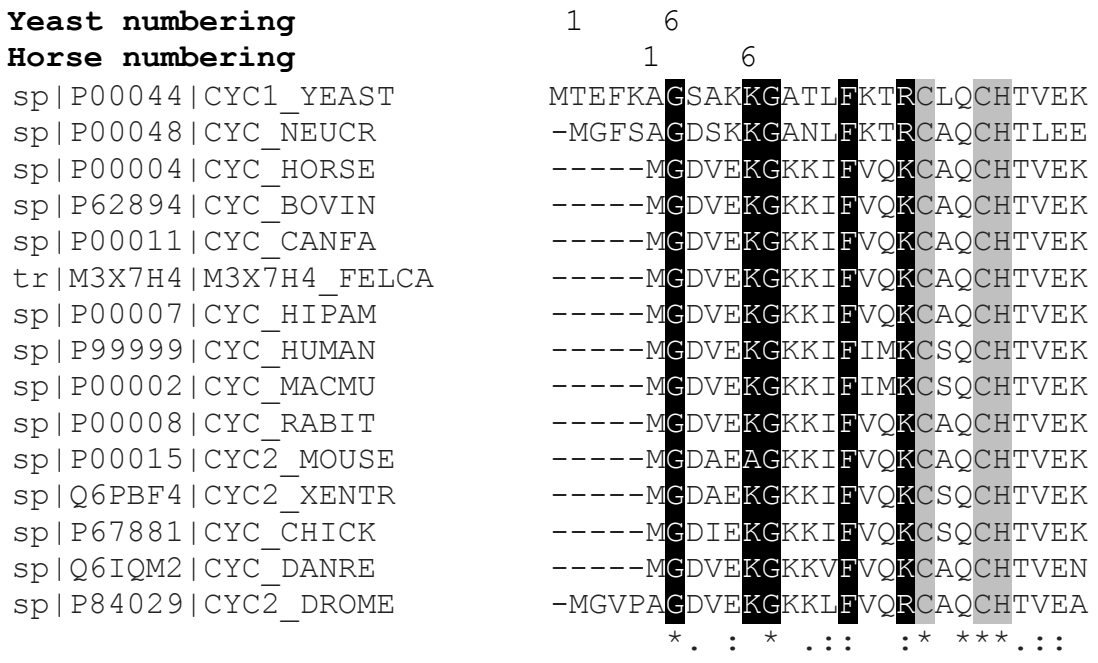


Figure 4.1 Multiple sequence alignment of the N-terminal region of mitochondrial cytochrome *c* in 16 eukaryotic organisms (animal and fungi). The organisms are: *Saccharomyces cerevisiae* (baker's yeast), *Neurospora crassa* (orange bread mold), *Equus caballus* (horse), *Bos Taurus* (bovine), *Canis familiaris* (dog), *Felis catus* (cat), *Hippopotamus amphibious* (Hippopotamus), *Homo sapiens* (human), *Macaca mulatta* (Rhesus monkey), *Oryctolagus cuniculus* (rabbit), *Mus musculus* (mouse), *Xenopus tropicalis* (Western clawed frog), *Gallus gallus* (chicken), *Danio rerio* (zebrafish) and *Drosophila melanogaster* (common fruit fly). The highly conserved cysteine and histidine residues in the CXXCH motif are depicted in grey boxes, and the other highly conserved residues chosen in this study are depicted in black boxes.

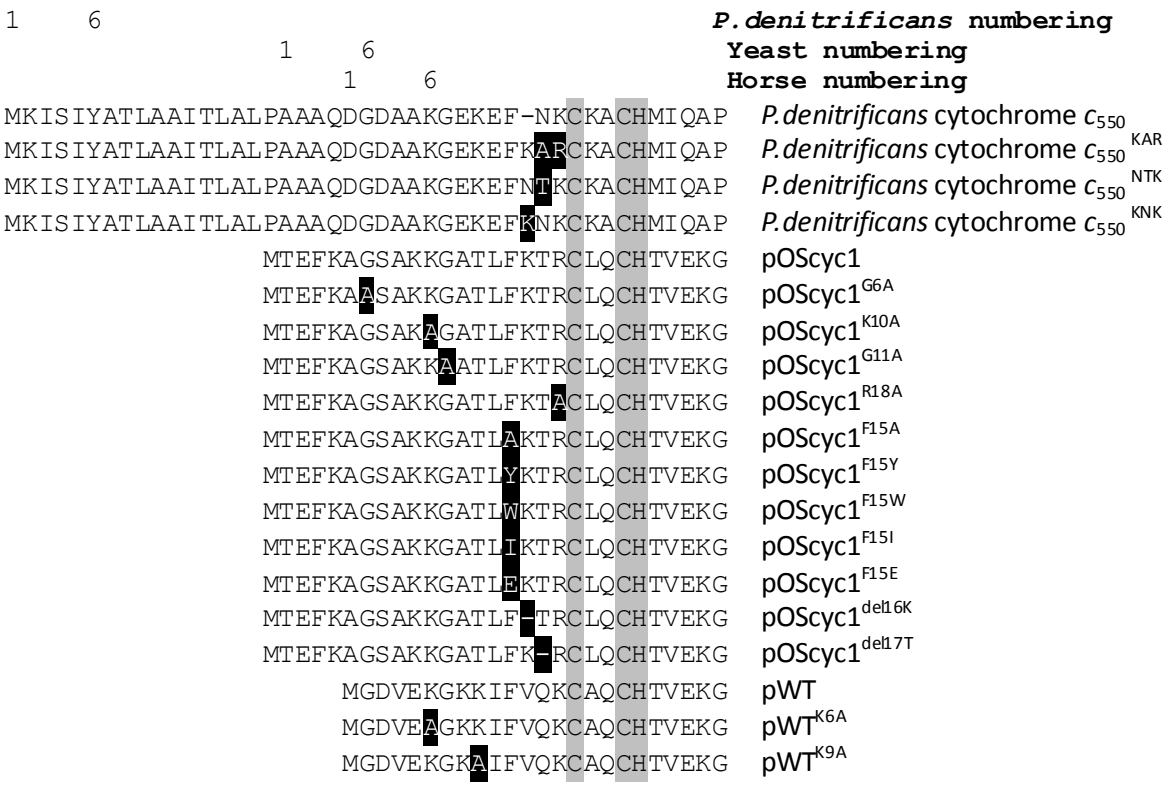


Figure 4.2 Multiple sequence alignment of N-terminal sequences of *S. cerevisiae* cytochrome *c* iso-1, *E. caballus* cytochrome *c* and their variants with *P.denitrificans* cytochrome c_{550} for comparison. The cysteines and histidines in the CXXCH motif are depicted in grey boxes, and the mutations are depicted in black boxes. KAR, NTR and KNR are *P.denitrificans* cytochrome c_{550} insertion variants; pOScyc1 del16K and del17T are deletion variants of S. cerevisiae cytochrome *c* (see Section 4.2.2).

ClustalW. Mutagenesis studies were carried out in order to narrow down the specificity elements in the N-terminal sequence of mitochondrial cytochrome *c*.

4.2.1.1 Conserved residues in *E. caballus* cytochrome *c*

In the published work, seven mutations were made in the horse cytochrome *c* gene in the single plasmid system [111] from which the *E. caballus* cytochrome is co-expressed with *S.*

cerevisiae HCCS [117] initially in collaboration with Dr Julie Stevens. The extent of sequence similarity in the conserved N-terminal regions of the horse and *S. cerevisiae* HCCS is 61%, which is consistent with the ability of the yeast HCCS to process horse cytochrome *c*. The overexpression of cytochrome *c* in this system would enable any perturbations or partial attenuations of holocytochrome *c* production to be identified. These residues are among the most highly conserved in the mitochondrial cytochromes N-terminal region sequence. Two out of these seven, K6A, and K9A (in Stevens *et al.* [117], a different numbering system was used) were made in the present work, expressed and analysed (N-terminal region sequence shown in Figure 4.2).

The cytoplasmic extracts of the *E. coli* cells expressing these variants were analysed by heme-staining on SDS-PAGE. Out of these seven variants, K6A and K9A of horse cytochrome *c* were matured at similar or even higher levels to the wild-type protein, and other variants has similar or higher maturation levels too, except G7A and F11A [117]. The results for K6A and K9A are shown in Figure 4.3. The bands for K6A and K9A, as well as the band for the wild type, are quite intense; it was difficult to distinguish the relative difference in their intensities. According to the intense red colour of the cell pellets (not shown) of these variants and the wild-type, the difference in maturation level should not be very significant. Nevertheless, it is possible that these two alanine mutations may lead to a potential increase in the maturation level of holocytochrome *c*.

The variants D3A and E5A were produced in very similar quantities to the wild type protein whereas G7A and F11A were not produced [117]. Of these results, that with K6A

was perhaps surprising, as this is a very well conserved residue among the mitochondrial cytochromes *c* (whereas K9A is not as well conserved as it). Although the efficient processing of apocytochrome *c* by *S. cerevisiae* HCCS is well established, a possible criticism of these results is that artefacts could arise from using proteins from two different organisms. Consequently, subsequent work was carried out studying the effects of variants of *S. cerevisiae* cytochrome *c* with the processing by *S. cerevisiae* HCCS.

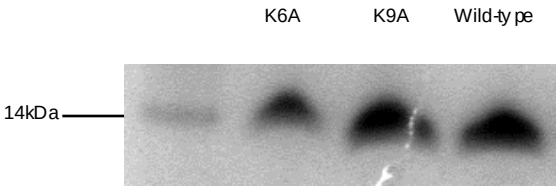


Figure 4.3 Heme-stained SDS-PAGE analysis of cytoplasmic extracts of *E. coli* containing wild-type and lysine 6 and lysine 9 variants of *E. caballus* cytochrome *c*. The volumes of cytoplasmic extraction loaded on each lane are equal and normalised according to wet cell mass. The molecular markers are labelled as shown in the figures.

4.2.1.2 Conserved residues in *S. cerevisiae* cytochrome *c*

4.2.1.2.1 Mutagenesis of five conserved residues

Following the work on the effect of mutations in the N-terminal region of *E. caballus* cytochrome *c* on the ability of *S. cerevisiae* HCCS to catalyse heme attachment, it was decided to continue the work using *S. cerevisiae* cytochrome *c*. This was to respond to the criticism that although *S. cerevisiae* HCCS clearly recognises *E. caballus* cytochrome *c*

successfully, there could have been some misleading results owing to the use of two proteins from different species.

Single residue variants of five conserved N-terminal residues G6A, K10A, G11A, F15A and R18A of *S. cerevisiae* iso-1 cytochrome *c* were made and tested for maturation by *S. cerevisiae* HCCS. The positions of these five residues are shown in the pymol-generated cartoon for the product holoprotein in Figure 4.4. The HCCS and cytochrome proteins

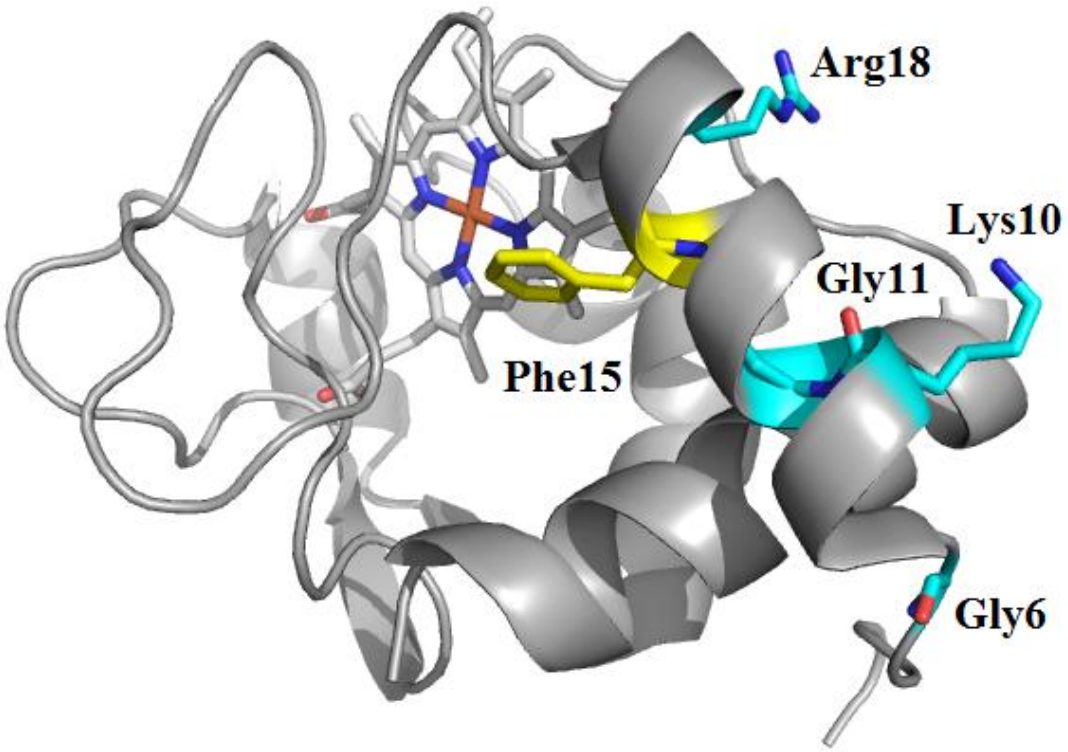


Figure 4.4 3-D rendering (Pymol) of the conserved amino acid residue positions in the N-terminal sequence of *S. cerevisiae* cytochrome *c* for mutagenesis studies. The positions are highlighted in cyan and yellow colours. This structure is for the holoprotein, and it is not know whether the N-terminal helix is found in the apoprotein.

were produced from separate plasmids in the *E. coli* cytoplasm (two-plasmid system). A heme-stained SDS-PAGE of the cytoplasmic extracts of cells expressing these variants is shown in Figure 4.6 (B). The F15A replacement (corresponding to F11 in the horse cytochrome *c*, shown in Figure 4.2) again caused complete loss of cytochrome *c* production, as seen in lane 5, compared with the WT in lane 1. However, the yeast G11A variant (lane 4, corresponding to horse G8A in 4.2.2.1) was matured, as other G11 variants are in yeast [128], implying that the context of this G is important for its interaction with HCCS. The continued production of cytochrome *c* in the R18A variant (lane 6) is notable as the residue next to the first C of the CXXCH heme-binding motif is always positively charged in mitochondrial cytochromes *c*. HCCS appears to be indifferent to the presence of this positive charge. The other two replacements, G6A corresponding to horse G4, and K10A corresponding to horse K7, did not abolish the maturation of the cytochrome by the HCCS [111]. The UV–vis spectra and pyridine hemochrome results in Figure 4.5 (A) (B) and Figure 4.6 (A) confirm the observation on the heme-stained SDS-PAGE, as shown in Figure 4.6 (B) – two vinyl groups of the heme are saturated and two thioether bonds are formed between the heme group and the polypeptide chain. A negative control was made and the spectra of the wild-type and the variants had the control subtracted.

4.2.1.2.2 Mutagenesis of the phenylalanine residue

Further exploration was carried out about what, if any, other residues, could substitute for the crucial F15 as key for recognition by HCCS. Four additional replacements of the phenylalanine in the N-terminal region of *S. cerevisiae* cytochrome *c* were made (F15Y, F15W, F15I and F15E). These residues were chosen because they represent a range of

amino acids with different characteristics: tyrosine and tryptophan are aromatic (and therefore similar to the properties of phenylalanine), isoleucine is non-polar/bulky, and glutamate is polar. The N-terminal sequences of these variants are shown in Figure 4.2. A F15A variant was also included in this analysis for comparison with these variants. The results of SDS–PAGE analysis and UV–vis absorption spectra in Figure 4.7 and Figure 4.8 show that the F15A and F15E replacements abolished holocytochrome *c* production completely, whereas the maturation of cytochrome *c* is attenuated, but not abolished, to different extents in both the F15W and F15I variants. To compare the levels of holocytochrome *c*, Figure 4.7 (B) and Figure 4.11 (A) provides the relative expression and maturation level among the variants and wild-type. The F15Y variant retained a high level of heme attachment and therefore also protein production. This implies the aromatic R-group of residue F15 is important in recognition by HCCS, as replacement of phenylalanine by both tryptophan and tyrosine had smaller impacts on the level of heme attachment than replacement with other residues including the hydrophobic isoleucine. This might indicate the importance of involvement of an aromatic residue in a π–π or π-cation interaction with HCCS. The wavelengths of the peaks in the α-band region of the pyridine hemochrome UV–vis absorption spectra of the variants were mostly at 550 nm as for the wild-type shown in Figure 4.9, indicating that the two heme-to-protein thioether bonds had formed in these cases. Only the wavelength of α peak in the spectrum of F15W was measured at 550.8~550.9 nm, which slightly deviates from 550 nm of the wild-type. However, the Western blot result in Figure 4.10 (A) shows that while F15A and F15E variants produce no holocytochrome *c*, there was a significant level of apocytochrome *c* produced in the cytoplasmic extracts for both variants. In contrast, the overall cytochrome *c* (both apo and holo) protein levels of F15W and F15I variants are lower than wild-type,

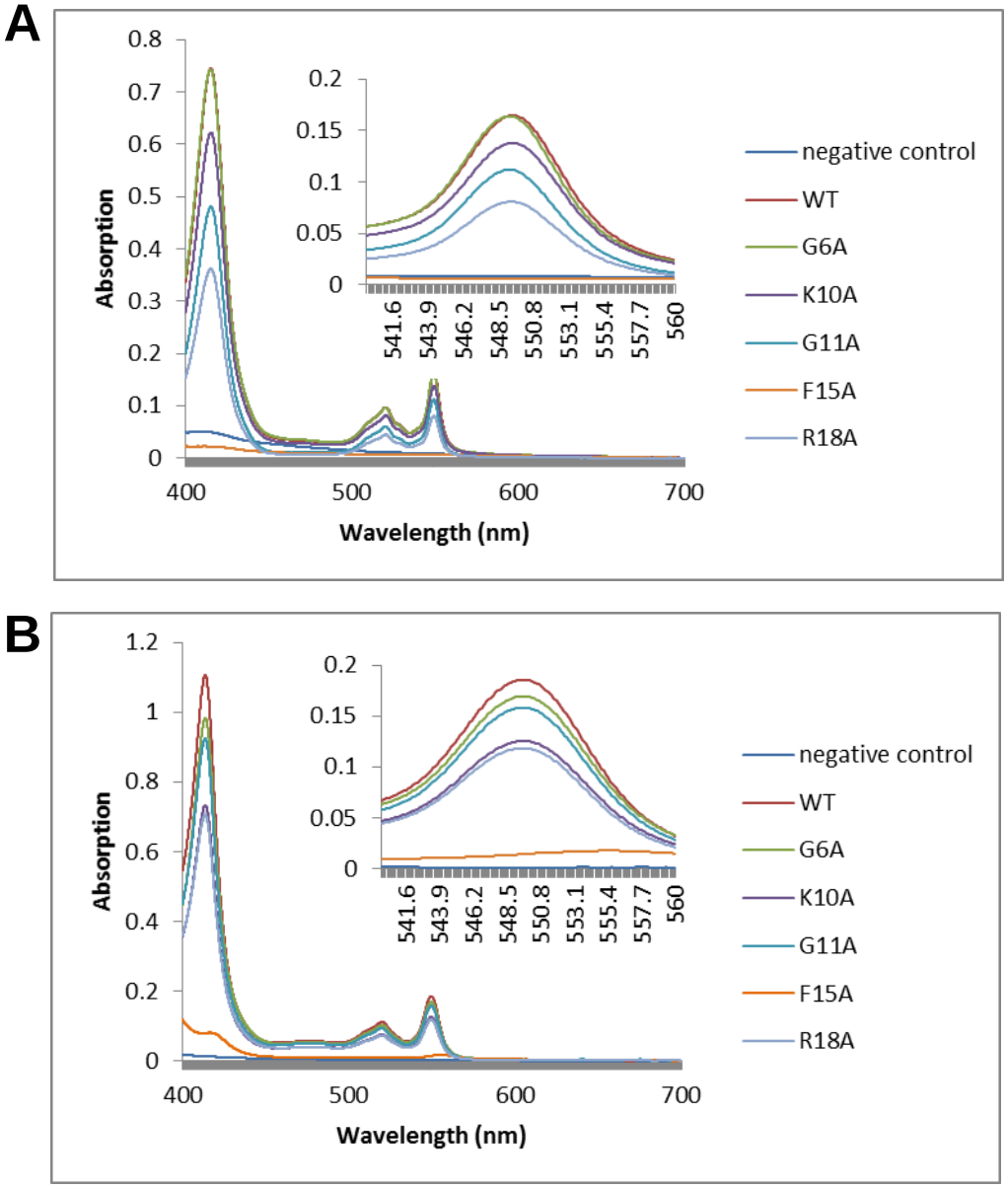
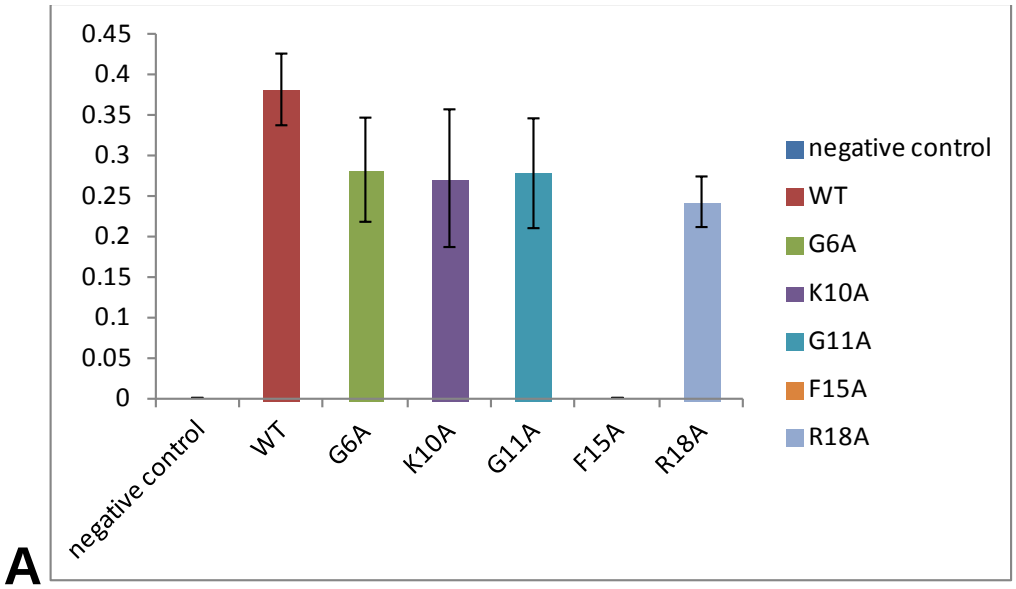


Figure 4.5 (A) UV-vis spectra of reduced cytoplasmic extracts of wild-type *S. cerevisiae* cytochrome *c*, N-terminal region variants and the negative control. **(B)** Pyridine hemochrome spectra of cytoplasmic extracts of wild-type *S. cerevisiae* cytochrome *c*, N-terminal region variants and the negative control. Both figures have the magnified spectra around 550nm wavelength inset.



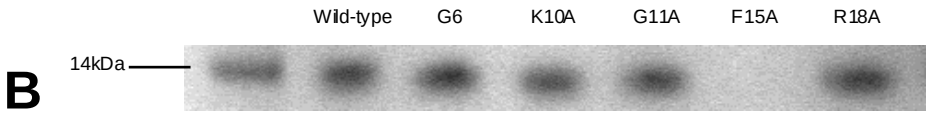


Figure 4.6 (A) Average absorption maxima of α peak of the UV-vis spectra of wild-type *S. cerevisiae* cytochrome *c*, N-terminal region variants and the negative control. **(B)** Heme-stained SDS-PAGE analysis of cytoplasmic extracts of *E. coli* coexpressing *S. cerevisiae* HCCS with wild-type *S. cerevisiae* cytochrome *c* and N-terminal region variants. Lane labelled M shows the 14 kDa molecular mass marker.

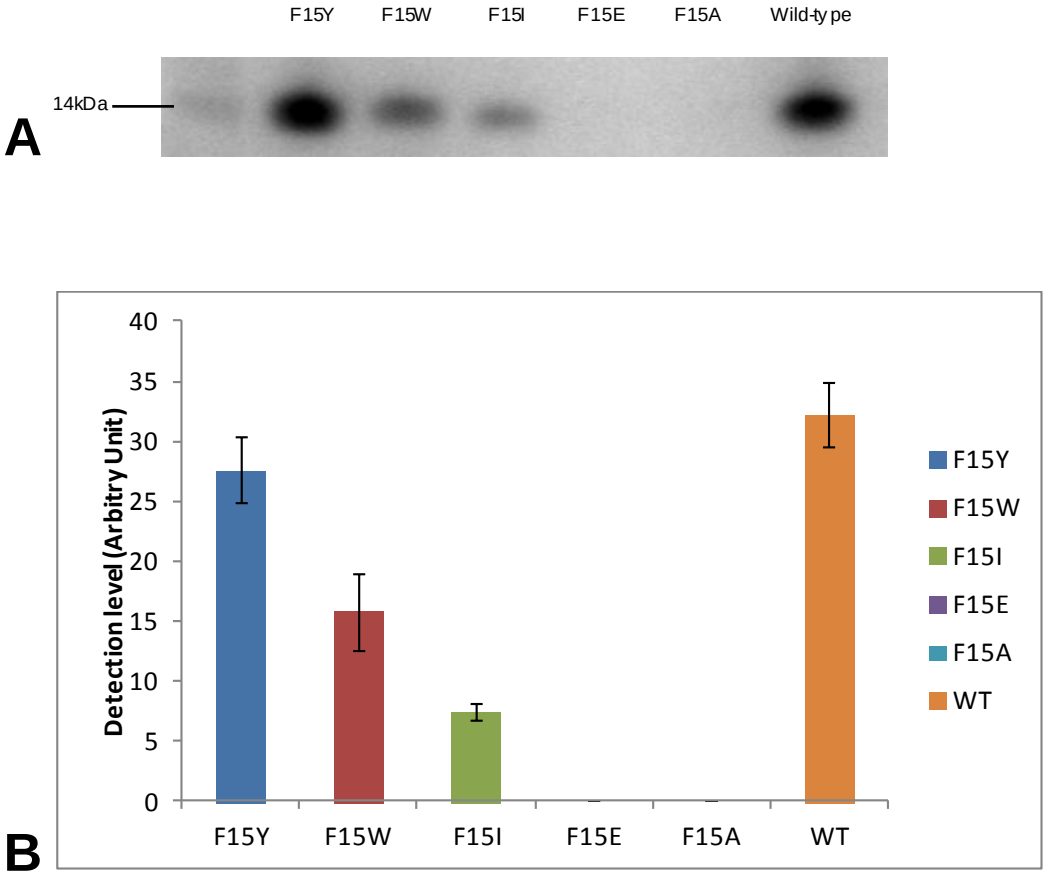


Figure 4.7 (A) Heme-stained SDS-PAGE analysis of cytoplasmic extracts of *E. coli* containing wild-type and phenylalanine 15 variants of *S. cerevisiae* cytochrome *c*. The volumes of cytoplasmic extraction loaded on each lane are equal and normalised according to wet cell mass. Lane M shows the 14 kDa molecular marker. **(B)** Average densitometry analysis recordings for the heme-staining on SDS-PAGE of cytoplasmic extracts of *E. coli* containing wild-type and phenylalanine 15 variants of *S. cerevisiae* cytochrome *c*.

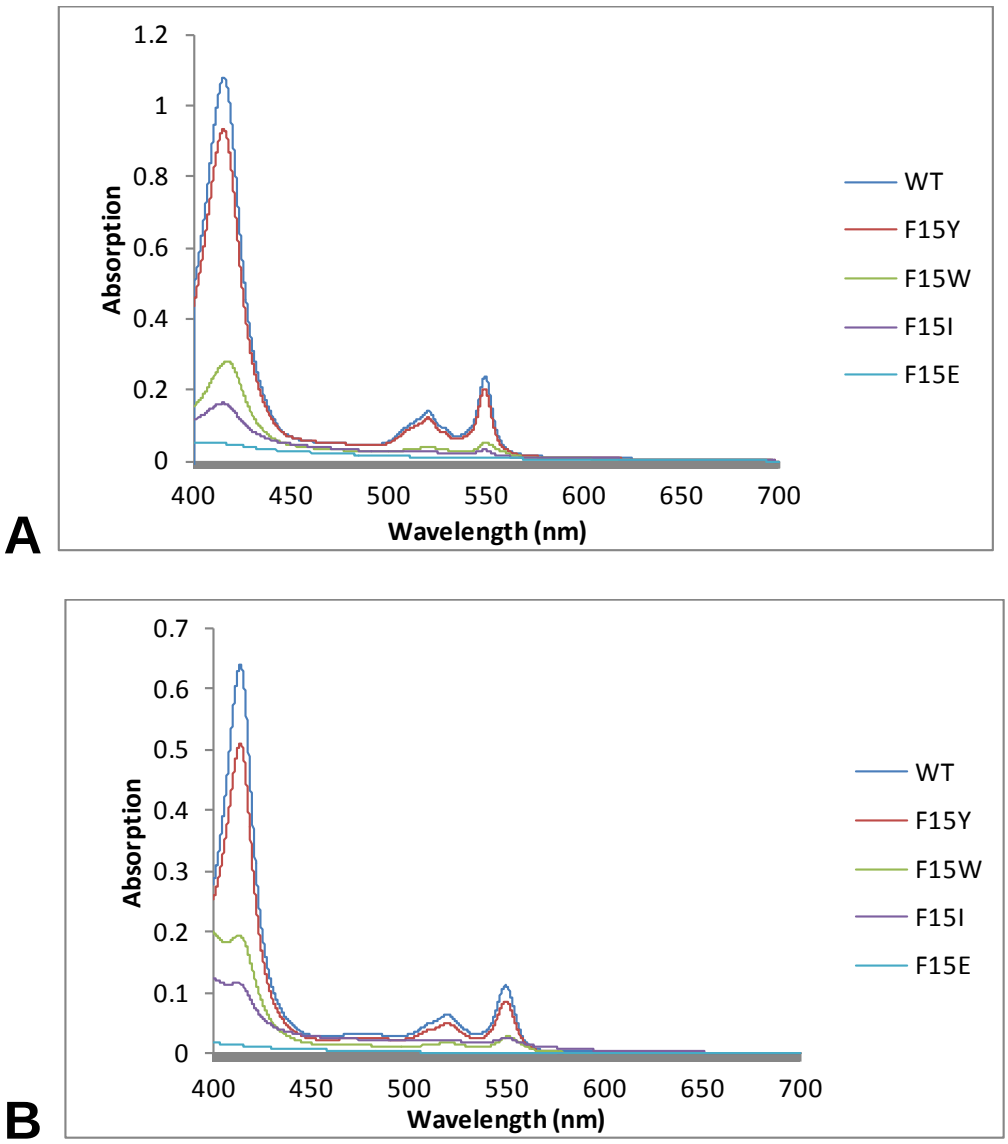


Figure 4.8 (A) UV-vis spectra of cytoplasmic extracts of wild-type and phenylalanine 15 variants of *S. cerevisiae* cytochrome *c*. **(B)** Pyridine hemochrome spectra of cytoplasmic extracts of *E. coli* containing wild-type and phenylalanine 15 variants of *S. cerevisiae* cytochrome *c*.

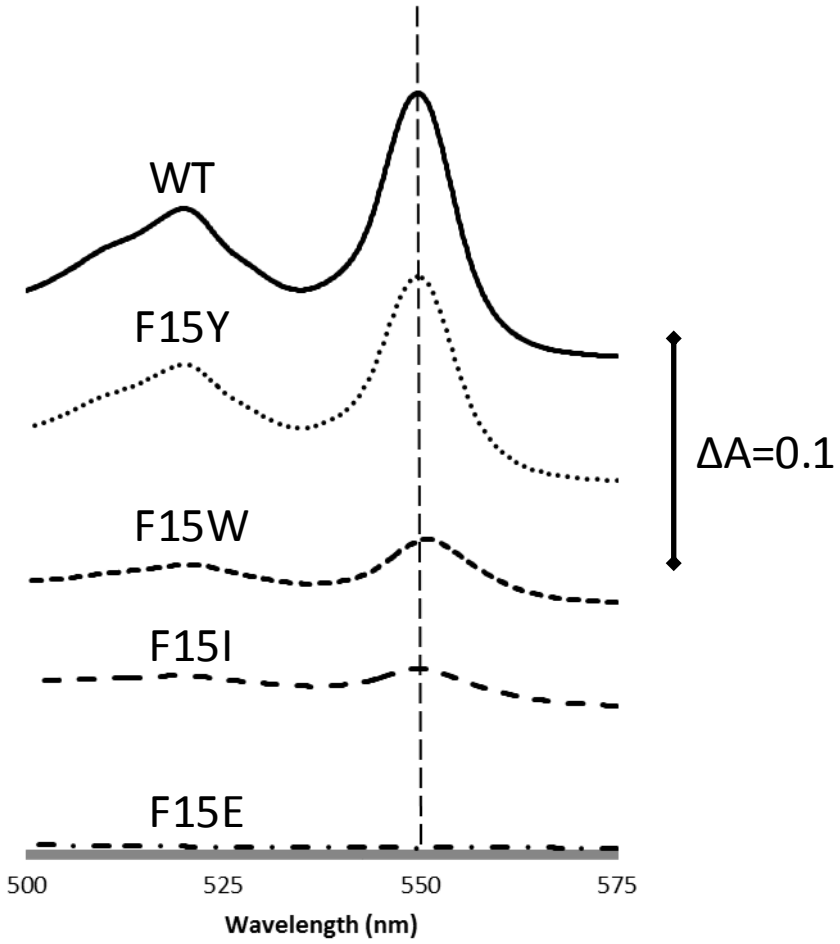


Figure 4.9 α-band region of pyridine hemochrome UV-vis absorption spectra of reduced cytoplasmic extracts of *E. coli* containing wild-type and phenylalanine 15 variants of *S. cerevisiae* cytochrome.

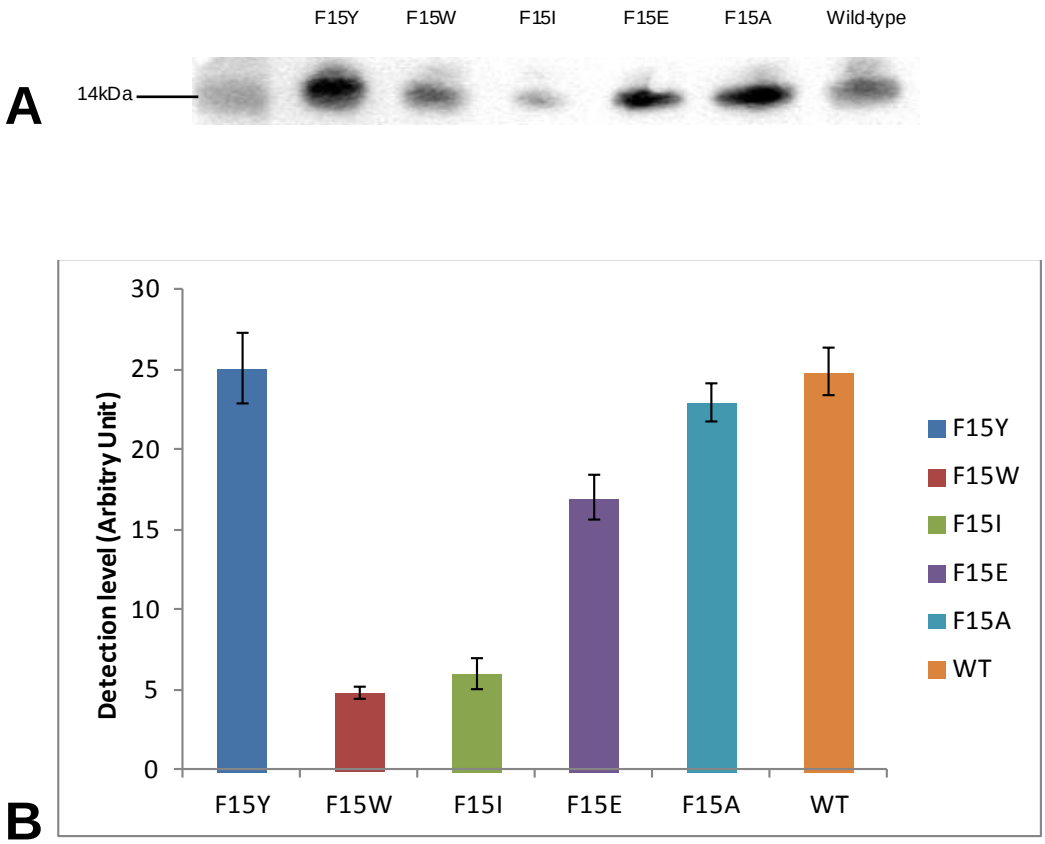
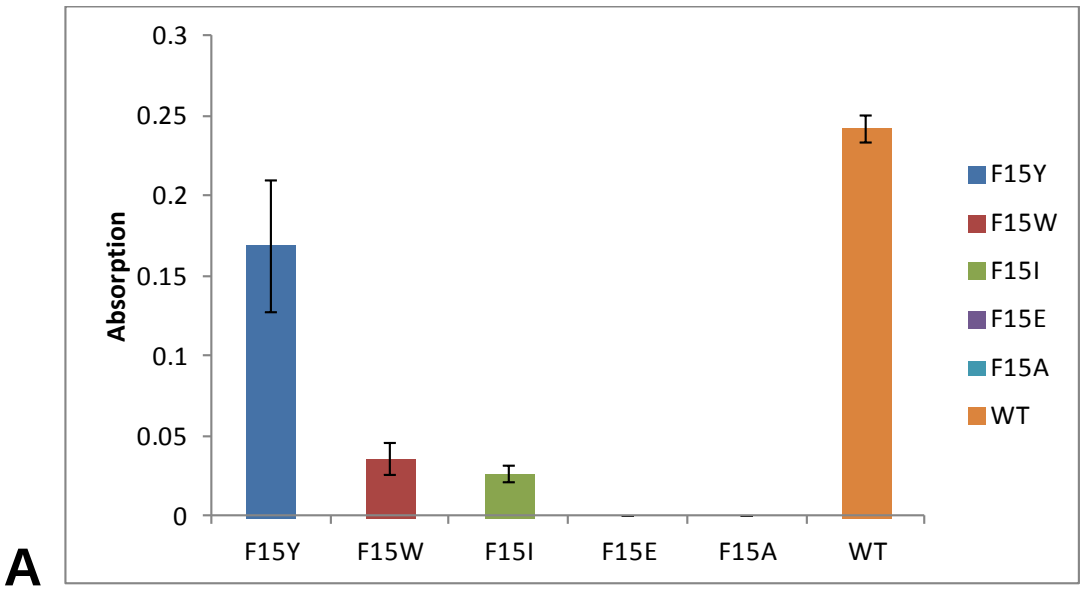


Figure 4.10 (A) Western blotting analysis of cytoplasmic extracts of *E. coli* containing wild-type and phenylalanine 15 variants of *S. cerevisiae* cytochrome *c*. An anti-Streptactin antibody was used in the blotting. The volumes of cytoplasmic extractions loaded in each lane are equal (diluted so as to contain equal amounts of material from wet cell mass). **(B)** Average densitometry analysis recordings for the Western blot of SDS-PAGE of cytoplasmic extracts of *E. coli* containing wild-type and phenylalanine 15 variants of *S. cerevisiae* cytochrome *c*.



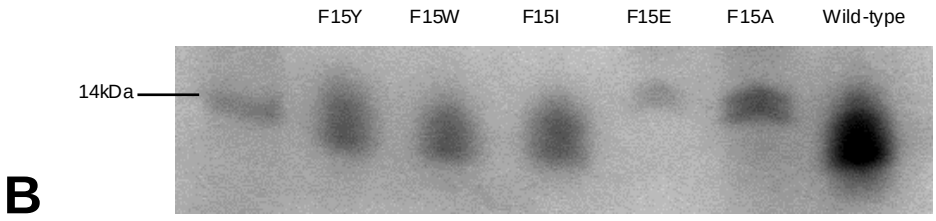


Figure 4.11 (A) Average absorption intensity of α peak of the UV-vis spectra of wild-type and phenylalanine 15 variants of *S. cerevisiae* cytochrome. **(B)** Heme-stained SDS-PAGE analysis of whole-cell extract of *E. coli* containing extracts of wild-type and phenylalanine 15 variants of *S. cerevisiae* cytochrome *c*. The volumes of cytoplasmic extractions loaded in each lane are equal (diluted so as to contain equal amounts of material from wet cell mass).

and F15Y. The overall protein level was measured using the Western blot technique with antibody against the N-terminal Strep-tag on the cytochromes *c*. The densitometry analysis result is shown in Figure 4.10 (B). One possibility for the abolition of the holocytochrome *c* production for F15E and F15A could be due to intrinsic structural disturbance in holocytochrome *c* introduced by the replacements, instead of the failure of recognition by HCCS. The observation of readily detectable levels of apocytochrome of the F15A and F15E variants in cytoplasmic extracts suggests that the lack of heme attachment to these two variants is caused by loss of recognition by HCCS, rather than structural instability of the proteins in either the apo or holo forms.

4.2.1.2.3 Periplasmically targeted phenylalanine variants with Ccm system

In order to further demonstrate the importance of the F15 residue in the context of recognition but not structural significance, a periplasmically targeted yeast cytochrome *c* was constructed and tested by co-expressing with the Ccm system encoded on plasmid pEC86. As introduced in Chapter 3, periplasmic production of a holocytochrome *c* using the Ccm system can in principle demonstrate that a variant form of cytochrome is a stable protein, and any failure to produce the same variant using HCCS can be attributed to the failure of HCCS to recognise the substrate apoprotein.

The 23-amino-acid long periplasmic targeting sequence from *P. aeruginosa* cytochrome $c_{551/552}$ (MKPYALLSLLATGTLLAQGAWAED) was inserted at the beginning of the sequence of Strep-tagged *S. cerevisiae* cytochrome *c*. Periplasmically targeted variants

F15A, F15Y, F15W, F15I and F15E were made from this construct. Despite efforts to optimise the procedure, periplasmic extraction of the wild-type and variants were incomplete as large amounts of periplasmically targeted cytochrome remained as contaminants in the cytoplasmic portion after the extraction. All periplasmically targeted variants expressed readily detectable quantities of holocytochrome *c*, but the level of holocytochrome *c* of F15E was low. Figure 4.11 (B) shows SDS–PAGE heme-staining analysis of whole-cell-lysis extracts for the five variants and wild-type cytochrome *c*. The periplasmically targeted F15I variant was matured to a similar level as the F15Y and F15W variants. The larger size of bands of F15Aperi and F15Eperi on the gel is attributed to incomplete cleavage of the targeting signal sequence as observed previously [12]. Both of them also show a much lower level of production of holocytochrome *c* than wild-type. Taken together with the detection of apoproteins of F15A and F15E, the successful formation of periplasmic holoprotein of F15A via the action of System I, is strong evidence that the requirement for a particular residue at position 15 is a consequence of the demands of HCCS and not the demands for structural integrity of holocytochrome *c*.

4.2.2 The spacing between CXXCH motif and the conserved phenylalanine in N-terminus

Amino acid R18 in *S. cerevisiae* cytochrome *c* is a conserved residue, and is directly adjacent to the CXXCH motif in the amino acid sequence on the N-terminal side. Its significance in HCCS substrate recognition was analysed in Section 4.2.1.2.1. The equivalent residue to this arginine in the N-terminal region in *P. denitrificans* c_{550} is a lysine which has an adjacent asparagine residue; however, in *S. cerevisiae* cytochrome *c*

there are three residues (KTR) between the phenylalanine F15 and CXXCH motif but in *P. denitrificans* cytochrome c_{550} just two residues (NK) lie between the F15 residue and the CXXCH motif. Therefore this difference in spacing might explain why *P. denitrificans* cytochrome c_{550} is not a substrate for HCCS. Two deletion variants of *S. cerevisiae* cytochrome *c*, as well as the *P. denitrificans* cytochrome c_{550} insertion variants were constructed and their N-terminal sequences are shown in Fig. 4.2. The N-terminal region sequences of wild-type *S. cerevisiae* cytochrome *c* and the *P. denitrificans* c_{550} are also included in the same figure for reference.

4.2.2.1 The deletion variants of *S. cerevisiae* cytochrome *c*

To test the importance of this spacing difference and its possible implication in the substrate recognition by HCCS, two variants del16K and del17T in yeast cytochrome *c* were made by deleting either lysine or threonine. Figure 4.12 shows the Western Blot for the analysis of the cytoplasmic extracts of cells expressing the deletion variants co-expressed with yeast HCCS, and shows the heme-stained SDS–PAGE. Figure 4.13 shows UV–vis absorption spectra. Neither of them had heme attachment completely abolished; however, the levels of holoprotein were significantly diminished. The total cytochrome *c* levels of the two deletion variants are also significantly lower than wild-type, according to the Western Blot in Figure 4.12 (B). From the result, it could also be deduced that the information about the degree of attachment from Figure 4.12 (A) and the total protein levels from Figure 4.12 (B) are relatively similar, although the total protein level of del16K is slightly lower than del17T, whereas the protein attachment level is the other way around. Based on the observations, the relative heme attachment level compared to protein

expression of the wild-type and the deletion variants are roughly comparable. Therefore, the relative maturation levels of the deletion variants and the level of the wild-type are similar too, which means that a reasonable amount of the cytochrome *c* deletion variants expressed had the heme covalently attached to them.

The pyridine hemochrome UV–vis spectra in Fig. 4.13 (A, inset) shows while the levels of the two deletion variants were significantly lower than that of wild-type, the wavelengths of the absorption maxima of the variants were shifted from 550 nm (wild-type) to 552.5 nm (α peak) which indicates a change of chemical environment around the heme within the cytochrome. The comparison of the maturation level is illustrated in Figure 4.13 (B). As UV-vis spectroscopy analysis is relatively more reliable than the empirical observation

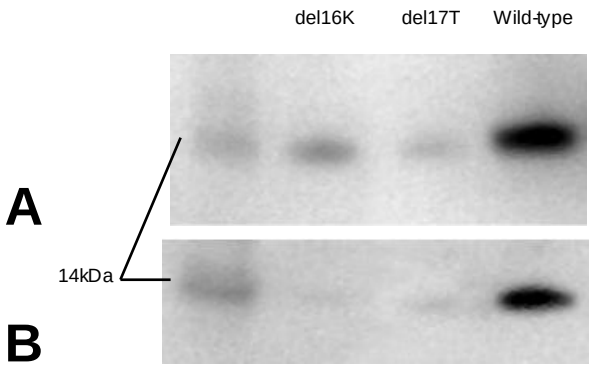


Figure 4.12 (A) Heme-stained SDS-PAGE analysis of cytoplasmic extracts of *E. coli* containing wild-type and N-terminal deletion variants of *S. cerevisiae* cytochrome *c*. Equal volumes of cytoplasmic extraction were loaded on each lane and normalised according to material from wet cell mass. **(B)** Western Blot of cytoplasmic extracts of wild-type and deletion variants of *E. coli* containing *S. cerevisiae* cytochrome *c*. The volumes of cytoplasmic extraction loaded on each lane are equal (diluted so as to contain equal amounts of material from wet cell mass).

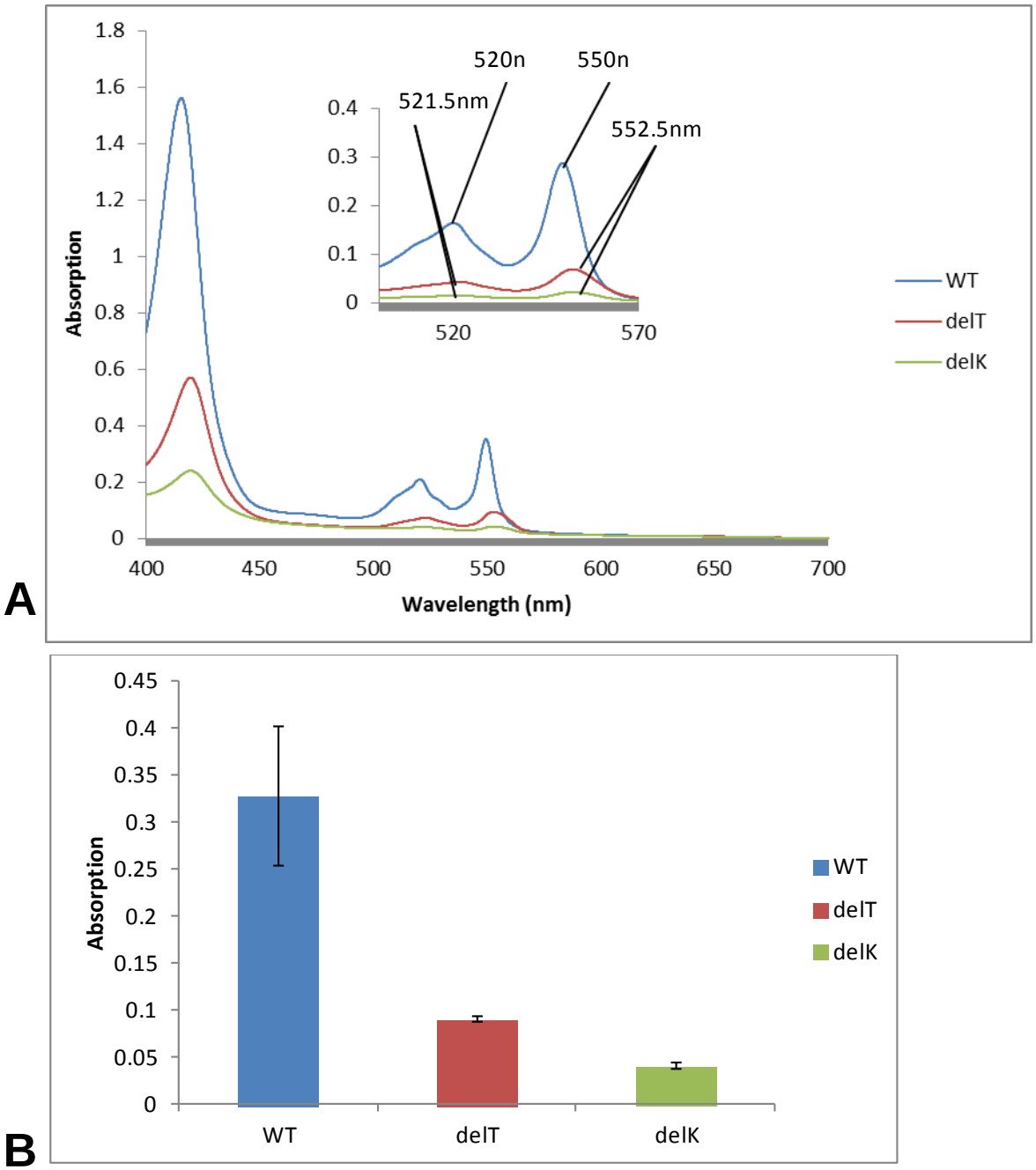


Figure 4.13 (A) UV-vis absorption spectra of cytoplasmic extracts of *E. coli* containing wild-type, delK16 and delT17 of *S. cerevisiae* cytochrome *c*. The reduced spectra are shown as the main figure and the inset shows the α-band region of the pyridine hemochrome spectra on an expanded scale. **(B)** Average absorption maxima of α peak of the UV-vis spectra of the wild-type *S. cerevisiae* cytochrome *c* and the del16K and del17T variants.

of the heme staining SDS-PAGE analysis, it can be demonstrated that both the heme attachment level and total protein expression of del17T is higher than that of del16K.

As it was necessary to examine the possible role of intrinsic instability in the variants, the same deletion variants were also made in the periplasmically targeted *S. cerevisiae* cytochrome *c*. These variants, along with the wild-type, were co-expressed with the Ccm system from pEC86 plasmid. In Figure 4.14 (B), the wavelengths of the peaks in the α-band region of the pyridine hemochrome UV–vis absorption spectra of the variants are same as the wild-type value at 550 nm. The heme-stained SDS–PAGE analysis of cytoplasmic extracts results shows that both of the variants are produced in considerable amounts by the Ccm system as shown in Figure 4.15. Thus cytochrome *c* seems capable of folding to a stable structure once heme is added to the deletion proteins by the Ccm system.

In summary, HCCS appears to have misattached heme to the two deletion mutants but when the same two proteins were expressed in the periplasm, the Ccm system processed them 'normally' as judged by both a pyridine hemochrome spectra and relative intensities on a heme-stained gel. Thus the cytochrome *c* itself does not appear to be compromised by the deletions but rather it is the action of HCCS on the apoprotein that is affected.

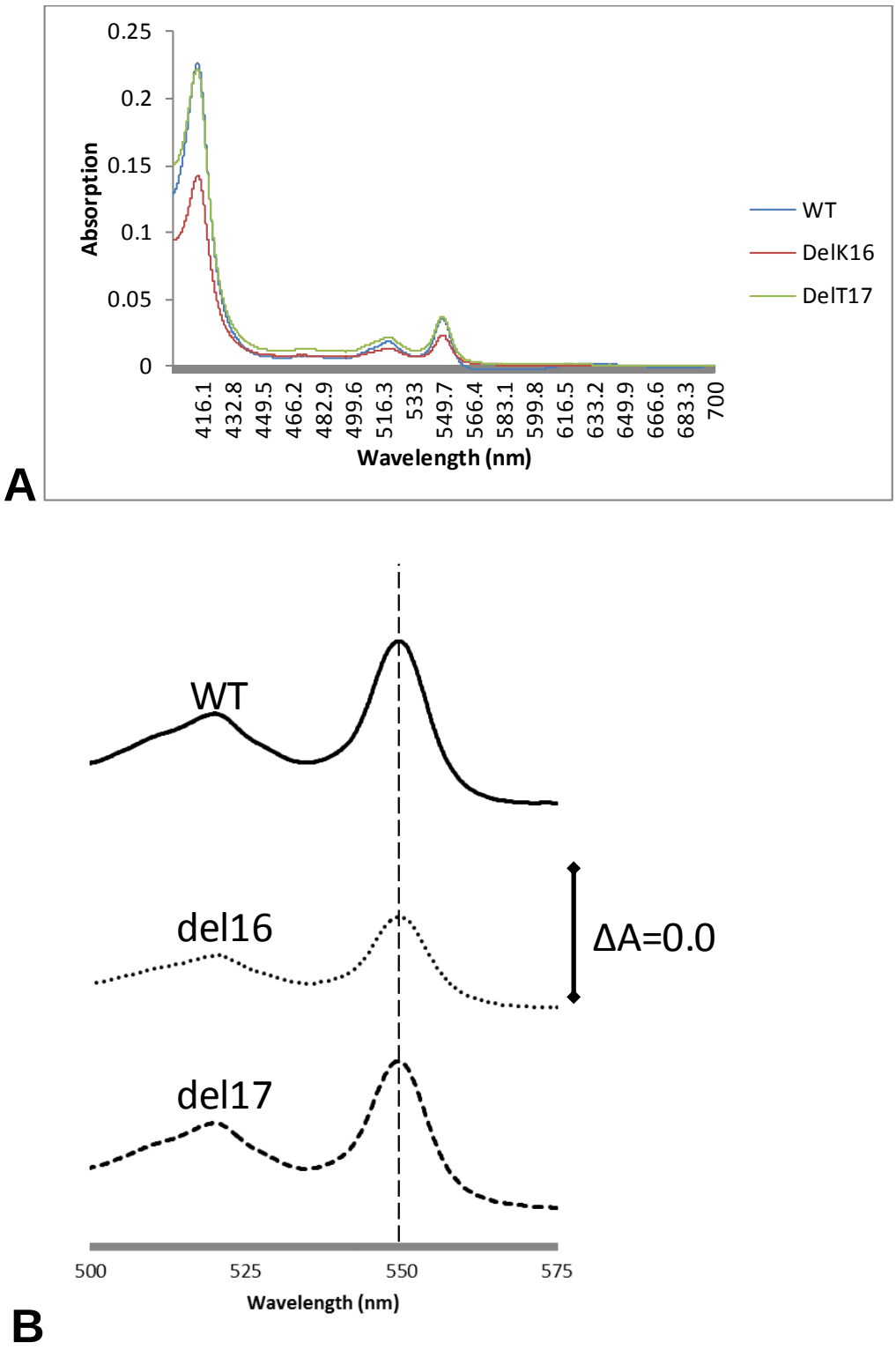


Figure 4.14 (A) UV-vis absorption spectra of the reduced periplasmic extracts of *E. coli* containing periplasmically targeted wild-type and deletion variants (del16K and del17T) of *S. cerevisiae* cytochrome. **(B)** α-band region of pyridine hemochrome UV-vis absorption spectra of reduced periplasmic extracts of *E. coli* containing periplasmically targeted wild-type and deletion variants of *S. cerevisiae* cytochrome.

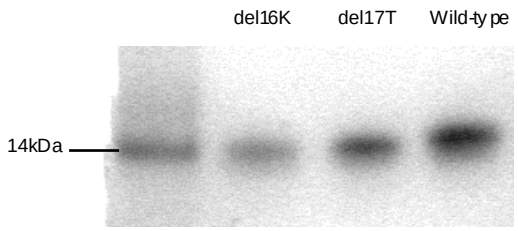


Figure 4.15 Heme-stained SDS-PAGE analysis of periplasmic extracts of *E. coli* containing wild-type and N-terminal deletion variants of periplasmically targeted *S. cerevisiae* cytochrome *c*. Equal amounts of total protein were loaded into each lane. Lane M shows the protein marker 14kDa.

4.2.2.2 The insertion variants of *P. denitrificans* cytochrome c_{550}

As mentioned in Chapter 3, HCCS is not able to mature *P. denitrificans* cytochrome c_{550} [110]. The spacing difference between the eukaryotic and bacterial cytochrome *c* was also addressed by reciprocal experiments in which an insertion was introduced into the *P. denitrificans* cytochrome c_{550}. This is in order to increase the spacing between the highly conserved phenylalanine and CXXCH from two residues to three residues. A truncated version of the *P. denitrificans* cytochrome c_{550} was obtained by the removal of the N-terminal signal sequence, and a Strep-tag (the same as for the yeast cytochrome *c* used in this study) was added. Three insertion variants, KAR, KTR and NKR were expressed with yeast HCCS in *E. coli*. Their sequences are shown in Figure 4.1. None of these three variants had detectable heme maturation according to heme-staining analysis on SDS-PAGE, which is shown in Figure 4.16 (A). From the Western blot result in Figure 4.16 (B) it can be observed that there was also no detectable expression of any apocytochrome in any of those variants.

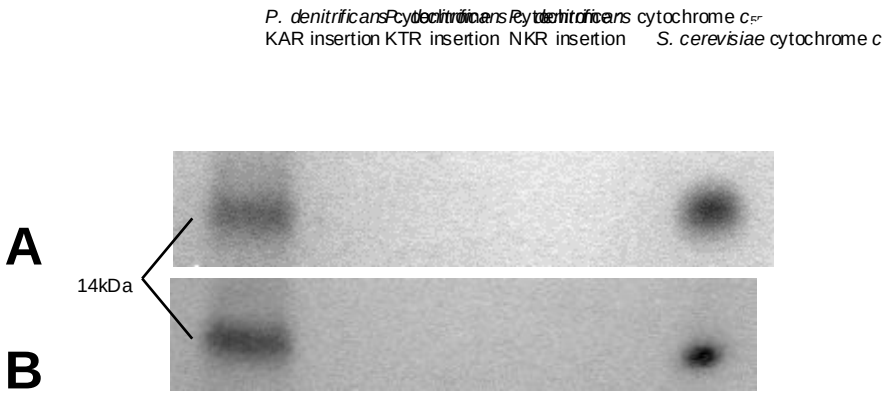


Figure 4.16 (A) Heme-stained SDS-PAGE analysis of cytoplasmic extracts of *E. coli* containing *P. denitrificans* cytochrome c_{550} variants and a *S. cerevisiae* cytochrome *c* as a positive control. Equal volumes of cytoplasmic extracts - diluted so as to contain equal amounts of material from wet cell mass - were loaded on each lane. **(B)** Western Blot of cytoplasmic extracts of cytoplasmic extracts of *E. coli* containing *P. denitrificans* cytochrome c_{550} variants and a *S. cerevisiae* cytochrome *c* as a positive control. The volumes of cytoplasmic extractions loaded in each lane are equal (diluted so as to contain equal amounts of material from wet cell mass).

4.2.3 Control experiments for apocytochrome *c* expression

One of the possibilities to be considered further is the expression of apocytochrome *c* in *E. coli*. The expression of HCCS within the two-plasmid system in *E. coli* has been analysed in Section 3.2.4 – HCCS can be overexpressed in *E. coli*. On top of the observation of large amount of expressed apocytochrome *c* in Figure 4.10 (A), the expression of apocytochrome *c* without HCCS in the cell should also be considered.

Either of the plasmids of pACcyc3 (expressing *S. cerevisiae* cytochrome *c*) and pOScyc1 (expressing *S. cerevisiae* HCCS) was transformed in *E. coli* and had the whole-cell extract analysed, along with the positive control of double transformation. The result is shown in Figure 4.17. It shows that the apocytochrome alone, lane marked pOScyc1, has been produced but not matured in the absence of HCCS in the cell, which is consistent with the observations of figure 4.10, i.e. the apocytochrome *c* is readily over-expressed in the two-plasmid system.

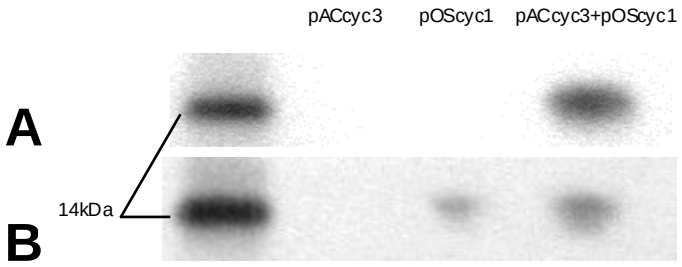


Figure 4.17 (A) Heme-stained SDS-PAGE analysis of whole-cell extracts of controls of expression of *S. cerevisiae* apocytochrome *c*. Equal amounts of total protein were loaded into each lane. Lane M shows the protein marker 14kDa. **(B)** Western blotting analysis (Strep-tag) of controls of whole-cell extracts of expression of *S. cerevisiae* apocytochrome *c*. Equal amounts of total protein were loaded into each lane. Lane M shows the protein marker 14kDa.

4.3 Discussion

4.3.1 Conserved residues in the N-terminal region of *S. cerevisiae* cytochrome *c*

It has been previously reported by Wang *et al.* [108] that F15 is possibly important for import and/or processing of mitochondrial cytochrome *c* in *S. cerevisiae*. From the result of this thesis, the evidence indicates that a disruption of apocytochrome-HCCS interaction in the absence of this side chain could explain the observation. It is worth noting that the side-chain of F15 (yeast numbering) / F11 (horse numbering) is relatively buried within the folded cytochrome *c* protein structure [146], but it could be exposed before heme attachment when the protein is still in the apoform without a fold. The importance of this highly conserved phenylalanine was also confirmed by Kleingardner and Bren [140] as they also reported that the F11A variant of horse heart cytochrome *c* could not be matured by yeast HCCS, but was assembled as a holoprotein in the periplasm. This paper was published after the completion of the work in this section. After the completion of work described in this section, the paper published by San Francisco *et al.* [142] reported the successful attachment of heme to the tyrosine variant of the phenylalanine in human cytochrome *c*.

As has been discussed above, within the N-terminus of mitochondrial cytochrome *c*, the importance of the phenylalanine residue F15 (yeast numbering) for maturation has been firmly established. The formation of stable and folded F15A (or F11A for the horse heart and human proteins) cytochrome *c* through periplasmic maturation using the Ccm system

has shown that, contrary to a long-held assumption, the conserved phenylalanine at position 11 is not crucial for the fold of the protein but rather for the action of HCCS. The consequences of substitution of F15 led to mixed results – F15Y variant retained a similar (80–90%) level, as judged by pyridine hemochrome analysis of holocytochrome production as the wild-type, whereas F15W and F15I variants had levels around 40% and 20% of that in the wild-type (Figure 4.10 (A)). The peak at the α-band region of the UV-vis spectrum of variant F15W is at 550.8~550.9 nm when the α-band region is magnified, indicating that the substitution of phenylalanine 15 to the more bulky tryptophan may probably induce a kink and steric hindrance to the pocket which heme binds. Even though the heme would be still bound by two thioether bonds to the polypeptide chain of the F15W variant, the distortion of the heme conformation may cause the shift of the absorption maximal of the peaks in the UV-vis spectra.

In an old study by random mutagenesis of residues that could substitute for the conserved glycine and phenylalanine residues in the N-terminus (equivalent to the G11 and F15 mentioned in Section 4.2.1.1) in *S. cerevisiae* cytochrome *c* iso-1, Auld and Pielak [129] found that serine, aspartic acid, valine and methionine mutations at the glycine position gave some functional cytochrome *c*, whereas leucine, phenylalanine and lysine mutations rendered the cytochrome *c* non-functional. At the conserved phenylalanine position, the leucine, isoleucine, tyrosine, tryptophan, valine, methionine and cysteine mutations were compatible with functional cytochrome *c*, but the variants with serine, arginine and glycine mutations were non-functional. They made the assumption that their results could be explained in terms of effects on the stability of cytochrome *c*; but from the results in this chapter, it is now clear that sequence changes at these two positions would have had an

effect on the action of HCCS. Thus it is established in this chapter that HCCS can tolerate tyrosine, tryptophan and to a small extent, isoleusine, at this highly conserved phenylalnine position, which is consistent with the results of Auld and Pielak [129]. Our periplasmic expression strategy would allow the effects of all these substitutions on stability alone of cytochrome *c* to be assessed. Thus the F15A variant which cannot be made via HCCS, but it could be made via the Ccm system and the stability assessed in other ways, for example, by determining the melting temperature.

One possible explanation for the nature of those residues substituting for phenylalanine is the aromatic nature of the R-groups of tyrosine (substantially) and tryptophan (moderately so) are recognised by HCCS and are compatible with a stable folded holocytochrome *c*. It should also be noted that in mitochondrial cytochromes c_1 there is also a highly conserved tyrosine in the equivalent position of the mitochondrial *c*-type cytochromes, and this feature is further discussed in Chapter 6. It was thought initially that there was no tyrosine found present at this conserved phenylalanine position for any known mitochondrial cytochrome *c*. However, following the completion of the experimental work in this chapter, a more detailed analysis of mitochondrial cytochrome *c* sequences was made. It was noted that one of the two cytochrome *c* sequences (UniProt entry P19974) in the well-known model organism *C. elegans*, designated as *cyc-2.1* (the other sequence is called *cyc-2.2*), actually has a tyrosine residue instead of the normal phenylalanine at postion 15 (Figure 4.17). Vincelli *et al.* [87] expressed this *cyc-2.1* gene using the Ccm system and found a 'normal' cytochrome *c*. Strangely, they were unable to express *cyc-2.2* via the Ccm system but were able to use *S. cerevisiae* HCCS to express it. From the results in this thesis, it is likely that they would have been able to express *cyc-2.1* via System III. Further

BLASTp analysis has shown that the Y for F substitution is found elsewhere amongst species in the phylum *Nematoda* (Chapter 6, Section 6.2.3).

```
CYC_HORSE        ---------------MGDVEKGKKIFVQKCAQCHTVEKG
CYC1_YEAST       ----MTE------FKAGSAKKGATLFKTRCLQCHTVEKG
CYC21_CAEEL      ----MSD------IPAGDYEKGKKVYKQRCLQCHVVD-S
CYC22_CAEEL      MGKKKSDTASGGAIPEGDNEKGKKIFKQRCEQCHVVN-S
                                *. :** .::  :* ***.*: .
```

Figure 4.18 Multiple sequence alignment of the N-terminal region of mitochondrial *S. cerevisiae* cytochrome *c*, *E. caballus* cytochrome *c* and. *C. elegans* cytochromes $c_{2.1}$ and $c_{2.2}$. The highly conserved cysteine and histidine residues in the CXXCH motif are depicted in grey boxes, and the highly conserved phenylalanine residues are depicted in black boxes. The tyrosine residue in is *C. elegans* cytochromes $c_{2.1}$ depicted in a yellow box.

Compared to the aromatic side-chains, the bulky isoleucine is less successful in imitating phenylalanine in tertiary structure than tyrosine or tryptophan. Nevertheless, it can be demonstrated that the phenylalanine is the preferred residue at this position, on the basis of experiments described in this thesis, and that can explain its conservation. Glutamate and alanine, with very different R-group characteristics from the wild-type phenylalanine, are not able to replace the aromatic phenylalanine, hence holocytochrome production was abolished. However, as the presence of high levels of accumulated apocytochromes F15E and F15A variants demonstrates, these apoforms were available as potential substrates for HCCS and thus the absence of cytochrome maturation in these variants may be explained by lack of substrate recognition by HCCS but not the instability of the protein itself (apocytochrome *c* is unfolded and subject to degradation within the cell). The finding that

the F15Y variant is well tolerated by HCCS from yeast is of interest in respect of the recent report of a similar finding for human HCCS [142]. In the latter case it was speculated that this might be an adaptation in the human enzyme to allow it to recognise cytochrome *c* and cytochrome c_1, the latter having Y at the corresponding position. However, our work shows that even though yeast has a dedicated enzyme for cytochrome c_1 assembly the HCCS can still tolerate Y in that position.

4.3.2 The spacing between CXXCH motif and the conserved phenylalanine in the N-terminus

It is found that deleting a lysine or a threonine residue between the highly conserved phenylalanine in the N-terminus and the CXXCH motif in the mitochondrial cytochrome, and thus making the spacing similar to the bacterial proteins, attenuated the production of holocytochrome *c* by HCCS but far from abolished it. The shift of wavelength of absorption peaks in the UV–vis spectra (550nm to 552nm) fits the pattern of absorption spectra of *c*-type cytochromes with a single thioether bond [147], hence the misattachment is most likely the formation of a single thioether bond rather than the usual two in the HCCS-processed proteins.

On the other hand, we had found that inserting an alanine in the *P. denitrificans* sequence between the F and the CXXCH did not result in successful recognition by HCCS, which agrees with the results of a similar experiment [142] using human HCCS and *R. capsulatus* cytochrome c_2. Apocytochromes c_2 of both *P. denitrificans* and *Rhodobacter capsulatus* each contain a phenylalanine residue, but at the equivalent of position 16; HCCS is not

able to recognise and mature these apoproteins [142,143]. The sole insertion of a threonine or a glutamine also results in no recognition by HCCS and maturation of apocytochrome. In fact, no detectable apocytochrome production could be found in these insertion variants. However, in two recent related studies that appeared after completion of the present work, recognition of a bacterial (*R. capsulatus*) cytochrome *c* by HCCS could be achieved either by a K11 insertion in combination with an E9L substitution [143] or insertion of an A together with mutations of two glutamate residues at the N-terminus (E8K_E10I_Ala) in *R. capsulatus* cytochrome c_2 [142]. In both the latter cases the spacing between F and CXXCH was increased to three residues but there are clearly additional factors at work than just the spacing difference in different cytochrome amino acid sequences. The shift in wavelength of the alpha peak of the spectra of the deletion variants of *S. cerevisiae* indicates a change of nature in heme binding to the CXXCH motif locally, due to the change in spacing between residues F15 and C19. The pyridine hemochrome spectra of the deletion variants are consistent with only one thioether bond being formed between CXXCH and heme by HCCS in these proteins. This would be because the CXXCH sequence is incorrectly aligned at the binding pocket of HCCS in the presence of the deletions between the phenylalanine and CXXCH.

The proposal that heme is attached by only one thioether bond in the two deletion variants, del16K and del17K needs to be confirmed by further work. Given that HCCS does attach heme to mitochondrial apocytochrome *c* possessing only one cysteine, then the plasmids encoding variant forms of cytochrome *c* with C15A or C18A changes alongside either of the two deletion variations could be constructed. Expression of such plasmids should show the absence of one or the other of the cysteines results in failure to attach heme. Other

approaches would be possible: N-terminal sequencing by the Edman degradation method, coupled with carboxymethylation of free cysteines, should also show which cysteine binds to the heme.

4.4 Perspectives

In this chapter, the vital role of the phenylalanine among five highly conserved residues in the N-terminal region of *S. cerevisiae* cytochrome *c* on HCCS substrate recognition and the inhibitory effect of varying this conserved phenylalanine were observed, but it seems that the substrate recognition by HCCS does not depend solely on the phenylalanine. The finding that a glycine is critical for heme attachment to the horse [117], but not the yeast protein indicates the overall organisation of the N-terminal region is important, as does the failure of the phenylalanine in the *P. denitrificans* c_{550} N-terminal region, at a different spacing relative to the CXXCH motif, to ensure heme attachment by HCCS. From the results of the spacing variants, it could be suggested that the phenylalanine residue and spacing in the N-terminal region of cytochrome *c* are both essential for optimal heme attachment performance by HCCS, although some heme attachment can still occur with a spacing of two; thus there are multiple deciding factors existing in the N-terminus for the substrate recognition of HCCS. In the following chapter, the investigation of the role of the heme-binding binding motif CXXCH is presented.

Chapter 5

The CXXCH motif in the substrate recognition by HCCS

5.1 Introduction

5.1.1 CXXCH motif in mitochondrial cytochrome *c*

The CXXCH motif is nearly universal in the family of *c*-type cytochromes. Located in the N-terminal region, the motif is the site of heme binding to the polypeptide. The cysteines bind to the ethene groups on the porphyrin ring, and the histidine ligates the iron ion in the centre of the ring, along with the distal methionine which is in the C-terminal region of the polypeptide chain of the cytochrome [35].

There have been some previous studies which have focused on the importance of the complete CXXCH motif in mitochondrial cytochromes *c*. There have been reports of both success [148,149] and failure [108] of heme attachment to SXXCH and CXXSH variants of yeast and human cytochrome *c*. A CXXCR variant of yeast cytochrome *c* was also reported to be able to be expressed by plasmid-based expression systems, but not by the native cytochrome biosynthetic system of yeast itself [109,150,151].

5.1.2 Variations of the CXXCH motif

Despite being highly conserved across a very broad range of species, including most organisms requiring oxygen respiration [7], the CXXCH motif in *c*-type cytochromes still has some rare variants found in particular types of species, which usually correlates with

differences in the mechanism of cytochrome *c* biogenesis. Studies of the handling of the CXXCH motif in mitochondrial cytochrome *c* might be advanced by assessing the ability of HCCS to process similar variants.

5.1.2.1 Multiple X variations

There are variations of the CXXCH heme binding motif in *c*-type cytochromes in many microorganisms. One form of variation is the difference in the number of residues between the two cysteines in the motif. Examples include a CX_3CH motif in a diheme cytochrome c_3 from *Desulfovibrio gigas* [152] and in a diheme cytochrome c_{552} from *Pseudomonas stutzeri* [153], a CX_4CH motif in a tetraheme cytochrome c_3 from *Desulfovibrio desulfuricans* [154,155], a $CX_{15}CH$ motif for one heme within an octaheme *c*-type cytochrome MccA from *W. succinogenes* [156].

5.1.2.2 AXXCH/FXXCH motif

In organisms belonging to the *Euglenozoa* phylum in the kingdom of *Excavata*, the mitochondrial cytochrome *c* and cytochrome c_1 contain an AXXCH or a FXXCH single-bond linked motif in place of the CXXCH motif, respectively [40,60,157-159]. The sequences are unique in these organisms, and no other known organisms have been reported to contain these motifs in their cytochromes. A mitochondrial cytochrome *c* biogenesis System V is expected to occur in these organisms [35].

5.1.2.3 Proximal ligand variations

Not only the positions of the cysteines in this motif can be varied, but the proximal ligand to the heme within this motif can sometimes also be different. A CXXCK motif in the NrfA protein from *W. succinogenes* [160] and a CXXCHXM motif in the NrfH protein from *D. vulgaris* [161]. Some of these rare cytochromes are supposed to be linked with System I and System II for heme attachment [91], as mentioned in Chapter 1. Other heme-coordinating residues include arginine, which presents in a WXXXR motif that interacts with a side chain of the heme group in cytochrome P450 [162].

Here, a series of CXXCH motif residue mutations which are based on the naturally existing variations mentioned above were made on horse heart cytochrome *c*. They were expressed with yeast HCCS in the single-plasmid system in order to examine the range of substrate recognition of the HCCS enzyme, specifically of the CXXCH motif in this case.

5.2 Results

5.2.1 Cysteine and histidine variants in CXXCH

In the CXXCH motif of the cytochrome *c* of *E. caballus*, the cysteine residues were substituted with alanine to produce variants C15A and C18A; the histidine residue was substituted with lysine, methionine or arginine producing the variants H19K, H19M and H19R (as shown in Figure 5.1). All of these five variants were co-expressed along with wild-type horse heart cytochrome *c* with *S.cerevisiae* HCCS and analysed by heme-stained SDS–PAGE, as well as UV-vis spectroscopy, as shown in Figure 5.2, Figure 5.3 and Figure 5.4.

5.2.1.1 Heme-stained SDS PAGE analysis

The C15A variant (AXXCH) alone was matured, as seen by a heme-staining band at the expected molecular mass of 14 kDa, but all the other variants did not undergo heme attachment, as shown in Figure 5.2. The gels were overloaded to maximise detection. The level of holocytochrome *c* for the C15A variant is lower than wild-type according to the intensity of the bands normalised for cell mass.

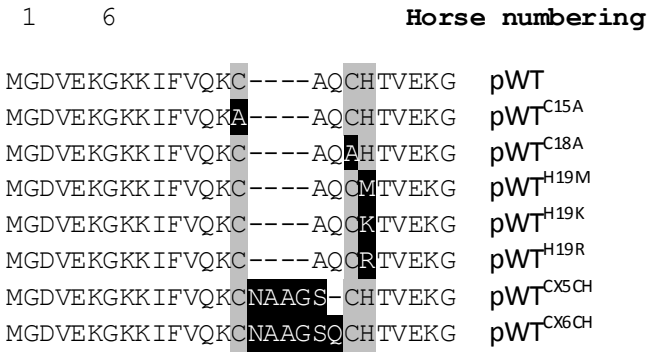


Figure 5.1 Multiple sequence alignment of N-terminal region sequences of *E. caballus* cytochrome *c* and its variants in the plasmids. The cysteines and histidines in the CXXCH motif are depicted in grey boxes, and the mutations are depicted in black boxes.

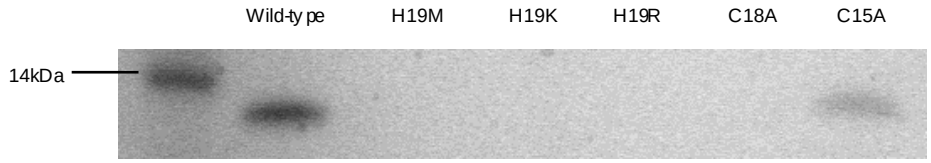


Figure 5.2 Heme-stained SDS-PAGE analysis of cytoplasmic extracts of *E. coli* containing wild-type and cysteine or histidine variants within the CXXCH motif of *E. caballus* cytochrome *c*. Equal volumes of cytoplasmic extracts - diluted so as to contain equal amounts of material from wet cell mass - were loaded on each lane.

5.2.1.2 UV-vis spectra

According to the UV–vis spectra, as shown in Figure 5.3 (A), the wavelengths of the absorption maxima of the C15A variant protein in the pyridine hemochrome spectra were

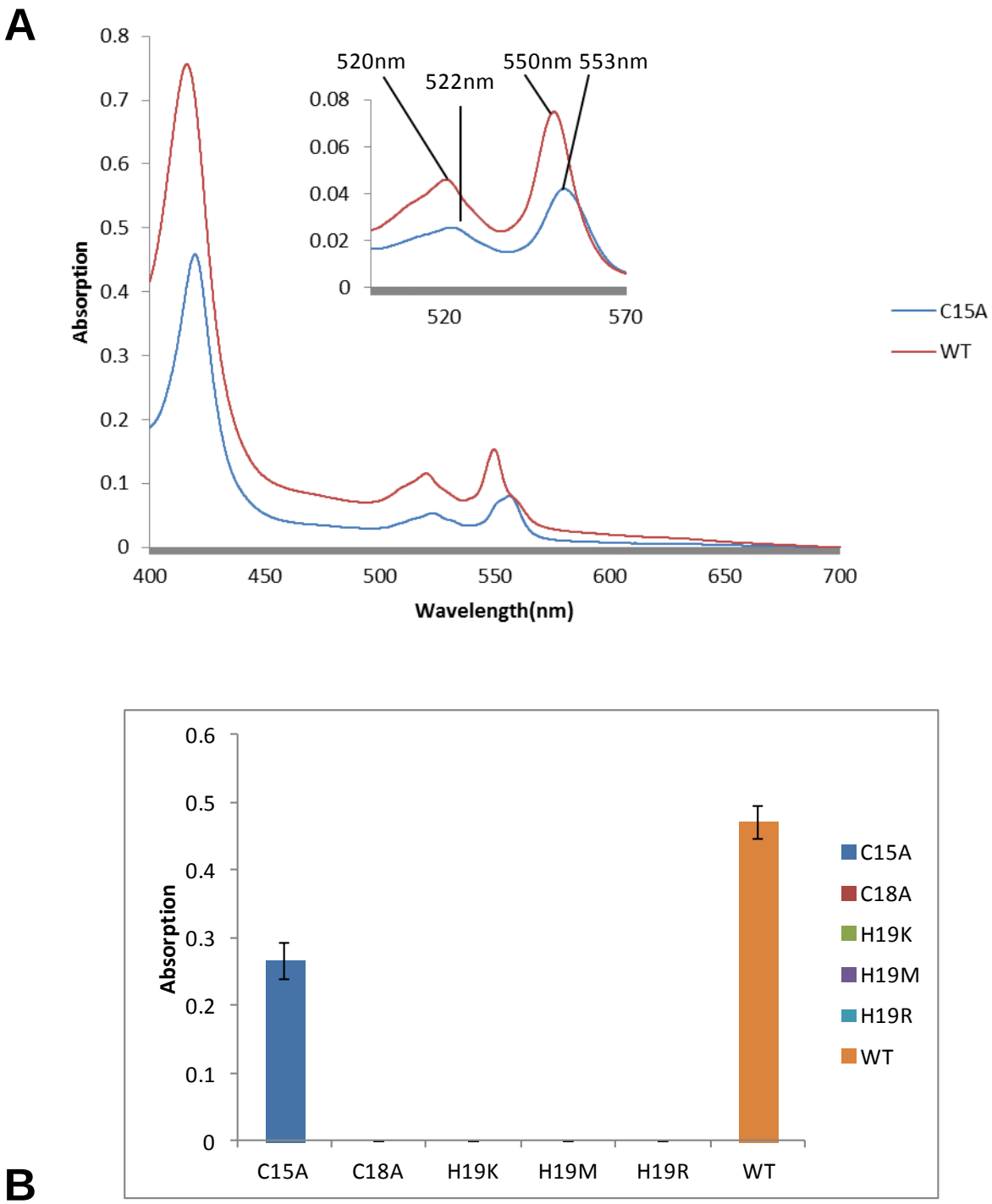


Figure 5.3 (A) UV-vis absorption spectra of similar volumes of cytoplasmic extracts of wild-type and C15A variant of *E. caballus* cytochrome *c*. The reduced spectra are shown as the main figure and the inset shows the α-band and β-band region of pyridine hemochrome on an expanded scale. **(B)** Average absorption intensity of α peak of the UV-vis spectra of the cysteine or histidine variants and the wild-type *E. caballus* cytochrome *c*.

shifted, i.e. the absorption maxima of the α peak shifted from 550 nm to 553 nm, as would be expected for a cytochrome *c* with a single bond between the heme and the polypeptide chain. From the average intensity of the α peak shown in Figure 5.3 (B) the level of maturation of holocytochrome *c* protein also dropped by approximately 40%. The heme-stained SDS-PAGE analysis and UV-vis spectroscopic analysis are consistent qualitatively, as both of them show a drop in the level of cytochrome *c* maturation by HCCS with the cysteine substitution.

5.2.2 CXnCH variants

Two variants with large numbers of X residues between the cysteines, i.e. CX_5CH (CNAAGSCH) and CX_6CH (CNAAGSQCH) motifs (Figure 5.1) were designed based on previous work [163], made and co-expressed with *S. cerevisiae* HCCS in *E. coli*. Neither of them was found to be matured by the HCCS enzyme, and as shown by the heme-stained gel in Figure 5.4.

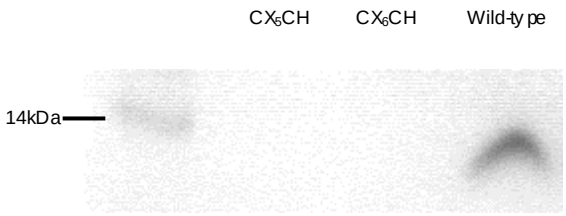


Figure 5.4 Heme-stained SDS-PAGE analysis of cytoplasmic extracts of *E. coli* containing wild-type and multiple X variants within the CXXCH motif of *E. caballus* cytochrome *c*. Equal volumes of cytoplasmic extracts - diluted so as to contain equal amounts of material from wet cell mass - were loaded on each lane.

5.3 Discussion

5.3.1 Cysteine and histidine variants of the CXXCH motif of *E. caballus* cytochrome *c*

The results confirm conclusions from several previous studies which showed that HCCS is able to attach heme to an AXXCH motif but no detectable heme attachment to a CXXAH motif [148]. The UV-vis absorption spectra of the AXXCH variant are consistent with a single thioether bond with a shift of wavelengths of absorption maxima of peaks in pyridine hemochrome spectra. The spectrum for the variant is also similar to that of the wild-type *c*-type cytochromes from *Trypanosoma* species that have a single thioether bond. Single bond SXXCH and CXXSH variants were reported to be produced in much lower levels than wild-type cytochrome *c* in both *S. cerevisiae* and *Drosophila melanogaster* [148,164]. After the completion of the work described in this section, San Francisco *et al.* [142] reported the successful heme attachment to single bond variants of both AXXCH and CXXAH of *H. sapiens* cytochrome *c*; however, in the case of the latter the heme attachment was very low. The absence of the second thioether bond in the latter type of protein has been shown to cause little perturbation to the overall structure of the cytochrome [116]. However, although from these experiments it is clear that HCCS can mature cytochrome *c* with one thioether bond to mimic the trypanosome *c*-type cytochromes, euglenozoid organisms clearly lack the HCCS system but contain a putative novel System V for cytochrome *c* maturation, with unknown components [35]. Interestingly, a recent study has shown that the cytochrome *c* maturation system in those organisms cannot produce an ordinary version of a CXXCH-containing cytochrome [147], thus adding to the evidence that a system distinct from HCCS is functional in that group of organisms and that HCCS needs the entire CXXCH motif for optimal performance. The

lack of covalent heme attachment to the H19M, H19K and H19R variants further demonstrates the significance of the histidine residue within the CXXCH motif. Thus, it seems that the HCCS system cannot process a CXXCM, CXXCR, or a CXXCK motif which presents in an NrfA nitrite reductase mentioned earlier. In conclusion, no cytochrome *c* biogenesis specific system, except specific variants of Systems I or II in certain species, will recognise any other residues except histidine. However, the evolutionary implication and the basis of this finding are still unknown.

After the completion of this work, Babbitt *et al.* [165] reported detection of heme attachment to a CXXCM motif. However, the attachment was at a very low level, approximately 1-5% of the level of the wild-type. If under our conditions, the yield of the CXXCM protein was 1%, then it may have evaded detection. On the other hand, that work [165] also failed to detect heme attachment to a CXXCR motif.

The 3-D crystal structure of *W. succinogenes* NrfA was solved and is available in the Protein data bank (PDB entry 3BNG). The sequence next to the CXXCK heme binding motif is shown in Figure 5.5. According to Figure 5.5, the CXXCH and AXXCH motif on mitochondrial cytochromes *c* from horse and trypanosome are both next to a strand of α-helix in the N-terminal region, with the highly conserved glycine G6 (horse numbering). In the case of the CXXCK motif in *W. succinogenes* NrfA shown in Figure 5.5, the N-terminal sequence preceding the CXXCK motif forms an α-helix but with a different conformation and the helix itself is extremely short. It is possible that the local environment presented in the *E. caballus* variants is significantly different from the wild-

type, and with the conformation of N-terminal region altered, the N-terminus fails to fit into the binding pocket on HCCS and HCCS cannot recognise the variant cytochrome *c*.

5.3.2 CX_nCH variants of *E. caballus* cytochrome *c*

The absence of heme attachment in the CX_5CH and CX_6CH suggests that it is possible that the longer sequence between the cysteines results in conformational change of the polypeptide chain, which probably hinders the protein-protein interaction between HCCS and the cytochrome polypeptide chain, hence blocks the substrate recognition.

Kleingardner and Bren [140] reported the heme maturation levels on some CX_nCH variants of *Hydrogenobacter thermophiles* cytochrome c_{552} and *E. caballus* cytochrome *c*. A CX_3CH variant (CAQGCH) and a CXCH variant (CACH) of *E. caballus* cytochrome were co-expressed with either *S. cerevisiae* HCCS in the cytoplasm *E. coli* or Ccm system in the periplasm of *E. coli*. The CXCH variant was not able to be matured by HCCS, but the CX_3CH variant could be matured. Both of them were successfully matured by Ccm System. Two variants of *H. thermophilus* cytochrome c_{552} with CAQGCH and CACH, along with a CX_4CH (CAMGACH) variant were also successfully produced on with Ccm system.

In another study, it was reported that CXCH, CX_5CH and CX_6CH variants of bacterial cytochrome *c*-b_{562} (an artificial *c*-type cytochrome synthesised by an insertion of a *c* heme

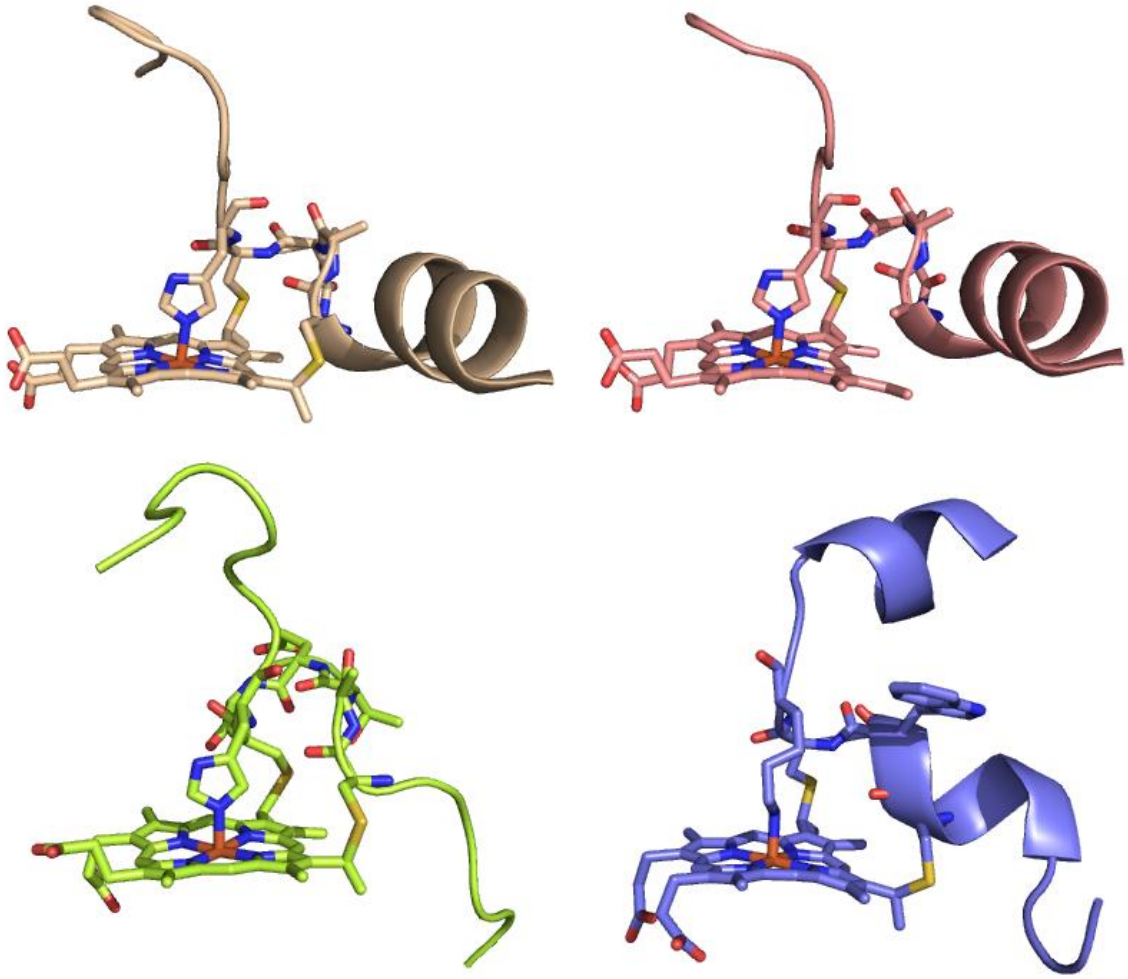


Figure 5.5 3-D rendering of different heme binding motifs of *c*-type cytochromes in different organisms: CXXCH in *Equus caballus* heart cytochrome *c* (Protein Data Bank code: 1HRC) in light orange; AXXCH in *Crithidia fasciculata* cytochrome *c* (Protein Data Bank code: 2YK3) in flesh tint; CXXXXCH in *Desulfovibrio africanus* cytochrome c_3 (Protein Data Bank code: 2BQ4) in grass green; CXXCK in *Wolinella succinogenes* NrfA (Protein Data Bank code: 3BNG).

binding motif into the *E. coli* cytochrome b_{562}) were able to be produced with Ccm system with heme attachment confirmed by high-accuracy mass spectrometry. The heme is also linked to the CX_nCH (n=1, 5, 6) motif by a persulfide bond on some occasions, which means an extra sulphur being incorporated into the linkage between one of the cysteines [163]. This persulfide bond was detected by mass spectroscopy based on the extra mass of this extra sulphur atom within the species containing this bond in the mixture. This aberrant linkage indicates that a chemical modification has occurred within the variant, possibly giving some insight about the mechanism of heme attachment in System I.

As part of the HCCS enzyme recognition process relies on the CXXCH motif, the length of the sequence between the two cysteines also has significance in this process. The consideration of the hindrance poised by long sequences (5 or 6 residues) is comparable to the reported failure of heme attachment to the CXCH motif, which has too short a space between the two cysteines to bind the two vinyl groups of the heme. The variants on cytochrome *c*-b_{562} have been tested on the Ccm system [163], and were found to be matured. Hence, the results here verify the narrower substrate specificity of HCCS when compared to the Ccm system.

Similar to the case of the CXXCK motif in NrfA, the amino acid sequence preceding the CX_nCH motif in multiheme cytochromes, such as the example of the CX_4CH motif in *Desulfovibrio africanus* cytochrome c_3 (Figure 5.5) also forms a drastic different local environment from the mitochondrial cytochromes *c*. The tertiary structure for the N-terminal sequence is a coil instead of the helical structure in mitochondrial cytochromes *c*.

Moreover, the extra two X residues between the two cysteines makes these four residues form a loop between the thioether bonds to the heme. Similar to the histidine variants, this type of unique environment around the heme group potentially causes the distortion of N-terminal region of the horse heart cytochrome *c* variants.

As described in previous chapters, mitochondrial apocytochrome *c* is relatively unstructured before the attachment of heme, and it is the apoprotein that is recognised by HCCS. It has to be admitted that there is no evidence concerning the relative levels of the various apocytochromes *c* that were tested as possible substrates for HCCS in this chapter. However, it seems very unlikely that the changes to CXXCH of AXXCH, CXXAH, CXXCM and other variations would cause substantial changes to the stability of the apoproteins. It is notable that in the study of Babbitt *et al.* [165], for the production of a CXXCM variant using a variant form of human HCCS, it was not investigated whether the low yields were a consequence of low levels of apoprotein rather than the low catalytic efficiency of the HCCS towards the CXXCM variant. It is perhaps more probable that alerting the spacing between the two cysteines could alter the stability of the apoprotein as conceivable that could induce a change in local conformation even in the apoprotein. Whilst such a factor might account for the failure of HCCS to attach heme, it could also alter the stability of the apoprotein. Variation in the stability of different apocytochromes *c* could be investigated by repeating the work with a suitable tag (such as Step-tag), such that an antibody could be used to detect the apoprotein and give some qualitative estimate of the amounts of protein produced.

5.4 Perspectives

The significance of the cysteine and histidine residues in the CXXCH motif was demonstrated in this chapter, in agreement with older works. In System III, the cysteine residue (s) may show flexibility to some extent; however, the histidine is generally required for the heme attachment by HCCS, except some rare cases. In the following chapter, a broader perspective is taken for the sequence analysis for the N-terminal region of cytochrome *c* from different organisms, focusing on species in plants, trypanosomes and roundworms.

Chapter 6

Sequence analysis of plant, trypanosome and nematode cytochromes *c*

6.1 Introduction

As has already been discussed extensively in Chapter 4, it is firmly established that the near-universal phenylalanine in the N-terminal sequences of the *c*-type mitochondrial cytochromes performs a very important role in the substrate recognition by HCCS. The only variation in this position is an occasional variation in nematodes to tyrosine. But as discussed in Chpater 4, Section 4.3.1, this substitution is readily accepted by HCCS and thus can be acted on by the HCCS proteins in these roundworms. The features of the N-terminal region now known to be important for recognition by HCCS were noted in Chapter 3 to be present in sequences of cytochromes *c* from *A. thaliana* and *T. brucei*. These sequences appeared not to be recognised by *S. cerevisiae* HCCS, possibly because the N-termini of the cytochromes in these organisms were extended at the extreme N-terminus. Nevertheless, the similarity of the ten residues immediately to the N-terminal side of the CXXCH motif in apocytochromes *c* not processed by HCCS to those processed by HCCS suggested that this sequence feature may have evolved before the rise of HCCS and the putative System V in trypanosomes. Such a view would be supported if the features of the sequences were highly conserved amongst all eukaryotic cytochromes *c* and so this chapter contains an analysis of a large number of cytochrome *c* sequences from organisms which do not use System III. Some comparisons of sequences for cytochromes using HCCS were given in Chapter 4, Figure 4.1.

As mentioned previously (Chapter 1, Section 1.6.3), the processing of cytochrome c_1 is more complex than that of cytochrome *c* but it is believed that a single HCCS can process both cytochrome *c* and cytochrome c_1 in higher eukaryotes whereas the separate HCC_1S is

dedicated to the maturation of cytochrome c_1 in fungi. Therefore it was of interest to compare the ten or so residues immediately to the N-terminal side of the CXXCH motif in the cytochromes c_1 of many distinct organisms.

However, apart from studies aiming to characterise the region with sequences of mitochondrial cytochrome *c* of relatively common species [140,143], a comprehensive analysis of the amino acid sequences of rarer mitochondrial cytochromes *c* has not previously been conducted due to the lack of number of sequences from genomes, especially from more 'exotic' species. As of 2014, it is reported that there are around 285 manually reviewed or '*cyc*' gene originated protein sequences of cytochrome *c* found in UniProt [166]. This is a tiny fraction of all the *c*-type cytochromes considering the number of species and the number of different *c*-type cytochromes in organisms, including isoforms.

In recent times, genome sequencing of different organisms has been carried out on model organisms, as well as common species, such as baker's yeast (*S. cerevisiae*) [167], horse [168] and cow [169]. There have also been developments in the analysis of transcriptome data of some species, e.g. Yamada *et al.* [130]. An analysis of the putative cytochrome *c* sequences identified from these whole genome sequencing may be performed by bioinformatics tools. The BLAST software at the UniProt website provides a convenient way to the search for the homologues/orthologues of any given protein sequence. To test the extent and scale of the conservation of the phenylalanine (residue 11 in horse protein and residue 15 in *S. cerevisiae*) and its variations in different *c*-type cytochromes, a

number of global sequence searches and sequence alignments were performed electronically on the UniProt site with BLAST and the ClustalW software.

6.2 Results and discussion

6.2.1 Anomaly of an alanine in position of consensus phenylalanine position in the N-terminal region of mitochondrial cytochrome *c*

A global sequence search by BLAST for plant cytochrome *c* sequences was performed against the *S. cerevisiae* iso-1 cytochrome *c*. Out of the 155 gene products of *Viridiplantae* (Green plants) from the BLAST search result (1000 hits), one suspected cytochrome *c* gene (UniProt entry K7UZX9) from the B73 *Zea mays* (maize) genome [170] contains a so far unique feature within the amino acid sequences, as shown in Figure 6.1. It has an alanine residue in the position of the usual phenylalanine which is otherwise highly conserved in most other *c*-type cytochromes. According to the raw data obtained from EMBL (CM000784 Genomic DNA. Translation: AFW83374.1), the genetic triplet code for this alanine is GCA. As the genetic code for phenylalanine is either TTT or TTC, it is unlikely that this unusual sequence is due to sequencing error in the sequencing process, as it requires that all three bases of this triplet to be incorrectly sequenced (shown in appendix). This suspected cytochrome also has an unusual 153-aa long valine rich N-terminal sequence next to the CXXCH motif (Figure 6.2 (A)), which is much longer than the more common, 22-aa long N-terminal sequence of other *c*-type cytochrome sequences. There are a few sequences with an N-terminal region longer than the typical length; however, the highly conserved phenylalanine is always present in the position within the

Figure 6.1 Multiple sequence alignment of 155 N-terminal sequences of mitochondrial cytochrome *c* and other sequences in the result of BLASTp from organisms in *Viridiplantae*. The unique sequence from *Zea mays* is coloured in yellow, with the alanine residue at the position of the highly conserved phenylalanine residue coloured green and the CXXCH motif coloured in cyan. The sequences of *S. cerevisiae* and *E. caballus* cytochromes *c* are both shown in grey boxes for reference. The CXXCH motifs in other sequences are coloured in dark grey. A unique SXXCH motif is coloured in red.

```
CYC1_YEAST    --------------------MTEFKA--GSAKKGATLFKTR-CLQCHT  25
D7LY03_ARALL  ------------------MASFDEAPP--GNPKAGEKIFRTK-CAQCHT  28
CYC3_ARATH    ------------------MASFDEAPP--GNPKAGEKIFRTK-CAQCHT  28
V4L8A1_EUTSA  ------------------MASFDEAPP--GNPKAGEKIFRTK-CAQCHT  28
M4C921_BRARP  ------------------MASFDEAPP--GNPKAGEKIFRTK-CAQCHT  28
R0GY45_9BRAS  ------------------MASFDEAPP--GNPKAGEKIFRTK-CAQCHT  28
V4TYB0_9ROSI  ------------------MASFDEAPP--GNAKAGEKIFKTK-CAQCHT  28
W9RNP9_9ROSA  ------------------MASFSEAPP--GNAKAGEKIFKTK-CAQCHT  28
V4T5D8_9ROSI  ------------------MASFEEAPP--GNPKAGEKIFKTK-CAQCHT  28
B2LXS2_VITVI  ------------------MASFDEAPP--GNPAAGEKIFKTK-CAQCHT  28
M5XNM3_PRUPE  ------------------MASFAEAPP--GDAKAGEKIFRTK-CAQCHT  28
I1HJA9_BRADI  ------------------MASFSEAPP--GNPKAGEKIFKTK-CAQCHT  28
A0A059BTW6_E  ------------------MASFAEAPP--GNPKVGEKIFKTK-CAQCHT  28
M4E6Y6_BRARP  ------------------MASFDEAPP--GNSKAGEKIFRTK-CAQCHT  28
M4D786_BRARP  ------------------MASFDEAPP--GNSKAGEKIFRTK-CAQCHT  28
A2Q5X9_MEDTR  ------------------MASFDEAPP--GNPAAGEKIFRTK-CAQCHT  28
W5AFP4_WHEAT  ------------------------------------FADVVGRAC--   9
D8RC36_SELML  ----------------MATSFQDAPVIVRNAANGRKIFMAK-CSDCHA  29
J3L6I6_ORYBR  ------------------MATFSEAPA--GDAAAGEKIFRTK-CAYCHS  28
D8T2Y2_SELML  ----------------MATSFQDAPVIVRNAANGRKIFMAK-CSDCHA  29
CYC_STELP     ----------------------MGFKE--GDAKKGANLFKTR-CAQCHT  24
CYC1_ARATH    ------------------MQVADISLQ--GDAKKGANLFKTR-CAQCHT  28
M0SZ69_MUSAM  ------------------MASFEDAPP--GDAKSGHKIFQSK-CAHCHT  28
S8C0E5_9LAMI  ------------------------------------------------   0
M7YRL6_TRIUA  ------------------------------------------MARC--   4
W5CBS2_WHEAT  ------------------------------------------------   0
K7UZX9_MAIZE  --------EEGLI---DRMNSIDDR-M--NSRAASTPPARTK-CAQCHT 161
W5BX87_WHEAT  ------------------MASFGEAPA--GNAASGEKIFKTK-CAQCHT  28
A2Q5Y1_MEDTR  ------------------------------------------------   0
C4J2R4_MAIZE  ------------------MASFSEAPP--GNPKAGEKIFKTK-CAQCHT  28
K3ZB68_SETIT  ------------------MASFSEAPP--GNPKAGEKIFKTK-CAQCHT  28
B3H4Y9_ARATH  ------------------MASFDEAPP--GNAKAGEKIFRTK-CAQCHT  28
W5BSC0_WHEAT  --------RSSRPS-ARSATRWSAAAP--TS----RAPTSAA-SXQCHT  72
M0YFK2_HORVD  ------------------MASFSEAPP--GNPASGEKIFKTK-CAQCHT  28
A2WXP5_ORYSI  ------------------MATFSDAPP--GDAAAGEKIFRTK-CAYCHA  28
M8C927_AEGTA  ------------------MASFSEAPP--GNPAAGEKIFKTK-CAQCHT  28
I1NU16_ORYGL  ------------------MATFSDAPP--GDAAAGEKIFRTK-CAYCHA  28
E1ZA99_CHLVA  ------------------MASFAEAPA--GDAAKGAKIFKTK-CAQCHT  28
Q8S0R8_ORYSJ  ------------------MATFSDAPP--GDAAAGEKIFRTK-CAYCHA  28
CYC_HORSE     -------------------------M--GDVEKGKKIFVQK-CAQCHT  20
```

```
CYC1_YEAST   --------------------MTEFKA--GSAKKGATLFKTR-CLQCHT  25
Q8GU28_9CHLO ------------------MASFADAPA--GNIANGEKIFKTK-CAQCHV  28
Q1WLY6_CHLIN ------------------MSTFAEAPA--GDLARGEKIFKTK-CAQCHV  28
A8J5T0_CHLRE ------------------MSTFAEAPA--GDLARGEKIFKTK-CAQCHV  28
CYC_CHLRE    ------------------MSTFAEAPA--GDLARGEKIFKTK-CAQCHV  28
D8RHT8_SELML ------------------MATFKDAPA--GNAASGEKLFKTK-CSACHT  28
C1EDX4_MICSR ------------------MASFAEAPA--GDAEKGAKIFKTK-CAQCHV  28
C1MXP8_MICPC ------------------MASFAEAPA--GDIAKGAKIFKTK-CAQCHV  28
K8EY41_9CHLO ------------------MASFAEAPA--GDVAKGAKIFKTK-CAQCHV  28
CYC_ULVIN    ------------------STFABAPP--GBPAKGAKIFKAK-CAZCHT  27
A9SLT5_PHYPA ------------------MATFGEAPK--GNGKAGEKIFKTK-CAQCHG  28
A4RW27_OSTLU ------------------MSTFSEAPA--GNAAKGAKIFKTK-CAQCHV  28
D8RFW8_SELML ------------------MATFGDAPA--GNPKNGEKIFKMK-CAQCHV  28
O22490_ORYSA ------------------MASFSEAPP--GNPKAGEKIFKTK-CAQCHT  28
B6T2W2_MAIZE ------------------MASFGEAPG--GDAGSGEKIFRTK-CAQCHT  28
I0YYR5_9CHLO ------------------MASFGEAPA--GDSAKGAKIFKTK-CAQCHV  28
B8Y8S4_CONCI ------------------MASFAEAPA--GNPKAGEKTFKTK-CAQCHA  28
A9SNF0_PHYPA ------------------MASFAEAPK--GNEKAGEKIFKTK-CAQCHA  28
A3AYZ5_ORYSJ ------------------MASFGVAPV--GNAASGKKIFRTK-CAQCHT  28
J3M320_ORYBR ------------------MASFGEAPG--GNAASGEKIFRTK-CAQCHT  28
M4F9Y7_BRARP ------------------MASFDEAPP--GNPKAGEKIFRTK-CAQCHT  28
W5BGZ9_WHEAT ------------------MASFGEAPA--GNAASGEKIFKTK-CAQCHT  28
M0UVA4_HORVD ------------------MASFGEAPA--GNAAGGEKIFKTK-CAQCHT  28
CYC_GUIAB    ------------------ASFAEAPA--GDAKAGEKIFKTK-CAZCHT  27
M8BUK1_AEGTA ------------------MASFGEAPA--GNAASGEKIFKTK-CAQCHT  28
CYC_GINBI    ------------------ATFSEAPP--GDPKAGEKIFKTK-CAZCHT  27
CYC_TROMA    ------------------ASFAEAPA--GDNKAGDKIFKNK-CAQCHT  27
CYC_FAGES    ------------------ATFSEAPP--GNIKSGEKIFKTK-CAQCHT  27
CYC_NIGDA    ------------------ASFBZAPA--GBSASGEKIFKTK-CAZCHT  27
A9SRQ9_PHYPA ------------------MATFAEAPK--GNDKAGEKIFKTK-CAQCHA  28
CYC_PASSA    ------------------ASFAEAPP--GDKDVGGKIFKTK-CAZCHT  27
F2EGP0_HORVD ------------------MASFGEAPA--GNAAGGEKIFKTK-CAQCHT  28
A9NKS1_PICSI ------------------MSTFSEAPP--GDTKVGEKIFKMK-CAQCHT  28
CYC_SPIOL    ------------------ATFSEAPP--GNKDVGAKIFKTK-CAQCHT  27
K7UNJ8_MAIZE ------------------MASFGEAPG--GDAGSGEKIFRTK-CAQCHT  28
K3ZAX4_SETIT ------------------MASFAEAPA--GDAGSGEKIFRTK-CAQCHT  28
C5YB66_SORBI ------------------MASSFGEAPV--GDAGSGEKIFRTK-CAQCHT  29
A9RJ09_PHYPA ------------------MATFAEAPK--GNEKAGEKIFKTK-CAQCHA  28
B9IQC2_POPTR ------------------MASFAEAPP--GDSKVGEKVFKTK-CAQCHT  28
CYC_HELAN    ------------------MASFAEAPA--GNPTTGEKIFKTK-CAQCHT  28
W4ZNQ8_WHEAT ------------------MASFSEAPP--GNPDAGAKIFKTK-CAQCHT  28
B9RFB0_RICCO ------------------MATFEQAPP--GESKAGEKIFKTK-CAQCHT  28
CYC_MAIZE    ------------------ASFSEAPP--GNPKAGEKIFKTK-CAQCHT  27
CYC_SOLTU    ------------------ASFGEAPP--GNPKAGEKIFKTK-CAQCHT  27
M8A7D1_TRIUA --------VRAGSAAAAAMASFSEAPP--GNPAAGEKIFKTK-CAQCHT 103
CYC_FRIAG    ------------------MASFSEAPP--GDFKSGEKIFKTK-CAQCHT  28
A9PJI0_9ROSI ------------------MASFAEAPP--GDSKVGEKVFKTK-CAQCHT  28
CYC_HORSE    -------------------------M--GDVEKGKKIFVQK-CAQCHT  20
```

```
CYC1_YEAST   --------------------MTEFKA--GSAKKGATLFKTR-CLQCHT  25
M0SUR9_MUSAM ------------------MASFAEAPP--GDPKAGEKIFKTK-CAQCHT  28
K4D814_SOLLC ------------------MASFGEAPA--GNEKAGEKIFKTK-CAQCHT  28
A9P8Y7_POPTR ------------------MASFAEAPP--GDSKVGEKVFKTK-CTQCHT  28
Q9ZSL2_CICIN ------------------MASFAEAPA--GNATAGEKIFKTK-CAQCHT  28
M0TK41_MUSAM ------------------MASFAEAPP--GDPKAGEKIFKTK-CAQCHT  28
CYC_ALLPO    ------------------ATFSZAPP--GBZKAGQKIFKLK-CAQCHT  27
D7TGR0_VITVI ------------------MATFADAPA--GNAAAGEKIFRTK-CAQCHT  28
W1NRT4_AMBTC ------------------MATFNEAPP--GNDKAGEKIFKTK-CAQCHT  28
M8BK26_AEGTA ------------------MASFSEAPP--GNPDAGAKIFKTK-CAQCHT  28
F2DJB5_HORVD ------------------MASFSEAPP--GNPDAGAKIFKTK-CAQCHT  28
CYC_WHEAT    ------------------ASFSEAPP--GNPDAGAKIFKTK-CAQCHT  27
B8LQM4_PICSI ------------------MASFSEAPP--GDAKAGEKIFKTK-CAQCHT  28
U7DZX5_POPTR ------KQKAKRIQRDEEMASFAEAPP--GDSKVGEKIFKTK-CAQCHT  78
Q6S4N3_HELAN ------------------MASFAEAPA--GNPTTGEKIFKTK-CAQCHT  28
A9P208_PICSI ------------------MASFSEAPP--GDAKVGEKIFKTK-CAQCHT  28
CYC_ACENE    ------------------ASFAEAPP--GNPAAGEKIFKTK-CAQCHT  27
CYC_GOSBA    ------------------ASFQZAPP--GBAKAGEKIFKTK-CAQCHT  27
A9PHL1_POPTR ------------------MASFAEAPP--GDSKAGEKIFRTK-CAQCHT  28
CYC_ARUMA    ------------------ASFAEAPP--GNPKAGEKIFKTK-CAQCHT  27
M0RTM5_MUSAM ------------------MASFAEAPP--GDPNAGEKIFKTK-CAQCHT  28
C0PK55_MAIZE ------------------MASFSEAPP--GNPKAGEKIFKTK-CAQCHT  28
M0ZKD3_SOLTU ------------------MASFGEAPP--GNEKAGEKIFKTK-CAQCHT  28
CYC_SESIN    ------------------ASFBZAPP--GBVKSGEKIFKTK-CAQCHT  27
CYC_CANSA    ------------------ASFBZAPP--GBSKAGEKIFKTK-CAECHT  27
A0A022R5V3_M ------------------MASFDEAPP--GDVKSGDKIFRTK-CAQCHT  28
CYC_ORYSJ    ------------------MASFSEAPP--GNPKAGEKIFKTK-CAQCHT  28
C6SZC5_SOYBN ------------------MASFEEAPP--GNSKNGEKIFKTK-CAQCHT  28
A0A022R107_M ------------------MASFDEAPP--GDGKAGEKIFKTK-CAQCHT  28
C5XEM6_SORBI ------------------MASFSEAPP--GNPTAGEKIFKTK-CAQCHT  28
CYC_VIGRR    ------------------ASFDEAPP--GNSKSGEKIFKTK-CAQCHT  27
M0YFK3_HORVD ------------------MASFSEAPP--GNPASGEKIFKTK-CAQCHT  28
M0T845_MUSAM ------------------MASFAEAPP--GDPKAGEKIFKTK-CAQCHT  28
B6TGS7_MAIZE ------------------MASFSEAPP--GNPKAGEKIFKTK-CAQCHT  28
V7B4A0_PHAVU ------------------MASFDEAPP--GNSKNGEKIFKTK-CAQCHT  28
U5FVI3_POPTR --------RPLLLAEPEEMASFAEAPP--GDSKAGEKIFRTK-CAQCHT  90
B6T4T7_MAIZE ------------------MASFSEAPP--GNPTAGEKIFKTK-CAQCHT  28
G7KSG6_MEDTR TFLYNSFATRCSFAELSQMASFDEAPP--GNPAAGEKIFRTK-CAQCHT 261
K3XMJ2_SETIT ----------GSPSCSAAMASFSEAPP--GNPAAGEKIFKTK-CAQCHT  88
B9SS73_RICCO ------------------MASFDQAPP--GDVKAGEKIFKTK-CAQCHT  28
K3ZAX0_SETIT ------------------MASFSEAPP--GNPKAGEKIFKTK-CAQCHT  28
K7XKN6_SOLTU ------------------MASFGEAPP--GNPKAGEKIFKTK-CAQCHT  28
CYC_RICCO    ------------------ASFBZAPP--GBVKAGEKIFKTK-CAQCHT  27
V7CI22_PHAVU ------------------MASFDLAPP--GDVKSGDKIFRTK-CAQCHT  28
B4FYS2_MAIZE ------------------MASFSEAPP--GNPTAGEKIFKTK-CAQCHT  28
B6SKR4_MAIZE ------------------MASFSEAPP--GNPKAGEKIFKTK-CAQCHT  28
C5YY64_SORBI ------------------MASFSEAPP--GNPKAGEKIFKTK-CAQCHT  28
CYC_HORSE    -------------------------M--GDVEKGKKIFVQK-CAQCHT  20
```

```
CYC1_YEAST    --------------------MTEFKA--GSAKKGATLFKTR-CLQCHT  25
I1HU25_BRADI  ------------------MASFSEAPP--GNQATGEKIFKTK-CAQCHT  28
M1CBK0_SOLTU  ------------------MASFGEAPP--GNPKAGEKIFKTK-CAQCHT  28
W5D5X9_WHEAT  ------------------MASFSEAPP--GNPAAGEKIFKTK-CAQCHT  28
K4B1P8_SOLLC  ------------------MASFAEAPP--GNPKAGEKIFKTK-CAQCHT  28
I3T359_LOTJA  ------------------MATFDEAPP--GDTKAGEKIFKIK-CAQCHT  28
CYC_ORYSI     ------------------MASFSEAPP--GNPKAGEKIFKTK-CAQCHT  28
B7E4T8_ORYSJ  ------------------MASFSEAPP--GNPKAGEKIFKTK-CAQCHT  28
I1PVR3_ORYGL  ------------------MASFSEAPP--GNPKAGEKIFKTK-CAQCHT  28
J3M7A4_ORYBR  ------------------MASFSEAPP--GNPKAGEKIFKTK-CAQCHT  28
V4KTA0_EUTSA  ------------------MASFDEAPP--GNPTAGDKIFRLK-CAQCHT  28
CYC_SOLLC     ------------------ASFNEAPP--GNPKAGEKIFKTK-CAQCHT  27
CYC2_ARATH    ------------------MASFDEAPP--GNAKAGEKIFRTK-CAQCHT  28
C6SYA8_SOYBN  ------------------MASFDQAPP--GDVKSGDKIFRTK-CAQCHT  28
CYC_BRANA     ------------------ASFDEAPP--GNSKAGEKIFKTK-CAQCHT  27
CYC_BRAOL     ------------------ASFDEAPP--GNSKAGEKIFKTK-CAQCHT  27
G7K7W9_MEDTR  ------------------MASFELAPP--GDAKSGDKIFRTK-CAQCHT  28
CYC_ABUTH     ------------------ASFQZAPP--GBAKAGEKIFKTK-CAQCHT  27
C6TJ91_SOYBN  ------------------MASFDEAPP--GNSKNGEKIFKTK-CAQCHT  28
R0I8G5_9BRAS  RIHFLLVVPDPVVGQAEKMASFDEAPP--GNGKAGEKIFRTK-CAQCHT  70
C6SX87_SOYBN  ------------------MASFDQAPP--GDVKSGDKIFRTK-CAQCHT  28
CYC_SAMNI     ------------------ASFAEAPP--GNPKAGEKIFKTK-CNQCHT  27
CYC_CUCMA     ------------------ASFDEAPP--GNSKAGEKIFKTK-CAQCHT  27
D7KMK0_ARALL  ------------------MASFDEAPP--GNAKAGEKIFRTK-CAQCHT  28
I3SL85_LOTJA  ------------------MASFDEAPP--GNSKSGEKIFKLK-CAQCHT  28
A5AQD0_VITVI  ------------------MASFDEAPP--GNPAAGEKIFKTK-CAQCHT  28
CYC_HORSE     -------------------------M--GDVEKGKKIFVQK-CAQCHT  20
```

consensus sequence in the N-terminal region of these proteins. Any evidence that this predicted protein with alanine rather than phenylalanine in this position is expressed, such as transcriptomic data, has yet to be shown.

```
1    ------------------------------------------------------------    0  K7UNJ8     K7UNJ8_MAIZE
1    ------------------------------------------------------------    0  B6T2W2     B6T2W2_MAIZE
1    ------------------------------------------------------------    0  B6T4T7     B6T4T7_MAIZE
1    ------------------------------------------------------------    0  B4FYS2     B4FYS2_MAIZE
1    ------------------------------------------------------------    0  B6TGS7     B6TGS7_MAIZE
1    ------------------------------------------------------------    0  B6SKR4     B6SKR4_MAIZE
1    ------------------------------------------------------------    0  P00056     CYC_MAIZE
1    ------------------------------------------------------------    0  C0PK55     C0PK55_MAIZE
1    MVDEVEEEVHAMVVQEPEVKDVLVTKVVEALEVTKVHPVLESEPRVDEVLVVNETPVVPE   60  K7UZX9     K7UZX9_MAIZE

1    ------------------------------------------------------------    0  K7UNJ8     K7UNJ8_MAIZE
1    ------------------------------------------------------------    0  B6T2W2     B6T2W2_MAIZE
1    ------------------------------------------------------------    0  B6T4T7     B6T4T7_MAIZE
1    ------------------------------------------------------------    0  B4FYS2     B4FYS2_MAIZE
1    ------------------------------------------------------------    0  B6TGS7     B6TGS7_MAIZE
1    ------------------------------------------------------------    0  B6SKR4     B6SKR4_MAIZE
1    ------------------------------------------------------------    0  P00056     CYC_MAIZE
1    ------------------------------------------------------------    0  C0PK55     C0PK55_MAIZE
61   IQEPEVKGVGATVVLKEQETNSGNVVVKDSDEALPSQEVAGVHTTDRGSCQAALFGALVI  120  K7UZX9     K7UZX9_MAIZE

1    --------------MASFGEAPGGDAGSGEKIFRTKCAQCHTVERGGAHRQGPNLHGLFG   46  K7UNJ8     K7UNJ8_MAIZE
1    --------------MASFGEAPGGDAGSGEKIFRTKCAQCHTVERGGAHRQGPNLHGLFG   46  B6T2W2     B6T2W2_MAIZE
1    --------------MASFSEAPPGNPTAGEKIFKTKCAQCHTVDKGAGHKQGPNLNGLFG   46  B6T4T7     B6T4T7_MAIZE
1    --------------MASFSEAPPGNPTAGEKIFKTKCAQCHTVDKGAGHKQGPNLNGLFG   46  B4FYS2     B4FYS2_MAIZE
1    --------------MASFSEAPPGNPKAGEKIFKTKCAQCHTVEKGAGHKQGPNLNGLFG   46  B6TGS7     B6TGS7_MAIZE
1    --------------MASFSEAPPGNPKAGEKIFKTKCAQCHTVDKGAGHKQGPNLNGLFG   46  B6SKR4     B6SKR4_MAIZE
1    ---------------ASFSEAPPGNPKAGEKIFKTKCAQCHTVEKGAGHKQGPNLNGLFG   45  P00056     CYC_MAIZE
1    --------------MASFSEAPPGNPKAGEKIFKTKCAQCHTVDKGAGHKQGPNLNGLFG   46  C0PK55     C0PK55_MAIZE
121  IFKNGTIEEGLIDRMNSIDDRM-NSRAASTPPARTKCAQCHTVERGGAHRQGPNLHDLFG  179  K7UZX9     K7UZX9_MAIZE
                   *: :    .  :.    :*********::*..*:*****. ***

47   RQSGTTLGYAYSTANKNMAVVWEEGTLYDYLFNPKKYIPGTKMVFPGLKKPKERTDLIAY  106  K7UNJ8     K7UNJ8_MAIZE
47   RHSGTTLGYAYSTANKNMAVVWEEGTLYDYLFNPKKYIPGTKMVFPGLKKPKERTDLIAY  106  B6T2W2     B6T2W2_MAIZE
47   RQSGTTPGYSYSSANKNMAVIWEENTLYDYLLNPKKYIPGTKMVFPGLKKPQERVDLIAY  106  B6T4T7     B6T4T7_MAIZE
47   RQSGTTPGYSYSSANKNMAVIWEENTLYDYLLNPKKYIPGTKMVFPGLKKPQERADLIAY  106  B4FYS2     B4FYS2_MAIZE
47   RQSGTTTGYSYSAGNKNKAVIWEENTLYDYLLNPKKYIPGTKMVFPGLKKPQERADLIAY  106  B6TGS7     B6TGS7_MAIZE
47   RQSGTTAGYSYSAGNKNKAVVWEENTLYDYLLNPKKYIPGTKMVFPGLKKPQERADLIAY  106  B6SKR4     B6SKR4_MAIZE
46   RQSGTTAGYSYSAANKNKAVVWEENTLYDYLLNPXKYIPGTKMVFPGLXKPQERADLIAY  105  P00056     CYC_MAIZE
47   RQSGTTAGYSYSAGNKNKAVVWEEDTLYEYLLNPKKYIPGTKMVFPGLKKPQERADLIAY  106  C0PK55     C0PK55_MAIZE
180  RQSGTTIGYAYSTANKNMAVVWEKGTMYDYLFNPKKYIPGTKMVFPRLKKPKERTDLIAY  239  K7UZX9     K7UZX9_MAIZE
     *:**** **:**:.*** **:**: *:*:**:** ********** * **:**.*****

107  LKESTA  112  K7UNJ8     K7UNJ8_MAIZE
107  LKESTA  112  B6T2W2     B6T2W2_MAIZE
107  LKNATV  112  B6T4T7     B6T4T7_MAIZE
107  LKNATV  112  B4FYS2     B4FYS2_MAIZE
107  LKEATA  112  B6TGS7     B6TGS7_MAIZE
107  LKEATA  112  B6SKR4     B6SKR4_MAIZE
106  LKEATA  111  P00056     CYC_MAIZE
107  LKEATA  112  C0PK55     C0PK55_MAIZE
240  LTY---  242  K7UZX9     K7UZX9_MAIZE
     *.
```

A

B

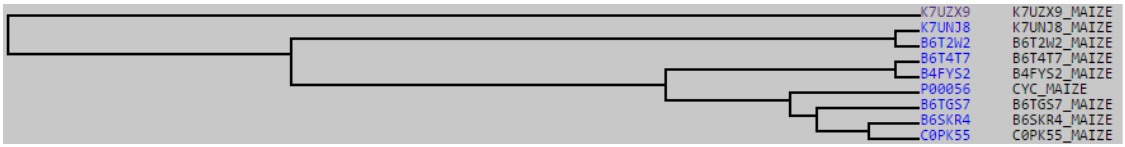


Figure 6.2 Multiple sequence alignment of the *Zea mays* putative cytochromes *c*. (A) the raw result of the Cluster Omega sequence alignment of 9 *Z. mays* cytochromes *c*; (C) The guide tree of this sequence alignment, showing the taxonomy between these predicted proteins.

All the other sequences from other subspecies of *Z. mays* share the common consensus sequence. From the guide tree of the alignment of these sequences from *Z. mays* (Figure 6.2 (B)), it can also be deduced that the gene encoding K7UZX9 is the most divergent one out of the 9 sequences with a CXXCH motif from *Z. mays*. With the long N-terminal sequence and this alanine in the conserved phenylalanine position, it can be suggested that K7UZX9 protein had undergone a series of mutations to become its current form.

Figure 6.1 shows that a large number of plant mitochondrial cytochrome *c* sequences share an FK/RTK/R sequence immediately to the N-terminal side of the CXXCH motif in common with the *S. cerevisiae* protein. In the horse protein, the sequence FVQK occurs in the equivalent position. All sequences have a hydrophobic residue L/I/V before the conserved F and a common G three residues further back in the sequence. Thus it seems that the main recognition features for HCCS in the N-terminal region of mitochondrial cytochrome *c* are present in the plant proteins even though they are processed by the Ccm system. However, for some unknown reasons, the attempt to experimentally support this proposal (Chapter 3, Section 3.2.3) was unsuccessful. Nevertheless, it is reasonable to propose that the HCCS system evolved to recognise elements of the N-terminal sequence in mitochondrial *c*-type cytochromes, for example the FK/RTK/RCXXCH sequence that is common to yeast and plants. Similarly, a recent review by Verissimo and Daldal [44] proposed a K/AGXXL/IFXXXCXXCH consensus sequence based on recent studies.

It is also notable in the CXXCH motif, the XX residues are very commonly AQ or LQ in mitochondrial *c*-type cytochromes. There are many bacterial *c*-type cytochromes with a

fold similar to mitochondrial cytochrome *c* which are commonly called cytochrome c_2 [5]. The XX is varied amongst these proteins, but LQ and AQ are very rare [171]. In addition, as noted in Chapter 3, Section 3.3.1, the bacterial c_2-type cytochromes do not have the recognition features for HCCS in the N-terminal region.

The protein sequences of cytochromes *c* in trypanosome and related species were also analysed. For the members of phylum *Euglenozoa* in kingdom *Exavata*, the mitochondrial cytochromes *c* are not synthesised by the typical biogenesis System III, but the putative System V. A global sequence search by BLAST software on the *T. brucei* cytochrome *c* results in a total of 22 gene products, and the sequence alignment shows none of them possess any other residue other than phenylalanine in the aforementioned highly conserved position (Figure 6.1 (A)). This group of proteins has an AAQCH heme-binding motif rather than the CAQCH motif found in many plant sequences, but it should be noted that AQ is well conserved in euglenoids. There are other similarities between euglenoid and plant N-terminal regions, such as the conserved EKIFR/KXK/R sequence next to the CXXCH motif, which can be observed from Figure 6.1 and 6.2.

However, a sequence from the review by Allen *et al.* mentions two isoforms of mitochondrial cytochrome *c* from a euglenoid species *Diplonema papillatum* that also have an alanine residue at the consensus phenylalanine position in the N-terminal region sequence. This sequence was not extracted from genomic data, but from peptide sequencing method of Edman degradation. Therefore this sequence cannot be traced in the

Figure 6.3 Multiple sequence alignment of 22 N-terminal sequences of mitochondrial cytochrome *c* and other sequences in the result of BLASTp from euglenoid organisms. Two extra sequences from *Diplonema papillatum* (adapted from Allen [38]) are coloured in yellow, with the alanine residue at the position of the highly conserved phenylalanine residue coloured in green and the AXXCH motif coloured in cyan. The sequences of *S. cerevisiae* and *E. caballus* cytochromes *c* are both shown in grey boxes for reference. The CXXCH motifs in other sequences are coloured in dark grey.

```
CYC1_YEAST      -----MTEFKAGSAKKGATLFKTRCLQCHTVEKGGPHKVGPNLHG  40
Dpiso1          -------GYKGGDETKGKKLAQTRAGQCHNFASGGPNGVGPNLFG  38
Dpiso2          -------GYKDGNVEKGKKLASTRAGQCHNFASGGPNGVGPNLFG  38
Q57UI4_TRYB2    MPPKERAALPPGDAVRGEKLFKGRAAQCHTGTKGGSNGVGPNLYG  45
C9ZVW7_TRYB9    MPPKERAALPPGDAARGEKLFKGRAAQCHTGTKGGSNGVGPNLYG  45
G0USA6_TRYCI    MPPKERPPLPPGDVARGEKLFKGRAAQCHTGTKGGSNGVGPNLYG  45
A4H8Q9_LEIBR    MPPKERPPLPPGDVARGEKLFKGRAAQCHTATKGGSNGVGPNLFG  45
G0U1A6_TRYVY    MPPKERPPLPPGDPVRGEKIFKGRAAQCHTGSRGGSNGVGPNLYG  45
S9USL6_9TRYP    MPPKAREPLPPGDAAKGEKIFKGRAAQCHTGTKGGANGVGPNLFG  45
Q4CV48_TRYCC    MPPKARPPLPPGDAAKGEKIFKGRAAQCHTGTKGGANGVGPNLYG  45
Q4D480_TRYCC    MPPKARPPLPPGDAAKGEKIFKGRAAQCHTGTKGGANGVGPNLYG  45
A4H8R0_LEIBR    MPPKARAPLPPGDVERGEKLFKGRAAQCHTATKGGSNGVGPNLFG  45
Q4QEN5_LEIMA    MPPKARAPLPPGDVERGEKLFKGRAAQCHTATKGGSNGVGPNLFG  45
E9AQU2_LEIMU    MPPKARPPLPPGDVERGEKLFKGRAAQCHTATKGGSNGVGPNLFG  45
CYC_CRION       MPPKAREPLPPGDAAKGEKIFKGRAAQCHTGAKGGANGVGPNLFG  45
CYC_CRIFA       MPPKARAPLPPGDAARGEKLFKGRAAQCHTANQGGANGVGPNLYG  45
A4HX29_LEIIN    MPPKARAPLPPGDVERGEKLFKGRAAQCHTATKGGANGVGPNLFG  45
E9BCZ1_LEIDB    MPPKARAPLPPGDVERGEKLFKGRAAQCHTATKGGANGVGPNLFG  45
S9V7Q1_9TRYP    MPPKARAPLPPGDVERGEKLFKGRAAQCHTATKGGANGVGPNLFG  45
S9WL95_9TRYP    MPPKAREPLPEGDATRGEKIFKGRAAQCHTATKGGANGVGPNLFG  45
E9BCZ0_LEIDB    MPPKARAPLPPGDVERGEKLFKGRAAQCHTATKGGANGVGPNLFG  45
A4HX28_LEIIN    MPPKARAPLPPGDVERGEKLFKGRAAQCHTATKGGANGVGPNLFG  45
CYC_TRYBB       ----------MGDA---------------AAAQGGSNGVGPNLYG  24
CYC_EUGVI       -----------GDAERGKKLFESRAGQCHSSQK-GVNSTGPALYG  33
CYC_EUGGR       -----------GDAERGKKLFESRAAQCHSAQK-GVNSTGPSLWG  33
CYC_HORSE       ----------MGDVEKGKKIFVQKCAQCHTVEKGGKHKTGPNLHG  35
```

UniProt database or other online sources, but the genes are expressed and matured to give holoprotein.

The importance of the highly conserved F in the N-terminal region of mitochondrial cytochromes *c* for HCCS substrate specificity has been established (Chapter 4, Section 4.2.1.2.1). Given that the vast majority of plants use the Ccm system but not HCCS for the heme attachment to mitochondrial cytochromes *c* [38], the N-terminal region of plant cytochromes *c* still preserves this F at the equivalent position. Thus, the evolutionary implication is that the phenylalanine residue, as well as the other highly conserved residues in the N-terminal region, probably evolved in an earlier age than divergence of different cytochrome *c* biogenesis systems. The use of the uncharacterised System V can be deduced from the *D. papillatum* sequences which possess no aromatic residue equivalent to F15 in *S. cerevisiae*. It is also notable that while there is a K or R next to the F in most euglenoids and plants, in *D. papillatum* there is a Q following this A at the conserved F position (Figure 6.3), suggesting a possible smaller selection-pressure on this organism than trypanosomes caused evolutionary shift for this protein which is not produced by HCCS. Presumably System V, in common with System I, does not require an F residue in this position for substrate recognition. However, since System V is only a speculated system, there is no evidence that all the euglenoid organisms use the same system. It is possible that *D. papillatum* (and related species) utilise the Ccm system instead; however, it is known that the recognition of the AXXCH motif by the Ccm system is poor [119]. It is also possible that these species use another yet unknown system different from System V.

The presence of phenylalanine in various *c*-type cytochromes in different organisms which include bacteria, archea, *Exavata*, other prototists, yeast, plant and animals is wide spread. However, the N-terminal sequence of cytochromes *c* in different species is far more varied

and diverse, with different substrate specificity for different cytochrome *c* biogenesis systems. A possible common ancestor of the cytochromes *c* with this phenylalanine had a first cytochrome *c* biogenesis system, and different systems evolved later on. Also, even though this phenylalanine position is not critical for the structural integrity of the protein (as discussed in Chapter 4), it is deeply embedded in the aromatic patch within the fold of cytochrome *c*, and its absence possibly can cause a conformational change and/or disruption in function of the protein.

6.2.2 Anomaly of a tyrosine in position of consensus phenylalanine position in the N-terminal region of mitochondrial cytochrome *c*

C. elegans, a well-known and well-characterised model organism, has two isoforms of cytochromes *c* genes in its genome which are both expressed. One of these isoforms (UniProt entry P19974/CYC21_CAEEL, termed cytochrome $c_{2.1}$) of *C. elegans* was found to contain a tyrosine residue at the position of the highly conserved phenylalanine residue in the N-terminal region. The other isoform, cytochrome $c_{2.2}$, (UniProt entry Q23240/CYC22_CAEEL) has the typical phenylalanine at this position, but it has an unusually long N-terminus. A multiple sequence alignment of these two sequences can be found in Figure 4.18 in Chapter 4. The result of the subsequent global sequence search by BLAST performed on the *C. elegans* cytochrome $c_{2.1}$ gives 39 hits in the phylum *Nematoda*. In total, eight sequences other than *C. elegans* cytochrome $c_{2.1}$were found to contain this tyrosine at the conserved phenylalanine position ((UniProt entries: E3M649, G0PK93, A8WQY3/CYC21_CAEBR, H2WIC6, A0A016WP84, W2TKL3 and W2SWV9) and no hits from organisms other than nematodes (or roundworms, Figure 6.4).

All of these sequences, except H9X295, have a methionine at the start indicating an open reading frame. Four sequences are from organisms in genus *Caenorhabditis*: *Caenorhabditis remanei*, *Caenorhabditis brenneri*, *Caenorhabditis briggsae* and *Caenorhabditis japonica*. The other three sequences (Figure 6.4) are from *Ancylostoma ceylanicum* and *Necator americanus* (two proteins).

As has already been discussed in Chapter 4, Section 4.3.1, it has been confirmed that the F15Y variant of *S. cerevisiae* cytochrome *c* could be processed by HCCS. As *C. elegans* has one HCCS gene in its genome, it can be deducted that it uses System III to attach heme to the cytochromes *c* and c_1. The discovery of the presence of tyrosine at this position in the natural world agrees with the conclusion that the N-terminal for HCCS recognition of apocytochrome *c* is complex and multi-factored. It also implies that these species underwent particular selection pressure in its evolutionary history to somehow acquire and maintain such a substitution. It should also be noted, however, that the consensus sequence, such as the I/V/L next to the usual F, as well as the G three residues away, is present in most of the sequences; thus the HCCS recognition for the apocytochorme *c* is not affected. Some of the other elements in the N-terminus, such as the K and G mentioned in the previous proposal [44], are also highly conserved in most sequences.

Figure 6.4 Multiple sequence alignment of 39 N-terminal sequences of mitochondrial cytochrome *c* and other sequences in the result of BLASTp from organisms in the phylum *Nematoda*. The 8 unique sequences are coloured in yellow, with the tyrosine residues at the position of the highly conserved phenylalanine residue coloured in green and the CXXCH motifs coloured in cyan. The sequences of *S. cerevisiae* and *E. caballus* cytochromes *c* are both shown in grey boxes for reference. The CXXCH motifs in other sequences are coloured in dark grey.

```
CYC1_YEAST    ------------------MTEFKAGSAKKGATLFKTRCLQCHTVEKGGP  31
W2TKL3_NECAM  ----------------------FFIITLRLIQIYKYRCAQCHSINS-TR  26
W2SWV9_NECAM  AAKAGSIPEGDYEKGKKIFSLKFFIITLRFIQIYKYRCAQCHSINS-TR  52
CYC21_CAEEL   ------------------MSDIPAGDYEKGKKVYKQRCLQCHVVDS-TA  30
E3M649_CAERE  ------------------MADIPAGDYEKGKKVYKQRCLQCHVVDS-TA  30
G0PK93_CAEBE  ------------------MADIPAGDYEKGKKVYKQRCLQCHVVDS-TA  30
CYC21_CAEBR   ------------------MADIPAGDYEKGKKVYKQRCLQCHVVDS-TA  30
A0A016RX68_9  ------------MFKVAEMADIPEGDYEKGKKVFKQRCLQCHVVDS-KA  36
H2WIC6_CAEJA  ------------------MADIPAGDYEKGKKVYKQRCLQCHVVDS-TA  30
W2SJX0_NECAM  ------------------MADIPEGDYERGKKVFKQRCLQCHVVDS-KA  30
B6D1W8_HAECO  ------------------MGDIPEGDYEKGKKVFKQRCLQCHVVDS-KA  30
F1LBJ8_ASCSU  ------------------MPEIPEGDYERGKKVFKQRCLQCHVVDS-KA  30
F1LFU2_ASCSU  MAFTLREE----FGNSAKMPEIPEGDYERGKKVFKQRCLQCHVVDS-KA  44
CYC1_ASCSU    ------------------MPEIPEGDYESGKKVFKQRCLQCHVVDS-KA  30
H3FYW3_PRIPA  ------------------MAAIPDGDYERGKKVFKQRCLQCHVADS-KA  30
B6DQJ7_9BILA  ------MG-----KGSDSGSGIPEGDYEKGKKVFKQRCAQCHVIDS-MA  37
E3LNM6_CAERE  ----MGKK----KSDSASGGAIPEGDYEKGKKIFKQRCEQCHVVNS-LQ  40
G0P548_CAEBE  ----MGKK----KSDSASGGAIPEGDYEKGKKIFKQRCEQCHVVNS-LQ  40
H2VMB2_CAEJA  ----MGKK----KSD-AGAGAIPEGDYEKGKKIFKQRCEQCHVVNS-LQ  40
A0A016WRJ0_9  ----------------MGSGGTPEGDYEKGKKVFKQRCLQCHVIDS-MA  32
E3NU84_CAERE  ----MGKK----NSDSASGGAIPEGDYEKGKKIFKQRCEQCHVVNS-LQ  40
G0PP39_CAEBE  ----MGKK----KSDSASGGAIPEGDYEKGKKIFKQRCEQCHVVNS-LQ  40
A0A044T2C1_O  ------------LNLHIVMSVPEGDYERGKKLFKMRCLQCHVIDS-DA  35
CYC22_CAEEL   ----MGKK----KSDTASGGAIPEGDNEKGKKIFKQRCEQCHVVNS-LQ  40
A8PJQ3_BRUMA  ------------------MSIPEGDYERGKKLFKMRCLQCHVIDS-DA  29
E1FYJ0_LOALO  ------------------MSIPEGDYERGKKLFKMRCLQCHVIDS-DA  29
CYC22_CAEBR   ----MGKK----KSDSASGGAIPEGDYEKGKKIFKQRCEQCHVVNS-LQ  40
A0A016UN64_9  ----------------MGSEDIPEGDYEKGKKIFKTRCVSCHVIDS-KS  32
U6PL52_HAECO  ------MG-----KSKKSEEGIPEGDYEKGKKIFKQRCLQCHAIDS-MA  37
F1LGX4_ASCSU  ------------------MTEIPEGDYERGKQIFKTRCLMCHVVDS-DK  30
CYC2_ASCSU    ------------------MTEIPEGDYESGKQIFKTRCLMCHVVDS-DK  30
A0A016WP84_9  ---TAGAA----KGDASSPKSIPEGDYEKGKKIYKYRCAQCHTINS-KR  40
U1LRA0_ASCSU  ----------------------------------------------IE 105
E5S446_TRISP  ------------------MSQIPKGDPEKGKKLFVQRCAQCHTVEKGGG  31
U6PIW5_HAECO  ----MGIWKKDKSDSKTESTSIPEGDYERGKKLFKYRCAQCHAINT-TR  44
F1LCW4_ASCSU  ----------------------------------------------IA 107
```

```
U1MK72_ASCSU ----------------------------------------IA 107
CYC_HORSE    ----------------------MGDVEKGKKIFVQKCAQCHTVEKGGK  26
```

6.2.3 Variation of the highly conserved tyrosine residue in N-terminus of mitochondrial cytochrome c_1 from selected species

A similar global sequence search was carried out using the *S. cerevisiae* cytochrome c_1 for the sequences of cytochrome c_1 from diverse species. According to a sequence alignment by ClustalW, the mitochondrial cytochrome c_1 sequences have a highly conserved tyrosine at the corresponding position of the highly conserved phenylalanine in cytochromes *c* (discussed in Chapter 4, Section 4.3.1). This phenomenon exists across a broad range of species, including animals, plants, bacteria and protozoa. However, some exceptions exist and in the sequence alignment carried out in this work, a phenylalanine presents in the equivalent position for this tyrosine in eight sequences within the alignment (UniProt entries: M0S0X3, A9NX47, A9NMY7, A9RCI4, A9SAZ2, A9SB01, A0A059AGZ7, H9X295), as shown in the multiple sequence alignment in Figure 6.5. It should be noted that all of these eight sequences have not been manually reviewed, and they are just putative proteins reported on the transcription level (A9NX47 and A9NMY7), or simply predicted which is similar to the situation for K7UZX9 sequence in maize which is mentioned in Section 6.2.1. All of these sequences, except H9X295, have a methionine at the beginning for the validity of an open reading frame. These sequences are from a diverse collection of species: *Musa acuminate* (banana), *Picea sitchensis* (*Sitka* spruce, *2* proteins), *Physcomitrella patens patens* (Moss, 3 proteins), *Eucalyptus grandis* (Flooded gum) and *Pinus taeda* (Loblolly pine). Inspection of Figure 6.5 shows that if cytochrome c_1 is recognised by HCCS (or HCC_1S, depending on the organism) by sequence

immediately on the N-terminal side of CXXCH, then this sequence is in plant cytochrome *c*, that are processed by System I.

Figure 6.5 Multiple sequence alignment of 155 N-terminal sequences of mitochondrial cytochrome c_1 and other sequences in the result of BLASTp from organisms in *Viridiplantae*. The 8 unique sequences are depicted in yellow background, with the tyrosine residues at the position of the highly conserved phenylanine residue coloured in green and the CXXCH motifs coloured in cyan. The sequences of *S. cerevisiae* and *H. sapiens* cytochromes c_1 are both shown in grey boxes for reference. The CXXCH motifs in other sequences are coloured in dark grey.

```
CY1_YEAST     EAEHGLHAPAYAWSHNGPFETFDHASIRRGYQVYREVCAACHSLDRVAW 132
CYC1A_ARATH   EAEHGLESPEYPWPHDGILSSYDHASIRRGHQVYQQVCASCHSMSLISY 114
D7LLY4_ARALL  EAEHGLECPDYPWPHDGILSSYDHASIRRGHQVYQQVCASCHSMSLISY 114
V4KZ70_EUTSA  EAEHGLECPDYPWPHDGILSSYDHASIRRGHQVYQQVCASCHSMSLISY 114
V4KN96_EUTSA  EAEHGLECPDYPWPHDGILSSYDHASIRRGHQVYQQVCASCHSMSLISY 122
M4EW08_BRARP  EAEHGLACPDYPWPHDGILSSYDHASIRRGHQVYQQVCASCHSMSLISY 114
M4FDB8_BRARP  EAEHGLACPDYPWPHDGILSSYDHASIRRGHQVYQQVCASCHSMSLISY 114
D7MJ34_ARALL  EAEHGLECPNYPWPHEGILSSYDHASIRRGHQVYQQVCASCHSMSLISY 114
CYC1B_ARATH   EAEHGLECPNYPWPHEGILSSYDHASIRRGHQVYQQVCASCHSMSLISY 114
R0GTE4_9BRAS  EAEHGLECPNYPWPHEGILSSYDHASIRRGHQVYQQVCASCHSMSLISY 114
V4NKH0_EUTSA  EAEHGLECPNYPWPHDGILSSYDHASIRRGHQVYQQVCASCHSMSLISY 114
A9PE48_POPTR  EAEHGLECPSYPWPHNGILSSYDHASIRRGHQVYQQVCASCHSMSLISY 117
V4TXZ0_9ROSI  EAEHGLACPSYPWPHKGILSSYDHASIRRGHQVYQQVCASCHSMSLISY 117
B9HPZ7_POPTR  EAEHGLECPSYPWPHKGILSSYDHSSIRRGQQVYQQVCASCHSMSLISY 117
M5WTA2_PRUPE  EAEHGLECPNYPWPHQGILSSYDHASIRRGHQVYTQVCASCHSMSLISY 114
B9RBM7_RICCO  EAEHGLECPNYPWPHKGILSSYDHSSIRRGQQVYQQVCASCHSMSLISY 117
G7KN93_MEDTR  EAEHGLASPHYPWPHEGILSSYDHASIRRGHQVYTQVCASCHSMSLISY 113
M1BST6_SOLTU  EAEHGLECPSYPWPHAGILSSYDHASIRRGHQVYQQVCASCHSMSLVSY 117
A0A022RXR3_M  EAEHGLDAPSYPWPHKGILSSYDHASIRRGHQVYQQVCASCHSMSLVSY 117
C6TIQ7_SOYBN  EAEHGLACPSYPWPHKGILSSYDHASIRRGHQVYTEVCASCHSMSLISY 114
C6TJ17_SOYBN  EAEHGLACPSYPWPHKGILSSYDHASIRRGHQVYTEVCASCHSMSLISY 114
D7UBF7_VITVI  EAEHGLECANYPWPHSGILSSYDHASIRRGHQVYQQVCASCHSMSLISY 117
A0A059AG57_E  EAEHGLECPSYPWPHKGILSSYDHASIRRGHQVYQQVCASCHSMSLISY 115
I3S7F6_LOTJA  EAEHGLSCPSYPWPHEGILSSYDHASIRRGHQVYQQVCASCHSMSLISY 114
K3YIS4_SETIT  EAEHGLAAPDYPWPHAGIMSSYDHASIRRGHQVYTQVCASCHSMSLISY 111
U5GJY0_POPTR  EAEHGLECPSYPWPHNGILSSYDHASIRRGHQVYQQVCASCHSMSLISY 150
M4E904_BRARP  EAEHGLACPDYPWPHDGILSSYDHASIRRGHQVYQQVCASCHSMSLISY 114
V7C3R3_PHAVU  EAEHGLACPSYPWPHKGILSSYDHASIRRGHQVYQEVCASCHSMSLISY 113
I1K147_SOYBN  EAEHGLACPSYPWPHKGILSSYDHASIRRGHQVYQEVCASCHSMSLISY 114
V7B8T3_PHAVU  EAEHGLACPSYPWPHKGILSSYDHASIRRGHQVYTEVCASCHSMSLISY 114
A0A022R3I6_M  EAEHGLDAPNYPWPHKGILSSYDHSSVRRGHQVYQQVCASCHSMSLVSY 117
```

```
Q0DJC3_ORYSJ EAEHGLAAAEYPWPHAGILSSYDHASIRRGHQVYTQVCASCHSMSLISY 111
I1PU78_ORYGL EAEHGLAAAEYPWPHAGILSSYDHASIRRGHQVYTQVCASCHSMSLISY 111
I1I9E4_BRADI EAEHGLAAAEYPWPHAGILSSYDHASIRRGHQVYQQVCASCHSMSLISY 112
C4J0P7_MAIZE EAEHGLAAPDYPWPHAGIMSSYDHASIRRGHQVYTQVCASCHSMSLISY 111
B9FNR2_ORYSJ EAEHGLAAAEYPWPHAGILSSYDHASIRRGHQVYTQVCASCHSMSLISY 124
B8AWB5_ORYSI EAEHGLAAAEYPWPHAGILSSYDHASIRRGHQVYTQVCASCHSMSLISY 124
B6SWR9_MAIZE EAEHGLAAPDYPWPHAGIMSSYDHASIRRGHQVYTQVCASCHSMSLISY 111
CY1_HUMAN    ASDLELHPPSYPWSHRGLLSSLDHTSIRRGFQVYKQVCASCHSMDFVAY 132
CY1_YEAST    EAEHGLHAPAYAWSHNGPFETFDHASIRRGYQVYREVCAACHSLDRVAW 132
I1N7F4_SOYBN EAEHGLACPSYPWPHKGILSSYDHASIRRGHQVYQEVCASCHSMSLISY 114
J3M5T0_ORYBR EAEHGLAAADYPWPHAGILSSYDHASIRRGHQVYTQVCASCHSMSLISY 116
B4FHM1_MAIZE EAEHGLAAPDYPWPHAGIMSSYDHASIRRGHQVYTQVCASCHSMSLISY 111
K7TYI3_MAIZE EAEHGLAAPDYPWPHAGIMSSYDHASIRRGHQVYTQVCASCHSMSLISY 113
I1K146_SOYBN EAEHGLACPSYPWPHKGILSSYDHASIRRGHQVYQEVCASCHSMSLISY 115
G7KN99_MEDTR EAEHGLAAPSYPWPHEGILSSYDHASIRRGHQVYTQVCASCHSMSLISY 114
W5HEN1_WHEAT EAEHGLAAAEYPWPHAGILSSYDHASIRRGHQVYQQVCASCHSMSLISY 111
M8AAU3_TRIUA EAEHGLAAAEYPWPHAGILSSYDHASIRRGHQVYQQVCASCHSMSLISY 111
C6TKJ4_SOYBN EAEHGLAAAEYPWPHAGILSSYDHASIRRGHQVYQEVCASCHSMSLISY 114
W5HSM1_WHEAT EAEHGLAAAEYPWPHAGILSSYDHASIRRGHQVYQQVCASCHSMSLISY 116
Q942X6_ORYSJ EAEHGLEAPHYPWPHAGILSSYDHASIRRGHQVYTQVCASCHSMSLISY 111
I1NV02_ORYGL EAEHGLEAPHYPWPHAGILSSYDHASIRRGHQVYTQVCASCHSMSLISY 111
B8A8T9_ORYSI EAEHGLEAPHYPWPHAGILSSYDHASIRRGHQVYTQVCASCHSMSLISY 111
B6TKU5_MAIZE EAEHGLAAPDYPWPHAGIMSSYDHASIRRGHQVYTQVCASCHSMSLISY 111
W5CHZ4_WHEAT EAEHGLESPSYPWPHEGILSSYDHASIRRGHQVYQQVCASCHSMSLISY 128
K4C7W7_SOLLC EAEHGLECPSYPWPHEGILSSYDHASIRRGHQVYQQVCASCHSMSLISY 127
F2CW75_HORVD EAEHGLAAAEYPWPHAGILSSYDHASIRRGHQVYQQVCASCHSMSLISY 111
CYC1B_ARATH  EAEHGLECPNYPWPHEGILSSYDHASIRRGHQVYQQVCASCHSMSLISY 114
M8AFD2_TRIUA EAEHGLESPSYPWPHEGILSSYDHASIRRGHQVYQQVCASCHSMSLISY 114
CY11_SOLTU   EAEHGLECPNYPWPHEGILSSYDHASIRRGHQVYQQVCASCHSMSLISY 117
M1CSJ8_SOLTU EAEHGLECPSYPWPHEGILSSYDHASIRRGHQVYQQVCASCHSMSLISY 131
K4DF81_SOLLC EAEHGLECPSYPWPHEGILSSYDHASIRRGHQVYQQVCASCHSMSLISY 131
M0RYI1_MUSAM EAEHGLPAPSYPWPHKGILSSYDHASIRRGHQVYQQVCASCHSMSLISY 112
W5DDR2_WHEAT EAEHGLESPSYPWPHEGILSSYDHASIRRGHQVYQQVCASCHSMSLISY 114
W9SAX0_9ROSA EAEHGLECPNYPWPHEGILSSYDHASIRRGHQVYQQVCASCHSMSLISY 114
K4BDI6_SOLLC EAEHGLECPSYPWPHAGILSSYDHASIRRGHQVYQQVCASCHSMSLVSY  76
J3L7L4_ORYBR EAEHGLEAPHYPWPHAGILSSYDHASIRRGHQVYQQVCASCHSMSLISY 135
M0TEY7_MUSAM EAEHGLPAPSYPWPHDGILSSYDHASIRRGHQVYQQVCASCHSMSLISF 115
M0S0X3_MUSAM EAEHGLAAPIYPWPHKGILDSYDHASIRRGHQVFQQVCASCHSMSLISY 114
U5DEW6_AMBTC EAEHGLACPSYPWPHSGILSSYDHASIRRGHQVYQQVCASCHSMSLISY  96
Q41450_SOLTU EAEHGLECPSYPWPHAGILSSYDHASIRRGHQVYQQVCASCHSMSLVSY  67
V4SIG8_9ROSI EAEEGLHCPNYPWPHKGILSSYDHASIRRGHQVYQQVCASCHSMSLISF 130
K3XKU7_SETIT EAEHGLEAPNYPWPHAGILSSYDHASIRRGHQVYQQVCASCHSMSLISY  82
D7U0V9_VITVI EAEHGLACPNYPWPHSGILSSYDHASIRRGHQVYQQVCASCHSMSLISF 123
CY12_SOLTU   EAEHGLECPSYPWPHAGILSSYDHASIRRGHQVYQQVCASCHSMSLVSY  67
A0A022QBN3_M EAEHGLDSPSYPWPHKGILNSYDHASIRRGHQVYQQVCASCHSMSLISY 125
A0A022QAN0_M EAEHGLDSPSYPWPHKGILNSYDHASIRRGHQVYQQVCASCHSMSLISY 121
B9RUZ9_RICCO EAEHGLECPSYPWPHKGILSSYDHASIRRGHQVYQEVCASCHSMSLISY  72
M8ASU5_AEGTA EAEHGLESPSYPWPHEGILSSYDHAISRRGHQVYQQVCASCHSMSL--- 111
```

```
S8E5S1_9LAMI EAEHGLDCPSYPWPHKGILSSYDHSSIRRGHQVYQQVCASCHSMSLVSY  79
A9NX47_PICSI ETEHGLDCPSYPWPHRGILSSYDHSSIRRGHQVFQQVCASCHSMSLVTY 129
A9NMY7_PICSI ETEHGLDCPSYPWPHKGILSSYDHASIRRGHQVFQQVCASCHSMSLVTY 138
U5GMW1_POPTR EAEHGLECPSYPWPHNGILSSYDHASIRRGHQVYQQVCASCHSMSLISY 117
A0A059APG2_E EAEHGLECPNYPWPHKGILSSYDHASIRRGHQVYQQVCASCHSMSLISY 236
A0A059AR22_E EAEHGLECPNYPWPHKGILSSYDHASIRRGHQVYQQVCASCHSMSLISY 237
A0A059AQ95_E EAEHGLECPNYPWPHKGILSSYDHASIRRGHQVYQQVCASCHSMSLISY 237
CY1_HUMAN    ASDLELHPPSYPWSHRGLLSSLDHTSIRRGFQVYKQVCASCHSMDFVAY 132
CY1_YEAST    EAEHGLHAPAYAWSHNGPFETFDHASIRRGYQVYREVCAACHSLDRVAW 132
W9QQ48_9ROSA EAEHGLDSPSYPWPHKGILSSYDPASIRRGHQVYQQVCASCHSMSLISY 104
A9RCI4_PHYPA EAEHGLAAPQYPWSHNGVLSGYDHASIRRGHQVFQQVCAACHSASLVAY 120
A9SAZ2_PHYPA EAEHGLAPPQYPWSHNGVLSGYDHASIRRGHQVFQQVCAACHTASLVAY 128
H6TNN5_ELAGV EAEHGLPCPSYPWPHKGILSSYDHASIRRGHQVYQQVCASCHSMSLISY 117
A9SB01_PHYPA EAEHGLAPPQYPWSHNGVLSGYDHASIRRGHQVFQQVCAACHSASLVAY 128
A0A059AGZ7_E EAEHGLECPSYPWPHKGILT----------GVFQQACASCHSMSLVSF 105
R0H5Y6_9BRAS EAEHGLESPNYPWPHDGILSSYDHASIRRGHQVYQQVCASCHSMSLVSY 114
I1I9E6_BRADI EAEHGLAAAEYPWPHAGILSSYDHASIRRGHQVYQQVCASCHSMSLISY 112
E1ZJQ7_CHLVA EAEHGLHAAAFPWPHEGFFSSYDHRSIRRGFQVYQQVCATCHSLDLVHY 109
C1E790_MICSR EAEHGLHSAAYPWPHEGMFDSYDHASIRRGHQVYQQVCAACHSLNLVAY  67
C1N8D8_MICPC EAEHGLHSPQMPWPHEGMFDSYDHASIRRGHQVYQQVCAACHSLNLVAY  67
I0Z5X3_9CHLO EAELGLHAPHYPWNHAGLFSAYDHASIRRGHQVYSQVCAACHSVNQLHY 136
Q00TB0_OSTTA EAEHGLALPQYAWPHDGVFSSYDHASIRRGHQVYAQVCAACHSLNQIKY  91
R7WAW7_AEGTA EAEHGLAAAEYPWPHAGILSSYDHASIRRGHQVYQQVCASCHSMSLISY 111
A4S995_OSTLU EAEHGLALPRYEWPHDGLFSSYDHASVRRGHQVYTQVCAACHSLNQIAY  58
A0A059LE04_9 EAEHGLAAAQYPWPHEGWFSAYDHASIRRGYQVYTQVCATCHSMSAIHF  91
D8T433_SELML EADHGLPPPSYKWPYNGIFEAQDHASIRRGHQVYAQVCAACHSMQYISY  91
K8ELD3_9CHLO EAEHGLHLPQYSWPHDGLFDAYDHASIRRGHQVYQQVCAACHSLNQICY 137
A9SHR8_PHYPA EAEHGLQAPSYPWSHSGWFSSYDHASIRRGHQVYKEVCAACHAMSLVTY  62
D8S7R5_SELML EADHGLPPPSYKWPYNGFFEAQDHASIRRGHQVYAQVCAACHSMQYISY 103
Q9FQ96_CHLRE EAADGLHAPHYPWGHEGVLDSYDHAAIRRGHKVYQQVCAACHSMQYLHW 121
M0V9E6_HORVD ----------------------------------------MSLISY   6
F2DK74_HORVD ----------------------------------------MSLISY   6
D8UA93_VOLCA EVADGLHAPHYPWPHEGVLDSYDHAAIRRGHKVYQEVCAACHSMQYLHW 123
H8PGG2_9CHLO EAGDGLHPVSYPWSHEGMASSFDHSAIRRGHQVYQQVCAACHSMNYTHW 116
B7FJZ7_MEDTR EAEHGLASPHYPWPHEGILSSYDHASIRRGHQVYTQVCASCHSMSLISY 113
I1I9E5_BRADI EAEHGLAAAEYPWPHAGILSSYDHASIRRGHQVYQQVCASCHSMSLISY 112
A8JIU5_CHLRE EAADGLHAPHYPWGHEGVLDSYDHAAIRRGHKVRGALRARCRENWQVAW 121
B7ELQ3_ORYSJ EAEHGLEAPHYPWPHAGILSSYDHASIRRGHQVYTQVCASCHSMSLISY 111
Q2VU88_9ASPA ------------------------------------------------   0
B4FV18_MAIZE ASDLELHPPSYPWSHRGLLSSLDHTSIRRGFQVYKQVCSSCHSMDYVAY 133
C4IZW4_MAIZE EAEHGLAAPDYPWPHAGIMSSYDHASIRRGHQVYTQVCASCHSMSLISY 111
H9X295_PINTA ----------------------------VFQQVCASCHSMSLVTY  17
M1CSJ7_SOLTU EAEHGLECPSYPWPHEGILSSYDHASIRRGHQVYQQVCASCHSMSLISY 131
W1PD74_AMBTC EAEHGLECPSYPWPHEGILSSYDHASIRRGHQVYQQVCASCHSMSLIAY 113
I6QZ52_9ORYZ ----------YPWPHAGILSSYDHASIRRGHQVYTQVCASCHSMSLISY  39
CY1_HUMAN    ASDLELHPPSYPWSHRGLLSSLDHTSIRRGFQVYKQVCASCHSMDFVAY 132
```

Figure 6.6 Multiple sequence alignment of immediate N-terminal sequences to CXXCH motif of *Paracoccus denitrificans*, *Rhodobacter sphaeroides*, *Brucella abortus*, *E. caballus* (horse), *Bos mutus* (cow), *Homo sapiens* (human), *S. cereveisiae* (baker's yeast), *C. elegans* and *Pseudomonas aeruginosa*) and cytochromes c_1. The conserved residues are shown in grey boxes. The sequence of horse from the database has no histidine in the position within the CXXCH motif. This is probably due to a sequencing error.

```
A1B4F4_PARDP  AQEAGDSHAAAHIEDISFSFEGPFGKFDQH-QLQRGLQVYTEVCSACH 249
CY1_RHOSH     AMAAGG----GHVEDVPFSFEGPFGTFDQH-QLQRGLQVYTEVCAACH  62
N8AKH9_BRUAO  NPAEPDHFPIHHPKQESWSFAGPFGTYDKG-QLQRGLKVYKEVCSACH  79
F6VRY7_HORSE  HSAVSASDLELHPPSYPWSHRGLLSSLDHTRSSGRGFQVYKQVCSSC-  75
CY1_BOVIN     HSAVSASDLELHPPSYPWSHRGLLSSLDHT-SIRRGFQVYKQVCSSCH 125
CY1_HUMAN     HSAVSASDLELHPPSYPWSHRGLLSSLDHT-SIRRGFQVYKQVCASCH 125
CY1_YEAST     AEAMTAAEHGLHAPAYAWSHNGPFETFDHA-SIRRGYQVYREVCAACH 125
Q18853_CAEEL  ENSVSASGDNVHPYALPWAHSGPFSSFDIA-SVRRGYEVYKQVCAACH  78
Q02H11_PSEAB  ----------PQLDHVDIDLT------DKA-AMQDGARTFANYCMGCH  55
```

It is notable that whereas bacterial cytochromes *c* do not have N-terminal sequences resembling those of mitochondrial cytochromes, at least some bacterial cytochromes c_1 do resemble their mitochondrial counterparts. Figure 6.6 shows the multiple sequence alignment of some bacterial cytochrome c_1 sequences with the cytochrome c_1 sequences from the genomes of some eukaryotic organisms. The similarity between the N-terminal residues next to the CXXCH motif suggests that these recognition features for HCCS/HCC_1S were already in bacterial cytochrome c_1 before the rise of different cytochrome *c* biogenesis systems.

Figure 6.7 Multiple sequence alignment of 22 N-terminal sequences of mitochondrial cytochrome c_1 and other sequences in the result of BLASTp from euglenoid organisms. The highly conserved tyrosine residue and the FXXCH motif of the first sequence in the alignment are coloured in green and cyan respectively as an example. The sequences of *S. cerevisiae* and *H. sapiens* cytochromes c_1 are both shown in grey boxes for reference. The CXXCH and FXXCH motifs in other sequences are coloured in dark grey.

```
CY1_YEAST     HGLHAPAYAWSHNG---PFETFDHASIRRGYQVYREVCAACHSLDRVAW 112
O97134_9TRYP  AHPIKRDWYWNHNDRFEIWHSLDWPSVRRGRQIYTEVFAPCHPLGRMTF  54
D6XKS6_TRYB   AHPIKRDWYWNHNDRFEIWHSLDWPSVRRGRQIYTEVFAPCHPLGRMTF  54
C9ZUW4_TRYB   AHPIKRDWYWNHNDRFEIWHSLDWPSVRRGRQIYTEVFAPCHPLGRMTF 148
V5D9C0_TRYCR  AHPIKRDWYWNHNDRFEIWHSLDWPSVRRGRQIYTEVFAPCHPLGRMTF  54
K4DMD2_TRYCR  AHPIKRDWYWNHNDRFEIWHSLDWPSVRRGRQIYTEVFAPCHPLGRMTF  54
Q4DU59_TRYCC  AHPIKRDWYWNHNDRFEIWHSLDWPSVRRGRQIYTEVFAPCHPLGRMTF  54
G0U0C2_TRYVY  AHPIKRDWYWNHNDRFEIWHSLDWPSVRRGRQIYTEVFAPCHPLGRMTF  54
G0URF3_TRYCI  AHPIKRDWYWNHNDRFEIWHSLDWPSVRRGRQIYTEVFAPCHPLGRMTF 171
K2NDI2_TRYCR  AHPIKRDWYWNHNDRFEIWHSLDWPSVRRGRQIYTEVFAPCHPLGRMTF 118
E9B917_LEIDB  AHPIKRDWYWNHNDRFEIWHSLDWPSVRRGRQIYTEVFAPCHSLGRMTF  54
S9WBX7_9TRYP  AHPIHRDWYWNHNDRFEIWHSLDWPSVRRGRQIYTEVFAPCHSLGRMTF  54
A4HT63_LEIIN  AHPIKRDWYWNHNDRFEIWHSLDWPSVRRGRQIYTEVFAPCHSLGRMTF  54
Q4QIT8_LEIMA  AHPIKRDWYWNHNDRFEIWHSLDWPSVRRGRQIYTEVFAPCHSLGRMTF  54
S9WBW1_9TRYP  AHPIHRDWYWNHNDRFEIWHSLDWPSVRRGRQIYTEVFAPCHSLGRMTF  54
A4H4U9_LEIBR  AHPIKRDWYWNHNDRFEIWHSLDWPSVRRGRQIYTEVFAPCHSLGRMTF  54
E9AL16_LEIMU  AHPIKRDWYWNHNDRFEIWHSLDWPSVRRGRQIYTEVFAPCHSLGRMTF  54
S9WSZ8_9TRYP  ------------------------------------------MTF   3
S9VUH4_9TRYP  ------------------------------------------MTF   3
Q26325_CRIFA  AHPIKRDWYWNHNDRFEIWHSLDWPSVRRGRQIYTEVFAPCHSLGRMTF  54
Q26326_9EUGL  AHPVHRDWYWEHNDHWGVWKSLDWPSVRRGRQVYAEVFAPCHPLGKLTF  54
CY1_EUGGR     SHPPALPW-----PHFQWFQGLDWRSVRRGKEVYEQVFAPCHSLSFIKY  47
CY1_HUMAN     LELHPPSYPWSHRG---LLSSLDHTSIRRGFQVYKQVCASCHSMDFVAY 132
```

A BLAST global search was also performed using the *T. brucei* cytochrome c_1, leading to a search result of 21 hits in phylum *Euglenozoa*, as shown in Figure 6.7. The result is also similar to what has been discussed in Section 6.2.1, since all of these sequences contain a tyrosine residue at the conserved position in the N-terminus (Figure 6.7). So far no cytochrome c_1 sequence in phylum *Euglenozoa* in the UniProt database has a unique residue (e.g. phenylalanine) at this conserved position in N-terminal Region. The presence of the tyrosine, believed to be important for the processing by HCCS, is striking given that these organisms do not utilise HCCS for heme attachment to cytochromes *c* and c_1.

Finally, a comparison of the N-terminus of human cytochrome *c* with the immediate N-terminal residues of human cytochrome c_1 is shown in Figure 6.8. As human HCCS processes the heme attachment for both cytochrome *c* and cytochrome c_1 [35], the residues at the positions depicted in grey boxes are presumably recognition points by human HCCS.

Figure 6.8 Pairwise sequence alignment of immediate N-terminal sequences to CXXCH motif of human cytochrome *c* and cytochromes c_1. The residues on the equivalent positions and sharing similar characterisitics are shown in grey boxes. The CXXCH motifs in other sequences are coloured in dark grey.

```
CYC_HUMAN ------------------------------MGDVEKGKKIFIMKCSQCH  19
CY1_HUMAN AMALHSAVSASDLELHPPSYPWSHRGLLSSLDHTSIRRGFQVYKQVCASCH  125
```

However, it should be noted that most of the sequences listed in the sequence alignments in Figure 6.1, Figure 6.3, Figure 6.4, Figure 6.5 and Figure 6.7 are only confirmed in the bioinformatics level, and a few of them have evidence on the transcriptomic level. For some sequences with a very long domains that are similar to sequences of other protein families, the possibility remains that they might lose the function of the cytochromes and may play different roles in the cell. Unfortunately, there is no sequence information available for *D. papillatum* cytochrome c_1; it would be interesting to know if it has a Y, F or A at the critical position mentioned in previous sections.

6.2.4 Sequence analysis of HCCS and HCC_1S in different species

As mentioned in Chapter 1, the structure, taxonomy and mechanism of HCCS (and its isomer HCC_1S in some organisms) remains an enigma. Since in yeast, the HCCS is able to mature both cytochrome *c* and cytochrome c_1, and the HCC_1S can only mature cytochrome c_1 [79], this difference in substrate specificity may be represented in the amino acid sequences of the two proteins. Therefore, a sequence analysis of both of these two enzymes in different organisms is necessary to give insight to this problem.

6.2.4.1 Sequence analysis of HCCS and its isoforms in different species

In order to investigate the variation of substrate specificity of HCCS and HCC_1S in System III in various organisms, BLAST searches were carried out. The two query sequences were

chosen: for *S. cerevisiae* HCCS (UniProt entry P06182) and HCC_1S (UniProt entry Q00873) because in the present study, HCCS from *S. cerevisiae* was used.

The sequence alignments in Figures 6.9, 6.10 and 6.11 were generated from Global sequence searches by BLASTp programme for *S. cerevisiae* HCCS and HCC_1S, as well as *E. caballus* HCCS with the 20 sequences with the highest score respectively. The sequences in Figures 6.9 and 6.10 are all from fungal organisms (such as *Issatchenkia orientalis* and *Ogataea parapolymorpha*), and the sequences in Figure 6.11 are all from animals. The sequence of human HCCS (Uniprot entry P53701) was included in the sequence alignments too.

According to the results shown in Babbitt *et al.* [172] , the following highly conserved human HCCS residues Trp118, Tyr120, Pro121, Met126, Asn128, His154 Asn155, Gln159, Trp162 and Arg217 have been identified as important in the substrate recognition by HCCS in holocytochrome *c* synthesis. The positions of these residues are highlighted in yellow in the sequence alignments in Figures 6.9, 6.10 and 6.11.

These residues listed above are highly conserved in the HCCS sequences, and they are highlighted in the alignment in Figures 6.9 and 6.11. Similarly, most of these residues are also highly conserved in the HCC_1S sequences listed in the alignment in Figure 6.10 (and highlighted too), save the position of Asn128, which can be glutamine, aspartate and glutamate. Therefore, it is possible that the variation of the R-group of the residue at this position in HCC_1S proteins from different organisms is one of the explanations why the

HCC_1S protein functions differently from the HCCS in System III. However, the HCC_1S of *O. parapolymorpha* appears to have Asn at this position. This renders this suggestion less convincing.

Figure 6.9 Multiple sequence alignment of *S. cerevisiae* HCCS and the HCCS sequences from the BLASTp result from its sequence (*I. orientalis* (A0A099P019), *O. parapolymorpha* (W1QH41), *Cyberlindnera fabianii* (A0A061ARF0), *Wickerhamomyces ciferrii* (K0KLR2), *Candida albicans* (A0A0A6KFG2), *Lodderomyces elongisporus* (A5E282), *Scheffersomyces stipitis* (A3LZC6), *Spathaspora passalidarum* (G3ANH2), *Pichia sorbitophila* (G8XZ88), *Meyerozyma guilliermondii* (A5DJC2), *Pichia pastoris* (C4R291), *Kazachstania africana* (H2B1A5), *Lachancea thermotolerans* (C5DG01), *Kluyveromyces marxianus* (W0TK60), *Eremothecium cymbalariae* (I6NCP3), *Ashbya gossypii* (M9MZH1), *Zygosaccharomyces bailii* (W0VT48), *Torulaspora delbrueckii* (G9A078), *Tetrapisispora phaffii* (G8C0Y7) and *Naumovozyma castellii* (G0V768)), as well as *E. caballus* and *H. sapiens*. Important residues according to Babbitt *et al.* [172] is highlighted in yellow, and the *S. cerevisiae* , *E. caballus* and *H. sapiens* HCCS sequences are coloured in light grey. The stretch of the sequences of the N-terminal and the C-terminal tails are long and have lower similarities, therefore they are omitted from this figure.

```
P06182       ELAASK-QPGQKMDLPVDRTISSIPKS----PDSNEFWEYPSPQQMYNAM 104
A0A099P019   YDISVQKQPGQKIDLPTDRTLSTIPRGL---DTKEGVWEYPSPQQMFNAM 78
W1QH41       YVLPTDKVPGQKIDLPTERTLSSIPRGD---VPDQGVWEYPSPQQMFNAM 100
A0A061ARF0   HVISASKLPGQKLDLPTERTFSTIPRGT---DEGEGVWEYPSPQQMFNAM 115
K0KLR2       SFISSSKLPGQKLDLPTDRTFSTIPRGE---DEDEGVWEYPSPQQMFNAM 125
A0A0A6KFG2   MAISSERAPGQRIKLSTERTISTIPRGE---SEDQGLWEYPSPQQMLNAM 106
A5E282       MAISSELAPGQKIKLSTERTISTIPRGE---NDDQGLWEYPSPQQMLNAM 127
A3LZC6       MAISSERLPGQKVSLSTERTISTIPRGE---SDDQGLWEYPSPQQMLNAM 93
G3ANH2       VNLSSQRAPGQKLALSTTRTFSTIPRGE---TEDNGVWEYPSPQQMLNAM 102
G8XZ88       MDLSSERAPGQKITLSTDRTISSIPRGT---SEDQGLWEYPSPQQMLNAM 123
A5DJC2       MDISSERAPGQKVVLPTERTVSSIPRGE---SADMGFWEYPSPQQMLNAM 108
C4R291       YDLSSARAPGQKISLPTQRTLSTIPKGE---TDAQGVWEYPSPQQMLNAM 112
```

```
H2B1A5          HDLTKV-HEGQKLDLTTQRTISSIPKG----SEKGK-WEYPSAQQMYNAM 95
C5DG01          KNLSHEKHIGQKLELPKDRTTSSIPKG---NPGEEEFWEYPSPQQMYNAM 107
W0TK60          VGLTHAKQAGQKLDLPTERTVSSIPKG---KASDEEFWEYPSPQQMYNAM 119
I6NCP3          SFISDRKQPGQKLDLPTCRTTSSIPKGSCDQMGKEEFWEYPSPQQMYNAM 123
M9MZH1          NLSAHR-QAFQKLELPTERTVSSIPKGNTDAAGEGEFWEYPSPQQMYNAM 114
W0VT48          LMISAAKQHGQTLDLPTERTISSIPKG----ADPDENWEYPSPQQMYNAM 107
G9A078          NFLSTKRQPGQKLNLPTDRTVSSIPKG----EDDQSNWEYPSPQQMYNAM 106
G8C0Y7          KALPVGKQEGQKLDLPQERTVSSIPKG----SDSTEFWEYPSPQQMYNAM 120
G0V768          FSISSDKHSTQRLELPKDRTISSIPRG----SDPSQNWEYPSPQQMYNAM 111
P53701          PP-NQTPAPDQPFALSTVREESSIPRAD-----SEKKWVYPSEQMFWNAM 130
F6YF82          PPANQTPAPDQPFPLSTVREESSIPRAD-----SEKKWVYPSEQMFWNAM 134
                          * . *.  *  *:**:.          * *** * : ***

P06182          VRKGKIGGSG--EVAEDAVESMVQVHNFLNEGCWQEVLEWEKPHTDESHV 152
A0A099P019      IRKG-----KGD-VPEDAVESMVSIHNFLNEGAWEEIEEWEKPYTEQTHI 122
W1QH41          IRKG-----KGD-VPEDAVESMVNIHNFLNEGAWSEILKWEKPHTEQTKQ 144
A0A061ARF0      VRKG-----KAEDIPEDAVESMVDVHNFLNEGAWQEILEWEKPYVEQYGK 160
K0KLR2          MKKG-----KGDGIPEDAVESMVDVHNFLNEGAWQEILEWEKPYESLYNK 170
A0A0A6KFG2      LSKG-----KGDGVPEDAVESMVEVHNFLNEGAWQQILTWEDQYTQQTKV 151
A5E282          LAKG-----KGDGIPEDAVESMVEVHNFLNEGAWQQILEWEDKYTQETKV 172
A3LZC6          VRKG-----KAKDIPEDAVESMVDVHNFLNEGAWQQILDWEDEYTKETKI 138
G3ANH2          LRKG-----KGDGIPEDAVESMVDVHNFLNEGAWQQILDWENKYTLETKV 147
G8XZ88          RRKG-----KGDGIPEDAVESMVEVHNFLNEGAWQQILDWERKYTQQTRV 168
A5DJC2          LRKG-----KGSGIPEEAVESMVGVHNFLNEAAWQQISEWEAPYTQSTNV 153
C4R291          LRKG-----KGE-IAEDAVESMVDVHNFLNEGAWQQILAWEKRHTEENKV 156
H2B1A5          MRKGKINGN---EGDEDTIESMVQIHNFLNESCWQEILRWENRYVQRTNV 142
C5DG01          IRKGKIDPNTGEEIPEDAVESMVFVHNFLNEGCWQEILDWEKKYSEQTEM 157
W0TK60          VRKGKIDPYTGEEIPEDAVESMVFVHNFLNEGCWQEILDWEKKYVQQTQS 169
I6NCP3          VRKGKVDPNTGEEIPEDAVESMVFVHNYLNEGCWEEILDWEKPHIEATAR 173
M9MZH1          VRKGKIDPYTGEEIPEDAVESMVYVHNYLNEGCWEEILDWERRYSEAAAA 164
W0VT48          LKKGKIDPNTGEEIPEDAVESMVQVHNFLNEGCWQEVLEWEKPHTEQTHV 157
G9A078          VRKGKIDPNTGEEIPEDAVESMVFVHNFLNEGCWEEVLEWEKPYTEQTKI 156
G8C0Y7          IRKGKIDPNTGEEIPEDAVESMVFVHNFLNEGCWQEVLKWEKPYTDQTNV 170
G0V768          VRKGKIDQNTGEEIPEDAVESMVYVHNFLNEGSWEEILKWEKPYTEQNQT 161
P53701          LKKG--WKWKDEDISQKDMYNIIRIHNQNNEQAWKEILKWEALHAAECPC 178
F6YF82          LRKG--WKWKEEDISQKDMYNIIRIHNQNNEQAWQEILKWEALHAAECPC 182
                  **          :. : .:: :**  ** .*.::  **  :

P06182          QPKLLKFMGKPGVLSPRARWMHLCGLLFPSHFSQELPFDRHDWIVLRGER 202
A0A099P019      EPRLLKFTGRPDALSPRARYYNFMSKIFPDKYGTELPFDRHDWTVLRN-- 170
W1QH41          EPRLLKFTGRPHDLTPRARFYQIMGKLFPDKYATEPPFDRHDWTVLRG-- 192
A0A061ARF0      MPRLEQFTGRPHDLSPRARMYLFLSKFFPETFNTEPPFDRHDWYVLRP-- 208
K0KLR           APRLAQFTGRPHDLSPRARMYLQLSKIFPKTFNTDPPFDRHDWYVLRS-- 218
A0A0A6KFG2      EPRLKKFTGRPHDLSPKARMYLWLGQLFPETFNTIPPFDRHDWTVLRS-- 199
A5E282          EPRLKKFTGKPHDLSPKAKMYLWLGSLFPDTFNTVPPFDRHDWTVLRS-- 220
A3LZC6          EPRLKKFTGRPNDLSPRAQMYLWLGKLFPDTFNTQPPFDRHDWTVLRS-- 186
G3ANH2          EPRLKQFTGKPHDLSPRAQLFLWLGKFFPEQFNTQPPFDRHDWTVLRS-- 195
G8XZ88          EPRLLKFTGRPNDLSPRAAMYLGLGKIFPETFNTQPPFDRHDWTVLRS-- 216
```

```
A5DJC2      SPRLLKFTGRPHDLSPRARMFLALGKLFPETFNTEPPFDRHDWTVLRS-- 201
C4R291      EPRLLKFTGRPYDLSPRAAMYQYLGKLFPSHFSTQPPFDRHDWTVLRSKN 206
H2B1A5      QPKLLKFMGKPDNLSPRARMHYYLSYIFPNTFARELPFDRHDWLVLRGDP 192
C5DG01      EPKLLQFMGKPSQLSPRARCFSILGTLFPSKFSAELPFDRHDWVVLRPDP 207
W0TK60      YPKLLQFTGKPDQLSPRARCFNILGTLFPGHFSSELPFDRHDWVVLRPDP 219
I6NCP3      YPKLLKFMGKPSQVTPRARWYHLLGTLFPSKFCGELPFDRHDWIVLRPDP 223
M9MZH1      HPKLLQFMGKPSQLTPRARWYHLLGTLFPSKFSAELPFDRHDWTILRPDP 214
W0VT48      EPKLLKFTGKPGQLSPRARWYGILSKLFPKTFDGNPPFDRHDWIVLRADP 207
G9A078      QPKLLKFMGKPHEMSPRARYYQMMSKVFPKQYSGEPPFDRHDWTVLRADS 206
G8C0Y7      EPKLLRFQGKPNEMSPRAKFYNTLGKLFPNYFSSDPPFDRHDWVVLRADP 220
G0V768      MPKLLKFTGRPDSLSPKARWHHFLSNIFPSKYSSELPFDRHDWVVLRGEP 211
P53701      GPSLIRFGGKAKEYSPRARIRSWMG--------YELPFDRHDWIINRCG- 219
F6YF82      GPSLIRFGGKAKEYSPRARIRSWMG--------YELPFDRHDWIINRCG- 223
                * * :* *:.   :*:*      .           ******* : *
```

Figure 6.10 Multiple sequence alignment of *S. cerevisiae* HCC_1S and the HCC_1S sequences from the BLASTp result from its sequence (*E. cymbalariae* (G8JRX1), *Z. bailii* (W0W7V2), *Vanderwaltozyma polyspora* (A7TR56), *Kluyveromyces dobzhanskii* (A0A0A8LDM2), *L. thermotolerans* (C5DDH8), *T. phaffii* (G8C1C3), *T. delbrueckii* (G8ZQH3), *Ashbya aceri* (R9XCI3), *N. castellii* (G0VJG5), *Hanseniaspora uvarum* (A0A0F4X6C8), *O. parapolymorpha* (W1QF50), *C. fabianii* (A0A061BK07), *P. pastoris* (C4R295), *Komagataella pastoris* (F2QRH3), *C. albicans* (C4YHN3), *W. ciferrii* (K0KHV4), *Debaryomyces hansenii* (Q6BHJ2), *S. passalidarum* (G3AH05), *Kuraishia capsulata* (W6MR39), *S. stipitis* (A3GGM8)), as well as *E. caballus* HCCS and *H. sapiens* HCCS. Important residues proposed [172] are highlighted in yellow, and the *S. cerevisiae* HCC_1S and *H. sapiens* HCCS sequences are coloured in light grey. The stretch of the sequences of the N-terminal and the C-terminal tails are long and have lower similarities, therefore they are omitted from this figure.

```
Q00873      --QLECSANPQDN-DKTPEYHTTVDLSQSREVSTIPRTNSD-RNWIYPSE 83
A7TR56      --DIECSSD--AI-PAHPVYKTDVKLPDHRELSSIPRTGAS-SNWVYPSE 152
G8C1C3      --DVECSSD--EI-PQTPVYKTDVKLPDEREVSSIPRTGHG-SNWIYPSE 136
G0VJG5      --PLECSSETLKQ-PTP---------ASNREISTIPRTGTE-QNWIYPSE 80
A0A061BK07  PTAIEHSSDTLP---DQPKYVTDVALPEDREISSIPRTGSD-SNWIYPSQ 144
K0KHV4      TSNTNHSSENLSN--DPPIYLTNTPLPEDREISSIPRTGAD-ANWIYPSQ 129
W6MR39      EPGESCSSGTLQ---NTPAFTSTVSLPTDREISTIPRTGAG-ENWVYPSQ 126
C4R295      -----CSSDALP---EQPTYSSNIDLPTDREISSIPRTGAE-LNWIYPSQ 135
F2QRH3      -----CSSDALP---EQPTYSSNIDLPTDREISSIPRTGAE-LNWIYPSQ 135
C4YHN3      ----TCDSSKIP-NTLEDTTTTNIDLPGERELSSIPRTSSN-TNWIYPSQ 116
G3AH05      ----GCSSNAVN-QDLT-TAGAPSDLATEREISSIPRTSAD-SNWIYPSQ 125
```

```
Q6BHJ2       ---T-CSSDKLDTSAAHIS-DSNSKLPTEREISSIPRTSGQ-SNWIYPSQ 116
A3GGM8       ---EGCSSDTLSSVAADSAIASNVDLPTERATSSIPRTAAG-SNWVYPSQ 118
W1QF50       ----TCSSDDLD---SPKPIVVDVDLPEEREVSSIPRTGAQ-SNWIYPSQ 147
G8JRX1       -EEVECSSDKLL---TRPEYSVDVNLPTEREVSSIPRTGTT-HNWVYPSE 55
R9XCI3       -EDVECSSDKLA---DMPHYSTDVPLPTEREQSSIPRTGTE-HNWVYPSQ 114
A0A0A8LDM2   TEDIECSSDEIP---AQPNYTTTVNLPTEREKSSIPRTGSS-DNWVYPSE 127
W0W7V2       GEAVECSSDELP---TIPKYNTDVELPENREVSSIPRTGAQ-ANWVYPSE 123
G8ZQH3       -EDVECSSDQLP---EVPQYHSDVKLPEKREISSIPRTGSN-SHWVYPSE 101
C5DDH8       PEAVECTSDQIP---NTPQYKTDVKLPTEREVSSIPRTGTEGQHWVYPSQ 106
A0A0F4X6C8   DVELSCDSGKLSN--KAPTYKTDVNLSEDREISSIPRTNAD-GNWIYPSQ 183
P53701       NKENLDPSNLMPPPNQTPAPDQPFALSTVREESSIPRADSE-KKWVYPSE 123
P06182       DNDRINPLNNMPELAASKQPGQKMDLPVDRTISSIPKSPDSNEFWEYPSP 97
                       :                    .  *  *:***:     :*:***:

Q00873       KQFYEAMMKKNWDPNSD-----DMKVVVPLHNSINERVWNYIKSWEDK-- 126
A7TR56       KQFYEAMVRKNWDPEAD-----DMKVVVPIHNQVNERVWNYIKRWEEN-- 195
G8C1C3       KQFYEAMIRKKWNPEAD-----DMKTVVPLHNSVNERVWNCIKNWETA-- 179
G0VJG5       SQFYQAMKRKNWNPDLK-----DMQTVVPLHNSINERVWHYIKNWEKD-- 123
A0A061BK07   KQFFEAMKRKNWNPEAD-----VMAAVVPIHNAVNERAWYQILKWEEG-- 187
K0KHV4       KQFFEAMKRKNWDPNAS-----DMGTVVPIHNAVNERAWYHILKWEEG-- 172
W6MR39       KQFFEAMQRKKWNPEAQ-----DMKSIVPIHNAVNERAWYHIRLWELG-- 169
C4R295       KQFFEAMRRKDWNPHAK-----DMKSIVPIHNAVNEQAWKYIQMWEHG-- 178
F2QRH3       KQFFEAMRRKDWNPHAK-----DMKSIVPIHNAVNEQAWKYIQMWEHG-- 178
C4YHN3       KQFFEAMKRKNWDPQQE-----DMKVIVPIHNLVNERAWKHILMWEKPYA 161
G3AH05       KQFFDAMKRKQWNPEKE-----DMKTVVPIHNAVNEQAWRHIQMWEAPYA 170
Q6BHJ2       KQFFEAMKRKNWEPEAQ-----DMKTVVPIHNAVNEKAWSHILNWERSHY 161
A3GGM8       KQFYDAMKRKKWNPEAK-----DMETIVPIHNAVNERAWIHILNWERNNY 163
W1QF50       KQFFNAMRRKDWNPDAG-----DMKTIVPIHNMVNERAWAYILMWENG-- 190
G8JRX1       KQFYEAMLRKNWDPEVQ-----DMKAVIPIHNTVNERVWSYIKSWEKD-- 98
R9XCI3       KQFFEAMLRKNWNPHAE-----DMKTVIPLHNSVNERVWNYIKLWENG-- 157
A0A0A8LDM2   KQFFEAMKRKNWDPEAV-----DMKAVVPIHNSVNERVWNYIKIWENN-- 170
W0W7V2       KQFYEAMLRKKWDPNAI-----DMRTVVPLHNSVNERVWNCIKTWEKD-- 166
G8ZQH3       KQFFEAMLRKKWDPHSE-----DMKTVVPLHNTVNERVWNSIKLWEQG-- 144
C5DDH8       KQFFEAMLRKQWNPNSD-----DMKTVVPIHNQVNERVWNYIRLWEKG-- 149
A0A0F4X6C8   KQFYEAMVRKEFNPEAQ-----DMKTIIPIHNNINERCWNFIKNWEKG-- 226
P53701       QMFWNAMLKKGWKWKDEDISQKDMYNIIRIHNQNNEQAWKEILKWEAL-- 171
P06182       QQMYNAMVRKGKIGGSGEVAEDAVESMVQVHNFLNEGCWQEVLEWEKP-- 145
             . *::** :* :. .        *  :: :**  **: *  *  **

Q00873       -QGGEACGGIKLTNFKGDSKKLTPRAWFRSRILH--------LAKPFDRH 167
A7TR56       -EDSESCGGLKLTSFKGDSKKLTPRAWIRSTLLG--------YSKPFDRH 236
G8C1C3       -EEND----LKLTSFKGDAKKVTPRAWIRSNILG--------LSKPFDRH 216
G0VJG5       -AGGDACGGIKLSTFKGDSHKWTPRAWIRYNIFQ--------MSKPFDRH 164
A0A061BK07   -EGSEKCGGIQLTSFKGDSKKLTPKAALKWALLG--------YQKPFDRH 228
K0KHV4       -QGSDECGGVKLTSFKGDSKALTPKARLK-LLLG--------YSKPFDRH 212
W6MR39       -QGGDKCGGIQLTSFKGESKKLTPRARMK-CFFG--------YEKPFDRH 209
C4R295       -KGGQSCGGIQLTSFKGDSKKLTPRARWRL-FLG--------YERPFDRH 218
F2QRH3       -KGGQSCGGIQLTSFKGDSKKLTPRARWRL-FLG--------YERPFDRH 218
C4YHN3       EDTQQKCGGITLTSFKGDSKKLTPRAWLKS-IFG--------YDKPFDRH 202
```

```
G3AH05       EETHQKCGGVSLTSFKGDSKKLTPRAWFKSSVLG--------YAKPFDRH 212
Q6BHJ2       QQSLAQCGGIKLTSFKGDSKKLTPRAWFNSTILG--------YEKPFDRH 203
A3GGM8       DEAVKMCGGITLTSFKGDSKKLTPRAWINSTFLG--------YQKPFDRH 205
W1QF50       -QGGEKCGGVQLTSFKGDSKKITPRARIRNLIFG--------ADLPFDRH 231
G8JRX1       -QGGDACGGIKLTSFKGNSKQLTPRAWFRSTILG--------LSKPFDRH 139
R9XCI3       -RGGESCGGIKLTSFKGSAKDLTPRAWFRSTVLG--------YTPPFDRH 198
A0A0A8LDM2   -QGGDECGGIKLTSFKGDSKKLTPRAWFRSSILG--------MNKPFDRH 201
W0W7V2       -QGGDACGGVQLTSFKGDSKKLTPRAWFRSTILG--------LSKPFDRH 197
G8ZQH3       -QGGESCGGIQLTSFKGDAKKLTPRAWFRSAILG--------FSKPFDRH 185
C5DDH8       -QGGEECGGVQLTSFKGDSKKLTPRAWFRSSILG--------YSKPFDRH 190
A0A0F4X6C8   -LGGDTCGGIKLTSFKGDSKKMTPRAMLRTYILG--------YEKPFDRH 267
P53701       -HAAECPCGPSLIRFGGKAKEYSPRARIRS-WMG--------YELPFDRH 201
P06182       -HTDESHVQPKLLKFMGKPGVLSPRARWMH-LCGLLFPSHFSQELPFDRH 193
                          *  * *.::  :*:*  .   :           *****

Q00873       DWQIDRCG 175
A7TR56       DWTINRCG 244
G8C1C3       DWTINRNG 224
G0VJG5       DWQIDRCG 172
A0A061BK07   DWVIDRCG 236
K0KHV4       DWTINRCG 220
W6MR39       DWTINRCG 217
C4R295       DWTIDRCG 226
F2QRH3       DWLINRCG 210
C4YHN3       DWTVDRCG 220
G3AH05       DWIIDRCG 211
Q6BHJ2       DWKIDRCG 213
A3GGM8       DWIVDRCG 239
W1QF50       DWIVDRCG 147
G8JRX1       DWRVDRCG 206
R9XCI3       DWVVDRCG 219
A0A0F4X6C8   DWTVNRCG 215
G8ZQH3       DWTVNRCG 193
C5DDH8       DWTVNRCG 198
A0A0F4X6C8   DWTVNRCG 275
P53701       DWIINRCG 219
P06182       DWIVLRGE 201
             ** ::* *
```

It should be noted that there might be other unspecified elements in the amino acid sequence of HCC_1S that are important in substrate recognition of cytochrome c_1. The reference mentioned [172] was only concerned with the maturation of cytochrome *c* by

human HCCS, and the maturation of cytochrome c_1 by human HCCS was not investigated. Therefore, critical residues in human HCCS for the production of holocytochrome c_1 may be overlooked in the list of important residues and the mechanism proposed [172]. Future studies on the maturation of holocytochrome c_1 are required to give insight to answer this question.

Figure 6.11 Multiple sequence alignment of *E. caballus* (horse) HCCS and the HCCS sequences from the BLASTp result from its sequence (*Ailuropoda melanoleuca* (giant panda, D2H0T7), *C. familiaris* (dog, E2R7M2), *Ictidomys tridecemlineatus* (thirteen-lined ground squirrel, I3MM59), *Pteropus alecto* (black flying fox, L5L251), *Neovison vison* (American mink, U6DXA2), *Felis catus* (cat, M3XCQ6), *Myotis brandtii* (Brandt's bat, S7NJM1), *Callithrix jacchus* (white-tufted-ear marmoset, F6XSM6), *Bos mutus* (yak, L8I4R6), *Pan troglodytes* (chimpanzee, K7DKZ2), *Desmodus rotundus* (vampire bat, K9IX17), *Mus musculus* (P53702), *Sarcophilus harrisii* (Tasmanian devil, G3WAQ4), *Nipponia nippon* (crested ibis, A0A091WFQ9), *Meleagris gallopavo* (common turkey, G1N619), *Xenopus laevis* (African clawed frog, Q7ZTL8), *Boiga irregularis* (brown tree snake, A0A0B8RYN4), *Ornithorhynchus anatinus* (duckbill platypus, F7FH91), *Lepisosteus oculatus* (spotted gar, W5MNJ5), *Strongylocentrotus purpuratus* (purple sea urchin, W4ZC02), *Apis mellifera* (honey bee, A0A087ZSK5), *Amphimedon queenslandica* (sponge, I1G1T9) and *Trichinella spiralis* (trichina worm, E5SBS8)), as well as *H. sapiens* and *S. cerevisiae*. Important residues according to Babbitt *et al.* [172] is highlighted in yellow, and the *S. cerevisiae*, *E. caballus* and *H. sapiens* HCCS sequences are coloured in light grey. The stretch of the sequences of the N-terminal and the C-terminal tails are long and have lower similarities, therefore they are omitted from this figure.

```
F6YF82       VREESSIPRA--DSEKKWVYPSEQMFWNAMLRKGWKWKEEDISQ--KDMY 152
Q7ZTL8       DREESTIPRS--STEKNWVYPSEQMFWNAMLRKGWRWKEDDLKP--EDMT 181
W5MNJ5       AREESNIPRA--GSEKNWVYPSEQMFWNAMLRKGWRWKDDDLSP--GDMT 158
A0A091WFQ9   VREESSIPRA--HSDKNWVYPSEQMFWNAMLRKGWRWKDDDITG--EDMT 152
G1N619       VREESSIPRA--HSDKKWVYPSEQMFWNAMLRKGWRWKDDDITS--EDMT 178
G3WAQ4       VREESSIPRA--DSDKKWVYPSEQMFWNAMLRKGWKWKDDDISQKEKDMY 154
F7FH91       VREESSIPRA--DSEKKWVYPSEQMFWNAMLRKGWRWKDDEINQ--KDMY 152
K7DKZ2       VREESSIPRA--DSEKKWVYPSEQMFWNAMLKKGWKWKDEDISQ--KDMY 148
F6XSM6       IREESSIPRA--DSEKKWVYPSEQMFWNAMLRKGWKWKDEDISQ--KDMY 148
I3MM59       VREESSIPRA--DSEKKWVYPSEQMFWNAMLRKGWKWKDEDISQ--KDMY 152
D2H0T7       VREESSIPRA--DSEKKWVYPSEQMFWNAMLRKGWKWKDDDISQ--KDMY 152
```

```
U6DXA2       AREESSIPRA--DSEKNWVYPSEQMFWNAMLRKGWKWKDDDISQ--KDMY 152
E2R7M2       VREESSIPRA--DSEKKWVYPSEQMFWNAMLRKGWKWKDDDISQ--KDMY 152
M3XCQ6       VREESSIPRA--DSDKKWVYPSEQMFWNAMLRKGWKWKDDDISQ--RDMY 152
L8I4R6       VREESSIPRA--DSDKKWVYPSEQMFWNAMLRKGWKWKDEDISQ--KDMY 155
K9IX17       VREESSIPRA--DSEKKWVYPSEQMFWNAMLKKGWKWKDDDISQ--KDMY 149
L5L251       VREESSIPRA--DSEKKWVYPSEQMFWNAMLRKGWKWKDDDISQ--RDMY 152
S7NJM1       VREESSIPRA--DSEKKWVYPSEQMFWNAMLKKGWKWKDDDISQ--KDMY 151
P53702       SREESSIPRA--DSEKKWVYPSEQMFWNAMLRKGWKWKDDDISQ--KDMY 152
A0A0B8RYN4   VREESTIPRA--YSEQKWVYPSEQMFWNAMLRKGWRWKDDDMSP--EDMS 145
A0A087ZSK5   ERQVSSIPKAT-GEGEFWVYPSQQMFWNAMLRKGWRWKNDDITP--KDMD 140
E5SBS8       ERELSSIPKA--GSNENWMYPSAQMFWNAMLRKGWQWKESDIQP--DDMD 65
W4ZC02       ERKTSGIPRA--GKDENWVYPSQQMFWNAMLRKGWRWQPEDIQQ--EDMH 171
I1G1T9       EREVSSIPKSGGDKDDKWVYPSEQMFFNAMTRKGWKWTEAELKQ--EDMK 103
P53701       VREESSIPRA--DSEKKWVYPSEQMFWNAMLKKGWKWKDEDISQ--KDMY 148
P06182       DRTISSIPKSP-DSNEFWEYPSPQQMYNAMVRKGKIGGSGEVAE--DAVE 122
                *  * **::     . * *** * ::*** :**      ::      :

F6YF82       NIIRIHNQNNEQAWQEILKWEALHAAECPCGPSLIRFGGKAKEYSPRARI 202
Q7ZTL8       NIIKIHNKNNEQAWSEILKWEALHAKECPCGPSLVRFGGKAKEFSPRARM 231
W5MNJ5       NIIKIHNQNNEQAWHEILKWEALHAKECPCGPTLVRFGGKAKEYSPRARI 208
A0A091WFQ9   NIIKIHNQNNEQAWKEILKWEALHAVECPCGPSLMRFGGKAKEYSPRARI 202
G1N619       NIIKIHNQNNEQAWKEILKWEALHAMECPCGPSLMRFGGKAKEYSPRARI 228
G3WAQ4       NIIKIHNQNNEQAWKEILKWEALHATECPCGPTLIRFGGKAKDYSPRARI 204
F7FH91       NIIKIHNQNNEQAWKEILKWEALHATECPCGPTLVRFGGKAKDYSPRAKI 202
K7DKZ2       NIIRIHNQNNEQAWKEILKWEALHAAECPCGPSLIRFGGKAKEYSPRARI 198
F6XSM6       NIIRIHNQNNEQAWKEILKWEALHAAECPSGPSLIRFGGKAKEYSPRARI 198
I3MM59       NIIKIHNQNNEQAWKEILKWEALHAAECPCGPSLIRFGGKAKQYSPRARI 202
D2H0T7       NIIRIHNQNNEQAWKEILKWEALHAAECPCGPSLIRFGGKAKEYSPRARI 202
U6DXA2       NIIRIHNQNNEQAWKEILKWEALHAAECPCGPSLIRFGGKAKEYSPRARI 202
E2R7M2       NIIRIHNQNNEQAWKEILKWEALHAAECPCGPSLIRFGGKAKEYSPRARI 202
M3XCQ6       NIIKIHNQNNEQAWKEILKWEALHAAECPCGPSLIRFGGKAKEYSPRARI 202
L8I4R6       NIIRIHNQNNEQAWKEILKWEALHAAECPCGPSLIRFGGKAKEYSPRARI 205
K9IX17       NIIRIHNQNNEQAWKEILKWEALHAAECPCGPSLVRFGGKAKEYSPRARI 199
L5L251       NIIRIHNQNNEQAWKEILKWEALHAAECPCGPSLIRFGGKAKEYSPRARI 202
S7NJM1       NIIKIHNQNNEQAWKEILKWEALHAAECPCGPSLVRFGGKAKEYSPRARI 201
P53702       NIIRIHNQNNEQAWKEILKWEALHAHECPCGPSLVRFGGKAREYSPRARI 202
A0A0B8RYN4   NIIKIHNRNNEQTWKEILKWELLHLRDCPCGPTLVRFGGKAKEYSPRARI 195
A0A087ZSK5   DIIKIHNANNELAWQEVLKWEALHVKECGT-PKLRSFGGKAKQYSPRARI 189
E5SBS8       SIIRIHNSNNEQAWREILRWEALHFKECDC-PKLKSFRGDATKYTPRARI 114
W4ZC02       NIIHIHNANNEQAWQEVLKWEAMHAKECPC-PKLASFGGRYNDLSPRAKI 220
I1G1T9       HIISIHNRNNEEAWREVLKWEQFHAHECTS-PKLVSFRGRAKDFTPRARL 152
P53701       NIIRIHNQNNEQAWKEILKWEALHAAECPCGPSLIRFGGKAKEYSPRARI 198
P06182       SMVQVHNFLNEGCWQEVLEWEKPHTDESHVQPKLLKFMGKPGVLSPRARW 172
              :: :**  **  * *:*.**  *  :.   *.*  * *     :***:

F6YF82       RSWMG--------YELPFDRHDWIINRCG----------TEVRYVIDYYD 234
Q7ZTL8       RSWMG--------YELPFDRHDWIVDRCG----------RKVRYVIDYYD 263
W5MNJ5       RSWMG--------YELPFDRHDWIVDRCG----------KEVRYVIDYYD 240
A0A091WFQ9   RSWMG--------YELPFDRHDWIVDRCG----------KEVRYVIDYYD 234
G1N619       RSWMG--------YELPFDRHDWIVDRCG----------KEVRYVIDYYD 260
G3WAQ4       RSWMGKR------YELPFDRHDWIVNRCG----------TEVRYVIDYYD 238
F7FH91       RSWMG--------YELPFDRHDWIVDRCG----------TEVR------- 227
K7DKZ2       RSWMG--------YELPFDRHDWIINRCG----------TEVRYVIDYYD 230
F6XSM6       RSWMG--------YELPFDRHDWIINRCG----------TEVRYVIDYYD 230
```

```
I3MM59          RSWMG--------YELPFDRHDWIINRCG----------TEVRYVIDYYD 234
D2H0T7          RSWMG--------YELPFDRHDWIINRCG----------TEVRYVIDYYD 234
U6DXA2          RSWMG--------YELPFDRHDWIINRCG----------TEVRYVIDYYD 234
E2R7M2          RSWMG--------YELPFDRHDWIINRCG----------TEVRYVIDYYD 234
M3XCQ6          RSWMG--------YELPFDRHDWIINRCG----------TEVRYVIDYYD 234
L8I4R6          RSWMG--------YELPFDRHDWIINRCG----------TEVRYVIDYYD 237
K9IX17          RSWMG--------YELPFDRHDWIINRCG----------TEVRYVIDYYD 231
L5L251          RSWMG--------YELPFDRHDWIINRCG----------TEVRYVIDYYD 234
S7NJM1          RSWMG--------YELPFDRHDWIINRCG----------TEVRYVIDYYD 233
P53702          RSWMG--------YELPFDRHDWIINRCG----------TEVRYVIDYYD 234
A0A0B8RYN4      RSWMG--------YELPFDRHDWIVDRCG----------TQVRYIIDYYD 227
A0A087ZSK5      RYWMG--------YELPFDRHDWIIDRCG----------KDVRYVIDYYD 221
E5SBS8          RRALG--------YELPFDRHDWIIDRCG----------KEVHYVIDYYD 146
W4ZC02          RCWMG--------YEKPFDRHDWIVDRCG----------RKVRYVIDYYD 252
I1G1T9          RYWMG--------YELPFDRHDWIVDRCG----------RHVRYVIDYYG 184
P53701          RSWMG--------YELPFDRHDWIINRCG----------TEVRYVIDYYD 230
P06182          MHLCGLLFPSHFSQELPFDRHDWIVLRGERKAEQQPPTFKEVRYVLDFYG 222
                     *         * ********: *              .*:
```

6.2.4.2 Variation of the number of the isoforms of HCCS in different species

Pairwise alignments by ClustalW were conducted among the *S. cerevisiae* HCCS and HCC_1S, *E. caballus* HCCS and human HCCS, as well as among the isoforms of HCCS in *C. albicans* and in *D. rerio*. Table 6.1 shows the deducted identity and similarity scores of each alignment. (In my study, the *S. cerevisiae* HCCS and HCC_1S were chosen for the systems in the experiments, as described in Chapters 2 and 3). Among the results of these alignments, the *E. caballus* HCCS and the *H. sapiens* HCCS have high identity and similarity, but the *E. caballus* HCCS have relatively lower identity and similarity with both of the *S. cerevisiae* HCCS and HCC_1S. *S. cerevisiae* HCCS and HCC_1S have the lowest sequential similarity and identity. Therefore, it seems that the HCCS and HCC_1S in yeast *S. cerevisiae* diverged from each other very early.

Sequence	Identity	Similarity
S. cerevisiae HCCS (P06182) + *S. cerevisiae* HCC_1S (Q00873)	26%	38.2%
E. caballus HCCS (F6YF82) + *S. cerevisiae* HCCS (P06182)	33.6%	45.1%
E. caballus HCCS (F6YF82) + *S. cerevisiae* HCC_1S (Q00873)	31.3%	44.2%

E. caballus HCCS (F6YF82) + *H. sapiens* HCCS (P53701)		90.1%	93.4%
C. albicans	Q5AMR8 + Q5A1U6	28.5%	38.4%
	Q5A1U6 + Q5A1N9	70.8%	71.2%
	Q5A1N9 + Q5AMR8:	23.1%	31.5%
D. rerio	F1QPK2 + Q803H2	68.0%	77.6%
	Q803H2 + F1QN38	32.9%	37.5%
	F1QN38 + F1QPK2	48.9%	50.0%

Table 6.1 The identities and similarities of pairwise alignements of different sequences of different organisms. The UniProt entries are listed in the table.

Species' taxonomical name	No. of genes of HCCS	Species' common name
Anolis carolinensis	1	Animal, Lizard
Anopheles gambiae	1	Animal, Mosquito
Ashbya gossypii	2	Fungus, Yeast
Batrachochytrium dendrobatidis	2	Fungus, Chytrid
Bos taurus	1	Animal, Ox
Branchiostoma floridae	2	Animal, Lancelet
Caenorhabditis briggsae	1	Animal, Roundworm
Caenorhabditis elegans	1	Animal, Roundworm
Candida albicans	3	Fungus, Yeast
Canis familiaris	1	Animal, Dog
Chlamydomonas reinhardtii	10	Green Alga
Ciona intestinalis	1	Animal, Sea squirt
Cryptococcus neoformans	2	Fungus, Yeast
Danio rerio	3	Animal, Zebrafish
Daphnia pulex	1	Animal, Water flea
Dictyostelium discoideum	1	Amoeba
Dictyostelium purpureum	1	Amoeba
Drosophila melanogaster	1	Animal, Fruit Fly
Emericella nidulans	2	Fungus, Ascomycete
Equus caballus	1	Animal, Horse
Gallus gallus	1	Animal, Chicken
Homo sapiens	1	Animal, Human
Ixodes scapularis	1	Animal, Tick
Macaca mulatta	1	Animal, Monkey
Monodelphis domestica	1	Animal, Opossum
Monosiga brevicollis	2	Choanoflagellate
Mus musculus	1	Animal, Mouse
Nematostella vectensis	1	Animal, Sea anemone
Neurospora crassa	2	Fungus, Ascomycete
Ornithorhynchus anatinus	1	Animal, Platypus
Pan troglodytes	1	Animal, Chimpanzee
Phaeosphaeria nodorum	2	Fungus, Ascomycete
Phytophthora infestans	2	Chromalveolata, Oomycete
Plasmodium falciparum	2	Chromalveolata, Apicomplexa
Pristionchus pacificus	1	Animal, Roundworm

Puccinia graminis	2	Fungus, Basidiomycota
Rattus norvegicus	1	Animal, Rat
Saccharomyces cerevisiae	2	Fungus, Yeast
Schistosoma mansoni	2	Animal, Fluke
Schizosaccharomyces pombe	2	Fungus, Yeast
Sclerotinia sclerotiorum	2	Fungus
Strongylocentrotus purpuratus	1	Animal, Sea urchin
Sus scrofa	1	Animal, Boar
Takifugu rubripes	1	Animal, Pufferfish
Thalassiosira pseudonana	2	Chromalveolata, Diatom
Trichoplax adhaerens	1	Animal, Placozoa
Ustilago maydis	2	Fungus, Basidiomycota
Xenopus tropicalis	1	Animal, Frog
Yarrowia lipolytica	2	Fungus, yeast

Table 6.2 The number of orthologues of HCCS genes in various model organisms, according to data from Panther classification system (http://www.pantherdb.org/)

The Panther classification system (http://www.pantherdb.org/) provides an overview for the classification of orthologues of certain proteins in different organisms. According to the genomic data, the fungi model organisms chosen (such *as S. cevesisiae* and *N. crassa*) usually have two isoforms of HCCS gene in their genomes, presumably coding for two different homologues (HCCS and HCC_1S). An exception is *C. albicans* which has three orthologues (*Emericella nidulans* actually only has two isoforms of HCCS and the extra hit is a false positive, data not shown). The choanoflagellate *Monosiga brevicollis* also has two orthologues of HCCS-coding gene. Organisms under the supergroup *Chromalveolata* - oomycete *Phytophthora infestans*, apicomplexan *Plasmodium falciparum*, and diatom *Thalassiosira pseudonana* – all have two orthologues as well. After examining the sequences, it can be deduced that in the three sequences from *C. albicans*, Q5A1N9 is the same as Q5A1U6 but without a stretch in the C-terminal region. Most multicellular animals (including the primitive placozoan *Trichoplax adhaerens*) have only one orthologue of HCCS, but one notable exception is the zebrafish *D. rerio* with three orthologues. In *D. rerio*, one of the three sequences (F1QPK2) is labeled as 'HCCS', and another (Q803H2) shares a similar sequence with F1QPK2 but with some gaps. The other (F1QN38) seems to be a fragment with only N-terminal intact.

According to the entry on the Panther classification system, the green alga *Chlamydomonas reinhardtii* has ten sequences listed as orthologues of HCCS gene (UniProt entries: A8JB79, A8J5J2, A8J5I2, A8J5H0, A8J5G5, A8JJ99, A8J5G7, A8J5H3, A8J5H6, A8J5G9) in its genome, although it is questionable if all of these genes are expressed and functional. The alignment of the sequences of these orthologues is shown in Figure 6.12. In fact,

Figure 6.12 Multiple sequence alignment of the isoforms of HCCS genes (as shown in Panther classification system) in *C. reinhardtii* genome, along with *H. sapiens* HCCS and *S. cerevisiae* HCCS and HCC_1S (highlighted in grey). Important residues (in cytochrome biosynthesis) proposed [172] are highlighted in yellow. The codes for the UniProt entries have been described in the text.

```
A8J5I2   RSPGTVKLKD----EACMDD-------------------QKVKAFIKQH 169
A8J5H0   RSPGTVKLED----EACMDDQAQAICPCVHSLPPRALIPPQKVKAFIKQH 156
A8J5G5   DRPEGGKAQQEDVRQR-----------------------------KPA 115
A8JJ99   GRP--GRAGG-------------------------------------PP 73
A8JB79   ------HQAPGQ---------------------------QNWVYPSEQ 24
P53701   REESSIPRADSE---------------------------KKWVYPSEQ 124
Q00873   -----IPRTNSD---------------------------RNWIYPSEK 84
P06182   ----IPKSPDSN---------------------------EFWEYPSPQ 98
A8J5J2   RVASNIPKGGTE---------------------------STWLFPSPQ 120
A8J5G7   S-----------------------------------------QAGG 97
A8J5H3   S-----------------------------------------QAGG 97
A8J5H6   DSVEAALDRVKM---------------------------NIYLPQAGG 131
A8J5G9   ----------------------------------------LERSR 74

A8J5I2   GLTLCIPQRELGKGAFGHVEMVIITLPDGTTVKAARKTLHACETKGAVAA 219
A8J5H0   GLTLCIPQRELGKGAFGHVEMVIITLPDGTTVKAARKTLHACETKGAVAA 206
A8J5G5   GMNEATWQRVA------QWEMLHRGECDTPTVLRFQAKPHDLSSRLPPAA 159
A8JJ99   ALATAPFA-----------------------------------EPPAA 86
A8JB79   MFYNAMKRKGWDPQAEDMR-----SVVGIHNTVNEQAWHQVLAWERLHCD 69
P53701   MFWNAMLKKGWKWKDEDISQKDMYNIIRIHNQNNEQAWKEILKWEALHAA 174
Q00873   QFYEAMMKKNWDPNSDDMK-----VVVPLHNSINERVWNYIKSWEDKQGG 129
P06182   QMYNAMVRKGKIGGSGEVAEDAVESMVQVHNFLNEGCWQEVLEWEKPHTD 148
A8J5J2   MVFNALKRKG---KGDDVTEDDMDGFIAAHNSMNEATWQRVAQWEMLHRG 167
A8J5G7   SPSAAIPRAATAAAAACQQFAAGTSSVLQPPELNVGPLGVQARSHDVHMA 147
```

```
A8J5H3   SPSAAIPRAATAAAAACQQFAAGTSSVLQPPELNVGPLGVQARSHDVHMA 147
A8J5H6   SPSAAIPRAAT---AACQQFAAGTSSVLQPPELNVGPLGVQARSHDVHMA 178
A8J5G9   WARELIWPAAANQTHPRPIRISSRRTWPRPPASRASAAPVAPSLQAISSN 124

A8J5I2   LRRILNQELAGLAAAAGCEHAPPAAQR---------------------- 246
A8J5H0   LRRILNQELAGLAASAGCERAVQCLGYRLPTDDDERAELLLSCAEGGSVE 256
A8J5G5   ER------------------------------------------------ 161
A8JJ99   ER------------------------------------------------ 88
A8JB79   ECAT-PRLKRFQGRPSDLSPKARLLN---------------------- 94
P53701   ECPCGPSLIRFGGKAKEYSPRARIRS---------------------- 200
Q00873   EACGGIKLTNFKGDSKKLTPRAWFRSR---------------------- 156
P06182   ESHVQPKLLKFMGKPGVLSPRARWMHLCGLL----------------- 179
A8J5J2   ECDT-PTLLRFQGKPHDLSPLAWVRH---------------------- 192
A8J5G7   EAVQALHPALVDRVKTLREAAAKLQAAPSDAAKQAARRAFMNAKGKVAGL 197
A8J5H3   EAH------------------------------------------------ 150
A8J5H6   EAVQALHP---------------------------------------- 186
A8J5G9   SRVR---------------------------------------------- 128

A8J5I2   ----------------AARCAHDWVD---------------------- 256
A8J5H0   NLLLLTDWSPSAFFIQNAVAKDHFDRIDNAPERRRTGKNKDKYEMLPYPG 306
A8J5G5   ----------------AARCAHDWVI---------------------- 171
A8J5J99  ----------------AARCAHNWVI---------------------- 98
A8JB79   ------------FVGFGLPFDRHDWVVDRCG----------------- 113
P53701   ------------WMGYELPFDRHDWIINRCG----------------- 219
Q00873   ------------ILHLAKPFDRHDWQIDRCG----------------- 175
P06182   ----------FPSHFSQELPFDRHDWIVLRGERKAEQQ---------- 207
A8J5J2   ------------MLGGPAPFDRHDWVIDRCG----------------- 211
A8J5G7   MVVVDTMFNENSRDEVLALLSTSEVARAATTVRDAVASIQAVALQPAHVN 247
A8J5H3   CTVVSWVLGN-------VLLATGKVAPSQT------------------- 173
A8J5H6   -ALVDRAAAAAGAEPPSSCPVNPKYKNPAVYN----------------- 217
A8J5G9   ----------------CRSLSRPRRLAEAS----------------- 142
```

there are few or no domains or motifs shared by all of these ten sequences, and therefore it is possible that some of these sequences are related to other functions (or in fact, have little function at all).

The reported important residues as described in Section 6.2.4.1 were highlighted in Figure 6.12. Only A8JB79 and A8J5J2 have mostly the same residues at these positions respectively (A8J5J2 has a lysine at M126 and a phenylalanine at Y120). Others do not share much sequence similarity with any of these. Hence, it is possible that these sequences actually represent proteins for different functions other than cytochrome biogenesis. How this database classified these sequences into HCCS category is not yet clear.

It has been suggested that lateral gene transfer played a vital role in the duplication, and then the different isoforms of HCCS evolved to be specific in maturation of specific cytochromes *c* respectively [38]. Therefore, presumably the common ancestor of the eukaryotic organisms must have only had one HCCS in its genome. The HCCS isoforms could be categorised into HCCS and HCC_1S types and organisms with two HCCS orthologues, such as has one HCCS type and one HCC_1S type [38].These observations indicate that more diverse System III components and mechanisms potentially exist in different unicellular eukaryotes that are divergent from animals and fungi. Since the vast majority of the multicellular organisms only have one copy of HCCS (with exceptions like zebrafish, which might be a late multiplication since the primitive placozoa have only one copy of HCCS), this distinction between copy numbers of animals, fungi and other eukaryotic organisms could be important in the evolutionary progress of animals and divergence of these clades.

6.3 Perspectives

The sequences of the N-terminal region of the unique mitochondrial *c*-type cytochromes from *Z. mays*, *D. papillatum* and nematodes mentioned above suggest the possibility of more undiscovered mitochondrial cytochrome *c* sequences in species with their genome yet to be sequenced, with this alanine in place of the aforementioned phenylalanine. In the meantime, the level of conservation of particular residues at the N-terminal region of plant, euglenoid and nematode mitochondrial cytochromes *c* is significant and evolutionary implications can be deduced from the observations. Sequences of HCCS and HCC_1S from different organisms were aligned and compared, and potentially important features for the difference between their specificity have been identified.

Chapter 7

Expression of *Homo sapiens* holocytochrome *c* synthase

7.1 Introduction

As the only component in the cytochrome *c* biogenesis system III, holocytochrome *c* synthase (HCCS) protein has a crucial role in the assembly of cytochrome *c* which has vital functions in the mitochondrial respiratory chain and the apoptosis process. However, the studies of its recognition are mostly focused on its substrate, the mitochondrial cytochrome *c* [108,129,149,173]. In contrast, mutagenesis analysis and biochemical characterisation on the HCCS enzyme itself has been scarce.

One of the major obstacles for the study on HCCS enzyme is the difficulty of expression and purification of the enzyme itself. In the past, the *S. cerevisiae* (yeast) HCCS protein had only been partially purified by solubilization with Triton X-100 and ammonium sulfate fractionation [174]. The low expression level of HCCS - a mitochondrial protein - in bacteria and the difficulty with the purification techniques has made the task of overexpression and purification complicated. Recently, San Francisco *et al.* succeeded in purifying the *H. sapiens* (human) HCCS enzyme [142]. Their approach involved the fusion of HCCS with a Glutathione S-transferase (GST) tag. Evidently, in the results in San Francisco *et al.* [142] the degraded GST-tag is clearly visible in the figures, with bands at various molecular mass present in a ladder in the SDS-PAGE analysis or Western blot. This problem complicates the purification process and interrupts the further structural and functional studies on this tagged protein. As discussed above, the size of this tag also has implications on the analysis of this protein.

In this study, I attempt to solve this degradation problem of GST-tagged HCCS in order to express the protein in a better understood and more-frequently-used strain, and to assess the potential approaches to solve the degradation problem and to increase the yield.

7.2 Results

7.2.1 The expression of the GST-tagged *H. sapiens* HCCS in *E. coli* strain EC06

San Francisco *et al.* [142] used GST-tagged HCCS in their studies, and they chose the *E.coli* strain EC06 [113] for protein production. The sequence of the human HCCS is shown in Figure 7.1. In this work, the *H. sapiens HCCS* gene was synthesised, and the plasmid was constructed by Genescript commercially with the vector pGEX.4T1. It was tested and successfully expressed within *E. coli*. The band for the GST-tagged HCCS was detected from the Western blotting analysis shown in Figure 7.2.

```
MGLSPSAPAVAVQASNASASPPSGCPMHEGKMKGCPVNTEPSGPTCEKK
TYSVPAHQERAYEYVECPIRGTAAENKENLDPSNLMPPPNQTPAPDQPF
ALSTVREESSIPRADSEKKWVYPSEQMFWNAMLKKGWKWKDEDISQKDM
YNIIRIHNQNNEQAWKEILKWEALHAAECPCGPSLIRFGGKAKEYSPRA
RIRSWMGYELPFDRHDWIINRCGTEVRYVIDYYDGGEVNKDYQFTILDV
RPALDSLSAVWDRMKVAWWRWTS
```

Figure 7.1 The amino acid sequence of *H. sapiens* HCCS

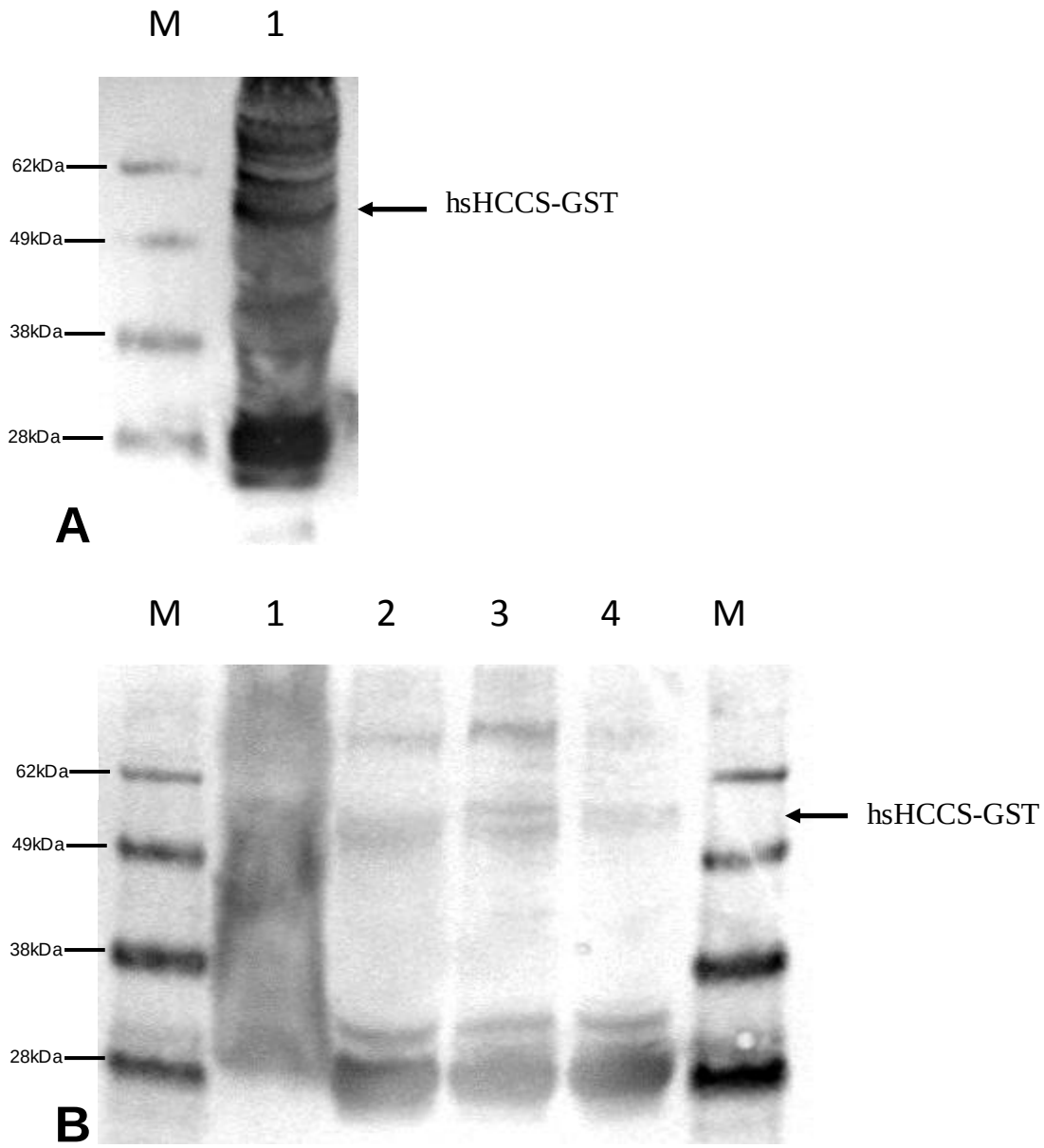


Figure 7.2 (A) The Commassie blue stained SDS-PAGE analysis of the cytoplasmic extraction of GST-tagged *H. sapiens* HCCS. The 28 kDa, 38 kDa, 49 kDa and 62 kDa molecular markers are shown in lane M. Lane 1 shows the expression with 10ml starting culture, 0.1mM IPTG induction and 5 hour growth at 37^{o}C. **(B)** Western blotting analysis of the cytoplasmic extraction of GST-tagged *H. sapiens* HCCS expressed in *E. coli* strain EC06 under different conditions. Equal volumes of cytoplasmic extracts - diluted so as to contain equal amounts of material from wet cell mass - were loaded on each. The 28 kDa, 38 kDa, 49 kDa and 62 kDa molecular markers are shown in lane M. Lane 1 shows the expression with 10ml starting culture, 0.1mM IPTG induction and 5 hour growth at 37^{o}C; Lane 2 expression with 10ml starting culture, 0.1mM IPTG induction and overnight growth at 37^{o}C; Lane 3 shows the expression with 1ml starting culture, 1mM IPTG induction and overnight growth at 30^{o}C and Lane 4 shows the expression with 1ml starting culture, 1mM IPTG induction and overnight growth at 30^{o}C. All of the cultures were induced at OD_{600} 0.6~0.7.

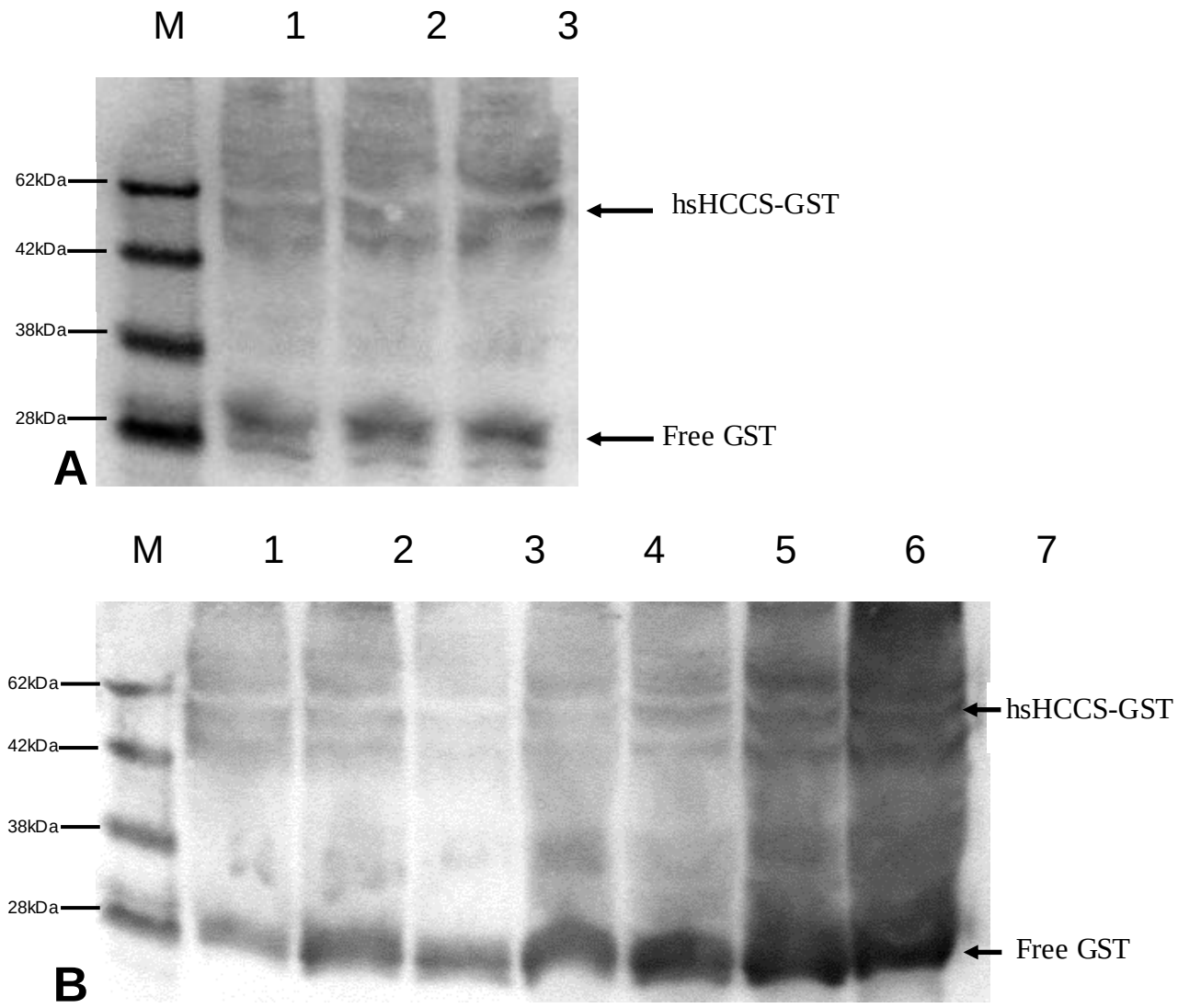


Figure 7.3 Western blotting analysis of GST tagged *S. cerevisiae* HCCS in *E. coli* expressed in *E. coli* strain EC06 under various conditions. Equal volumes of cytoplasmic extracts - diluted so as to contain equal amounts of material from wet cell mass - were loaded on each lane. The 28 kDa, 38 kDa, 49 kDa and 62 kDa molecular markers are shown in lane M. (A) Lane 1 shows the expression with 0.1mM IPTG induction and 5 hour growth; Lane 2 shows the expression with 0.5mM IPTG induction and 5 hour growth and Lane 3 shows the expression with 1mM IPTG induction and 5 hour growth. All of the cultures were grown at 30^{o}C and induced at OD_{600} 0.6~0.7. (B) Lane 1 shows the expression with 0.1mM IPTG induction and 5 hour growth; Lane 2 shows the expression with 0.5mM IPTG induction and 5 hour growth; Lane 3 shows the expression with 1mM IPTG induction and 5 hour growth and Lane 4-7 shows expressions with the same condition with Lane 1-4 in Figure 7.2 (A). All of the cultures in Lane 1-3 were grown at 30^{o}C and induced at OD_{600} 0.6~0.7.

Different conditions with a consortium of varied concentrations of IPTG for induction, varied temperature and varied growth time were tested in EC06 strain. IPTG concentrations (0.1mM, 0.5mM and 1mM) were tested in expressions, under different temperatures (20^{o}C, 30^{o}C and 37^{o}C) with mixed results. The results are shown in Figure 7.2 and Figure 7.3. In all the conditions, the free GST-tag bands (indicated on the gel images in the figures) are visible and the GST-tagged HCCS are mostly visible but sometimes faint.

7.3 Discussion and perspectives

The expression completed in [142] was successfully repeated and confirmed in the *E. coli* EC06 strain. The expression level was reasonable and it agrees with the literature. However, the degradation problem of the GST-tag persists in the expressions in *E. coli*, both in EC06 and BL21(DE3). As shown in the Western blot results in Figure 7.3, in all the lanes, the bands of free GST (often with a high intensity) can be observed from the result due to this degradation, along with a number of other bands. The bands of GST-tagged HCCS are observable in most lanes, but are usually faint, indicating the degradation of the protein results in a low level of binding on the blot. Hence, under all of the varied growth temperature and IPTG concentration conditions, there was no apparent improvement of the degradation problem with the GST-tag.

The choice of the GST tag in the study of San Francisco *et al.* [142] is worthy for discussion: GST is a relatively large protein with the molar mass of 26 kDa. As the HCCS protein has a molecular weight of 30.6 kDa, the fusion protein or GST-tagged HCCS has a resultant molecular weight of ~57 kDa. The large GST-tag directly attached to HCCS protein has the possibility to interfere and disrupt its interaction with other species. Furthermore, this problem may hinder any analysis carried out on this GST-tagged HCCS which can be overexpressed and purified.

On the other hand, the degradation of GST-tag also poses a problem for this approach for the expression and purification of HCCS. The GST-tagged HCCS can readily have the large GST tag hydrolysed and dissociated from the fusion protein following production. A GST-tag is usually hydrolysed from the GST-tagged HCCS in a progressive fashion, with the GST bands clearly visible in a Western blot analysis. This degradation causes the loss of tag from HCCS protein, rendering the quantity of tagged protein to decrease, and may thus result in a low yield in purification.

Therefore, designing a new experiment for a new tag on the HCCS for a better purification yield and ambiguity would be an important step to optimise the expression of HCCS. One such candidate for the tag is the Strep-tag, as it has a much smaller size than the size of the GST tag. The Strep-tag has also been used in Chapter 4 for the Western blotting analyses, and therefore it would be a familiar tool for further investigations. The smaller tag may have the potential of having a lower susceptibility to degradation, thus the yield may be

further improved. Comprehensive design of trials with different conditions would enable the identification of the optimised condition for the expression of HCCS in *E. coli*.

Due to the time limit, it was not feasible to carry out the experiments that are described above. But the present work sets the scene for others to take forward such approaches.

Chapter 8

Concluding remarks and further research perspectives

8.1 General conclusions and implications

The conclusion that the vital deciding factors for the HCCS recognition of its substrate, apocytochrome *c*, lie in the N-terminal region preceding the heme-binding motif CXXCH, has been extensively demonstrated in Chapter 3 by the successful maturation of two chimeric cytochromes *c*, and the truncation variants of mitochondrial cytochrome *c*, by HCCS. The features in the N-terminal region sequence, namely the highly conserved residues and their spacing difference between the mitochondrial cytochromes *c* and bacterial cytochromes *c*, were investigated in Chapter 4. The highly conserved phenylalanine (F11 in *E. caballus* and F15 in *S. cerevisiae*) was found to be critical for the heme attachment on mitochondrial cytochrome *c* by HCCS. Further characterisation indicates the requirements of a residue at this position, for example the successful holocytochrome *c* production of the F15Y variant. This is supported by the observation of Y in the equivalent position in cytochromes c_1 in most of the sequences available, as well as in cytochromes *c* in some nematode species. Chapter 6 also presented a few sequences with an A in the equivalent position in the N-terminal region, suggesting an evolutionary shift in euglenoids.

The similarity between the highly conserved N-terminal sequence in yeast and animals which use System III, and that in euglenoids (using the speculated System V) and most plants (using System I) indicates that some unknown mechanism drives these species to preserve this N-terminal region. One possibility is that the components in System V havw recognition specificity narrower than HCCS and HCC_1S, therefore they might recognise both the N-terminal region sequence and some other unknown features on their own

cytochromes *c*. But the observation with *D. papillatum* suggests that System V does not require the crucial F. At this point it should be recognised that as there is no genome sequence for *D. papillatum* available, it is assumed that HCCS is absent in its genome, similar to other euglenoids. It is also possible that *D. papillatum* has a Ccm system/System I, except that this system functions poorly with cytochromes *c* with an AXXCH motif. In plants, there might be some undiscovered components in their mitochondria to recognise the N-terminal region sequence. This high conservation level of residues also indicates that this phenylalanine in the N-terminal region, and maybe some other features, evolved earlier than the rise of different cytochrome *c* biogenesis systems. Even though the expression of the mitochondrial cytochrome c_1 – HCC_1S system was not successful, the sequence analysis in Chapter 6 provides some insights which could be useful for future studies on the system.

The results in Chapter 5 give new evidence for the importance of the highly conserved cysteines and histidine residues in the CXXCH motif. The failure of cytochrome *c* maturation of all the histidine variants demonstrates the importance of this heme ligating residue, in agreement with previous studies.

Even though the results in this thesis define the key features in the N-terminal region of the sequence of mitochondrial cytochromes *c* for the substrate recognition of HCCS, many questions remain unanswered. The mechanism of HCCS recognition of mitochondrial cytochrome c_1, as well as the function of the isoform HCC_1S in fungi and other species [38], have not been investigated. Also, the components of the speculated System V in

euglenoids, if identified, would give a much better understanding of the evolutionary history of the cytochrome *c* biogenesis systems. An updated phylogenetic study on the distribution of these systems, including System III in all organisms, would be helpful too. With the development of more cost effective DNA sequencing technique, more genomes of diverse organisms would be sequenced. The taxonomy of these cytochromes *c* sequences could be established to give a clearer picture of the evolution of mitochondrial cytochrome *c*, and to tackle many unanswered questions.

The results of many experiments in this thesis have been interpreted to mean that HCCS is unable to catalyse heme attachment to various variant forms of mitochondrial cytochrome *c*, perhaps most importantly to proteins with F11 (horse) or F15 (*S.cerevisiae*) replaced by alanine. Other interpretations of the data have been considered and either excluded or considered improbable. To recap, one possible alternative explanation is that heme was attached to F11/15A variants, and that the resulting holoprotein was sufficiently unstable that it was rapidly degraded. This possibility was explored by seeking to produce the same variant protein in the periplasm of *E. coli* using the System I biogenesis proteins. The protein was produced by this route thus showing that the holocytochromes F11/15A variants are not unstable. Simultaneously with our work, Kleingardner and Bren [140] also reported the same 'control experiment' and they even took the work further by obtaining NMR spectra for the alanine variant produced in the periplasm. The protein appeared to be folded normally. The same type of control approach was done for other variants in the present work. Another possibility is that the HCCS was not expressed simultaneously with the F11/15A protein, whereas it clearly was with the wild type and other variants to which the heme was attached. Admittedly, this has not been controlled

against as it seems to be a very unlikely scenario, the mechanism for which is not easy to envisage. It could have been controlled as the HCCS has a polyHis tag, and Western blotting analysis could be carried out for this purpose. In fact, in work subsequent to that in this thesis, the Kranz group has shown that tagged human HCCS is expressed alongside apocytochromes *c* that are not, or poorly, processed [142,172]. Furthermore, Pollock *et al.* [82] showed that *E. coli* cannot produce mitochondrial cytochrome *c* in the absence of HCCS.

In the investigation of some variant proteins the 'periplasmic control' was not done, for example, the variation of CXXCH because of the previously recognised failure of System I to process such variants. Thus, when the failure of HCCS to attach heme to a CXXCM or CXXCR variant is reported, it cannot be excluded that the apoforms of these protein variants are extremely unstable or that indeed there is no expression of the mutated gene. However, this seems rather unlikely and it is noted that Babbit *et al.* [165] also recently concluded that heme cannot be attached to a CXXCR motif without seeking to detect the apoform of the cytochrome.

Overall, this thesis has provided insight into the significance of multiple features in the amino acid sequence of mitochondrial cytochrome *c* for HCCS substrate specificity: the N-terminal phenylalanine, the residue spacing difference between mitochondrial and bacterial *c*-type cytochromes and the histidine in CXXCH motif. Some features, such as the spacing at the N-terminus, imply that the recognition site of HCCS is complex.

Furthermore, with the proposed future optimisation experiments of the method of *H. sapiens* HCCS production and potential purification in Chapter 7 of this thesis, HCCS could not only potentially be characterised and crystallised to provide a better picture of its substrate recognition mechanism, but also have its relationship with the other cytochrome *c* biogenesis systems discovered. NMR spectroscopy or X-ray crystallography could be applied to the purified and concentrated samples of HCCS and apocytochrome *c* to investigate the protein-protein interactions between them, which would also be analysed with mutagenesis studies.

A better understanding of the mechanism of the HCCS and the evolutionary history of the holocytochrome *c* maturation systems in different organisms may enable a better perspective for the pharmaceutical focus on the trypanosomatids in the kingdom of *Exvata* (trypanosomes, leishmania and others), and other pathogens such as fungi (*P. nodorum, P. graminis* and others as mentioned in Chapter 6). Since they have distinct cytochrome *c* maturation systems, the better understanding of the whole picture of holocytochrome *c* synthesis may provide insights and opportunities to develop potential antiparasitic drugs and antifungal drugs that may target the synthesis of holocytochrome *c* in these pathogens. Moreover, more insight in the biosynthesis of holocytochrome *c*, in particular system III, may also provide opportunities in potential therapies in the MLS disease in humans as mentioned in Section 1.7.2.

REFERENCES

[1] **Pace, N.R.** (1997) A molecular view of microbial diversity and the biosphere. *Science*. *276*, 734-740.

[2] **Richardson, D.J.** (2000) Bacterial respiration: a flexible process for a changing environment. *Microbiology*. *146*, 551-571.

[3] **Mitchell, P.** (1961) Coupling of phosphorylation to electron and hydrogen transfer by a chemi-osmotic type of mechanism. *Nature*. *191*, 144-148.

[4] **Knowles, J.R.** (1980) Enzyme-catalyzed phosphoryl transfer reactions. *Annu. Rev. Biochem. 49*, 877-919.

[5] **Nicholls, D.G. and Ferguson, S.J.** (2013) Bioenergetics 4, Academic Press. *London*.

[6] **Thony-Meyer, L.** (1997) Biogenesis of respiratory cytochromes in bacteria. *Microbiol. Mol. Biol. Rev*. *61*, 337-376.

[7] **Bowman, S.E. and Bren, K.L.** (2008) The chemistry and biochemistry of heme *c*: functional bases for covalent attachment. *Nat. Prod. Rep*. *25*, 1118-30.

[8] **Danielson, P.B.** (2002) The cytochrome P450 superfamily: Biochemistry, evolution and drug metabolism in humans. *Curr. Drug. Metab*. *3*, 561-597.

[9] **Schneider, S. et al.** (2007) Diversity and conservation of interactions for binding heme in *b*-type heme proteins. *Nat. Prod. Rep*. *24*, 621-30.

[10] **O'Brian, M.R. and Thony-Meyer, L.** (2002) Biochemistry, regulation and genomics of haem biosynthesis in prokaryotes. *Adv. Microb. Physiol. 46*, 257-318.

[11] **Wijayanti, N., Katz, N. and Immenschuh, S.** (2004) Biology of heme in health and disease. *Curr. Med. Chem. 11*, 981-986.

[12] **Layer, G. et al.** (2010) Structure and function of enzymes in heme biosynthesis. *Protein Sci. 19*, 1137-61.

[13] **Kumar, S. and Bandyopadhyay, U.** (2005) Free heme toxicity and its detoxification systems in human. *Toxicol. Lett. 157*, 175-188.

[14] **Widger, W.R. et al.** (1984) Sequence homology and structural similarity between cytochrome *b* of mitochondrial complex-III and the chloroplast b_6f complex - position of the cytochrome *b* hemes in the membrane. *P. Natl. Acad. Sci-Biol. 81*, 674-678.

[15] **Esposti, M.D. et al.** (1993) Mitochondrial cytochrome *b* - evolution and structure of the protein. *Biochim. Biophys. Acta*. *1143*, 243-271.

[16] **von Jagow, G. and Sebald, W.** (1980) *B*-Type Cytochromes. *Annu. Rev. Biochem. 49*, 281-314.

[17] **Newton, N.** (1969) The two-haem nitrite reductase of Micrococcus denitrificans. *Biochim Biophys Acta*. *185*, 316-31.

[18] **Caughey, W.S. et al.** (1975) Heme *a* of cytochrome *c* oxidase - structure and properties - comparisons with heme *b*, Heme *c* and Heme *s* and Derivatives. *J. Biol. Chem. 250*, 7602-7622.

[19] **Kranz, R.G. et al.** (2009) Cytochrome *c* biogenesis: mechanisms for covalent modifications and trafficking of heme and for heme-iron redox control. *Microbiol. Mol. Biol. Rev. 73*, 510-28, Table of Contents.

[20] **Ferguson, S.J.** (2009) The Role of Heme d_1 in Denitrification. In Tetrapyrroles: Birth, Life and Death (Warren, M.J., Smith, A. G., ed.^eds), pp. 390-300. *Springer, New York*.

[21] **Ambler, R.P.** (1991) Sequence variability in bacterial cytochromes *c*. *Biochim. Biophys. Acta*. *1058*, 42-47.

[22] **Barker, P.D. and Ferguson, S.J.** (1999) Still a puzzle: why is haem covalently attached in *c*-type cytochromes? *Structure*. *7*, R281-R290.

[23] **Wood, P.M.** (1983) Why do *c*-type cytochromes exist. *FEBS. Lett. 164*, 223-226.

[24] **Wood, P.M.** (1991) Why do *c*-type cytochromes exist - reprise. *Biochim. Biophys. Acta*. *1058*, 5-7.

[25] **Richardson, D.J. et al.** (2012) The 'porin-cytochrome' model for microbe-to-mineral electron transfer. *Mol. Microbiol. 85*, 201-212.

[26] **Page, M.D., Sambongi, Y. and Ferguson, S.J.** (1998) Contrasting routes of *c*-type cytochrome assembly in mitochondria, chloroplasts and bacteria. *Trends Biochem. Sci. 23*, 103-108.

[27] **Allen, J.W.A., Barker, P.D. and Ferguson, S.J.** (2003) A cytochrome b_{562} variant with a *c*-type cytochrome CXXCH heme-binding motif as a probe of the *Escherichia coli* cytochrome *c* maturation system. *J. Biol. Chem. 278*, 52075-52083.

[28] **Tomlinson, E.J. and Ferguson, S.J.** (2000) Conversion of a *c* type cytochrome to a *b* type that spontaneously forms in vitro from apo protein and heme: Implications for *c*-type cytochrome biogenesis and folding. *Proc. Natl. Acad. Sci. U.S.A. 97*, 5156-5160.

[29] **Yamanaka, M. et al.** (2009) Heme is not required for *Aquifex aeolicus* cytochrome c_{555} polypeptide folding. *Biosci. Biotech. Bioch. 73*, 2022-2025.

[30] **Tomlinson, E.J. and Ferguson, S.J.** (2000) Loss of either of the two heme-binding cysteines from a class I *c*-type cytochrome has a surprisingly small effect on physicochemical properties. *J. Biol. Chem. 275*, 32530-32534.

[31] **Allen, J.W.A. et al.** (2005) Why isn't 'standard' heme good enough for *c*-type and d_1-type cytochromes? *Dalton T*. 3410-3418.

[32] **Yamanaka, M., Masanari, M. and Sambongi, Y.** (2011) Conferment of Folding Ability to a Naturally Unfolded Apocytochrome *c* through Introduction of Hydrophobic Amino Acid Residues. *Biochemistry*. *50*, 2313-2320.

[33] **Allen, J.W.A. et al.** (2003) *C*-type cytochromes: diverse structures and biogenesis systems pose evolutionary problems. *Phil. Trans. R. Soc. B. 358*, 255-266.

[34] **Kranz, R. et al.** (1998) Molecular mechanisms of cytochrome *c* biogenesis: three distinct systems. *Mol. Microbiol. 29*, 383-96.

[35] **Allen, J.W.** (2011) Cytochrome *c* biogenesis in mitochondria--Systems III and V. *FEBS. J. 278*, 4198-216.

[36] **Kranz, R.G., Beckett, C.S. and Goldman, B.S.** (2002) Genomic analyses of bacterial respiratory and cytochrome *c* assembly systems: *Bordetella* as a model for the system II cytochrome *c* biogenesis pathway. *Res. Microbiol. 153*, 1-6.

[37] **Stevens, J.M. et al.** (2004) *C*-type cytochrome formation: chemical and biological enigmas. *Accounts Chem. Res. 37*, 999-1007.

[38] **Allen, J.W.A. et al.** (2008) Order within a mosaic distribution of mitochondrial *c*-type cytochrome biogenesis systems? *FEBS. J. 275*, 2385-2402.

[39] **Mavridou, D.A.I., Ferguson, S.J. and Stevens, J.M.** (2013) Cytochrome *c* assembly. *Iubmb Life. 65*, 209-216.

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

[41] **Kuras, R. et al.** (2007) A specific *c*-type cytochrome maturation system is required for oxygenic photosynthesis. *Proc. Natl. Acad. Sci. U.S.A. 104*, 9906-9910.

[42] **Stevens, J.M. et al.** (2011) Cytochrome *c* biogenesis System I. *FEBS. J. 278*, 4170-4178.

[43] **Mavridou, D.A.I. et al.** (2013) Probing Heme Delivery Processes in Cytochrome *c* Biogenesis System I. *Biochemistry*. *52*, 7262-7270.

[44] **Verissimo, A.F. and Daldal, F.** (2014) Cytochrome *c* biogenesis System I: An intricate process catalyzed by a maturase supercomplex? *Biochim. Biophys. Acta. 1837*, 989-998.

[45] **Mavridou, D.A. et al.** (2012) A pivotal heme-transfer reaction intermediate in cytochrome *c* biogenesis. *J. Biol. Chem. 287*, 2342-52.

[46] **Schulz, H., Hennecke, H. and Thony-Meyer, L.** (1998) Prototype of a heme chaperone essential for cytochrome *c* maturation. *Science*. *281*, 1197-200.

[47] **Richard-Fogal, C.L., Frawley, E.R. and Kranz, R.G.** (2008) Topology and function of CcmD in cytochrome *c* maturation. *J. Bacteriol. 190*, 3489-93.

[48] **Ahuja, U. and Thony-Meyer, L.** (2005) CcmD is involved in complex formation between CcmC and the heme chaperone CcmE during cytochrome *c* maturation. *J. Biol. Chem. 280*, 236-243.

[49] **Di Matteo, A. et al.** (2007) A strategic protein in cytochrome *c* maturation: three-dimensional structure of CcmH and binding to apocytochrome *c*. *J. Biol. Chem. 282*, 27012-9.

[50] **Sanders, C. et al.** (2010) Cytochrome *c* biogenesis: the Ccm system. *Trends Microbiol. 18*, 266-274.

[51] **Ren, Q., Ahuja, U. and Thony-Meyer, L.** (2002) A bacterial cytochrome *c* heme lyase - CcmF forms a complex with the heme chaperone CcmE and CcmH but not with apocytochrome *c*. *J. Biol. Chem. 277*, 7657-7663.

[52] **Allen, J.W.A. et al.** (2006) A variant System I for cytochrome *c* biogenesis in archaea and some bacteria has a novel CcmE and no CcmH. *FEBS. Lett. 580*, 4827-4834.

[53] **Eaves, D. et al.** (1998) The nrfEFG gene products are required for the activity of the cytochrome c_{552} nitrite reductase from *Escherichia coli*. *Biochem. Soc. T. 26*, S216.

[54] **Braun, M. and Thony-Meyer, L.** (2004) Biosynthesis of artificial microperoxidases by exploiting the secretion and cytochrome *c* maturation apparatuses of *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A. 101*, 12830-5.

[55] **Mavridou, D.A.L. et al.** (2008) Avoidance of the cytochrome *c* biogenesis system by periplasmic CXXCH motifs. *Biochem. Soc. Trans. 36*, 1124-1128.

[56] **Beckett, C.S. et al.** (2000) Four genes are required for the system II cytochrome *c* biogenesis pathway in Bordetella pertussis, a unique bacterial model. *Mol. Microbiol. 38*, 465-81.

[57] **Simon, J. and Hederstedt, L.** (2011) Composition and function of cytochrome *c* biogenesis System II. *FEBS. J. 278*, 4179-88.

[58] **Frawley, E.R. and Kranz, R.G.** (2009) CcsBA is a cytochrome *c* synthetase that also functions in heme transport. *Proc. Natl. Acad. Sci. U.S.A. 106*, 10201-6.

[59] **Thony-Meyer, L. and Kunzler, P.** (1997) Translocation to the periplasm and signal sequence cleavage of preapocytochrome *c* depend on *sec* and *lep*, but not on the *ccm* gene products. *Euro. J. Biochem. 246*, 794-9.

[60] **Pettigrew, G.W., Aviram, I. and Schejter, A.** (1975) Physicochemical properties of 2 atypical cytochromes *c*, *Crithidia Cytochrome* c_{557} and *Euglena* cytochrome c_{558}. *Biochem. J. 149*, 155-167.

[61] **Yu, J. and Le Brun, N.E.** (1998) Studies of the cytochrome subunits of menaquinone: cytochrome *c* reductase (*bc* complex) of *Bacillus subtilis*. Evidence for the covalent attachment of heme to the cytochrome *b* subunit. *J. Biol. Chem. 273*, 8860-6.

[62] **de Vitry, C.** (2011) Cytochrome *c* maturation system on the negative side of bioenergetic membranes: CCB or System IV. *FEBS. J. 278*, 4189-4197.

[63] **Dumont, M.E. et al.** (1987) Identification and sequence of the gene encoding cytochrome *c* heme lyase in the yeast *Saccharomyces cerevisiae*. *EMBO J. 6*, 235-241.

[64] **Drygas, M.E., Lambowitz, A.M. and Nargang, F.E.** (1989) Cloning and analysis of the *Neurospora crassa* gene for cytochrome *c* heme lyase. *J. Biol. Chem. 264*, 17897-906.

[65] **Corvest, V. et al.** (2012) The flavoprotein Cyc2p, a mitochondrial cytochrome *c* assembly factor, is a NAD(P)H-dependent haem reductase. *Mol. Microbiol. 83*, 968-980.

[66] **Rothstein, R.J. and Sherman, F.** (1980) Genes affecting the expression of cytochrome *c* in yeast: genetic mapping and genetic interactions. *Genetics. 94*, 871-89.

[67] **Bernard, D.G. et al.** (2005) Cyc2p, a membrane-bound flavoprotein involved in the maturation of mitochondrial *c*-type cytochromes. *J. Biol. Chem. 280*, 39852-9.

[68] **Matner, R.R. and Sherman, F.** (1982) Differential accumulation of two apo-iso-cytochromes *c* in processing mutants of yeast. *J. Biol. Chem. 257*, 9811-21.

[69] **Bertini, I., Cavallaro, G. and Rosato, A.** (2006) Cytochrome *c*: occurrence and functions. *Chem. Rev. 106*, 90-115.

[70] **Hatefi, Y.** (1985) The mitochondrial electron transport and oxidative phosphorylation system. *Annu. Rev. Biochem. 54*, 1015-1069.

[71] **Sassa, S. and Nagai, T.** (1996) The role of heme in gene expression. *Int. J. Hematol. 63*, 167-178.

[72] **Leung, K.H. and Hinkle, P.C.** (1975) Reconstitution of ion transport and respiratory control in vesicles formed from reduced coenzyme Q-cytochrome *c* reductase and phospholipids. *J. Biol. Chem. 250*, 8467-71.

[73] **Jiang, X.J. and Wang, X.D.** (2004) Cytochrome *c*-mediated apoptosis. *Annu. Rev. Biochem. 73*, 87-106.

[74] **Ow, Y.L.P. et al.** (2008) Cytochrome *c*: functions beyond respiration. *Nat. Rev. Mol. Cell. Bio. 9*, 532-542.

[75] **Zollner, A., Rodel, G. and Haid, A.** (1992) Molecular cloning and characterization of the *Saccharomyces cerevisiae cyt2* gene encoding cytochrome c_1 heme lyase. *Euro. J. Biochem. 207*, 1093-1100.

[76] **Lill, R. et al.** (1992) Import of cytochrome *c* heme lyase into mitochondria - a novel pathway into the intermembrane space. *EMBO J. 11*, 449-456.

[77] **Nicholson, D.W. and Neupert, W.** (1989) Import of Cytochrome *c* into Mitochondria - Reduction of Heme, Mediated by Nadh and Flavin Nucleotides, Is Obligatory for Its Covalent Linkage to Apocytochrome *c*. *Proc. Natl. Acad. Sci. U.S.A. 86*, 4340-4344.

[78] **Nicholson, D.W., Stuart, R.A. and Neupert, W.** (1989) Biogenesis of cytochrome c_1 - role of cytochrome c_1 heme lyase and of the 2 proteolytic processing steps during import into mitochondria. *J. Biol. Chem. 264*, 10156-10168.

[79] **Bernard, D.G. et al.** (2003) Overlapping specificities of the mitochondrial cytochrome *c* and c_1 heme lyases. *J. Biol. Chem. 278*, 49732-42.

[80] **Steiner, H. et al.** (1996) Heme binding to a conserved Cys-Pro-Val motif is crucial for the catalytic function of mitochondrial heme lyases. *J. Biol. Chem. 271*, 32605-32611.

[81] **Moore, R.L., Stevens, J.M. and Ferguson, S.J.** (2011) Mitochondrial cytochrome *c* synthase: CP motifs are not necessary for heme attachment to apocytochrome *c*. *FEBS. Lett. 585*, 3415-3419.

[82] **Pollock, W.B. et al.** (1998) Bacterial expression of a mitochondrial cytochrome *c*. Trimethylation of lys72 in yeast iso-1-cytochrome *c* and the alkaline conformational transition. *Biochemistry*. *37*, 6124-31.

[83] **Meyer, E.H. et al.** (2005) AtCCMH, an essential component of the *c*-type cytochrome maturation pathway in *Arabidopsis* mitochondria, interacts with apocytochrome *c*. *Proc. Natl. Acad. Sci. U.S.A. 102*, 16113-16118.

[84] **Diekert, K. et al.** (2001) Apocytochrome *c* requires the TOM complex for translocation across the mitochondrial outer membrane. *EMBO J. 20*, 5626-35.

[85] **Steiner, H. et al.** (1995) Biogenesis of mitochondrial heme lyases in yeast - import and folding in the intermembrane space. *J. Biol. Chem. 270*, 22842-22849.

[86] **Laz, T.M., Pietras, D.F. and Sherman, F.** (1984) Differential regulation of the duplicated isocytochrome *c* genes in yeast. *Proc. Natl. Acad. Sci. U.S.A. 81*, 4475-9.

[87] **Vincelli, A.J. et al.** (2013) Recombinant expression, biophysical characterization, and cardiolipin-induced changes of two *Caenorhabditis elegans* cytochrome *c* proteins. *Biochemistry*. *52*, 653-66.

[88] **Huttemann, M. et al.** (2011) The multiple functions of cytochrome *c* and their regulation in life and death decisions of the mammalian cell: from respiration to apoptosis. *Mitochondrion*. *11*, 369-381.

[89] **Dumont, M.E., Corin, A.F. and Campbell, G.A.** (1994) Noncovalent binding of heme induces a compact apocytochrome *c* structure. *Biochemistry*. *33*, 7368-78.

[90] **Daltrop, O. and Ferguson, S.J.** (2003) Cytochrome *c* maturation - The in vitro reactions of horse heart apocytochrome *c* and *Paracoccus denitrificans* apocytochrome c_{550} with heme. *J. Biol. Chem. 278*, 4404-4409.

[91] **Ferguson, S.J. et al.** (2008) Cytochrome *c* assembly: A tale of ever increasing variation and mystery? *BBA - Bioenergetics. 1777*, 980-984.

[92] **Ikegami, I., Katoh, S. and Takamiya, A.** (1968) Nature of heme moiety and oxidation-reduction potential of cytochrome $_{558}$ in *Euglena* chloroplasts. *Biochim. Biophys. Acta. 162*, 604-&.

[93] **Pettigrew, G.W. et al.** (1975) Purification, properties and amino acid sequence of atypical cytochrome *c* from two protozoa, *Euglena gracilis* and *Crithidia oncopelti*. *Biochem. J. 147*, 291-302.

[94] **Dumont, M.E., Ernst, J.F. and Sherman, F.** (1988) Coupling of heme attachment to import of cytochrome *c* into yeast mitochondria - studies with heme lyase-deficient mitochondria and altered apocytochromes *c*. *J. Biol. Chem. 263*, 15928-15937.

[95] **Nicholson, D.W., Hergersberg, C. and Neupert, W.** (1988) Role of Cytochrome *c* Heme Lyase in the Import of Cytochrome *c* into Mitochondria. *J. Biol. Chem. 263*, 19034-19042.

[96] **Dumont, M.E. et al.** (1991) Role of cytochrome *c* heme lyase in mitochondrial import and accumulation of cytochrome *c* in *Saccharomyces cerevisiae*. *Mol. Cell. Bio. 11*, 5487-96.

[97] **Mayer, A., Neupert, W. and Lill, R.** (1995) Translocation of apocytochrome *c* across the outer-membrane of mitochondria. *J. Biol. Chem. 270*, 12390-12397.

[98] **Hakvoort, T.B.M., Sprinkle, J.R. and Margoliash, E.** (1990) Reversible Import of Apocytochrome *c* into Mitochondria. *Proc. Natl. Acad. Sci. U.S.A. 87*, 4996-5000.

[99] **Maccecchini, M.L.** (1981) Import of Proteins into Mitochondria. *Method Cell. Biol. 23*, 39-50.

[100] **Arnold, I. et al.** (1998) Two distinct and independent mitochondrial targeting signals function in the sorting of an inner membrane protein, cytochrome c_1. *J. Biol. Chem. 273*, 1469-76.

[101] **Garrido, C. et al.** (2006) Mechanisms of cytochrome *c* release from mitochondria. *Cell Death Differ. 13*, 1423-1433.

[102] **Mattson, M.P. and Chan, S.L.** (2003) Calcium orchestrates apoptosis. *Nat. Cell Biol. 5*, 1041-3.

[103] **Wimplinger, I. et al.** (2006) Mutations of the mitochondrial holocytochrome *c*-type synthase in X-linked dominant microphthalmia with linear skin defects syndrome. *Am. J. Hum. Genet. 79*, 878-89.

[104] **Wimplinger, I., Shaw, G.M. and Kutsche, K.** (2007) HCCS loss-of-function missense mutation in a female with bilateral microphthalmia and sclerocornea: a novel gene for severe ocular malformations? *Mol Vis. 13*, 1475-82.

[105] **Kiryu-Seo, S. et al.** (2006) Unique anti-apoptotic activity of EAAC1 in injured motor neurons. *EMBO J. 25*, 3411-21.

[106] **Prakash, S.K. et al.** (2002) Loss of holocytochrome *c*-type synthetase causes the male lethality of X-linked dominant microphthalmia with linear skin defects (MLS) syndrome. *Hum Mol Genet. 11*, 3237-48.

[107] **Schaefer, L., Ballabio, A. and Zoghbi, H.Y.** (1996) Cloning and characterization of a putative human holocytochrome *c*-type synthetase gene (HCCS) isolated from the critical region for microphthalmia with linear skin defects (MLS). *Genomics. 34*, 166-72.

[108] **Wang, X., Dumont, M.E. and Sherman, F.** (1996) Sequence requirements for mitochondrial import of yeast cytochrome *c*. *J. Biol. Chem. 271*, 6594-604.

[109] **Garcia, L.L. et al.** (1992) Site-directed mutagenesis of the histidine heme ligand in iso-1-cytochrome *c* of *Saccharomyces cerevisiae*. *New J. Chem. 16*, 629-632.

[110] **Sanders, C. and Lill, H.** (2000) Expression of prokaryotic and eukaryotic cytochromes *c* in *Escherichia coli*. *BBA - Bioenergetics. 1459*, 131-138.

[111] **Rumbley, J.N., Hoang, L. and Englander, S.W.** (2002) Recombinant equine cytochrome *c* in *Escherichia coli*: High-level expression, characterization, and folding and assembly mutants. *Biochemistry. 41*, 13894-13901.

[112] **Hanahan, D.** (1983) Studies on transformation of *Escherichia coli* with plasmids. *J. Mol. Biol. 166*, 557-80.

[113] **Thony-Meyer, L. et al.** (1995) *Escherichia coli* genes required for cytochrome *c* maturation. *J. Bacteriol. 177*, 4321-6.

[114] **Arslan, E. et al.** (1998) Overproduction of the *Bradyrhizobium japonicum c*-type cytochrome subunits of the cbb(3) oxidase in *Escherichia coli*. *Biochem. Biophys. Res. Commun. 251*, 744-747.

[115] **Goddard, A.D. et al.** (2010) *C*-type cytochrome biogenesis can occur via a natural Ccm system lacking CcmH, CcmG, and the heme-binding histidine of CcmE. *J. Biol. Chem. 285*, 22880-22887.

[116] **Fulop, V. et al.** (2009) Structure of a trypanosomatid mitochondrial cytochrome *c* with heme attached via only one thioether bond and implications for the substrate recognition requirements of heme lyase. *FEBS. J. 276*, 2822-2832.

[117] **Stevens, J.M. et al.** (2011) The mitochondrial cytochrome *c* N-terminal region is critical for maturation by holocytochrome *c* synthase. *FEBS. Lett. 585*, 1891-6.

[118] **Mizuguchi, H. et al.** (1999) Characterization and application to hot start PCR of neutralizing monoclonal antibodies against KOD DNA polymerase. *J. Biochem. 126*, 762-8.

[119] **Allen, J.W.A. et al.** (2002) The *Escherichia coli* cytochrome *c* maturation (Ccm) system does not detectably attach heme to single cysteine variants of an apocytochrome *c*. *J. Biol. Chem. 277*, 33559-33563.

[120] **Bartsch, R.G.** (1971) Cytochromes: Bacterial. *Methods Enzymol. 23*, 344-363.

[121] **Berry, E.A. and Trumpower, B.L.** (1987) Simultaneous determination of hemes *a*, *b*, and *c* from pyridine hemochrome spectra. *Anal. Biochem. 161*, 1-15.

[122] **Goodhew, C.F., Brown, K.R. and Pettigrew, G.W.** (1986) Heme staining in gels, a useful tool in the study of bacterial *c*-type cytochromes. *Biochim. Biophys. Acta*. *852*, 288-294.

[123] **Altschul, S.F. et al.** (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. *25*, 3389-402.

[124] **Thompson, J.D., Higgins, D.G. and Gibson, T.J.** (1994) Clustal-W - improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*. *22*, 4673-4680.

[125] **Allen, J.W.A. et al.** (2009) Variant *c*-type cytochromes as probes of the substrate specificity of the *E. coli* cytochrome *c* maturation (Ccm) apparatus. *Biochem. J. 419*, 177-184.

[126] **Tanaka, N. et al.** (1975) The crystal structure of bonito (katsuo) ferrocytochrome *c* at 2.3 A resolution. II. Structure and function. *J. Biochem. 77*, 147-62.

[127] **Timkovich, R. and Dickerson, R.E.** (1976) The structure of Paracoccus denitrificans cytochrome c_{550}. *J. Biol. Chem. 251*, 4033-46.

[128] **Hampsey, D.M., Das, G. and Sherman, F.** (1988) Yeast iso-1-cytochrome *c*: genetic analysis of structural requirements. *FEBS Lett. 231*, 275-83.

[129] **Auld, D.S. and Pielak, G.J.** (1991) Constraints on amino acid substitutions in the N-terminal helix of cytochrome *c* explored by random mutagenesis. *Biochemistry*. *30*, 8684-90.

[130] **Yamada, K. et al.** (2003) Empirical analysis of transcriptional activity in the *Arabidopsis* genome. *Science*. *302*, 842-6.

[131] **Bienvenut, W.V. et al.** (2012) Comparative large scale characterization of plant versus mammal proteins reveals similar and idiosyncratic N-alpha-acetylation features. *Mol. Cell, Proteomics*. *11*

[132] **Nicholson, D.W., Hergersberg, C. and Neupert, W.** (1989) Import of cytochrome *c* into mitochondria - role of the cytochrome *c* heme lyase. *Biol Chem H-S. 370*, 627-627.

[133] **Glick, B.S. et al.** (1992) Cytochromes c_1 and b_2 are sorted to the intermembrane space of yeast mitochondria by a stop-transfer mechanism. *Cell. 69*, 809-22.

[134] **Veloso, D., Juillerat, M. and Taniuchi, H.** (1984) Synthesis of a heme fragment of horse cytochrome *c* which forms a productive complex with a native apofragment. *J. Biol. Chem. 259*, 6067-6073.

[135] **Stellwagen, E., Rysavy, R. and Babul, G.** (1972) The conformation of horse heart apocytochrome *c*. *The Journal of biological chemistry*. *247*, 8074-7.

[136] **Ozturk, Y. et al.** (2008) Soluble variants of *Rhodobacter capsulatus* membrane-anchored cytochrome c_y are efficient photosynthetic electron carriers. *J. Biol. Chem. 283*, 13964-13972.

[137] **Nye, S.H. and Scarpulla, R.C.** (1990) Mitochondrial targeting of yeast apoiso-1-cytochrome *c* is mediated through functionally independent structural domains. *Mol. Cell. Biol. 10*, 5763-5771.

[138] **Silkstone, G. et al.** (2002) Production and characterisation of Met80X mutants of yeast iso-1-cytochrome *c*: spectral, photochemical and binding studies on the ferrous derivatives. *Biophys. Chem. 98*, 65-77.

[139] **Daltrop, O. et al.** (2002) In vitro formation of a *c*-type cytochrome. *Proc. Natl. Acad. Sci. U.S.A. 99*, 7872-7876.

[140] **Kleingardner, J.G. and Bren, K.L.** (2011) Comparing substrate specificity between cytochrome *c* maturation and cytochrome *c* heme lyase systems for cytochrome *c* biogenesis. *Metallomics*. *3*, 396-403.

[141] **Asher, W.B. and Bren, K.L.** (2012) Cytochrome *c* heme lyase can mature a fusion peptide composed of the amino-terminal residues of horse cytochrome *c*. *Chem. Commun*. *48*, 8344-8346.

[142] **San Francisco, B., Bretsnyder, E.C. and Kranz, R.G.** (2013) Human mitochondrial holocytochrome *c* synthase's heme binding, maturation determinants, and complex formation with cytochrome *c*. *Proc. Natl. Acad. Sci. U.S.A. 110*, E788-97.

[143] **Verissimo, A.F. et al.** (2012) Engineering a prokaryotic apocytochrome *c* as an efficient substrate for *Saccharomyces cerevisiae* cytochrome *c* heme lyase. *Biochem. Biophys. Res. Commun. 424*, 130-135.

[144] **Bren, K.L. and Gray, H.B.** (1993) Structurally engineered cytochromes with novel ligand-binding sites - oxy and carbonmonoxy derivatives of semisynthetic horse heart Ala80 cytochrome *c*. *J. Am. Chem. Soc. 115*, 10382-10383.

[145] **Moreno-Beltran, B. et al.** (2014) Cytochrome c_1 exhibits two binding sites for cytochrome *c* in plants. *BBA - Bioenergetics*. *1837*, 1717-1729.

[146] **Banci, L. et al.** (1999) Solution structure of reduced horse heart cytochrome *c*. *J. Biol. Inorg. Chem. 4*, 21-31.

[147] **Ginger, M.L., Sam, K.A. and Allen, J.W.** (2012) Probing why trypanosomes assemble atypical cytochrome *c* with an AxxCH haem-binding motif instead of CxxCH. *Biochem. J. 448*, 253-60.

[148] **Rosell, F.I. and Mauk, A.G.** (2002) Spectroscopic properties of a mitochondrial cytochrome *c* with a single thioether bond to the heme prosthetic group. *Biochemistry*. *41*, 7811-8.

[149] **Tanaka, Y. et al.** (1990) Site-directedly mutated human cytochrome *c* which retains heme-*c* via only one thioether bond. *J. Biochem. 108*, 7-8.

[150] **Sorrell, T.N., Martin, P.K. and Bowden, E.F.** (1989) A novel, functional variant of cytochrome *c* - replacement of the histidine ligand with arginine via site-directed mutagenesis. *J. Am. Chem. Soc. 111*, 766-767.

[151] **Fumo, G., Spitzer, J.S. and Fetrow, J.S.** (1995) A method of directed random mutagenesis of the yeast chromosome shows that the iso-1-cytochrome *c* heme ligand His18 is essential. *Gene*. *164*, 33-39.

[152] **Aragao, D. et al.** (2003) Structure of dimeric cytochrome c_3 from *Desulfovibrio gigas* at 1.2 A resolution. *Acta crystallographica. Section D, Biological crystallography*. *59*, 644-53.

[153] **Jungst, A. et al.** (1991) The nirSTBM region coding for cytochrome cd_1-dependent nitrite respiration of *Pseudomonas stutzeri* consists of a cluster of mono-, di-, and tetraheme proteins. *FEBS Lett*. *279*, 205-9.

[154] **Jentzen, W., Ma, J.G. and Shelnutt, J.A.** (1998) Conservation of the conformation of the porphyrin macrocycle in hemoproteins. *Biophys. J. 74*, 753-763.

[155] **Morais, J. et al.** (1995) Structure of the tetraheme cytochrome from *Desulfovibrio desulfuricans* ATCC 27774: X-ray diffraction and electron paramagnetic resonance studies. *Biochemistry*. *34*, 12830-12841.

[156] **Hartshorne, R.S. et al.** (2007) A dedicated haem lyase is required for the maturation of a novel bacterial cytochrome *c* with unconventional covalent haem binding. *Mol. Microbiol. 64*, 1049-1060.

[157] **Ambler, R.P. et al.** (1991) Amino acid sequences of *Euglena viridis* ferredoxin and cytochromes *c*. *Biochem. J. 276 (Pt 1)*, 47-52.

[158] **Priest, J.W. and Hajduk, S.L.** (1992) Cytochrome *c* reductase purified from *Crithidia fasciculata* contains an atypical cytochrome c_1. *J. Biol. Chem. 267*, 20188-95.

[159] **Mukai, K. et al.** (1989) An atypical heme-binding structure of cytochrome c_1 of *Euglena gracilis* mitochondrial complex III. *European journal of biochemistry / FEBS*. *178*, 649-56.

[160] **Einsle, O. et al.** (1999) Structure of cytochrome *c* nitrite reductase. *Nature*. *400*, 476-480.

[161] **Todorovic, S. et al.** (2012) Redox properties of lysine- and methionine-coordinated hemes ensure downhill electron transfer in $NrfH_2A_4$ nitrite reductase. *J. Phys. Chem. B. 116*, 5637-43.

[162] **Hasemann, C.A. et al.** (1995) Structure and function of cytochromes P450: a comparative analysis of three crystal structures. *Structure*. *3*, 41-62.

[163] **Sawyer, E.B. et al.** (2010) Aberrant attachment of heme to cytochrome by the Ccm system results in a cysteine persulfide linkage. *J. Am. Chem. Soc. 132*, 4974-5.

[164] **Tong, J.C. and Margoliash, E.** (1998) Cytochrome *c* heme lyase activity of yeast mitochondria. *J. Biol. Chem. 273*, 25695-25702.

[165] **Babbitt, S.E. et al.** (2014) Mechanisms of mitochondrial holocytochrome *c* synthase and the key roles played by cysteines and histidine of the heme attachment site, CysXXCysHis. *J. Biol. Chem.*

[166] **Zaidi, S. et al.** (2014) The role of key residues in structure, function, and stability of cytochrome *c*. *Cell. Mol. Life Sci. 71*, 229-55.

[167] **Goffeau, A. et al.** (1996) Life with 6000 genes. *Science*. *274*, 546-&.

[168] **Wade, C.M. et al.** (2009) Genome Sequence, Comparative Analysis, and Population Genetics of the Domestic Horse. *Science*. *326*, 865-867.

[169] **Zimin, A.V. et al.** (2009) A whole-genome assembly of the domestic cow, Bos taurus. *Genome Biol. 10*

[170] **Schnable, P.S. et al.** (2009) The B73 maize genome: complexity, diversity, and dynamics. *Science*. *326*, 1112-5.

[171] **Moore, G.R.P., G. W.** (1990) Cytochromes *c*: evolutionary, structural and physicochemical aspectsed.^eds), pp. 478. *Springer-Verlag, New York*.

[172] **Babbitt, S.E. et al.** (2014) Conserved residues of the human mitochondrial holocytochrome *c* synthase mediate interactions with heme. *Biochemistry*. *53*, 5261-71.

[173] **Allen, J.W.A. and Ferguson, S.J.** (2003) Variation of the axial haem ligands and haem-binding motif as a probe of the *Escherichia coli c*-type cytochrome maturation (Ccm) system. *Biochem. J. 375*, 721-728.

[174] **Taniuchi, H. et al.** (1983) Evidence for formation of two thioether bonds to link heme to apocytochrome *c* by partially purified cytochrome *c* synthetase. *J. Biol. Chem. 258*, 10963-6.

APPENDIX

1. The order for the gene synthesis of *S. cerevisiae* cytochrome *c* and HCC_1S at Genescript

Order Items:

1 gene 1:
Gene Synthesis:
Gene name: gene 1,
Length: 717 bp,
Vector name: pUC57,
Cloning strategy: Other,
Plasmid preparation: Standard delivery: 4 ug (Free of charge),
Other Requirements: I do not have the full sequence of vector pLysS – the sequence of the region surrounding the Nde1 site is:

TAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACATATGAGAGGCTCGCATCATCA
TCATCATCATGGTATGGGTTGGTTTTGGGCAGATC

and the sequence around the *BamH*1 site is:

GGTCCGTCCTCTTCGTCCTCCGCCCCTTAAGGATCCGGCTGCTAACAAAGCCCGAAAGGA,

Sequence:

CATATGAGAGGCTCGCATCATCATCATCATCATGGTATGATGTCTTCAGACCAACAGGGG
AAATGCCCCGTAGATGAGGAAACCAAAAAGTTATGGCTACGAGAACATGGCAACGAAGCG
CATCCTGGTGCTACTGCCCCAGGCAATCAACTAGAGTGCTCTGCAAACCCACAAGATAAC
GATAAAACGCCTGAATATCACACCACGGTGGATCTCTCTCAGTCCAGAGAGGTTTCCACC
ATACCAAGGACGAATTCTGACAGAAACTGGATATACCCGTCGGAGAAACAATTTTACGAG
GCAATGATGAAAAAAAATTGGGATCCGAACTCAGATGACATGAAGGTGGTTGTGCCTTTA
CATAACTCCATCAATGAGCGGGTTTGGAACTACATTAAAAGCTGGGAAGACAAGCAGGGT
GGTGAAGCGTGTGGCGGTATCAAATTAACAAATTTTAAAGGTGATTCCAAAAAACTGACA
CCAAGGGCCTGGTTCAGGTCCCGTATCTTGCACCTGGCCAAACCTTTTGATAGACATGAT
TGGCAGATAGATAGATGCGGTAAAACCGTCGACTACGTAATCGATTTCTATTCCACCGAT
CTGAACGATGCAAACTCGCAGCAACAACCACTTATTTATCTCGATGTTAGACCAAAATTA
AACAGTTTTGAGGGCTTTAGATTACGGTTCTGGAAATCTTTAGGCTTTTGAGGATCC

Price: £171.35

2 147004_2:
Name: gene 1-Subcloning,
Vector: pLysS, Quantity: 4 ug
Clone name: gene 1-Subcloning,
Vector name: pLysS,
Vector size: -, Resistance: -,
Copy number: High,
Cloning method: Cloning site: NdeI-BamHI,
Gene usage: Protein expression/analysis,
Reading frame: Should be consistent with client's requirement:
Plasmid preparation: Standard delivery: 4 ug (Free of charge),

Comments: The client will provide the vector.
The client will provide the vector.
Unit Price: £101.74;
Quantity: 1;

3 gene 2:
Gene Synthesis:
Gene name: gene 2, Length: 654 bp,
Vector name: pUC57,
Cloning strategy: Other,
Plasmid preparation: Standard delivery: 4 ug (Free of charge),
Sequence:

CATATGGCTTCTTGGTCTCATCCTCAATTTGAGAAGATTGAAGGTCGTATGACCGCAGCT
GAACACGGATTGCACGCCCCAGCATATGCTTGGTCCCACAATGGGCCTTTTGAAACATTT
GATCATGCATCCATTAGAAGAGGTTACCAGGTTTACCGTGAAGTTTGTGCCGCCTGCCAT
TCTCTTGACAGAGTTGCTTGGAGAACTTTGGTTGGTGTTTCTCATACCAACGAAGAGGTT
CGTAATATGGCCGAAGAATTTGAATACGATGACGAACCTGATGAACAAGGTAACCCTAAA
AAGAGACCAGGTAAGTTGTCCGATTACATCCCTGGCCCATACCCAAACGAACAGGCTGCA
AGAGCTGCTAATCAAGGTGCCTTGCCACCTGATCTATCTTTGATCGTGAAAGCTAGACAC
GGTGGTTGTGACTACATTTTCTCTTTGTTGACCGGTTATCCTGATGAACCTCCTGCTGGT
GTGGCTTTACCACCAGGTTCTAATTATAACCCTTACTTCCCAGGTGGTTCCATTGCAATG
GCAAGAGTCTTGTTTGATGACATGGTTGAGTACGAAGATGGTACCCCCGCAACGACATCT
CAAATGGCAAAGGACGTTACCACCTTTTTAAACTGGTGTGCCGAATGAGGATCC

Price: £156.29

4 147004_4:
Name: gene 2-Subcloning,
Vector: pET-3a, Quantity: 4 ug
Clone name: gene 2-Subcloning,
Vector name: pET-3a,
Cloning method: Cloning site: NdeI-BamHI,
Gene usage: Protein expression/analysis,
Reading frame: Should be consistent with client's requirement: ,
Plasmid preparation: Standard delivery: 4 ug (Free of charge),
Comments: The client will provide the vector.
The client will provide the vector.
Unit Price: £101.74;
Quantity: 1;

2. The DNA open reading frame of the *Z. mays* sequence K7UZX9

The A presenting in the position of the consensus F in K7UZX9 sequence, as well as the F in horse heart cytochrome *c* sequence, are depicted in yellow boxes.

5'3' Frame 1

```
atggtggatgaggtagaggaggaggtgcacgctatggtggtgcaggaaccagaggtaaag
 M  V  D  E  V  E  E  E  V  H  A  M  V  V  Q  E  P  E  V  K
gatgtgttggtgaccaaggtggtggaggctttggaggtgaccaaggtgcatccggtgctt
 D  V  L  V  T  K  V  V  E  A  L  E  V  T  K  V  H  P  V  L
gaatctgagccaagggttgatgaagtgttggtggtcaacgaaacacctgtagtgcctgag
 E  S  E  P  R  V  D  E  V  L  V  V  N  E  T  P  V  V  P  E
atacaggaaccagaggtgaaaggcgtcggtgctaccgttgtgttgaaggaacaagagacg
 I  Q  E  P  E  V  K  G  V  G  A  T  V  V  L  K  E  Q  E  T
aatagtggcaatgtggtggtcaaggactctgatgaggcgttgccatctcaggaggttgct
 N  S  G  N  V  V  V  K  D  S  D  E  A  L  P  S  Q  E  V  A
ggtgtgcatactacagatcgaggaagttgccaggcagctttgtttggagcattggttatc
 G  V  H  T  T  D  R  G  S  C  Q  A  A  L  F  G  A  L  V  I
attttcaaaaatggaacaatagaagaaggtctgatcgacaggatgaactccattgacgac
 I  F  K  N  G  T  I  E  E  G  L  I  D  R  M  N  S  I  D  D
aggatgaactcacgagcagctagcaccccacctgcacgcaccaagtgcgcgcagtgccat
 R  M  N  S  R  A  A  S  T  P  P  A  R  T  K  C  A  Q  C  H
accgtggagcgaggtggcgcgcacaggcaggggcccaacctgcacgacctcttcggtcgt
 T  V  E  R  G  G  A  H  R  Q  G  P  N  L  H  D  L  F  G  R
cagtcaggcaccaccatcggctatgcctactccacggccaataagaacatggctgtcgtc
 Q  S  G  T  T  I  G  Y  A  Y  S  T  A  N  K  N  M  A  V  V
tgggagaagggcaccatgtacgactacctcttcaaccccaagaagtacattccaggcaca
 W  E  K  G  T  M  Y  D  Y  L  F  N  P  K  K  Y  I  P  G  T
aagatggtcttcccgaggctcaagaagcccaaggagcgaaccgatctcatcgcctacttg
 K  M  V  F  P  R  L  K  K  P  K  E  R  T  D  L  I  A  Y  L
acctattga
 T  Y  -
```

5'3' Frame 1

```
atgggtgacgttgaaaagggtaagaagatttccgttcaaaagtgtgctcaatgtcacact
 M  G  D  V  E  K  G  K  K  I  F  V  Q  K  C  A  Q  C  H  T
gtcgaaaaaggtggtaagaacaagaccggtccaaacttgaacggtttgttcggtcgtaag
 V  E  K  G  G  K  N  K  T  G  P  N  L  N  G  L  F  G  R  K
accggtcaagctccaggtttcacttacactgacgctaacaagaacaagggtatcacctgg
 T  G  Q  A  P  G  F  T  Y  T  D  A  N  K  N  K  G  I  T  W
aaggaagaaactttgatggaatacttggaaaacccaaagaagtacattcctggtactaag
 K  E  E  T  L  M  E  Y  L  E  N  P  K  K  Y  I  P  G  T  K
atgattttcgctggtattaagaagaagactgaaagagaagacttgattgcttacttgaag
 M  I  F  A  G  I  K  K  K  T  E  R  E  D  L  I  A  Y  L  K
aaggctactaacgaataa
 K  A  T  N  E  -
```