Inês Tomás de Aquino Raposo
Bachelor's Degree in Biochemistry

Exploring cytochrome-*c*'s biogenesis in eukaryotes

Dissertation to obtain the Master's Degree in Biochemistry for Health

Supervisor: Ricardo Oliveira Louro, PhD

February, 2019

Inês Tomás de Aquino Raposo
Bachelor's Degree in Biochemistry

Exploring cytochrome-*c*'s biogenesis in eukaryotes

Dissertation to obtain the Master's Degree in Biochemistry for Health

Supervisor: Ricardo Oliveira Louro, PhD

Jury:

President: Pedro Matias, PhD
Opponent: Filipe Folgosa, PhD
Members of the jury: Margarida Archer, PhD

Instituto de Tecnologia Química e Biológica António Xavier

February, 2019

Exploring cytochrome-*c*'s biogenesis in eukaryotes

Acknowledgements

Firstly, I would like to thank my supervisor, Ricardo Louro, for the opportunity to pursue this project in his laboratory of Inorganic Biochemistry and NMR (IBN). For believing I had something more to offer and that my abilities would allow me to contribute towards the gained knowledge within his unit. I hope that I lived up to your expectations to the extent of my capabilities and that I will leave your laboratory a better professional than when I first started. Thank you for not giving up on me during these last few months.

Secondly, I would like to thank everyone at IBN, from the ones who have already left to the ones that still remain. Thank you, Ana, for being my unofficial co-supervisor (but for becoming as invested in my work as if it were your own), for teaching me everything that you learned with such care and detail, and for challenging me to always improve my methods. Thank you for every time that we discussed dubious results, as these conversations always sparked a new idea within me. Furthermore, I would like to thank Bruno, who took me "under his wing" and took care of me as if I was his younger sister, showing me that life in the lab can be fun but, whenever I needed, also gave me his scientific advice and offered me new challenges to tackle alone. Thank you, Catarina, for being the "mother" of the lab, and for your career and life advice when I felt lost. Thank you, Inês, for showing me what true dedication is, and that a scientific career entails not giving up in the face of adversity. Fortune does favour the bold. Furthermore, I would like to thank everyone else for not despairing whenever I went into a cleaning spree and for accommodating my material organisation, I know it can be hard sometimes.

I would also like to thank my "lab away from the lab", colleagues who became something more than that. To Filipa, Filipe, Patrícia, Fernanda, Beatriz and Ricardo. For every coffee break, every ludicrous lunch discussion, every break that allowed me to take a step back and refocus on my work. Filipe, thank you for making me realise that not everything is as lost and hopeless as it seems, and for showing me there is always a backup option where I never thought to look, your computer does hold a bit of magic.

To Isabel Pacheco and Paula Chicau, the most dedicated technicians I have ever had the pleasure to meet. Thank you, Isabel, for everything you taught me "by the book", for your suggestions whenever I reached a dead-end (or an air bubble inside the Äkta lines) and for our chats on the train ride back home. Thank you, Paula, for attempting the impossible to revive a mistreated apparatus to give me some results, and for suffering as much as I did until I had them. Thank you for all your support during the last part of my project.

To Teresa Catarino, the happiest and most talkative master's coordinator ever. Thank you for all your help and concern.

To ITQB and FCT for being institutions of excellence, which enabled me to deepen my love for science with great lectures and professors, and for providing their students with challenging and interesting dissertation projects. For giving me the starting tools to excel in a scientific career, whether it be a national or international one.

To my friends, especially Beatriz and Simone, for always facing life with a smile. For showing me that all choices in life are our own. For being by my side without even being asked, and for ensuring me that everything will turn out just fine.

To my family. The fiercest supporters I will ever have. To my aunt Anabela and my grandmother Rosa, who were always like mothers to me in their own special way. Thank you for nurturing me and my interests, even if you don't understand everything. For taking pride in every step I take, even if it seemed like the wrong one. To the newest additions to my family, Francisco and Helena. For never hesitating to bring me into your lives, and for treating me as if I were your own.

For my brother, Guilherme, my pride and joy. For simply being there, even if we just argue. For your daily reminder that life really is simple, and that I need to relax and let go. No matter what, I will always take care of you.

For Ricardo. Words can’t begin to describe everything you did for me during this past year. Thank you for being my strength when I had none and my happiness when I felt miserable. For making me believe in myself and for never giving up on me, always telling me that everything will be alright. For keeping me company during late nights and for helping me enjoy life. You are the sunlight in my growing.

Finally, I want to thank my parents. For their unconditional love and support, for every sacrifice they made. I love you both.

For my father, Rui. In spite of the distance, you did your best to ensure I knew you were there. Thank you for your words of wisdom, for your never-ending strength and for being the most resilient person I know in the face of adversity. For understanding my decisions while giving me space to grow into my own person.

For my mother, Idalina. Thank you for sparking my interest in life’s wonders, for encouraging me to question everything and to learn for myself. For being my role-model when I was growing, and for providing me with every opportunity to improve myself. This is my last gift to you, I hope it would make you proud.

Resumo

Citocromos tipo-*c* são proteínas ubíquas a organismos de todos os domínios da Vida. Estas proteínas desempenham papéis essenciais em processos diversos, desde transporte de electrões a participação em sítios activos enzimáticos no metabolismo de organismos e ao desencadear da apoptose celular. Contrariamente a outros tipos de citocromos, o cofactor hémico está ligado covalentemente à cadeia polipeptídica da holoproteína. Para que este processo ocorra, é necessário um sistema de maturação capaz de realizar esta reacção. Embora estes sistemas estejam bem estudados e caracterizados noutros organismos, existe uma clara lacuna relativamente ao Sistema III, presente em eucariotas. O Sistema III é composto por uma única proteína, CCHL (em leveduras) ou HCCS (em humanos). Isto torna-o o mais simples, estando a CCHL envolvida tanto na importação do polipéptido como na ligação covalente do hemo ao polipéptido após reconhecimento de um motivo de ligação específico, CXXCH.

Neste trabalho, procurou efectuar-se a purificação seguida de uma caracterização estrutural e funcional completa da CCHL. Para tal, o gene de CCHL foi inserido em diferentes vectores plasmídicos e transformado em diversas estirpes de *Escherichia coli* que deveriam sobreexpressar a proteína alvo. Seguidamente, realizou-se produção heteróloga e purificação, culminando numa identificação proteica utilizando métodos de Espectrometria de Massa e sequenciação por N-terminal. Uma proteína de *E. coli* com características semelhantes às de CCHL foi obtida, mas identificaram-se diversas estratégias para o isolamento desta no futuro. Também foi efectuada uma breve tentativa de isolamento do gene mitocondrial HCCS homólogo a partir de sangue humano.

Apesar dos resultados obtidos, as implicações do citocromo-*c* e, portanto, da CCHL para a saúde humana tornam esta linha de investigação extremamente importante a longo prazo, com descobertas passíveis de contribuir para o desenvolvimento de novas terapêuticas e fármacos que tenham as doenças relacionadas com citocromos como alvos.

Palavras-chave: CCHL, citocromo-*c*, eucariotas, hemo, sistema de maturação.

Abstract

C-type cytochromes are proteins ubiquitous to organisms from all domains of Life. They play essential roles in diverse processes, ranging from electron shuttles to participating in enzymatic active sites in organism's metabolism and triggering apoptosis within cells. Contrary to other types of cytochromes, the haem cofactor is bound to the polypeptide chain of the holoprotein. For this process to occur, a maturation system capable of catalysing this reaction is required. Although these systems are well studied and characterised in other organisms, there is a clear lack of knowledge regarding System III, which is present in eukaryotes. System III is composed of a single protein, CCHL (in yeast) or HCCS (in humans). This makes it the simplest one, with CCHL being involved in both the import of the polypeptide and the covalent attachment of the haem to the polypeptide after recognition of a specific binding motif, CXXCH.

In this work, purification followed by structural and full functional characterisation of CCHL was pursued. To this end, a construct of CCHL was inserted into different plasmid vectors and transformed into several *Escherichia coli* strains that should overexpress the target protein. Afterwards, heterologous production and purification was performed, culminating with protein identification using Mass Spectrometry and N-terminal sequencing methods. An *E. coli* protein with similar characteristics to CCHL was obtained, but several strategies were identified to isolate CCHL in the future. A brief attempt to isolate the homologous mitochondrial HCCS gene from human blood was also performed.

Regardless of the obtained results, the implications of cytochrome-*c* and thus CCHL for human health make this line of research extremely important in the long term, with discoveries contributing to the development of new therapeutics and drugs that can target cytochrome-related diseases.

Keywords: CCHL, cytochrome-*c*, eukaryotes, haem, maturation system.

Index

Acknowledgements ..... **I**
Resumo ..... **III**
Abstract ..... **V**
Index of figures ..... **IX**
Index of tables ..... **XIII**
List of Abbreviations ..... **XV**
1. Introduction **1**
1.1. *C*-type cytochromes 1
1.2. Maturation Systems 4
1.3. Lyase 6
1.4. Lyase and Health 8
2. Materials and Methods **11**
2.1. Molecular Biology 11
2.2. Expression tests 17
2.3. Heterologous Protein Expression and Purification 20
3. Results and Discussion **23**
3.1 Prior protein analysis 23
3.2 Expression tests 24
3.3 Growth and purifications 24
3.4 First Mass Spectrometry results 32
3.5 Expression tests with multiple *E. coli* strains 33
3.6 Molecular Biology solutions 36
3.7 pETBlue-2 Plasmid 37
3.8 Mass Spectrometry and N-terminal results 41
4. Conclusion **43**
5. References **45**

Index of figures

Figure 1.1 Haem (Protoporphyrin IX) coordinated with a ferrous iron (Fe^{2+}) atom[5] and haem formation. On the left, the conjugated double-bond system is responsible for the interaction of haem with visible light. On the right are represented the necessary chemical modifications to convert haem b into the different types of haem. Figure adapted from Koch (2007).[6] 1

Figure 1.2 Schematic representation of the electron transport chain in eukaryotes. Cytochrome c (abbreviated as cyt. c) is used as an electron shuttle in oxidative phosphorylation to generate ATP. 3

Figure 1.3 Schematic diagrams depicting cytochrome c's main maturation systems. System I is based on E. coli, System II on B. subtilis and System III on S. cerevisiae. Question marks indicate where the haem route is unknown. Image adapted from Mavridou (2013).[25] 5

Figure 1.4 Schematic diagram depicting cytochrome c assembly mechanism via the HCCS enzyme from system III. The red square represents the hemic prosthetic group while the black sequence above represents the c-type apocytochrome, highlighting the haem binding motif's amino acid residues (CXXCH). Numbers correspond to human HCCS protein, with the initial methionine as residue 1. Image adapted from Francisco (2012).[30] 6

Figure 1.5 Amino acid residue sequence of the human HCCS. The second residue, G, can suffer post-translational modification by addition of an N-myristoyl group (lipidation). [21] 7

Figure 1.6 Amino acid residue distribution of the human HCCS. Graph obtained using information provided by ExPASy software and UniProtKB.[21] 8

Figure 1.7 Schematic mechanism of cytochrome c interacting with the mitochondrial membrane. The purple-containing molecule is cytochrome c, while the green-headed phospholipids represent cardiolipin. Figure adapted from Thong (2018).[39] 9

Figure 2.1 Plasmid pBTR1. CYC1 is flanked by restriction enzyme's Nco I, EcoRI and Kpn I cleavage sites. CYC3's sequence appears after, being flanked by EcoRI and XbaI. This plasmid confers resistance against ampicillin. 11

Figure 2.2 Plasmid pBAD202. CYC3 is flanked by TOPO's overhang sequences, with a V5 epitope and extra nucleotides between itself and the histidine tail. 13

Figure 2.3 Plasmid pETBlue-2. CYC3 is contained within the highlighted area. Image adapted from Novagen. 13

Figure 2.4 NZYDNA Ladder III. Relative band intensities are due to different amounts of DNA, as demonstrated on the right side of the figure. 15

Figure 2.5 Plasmid pBAD/Thio-TOPO. CYC3 is contained within the above highlighted area. Image adapted from Invitrogen. 16

Figure 2.6 NZYColour Protein Marker II. Composed of 12 extremely pure pre-stained proteins, with two reference bands having a red (75 kDa) or green (25 kDa) covalently bound chromophore. 19

Figure 2.7 Western Blot cassette assembly system. The "cassette" is the perforated component in the figure, where every other component will be inserted and closed. 19

Figure 3.1 SDS-PAGE gel with partially purified unknown protein. Polymerised with 12% acrylamide, the first well has 10 µL of the aforementioned sample, while the second has 20 µL. The black arrow points to the excised band sent for sequencing. 23

Figure 3.2 Image of the membrane with best conditions obtained from expression tests. Three conditions were simultaneously tested, medium (LB and TB) with and without induction (with arabinose), temperature (37 ºC and 37/30 ºC) and time (4, 24 and 48 hours). 24

Figure 3.3 Chromatogram of first affinity purification. Chromatogram representative of the entire purification process, with several peaks which might correspond to lyase. 25

Figure 3.4 Chromatogram of first successful affinity purification. Representative portion of the entire chromatogram, with peak obtained using 10% elution buffer. This implies a low affinity of the protein for the used column. 25

Figure 3.5 DLS results of fraction 30. On the left, size distribution by intensity of fraction 30 had two peaks with 25.6% and 74.4, indicating the possibility of a dimer. On the right, size distribution by volume had one peak, which points towards fraction 30 existing as a monomer with a diameter of almost 20 nm. 26

Figure 3.6 Western blot of HisTrap fractions. Representative portion of the entire chromatogram, with peak obtained using 10% elution buffer. This implies a low affinity of the protein for the used column. 27

Figure 3.7 SDS-PAGE of concentrated samples resulting from size-exclusion chromatography. "With" – Samples with β-mercaptoethanol; "Without" – Samples without β-mercaptoethanol. Although samples seem more contaminated due to higher background noise, this is due to the use of BlueSafe staining instead of Coomassie Blue. While the latter is more readily available, the former is more sensitive to lower concentrations of protein. 28

Figure 3.8 Images of SDS-PAGE and Western Blot done on fraction 30. On the left, the SDS-PAG shows that the most noticeable bands are below the 20 kDa molecular marker, which is 10 kDa less than expected, as lyase is around 30 kDa. On the right, the same sample used in the third lane of the SDS-PAGE, 20 µL, was subjected to a Western Blot assay, which shows, once more, that new bands appeared and that the clearest one has an incorrect mass for it to be lyase. 29

Figure 3.9 Chromatogram of third affinity purification process. Representative portion of the entire chromatogram, with peak obtained using 40% elution buffer. Although not as intense as in the previous purifications, the peak is sharper and demonstrates that the sample has more affinity for the used column. 30

Figure 3.10 Chromatogram and SDS-PAGE of the samples obtained from the anionic purification. On the left, representative portion of the entire anionic purification process, with peak 1 obtained using 30% elution buffer and peak 2 obtained using 40% elution buffer. On the right, SDS-PAGE with representative fractions of the entire purification. The first lane corresponds to the sample that was obtained and concentrated from the affinity purification, and the fractions called "peak" correspond to the most intense portion of the peak. 30

Figure 3.11 Chromatogram and SDS-PAGE of the samples obtained from the size-exclusion purification. On the left, representative portion of the entire anionic purification process, with the smallest peak (in blue) being the most likely to be lyase, according to the column's elution profile. On the right, SDS-PAGE with representative fractions of the entire purification. The first lane corresponds to the sample that was obtained and concentrated from the anionic purification, and the other lanes correspond to the fractions obtained from both peaks. 31

Figure 3.12 SDS-PAGE of the best E. coli strains in the expression tests. Results from strain JM109 are on the left, from strain LMG194 on the right and from XL2 Blue on the bottom. The desired band, corresponding to lyase, should appear between the green molecular marker and the black one directly above it, which corresponds to a range between 25 and 35 kDa. 34

Figure 3.13 Western Blot of the best E. coli strains in the expression tests. All samples represented correspond to the 48-hour collection period. With a more sensitive assay which uses

an antibody that detects the histidine tail in the recombinant protein, there is no clear band in the area of interest. Every band appeared after the membrane was allowed to develop for several minutes, which indicates non-specific binding of the antibody. 35

Figure 3.14 Colony screenings of colonies transformed with pETBlue-2 ligated with CYC3. On the left is the colony screening of the original five colonies obtained from the transformation of pETBlue-2 with CYC3 into E. coli NovaBlue Singles competent cells, with colonies 2 and 5 displaying positive results. On the right is the result of the colony screening of the re-inoculated colonies 2 and 5, with positive results existing in colonies 2 through 6. 38

Figure 3.15 Image of the Western Blot membrane with pETBlue-2's expression tests. Three conditions were simultaneously tested, medium (LB and TB), presence of an inducing compound (IPTG) and time (4,8, 24, 30 and 48 hours). On the left are conditions collected from E. coli BL21 (DE3) and on the right are conditions collected from E. coli JM109. 39

Figure 3.16 Chromatogram and SDS-PAGE of the samples obtained from the affinity purification. On the left, representative portion of the entire affinity purification process, with a peak obtained using 20% elution buffer. On the right, SDS-PAGE with representative fractions of the entire purification. The first lane corresponds to the raw protein extract, while the second corresponds to the flowthrough – an almost identical elution profile is extremely evident. 40

Figure 3.17 SDS-PAGE images representative of the entire purification process. On the left are represented the fractions obtained from the size-exclusion purification step, with F11-12 belonging to a more intense peak with a higher molecular weight than lyase and F16-18 belonging to the peak with a molecular weight closer to lyase. On the right are fractions belonging to the second anionic chromatography's fronting peak, which have a molecular weight close to 75 kDa (red marker). 40

Index of tables

Table 2.1 Designed Primers. List of designed primers and important characteristics. Synthetic oligonucleotides were provided by NZYTech, except for the last two pairs, provided by Eurofins Genomics. The latter company does not specify melting temperatures above 75 ºC. 12

Table 2.2 Expression vectors. List of every utilised expression plasmid and their most relevant characteristics..... 12

Table 2.3 PCR Parameters. Cycling parameters for CYC3 amplification using NZY-Tech's Proof Polymerase for a total volume of 50 µL..... 14

Table 2.4 PCR Parameters. Cycling parameters for site-directed mutagenesis using NZY-Tech's protocol and Proof Polymerase for a total volume of 50 µL. 15

Table 2.5 PCR Parameters. Initial cycling parameters for HCCS amplification using a mixture of DNA polymerases for a final volume of 25 µL. 17

Table 2.6 Escherichia coli strains. List of every utilised bacterial strain and their most relevant characteristics, based on Casali (2003).[57,58,59,60,61,62] 17

List of Abbreviations

Ara – Arabinose

BCA – Bicinchoninic acid

BSA – Bovine serum albumin

CCHL – Cytochrome *c* haem lyase

CL – Cardiolipin

Cys – Cysteine

DAB – 3,3' – diaminobenzidine

DLS – Dynamic light scattering

***E. coli* –** *Escherichia coli*

HCCS – Holocytochrome *c* synthase

His – Histidine

IMS – Intermembrane space

IPTG - Isopropyl β-D-1-thiogalactopyranoside

IUBMB – International Union of Biochemistry and Molecular Biology

LB – Lysogeny broth

mDNA – Mitochondrial DNA

Met – Methionine

MHC – Multiheme cytochrome

MLS – Microphthalmia with linear skin defects syndrome

MS – Mass Spectrometry

nDNA – Nuclear DNA

OD – Optical Density

PBST – Phosphate buffered saline with Tween

PCR – Polymerase chain reaction

PVDF – Polyvinylidene difluoride

SDS-PAGE – Sodium dodecyl sulphate polyacrylamide gel electrophoresis

SEC – Secretory pathway

TB – Terrific broth

TOM – Translocase of the outer membrane

1. Introduction

1.1. *C*-type cytochromes

What is a cytochrome? Originating from the combination of the words *cyto* (cell) and *chrome* (colour), it refers to a haem-containing red protein ubiquitous to most organisms, from Archaea to Eukarya.

First described by MacMunn in 1884 as histohaematin, it was discovered through spectroscopic examination of several tissues from a wide range of organisms and classified as a respiratory pigment.[1] Only after approximately four decades were they mentioned again, or "rediscovered", by entomologist David Keilin. In the 1920's, he ascribed these molecules with their current designation and classified them according to the position of their lowest energy absorption band in the reduced state. Thus, he classified these "cellular pigments" as cytochromes *a* (605 nm), *b* (around 565 nm) and *c* (550 nm).[2]

A cytochrome is an electron shuttle/relay which is constituted by a polypeptide chain and by one or several metal-containing cofactors, which are haems. Regarding the polypeptide backbone, it can exist as both a globular or a membrane protein within organisms. In prokaryotes, such as bacteria, it is encoded and expressed in the cytoplasm, being transported into the periplasm. In eukaryotes, it is encoded within nuclear DNA and is expressed in its apo- form by cytoplasmic ribosomes. After binding to its prosthetic group in the mitochondria's intermembrane space, it remains there, being loosely associated with the inner membrane. In photosynthetic organisms, apocytochrome's binding to haem occurs in the lumen of chloroplasts and in the mitochondrial intermembrane space.[3]

The cytochrome cofactor, the haem (Figure 1.1), is a highly conjugated ring system which coordinates a metal ion with the tetradentate porphyrin ring and up to two axial ligands. It has been used in nature as a cofactor, with all of its modifications (farnesyl addition, oxidation, thiol reduction, formylation, double hydroxylation, etc.), for over 3 billion years.[4]

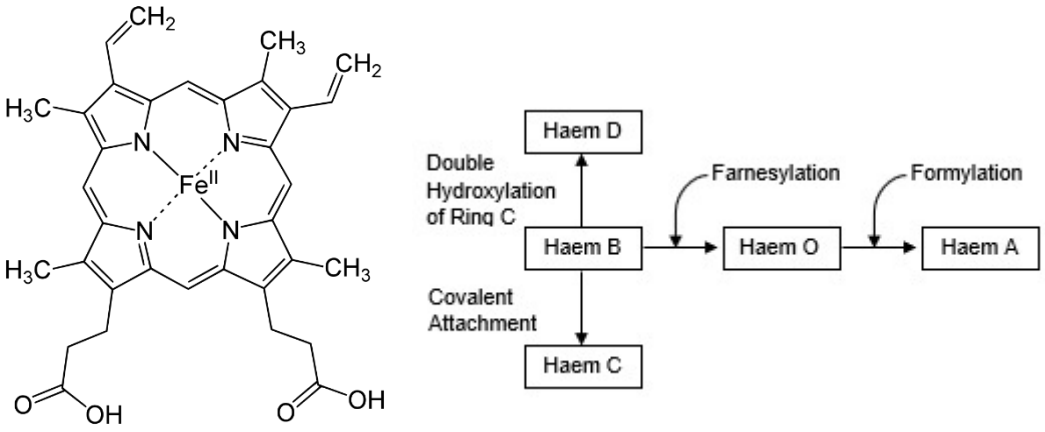

Figure 1.1 Haem (Protoporphyrin IX) coordinated with a ferrous iron (Fe^{2+}) atom[5] and haem formation. On the left, the conjugated double-bond system is responsible for the interaction of haem with visible light. On the right are represented the necessary chemical modifications to convert haem *b* into the different types of haem. Figure adapted from Koch (2007).[6]

Also known as protohaem IX, this prosthetic group was discovered and characterised according to its specific chemical environment which, in turn, yields distinctive spectral absorptions. The most biologically prominent haems are haem *a*, *b*, *c* and *o* – when isolated, they

are named with a capital letter and, when interacting with proteins, they are named with lower case letters.[7]

Haem *B* is the most common[8] and abundant type[9], with iron in either a pentacoordinate or hexacoordinate state due to the single or double coordination bonds between itself and the protein side-chain, respectively. This prosthetic group is synthesised through an enzymatic pathway called porphyrin synthesis, which begins with succinyl-CoA (supplied by the citric acid cycle) and the amino acid glycine to form aminolaevulinic acid in higher eukaryotes. In organisms such as plants, most bacteria and algae, glutamate is used to form aminolaevulinic acid as a haem precursor.[10] After a series of complex enzymatic reactions, haem *B* is obtained within the mitochondria (or in the cytoplasmic membrane)[4] with Fe^{2+} being added into the tetrapyrrole ring by ferrochelatase, the last enzyme involved in such a process. However, some bacteria can also use enzymes orthologous to ferrochelatase under conditions with different concentrations of oxygen, as an alternative to the classical pathway described above.[11]

Haem *A* can be obtained from haem B through an oxidation of a methyl side chain to an aldehyde and a farnesyl addition to a vinyl group; Haem *O*, on the other hand, only requires the second modification, and can also function as substrate for haem *A* synthesis[4] (Figure 1.1).

As for haem *C*, the only difference with haem *B* is that the two available vinyl groups become involved in thioether covalent bonds with cysteine residues of the protein scaffold. As a result, cytochromes can be classified according to their haem environment, originating three major classes of cytochromes: cytochrome *a*, *b* and *c*.

Cytochrome *c* belongs to a superfamily of all-alpha proteins[12] – proteins comprised entirely of alpha helices with the possibility of a few isolated beta sheets. These proteins can either be monohaem cytochromes, such as the mitochondrial cytochrome *c*, or multihaem cytochromes (MHCs)[13], such as MtrA or CymA from *Shewanella spp*, which have 10 and 4 haems *c*, respectively.[14]

Furthermore, they can still be subdivided into four[15] classes according to their sequences, even if such classification attempts within the realms of Biology never seemed to gain much consensus. Even so, the classes that do have an agreement to a certain extent are the following: Class I encompasses low-spin soluble cytochromes *c* from both bacteria and mitochondria with a hexacoordinated (through a methionine close to the C-terminal) haem near the N-terminal[15], such as the mitochondrial cytochrome *c*; Class II includes both high-spin and low-spin helical bundle-forming cytochromes with haems located near the C-terminal, although the former is pentacoordinated while the latter is hexacoordinated[16], such as cytochrome *c*' and cytochrome c_{556}, respectively; Class III pertains to low redox potential cytochromes with multiple haems which possess around 30 to 40 amino acid residues per bis-histidine[15,16] or histidine-methionine (His-Met)[17] coordinated prosthetic group, such as cytochrome c_7. Finally, if an additional class is accepted, Class IV, it incorporates cytochromes with extra prosthetic groups besides haem *c*, such as flavocytochromes.[15] However, some authors propose that this last group presents quite a heterogeneous arrangement of cytochromes, thus reserving this subdivision for when a more regular group is proposed.[16]

This kind of classification also provides some insight as to what to picture regarding cytochrome *c*'s expected absorption spectra. A typical cytochrome *c* absorption spectrum will possess three peaks in the visible range – α, β and γ – with α being located near 550 nm[16,18], β near 520 nm and γ near 416 nm.[19] Given the haem's apparently constant structure, the holoprotein will have a visible absorption spectrum which will only depend on its surrounding environment and backbone ligands with the haem prosthetic group, which will allow to further differentiate between these different types of molecules.[16]

Cytochrome *c* is capable of having different reduction potentials in nature, ranging from -550 mV to +450 mV (versus the standard hydrogen electrode)[9], which enables this molecule to have several functions as well. Cytochrome *c*'s action mechanism depends on the interconversion of its core iron reduction state (between Fe^{2+} and Fe^{3+}), whose ionisation energy suffices for cytochrome *c* to be active.[18] Thus, it is mostly used by organisms in electron transfer processes, such as respiration, photosynthesis and detoxification (via electron transport in prokaryotes)[3].

In the mitochondria of eukaryotes, it is responsible for transferring electrons between complex III (ubiquinone: cytochrome *c* oxidoreductase) and complex IV (cytochrome oxidase).[18]

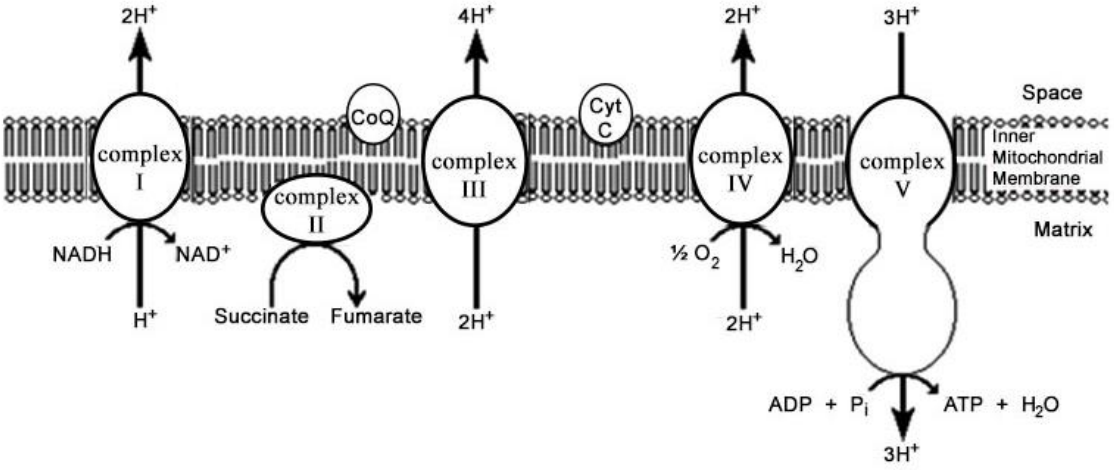

Figure 1.2 Schematic representation of the electron transport chain in eukaryotes. Cytochrome *c* (abbreviated as cyt. c) is used as an electron shuttle in oxidative phosphorylation to generate ATP.

Besides their functions as electron carriers, *c*-type cytochromes can also be found in enzymatic active sites and can also trigger apoptosis within cells.

Human cytochrome *c* is encoded by the gene CYCS, located in chromosome 7[20]; it has 105 amino acid residues and a mass of 11.749 Da. [21] Being encoded within the nuclear genome and synthesised in the cytosol, but functioning in mitochondria, it requires import into its final organelle. Furthermore, it is also synthesised lacking its necessary cofactor, which is added as a post-translational modification. How does cytochrome *c* maturation occur?

1.2. Maturation Systems

Cytochrome *c* biogenesis systems, also known as maturation systems, catalyse the post-translational processing required to obtain a fully-functioning cytochrome *c* holoprotein, that is, the covalent attachment of haem to the protein backbone.[4]

This cytosolic polypeptide has in its amino acid sequence a putative haem binding motif which is then recognised within each system to ligate the vinyl groups of the haems to the reduced thiols from cysteines (Cys) and the histidine to the core iron – Cys-Xxx-Xxx-Cys-His, also presented as CXXCH.[22] In addition to this consensus sequence, other unconventional binding motifs exist in nature, such as CXXXCH, CXXXXCH[23], $CX_{15}CH$ (*Wolinella succinogenes*)[24], A/FXXCH (eukaryotes), CXXCK (bacteria)[23] and AA/GQCH (Euglenozoa phylum)[5].

A minimum of six systems are recognised[25] by some authors. Once again, classifications and divisions are not final and are always subject to changes. As such, from the aforementioned six systems, the first three main maturation systems are universally accepted and recognised: Systems I, II and III. The first two are used by prokaryotes, while the last is used by eukaryotic organisms. Moreover, Systems I and II require the secretory pathway (Sec) to translocate the apoprotein, which always contains a signal sequence[4], into the periplasm.

Regarding other lesser known/more specialised systems, some authors are keen to subdivide them into up to three additional groups. System IV is present in oxygenic phototrophs, and requires four integral membrane proteins (Ccb1-Ccb4) to mature cytochrome b_6 with a single thioether bond, bearing some resemblance to System II; System V is present throughout the mitochondria of the Euglenozoa phylum, applying post-translational modifications with only one cysteine residue in the binding motif; System VI is yet another variant of System IV, present in Firmicutes, that attaches the protein via a single-cysteine bond with a different set of proteins, not related to Ccb1-Ccb4. [4,5,25]

As for the main groups (Figure 1.3), System I, also known as Ccm (cytochrome *c* maturation)[23], is composed by either eight[4] or nine[26] integral membrane proteins (CcmA to CcmH or CcmI, respectively) and is used by α-, γ-, some β- and some δ-proteobacteria, by plant mitochondria, some protozoal mitochondria, Deinococci and Archaea.[4,23] It is the most complex system, and it is possible to organize it in three distinct functional modules, with each one of them being responsible for one step of the cytochrome *c* assembly process. Module 1 (CcmA-CcmE) spends ATP molecules within the cell to translocate haem *b* across the membrane with the final goal of preparing and relaying the prosthetic group to Module 3; Module 2 (CcmG) is responsible for the transportation and chaperoning of the reduced state of the other fraction of this molecule, apocytochrome *c*; Module 3 (CcmFH/I) is therefore left with the task of catalysing the covalent ligation reaction between both components.[26]

System II, called Ccs (cytochrome c synthesis), has at least three or four essential proteins – CcdA, ResA and either CcsA and CcsB or a fusion of both, CcsAB – and occurs in β-, δ- and ε-proteobacteria, Gram-positive bacteria, cyanobacteria and in plant and algal chloroplasts. In those organisms, the CcsAB complex is responsible for haem transportation to the *p*-side of the membrane, recognition of the apocytochrome substrates and catalysis of the covalent bonds. All

that is left for the other two components of this simpler system, CcdA and ResA, is to reduce the cysteine residues in the protein involved in bond formation, which occurs before CcsAB catalyses the formation of the covalent bonds.[26]

System III is the most simple and elegant solution that organisms developed to fully mature a *c*-type cytochrome. To date, only one protein, cytochrome *c* haem lyase/holocytochrome *c* synthase (CCHL/HCCS, depending on the organism) has been identified as essential, and the apocytochrome import into mitochondria needs no membrane potential to be accomplished nor signal sequence.[27] Nevertheless, some disadvantages arise due to this simplicity: it is the most difficult system to study and thus, the most poorly understood one. The enzyme has never been completely purified[23], and knock-out mutation assays are not possible when only one protein comprises the entire system to be studied.

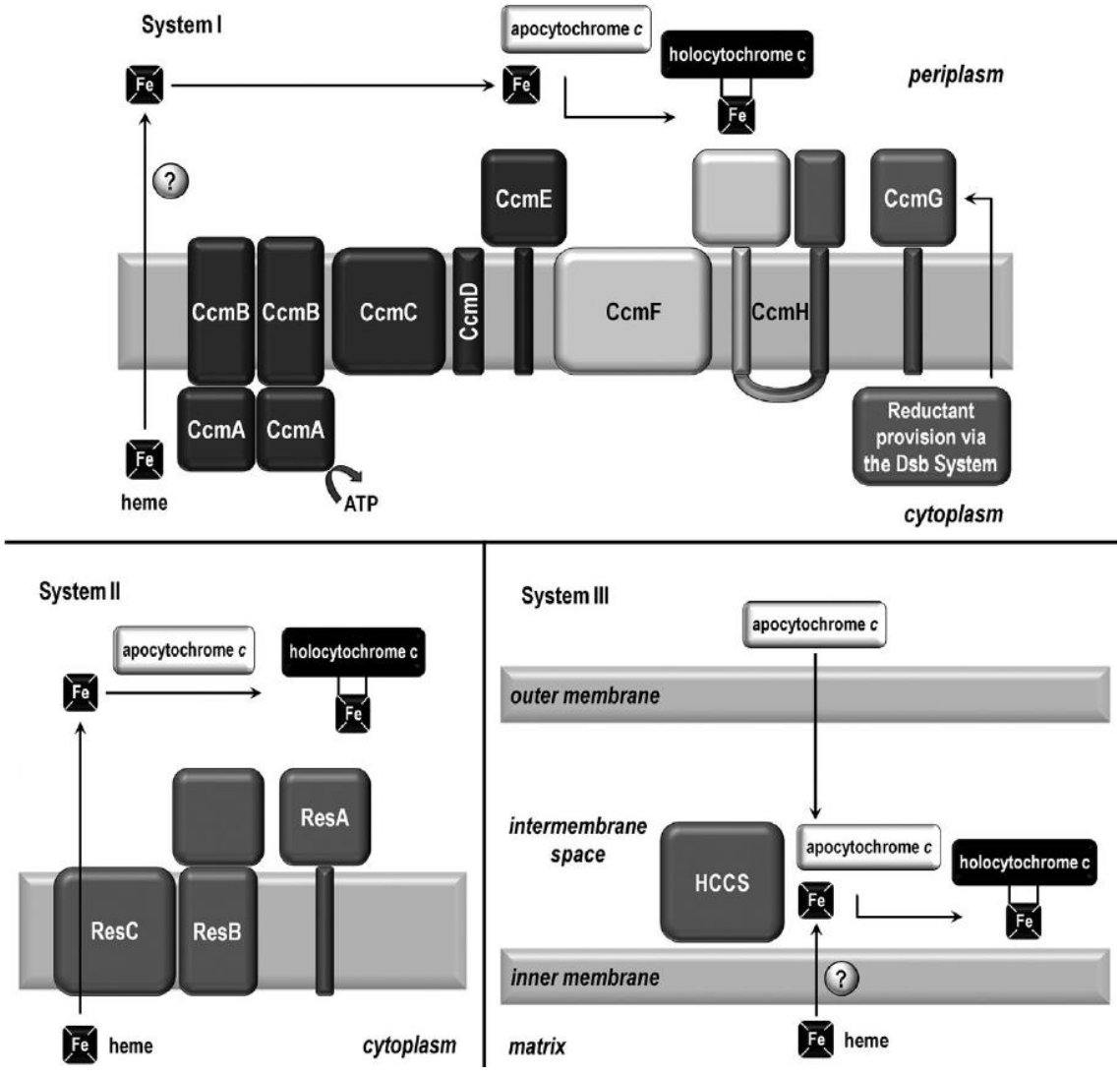

Figure 1.3 Schematic diagrams depicting cytochrome *c*'s main maturation systems. System I is based on *E. coli*, System II on *B. subtilis* and System III on *S. cerevisiae*. Question marks indicate where the haem route is unknown. Image adapted from Mavridou (2013).[25]

1.3. Lyase

Cytochrome *c* haem lyase (CCHL, in yeast), also known as holocytochrome *c* synthase (HCCS, in humans) or simply lyase, is the only protein comprised in maturation system III, i.e., lyase is system III. As such, this makes it a protein preceded by its fame, not only because it relates to eukaryotic cytochrome *c* maturation mechanisms, but also due to aforementioned reasons, as it keeps on eluding researchers during the recombinant expression and purification procedures[28].

In this system, the necessary haem binding motif which serves as the recognised substrate to activate the enzyme's catalytic activity is comprised of fourteen amino acid residues (instead of the usual CXXCH) located near cytochrome *c*'s N-terminal – K/AGXXL/IFXXXCXXCH.[26,29]

Several lines of research regarding the human version, HCCS, point towards a four-step assembly mechanism involving lyase (Figure 1.4).

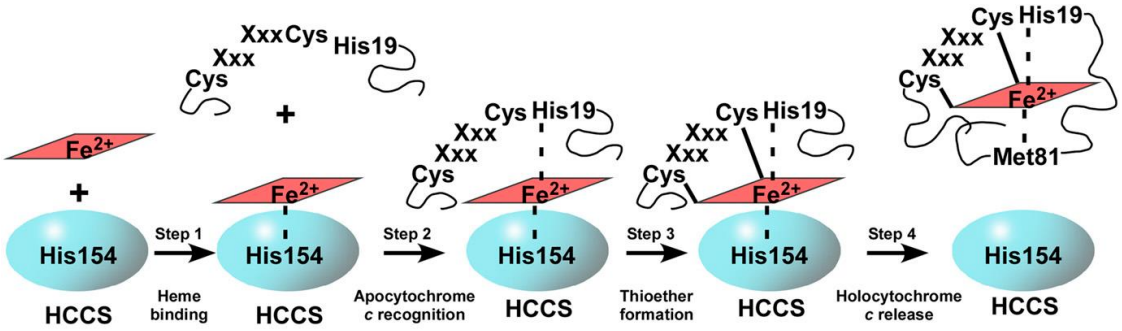

Figure 1.4 Schematic diagram depicting cytochrome *c* assembly mechanism via the HCCS enzyme from system III. The red square represents the hemic prosthetic group while the black sequence above represents the *c*-type apocytochrome, highlighting the haem binding motif's amino acid residues (CXXCH). Numbers correspond to human HCCS protein, with the initial methionine as residue 1. Image adapted from Francisco (2012).[30]

The first step is the haem binding by lyase. After Fe^{2+} addition to a haem in the mitochondrial matrix by ferrochelatase, haem *b* needs to cross the internal membrane, as this binding takes place in the mitochondrial intermembrane space (IMS). There are no certainties as to how haem is transported across the membrane, but yet another peptide motif (Cysteine-Proline-Valine, haem-regulatory) was suggested, as was the residue His154 from HCCS.[30] While the motif is not present in all lyases, nearly all cytochromes have at least one histidine residue capable of interacting as first axial ligand. Likewise, His154 was shown to establish this type of interaction with a haem molecule, suggesting it occupies a position in a haem-binding crevice within the protein's structure. [28,30–32]

The second step is apocytochrome *c* binding to lyase. Since this apoprotein is synthesised in the cytosol, it needs to be transported across the mitochondrial membrane, which is a task performed by the translocase of the outer membrane (TOM) without the aid of a signal sequence[5]. However, lyase also plays a role in cytochrome *c* translocation, as the covalent attachment of haem is intrinsically coupled to the apoprotein's import – the concentration of lyase appears to be correlated with the amount of imported apocytochrome *c*.[3] For this, one requirement need to be met, which is the maintenance of the haem iron in the reduced state. *In vitro* studies also showed the presence of both NADH and a flavin nucleotide, either FAD or FMN, as being capable of

controlling this reduced state of the prosthetic group.[27,33] In yeast, an accessory protein called Cyc2p, encoded by the gene CYC2, also seems to be involved in haem reduction.[32] After entering the IMS, His19 from the apoprotein provides the second axial ligand for the haem, which promotes the stereospecificity of the covalent bonds as it helps to position the following interacting ligands. If the haem iron is improperly bound, i.e. in the wrong orientation, the next step in the process won't occur. [31]

The third step is the formation of the thioether covalent bonds between the cysteines of the apoprotein and the vinyl groups of the haem. The first bond to be formed is between Cys18 and the haem's 4-vinyl, while the second occurs between Cys15 and the 2-vinyl group. This order is further supported by the existence of organisms in nature which only possess Cys18 to establish covalent interactions – consequently, CXXXH as a binding motif is absent in nature.[31]

The fourth and final step in cytochrome *c* maturation is the release of the holoprotein from lyase. After covalent interactions occur, the haem *c* is left distorted or, more specifically, "puckered". This event weakens the interaction between cytochrome *c* and haem lyase, promoting dissociation of both proteins and subsequent exchange of the second axial ligand – His154 from lyase is replaced by Met81 from the polypeptide backbone of the holoprotein – for the proper final folding to take place. [32] In addition to these residues, it was also shown that iron in its reduced state contributes towards release of the holoprotein by providing stability for the subsequent folding events[30] and by accelerating this process.[32]

Such a crucial enzyme was first identified back in 1987 as being encoded by the nuclear gene CYC3.[34] This gene is located in chromosome X and, after translation, it originates a 268 amino acid polypeptide (Figure 1.5) with a molecular mass of 30602 Da.[35]

```
        10         20         30         40         50         60         70         80         90        100
MGLSPSAPAV AVQASNASAS PPSGCPMHEG KMKGCPVNTE PSGPTCEKKT YSVPAHQERA YEYVECPIRG TAAENKENLD PSNLMPPPNQ TPAPDQPFAL
       110        120        130        140        150        160        170        180        190        200
STVREESSIP RADSEKKWVY PSEQMFWNAM LKKGWKWKDE DISQKDMYNIIRIHNQNNEQ AWKEILKWEA LHAAECPCGP SLIRFGGKAK EYSPRARIRS
       210        220        230        240        250        260
WMGYELPFDR HDWIINRCGT EVRYVIDYYD GGEVNKDYQF TILDVRPALD SLSAVWDRMK VAWWRWTS
```

Figure 1.5 Amino acid residue sequence of the human HCCS. The second residue, G, can suffer post-translational modification by addition of an N-myristoyl group (lipidation). [21]

According to the International Union of Biochemistry and Molecular Biology (IUBMB) enzyme nomenclature system, enzyme EC 4.4.1.17 catalyses the reaction between apocytochrome *c* and haem, thus being called holocytochrome-*c* synthase.[36] It was first named cytochrome *c* synthetase or cytochrome *c* haem lyase, although the most accepted terminology in the present is the one used by IUBMB. However, existing literature most commonly uses the term "haem lyase", which can be misleading, as the lyase function is the reverse of the enzyme's usual reaction.[5]

Within cells, it is located in the outer face of the inner mitochondrial membrane, but several studies indicate no transmembrane helices in its structure, suggesting a membrane association but not integral location.[5,26]

According to ExPASy software, ProtParam[37], HCCS has a theoretical pI of 6.25 with a total of 35 negatively charged amino acid residues and 33 positively charged residues (Figure 1.6).

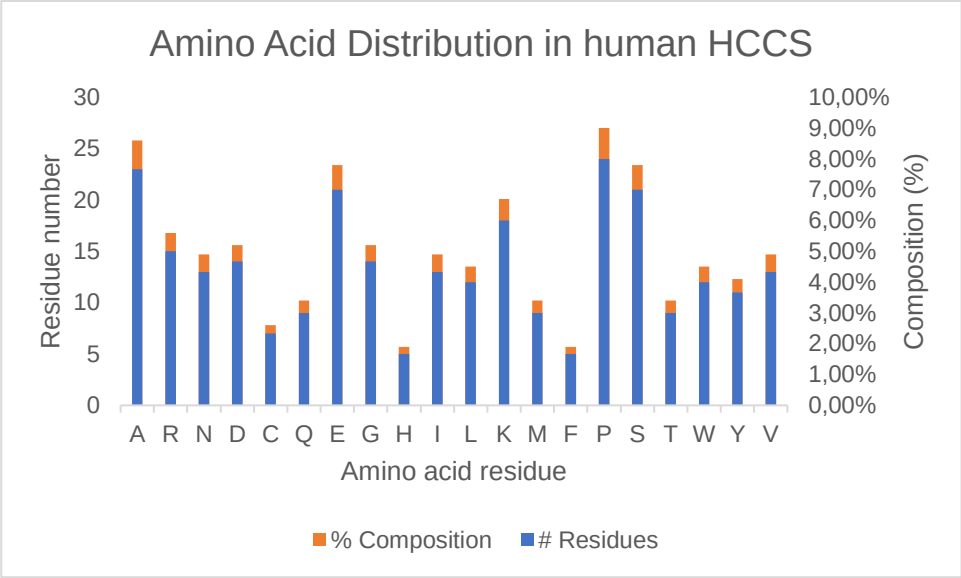

Figure 1.6 Amino acid residue distribution of the human HCCS. Graph obtained using information provided by ExPASy software and UniProtKB.[21]

Its extinction coefficient, ε_{280}, is 82765 $M^{-1}cm^{-1}$ if all Cys residues form cystines and 82390 if all Cys residues are reduced. Its estimated half-life in mammals is around 30 hours *in vitro*, while in *Saccharomyces cerevisiae* this decreases to around 20 hours *in vivo* and in *E. coli* to around 10 hours *in vivo*. The computed instability index (60.75) classifies this protein as being unstable, which is in agreement with the difficulties encountered when trying to study System III.[23]

When using the same software for the yeast version of the protein, CCHL, it computed a theoretical pI of 5.28, 37 negatively charged amino acids and 27 positively charged amino acids. Its extinction coefficient, ε_{280}, is 44710 $M^{-1}cm^{-1}$ if all Cys residues form cystines and 44460 if all Cys residues are reduced. As for its estimated half-life, it is the same as for HCCS, and its instability index is very similar (61.44) to its human counterpart.

1.4. Lyase and Health

As both a maturation system and a protein itself, cytochrome *c* haem lyase is involved in a multitude of health-related issues. Since it has a role in holocytochrome *c* formation, a deeper understanding of its mechanism and all that arises from it as a consequence can be manipulated to prevent cytochrome *c*'s actions.

Membrane-bound cytochrome *c* is an activator of apoptosis or programmed cell death. When this cellular mechanism suffers deregulation, it can cause several diseases, such as cancers and neurodegenerative disorders.[38] In humans, a major cytochrome *c*-mediated apoptosis pathway relies on caspases, a subfamily of cysteine proteases. To activate this function, the holoprotein binds to acidic phospholipids in the inner mitochondrial membrane, such as cardiolipin (CL), via four binding sites at the opening of its hydrophobic channel, near the haem crevice.

Binding of cytochrome *c* through electrostatic interactions is responsible for the initial association of the protein with the acidic phospholipid domain, followed by a partial penetration of the membrane and extended lipid anchorage with an acyl chain of CL. This binding is believed

to destabilise the CL packing in the membrane, inducing demixing and sorting of membrane phospholipids. As a result, CL-rich domains with altered phases form, which may help to extract CL from its environment and contribute to its apoptotic migration. Insertion of cytochrome *c* induces the negative curvature in the CL-rich phase, triggering a local transition to a folded state in the membrane, thus encouraging membrane budding with the formation of vesicles engulfing the haem protein (Figure 1.7).[39]

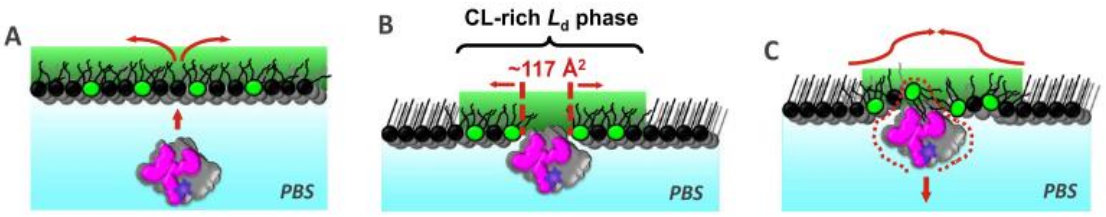

Figure 1.7 Schematic mechanism of cytochrome *c* interacting with the mitochondrial membrane. The purple-containing molecule is cytochrome *c*, while the green-headed phospholipids represent cardiolipin. Figure adapted from Thong (2018).[39]

Furthermore, cardiolipin is oxidised by a peroxidase function of the cytochrome *c*-CL complex, contributing to the protein's translocation into the cytoplasm. After a small amount of cytochrome *c* escapes from the mitochondrial outer membrane into the cytoplasm, the endoplasmic reticulum starts releasing calcium. This overall increase in calcium cations functions in a positive feedback loop, triggering a massive release of cytochrome *c* which activates caspase-9. [38,40] As such, regulating cytochrome *c* formation through HCCS could help to prevent its escape and contribution to the formation of cancer cells in humans.

Regarding HCCS itself, a mutation in its sequence is responsible for a condition in females called microphthalmia with linear skin defects syndrome (MLS). Given that the disease is caused by mutations in CYC3, which is located in the X chromosome, it is a lethal genotype for males.[32] These mutations usually involve chromosomal rearrangements and heterozygous deletion of the gene, while some point mutations, such as E159K, R197X or R217C, can affect the normal function of the enzyme/system III.[5,32] With a lack of functioning lyase, cells become unable to produce energy efficiently. Furthermore, given the importance of cytochrome *c* in apoptosis processes, cells lacking a functioning biogenesis system that can assemble this molecule may be prevented from entering programmed cell death. If this happens, cells undergo necrosis instead, which damages adjacent cells due to inflammation processes. Necrosis resulting from a lack of functional HCCS during early stages of embryonic development may lead to characteristic symptoms of MLS, such as visible eye and skin abnormalities, nail dystrophy, short stature, diaphragmatic hernias and brain, heart and genitourinary defects.[41]

Moreover, HCCS has also been mentioned as having more than one apoptotic role, being implicated in a caspase-independent cell death pathway in injured neurons[32] and in a caspase-dependent pathway through translocation to the cytosol and inhibition of the X-linked inhibitor of apoptosis protein with resulting activation of caspase-3.[5]

As such, if HCCS mechanism of action can be unravelled and thus fully understood, an extremely important line of action can be taken to target such HCCS and cytochrome-*c* related conditions. In the future, specific drugs and genetic therapies targeting either the flawed protein

or the mutated gene, respectively, could aid thousands of people. Although MLS is a rare disease, cancer is the second leading cause of death on a global level, causing almost 10 million deaths in 2018.[42]

Lyase is a never-before completely purified protein that comprises an entire biogenesis system in itself. Furthermore, it has many implications in several health conditions, as previously mentioned, due to being responsible for cytochrome-*c* assembly in the organism. With this in mind, the aim of this project is to understand the molecular mechanisms of cytochrome-*c* as performed by System III. To accomplish such a goal, a complete purification of lyase and its structural and functional characterisation will be pursued.

2. Materials and Methods

2.1. Molecular Biology

Upon project start, plasmid pBTR1 (Figure 2.1) was available at the hosting laboratory. This plasmid contains the gene which encodes *Saccharomyces cerevisiae*'s cytochrome *c*-type haem lyase, known as CYC3, which can be co-expressed with human cytochrome c_1, with gene name CYC1.[43]

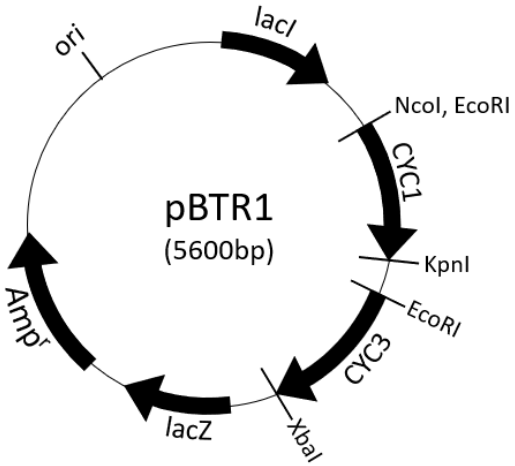

Figure 2.1 Plasmid pBTR1. CYC1 is flanked by restriction enzyme's Nco I, EcoRI and Kpn I cleavage sites. CYC3's sequence appears after, being flanked by EcoRI and XbaI. This plasmid confers resistance against ampicillin.

This plasmid was kindly gifted by the laboratory of Dr. Irene Díaz-Moreno (Universidad de Sevilla) and it was deposited by the laboratory of Gary Pielak (Addgene plasmid #22468). Despite being high-copy and conferring bacterial resistance to ampicillin, it lacks any type of affinity tag, which would greatly help with lyase's purification process.

An affinity tag is a polypeptide sequence fused to the protein of interest which enables it to be purified using affinity chromatography protocols. If needed, the tag can either be chemically or enzymatically cleaved, given the possibility of interfering with the native protein's structure and interactions.[18]

As such, it was decided within the hosting laboratory to insert CYC3 into pBAD202, which features the sequence for a polyhistidine region (called polyhistidine-tag, trademarked as His-Tag) after the gene of interest. A polyhistidine tag usually consists of six amino acid residues which, due to histidine's imidazole ring, exhibit high affinity for divalent nickel column matrices. However, usage of this tag with consequent affinity chromatography purification can lead to co-elution of untagged proteins due to nonspecific binding – two or more contiguous histidine residues suffice. [44]

To accomplish this, synthetic oligonucleotide sequences (known as primers) were designed (Table 2.1) according to the plasmid kit's specifications and a Polymerase Chain Reaction (PCR) was performed to amplify CYC3 from pBTR1 with pBAD202's primers.

This technique allows specific amplification of a DNA sequence through several cycles of DNA denaturation, primer annealing to single-stranded DNA and *de novo* synthesis catalysed by a thermostable DNA polymerase, such as *Taq*[45], or Proof DNA Polymerase.

The former was first isolated from a bacterial species which grows in hot springs, *Thermus aquaticus*, while the latter is an engineered recombinant DNA polymerase. It provides amplification with high fidelity and has 3'-5' exonuclease proofreading capacity. Furthermore, it is highly efficient in amplification of longer PCR products and site-directed mutagenesis, generating blunt-ended products with a significantly lower error rate when compared to that of Taq polymerase.

Table 2.1 Designed Primers. List of designed primers and important characteristics. Synthetic oligonucleotides were provided by NZYTech, except for the last two pairs, provided by Eurofins Genomics. The latter company does not specify melting temperatures above 75 ºC.

Primer Name	Sequence (5' – 3')	Number of bases	Molecular Weight (g/mol)	Melting Temperature (ºC)
Forward_pBAD202	CAC CTA AGA AGG AGA TAT ACA TCC CAT GGG TTG GTT TTG GGC A	43	13307	69.1
Reverse_pBAD202	GGG GCG GAG GAC GAA GAG GAC GGA CC	26	8185	67.2
Forward_pETBlue_2	GAT GGG TTG GTT TTG GGC AGA TCA AAA AAC TAC GGG CAA AGA TAT TGG TGG GGC AGC AGT ATC	63	19695	74.8
Reverse_pETBlue_2	AAT TAA GGG GCG GAG GAC GAA GAG GAC GGA CCC GAG AT	38	11925	71.9
Forward_pBTR1	TCC TCC GCC CCT TTG GTG CCA CGC GGC TCG TGG TCG CAC CCG CAA TTC GAG AAG TAA GAA ATA ATG	66	20251	80.0
Reverse_pBTR1	CAT TAT TTC TTA CTT CTC GAA TTG CGG GTG CGA CCA CGA GCC GCG TGG CAC CAA AGG GGC GGA GGA	66	20411	80.0
Forward_HCCS	TAA ATG GGT TTG TCT CCA TCT GCT CCT GCT GTT GCA GTT	39	11937	71.6
Reverse_HCCS	CGA GGT CCA ACG CCA CCA AGC GAC TTT CAT TCT	33	100018	72.0
Forward_Strep	CCG CCC CTT CAT TGA GTC ACC CCC AGT TTG AGA AGT AGT AAG TAG CC	47	14359	>75.0
Reverse_Strep	GGC TAC TTA CTA CTT CTC AAA CTG GGG GTG ACT CAA TGA AGG GGC GG	47	14559	>75.0

After DNA amplification, the gene was inserted into pBAD202 (Table 2.2), a commercially available system that exploits Topoisomerase I obtained from a *Vaccinia* virus to clone a blunt-end PCR product in the correct direction (Directional TOPO Cloning) with an efficiency of over 90% [46] (Figure 2.2).

Table 2.2 Expression vectors. List of every utilised expression plasmid and their most relevant characteristics.

Plasmid Name	Size (bp)	Supplier	Sequence Information	Antibiotic Resistance	Promoter	Copy Number	C-terminal Tag
pBTR1	5600	Laboratory of Irene Díaz-Moreno	Partial	Ampicillin	IPTG	High	None
pBAD202	4448	Invitrogen	Total	Kanamycin	Arabinose	Low	His-Tag
pETBlue-2	3653	Novagen	Total	Ampicillin	IPTG	High	His-Tag
pBAD/Thio-TOPO	4454	Invitrogen	Total	Ampicillin	Arabinose	High	His-Tag

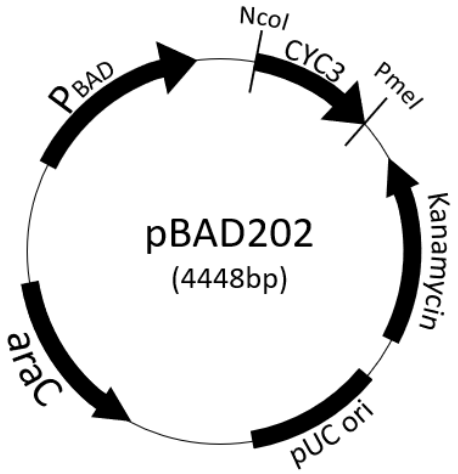

Figure 2.2 Plasmid pBAD202. CYC3 is flanked by TOPO's overhang sequences, with a V5 epitope and extra nucleotides between itself and the histidine tail.

Directional TOPO Cloning relies on double-stranded DNA with a CCCTT cleavage site[47] and a GTGG overhang. These additions permit the creation of an adduct between the nucleotides and the enzyme that can ligate to complementary DNA inserts – with a forward primer containing a CACC in the 5' end. This complementarity between the adduct and the insert enables directionality in the amplification process, thus originating correctly-oriented clones in at least 90% of the cases.[46]

Another molecular biology attempt was to insert CYC3 into pETBlue-2 (Figure 2.3). This plasmid, also commercially available, similarly provides a C-terminal histidine tag whilst maintaining a high-copy number and conferring resistance to ampicillin (Table 2.2).

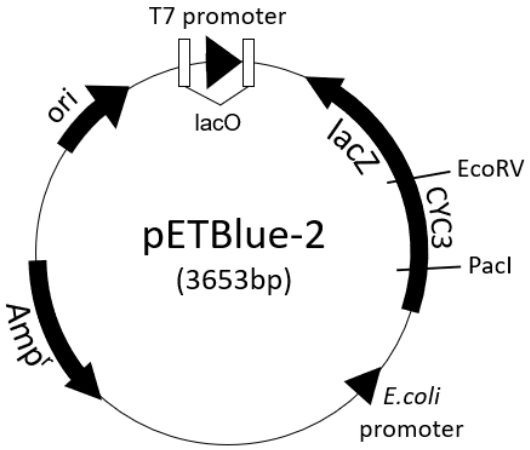

Figure 2.3 Plasmid pETBlue-2. CYC3 is contained within the highlighted area. Image adapted from Novagen.

Following the same steps as in the previous attempt, a pair of synthetic oligonucleotides were designed according to needed specifications, such as requiring two extra nucleotides in the reverse primer sequence to fuse the translated protein with the histidine tag hexapeptide. After designing the primers, the forward sequence was aligned with pETBlue-2's and lyase's sequences, and the melting temperature, T_m, was calculated with the help of an online software

(T_m Calculator, provided by Thermo Fisher)[48]. This sequence alignment was done to verify beforehand if the primers would anneal to the recombinant vector's sequence.

To insert CYC3 into pETBlue-2, the gene was amplified from pBTR1 using pETBlue-2's primers with the following master mixture: 5 µL Reaction Buffer 10x, 1 µL dNTP mix, 10 µL Stabiliser Solution 5x, 1.7 µL dsDNA (to obtain 60 ng of DNA from a 34.4 ng/µL sample), 1 µL Forward Primer and 1 µL Reverse Primer (from a 1:10 working solution prepared with distilled, nuclease-free water), 29.3 µL Milli-Q H_2O and 1 µL of Proof DNA Polymerase. PCR was performed with the following cycle parameters in a T100 thermal cycler (Bio-Rad):

Table 2.3 PCR Parameters. Cycling parameters for CYC3 amplification using NZY-Tech's Proof Polymerase for a total volume of 50 µL.

Cycle Step	Temperature (ºC)	Time (s)	Cycles
Initial Denaturation	95	180	1
Denaturation	95	30	30
Annealing	68	30	
Extension	72	60	
Final Extension	72	600	1
Save/Storage	4	∞	1

After obtaining the amplified fragment, the protocol supplied by Novagen[49] was followed without alterations. For the End Conversion Step required before inserting CYC3 into pETBlue-2, an online calculator[50] was used to compute the necessary amount of amplified DNA according to CYC3's length of 807 bp and the specified 0.05 pmol of dsDNA, which was 27 ng. As such, this End Conversion was done using 2 µL of a 14 ng/µL solution of insert, 3 µL of Milli-Q H_2O and 5 µL of End Mix solution. For the colony screening of transformed colonies, NZYTech's NZYTaq II 2x Green Master Mix kit was used, as well as the provided protocol.

As a third option, a streptavidin tag (Strep-Tag II) was to be inserted into the original plasmid, pBTR1. Strep-Tag II is an eight amino acid fusion protein (WSHPKFEQ) which relies on the biotin/streptavidin binding specificity, allowing elution of low-abundance proteins under mild, physiological conditions.[51]

Therefore, a Strep-Tag II was inserted into pBTR1 with an additional thrombin cleavage site (LVPR/GS) between its sequence and CYC3 to achieve tag cleavage after purification. This was accomplished by site-directed mutagenesis, which creates targeted mutations in double-stranded DNA. Using NZYMutagenesis kit (NZYTech), two synthetic, mutated, primers and Proof DNA Polymerase are used together with a restriction enzyme, Dpn I, to make point mutations in pBTR1. After PCR with Proof Polymerase, the new mutated plasmid has staggered nicks which can resist digestion with Dpn I, an endonuclease capable of recognising and cutting methylated DNA.[52] As such, only the mutated DNA will be available to be introduced into the expression organism.

To maintain the appropriate conditions for primer annealing, the aim was to introduce the least amount of point mutations possible to achieve the desired codons for the proper amino acids. For this, every codon for each amino acid was compared to pBTR1's sequence, with an overhang including part of each component's DNA. Afterwards, the forward primer and the original sequence were aligned to verify if they would still anneal properly to one another, and melting temperatures were calculated.

Firstly, pBTR1 was extracted and cleaned with commercial kits (Miniprep, NZYTech) before site-directed mutagenesis was performed. A master mixture was prepared with the following components: 5 µL Reaction Buffer 10x, 1 µL dNTP mix, 1.7 µL dsDNA (to obtain 60 ng of DNA from a 34.4 ng/µL sample), 0.6 µL Forward Primer and 0.6 µL Reverse Primer (from a 1:10 working solution prepared with distilled, nuclease-free water), 40.1 µL Milli-Q H_2O and 1 µL of Proof DNA Polymerase. PCR was performed with the following cycle parameters in a T100 thermal cycler (Bio-Rad):

Table 2.4 PCR Parameters. Cycling parameters for site-directed mutagenesis using NZY-Tech's protocol and Proof Polymerase for a total volume of 50 µL.

Cycle Step	Temperature (ºC)	Time (s)	Cycles
Initial Denaturation	95	120	1
Denaturation	95	60	18
Annealing	60	60	
Extension	68	480	
Final Extension	68	900	1
Save/Storage	4	∞	1

To verify the efficiency of every procedure, an agarose gel was prepared with 25 mL 1% agarose and 0.4 µL GreenSafe Premium (NZYTech). After polymerisation, the gel was placed in the electrophoresis chamber (Mini-Sub Cell GT, Bio-Rad) and covered with 40 mM Tris (Sigma-Aldrich), 20 mM acetic acid (Scharlau) and 1 mM EDTA (VWR) (TAE) buffer 1x. Samples were prepared with 5 µL DNA and 1 µL NZYDNA Loading Dye 6x (NZYTech). Afterwards, 5 µL were applied in the gel, as are 5 µL of a molecular marker, NZYDNA Ladder III (NZYTech) (Figure 2.4).

75 Volt were then applied to the gel for 40 minutes, and results were revealed with UV light, using the Gel Doc XR+ System (Bio-Rad).

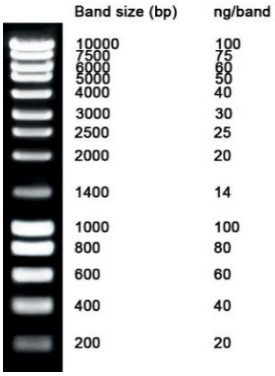

Figure 2.4 NZYDNA Ladder III. Relative band intensities are due to different amounts of DNA, as demonstrated on the right side of the figure.

Besides every Molecular Biology attempt mentioned to date, a different approach was tested: extraction and amplification of the human version of lyase's gene, HCCS. For the extraction, a trained professional collected two 3 mL blood samples within BD Plastic Vacutainer K_2EDTA tubes with lavender Hemogard closures (Fisher Scientific). These tubes contain K_2EDTA sprayed onto the walls, which is a strong anticoagulant used to obtain whole blood samples.[53] They can also be used for some blood bank procedures, such as blood typing and screening, and are preferred by most molecular genetics laboratories for DNA and RNA studies. After sample

collection, the tubes were inverted 180º up to ten times, followed by a centrifugation at 134 x *g* (Nahita 2650 angle centrifuge) for 10 minutes.

The subsequent DNA extraction was performed according to two other protocols[54,55], and the extraction efficiency was verified with an agarose gel, as previously explained. As for the DNA's concentration and purity, they were measured using a NanoDrop apparatus (Thermo Fisher Scientific).

After obtaining the human mitochondrial DNA (mDNA) fraction from the DNA extraction protocol, HCCS was to be inserted into a new vector, pBAD/Thio-TOPO (Figure 2.5). Similar to pBAD202, this commercially available vector also explores the enzyme Topoisomerase I, together with *Taq* polymerase's nontemplate-dependent terminal transferase activity. For this, *Taq* is used to create the insert by PCR, in which it adds a single adenine residue overhang to each 3' of the sequence. This gene sequence will then be inserted into the linearized vector, which is ligated with Topoisomerase I in a single 3' deoxythymidine residue overhang. The complementarity between the bases located at the overhangs allows the insert to be efficiently ligated to the vector in the absence of a DNA ligase, as Topoisomerase I's release from the 3' deoxythymidine supplies the necessary energy for this reaction to occur.[56] However, this high-copy plasmid confers the organism with resistance to ampicillin, whereas pBAD202 is low-copy, with resistance to kanamycin.

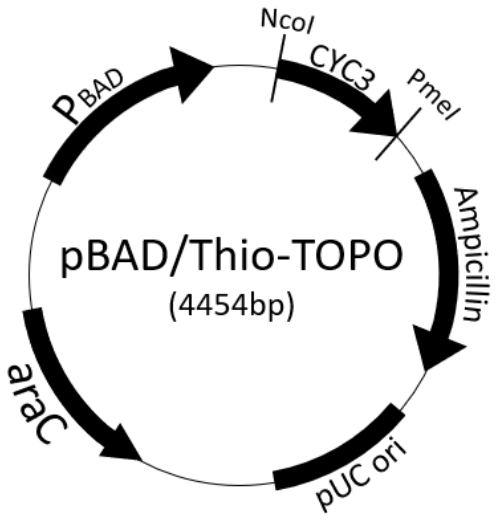

Figure 2.5 Plasmid pBAD/Thio-TOPO. CYC3 is contained within the above highlighted area. Image adapted from Invitrogen.

To amplify HCCS from the extracted mDNA, a pair of synthetic oligonucleotides were designed according to needed specifications, which were to remove the native stop codon from the gene of interest with the reverse primer whilst the original reading frame was maintained. With this, HCCS would be obtained as construct with a polyhistidine tag after protein expression. Primer design was checked for a proper alignment with the plasmid and the gene, and melting temperatures were calculated using the same software as previously mentioned (T_m calculator).

For the PCR using *Taq* polymerase, a master mixture was prepared with the following components: 5 µL Reaction Buffer 10x, 1 µL dNTP mix, 3.8 µL dsDNA (to obtain 60 ng of DNA from a 15.8 ng/µL sample), 1 µL Forward Primer and 1 µL Reverse Primer (from a 1:10 working

solution prepared with distilled, nuclease-free water), 2.5 µL Milli-Q H_2O, 2.5 µL $MgCl_2$, 1 µL of *Taq* DNA Polymerase and 0.1 µL of Proof DNA Polymerase. PCR was performed with the following cycle parameters in a T100 thermal cycler (Bio-Rad):

Table 2.5 PCR Parameters. Initial cycling parameters for HCCS amplification using a mixture of DNA polymerases for a final volume of 25 µL.

Cycle Step	Temperature (ºC)	Time (s)	Cycles
Initial Denaturation	95	120	1
Denaturation	95	30	30
Annealing	67	30	
Extension	72	60	
Final Extension	72	600	1
Save/Storage	4	∞	1

The parameters established in Table 2.4 were subjected to changes required to optimise the amplification of HCCS, as well as the mixture of polymerases. As such, annealing temperatures were changed, gradients were tested, Xpert High Fidelity DNA Polymerase (GRiSP) was used instead of Proof DNA Polymerase and *Taq* was used alone.

2.2. Expression tests

Using a heat shock protocol, several *Escherichia coli* (*E. coli*) strains (Table 2.5) were transformed with pBAD202 or pETBlue-2 to perform expression tests before a full scale growth.

Table 2.6 *Escherichia coli* strains. List of every utilised bacterial strain and their most relevant characteristics, based on Casali (2003).[57,58,59,60,61,62]

Strain Name	Suppression of amber mutations	Transformation of large plasmids	Production of ssDNA	Reduced recombination	Generation of unmethylated DNA	Cloning methylated DNA	Blue-White Screening	High protein yield	Tight control of protein expression	Enhanced eukaryotic protein expression
BL21 (DE3)								x	x	
BL21 Star Rosetta 2								x	x	x
DH5-α	x			x			x			
HB101	x			x						
JM109	x		x	x			x			
LMG194									x	
RP4182	x				x					
Top 10				x		x	x			
Tuner (DE3) pLacI								x	x	x
XL1-Blue	x	x	x	x		x	x			
XL2-Blue	x	x	x	x		x	x			

A heat shock protocol is a standard biology technique performed to promote plasmid uptake by bacteria. More specifically, the sudden change in temperature alters membrane fluidity[63], thus creating pores which permit plasmid DNA entry in the cell. To begin, DNA is inserted into a tube containing competent cells (prepared with a rubidium chloride treatment) and, after 20 minutes in

ice, the tube is shocked for 45 seconds at 42ºC. Finally, cells are placed on ice once more and, after an outgrowth step of 90 minutes with liquid medium, they are plated in agar Petri dishes with the adequate antibiotics.

To begin expression tests, a pre-inoculum was prepared with 20 mL of Lysogeny Broth (LB) liquid medium, 0.1 mg/mL kanamycin (PanReac AppliChem) or ampicillin (Bio-Rad) (for pBAD202 or pETBlue-2, respectively) and a colony resulting from the heat shock transformation protocol in a 100 mL Erlenmeyer flask. LB is a nutritionally rich medium composed of 1% tryptone (Scharlau), 0.5% yeast extract (Scharlau) and 1% NaCl (Fluka)[64].

This inoculum was allowed to grow overnight at 37ºC and 160 rpm a Certomat IS incubator shaker (Sartorius), to ensure proper aeration. On the following day, two inocula were prepared using the previously obtained culture: one with 100 mL LB, 100 µL of kanamycin or ampicillin and 1 mL of pre-inoculum; the second one with the same components, but with 100 mL of Terrific Broth (TB) liquid medium instead of LB. TB is a richer medium when compared with LB, enhancing the yield of plasmid DNA and recombinant proteins by supporting high cellular density, thus extending logarithmic growth phase. It is made with 1.2% tryptone, 2.4% yeast extract, 0.4% glycerol (NZYTech), 0.17 M KH_2PO_4 (PanReac AppliChem) and 0.72 M K_2HPO_4 (PanReac AppliChem).[65]

Optical density at 600 nm (OD_{600}) was measured using a spectrophotometer (Shimadzu UV-1800) and, when it reached a value of around 0.6, both cultures were divided once more, resulting in a total of four different Erlenmeyer flasks. An optical density of 0.6 is used to induce protein expression, assuming that cell cultures are in mid-logarithmic growth phase.[66]

Afterwards, one culture from each media was induced with 1 mM of arabinose (Ara) (Sigma-Aldrich) 1 M for pBAD202 or 1 mM of Isopropyl β-D-1-thiogalactopyranoside (IPTG) (Sigma-Aldrich) 1 M for pETBlue-2. As such, this assay tested four different conditions for each plasmid, named “LB”, “LB + Ara”, “TB” and “TB + Ara” or “LB”, “LB + IPTG”, “TB” and “TB + IPTG”, respectively.

After 4, 8, 24- and 48-hours post-induction (with an additional 30-hour sample taken for pETBlue-2 transformed strains), a 1 mL sample was collected from each condition. Cells were harvested by centrifugation for 4 minutes at 14.000 x *g*, and stored at -20 ºC. Each sample was then treated with 100 µL of NZY Bacterial Cell Lysis Buffer (NZYTech), followed by 20 minutes in a vortex mixer (Labnet) and 20 minutes of centrifugation at 14.000 x *g*. This was to ensure that the obtained supernatant would have the over-expressed protein without it being denatured, due to the detergent’s gentle disruption of the bacterial cell wall while maintaining a physiological pH value.[67]

Finally, the resulting supernatant was applied in a 12% sodium-dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) gel. 20 µL of supernatant were mixed with 5 µL of SDS-PAGE Sample Loading Buffer 5 x (NZYTech). 120 V were applied to the gel submerged in running buffer 1 x – made from a 5 x stock solution composed of 0.125 M Tris base, 0.96 M glycine (PanReac AppliChem), and 0.5% sodium dodecyl sulphate (Sigma-Aldrich) – within a Hoefer SE 260 Mini-Vertical Gel Electrophoresis Unit system (Amersham Biosciences). NZY

Colour Protein Marker II (NZYTech) (Figure 2.6) was used as a standard and, to stain the gel and choose the most appropriate condition, BlueSafe (NZYTech) was used and allowed to develop overnight. This is a one-step protein stain, meaning it does not require a destaining solution to visualise the obtained bands. Moreover, it is safer than Coomassie Blue, as it does not contain dangerous reagents (such as methanol), and it is more sensitive than the traditional stain, since it can reveal 10 ng of protein after only 60 minutes.[68]

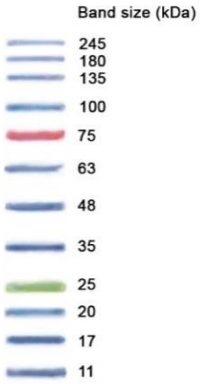

Figure 2.6 NZYColour Protein Marker II. Composed of 12 extremely pure pre-stained proteins, with two reference bands having a red (75 kDa) or green (25 kDa) covalently bound chromophore.

When further confirmation was needed, a Western Blot protocol was applied in addition to BlueSafe, using a Ni-NTA HRP Conjugate antibody (Qiagen). Western Blot, also called immunoblot, is a technique that allows protein separation by size and subsequent identification with primary and secondary antibodies. This is accomplished with protein transfer to a membrane after gel electrophoresis and incubation with antibodies specific for the protein of interest.[69] For this, after SDS-PAGE finished, a cassette is assembled as follows (Figure 2.7):

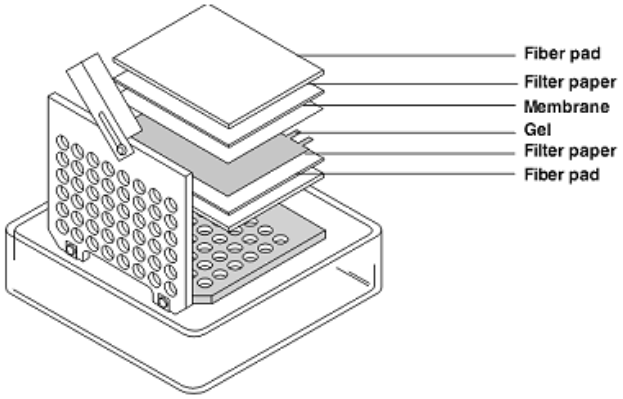

Figure 2.7 Western Blot cassette assembly system. The “cassette” is the perforated component in the figure, where every other component will be inserted and closed.

Open cassette with the black side facing down on a clean, disinfected surface. Afterwards, place a previously transfer buffer-soaked fibre pad on top of it, followed by three soaked filter paper squares (cut to fit the pad), the gel, both Immobilon-P 0.45 µm polyvinylidene difluoride (PVDF) membranes (Merck Millipore) (soaked in both methanol, for 30 seconds, and buffer), the other three filter paper squares and, finally, the second fibre pad. 1 L of transfer buffer was made with 200 mL running buffer 5 x, 200 mL methanol 100% (Riedel-de Haën) and 600 mL H_2O. While

assembling the system, one must take care not to leave air bubbles between layers. As such, a test tube was used to roll over each component (except for the membranes).

The gel was then run for 30 minutes at 100 V with an ice pack and a magnetic stirrer to transfer the proteins from the gel to the PVDF membranes. After transferring, the membrane with most protein was blocked for 1 hour with blocking solution to inhibit non-specific protein binding – phosphate buffered saline with 0.1% Tween 20 (Carl Roth) (PBST), 3% bovine serum albumin (BSA) (Fisher Scientific) and 5% milk powder (Molico, Nestlé) – and then washed with PBST three times for 5 minutes.

Afterwards, it was incubated with 25 mL primary antibody solution (25 mL PBS, 25 µL Tween, 25 µL Ni-NTA HRP Conjugate antibody) for two hours followed by yet again another three washes with PBST, as before. Finally, the membrane was revealed with 10 mL of a 3,3' – diaminobenzidine (DAB, Sigma-Aldrich) and 0.02% hydrogen peroxide (PanReac AppliChem) 1:1 solution. To prepare this final solution, DAB has to reach room temperature and 5 mg of the reagent are dissolved in 5 mL Tris 50 mM pH 7.2.

Membrane blocking and washing were done at room temperature on a stirring platform (ProBlot Rocker 25, Labnet).

2.3. Heterologous Protein Expression and Purification

After choosing the best growth conditions, a scale-up growth was performed. For this, cells were grown in 2 L TB medium, within 5 L Erlenmeyer flasks, with 2 mL of 100 mg/mL kanamycin (to achieve a final concentration of 0.1 mg/mL) at 37 ºC and 160 rpm inside a Certomat IS incubator shaker (Sartorius). 24 hours after induction with 2 mL of arabinose 1 M (to achieve a final concentration of 0.1 M), cultures were harvested by centrifugation at 11 325 x *g* for 10 minutes using an Avanti J-26I centrifuge with a JA-10 rotor (Beckman Coulter). The resulting pellet was stored overnight in a -20 ºC freezer.

The thawed pellet was solubilised with 20 mM phosphate buffer at pH 7.6 containing a protease inhibitor cocktail tablet (cOmplete EDTA-free Protease Inhibitor Cocktail, Roche) and DNase I from bovine pancreas (Sigma-Aldrich). This mixture suffered mechanical disruption within a French press at 16 000 psi and was ultracentrifuged (Optima LE-80K with a 45Ti rotor, Beckman Coulter) at 198 735 x *g* (calculated with a centrifuge rotor speed calculator[70]) for 60 minutes.

The resulting supernatant was used to start the purification process. As such, it was injected into an affinity column, HisTrap HP 5 mL (GE Healthcare), and eluted with an imidazole gradient ranging from 20 mM to 500 mM in a 20 mM potassium phosphate buffer pH 7.4 with 500 mM NaCl. The eluted fractions containing lyase were concentrated and dialysed overnight with 20 mM Tris buffer pH 7.6 and agitation, at 4 ºC. These fractions were chosen according to results obtained with either BlueSafe, Western Blot or a combination of both.

Before injection into the second chromatographic column due to the presence of impurities, the sample was concentrated with a 20 mL 10 kDa Vivaspin (Sartorius) in a 5430R centrifuge

(Eppendorf) at 7142 x *g* and 4 ºC. Afterwards, the concentrated fraction was injected into an anionic exchange chromatographic column, a Q Sepharose Fast Flow (GE Healthcare), and eluted with a NaCl gradient ranging from 0 to 1 M in a 20 mM Tris buffer pH 7.6 and a flow rate of 3 mL/min.

The third and final chromatographic step was a size-exclusion column, Superdex 75 10/300 GL (GE Healthcare), and the elution process used a buffer with 150 mM NaCl and 20 mM Tris pH 7.6 and a flow rate of 0.1 mL/min.

Every column was connected to an Äktaprime plus system (GE Healthcare), which allowed to measure absorbances at 280 nm throughout the entire purification procedure.

To confirm the presence of lyase in every step of the purification, a small amount of the eluted fractions was applied in 12% SDS-PAGE gels. The obtained protein was stored at -20 ºC.

Finally, protein concentration was calculated with a Bicinchoninic Acid (BCA) Protein Assay kit (Novagen), which is a method based on the reduction of divalent copper cations with concentration-dependent chelation of the monovalent cations by bicinchoninic acid, which produces a purple complex with absorbance maximum at 562 nm.[71]

3. Results and Discussion

3.1 Prior protein analysis

When starting the project, the hosting laboratory already had an almost purified protein stored at -20 ºC which was thought to be lyase, obtained after heterologous expression in *E. coli* BL21 using pBAD202 as expression plasmid. To confirm if this protein was lyase, a pure sample (Figure 3.1) obtained with Milli-Q water, was sent to N-terminal sequencing, which provides the first five amino acids of the sample's sequence. This sample was obtained by excision of the appropriate band, which was separated in an SDS-PAGE gel using Milli-Q water to ensure reproducibility of the results.

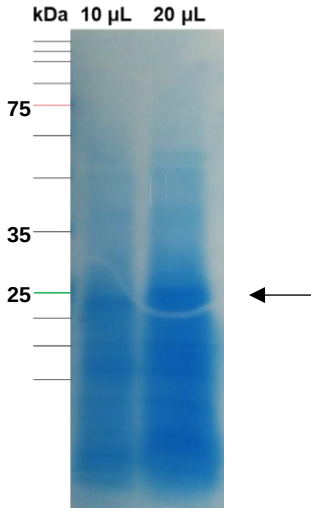

Figure 3.1 SDS-PAGE gel with partially purified unknown protein. Polymerised with 12% acrylamide, the first well has 10 µL of the aforementioned sample, while the second has 20 µL. The black arrow points to the excised band sent for sequencing.

A more intense, noticeable band can be seen in the well containing 20 µL of the applied sample at around 25 kDa (next to the green band of the molecular weight marker). After excision and analysis of the band, the following sequence was obtained: E/I L/K T/C W/A H. From these results, one can verify that there is a clear uncertainty in amino acid determination. This might be related to the amount of protein in the band, which was insufficient for a clear identification, and/or to the impurity of the sample itself, which presented difficulties when trying to excise an isolated band.

Nonetheless, when comparing the obtained sequence with the expected sequence (G W F W A), only one residue, the fourth, provides a match. As such, the partially purified protein was not lyase, but either a methionine-tRNA ligase or an uracil-DNA glycosylase, according to results obtained with the BLAST tool.

3.2 Expression tests

With such a result, the next step was to begin expression tests so as to determine the best growth conditions to verify if it would be feasible to purify lyase with this plasmid and strain combination. These expression tests were accomplished by changing three conditions, time (from 4 until 48 hours after induction), temperature (either stabilised at 37ºC or lowered until 30ºC after induction) and the media for cell growth (LB or TB with and without an inducing compound, which can either be arabinose or IPTG). After 48 hours, the best conditions obtained with SDS-PAGE (data not shown) were tested with a Western Blot (Figure 3.2). This technique allowed to detect which tested condition was the best at overexpressing lyase, given that the concentration of the protein is proportional to the intensity of the band that appears in the membrane.

As can be seen in the membrane, the most noticeable band is located in the seventh well, in the range between 25 and 35 kDa, which lies within lyase's molecular weight of about 30 kDa. Although the image was taken after several non-specific bonds were allowed to form, giving it a high background noise, this band remains as the most intense, which led to a 6-litre growth according to these conditions.

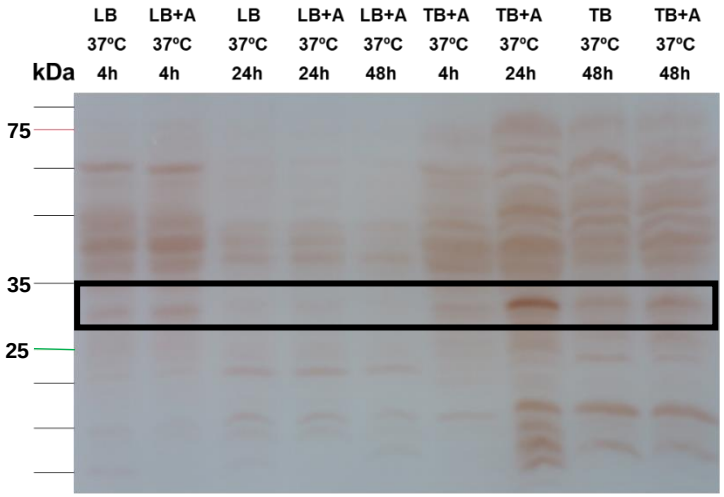

Figure 3.2 Image of the membrane with best conditions obtained from expression tests. Three conditions were simultaneously tested, medium (LB and TB) with and without induction (with arabinose), temperature (37 ºC and 37/30 ºC) and time (4, 24 and 48 hours).

3.3 Growth and purifications

After allowing *E. coli* to grow in TB medium 48 hours after arabinose induction at OD_{600} = 0.7 and 37 ºC, the obtained cell mass had 60 g, which corresponded to a yield of 6 g per litre of liquid medium.

Although a considerable mass of cells was obtained from 6 litres, the subsequent purification process was riddled with difficulties, starting with the chosen chromatograph, which was found to be faulty. This system wouldn't allow to perform a baseline for the resulting chromatograms, due to a fault in the UV flow cell, which resulted in chromatograms with an extremely low signal-to-noise ratio, such as the one obtained for the first affinity chromatography done using a HisTrap column (Figure 3.3).

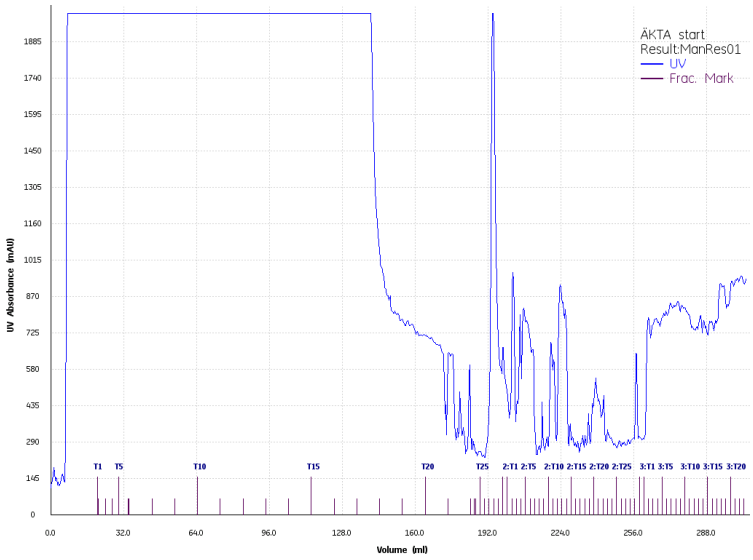

Figure 3.3 Chromatogram of first affinity purification. Chromatogram representative of the entire purification process, with several peaks which might correspond to lyase.

Given that lyase lacks colour, this inhibits one to follow visually its elution process through the columns. Nonetheless, most fractions obtained from the HisTrap column were tested using SDS-PAGE, which proved to be very arduous. Two of these samples were further purified with a Q Sepharose Fast Flow column. By the end of this process, whatever existed in the initial extract was completely lost, either due to lack of extensive testing (sample preparation for SDS-PAGE) or due to extreme dilutions caused by the usage of the aforementioned chromatographic columns.

As a result, an 8-litre growth was done to test the purification process in another Äkta system, an Äktaprime plus. Using the same conditions as previously stated, but with induction at OD_{600} = 0.55, this second growth using pBAD202 yielded a total of 97 g of cell mass, which corresponded to a yield of around 12 g per litre of TB.

The affinity purification using a HisTrap column resulted in a sharp peak (Figure 3.4), which corresponded to eight 2 mL fractions to be further purified.

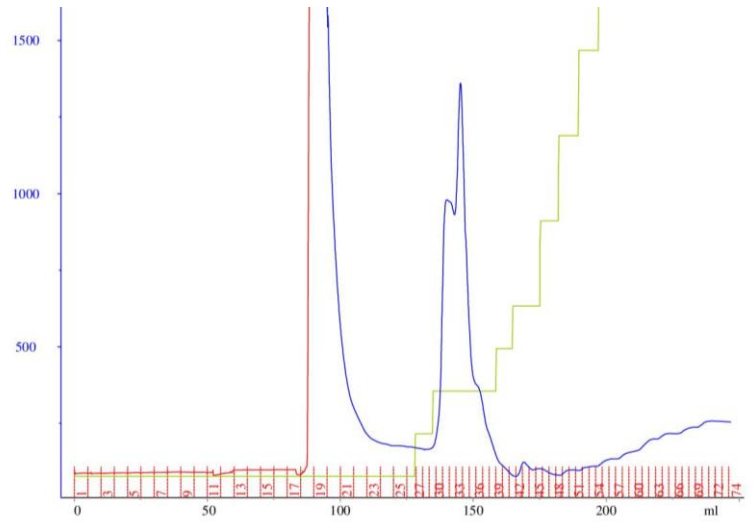

Figure 3.4 Chromatogram of first successful affinity purification. Representative portion of the entire chromatogram, with peak obtained using 10% elution buffer. This implies a low affinity of the protein for the used column.

Before starting a second purification process, the absorbance at 280 nm was measured for the obtained fractions so as to roughly verify if in an acceptable amount of protein was present, and each one had an absorbance higher than 1.5 (calculated considering the applied dilutions to guarantee that linearity between absorbance and the calculated concentration is maintained).[72]

Furthermore, a membrane fraction was obtained by solubilizing the cell pellet with 2 mL of stock Triton-X overnight, followed by an ultracentrifugation for 30 minutes to obtain a supernatant which was then analysed with a Western Blot assay. This fraction showed no protein enrichment around 30 kDa (Figure 3.5), which meant lyase was not being lost in the cell pellet.

In addition to this, the purest fraction (fraction 30, or F30, as identified in Figure 3.5) was chosen to do a UV-Vis spectrum, and it displayed an absorbance maximum at 413 nm, which could indicate the presence of oxidised haem. Given that *E. coli* only produces cytochromes *c* under anaerobic conditions and lyase was produced under aerobic conditions, aeration of the chosen media must have been less than perfect to allow the assembly of a small number of *c*-type cytochromes. Nonetheless, if *c*-type cytochromes are matured within the growing cells, lyase has to be present, even if it occurs in extremely low quantities.

Another attempt to confirm the existence of lyase as an oligomer was to use Dynamic Light Scattering (DLS) with fraction 30, the purest one of the entire process. When considering size distribution by intensity, two peaks were obtained (Figure 3.5, on the left) but, when considering the same characteristic by volume, only one peak appeared (Figure 3.5, on the right). Whilst handling fraction 30 and changing it to a different vessel, dust particles and other debris may have deposited within the sample, thus explaining the existence of two peaks in the first measurement. The second measurement demonstrated that this sample was monodisperse, which is an indicative of the protein in the sample existing as a monomer.

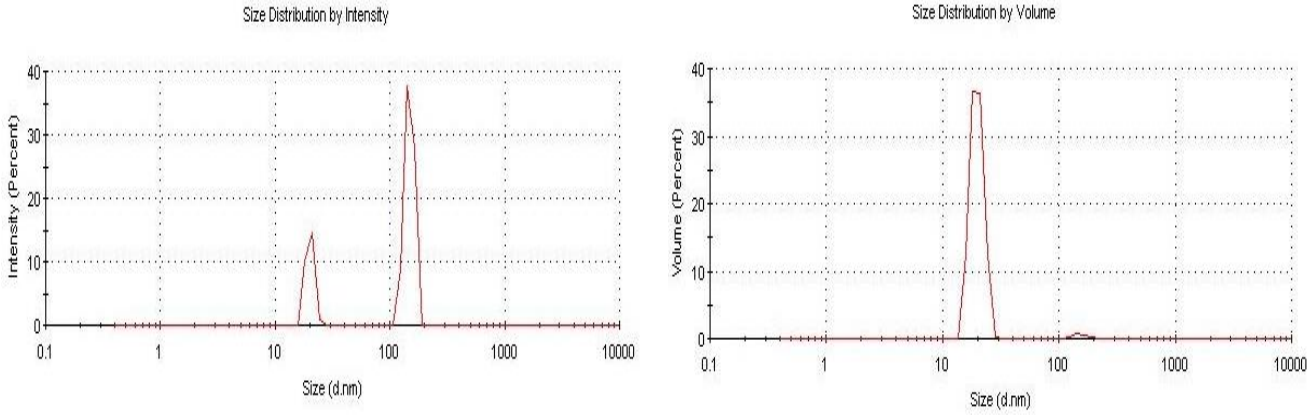

Figure 3.5 DLS results of fraction 30. On the left, size distribution by intensity of fraction 30 had two peaks with 25.6% and 74.4, indicating the possibility of a dimer. On the right, size distribution by volume had one peak, which points towards fraction 30 existing as a monomer with a diameter of almost 20 nm.

A weak band of the desired molecular weight could also be observed in the “Wash” fraction (Figure 3.6) and, to try and maximise protein yield, a second affinity chromatography step was performed, but this time collecting five 2 mL fractions at 50% elution buffer. The resulting SDS-PAGE obtained from this second purification showed a saturated band in two of the five collected fractions close to 30 kDa which, if corresponding to lyase, meant that some protein was being lost

in the washing step of the first HisTrap run. As a result, the final yield might be improved by resorting to an extra affinity purification step.

These two fractions had the same profile in a Western Blot as the samples in Figure 3.6, including an enriched band close to 75 kDa, which could illustrate the existence of a dimer. To test this hypothesis, one of the fractions was treated with a denaturing agent which reduces disulphide bonds, β-mercaptoethanol, followed by heating for 5 minutes at 99 ºC. Although this heavier band disappeared, no other band was enriched as a result, which contradicts the dimer hypothesis (data not shown).

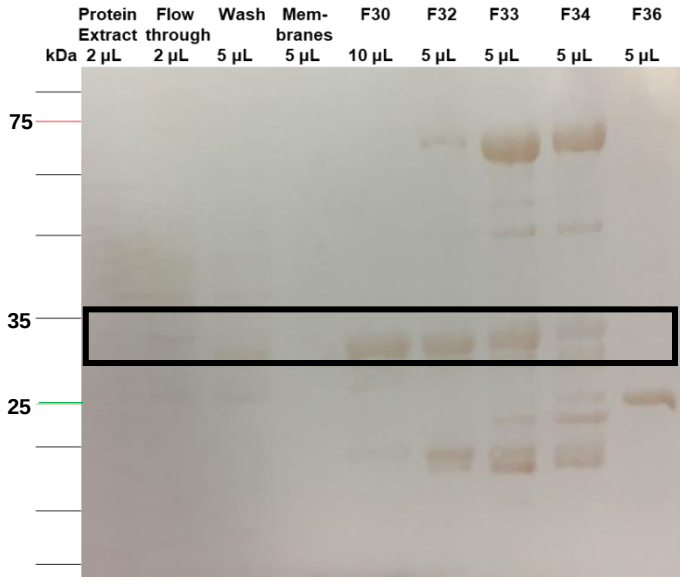

Figure 3.6 Western blot of HisTrap fractions. Representative portion of the entire chromatogram, with peak obtained using 10% elution buffer. This implies a low affinity of the protein for the used column.

The next step was to concentrate all samples with similar gel pattern (compared using SDS-PAGE) using VivaSpin vessels and centrifugation cycles at 4 ºC to fractions volumes of 1 mL. With this technique, two samples (F31 and F33+F34) were concentrated and applied into a Superdex 75 HiLoad 16/60 with a volume of 120 mL. According to the column's elution profile and calibration curve[73], lyase should be eluted between 70 to 80 mL of elution volume.

As expected, given the elution profile of the column, both samples had a peak at 70 mL (data not shown). However, the corresponding fractions still had the higher molecular weight band when analysed with SDS-PAGE. Given that their masses had quite some difference between them, they should have been successfully separated in the last column. As a result, this continued to indicate the possibility of an interaction between molecules or of a dimer.

To further test the existence of an interaction between molecules, similar fractions were concentrated into two final samples (sample 1 and sample 2) and were tested with β-mercaptoethanol (Figure 3.7). As with previous results, the higher molecular weight band disappeared, but no obvious enrichment is noticeable in other fractions present in the gel.

After measuring absorbances at 280 nm, sample 1 had a concentration of 1.2 $mg.mL^{-1}$, whereas sample 2 had more than double, with a concentration of 3.1 $mg.mL^{-1}$. As for fraction 30, its concentration was measured using a BCA assay, and it provided a concentration of 0.7 $mg.mL^{-1}$. Given that samples 1 and 2 had a total volume of 2 mL and fraction 30 had 0.5 mL, the 8-litre

growth using the combination of *E. coli* BL21 and pBAD202 had a final yield of 8.95 mg of protein, with only 0.35 mg accounting for putative pure lyase.

As this is an extremely low amount to use for biochemical characterization, a third growth was needed to both optimise the purification process and obtain more pure protein. However, some assurance that fraction 30 was, in fact, pure lyase was needed before proceeding with further testing.

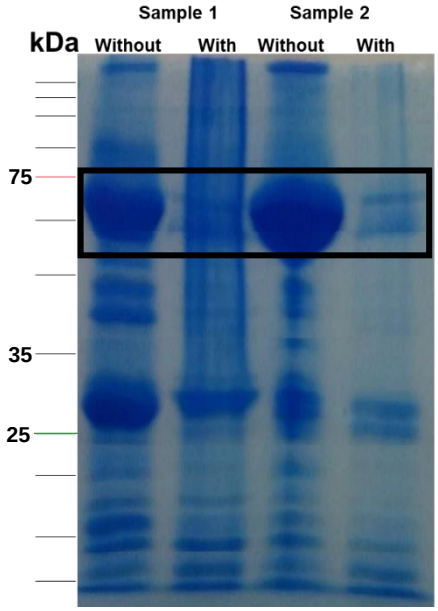

Figure 3.7 SDS-PAGE of concentrated samples resulting from size-exclusion chromatography. "With" – Samples with β-mercaptoethanol; "Without" – Samples without β-mercaptoethanol. Although samples seem more contaminated due to higher background noise, this is due to the use of BlueSafe staining instead of Coomassie Blue. While the latter is more readily available, the former is more sensitive to lower concentrations of protein.

For this purpose, an SDS-PAGE was done using three different volumes of fraction 30 (Figure 3.8, on the left) so as to send the best band to be analysed using mass spectrometry services. However, staining the gel revealed two problems: the sample was more contaminated than what was previously verified (Figure 3.6) and the most noticeable bands were nowhere near the expected mass. After performing a second SDS-PAGE and a new Western Blot, results showed that fraction 30 had, indeed, suffered degradation to a certain extent (Figure 3.8, on the right).

As can be seen after comparing fraction 30 in the Western Blots in Figures 3.6 and 3.8 (the fifth lane in Figure 3.6 and the image on the right in Figure 3.8), the elution profiles are completely different. The only band present in the first figure appears to have been broken into three new bands in the latter figure. Furthermore, the clearest band in the latter figure has suffered degradation and now appears at a position of lower molecular weight in the membrane when compared to the former figure. Taking these results into account, whatever small quantity of pure protein that was obtained in this 8-litre growth was insufficient and, in addition to this, it suffered degradation. Put differently, a new growth and purification process became essential to study lyase.

Simultaneously, plasmid pBTR1 was successfully transformed into *E. coli* BL21 so as to begin expression tests in case this 8-litre growth failed to provide a substantial amount of protein.

However, grown cells transformed with pBTR1 showed no discernible differences between conditions in an SDS-PAGE gel (not shown), with each and every one appearing almost identical. In addition to this, the aforementioned plasmid lacks an affinity tag, preventing detection with more specific assays. Thus, the option of trying to work with pBTR1 was abandoned.

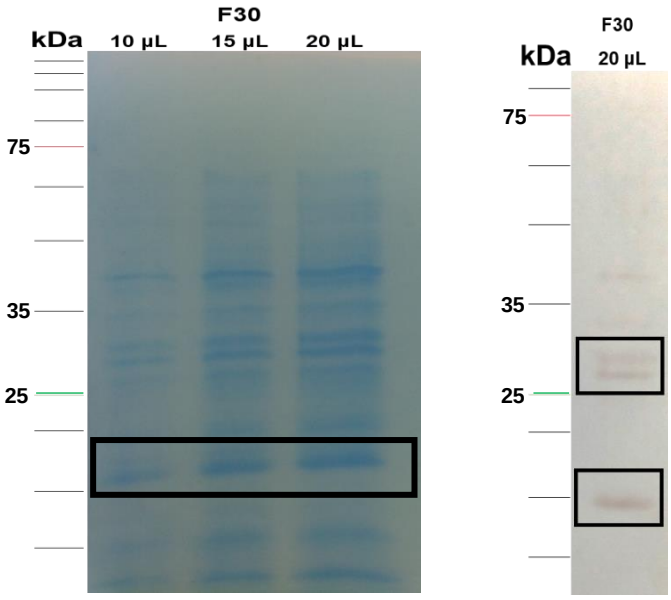

Figure 3.8 Images of SDS-PAGE and Western Blot done on fraction 30. On the left, the SDS-PAG shows that the most noticeable bands are below the 20 kDa molecular marker, which is 10 kDa less than expected, as lyase is around 30 kDa. On the right, the same sample used in the third lane of the SDS-PAGE, 20 µL, was subjected to a Western Blot assay, which shows, once more, that new bands appeared and that the clearest one has an incorrect mass for it to be lyase.

For the third growth, cells in 12 litres of TB media were induced with arabinose at OD_{600} = 0.7, and the same conditions as previously described were maintained for the growth and cell collection process, yielding a cell mass of 115 g. The obtained cell mass was divided and each half was purified separately. For the first half (m = 58 g), in the course of loading the crude protein extract from a Schott bottle into the column, around 40 mL of extract were lost by accident, which ended up being detrimental for the final yield.

Using a freshly stripped and recharged His-trap column, a sharp peak (Figure 3.9) was obtained with 40% elution buffer, which was an improvement when compared to the previous purifications. Besides there being a unique, sharp peak, the sample had higher affinity for the column than before.

When comparing with the peak represented in Figure 3.4, this one is undeniably sharper, although with a lower intensity. From this peak, 12.5 mL were collected and dialysed overnight with agitation in the buffer used for the following anionic purification with a Q Sepharose column.

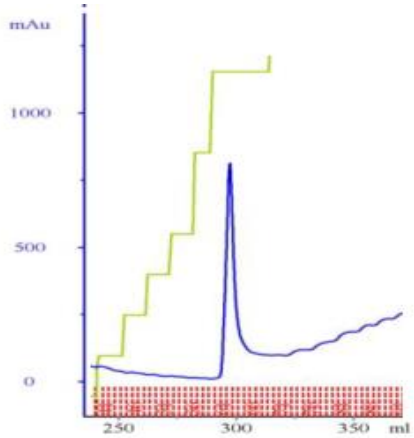

Figure 3.9 Chromatogram of third affinity purification process. Representative portion of the entire chromatogram, with peak obtained using 40% elution buffer. Although not as intense as in the previous purifications, the peak is sharper and demonstrates that the sample has more affinity for the used column.

In the anionic purification, two peaks were obtained (Figure 3.10, left panel), one at 30% (peak 1) and the other at 40% elution buffer (peak 2). After analysis with SDS-PAGE, the majority of the protein with the correct mass was located within peak 2 (Figure 3.10, right panel). As such, only peak 2 would be further purified with a Superdex column. Nonetheless, both samples were concentrated to 200 µL and stored.

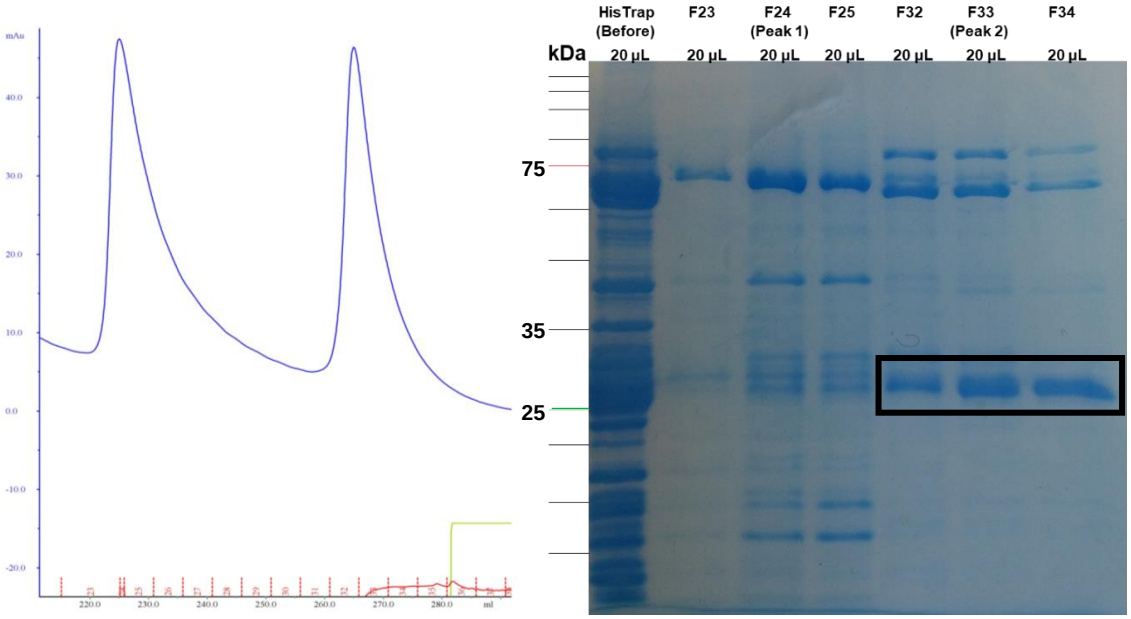

Figure 3.10 Chromatogram and SDS-PAGE of the samples obtained from the anionic purification. On the left, representative portion of the entire anionic purification process, with peak 1 obtained using 30% elution buffer and peak 2 obtained using 40% elution buffer. On the right, SDS-PAGE with representative fractions of the entire purification. The first lane corresponds to the sample that was obtained and concentrated from the affinity purification, and the fractions called “peak” correspond to the most intense portion of the peak.

The anionic column provided relatively clean fractions from the injected affinity chromatography sample, which can be verified when comparing the first lane of Figure 3.10 with all of the others. Moreover, the fractions that formed the sample from peak 2 to be concentrated and injected into the Superdex column (fractions 32 to 35) had bands that were considerably

different in molecular mass, which meant that a size-exclusion chromatography should be able to resolve them rather well.

This time around, a Superdex 75 10/300 was used and, according to column specifications, a protein with lyase's mass should be eluted from around 10 to 15 mL. Once again, two distinct, well-resolved, peaks were obtained in the chromatogram (Figure 3.11, left panel) but, due to their elution volumes, only one was most likely to be lyase. Contrary to what was expected, the SDS-PAGE (Figure 3.11, right panel) showed fractions almost identical to the ones in Figure 3.10, that is, fractions containing rather differently sized proteins. One possible explanation might be the existence of strong interactions between the proteins which compose these bands.

From this last column, five fractions (fractions 5 to 9) totalling 5.5 mL were collected and their concentrations were calculated using a BCA enhanced assay to guarantee that even low concentration values would be possible to interpolate from the standard curve. Furthermore, the purest fraction, fraction 9, which contained a single band with the desired molecular mass, was also analysed in a NanoDrop and using an extinction coefficient adapted from Expasy's supplied value (ε = 44710 $M^{-1}.cm^{-1}$, which corresponds to lyase's oxidised, cystine forming iteration due to the existing oxidising atmosphere and biological pH value) so as to compare all chosen methods for future determinations.

As a matter of fact, all three methods gave completely different results, which were the following: 1.5 µM using a NanoDrop apparatus, 3.6 µM using a BCA assay and 1 µM using Expasy's provided extinction coefficient. The NanoDrop was chosen to quantify the remaining fractions, as it corresponds to the median of the measurements of the three methods.

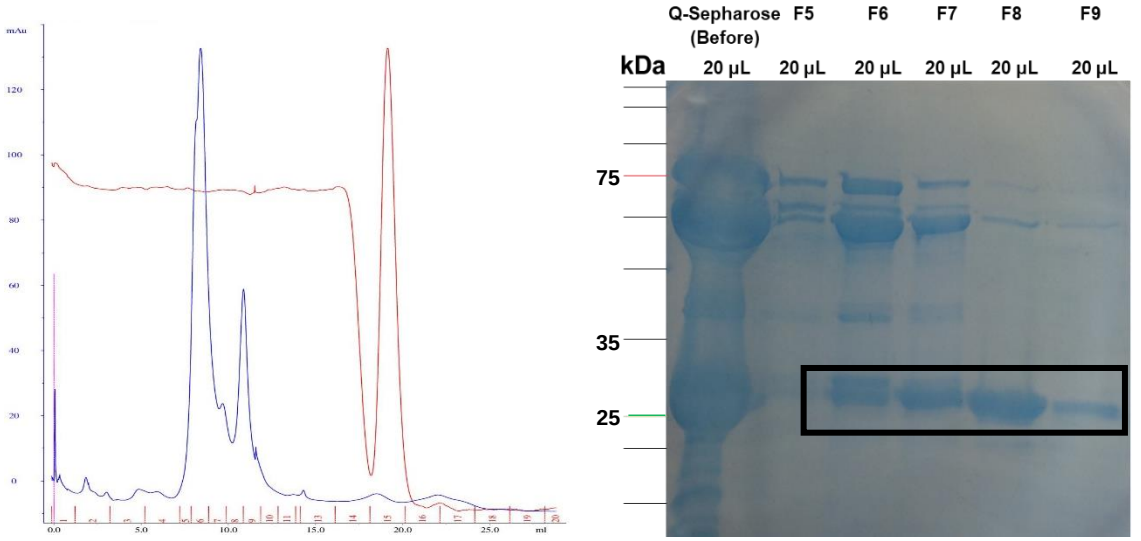

Figure 3.11 Chromatogram and SDS-PAGE of the samples obtained from the size-exclusion purification. On the left, representative portion of the entire anionic purification process, with the smallest peak (in blue) being the most likely to be lyase, according to the column's elution profile. On the right, SDS-PAGE with representative fractions of the entire purification. The first lane corresponds to the sample that was obtained and concentrated from the anionic purification, and the other lanes correspond to the fractions obtained from both peaks.

To sum up, these five fractions yielded a grand total of 1.38 mg of protein, which represents half of the obtained cell mass (m = 58 g), including a 40 mL cell extract loss. Given that one other

half still needed to be purified, an identical procedure would yield a final result of 2.76 mg of protein.

Using the remaining cell mass (m = 57 g), 185 mL of crude protein extract were loaded into the HisTrap column, without loss of volume. From this column, an identical chromatogram to Figure 3.8 was obtained, with the peak appearing at 40% elution buffer as well. Another 12.5 mL were dialysed as before and concentrated before injecting into a Q Sepharose column.

From this anionic purification, two peaks were obtained as before (Figure 3.10, left panel), but with much less intensity, which indicated that a lower quantity of protein was present in the fractions. Nonetheless, all fractions from the peak on the right were collected and concentrated to be injected into the size-exclusion column. This final step occurred exactly as before but, once again, with a peak with much less intensity.

From this third and last column, three fractions (fractions 12 to 14) totalling 3 mL were collected and their concentrations measured using the NanoDrop device. With the purification of the second half of the obtained cell mass, a yield of 0.19 mg of total protein was attained. Given that this yield is more than eight times less than the one obtained from the first half of the cell mass, the origin of this difference was investigated.

Due to the differences in intensity only becoming apparent in the second column's results, HisTrap's flow-through was tested. This flow-through was the solution that remained after concentrating the sample obtained from the affinity chromatography with a VivaSpin vessel. If working properly, no protein with a mass above the filter's cut-off molecular mass (30 kDa) should be present in the flowthrough. However, the exact opposite occurred, with the flow-through having most of the protein that should have been present in the sample to be injected in the anionic column. As such, an even smaller yield than what was expected was obtained. Nonetheless, even if the expected 2.76 mg of protein had been obtained, this would still be an unacceptable result to have from cells grown in 12 litres of media and to be able to use in biochemical characterisation assays.

3.4 First Mass Spectrometry results

Before attempting yet another growth and subsequent purification protocol, the purest fraction obtained from this 12-litre growth, fraction 9, was sent to be tested by in-house mass spectrometry (MS) services. For this, a liquid sample was taken from the same fraction applied in the SDS-PAGE pictured in Figure 3.11 right panel. As such, a peak between 25 and 35 kDa was to be expected, and perhaps a smaller peak between 63 and 75 kDa. Instead, two peaks outside of the aforementioned ranges were obtained, the first with 38.9 kDa and the second, more intense, with 41.5 kDa. No bands of such sizes were detected in the fraction applied in the SDS-PAGE. Nonetheless, the used plasmid, pBAD202, doesn't have lyase's gene directly connected to the tag, having extra amino acids (KGELKLE) and a V5 epitope between them. The sum of these extra components (0.8 kDa for the extra amino acids and 1.7 kDa for the V5 epitope) with lyase and the tag's mass results in a total mass of 33.5 kDa for the recombinant protein, which is still incompatible with the mass of the obtained peak. Even when considering the protein as being purified as a dimer, as previously thought, this would result in a peak around 67 kDa, which was

not detected in the tested sample. In summary, these results show that the desired protein was not obtained in the purification process at all, which indicated the need for a change and optimisation of the overall process.

3.5 Expression tests with multiple *E. coli* strains

One possibility for change was the combination of *E. coli* strain and plasmid used for the overexpression of lyase. For this purpose, several strains were transformed with plasmid pBAD202 so as to begin expression tests. Strains DH5-α, Rosetta 2 (DE3), Top 10, Tuner and XL2 Blue were transformed with a heat shock protocol, while strains JM109, LMG194 and XL1 Blue were transformed with an electrocompetence and electroporation protocol.

Besides these strains, HB101 and RP4182 were acquired for the same purpose, but were previously transformed with a different plasmid, pAV1. One method of forcing these strains to lose this plasmid was to remove the selective pressure needed to maintain it, which was growth in a medium containing ampicillin. As such, several passages alternating between liquid (LB without ampicillin) and solid (LB with agar) media were executed to force this loss of plasmid. A total of 9 cycles proved to be inefficient in the removal of this plasmid, as the colonies still grew in media containing ampicillin (positive control). The next attempt at plasmid curing (i.e., plasmid loss) consisted of 3 passages using the same solid medium as before, but with overnight growth at 43ºC. Still the cells maintained their plasmids, given that pAV1 is derived from a thermostable plasmid, pET-3a. Afterwards, a further 5 cycles of liquid/solid media passages were used, but failed to make the strains lose the plasmid. The last attempt was to use a high-voltage electroporation protocol adapted for plasmid curing, and still this gave no results, as colonies kept on growing within media treated with the antibiotic. Nevertheless, a heat shock transformation was still successful in inserting pBAD202 in both HB101 and RP4182 *E. coli* strains.

With a total of 10 transformed strains, some physical, temporal and material constraints were taken into account before starting expression tests. In fact, each strain would have required at least 7 Erlenmeyer flasks for the entirety of the process, with a total of 70 flasks to be used simultaneously. This would be extremely cumbersome to both move, control and put into incubator shakers. As far as these problems were concerned, one solution was to perform these expression tests in test tubes. To ensure proper aeration and thus, result reproducibility and possibility of comparison with Erlenmeyer flask growths, every tube was tilted 45º in a metal rack.

For these tests it was decided to grow every strain without pBAD202 as a negative control. When comparing SDS-PAGE results of both transformed and non-transformed strains, it was possible to conclude if the observed enriched bands belonged to lyase or to an *E. coli* putative protein. Furthermore, strain BL21 was grown as well to verify if the new chosen condition would be better than the one that had been used for the previous growth attempts. Most expression tests went as expected. Yet, some test tubes had media with a thick biofilm formation, which indicated either lack of aeration and/or media contamination. To solve this issue, it was decided that the best conditions as deduced from the SDS-PAGE results (Figure 3.12) would be re-grown in Erlenmeyer flasks and tested one more time using SDS-PAGE and Western Blot.

For the first round of tests with the test tubes, the best strains were JM109, LMG194 and XL2 Blue (Figure 3.12). As can be observed, every strain has some enrichment in the desired range of molecular mass.

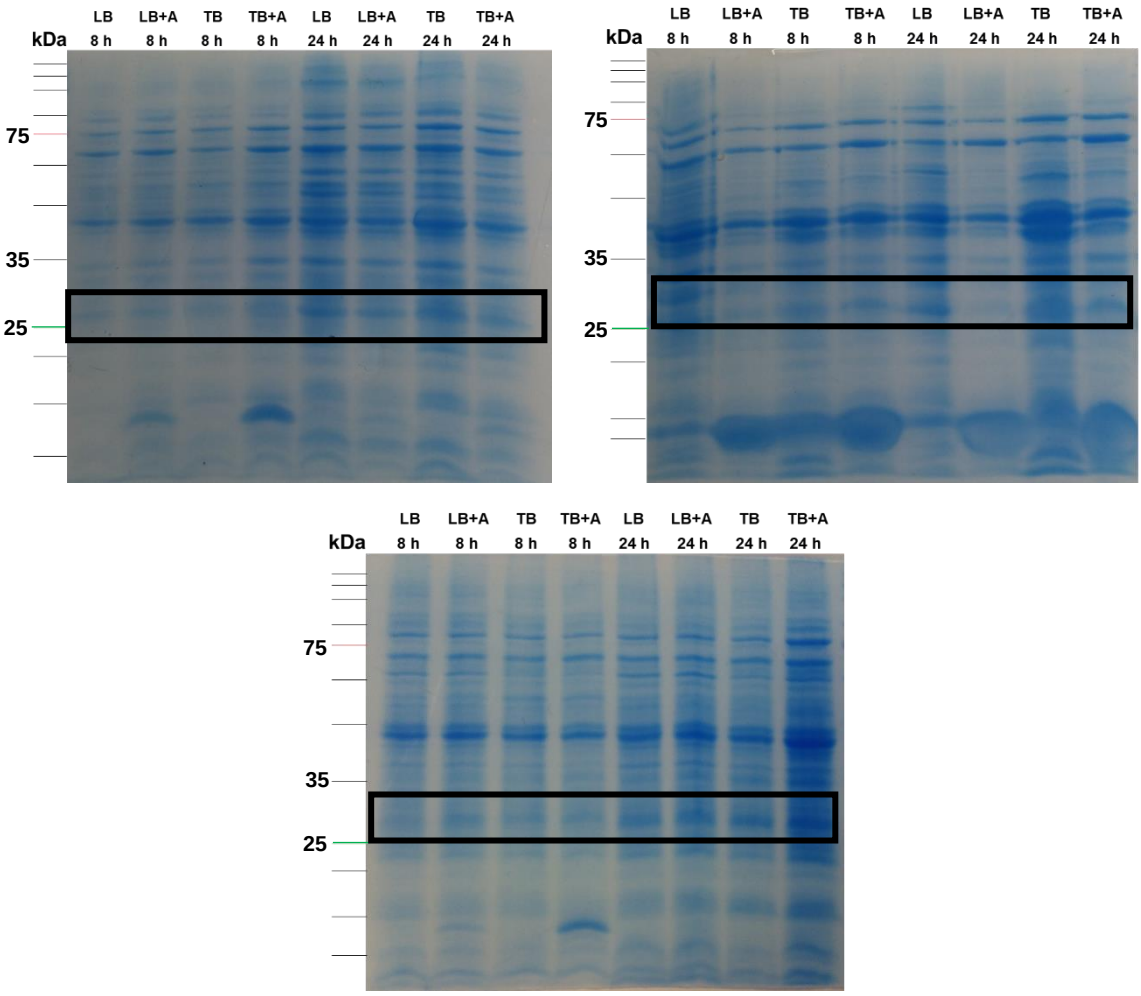

Figure 3.12 SDS-PAGE of the best *E. coli* strains in the expression tests. Results from strain JM109 are on the left, from strain LMG194 on the right and from XL2 Blue on the bottom. The desired band, corresponding to lyase, should appear between the green molecular marker and the black one directly above it, which corresponds to a range between 25 and 35 kDa.

Regarding the other strains, most had no clear bands in the desired area, whereas strains Rosetta 2, Top 10 and Tuner had extremely atypical SDS-PAGE patterns, with almost no bands whatsoever. Given that these expression tests analyse every protein obtained after lysis of *E. coli* cells, gels with high background noise are to be expected, but not gels with almost no bands. Even after repeating the analysis of these strains, still no bands were observed. Given that every strain went through the same lysis process, this indicates that the problem may have occurred during cell growth. Perhaps the contaminations within the test tubes led to protein degradation which, after 48 hours, can immensely affect the protein content outcome in the gel.

After completely excluding all strains but JM109, LMG194 and XL2 Blue, their gels were further analysed, and a longer growth period appeared to be correlated to a higher overexpression of protein. As such, the final samples collected from these strains, which occurred 48 hours after

induction, were also tested with SDS-PAGE to verify if this trend continued or if some protein degradation took place instead.

From these results (images not shown), the most promising ones belonged to *E. coli* XL2 Blue, given that there was clear overexpression close to the 25 kDa molecular marker associated with a relatively low background noise in the gel when compared to both JM109 and LMG194. To conclude these assays and pursue a test growth using one of these three strains, all of them were grown in test tubes without the plasmid and in Erlenmeyer flasks with and without the plasmid to verify if scaling-up maintained the same overexpression and background noise profiles.

After comparing the protein profile in each transformed strain to their stock counterpart, no discernible differences could be seen in the gel. Such a result clearly indicated that no overexpression of lyase occurred in these strains, which lead to believe that the previously observed protein enrichment in the gels were due to a constitutive *E. coli* protein with mass similar to lyase. To final cast aside the hypothesis of using pBAD202 with a different *E. coli* strain, and given the arrival of the antibody, every strain (both with and without the plasmid) was tested with a Western Blot protocol. As expected from the previous results, no overexpression of the recombinant protein was achieved (Figure 3.13), and bands that were clearly located within the area of interest in the SDS-PAGE completely disappeared in the Western Blot, confirming that it was indeed due to a putative protein of the chosen host organism.

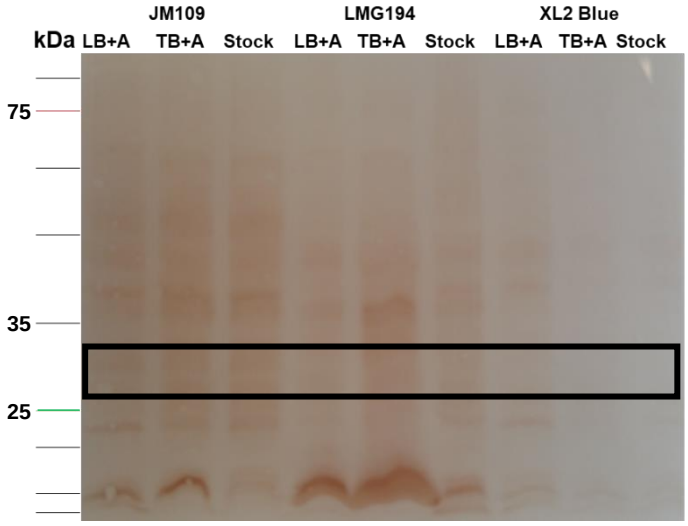

Figure 3.13 Western Blot of the best *E. coli* strains in the expression tests. All samples represented correspond to the 48-hour collection period. With a more sensitive assay which uses an antibody that detects the histidine tail in the recombinant protein, there is no clear band in the area of interest. Every band appeared after the membrane was allowed to develop for several minutes, which indicates non-specific binding of the antibody.

3.6 Molecular Biology solutions

The next option was to take one step back, which meant trying to insert the gene that codes for lyase, CYC3, into different plasmids. Three different approaches were attempted whilst dwelling in the realm of Molecular Biology.

One approach was to use pBTR1, given that it already contains CYC3, and use site-directed mutagenesis to insert a Strep-Tag II, as this plasmid lacks any kind of tag. Firstly, primers named Forward_pBTR1 and Reverse_pBTR1, with 66 bp and a melting temperature of 80 ºC were used to attempt site-directed mutagenesis in pBTR1. The aim of these primers was to introduce both a thrombin cleavage site (TCS) and a Strep-Tag II in lyase's C-terminal. After following NZYMutagenesis kit directions, the theoretically mutated recombinant DNA was transformed into *E. coli* XL2 competent cells. When several attempts to obtain colonies from the transformed cells failed, a new enzyme was tried. GRiSP's Xpert High Fidelity DNA Polymerase was then used to perform the mutagenesis protocol, but it also failed to produce recombinant colonies. As such, NZYTech's enzyme, Proof Polymerase, was used once again, but this time with a temperature gradient ranging from 68 ºC to 72 ºC for the annealing step of the protocol.

Once again, no results were obtained from the transformation. After verifying the site-directed mutagenesis protocol, it was decided that the aforementioned primers were too long for the PCR to work properly. To solve this issue, new primers were designed, Forward_Strep and Reverse_Strep, with 47 bp and a melting temperature of at least 75 ºC (exact value was not supplied by the company which synthesizes the oligonucleotides). These new primers would only insert the last amino acid residue required for the TCS and Strep-Tag II in lyase. If sequencing results would show these mutations as being successful, the entire TCS could be added later, which would permit to perform expression tests in the meantime.

With the improved primers, both enzymes were used to insert the mutations into the same pBTR1 source. While Proof Polymerase failed to yield transformed colonies, Xpert High Fidelity DNA Polymerase was successful, and two transformed colonies were obtained in a Petri Dish.

Both colonies were grown as pre-inocula in LB medium supplied with ampicillin to extract pDNA for sequencing purposes. Colony 1's sample was sent, as it had the best results after plasmid extraction was performed, with high concentration and low presence of impurities – 455.9 ng/µL, A_{260}/A_{280} = 1.9 and A_{260}/A_{230} = 2.1. Although only the mutation at the end of the gene needed to be checked, sequencing was done using a pair of stock primers supplied by the company, M13-FP and M13-RP. These primers were chosen as they align with pUC18, which is pBTR1's parent plasmid. Sequencing results showed that the plasmid was still in its original state, with not even one point mutation in the desired area of the gene.

This lack of results led to a second molecular biology attempt, which was to amplify and insert the human gene, HCCS, into pBAD/Thio-TOPO plasmid. For this, two 5-mL blood samples were collected within K_2EDTA sterile plastic test tubes (Vacutainers) for mitochondrial DNA extraction. Simultaneous nuclear DNA (nDNA) and mDNA extraction of one sample according to the altered protocol yielded samples with a concentration of 50.3 ng/µL and 15.8 ng/µL,

respectively, and absorbance ratio of A_{260}/A_{280} = 1.69 and A_{260}/A_{280} = 1.29, respectively. Both parameters were according to the expected results from the original protocol.

Afterwards, amplification of the HCCS gene within the obtained mDNA was attempted using the designed primers Forward_HCCS and Reverse_HCCS and a combination of two polymerases, *Taq* and Proof. Furthermore, two different annealing temperatures, 68 ºC and 70 Cº, as well as a different polymerase combination, *Taq* and Xpert High Fidelity, were used as variations of the first PCR protocol to achieve this amplification. None of these variants was conductive to a positive result.

The second blood sample was used together with a different extraction kit, NZYTech's gDNA Isolation Kit. Given that the provided version was intended for tissue samples instead of blood samples, some adaptations were made, resulting in a sample with 49.2 ng/µL and A_{260}/A_{280} = 1.5. This DNA sample was used to perform HCCS amplification with only Taq polymerase, but with an annealing temperature gradient, using 62 ºC and 67 ºC. Once again, no results were obtained.

This approach would have required some extra time and more samples, as well as the appropriate reagents, instead of alternatives and alterations to several protocols. Although both extraction methods yielded some amount of DNA, they had low purities, as verified by the absorbance ratios, which was insufficient for the required PCR protocols to function properly.

3.7 pETBlue-2 Plasmid

The third and final molecular biology attempt was to insert CYC3 into a new plasmid, pETBlue-2. At first, the primers created to insert CYC3 into pBAD202, Forward_pBAD202 and Reverse_pBAD202, were used to obtain a sufficient amount of insert for the new plasmid. However, this only led to its amplification without the proper ends required to ligate it with the new plasmid, rendering the insert unusable. For this purpose, primers Forward_pETBlue_2 and Reverse_pETBlue_2 were created.

For the first try using pETBlue-2's kit, the entire protocol was followed, including the creation of both a positive and a negative control for the transformation. Firstly, CYC3 was amplified using the newly created primers and Proof Polymerase. Afterwards, the obtained insert and the other alternatives (the supplied control insert acting as a positive control and sterile water as a negative control) suffered end-conversion and ligation as indicated before being transformed into the stock competent cells, *E. coli* NovaBlue Singles.

Seven Petri dishes were obtained from this process, three with different ratios of insert DNA to vector DNA (1:5, 1:8 and 1:10), one with the control insert (positive control), one with sterile water (negative control) and two with essential components missing (either the insert or the insert and the used ligase enzyme), needed to test the transformation efficacy.

From these seven Petri dishes, not one had a transformed colony. Given that every instruction was followed and that not even the positive control (supplied by the company) succeeded in obtaining colonies, it was hypothesised that one of the kit's components must have been defective.

With a newly opened kit, the entire process was repeated, but this time only five Petri dishes were made: one for the positive control, one for the negative control, and three with different volumes of applied insert. For these last three, either 25 µL or 50 µL of the original transformation media or 10 µL of centrifuged cells were inoculated into the solid medium. As expected, the positive control originated several colonies, whereas the negative control originated none. As for the plates inoculated with the vector ligated with CYC3, only the last one (inoculated after concentrating the cells via centrifugation) yielded colonies.

Afterwards, the five obtained colonies went through colony screening, and two tested positive for the desired insert (Figure 3.14, on the left). However, these colonies weren't completely isolated, and each one was re-inoculated into fresh solid media to re-test if positive.

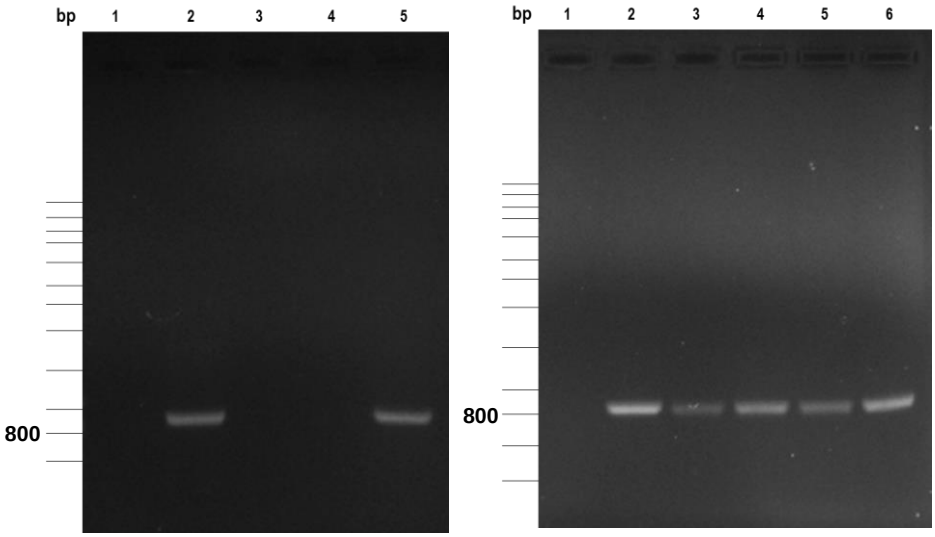

Figure 3.14 Colony screenings of colonies transformed with pETBlue-2 ligated with CYC3. On the left is the colony screening of the original five colonies obtained from the transformation of pETBlue-2 with CYC3 into *E. coli* NovaBlue Singles competent cells, with colonies 2 and 5 displaying positive results. On the right is the result of the colony screening of the re-inoculated colonies 2 and 5, with positive results existing in colonies 2 through 6.

From the re-inoculated colonies, completely isolated colonies were obtained, and these were once again tested via colony screening, with all but one possessing CYC3 (Figure 3.14, on the right). Colonies 2 and 5 were then chosen to extract plasmid in order to verify if the gene's sequence was correctly inserted into pETBlue-2.

Both colonies yielded the same sequencing results, which lacked around 50 bp from each end of CYC3. Moreover, there was one mutation in position 442 which, after sequence alignment with CYC3's deposited sequence, proved to be a silent mutation. The affected nucleotide was changed from a guanine to a cytosine, still originating a codon for the translation into serine to occur. Nonetheless, a complete sequence was required to ensure that CYC3 was intact and correctly inserted into the plasmid, and this was ordered with a different set of primers which would allow alignment with pETBlue-2 either upstream (T7) or downstream (pETBlue-RP) of CYC3's location.

While the sequencing process wasn't complete, given the positive partial results, the recombinant vector was successfully transformed into *E. coli* BL21 (DE3) and JM109 so as to begin expression tests. These tests were done as before, with variations in time (4, 8, 24, 30 and

48 hours after induction), liquid media (either LB or TB) and the absence of presence of an inducing compound (which would be IPTG for pETBlue-2). After 48 hours, some conditions from the middle time points were tested with both SDS-PAGE and Western Blot (Figure 3.15).

Regarding both tested strains, BL21 (DE3) had a clear, intense, band closer to 35 KDa on the second to last lane (Figure 3.15, on the left, labelled “TB+I” under 30h), which appeared immediately after the bands were revealed, whilst JM109 showed several intense bands much later after the bands were revealed. Given that more specific interactions appear first, the best result seemed to be the one belonging to BL21 (DE3), which was to allow cells to grow 30 hours in TB medium supplemented with IPTG.

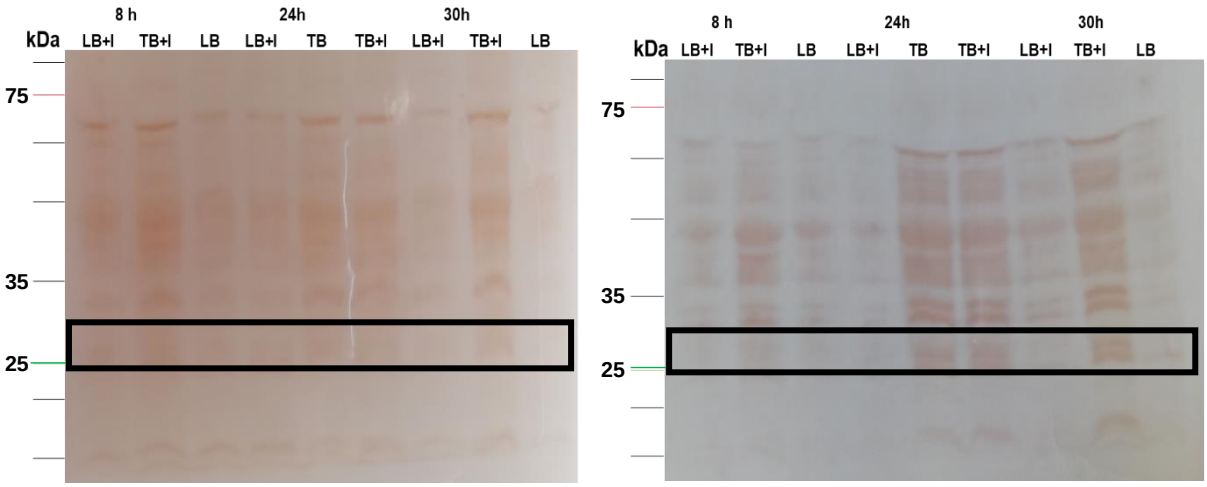

Figure 3.15 Image of the Western Blot membrane with pETBlue-2’s expression tests. Three conditions were simultaneously tested, medium (LB and TB), presence of an inducing compound (IPTG) and time (4,8, 24, 30 and 48 hours). On the left are conditions collected from *E. coli* BL21 (DE3) and on the right are conditions collected from *E. coli* JM109.

The information taken from these membranes led to a 6-litre growth using the previously mentioned conditions with a final yield of 46 g of cell mass. After obtaining a protein extract from this cell mass, a purification process began once more. Starting with an affinity chromatography, a small peak arose in the chromatogram with 20% elution buffer (Figure 3.16, on the left). However, after using SDS-PAGE to observe the obtained fractions, it is noticeable that most of the desired band had remained in the flowthrough (Figure 3.16, on the right). Before taking every relevant fraction and applying it into a second column, a second affinity chromatography (figure not shown) was done, and again a single peak appeared, but with only 10% elution buffer.

After overnight dialysis, around 10 mL from fractions 31 to 34 were injected into a Q Sepharose column, and this anionic process led to a less-than-perfect, fronting peak, obtained with 45% elution buffer, which caused two samples to be concentrated in separate: fractions from the peak’s “front” and the fractions which originated the peak itself. As for the fractions obtained from the second affinity chromatography, they were also injected into this anionic column, and resulted in exactly the same as in the first trial: a defective peak obtained with 45% elution buffer. So far, repetition of the columns with the original flow-through fraction led to the same result as with the raw protein extract, which meant that a significant amount of protein would have been lost otherwise.

To ensure that no protein was being lost in the process, both samples, that is, the peak's front and the peak itself, were injected into a Superdex 75 column which, given lyase's molecular weight, should give rise to a peak at around 12 mL of elution volume. As expected, this peak appeared, but it was preceded by a much more intense peak. For the first time since the beginning of this process, pure fractions, with only one band in the entire lane, were achieved (Figure 3.17).

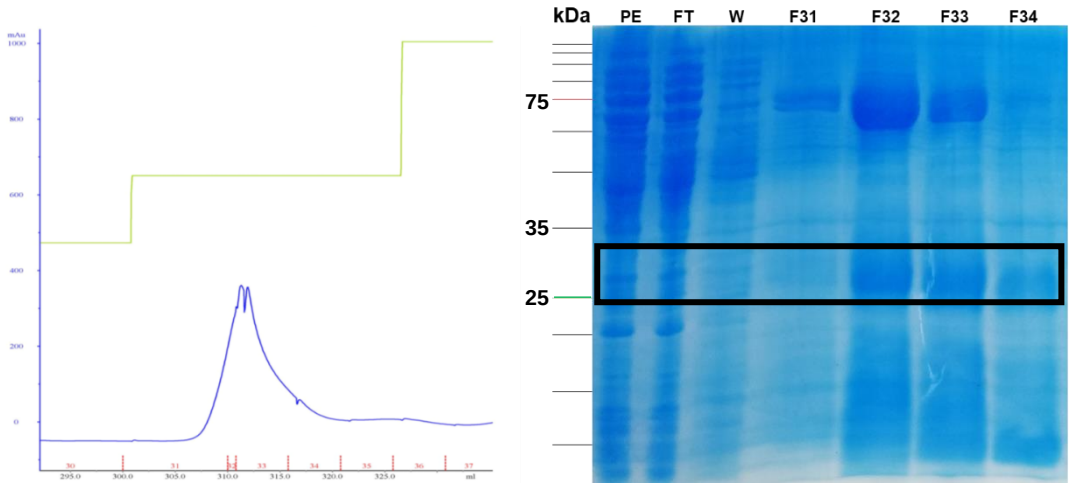

Figure 3.16 Chromatogram and SDS-PAGE of the samples obtained from the affinity purification. On the left, representative portion of the entire affinity purification process, with a peak obtained using 20% elution buffer. On the right, SDS-PAGE with representative fractions of the entire purification. The first lane corresponds to the raw protein extract, while the second corresponds to the flowthrough – an almost identical elution profile is extremely evident.

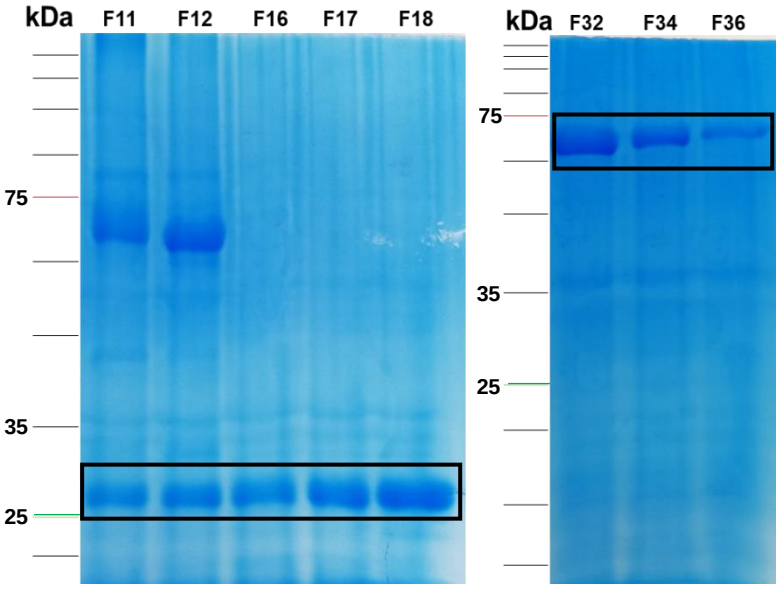

Figure 3.17 SDS-PAGE images representative of the entire purification process. On the left are represented the fractions obtained from the size-exclusion purification step, with F11-12 belonging to a more intense peak with a higher molecular weight than lyase and F16-18 belonging to the peak with a molecular weight closer to lyase. On the right are fractions belonging to the second anionic chromatography's fronting peak, which have a molecular weight close to 75 kDa (red marker).

From the size-exclusion chromatography, fractions were obtained with a protein of molecular mass close to lyase's molecular mass (Figure 3.17, on the left, lanes F16-F18), whereas from the

second anionic chromatography (from the original flowthrough) proteins closer to 75 kDa were purified (Figure 3.17, on the right). This 6-litre growth yielded pure protein fractions with a total of 2.145 mg.

3.8 Mass Spectrometry and N-terminal results

Finally, with pure protein fractions, it would be possible to unequivocally identify the protein which had been purified in every attempt. After concentrating fractions F16-18 (Figure 3.17, on the left), two samples were sent for further analysis, with one being analysed by MS services by MALDI-TOF and the other being sequenced by an N-terminal protocol.

According to MS results, the purified protein was a peptidyl-prolyl cis-trans isomerase. This result was further confirmed by the N-terminal sequencing of the first five amino acid residues, which fully supported the identification provided by MS. Whereas the first sample that was sent for sequencing via N-terminal had some uncertainty in amino acid residue identification (in section 3.1), this sample had no such issues. The provided sequence for the first five residues was K V A K D, which, after searching in *E. coli* genome databases, lead to the protein FKBP-type peptidyl-prolyl cis-trans isomerase, encoded by the gene slyD, with around 21 kDa and 196 amino acid residues.[74]

4. Conclusion

The main aim of this dissertation project was to study the molecular mechanisms of cytochrome *c* biogenesis in eukaryotes through structural and functional characterisation of lyase. To achieve this objective, it was necessary to overexpress a recombinant version of the protein, followed by successful purification.

However, as demonstrated in the previous chapter, none of these tasks proved effective. In spite of several attempts, the hardest task proved to be right at the start of the entire process, with no plasmid and *E. coli* strain combination displaying an unequivocal overexpression of protein. This would prove to negatively impact the following steps, as purification yielded extremely low amounts of protein. On the other hand, this shows that the enzyme must have a high catalytic activity, since it is capable of producing cytochromes in substantial amounts when co-expressed in the same plasmid.

Furthermore, MS and N-terminal sequencing techniques confirmed that the purification strategy used to obtain this protein was not successful, and the final product was but an isomerase from *E. coli*, FKBP-type peptidyl-prolyl cis-trans isomerase, and not the recombinant lyase. This protein has almost 9 kDa less than lyase, and a difference of 73 amino acid residues (196 instead of lyase's 269). Nonetheless, it appeared in SDS-PAGE gels and in Western Blot membranes in the molecular marker range of 25 to 35 kDa, where lyase should have been. In addition to this, it was purified using an affinity chromatography column that targeted the polyhistidine tag attached to lyase.

With the production of a recombinant form of lyase, using a polyhistidine affinity tag to ensure specificity for the purification steps that would follow, how could FKBP-type peptidyl-prolyl cis-trans isomerase have been purified in its place? This protein's secondary/alternative name, histidine-rich protein, may help provide an explanation. Besides its main purpose as isomerase, this protein has a metal-binding domain, located near the C-terminal. It contains 13 histidine residues, and binds with high affinity to up to seven nickel ions in a non-cooperative mode. As such, it could have been co-purified or purified in place of lyase, as it potentially binds with more affinity to the nickel column than lyase. If this is true, it should elute with a higher percentage of imidazole than lyase.

Ultimately, these setbacks led to the impossibility of purifying and fully characterising lyase. However, characterisation of lyase will lead to a deeper knowledge regarding System III and, finally, about cytochrome *c* maturation in eukaryotes. The importance of this topic makes it a necessity to overcome these difficulties.

As such, for the future, three main points could be changed to try and accomplish this goal: the recombinant gene, protein overexpression and, lastly, the purification process.

Firstly, the used construct of CCHL had a polyhistidine tail located in its C-terminal. One option would be to remove this tail and replace it with a Strep-tag II into plasmid pBTR1, given that it already contains the correct sequence encoding for CCHL. Another approach would be to insert the gene into a different vector which would already encode for this tag.

The latter alternative is compatible with the second point that could be subjected to change, which is protein overexpression. To ensure a greater protein yield in the end of the process, different *E. coli* and plasmid combinations in expression tests should prominently display a lyase band in SDS-PAGE gels and Western Blot membranes. If the new CCHL construct containing a Strep-tag would be obtained, it could participate in these new testing phases, as this tag is more specific than the previously used His-tag. A more specific tag, in combination with a clear overexpression of lyase would be a necessary improvement to the used protocol.

Finally, these changes would also enhance the following purification protocol, which is one of the main steps required to obtain pure lyase. With the fusion of a Strep-tag II with lyase, the first affinity chromatography would be much more specific than the one with a HisTrap column. Therefore, the resulting fractions would have less contaminants and a higher enrichment in lyase, thus requiring less chromatographic steps. By improving specificity and using less columns, the resulting purification process will have less steps, which leads to an overall reduction in protein waste and an improvement in the final yield.

Purification of lyase is not a straightforward approach, with opportunities for improvement and adaptation. Nonetheless, the physiological importance of such a protein makes this line of research extremely important, and crucial for a deeper understanding of cytochrome *c* maturation in eukaryotes. If this protein is purified and fully characterised, the possibility of cytochrome-related drug and therapeutic strategies would become viable, thus improving the lives of individuals affected with such ailments, like microphthalmia with linear skin defects syndrome[32,41], cancer and neurodegenerative disorders (these last two provoked by deregulation of cytochrome *c*-mediated apoptosis)[5,32,38]. This research would result at the very least in an improved knowledge of the process of cytochrome maturation, which would become clearer and from which the scientific community could benefit. In a favourable scenario, cytochrome-maturation related conditions could be cured and human lives would be improved.

5. References

1. Munn, C. A. Mac. Researches on Myohaematin and the Histohaematins. *Philos. Trans. R. Soc. London* 267–298 (1886). doi:10.1098/rstl.1886.0007

2. Reedy, C. J. & Gibney, B. R. Heme Protein Assemblies. *Chem. Rev.* **104,** 617–649 (2004).

3. Kranz, R., Lill, R., Goldman, B., Bonnard, G. & Merchant, S. Molecular mechanisms of cytochrome c biogenesis: Three distinct systems. *Mol. Microbiol.* **29,** 383–396 (1998).

4. Kranz, R. G., Richard-Fogal, C., Taylor, J.-S. & Frawley, E. R. Cytochrome c Biogenesis: Mechanisms for Covalent Modifications and Trafficking of Heme and for Heme-Iron Redox Control. *Microbiol. Mol. Biol. Rev.* **73,** 510–528 (2009).

5. Allen, J. W. A. Cytochrome c biogenesis in mitochondria - Systems III and V. *FEBS J.* **278,** 4198–4216 (2011).

6. Koch, H.-G. & Schneider, D. Folding , Assembly , and Stability of Transmembrane Cytochromes. (2007). doi:10.2174/187231307779814020

7. Puustinen, A. & Wikstrom, M. The heme groups of cytochrome o from Escherichia coli. **88,** 6122–6126 (1991).

8. Verner, Z. *et al.* Chapter Three - Malleable Mitochondrion of Trypanosoma brucei. in (ed. Jeon, K. W.) **315,** 73–151 (Academic Press, 2015).

9. Reedy, C. J., Elvekrog, M. M. & Gibney, B. R. Development of a heme protein structure–electrochemical function database. *Nucleic Acids Res.* **36,** D307–D313 (2008).

10. Layer, G., Reichelt, J., Jahn, D. & Heinz, D. W. Structure and function of enzymes in heme biosynthesis. *Protein Sci.* **19,** 1137–1161 (2010).

11. Lobo, S. A. L. *et al.* Staphylococcus aureus haem biosynthesis : characterisation of the enzymes involved in final steps of the pathway. **97,** 472–487 (2015).

12. PROSITE. C-type cytochrome superfamily profiles. (2018). Available at: https://prosite.expasy.org/PDOC00169.

13. ExPASy. C-type cytochrome superfamily profiles. (2004). Available at: https://prosite.expasy.org/PDOC00169. (Accessed: 12th June 2018)

14. Fonseca, B. M. *et al.* Mind the gap: cytochrome interactions reveal electron pathways across the periplasm of *Shewanella oneidensis* MR-1. *Biochem. J.* **449,** 101–108 (2013).

15. Ambler, R. P. Sequence variability in bacterial cytochromes c. *BBA - Bioenerg.* **1058,** 42–47 (1991).

16. Pettigrew, G. W. & Moore, G. R. *Cytochromes c: Biological Aspects*. (Springer-Verlag, 1987).

17. Saraiva, I. H., Newman, D. K. & Louro, R. O. Functional Characterization of the FoxE Iron Oxidoreductase from the Photoferrotroph Rhodobacter ferrooxidans SW2 *. **287,** 25541–25548 (2012).

18. Lehninger, A. L., Nelson, D. L., Cox, M. M. & others. *Principles of Biochemistry*. (W. H. Freeman and Company, 2008). doi:10.1017/CBO9780511794193

19. Lambeth, D. O. The Purification of Cytochrome C from Baker's Yeast - an undergraduate laboratory experiment in biochemistry. *J. Chem. Educ.* 270–272 (1979).

20. GeneCards - Human Gene Database. CYCS Gene (Protein Coding). Available at: https://www.genecards.org/cgi-bin/carddisp.pl?gene=CYCS#proteins.

21. UniProtKB - P99999 (CYC_HUMAN). Available at: http://www.uniprot.org/uniprot/P99999. (Accessed: 10th June 2018)

22. Allen, J. W. A., Daltrop, O., Stevens, J. M. & Ferguson, S. J. C-type cytochromes: diverse structures and biogenesis systems pose evolutionary problems. *Philos. Trans. R. Soc. B Biol. Sci.* **358,** 255–266 (2003).

23. Hartshorne, R. S., Richardson, D. J. & Simon, J. Multiple haem lyase genes indicate substrate specificity in cytochrome c biogenesis. *Biochem. Soc. Trans.* **34,** 146–149 (2006).

24. Kern, M. & Simon, J. Production of recombinant multiheme cytochromes c in Wolinella succinogenes. *Methods Enzymol.* **486,** 429–446 (2010).

25. Mavridou, D. A. I., Ferguson, S. J. & Stevens, J. M. Cytochrome c assembly. *IUBMB Life* **65,** 209–216 (2013).

26. Verissimo, A. F. & Daldal, F. Cytochrome c biogenesis system I: An intricate process catalyzed by a maturase supercomplex? *Biochim. Biophys. Acta - Bioenerg.* **1837,** 989–998 (2014).

27. Nicholson, D. W. & Neupert, W. 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 (1989).

28. Babbitt, S. E. *et al.* Mechanisms of mitochondrial holocytochrome c synthase and the key roles played by cysteines and histidine of the heme attachment site, Cys-XX-Cys-His. *J. Biol. Chem.* **289,** 28795–28807 (2014).

29. Verissimo, A. F., Sanders, J., Daldal, F. & Sanders, C. Engineering a prokaryotic apocytochrome c as an efficient substrate for Saccharomyces cerevisiae cytochrome c heme lyase. *Biochem. Biophys. Res. Commun.* **424,** 130–135 (2012).

30. Francisco, B. S., Bretsnyder, E. C. & Kranz, R. G. Human mitochondrial holocytochrome c synthase ' s heme binding , maturation determinants , and complex formation with cytochrome c. **110,** (2012).

31. Babbitt, S. E., Hsu, J., Mendez, D. L. & Kranz, R. G. Biosynthesis of Single Thioether c-Type Cytochromes Provides Insight into Mechanisms Intrinsic to Holocytochrome c Synthase (HCCS). *Biochemistry* **56,** 3337–3346 (2017).

32. Babbitt, S. E., Sutherland, M. C., Francisco, B. S., Mendez, D. L. & Kranz, R. G. Mitochondrial cytochrome c biogenesis: No longer an enigma. *Trends Biochem. Sci.* **40,** 446–455 (2015).

33. Dumont, M. E., Cardillo, T. S., Hayes, M. K. & Sherman, F. Role of cytochrome c heme lyase in mitochondrial import and accumulation of cytochrome c in Saccharomyces cerevisiae. *Mol. Cell. Biol.* **11,** 5487–5496 (1991).

34. Dumont, M. E., Ernst, J. F., Hampsey, D. M. & Sherman, F. Identification and sequence of the gene encoding cytochrome c heme lyase in the yeast Saccharomyces cerevisiae. *EMBO J.* **6,** 235–241 (1987).

35. GeneCards - Human Gene Database. HCCS Gene (Protein Coding). Available at: https://www.genecards.org/cgi-bin/carddisp.pl?gene=HCCS&keywords=cyc3.

36. IUBMB Enzyme Nomenclature. EC 4.4.1.17. (1990). Available at: http://www.sbcs.qmul.ac.uk/iubmb/enzyme/EC4/4/1/17.html.

37. ExPASy. ProtParam tool. Available at: https://web.expasy.org/protparam/.

38. Jiang, X. & Wang, X. Cytochrome C-Mediated Apoptosis. *Annu. Rev. Biochem.* **73,** 87–106 (2004).

39. Thong, A. & Tsoukanova, V. Cytochrome-c-assisted escape of cardiolipin from a model mitochondrial membrane. *Biochim. Biophys. Acta - Biomembr.* **1860,** 475–480 (2018).

40. Orrenius, S. & Zhivotovsky, B. Cardiolipin oxidation sets cytochrome c free. *Nat. Chem. Biol.* **1,** 188 (2005).

41. Genetics Home Reference. Microphthalmia with linear skin defects syndrome. (2018).

42. World Health Organization. Cancer. (2018). Available at: https://www.who.int/news-room/fact-sheets/detail/cancer.

43. Olteanu, A. *et al.* Stability and apoptotic activity of recombinant human cytochrome c. *Biochem. Biophys. Res. Commun.* **312,** 733–740 (2003).

44. Bornhorst, B. J. A. & Falke, J. J. Purification of Proteins Using Polyhistidine Affinity Tags. *Methods Enzymol.* **326,** 245–254 (2000).

45. Mullis, K. *et al.* Specific enzymatic amplification of DNA in vitro: The polymerase chain reaction. *Cold Spring Harbor Symposia on Quantitative Biology* **51,** 263–273 (1986).

46. Invitrogen. pBAD Directional TOPO ® Expression Kit: User Manual. (2009).

47. Shuman, S. Novel approach to molecular cloning and polynucleotide synthesis using vaccinia DNA topoisomerase. *J. Biol. Chem.* **269,** 32678–32684 (1994).

48. Thermo Fisher. Tm Calculator. Available at: https://www.thermofisher.com/pt/en/home/brands/thermo-scientific/molecular-biology/molecular-biology-learning-center/molecular-biology-resource-library/thermo-scientific-web-tools/tm-calculator.html.

49. Novagen. Perfectly Blunt Cloning Kits: User Protocol TB183. 1–17 (2011).

50. Promega. dsDNA: pmol to µg. (2018). Available at: https://www.promega.com/a/apps/biomath/. (Accessed: 20th April 2018)

51. Schmidt, T. G. M. & Skerra, A. The Strep-tag system for one-step purification and high-affinity detection or capturing of proteins. *Nat. Protoc.* **2,** 1528–1535 (2007).

52. NZYTech. NZYMutagenesis kit: User Manual MB012. 0–2 (2018).

53. Fisher Scientific. Tubes - blood collection. 153 (2003).

54. Lahiri, D. K. & Numberger, J. I. A rapid non-enzymatic method for the preparation of HMW DNA from blood for RFLP studies. *Nucleic Acids Res.* **19,** 5444 (1991).

55. Ahmad, S., Ghosh, A. & Nair, D. L. Simultaneous extraction of nuclear and mitochondrial DNA from human blood. **1,** 429–432 (2007).

56. Invitrogen. pBAD/TOPO® ThioFusion™ Expression Kit: User Manual. 45 (2009).

57. Casali, N. Escherichia coli Host Strains. *E. coli Plasmid Vectors* **235,** 27–48

58. Stratagene. Competent Cells - How do you choose the right competent cell? 1–22 (2008).

59. Invitrogen. One Shot ® TOP10 Competent Cells User Guide. (2013).

60. Novagen. Tuner™(DE3)pLacI Competent Cells. (2018). Available at: http://www.merckmillipore.com/PT/en/product/TunerDE3pLacI-Competent-Cells-Novagen,EMD_BIO-70625. (Accessed: 7th July 2018)

61. ATCC. Escherichia coli (Migula) Castellani and Chalmers (ATCC® 47090™). (2016). Available at: https://www.lgcstandards-atcc.org/en/Products/All/47090.aspx#characteristics. (Accessed: 7th July 2018)

62. Sohail, A., Lieb, M., Dar, M. & Bhagwat, A. K. A Gene Required for Very Short Patch Repair in Escherichia coli Is Adjacent to the DNA Cytosine Methylase Gene. *J. Bacteriol.* **172,** 4214–4221 (1990).

63. Panja, S., Aich, P., Jana, B. & Basu, T. How does plasmid DNA penetrate cell membranes in artificial transformation process of Escherichia coli? *Mol. Membr. Biol.* **25,** 411–422 (2008).

64. BERTANI, G. Studies on lysogenesis. I. The mode of phage liberation by lysogenic

Escherichia coli. *J. Bacteriol.* **62,** 293–300 (1951).

65. Tartof, K. & Hobbs, C. Improved media for growing plasmid and cosmid clones. *Focus* **9,** 10 (1987).

66. Kelley, K. D., Olive, L. Q., Hadziselimovic, A. & Sanders, C. R. Look and See if it is Time to Induce Protein Expression in Escherichia coli Cultures. *Biochemistry* **49,** 5405–5407 (2010).

67. NZYTech. NZY Bacterial Cell Lysis Buffer: User Manual MB17801. 1–2 (2014).

68. NZYTech. BlueSafe: User Manual MB15201. 1–2 (2018).

69. Mahmood, T. & Yang, P.-C. Western Blot: Technique, Theory, and Trouble Shooting. *N. Am. J. Med. Sci.* **4,** 429–434 (2012).

70. Science Gateway. Tools: Centrifuge Rotor Speed Calculator. Available at: http://www.sciencegateway.org/tools/rotor.htm.

71. Smith, P. K. *et al.* Measurement of protein using bicinchoninic acid. *Anal. Biochem.* (1985). doi:10.1016/0003-2697(85)90442-7

72. Mehta, A. Ultraviolet-Visible (UV-Vis) Spectroscopy – Limitations and Deviations of Beer-Lambert Law. *Analytical Chemistry Notes* (2012). Available at: https://pharmaxchange.info/2012/05/ultraviolet-visible-uv-vis-spectroscopy-–-limitations-and-deviations-of-beer-lambert-law/. (Accessed: 24th November 2018)

73. GE Healthcare Life Sciences. User Manual: HiLoad Superdex 30, 75, 200 prep grade. 1–6 (2011).

74. UniProtKB. SLYD_ECOLI. (2018). Available at: https://www.uniprot.org/uniprot/P0A9K9.