# SITE-DIRECTED SPIN-LABELLING OF PROTEINS FOR EPR SPECTROSCOPY:
## APPLICATION TO PROTEIN COMPLEXES AND DEVELOPMENT OF NEW METHODS FOR CYSTEINE RICH PROTEINS

**Stacey Bell**

**A Thesis Submitted for the Degree of PhD**
**at the**
**University of St Andrews**

**2016**

# Site-Directed Spin-Labelling of Proteins for EPR Spectroscopy: Application to Protein Complexes and Development of New Methods for Cysteine Rich Proteins.

Stacey Bell

University of St Andrews

This thesis is submitted in partial fulfilment for the degree of Doctor of Philosophy at the

University of St Andrews

September 2015

***Time for a little acknowledgement…***

First and foremost I would like to extend my gratitude to my Ph.D. supervisor Janet, for providing such stimulating problems to work on. For always motivating, asking insightful questions and for her invaluable advice throughout, I am forever grateful. During the course of my PhD, I have been very fortunate to attend several conferences, both in the U.K. and abroad, which would not have been possible without her support. Not only did these conferences expand my knowledge in the field of EPR and inspire my research, but also introduced me to the wider EPR community, with whom I made many great memories.

Throughout my PhD studies, I have been very fortunate to work with so many wonderful people, and form true friendships both in Edinburgh where it all began, and in St Andrews. To the 'Chamberlain Street Massive'; Jenny, Lizzie, Marc and Ravi, who have been friends since day one, and my lab mates Andri, Claire and Lee, I wish nothing but the best for each of you. We're all trying hard to find our own place in the world, but I take comfort in knowing we'll always be friends.

A special thank you must go to Ashley. We have shared so many memories over the years and I really can't put into words how comforting it's been going on the PhD journey with your best friend.

Lastly, but always most importantly, I want to thank Mum & Dad, without whom none of this would be possible. Thank you for your endless love and support. Achieving anything in life is impossible without encouragement, and I have been blessed with the best support network anyone could hope for. From holding my hand on my first days at school, and throughout the PhD process, you have been with me every step of the way. Finishing your PhD really can be all consuming, but when necessary, I'm grateful to you for always being able to bring my mind back from science to the humanities.

To my mum Rose; my biggest supporter in my studies, and in life. Thank you for always having my corner, thank you for being my rock, and most importantly thank you for being my friend.

To my dad Bobby, my hero; you've taught me the value of hard work and determination in achieving your goals, but most importantly you've taught me the importance in always being kind, and the simple beauty of laughter.

Words cannot express how grateful I am to you. I love you both.

For Erin and For Orla.

*Always*

**Abstract**

The work described in this thesis is an experimental study into the application of Electron Paramagnetic Resonance (EPR) Spectroscopy for the study of biological systems.

Using a variety of methods of site-directed spin-labelling (SDSL), this thesis aims to explore long range structure in an assortment of recombinant and native proteins, and complexes thereof.

The work described in this thesis covers all aspects of the work, from experimental design, molecular biology and cloning, protein expression and purification, as well as functional characterisation, and finally EPR distance measurements, data analysis and interpretation. Challenges and pitfalls will also be addressed.

Chapters **1** and **2** introduce EPR spectroscopy, and its application in the study of long range structure in biological systems. The experimental techniques employed throughout this thesis are also introduced.

Chapter **3** details an investigation into the complement C3b:factor H complex. This chapter addresses the challenges associated with the SDSL of cysteine rich proteins. Utilising hidden cysteine residues in native proteins for spin-labelling purposes will also be addressed.

Chapter **4** looks at the interactions of the human myosin regulatory light chain (RLC) with cardiac myosin binding protein C (cMyBP-C). Optimisation of expression and purification protocols will be the focus, as well as addressing issues with protein solubility and spin labelling efficiencies.

Chapter **5** explores the development of new methods of SDSL, for the specific labelling of cysteine rich proteins. The ability of *Escherichia coli* to read through the amber stop codon will be exploited for the incorporation of unnatural amino acids for labelling purposes, and novel spin labels, specific for labelling cysteine pairs tested in several model systems. Furthermore, native paramagnetic centres in recombinant proteins will be explored as potential labelling sites.

**Declaration**

I, Stacey Bell, hereby certify that this thesis, which is approximately 55,000 words in length, has been written by me, and that it is the record of work carried out by me, or principally by myself in collaboration with others as acknowledged, and that it has not been submitted in any previous application for a higher degree.

I was admitted as a research student in October 2011 and as a candidate for the degree of Doctor of Philosophy in September 2015, the higher study for which this is a record was carried out in the University of St Andrews between 2013 and 2015.

Date …… signature of candidate ……

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of Doctor of Philosophy in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date …… signature of supervisor ……

**Copyright Declaration**

**Abbreviations**

| | |
|---|---|
| $A_{280}$ | Absorbance at 280nm |
| $A_{412}$ | Absorbance at 412nm |
| AA | Amino Acid |
| ADP | Adenosine Di-Phosphate |
| AEC | Anion Exchange Chromatography |
| aHUS | atypical Haemolytic Uraemic Syndrome |
| Amp | Ampicillin |
| AP | Alternative Pathway |
| AMD | Age Related Macular Degeneration |
| AOX | Alcohol Oxidase |
| ATP | Adenosine Tri-Phosphate |
| BMG(Y) | Buffered Minimal Glycerol (supplemented with yeast) |
| BMM(Y) | Buffered Minimal MeOH (supplemented with yeast) |
| bp | base pairs |
| BSA | Bovine Serum Albumin |
| CaM | Calmodulin |
| C0 | (Myosin Binding Protein C) Domain 0 |
| CCP | Complement Control Protein |
| CEC | Cation Exchange Chromatography |
| Cm | Chloramphenicol |
| CV | Column Volume(s) |
| CVFBb | Cobra Venom Factor Bb convertase |
| CW | Continuous Wave |
| DAB | Di-amino Benzidine |

| | |
|---|---|
| $ddH_2O$ | double distilled Water |
| ddNTP | any deoxyribonucleotide triphosphate |
| DEER | Double Electron Electron Resonance |
| DMSO | Di-methyl Sulfoxide |
| DNA | Deoxyribonucleic Acid |
| DTNB | 5,5'-dithiobis-(2-nitrobenzoic acid) |
| DTT | Dithiothreitol |
| εACA | ε-Amino Caproic Acid |
| *E. coli* | *Escherichia coli* |
| EDTA | Ethylenediaminetetraacetic Acid |
| ELC | Essential light chain of myosin |
| ENDOR | Electron Nuclear Double Resonance |
| EPR | Electron Paramagnetic resonance |
| EtBr | Ethidium Bromide |
| EtOH | Ethanol |
| fB | Factor B |
| fH | Factor H |
| fI | Factor I |
| FPLC | Fast Protein Liquid Chromatography |
| FRET | Forster Resonance Energy Transfer |
| G | Gauss |
| GAG | Glycosaminoglycan |
| HCM | Hypertrophic Cardiomyopathy |
| HMM | Heavy Mero-Myosin |
| HPLC | High Performance Liquid Chromatography |

| | |
|---|---|
| hr | hour(s) |
| HRP | Horse Radish Peroxidase |
| HYSCORE | Hyperfine Sublevel Correlation |
| IEC | Ion Exchange Chromatography |
| IMAC | Immobilised Metal Affinity Chromatography |
| IPA | Isopropanol |
| IPTG | Isopropyl β-D-1-thiogalactopyranoside |
| J | Joules |
| Kan | Kanamycin |
| kDa | kiloDalton |
| LB | Luria Bertoni (Broth/Agar) |
| LMM | Light Mero-Myosin |
| LSLB | Low Salt Luria Bertoni (Broth/Agar) |
| MAC | Membrane Attack Complex |
| MCS | Multi-Cloning Site |
| MeOH | Methanol |
| min | minute(s) |
| MMM | Multi-scale Modelling of Macromolecular Systems |
| mRNA | messenger Ribonucleic Acid |
| MS | Mass Spectrometry |
| MTSSL | (1-Oxyl-2,2,5,5-tetramethyl-Δ3-pyrroline-3-methyl) Methanethiosulfonate |
| MW | Molecular Weight |
| MWCO | Molecular Weight Cut Off |
| MyBP-C | Myosin Binding Protein C |
| NaAc | Sodium Acetate |

| | |
|---|---|
| NMR | Nuclear Magnetic Resonance |
| NTB | 2-nitro-5-thiobenzoicacid |
| $OD_{600}$ | Optical density at 600nm |
| o/n | overnight |
| pAcPhe | p-acetyl phenylalanine |
| PBS | Phosphate Buffered Saline |
| PCR | Polymerase Chain Reaction |
| PDB | Protein Data Bank |
| $PEG_{3350}$ | Polythylene Glycol 3350 |
| Pi | Inorganic Phosphate |
| pI | Isoelectric Point |
| pK | propargyl lysine |
| PMSF | Phenylmethylsulfonyl Fluoride |
| *P. pastoris* | *Pichia pastoris* |
| PVDF | Polyvinylidene di-fluoride |
| RLC | Regulatory Light Chain of Myosin |
| RNA | Ribonucleic Acid |
| rpm | revolution per minute |
| r.t. | Room Temperature |
| s | second(s) |
| SAXS | Small Angle X-ray Scattering |
| SDM | Site-Directed Mutagenesis |
| SDSL | Site-Directed Spin Labelling |
| SDS-PAGE | Sodium Dodecyl sulphate Polyacrylamide Electrophoresis |
| SEC | Size Exclusion Chromatography |

| | |
|---|---|
| s/n | supernatant |
| SOB | Super Optimal Broth |
| SOC | Super Optimal Broth (with catabolite repression) |
| SPPS | Solid Phase Peptide Synthesis |
| T | Tesla |
| TBE | Tris-Borate EDTA |
| TBTA | Tris(benzyltriazolylmethyl)amine |
| TCEP | Tris(2-carboxyethyl) Phosphine |
| TE | Thioester |
| TED | Thioester Domain |
| TEMPO | (2,2,6,6-TetraMethylPiperidin-1-yl)Oxyl – a small molecule containing a nitroxide radical |
| Tet | Tetracycline |
| TEV | Tobacco Etch Virus |
| $T_{melt}$ | primer melting temperature |
| TOAC | 2,2,6,6-Tetramethylpiperidine-N-oxyl-4-amino-4-carboxylic Acid |
| tRNA | transfer Ribonucleic Acid |
| UAA | Unnatural Amino Acid |
| UV-Vis | Ultraviolet visible |
| wt | wild type |
| XRC | X-Ray Crystallography |
| YNB | Yeast Nitrogen Base |
| YPD | Yeast-Peptone Dextrose Media (broth/agar) |
| YPDS | Yeast-Peptone-Dextrose Media supplemented with 1M Sorbitol (broth/agar) |
| Zeo | Zeocin™ |

## Amino Acids

| | |
|---|---|
| Alanine (Ala) | **A** |
| Arginine (Arg) | **R** |
| Asparagine (Asn) | **N** |
| Aspartic Acid (Asp) | **D** |
| Cysteine (Cys) | **C** |
| Glutamine (Gln) | **Q** |
| Glutamic Acid (Glu) | **E** |
| Glycine (Gly) | **G** |
| Histidine (His) | **H** |
| Isoleucine (Ile) | **I** |
| Leucine (Leu) | **L** |
| Lysine (Lys) | **K** |
| Methionine (Met) | **M** |
| Phenylalanine (Phe) | **F** |
| Proline (Pro) | **P** |
| Propargyl Lysine | **pK** |
| Serine (Ser) | **S** |
| Threonine (Thr) | **T** |
| Tryptophan (Trp) | **W** |
| Tyrosine (Tyr) | **T** |
| Valine (Val) | **V** |

## List of Figures

**Figure 1.1:** The Electronic Zeeman Splitting Effect.
**Figure 1.2:** Near Isotropic 3 line Spectrum of 4-amino-TEMPO.
**Figure 1.3:** Spectral Simulations of a nitroxide radical (such as TEMPO), at X-band.
**Figure 1.4:** The Principle of the Hahn Spin Echo.
**Figure 1.5:** Schematic representation of the 3-pulse DEER sequence.
**Figure 1.6:** Schematic representation of the 4-Pulse DEER Experiment.
**Figure 1.7:** Chemical structure of the amino acid cysteine.
**Figure 1.8:** Structural composition of Pyrrolinoxyls, and SDSL of Cysteine Residues with MTSSL.
**Figure 1.9:** Chemical Structure of the unnatural amino acid p-acetyl-phenylalanine (pAcPhe).
**Figure 1.10:** Chemical structure of a cysteine-specific trityl radical.
**Figure 1.11:** Chemical structure of the UAA containing the 2,2,5,5-tetramethyl-pyrrolin-1-oxyl moiety.
**Figure 2.1:** The pPICZαB vector.
**Figure 2.2:** Ellman's Assay Reaction.
**Figure 2.3:** L-cysteine calibration curves for determining free cysteine concentration.
**Figure 2.4:** Schematic Representation of the Fluid-Phase Cofactor Assay.
**Figure 3.1:** 'Tick-Over' Mechanism of the AP of Complement Activation.
**Figure 3.2:** Conversion of Complement C3 to its activated form C3b.
**Figure 3.3:** Generalised Reaction Mechanism for the Reaction of a Nucleophile with the TE of C3.
**Figure 3.4:** Linker lengths and residues in Complement fH.
**Figure 3.5:** NMR derived three- dimensional cartoon representation of CCP 16 of fH in solution.
**Figure 3.6:** Schematic representation of a 'bent back' fH, showing distinct interaction sites for C3b.
**Figure 3.7:** Crystal Structures of fH bound to C3b, shown as cartoons, with accompanying schematics.
**Figure 3.8:** Superposition of the X-Ray Crystal structure of fH1-4 bound to C3b, with fH19-20 bound to the TED of C3b.
**Figure 3.9:** Model for the interaction of full length fH with C3b.
**Figure 3.10:** Schematic representation of fH bound to Complement C3 – as shown in Figure 3.9.
**Figure 3.11:** MMM Simulations of the DEER Time Traces (I) and Distance Distributions (II) for fH/C3b complexes.
**Figure 3.12:** Confirmation of the presence of a cysteine residue in the fH1-4 construct, K247C.
**Figure 3.13:** SDS-PAGE Analysis of small scale test expression of the fH1-4 K247C mutant.
**Figure 3.14:** Confirmation of the presence of a hexa-Histidine tag onto fH1-4 K247C.
**Figure 3.15:** SDS-PAGE Analysis of fH1-4 K247C Expression Tests.
**Figure 3.16:** SDS-PAGE and Western Blotting Analysis of Expression from a single fH1-4 K247C clone.
**Figure 3.17:** SDS-PAGE Analysis of the Elution of fH1-4 K247C from the Ni-NTA Superflow IMAC Cartridge.
**Figure 3.18:** Anion Exchange chromatography of fH1-4 K247C.
**Figure 3.19:** Size Exclusion Chromatography of the fH1-4 K247C Fragment.
**Figure 3.20:** Confirmation of the presence of a cysteine residue in the fH19-20 construct, G1107C.
**Figure 3.21:** SDS-PAGE Analysis of fH19-20 G1107C Expression Tests.
**Figure 3.22:** Example Fermentation Log from the Bioflo4500 Vessel.
**Figure 3.23:** Cation Exchange Chromatography of fH19-20 G1107C from 10 Litre Fermentation.
**Figure 3.24:** Reversible Oxidation of Cysteine to Cystine, visualised using Chemdraw ®.
**Figure 3.25:** Mechanism for reduction of a disulphide with DTT, visualised using Chemdraw®.
**Figure 3.26:** Mechanism for reduction of a disulphide bond with TCEP, visualised using Chemdraw ®.
**Figure 3.27:** SDS-PAGE Analysis of the reduction of fH19-20 G1107C with varying concentrations of DTT and TCEP.
**Figure 3.28:** Cation Exchange Chromatography of fH19-20 G1107C from Large Scale Flask Expression (SP Sepharose).

**Figure 3.29:** Cation Exchange Chromatography of fH19-20 G1107C from Large Scale Flask Expression (Source 15 S).
**Figure 3.30:** Size Exclusion Chromatography of the fH19-20 G1107C Fragment.
**Figure 3.31:** X-band CW EPR of fH1-4 K247C and fH19-20 G1107C constructs.
**Figure 3.32:** X-Band CW EPR Spectrum of the fH1-4 K247C construct following freezing.
**Figure 3.33:** CW EPR spectra and SDS-PAGE analysis of fH1-4K247C (A&D), fH19-20 G1107C (B&E) and fH19-20 R1210C (C&F).
**Figure 3.34:** Anion Exchange Chromatography of Plasma C3 following $PEG_{3350}$ Precipitation.
**Figure 3.35:** Anion Exchange Chromatography (Mono Q) of pooled C3 Fractions Following the QSFF Catchment Step.
**Figure 3.36:** Anion Exchange Chromatography (Mono Q) of Digested Plasma Purified C3, converted to C3b.
**Figure 3.37:** SDS-PAGE analysis of C3 Purification Troubleshooting.
**Figure 3.38:** Peptide Fingerprinting of the C3 α and β chains, as well as ceruloplasmin, following in-gel digest.
**Figure 3.39:** Structure of Plasma Protein Ceruloplasmin showing the copper centres (orange spheres).
**Figure 3.40:** Anion Exchange Chromatography of C3 pool, following QSFF, and before treatment with the nucleophile methylamine.
**Figure 3.41:** Anion Exchange Chromatography of Methylamine treated purified C3 and its conversion to C3(N).
**Figure 3.42:** X-band CW EPR of spin-labelled C3b and C3(N).
**Figure 3.43:** Fluid-Phase Cofactor Assay.
**Figure 3.44:** Raw data from the EPR experiments taken for the spin labelled fH fragments in complex with either labelled or unlabelled C3(N).
**Figure 4.1:** Schematic Representation of Actin and Myosin Positioned in the Sarcomere.
**Figure 4.2:** Schematic Representation of the Thick Myosin Filament Structure.
**Figure 4.3:** Schematic Representation of the Actin-Myosin Interaction in Striated Muscle.
**Figure 4.4:** Schematic representation of cMyBP-C positioning in the C zone of the sarcomere, in relation to the thick and thin filaments.
**Figure 4.5:** Schematic representation of cMyBP-C showing features unique to the cardiac isoform, including the 3D NMR Structure of domain C0.
**Figure 4.6:** Cartoon depicting the current understanding of cMyBP-C Function.
**Figure 4.7:** Model for the arrangement of the N' terminus of cMyBP-C around the myosin S1-S2 junction and the light chains.
**Figure 4.8:** SDS-PAGE Analysis Following Small-Scale Test Expression (100mL) of the C0 and RLC/MiniHMM Constructs.
**Figure 4.9:** Elution profile and corresponding SDS-PAGE analysis for the cMyBP-C Domain C0 Following IMAC.
**Figure 4.10:** Mass Spectrum of cMyBP-C, domain C0, performed at SIRCAMS, using FT–ICR mass spectrometry.
**Figure 4.11:** Schematic Representations of the Myosin Heavy and Light chains (A), zoomed in showing the MiniHMM Fragment, with bound RLC (B).
**Figure 4.12:** Crystal structure of human cardiac beta-myosin II S2Δ.
**Figure 4.13:** Schematic representation of the Leucine Zipper motif.
**Figure 4.14:** Sequence of the original MiniHMM construct aligned with the modified MiniHMM.
**Figure 4:15:** SDS-PAGE Analysis Following Large Scale Expression and Purification of WT RLC/MiniHMM Using IMAC.
**Figure 4.16:** Size Exclusion Chromatography elution profile (A) and SDS-PAGE analysis (B) for wt RLC/MiniHMM.
**Figure 4.17:** Sequence alignment of the Chicken RLC with the Human RLC.

**Figure 4.18:** Structures illustrating the positioning of the RLC in the S2 junction, in the vicinity of the S1 heads.
**Figure 4.19:** MMM Simulations of the DEER Time Traces (I) and Distance Distributions (II) shown alongside PyMOL cartoons illustrating the site for cysteine mutagenesis.
**Figure 4.20:** Sequencing chromatograms, comparing the wt sequence to the sequence following SDM, showing successful substitution for cysteine in all cases.
**Figure 4.21:** SDS-PAGE analysis following purification of each of the Cysteine RLC/MiniHMM Mutants following IMAC.
**Figure 4.22:** SDS-PAGE analysis of the E88C RLC/MiniHMM Mutant, showing purification using IMAC and elution from the Ni-NTA following Histag Cleavage.
**Figure 4.23:** SDS-PAGE analysis of the TEV-treated K30C mutant following passage over the IMAC column.
**Figure 4.24:** Elution profile for the K30C RLC/MiniHMM mutant following SEC and resulting SDS-PAGE analysis.
**Figure 4.25:** CW EPR spectrum for the E88C mutant.
**Figure 4.26:** (A) SDS-PAGE analysis and (B) CW EPR spectrum for the E88C mutant. of the elution of the new E88C sample following IMAC.
**Figure 4.27:** Raw DEER data obtained from the RLC E88C mutant**.**
**Figure 4.28:** Schematic representation of the effect phosphorylation of the RLC has on the RLC/MiniHMM complex**.**
**Figure: 4.29:** Schematic representations of the RLC/MiniHMM complex, showing how SDSL could be used to monitor dissociation of the S2 coiled coil upon phosphorylation of the RLC, both in the absence and presence of C0.
**Figure 5.1:** Schematic representation of mRNA processing at the ribosome, and resulting polypeptide synthesis.
**Figure 5.2:** X-Ray Crystal Structure of Sperm Whale Myoglobin.
**Figure 5.3:** Chemical Structures of (A) Lysine and (B) the UAA propargyl Lysine (pK), visualised using ChemDraw ®.
**Figure 5.4:** Confirmation of mutagenesis of S4C and S118C in sperm whale myoglobin.
**Figure 5.5:** Expression and purification of the Sperm Whale Myoglobin, S4CS118C mutant.
**Figure 5.6:** SDS-PAGE analysis following purification of the S4C Mutant following IMAC (A), shown alongside the elution profile (B) and resulting SDS-PAGE analysis following SEC (C).
**Figure 5.7:** SDS-PAGE analysis following purification of the S118C Mutant following IMAC (A), shown alongside the elution profile (B) and resulting SDS-PAGE analysis following SEC (C).
**Figure 5.8:** X-band CW EPR of myoglobin S4CS118C SL.
**Figure 5.9:** Confirmation of mutagenesis of S4pK and S118pK in sperm whale myoglobin.
**Figure 5.10:** SDS-PAGE Analysis of Small Scale Test Expressions OF the single pK mutants (A) S4pK and (B) S118pK**.**
**Figure 5.11:** SDS-PAGE analysis from large scale expression of the single mutants.
**Figure 5.12:** SDS-PAGE analysis from large scale expression of double pK mutant S4pKS118pK.
**Figure 5.13:** Reaction Mechanism for the copper catalyzed azide-alkyne cycloaddition of the pK containing peptide to spin label.
**Figure 5.14:** HPLC (Shimadzu) of the reaction mixture at t=0, 30, 90min using Phenomenex $C_{18}$ column 150*4.60mM.
**Figure 5.15:** X-band CW EPR of the S4pK myoglobin Mutant.
**Figure 5.16:** X-band CW EPR of the S4pKS118pK myoglobin Mutant.
**Figure 5.17:** LCT LC Mass Spectrometry analysis of the double pK myoglobin mutant S4pKS118pK**.**
**Figure 5.18:** Comparisons of the Background Corrected EPR Time-Traces and Distance Distributions, of the S4CS118C and S4pKS118pK doubly spin-labelled mutants, shown alongside MMM simulation.

**Figure 5.19:** Chemical Structures of novel nitroxide spin labels (A) BMSL, (B) POMSL, and (C) SLAsH. (visualised using ChemDraw™).
**Figure 5.20:** XRC Structures of the (A) 14-3-3ς and (B) Vps75.
**Figure 5.21:** Elution profile (A) and SDS-PAGE (B) analysis of 14-3-3ς following the final purification step (SEC).
**Figure 5.22:** Elution profile (A) and SDS-PAGE (B) analysis of Vps75 following the final purification step (SEC).
**Figure 5.23:** CW EPR of labelled 14-3-3 ς with (A) MTSSL, (B) SLAsH and (C) BMSL.
**Figure 5.24:** DEER time traces and associated distance distributions for the S4CS118C doubly labelled Met-myoglobin mutant, shown alongside MMM simulation.
**Figure 5.25:** DEER time traces and associated distance distributions for the S4C singly labelled Met-myoglobin mutant to the ferric heme, shown alongside MMM simulation.
**Figure 5.26:** DEER time traces and associated distance distributions for the S118C singly labelled Met-myoglobin mutant to the ferric heme, shown alongside MMM simulation.
**Figure 5.27:** Schematic representation of the NOS homodimer with NOS CaM bound.

## List of Tables

**Table 1.1:** Microwave bands found in commercial EPR spectrometers, with the static magnetic field required for resonance of a free electron**.**
**Table 2.1:** PCR Reaction Mixture for SDM.
**Table 2.2:** Programme used for Site-Directed Mutagenesis PCR.
**Table 2.3:** Final concentration of antibiotics for selection following *E. coli* transformation.
**Table 2.4:** Sequencing primers for pPICZα.
**Table 2.5:** Sequencing primers for T7 promotor and terminator.
**Table 2.6:** Programme used for Sequencing PCR.
**Table 2.7:** PCR Reaction Mixture for CCP Amplification.
**Table 2.8:** Programme used for fH CCP Amplification PCR.
**Table 2.9:** Programme used for PCR Colony Screen.
**Table 2.10:** Antibodies used for Blotting, company purchased from, and working dilution.
**Table 3.1:** Concentration of free cysteine in the fH19-20 G1107C samples following reduction, as determined by the Ellman's assay (DTNB).
**Table 3.2:** FH fragments yields following expression in *P. pastoris*, and purification.
**Table 3.3:** Summary of C3/C3b/C3(N) Yields following Isolation and purification From Plasma.
**Table 3.4:** Compositions of Samples analysed by DEER.
**Table 4.1:** Final Yields of RLC/MiniHMM Mutants per Litre of Cell Culture.
**Table 5.1:** Protein Yields for Myoglobin Mutants Per Litre of Cell Culture.
**Table 5.2:** Test Coupling Conditions of the Peptide to the Spin Label.

# Table of Contents

**Acknowledgement** 2

**Abstract** 4

**Declaration** 5

**Copyright Declaration** 6

**Abbreviations** 7

**Amino Acids** 12

**List of Figures** 13

**List of Tables** 16

**Table of Contents** 17

**Chapter 1: An Introduction to EPR Spectroscopy** 25

1.1 Distance Measurements in Proteins 25

1.2 EPR Theory - The Static Spin Hamiltonian 26

1.2.1 The Electronic Zeeman Splitting Effect 26

1.2.1.1 The Spectroscopic g-factor 27

1.2.1.2 Resonance and the CW EPR Experiment 28

1.2.2 The Hyperfine Splitting 29

1.2.2.1 Spin Label Dynamics 30

1.2.3 The Electron Electron Interaction 32

1.2.3.1 Exchange Coupling 32

1.2.3.2 Dipolar Coupling 32

1.2.4 Zero-Field Splitting 32

1.2.5 Nuclear Quadrupole Interaction 32

1.2.6 Nuclear Zeeman Interaction 33

1.3 Obtaining Distances Between Paramagnetic Centres Using EPR Spectroscopy 33

1.3.1 Distance Measurements Using CW EPR Spectroscopy 34

1.4 Pulsed Methods of EPR Spectroscopy 34

1.4.1 Distance Measurements Using Pulsed EPR Techniques – DEER 35

1.4.1.1 3-Pulse DEER 36

1.4.1.2 4-Pulse DEER 37

1.4.2 Enhancing DEER Sensitivity - Solvent, Temperature, and Deuteration Effects 39

1.4.3 DeerAnalysis 40

1.5 Site-Directed Spin Labelling (SDSL) – Cysteine Substitution Mutagenesis 40

1.5.1 Cysteine Chemistry 41

1.5.2 Nitroxide-Radical Spin Labels 41

1.6 SDSL – Incorporation of Unnatural Amino Acids (UAAs) 43

1.6.1 Solid Phase Peptide Synthesis (SPPS) (Merrifield Coupling) and Semi- Synthesis 43

1.6.2 Expanding the Genetic Lexicon 44

1.6.2.1 Reprogramming Translation Using Orthogonal Ribosomes 44

1.6.2.2 Orthogonal Amino-acyl tRNA Synthetase/ tRNA Pairs 45

1.6.2.3 Incorporation of the UAA p-acetyl Phenylalanine (pAcPhe) in response to the amber Stop Codon 45

1.7 Simulation of DEER Data 46

1.7.1 Multiscale Modeling of Macromolecules (MMM) 46

1.7.2 Mtssl Wizard 47

1.8 Advancing the Field of Structural Biology 47

1.8.1 EPR Spectroscopy in the Study of Large and Flexible Soluble Proteins 47

1.8.2 EPR in the Study of Membrane Proteins 48

1.8.3 EPR as a Stand Alone Technique 49

1.9 Beyond *in vitro* EPR, and towards physiological conditions 49

1.9.1 Trityl and Spiro Radical Spin Labels 50

1.9.2 *in vivo* EPR Studies 51

1.10 Application of EPR in this Thesis 52

**Chapter 2: Materials and Methods** 53

2.1 Mutagenesis, Purification, and Manipulation of DNA for Transformation into Chemically Competent *E. coli* 53

2.1.1 Site – Directed Mutagenesis (SDM) 53

2.1.1.1 Designing Mutants Suitable for EPR 54

2.1.1.2 Primer Design 54

2.1.2 Transformation of Plasmid DNA into Chemically Competent *E. coli* Cells 55

2.1.3 Isolation and Purification of Plasmid DNA 56

2.1.3.1 Mini-Prep 56

2.1.3.2 Maxi-Prep 57

2.1.3.3 Gel Extraction 57

2.1.4 Sequencing of Plasmid DNA 58

2.2 Mutagenesis, Purification, and Manipulation of DNA for Transformation into Electro-Competent *P. pastoris* 60

2.2.1 Amplification of specific CCPs from the full length codon optimised fH gene using PCR 60

2.2.2 DNA Agarose Gel Electrophoresis 61

2.2.3 Cloning of fH Fragments into pPICZαB Vector, Using Restriction Enzymes 61

2.2.3.1 Double Digest 62

2.2.3.2 Ligation of Digested Fragments into pPICZαB 62

2.2.3.3 PCR Colony Screen 62

2.2.4 Preparation of DNA for Transformation into *P. pastoris* 63

2.2.4.1 Phenol-Chloroform Extraction 63

2.2.4.2 EtOH Precipitation 64

2.2.5 Preparation of electro-competent *P. pastoris* cells 64

2.2.6 Electro-transformation of *P. pastoris* 65

2.3 Recombinant Protein Expression Using *E. coli* Expression Systems 65

2.3.1 Expression of RLC/MiniHMM Constructs 65

2.3.1.1 Cell Lysis of RLC/MiniHMM Constructs 66

2.3.2 Expression of Myoglobin Mutants and cMyBP-C Domain C0 66

2.3.2.1 Cell Lysis of Myoglobin Mutants and cMyBP-C Domain C0 by Sonication 67

2.3.3 Expression of 6xHis TEV Protease 67

2.3.3.1 Cell Lysis of 6xHis TEV Protease by Sonication 67

2.3.4 Expression of Myoglobin with the UAA pK Incorporated 67

2.3.4.1 Cell Lysis of Myoglobin containing the UAA, pK, 68

2.3.5 Expression of 14-3-3ς and Vps75 mutants 68

2.3.5.1 Cell Lysis of 14-3-3ς and Vps75 Mutants by Sonication 69

2.4 Recombinant Protein Expression Using the *P. pastoris* Expression System 69

2.4.1 The *P. pastoris* Expression System 69

2.4.2 Small-scale protein production 71

2.4.3 Large-scale protein production 72

2.4.4 Fermentation 72

2.5 Purification of Recombinant Proteins 73

2.5.1 Purification of 6xHIS TEV Protease Using IMAC 74

2.5.2 Purification of RLC/MiniHMM Constructs Using IMAC 74

2.5.2.1 Cleavage of the Histidine Tag 74

2.5.3 Purification of Wild Type, Cysteine and pK Containing Myoglobin Mutants and cMyBP-C Domain C0 using IMAC 75

2.5.3.1 Size Exclusion Chromatography 75

2.5.4 Purification of 14-3-3 ς Using IMAC 75

2.5.4.1 Size Exclusion Chromatography 76

2.5.5 Purification of Vps75 Using IMAC 76

2.5.5.1 Size Exclusion Chromatography 76

2.5.6 Purification of fH 1-4 K247C Fragment Using IMAC 76

2.5.6.1 Anion Exchange Chromatography 77

2.5.6.2 Size Exclusion Chromatography 77

2.5.7 Purification of fH19-20 G1107C Fragment Using SP Sepharose FastFlow Cation Exchange Chromatography Resin 77

2.5.7.1 Cation Exchange Chromatography 78

2.5.7.2 Size Exclusion Chromatography 78

2.6 Isolation and Purification of Complement Proteins From Human Blood/Plasma 78

2.6.1 Small-scale isolation of complement C3 78

2.6.1.1 Obtaining the Plasma Fraction from Whole Blood 79

2.6.1.2 Precipitation of Major Blood Proteins 79

2.6.1.3 Anion Exchange Chromatography – Q-Sepharose Fast Flow Resin 79

2.6.1.4 Anion Exchange Chromatography – Mono Q 79

2.6.1.5 Size Exclusion Chromatography 80

2.6.2 Large Scale Isolation of Complement C3 80

2.6.3 Conversion of C3 to C3b using Trypsin 80

2.6.4 Conversion of C3 to ‘C3 (N)’ – An Alternative to C3b 81

2.7 Protein Quantification 81

2.7.1 Sodium-Dodecyl Sulphate Polyacrylamide Gel Electrophoresis(SDS-PAGE) 81

2.7.2 Western Blotting 82

2.7.3 Determination of Protein Concentration Using Absorbance $A_{280}$ 83

2.7.4 Determination of Protein Concentration Using the Micro-Bradford Assay 83

2.8 Protein Characterisation 84

2.8.1 Mass Spectrometry (MS) Analysis 84

2.8.2 Ellman's Assay - For the Quantification of Free Cysteine 84

2.9 Protein Chemistry 86

2.9.1 SDSL of Incorporated Cysteine Residues 86

2.10 Functional Characterisation 87

2.10.1 Assessing Cofactor Activity of C3b, C3 (N) and fH1-4 K247C 87

2.11 EPR Spectroscopy 88

**Chapter 3: The Complement Factor H (fH) – C3b Complex: An EPR Study** 89

3.1 The Role of the Complement System in Innate Immunity 89

3.1.1 Mechanisms of Complement Activation 90

3.1.2 The Alternative Pathway of Complement Activation 90

3.1.3 Complement Activation – Driven by Conformational Change 91

3.1.4 Derivatization of Complement C3 – Labelling with Bio-Reporters 93

3.1.5 Regulators of Complement – fH 95

3.1.6 The fH-C3b Interaction 98

3.1.7 Aims of this work 101

3.2 Simulation of the DEER Experiment for the fH:C3b Complex Using MMM 103

3.3 Generation of Spin-Labelled fH1-4 K247C Fragment for EPR Studies, using the *P. pastoris* Expression System 105

3.3.1 Mutagenesis 105

3.3.2 Small Scale Test Expression of fH K247C 105

3.3.3 The need for an Affinity Tag? 106

3.3.4 Mutagenesis – Incorporation of a C' Histag onto fH1-4 K247C 107

3.3.5 Small Scale Test Expression of fH1-4 K247C 107

3.3.6 Large Scale Expression of fH1-4 K247C 108

3.3.7 Purification of fH1-4 K247C 109

3.3.7.1 IMAC 109

3.3.7.2 Anion Exchange Chromatography (AEC) Tricorn™ Mono Q 4.6/100 PE 110

3.3.7.3 Size Exclusion Chromatography 111

3.4 Generation of a Spin-Labelled fH19-20 Fragment for EPR Studies using the *P. pastoris* Expression System 111

3.4.1 Mutagenesis 111

3.4.2 Small Scale Test Expression of fH19-20 G1107C 112

3.4.3 Fermentation of fH19-20 G1107C 113

3.4.4 Purification of fH19-20 G1107C Following Fermentation 114

3.4.4.1 SP-Sepharose FastFlow™ CEC Resin 114

3.4.5 Selection of an Appropriate Reducing Agent 115

3.4.6 Purification of fH19-20 G1107C Following Large Scale Flask Expression 119

3.4.6.1 SP-Sepharose FastFlow™ CEC Resin 120

3.4.6.2 Resource 15S™ CEC 120

3.4.6.3 Size Exclusion Chromatography 122

3.5 Spin labelling of fH Fragments 1-4 K247C and 19-20 G1107C 122

3.6 Generation of Spin-Labelled C3b, Following Isolation and Purification of C3 126

3.6.1 Small Scale Isolation of C3/C3b 126

3.6.1.1 Q-Sepharose FastFlow™ AEC Resin 126

3.6.1.2 Anion Exchange Chromatography (AEC) – Tricorn™ Mono Q 4.6/100 PE 127

3.6.2 Conversion of C3 to C3b Using Limited Trypsin Digestion 128

3.6.3 Small Scale Isolation of C3/C3b – Troubleshooting 130

3.7 Generation of Spin-Labelled C3(N), an Alternative to C3b, following Isolation and Purification of C3 132

3.8 Spin labelling of C3b and C3(N) 135

3.9 Functional Characterisation – fI mediated cleavage of C3b and C3(N) in the presence of fH 136

3.10 The fH1-4. C3(N). fH19-20 Complex – An EPR Study 137

3.11 Outlook and Future Work 141

**Chapter 4: Probing the Interaction of the Cardiac Myosin RLC with cMyBP-C Using EPR Spectroscopy** 143

4.1 Muscle Contraction - The 'Dance' of Actin and Myosin, and regulation in the sarcomere 143

4.1.1 Cardiac Myosin Structure 145

4.1.2 The Light Chain Domains – ELC & RLC 147

4.1.3 Myosin Binding Protein-C (cMyBP-C) 147

4.1.4 The RLC-C0 interaction 151

4.1.5 Aims of this work 151

4.2 Small Scale Test Expressions of cMyBP-C C0 and RLC/MiniHMM 153

4.3 Expression and Purification of cMyBP-C Domain C0 154

4.4 Optimisation of Expression of the WT RLC/MiniHMM Complex 155

4.4.1 Modification of the MiniHMM Construct 156

4.5 Large Scale Expression of wt RLC/MiniHMM 157

4.5.1 IMAC of wt RLC/MiniHMM 158

4.5.2 Size Exclusion Chromatography of wt RLC/MiniHMM 159

4.6 Generation of Cysteine Mutants of the myosin RLC for SDSL Purposes 160

4.7 Simulation of the DEER Experiment for RLC/MiniHMM Cysteine Mutants 161

4.8 Generation of Cysteine Mutants of the myosin RLC for SDSL Purposes 164

4.9 Large Scale Expression and Purification of cardiac Myosin RLC Cysteine Mutants 165

4.9.1 Further Purification of K30C RLC/MiniHMM 167

4.10 Spin labelling of the E88C RLC/MiniHMM Mutant 169

4.11 DEER Measurements of E88C RLC/MiniHMM, with and Without cMyBP-C domain C0 170

4.12 DEER Measurements of E88C in High and Low NaCl 170

4.13 Outlook and Future Work 174

**Chapter 5: Development of New Strategies for the Site-Directed Spin –Labelling of Cysteine Rich Proteins** 177

5.1 The Limits of SDSL by Cysteine Substitution Mutagenesis 177

5.2 Site Directed Site Labelling *via* Reassignment of the Amber Stop Codon 178

5.2.1 Translation of mRNA into a Polypeptide at the Ribosome 178

5.2.2 Incorporation of Unnatural Amino Acids (UAAs) into Recombinant Proteins 180

5.3 Project Aims 180

5.4 Generation of Single (SC4 &S C118) and Double Cysteine (S4CS118C (or C4C118)) Myoglobin Mutants 183

5.4.1 Mutagenesis 183

5.4.2 Expression and Purification of the S4CS118C Myoglobin Mutant 184

5.4.3 Expression and Purification of the Single Cysteine Mutants S4C and S118C 185

5.4.4 Spin Labelling of Cysteine Myoglobin Mutants 187

5.5 Expression and Purification of Sperm Whale Myoglobin Mutants with the UAA, pK. 188

5.5.1 Mutagenesis 188

5.5.2 Small-Scale Expression of Myoglobin pK Single Mutants 189

5.5.3 Large Scale Expression and Purification of Myoglobin Single pK Mutants 190

5.5.4 Expression of the double mutant S4pKS118pK 191

5.6 Coupling of a Peptide Containing pK to an Azide Spin Label Using 'Click' Chemistry 192

5.7 Coupling of pK containing myoglobin to an azide spin label 194

5.7.1 From Peptide to Protein - Optimisation of Coupling 194

5.7.2 Spin Labelling of S4pK Using Copper Catalysed 'Click' chemistry 195

5.7.3 Spin Labelling of S4pKS118pK Using Copper Catalysed 'Click' chemistry 195

5.8 Confirmation of Spin Labelling of the S4pKS118pK mutant using Mass Spectrometry 196

5.9 DEER Experiments 197

5.10 Outlook and Future Work 199

5.11 A New Generation of Spin Label 202

5.11.1 Expression and Purification of the 14-3-3 ς and Vps75 Proteins 202

5.11.2 Spin Labelling of 14-3-3ς 204

5.12 Obtaining Nanometre Scale Distances Between the Ferric Heme of Sperm Whale Myoglobin and Nitroxide Spin Labels Using DEER 205

5.12.1 Aims of this work 206

5.13 Expression and Purification of Low Spin (s= ½) Myoglobin Mutants 207

5.14 DEER Experiments 207

5.14.1 Nitroxide-Nitroxide DEER Measurements (S4CS118C) at W-Band 208

5.14.2 Nitroxide-Heme ($Fe^{III}$) DEER Measurements at W-Band 209

5.15 Beyond MTSSL - Outlook and Future Work 211

**Appendices** 213

**References** 238

# Chapter 1: An Introduction to Electron Paramagnetic Resonance Spectroscopy (EPR) and its Application in Measuring Nanometre (nm) Scale Distances in Proteins

## 1 An Introduction to EPR Spectroscopy

Understanding how proteins and their complexes interact at a three dimensional, structural level is crucial in elucidating function. Obtaining distances between sites of interest in proteins and complexes thereof offers further insight into such interactions.

This thesis aims to explore the application of EPR Spectroscopy, in particular the use of Double Electron-Electron Resonance (DEER) Spectroscopy, in the study of protein-protein interactions in large functional bio-macromolecular complexes.

The introduction to EPR spectroscopy is summarised by Atkins & de Paula, 2011, and Brustolon & Giamello, 2009.

### 1.1 Distance Measurements in Proteins

Generally, spectroscopy can be defined as a measurement of the interaction of a given material with electromagnetic radiation. Molecules, or atoms, have discrete states, each of which has a corresponding energy. Spectroscopy can be used to measure and interpret the difference between these energy states ($\Delta E$), gaining further insight into the structural dynamics of the sample of interest. EPR is no exception, and involves the application of microwave frequency electromagnetic radiation to activate and specifically detect unpaired electrons.

Naturally occurring organic cofactors, *e.g.* flavin radicals, metal centres, *e.g.* copper, and metal clusters, *e.g.* ferric heme, in proteins are held in a fixed orientation with respect to the protein structure, and therefore are ideal spin probes, if paramagnetic. Very often, however, the EPR spectra of metal centres and naturally occurring radicals are too broad, and consequently very difficult to study by DEER (Bowen *et al*, 2013). Alternatively, spin labels containing radicals which are sensitive to their local environment can be incorporated site-specifically into proteins of interest, allowing indirect molecular observations of the protein/complex under study, at precisely defined locations within the protein structure.

Due to the relationship between the coupling of unpaired electrons and the distance between them, one can collect several long range nm distance constraints between spin labels. This allows one to yield global structural elements from proteins of interest, and gain further insight into how the small pieces come together, and how this contributes to overall biological function.

## 1.2 EPR Theory - The Static Spin Hamiltonian

These energy differences, $\Delta E$, can be measured due to the relationship between $\Delta E$ and the absorbance of electromagnetic radiation, as shown in equation 1.1

$$\Delta E = h\nu, \quad \text{(EQUATION 1.1)}$$

where $h$ is Plank's constant ($6.62607 \times 10^{-34}$ Js) and $\nu$ is the frequency of the radiation.
The EPR spectrum of an unpaired electron in an external magnetic field $B_0$, can be described by a Hamiltonian of the form described in equation 1.2:

$$H_0 = H_{EZ} + H_{HFS} + H_{EE} + H_{ZFS} + H_Q + H_{NZ} \quad \text{(EQUATION 1.2)}$$

where the above terms stand for the Electronic Zeeman Interaction ($_{EZ}$), Hyper-Fine Splitting ($_{HFS}$), the Electron-Electron Interaction ($_{EE}$), Zero-Field Splitting ($_{ZFS}$), the Quadrupolar Interaction($_Q$), and the Nuclear Zeeman Interaction ($_{NZ}$), respectively.

The greater majority of the work undertaken in this thesis has been done so using spin labels containing the nitroxide radical (**NO**$^{\cdot}$), with spin ½. For such spin systems the $H_{EZ}$, $H_{ZFS}$ and $H_{EE}$ contributions are the most significant, although all contributions to the spin Hamiltonian will be discussed.

### 1.2.1 The Electronic Zeeman Splitting Effect

Each unpaired electron has an intrinsic angular momentum, or 'spin', denoted $S$. Due to the charge carried by each unpaired electron, the angular motion of this particle generates a magnetic field, resulting in each unpaired electron displaying paramagnetism when aligned with an external magnetic field ($B_0$). The magnetic field, $B_0$, produces two energy levels for the magnetic moment, $\overline{\mu}$, of the electron, causing a splitting of the otherwise degenerate spin energy levels.

A single electron has a spin, S of ½, with magnetic components of $m_s \pm ½$, making it doubly degenerate, and restricted to two positions, referred to as up and down. When an external magnetic field is applied there is an absorption of energy, causing an EPR transition from the lower energy state (when aligned in the same direction as the field ($m_s$= - ½)) to the higher energy state (when aligned against the magnetic field direction ($m_s$ = +½)). This movement between energy levels is the Zeeman Splitting Effect, which is illustrated in Figure 1.1.

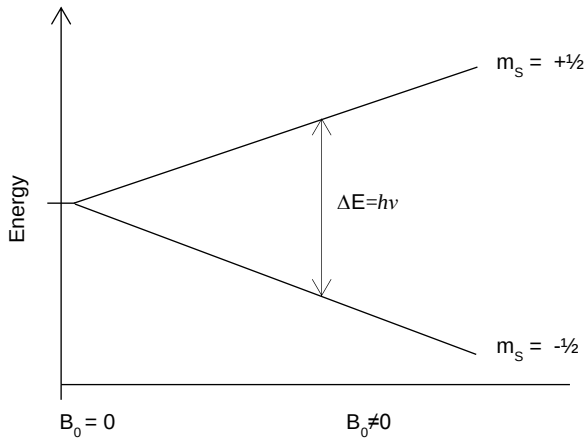



**Figure 1.1: The Electronic Zeeman Splitting Effect.** Divergence of energy levels of discrete spin states with application of magnetic field for a S= ½ system.

The energies for an electron with magnetic moments $m_s \pm ½$ are given by equations 1.3 and 1.4, respectively

$$E_{+½} = ½\, g_e \beta B_0 \quad \text{(EQUATION 1.3).}$$

$$\text{and} \quad E_{-½} = -½\, g_e \beta B_0 \quad \text{(EQUATION 1.4).}$$

for each electron, $\bar{\mu} = m_s g_e \beta$, where $\beta$ is a conversion constant called the Bohr magneton (9.27401 x$10^{-24}$ J x $T^{-1}$) and $g_e$ is the spectroscopic g-factor of the electron. The g factor is a constant of proportionality, whose value is the property of the electron in a certain environment.

### 1.2.1.1 The Spectroscopic g-factor

In order to interpret EPR spectra of radicals incorporated into proteins, one must first be able to compare the spectrum of the sample of interest with that of a free electron. The g-factor of a dimensionless electron in space is 2.003192778 (~2.00). Combining

equations 1.3 and 1.4, in a magnetic field ($B_0$) the energy levels of a free electron are split by the following, equation 1.5:

$$\Delta E = h\nu = g_e.\beta.B_0 \quad \text{(EQUATION 1.5).}$$

however, the magnetic moment of the nitroxide radical interacts with the external magnetic field in a different manner than that of a free electron.

Therefore, on account of local magnetic fields induced in the specific molecular framework of the nitroxide radical, the resonance condition is normally denoted by that described in equation 1.6:

$$\Delta E = h\nu = g.\beta.B_0 \quad \text{(EQUATION 1.6),}$$

where $g$ is equal to the g-value of the specific radical under study.

Typically, organic radicals have g-values of ~ 2.0027, inorganic radicals between 1.9-2.1, and paramagnetic d-metal complexes anywhere between 0 and 6. A measurement of a system's g factor yields some level of information about electronic structure, although it does not give much in way of structural dynamics.

#### 1.2.1.2 Resonance and the CW EPR Experiment

By keeping the electromagnetic radiation frequency constant and scanning the magnetic field, a peak in absorption will occur when the magnetic field tunes the two spin states, so that the difference in energy, $\Delta E$, matches the energy of the radiation. This is called the field for resonance and is the technique used in Continuous Wave (CW) EPR Spectroscopy. Typically for such experiments, in a 0.3 T static field, the frequency, $\nu$, of electromagnetic radiation required to cause an inversion of the unpaired electron spin is about 9 GHz. This is often referred to as X-band frequency.

The application of a resonant microwave frequency magnetic field induces the transitions between energy states, and is how CW EPR can be used to measure the resonance. Electrons are not isolated, and can exchange energy with their surroundings and with other spins in the system. Relaxation of the electron to its ground state is crucial in understanding magnetic resonance, as the shape of the EPR spectra obtained is determined by both the spin-lattice ($T_1$) and spin-spin ($T_2$) relaxation rates.

**Table 2.1 Microwave bands found in commercial EPR spectrometers, with the static magnetic field required for resonance of a free electron**

| Designation (band) | ν (GHz) | Field (T) |
|---|---|---|
| X-Band | 9 | 0.3 |
| Q-Band | 35 | 1.0 |
| W-Band | 95 | 3.4 |

Historically, microwave frequencies are divided into bands, with most commercially available EPR spectrometers operating at X, Q and W band, which are outlined in Table 1.1 with the corresponding magnetic field required for the resonance of a free electron. At higher field, the spectral resolution of the nitroxide radical improves since the Zeeman splitting increases with the applied magnetic field.

### 1.2.2 The Hyperfine Splitting

In order to interpret EPR spectra, one must take into account the effect that magnetic nuclei have on the energy of the unpaired electron. In EPR, the hyperfine structure is the result of the magnetic interaction between the electron spin, and the magnetic dipole moments of the nuclei, in the radical under study. Generally a nucleus with spin, $I$, splits the EPR spectrum into $2I+1$ hyperfine lines. For a nitroxide radical ($I$=1), the EPR spectrum consists of three lines of equal intensity (Figure 1.2), due to the three possible spin orientations of the $^{14}N$ nucleus.

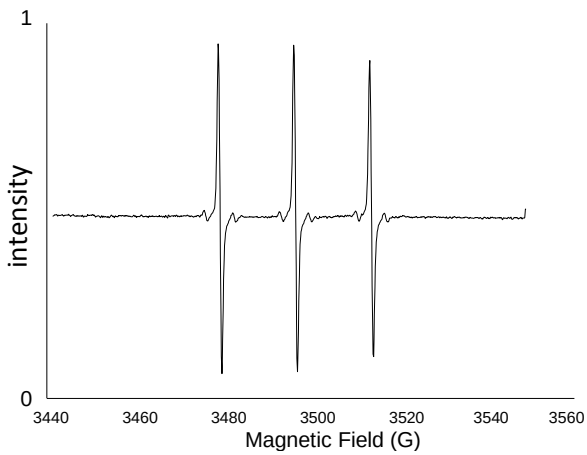



**Figure 1.2: Near Isotropic 3 line Spectrum of 4-amino-TEMPO.** The nitroxide containing small molecule TEMPO is in double distilled $H_2O$ ($ddH_2O$) (spectrum normalised to maxima).

In theory, the total energy absorbed by a sample is measurable, however, in practice the noise component significantly obscures the resonance signal. For this reason, the static magnetic field is modulated, and only the modulated part of the output voltage detected. The output is therefore the detection of the first derivative of the absorption, rather than the absorption itself, but with an enhanced signal to noise ratio. The hyperfine structure of the EPR spectra is a basic fingerprint of this interaction from which it is possible to gleam a host of information. Figure 1.2 shows the near isotropic CW spectra of the nitroxide containing 4-amino TEMPO.

The hyperfine coupling has anisotropic components which influence the CW EPR line-shape. A small molecule containing a nitroxide radical, such as TEMPO – see Figure 1.2, tumbling rapidly in solution has an average hyperfine splitting, typically of about 15G. The hyperfine anisotropy is due to the majority of the electron density being distributed in molecular orbitals (p orbitals of the nitrogen lone pair), aligned parallel to the molecular z axis of the paramagnetic moiety. Typically therefore $A_{zz}>A_{yy},A_{xx}$ and this becomes evident for crystal or powdered samples.

Furthermore the hyperfine coupling, and consequently the CW EPR linewidth is sensitive to variation from the spin label side chain sampling different water, protein and lipid environments. Bordignon & Steinhoff, 2007 confirmed this with their studies on solvent accessibility in membranes.

Generally as the polarity and proticity of the environment increases, so does the hyperfine coupling, showing the potential for defining the topology of the spin label side chain with respect to other proteins, as well as the identification of water-membrane boundaries.

Characterising specific regions of proteins in terms of their polarity/proticity profiles, as well as identifying hydrophobic barriers by monitoring changes in hyperfine coupling, may provide means to obtain structural and topological detail of proteins, and help elucidate further specific biological process.

#### 1.2.2.1 Spin Label Dynamics

The appearance of the EPR spectrum changes as the motion of the radical is restricted, and rotational correlation time increases due to the anisotropic nature of the hyperfine tensor. Figure 1.3 shows the effect of an increased rotational correlation time on the

EPR line-shape, as well as the effect of freezing. Restricted motion may be as a result of a reduced tumbling rate due to attachment of the spin label to a particularly large protein, or restriction of the label's rotation through binding position, temperature reduction or an increase in viscosity (Beier & Steinhoff, 2006).

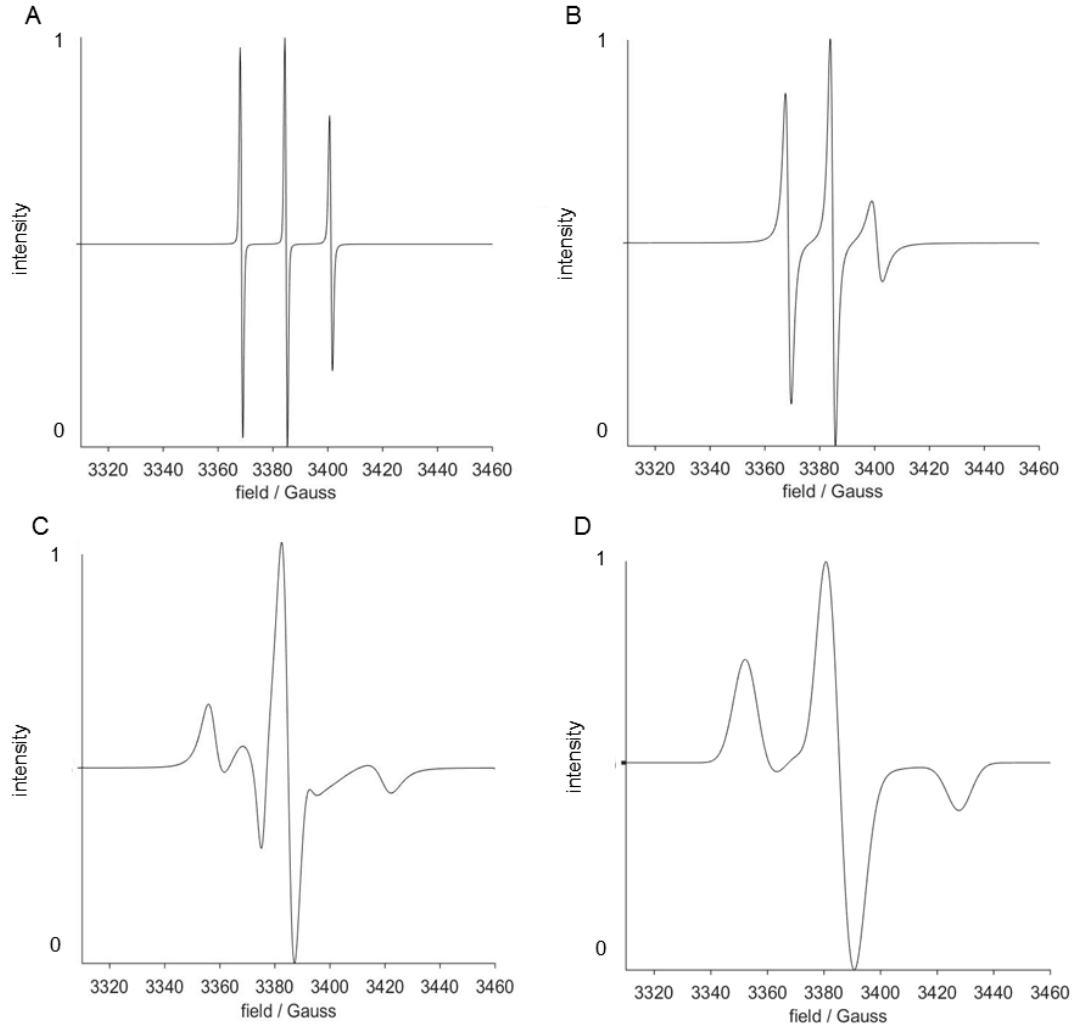



**Figure 1.3: Spectral Simulations of a nitroxide radical (such as TEMPO), at X-band.** Shows increasing correlation times, as simulated using the Easyspin Spectral Simulation Software (Chili and Pepper Functions) for MatLab® (spectra normalised to maxima). (A) correlation time of $0.1 \times 10^{-9}$s, (B) correlation time of $1 \times 10^{-9}$s, (C) correlation time of $10 \times 10^{-9}$s, (D) simulated powder spectrum of a nitroxide (*e.g.* frozen).

Ideally, spin labels should produce an EPR spectrum that significantly broadens as its motion is even slightly restricted. This is true for nitroxide spin labels, as we can see an increased rotational correlation time from Figures 1.3A to 1.3C. Assuming total intensities are the same across all simulated spectra, it is clear that an increase in

correlation time results in a broadening of the spectral lines due to an increased rigidity, and more restricted motion of the spin label.

### 1.2.3 The Electron-Electron Interaction

The interaction of two (or more) electrons results in a coupling of spin states. There are two magnetic interactions that operate between two interacting unpaired electrons, one being the isotropic exchange coupling interaction and the second the anisotropic through space dipole-dipole interaction.

#### 1.2.3.1 Exchange Coupling

The exchange interaction, characterised by the exchange coupling tensor, *J*, is dependent upon the electric interaction between electrons, from a combination of repulsion and overlapping of orbital densities. This interaction falls off exponentially as the distance between interacting centres increases, assuming no orbital overlap.

#### 1.2.3.2 Dipolar Coupling

The dipolar coupling interaction will be discussed in greater detail in 1.4. Briefly, the dipolar coupling is the interaction between two magnetic moments, which for the purpose of this thesis will be the magnetic moments between two interacting electrons separated in space.

### 1.2.4 Zero-Field Splitting

The ZFS interaction is only relevant when applied to high spin systems, *i.e.* $S > ½$. Zero-field splitting describes various interactions of the energy levels of an electron spin in the absence of an applied magnetic field. The electrons couple to one another, causing an energy level splitting called the ZFS. This can be very large, *e.g.* common transition metal ions in biological systems, such as iron, have such large ZFS energies that their spectra have still to be fully measured even at very high magnetic fields (Bowen *et al*, 2013).

### 1.2.5 Nuclear Quadrupole Interaction

The Nuclear quadrupole interaction is analogous to the zero-field interaction, as these transitions can be detected in the absence of an applied external field. Nuclei with $I \geq 1$,

possess an electrical quadrupole moment *Q*, arising from their non-spherical charge distribution. This charge distribution then interacts with the electric-field gradient produced by the electrons in the vicinity.

In EPR spectra, nuclear quadrupole interactions can cause a shift in resonance, and the appearance of forbidden transitions, however, such effects are usually of the second order and difficult to observe.

### 1.2.6 Nuclear Zeeman Interaction

This contribution is analogous to the electronic Zeeman, but addresses the coupling of nuclear spins to the external magnetic field. For the most part, this can be considered isotropic. This small energy contribution has little influence on EPR spectra, however, may affect nuclear frequency spectra measured by EPR techniques, such as Electron-Nuclear Double Resonance (ENDOR) Spectroscopy.

## 1.3 Obtaining Distances Between Paramagnetic Centres Using EPR

Distance determination across all methods of EPR relies upon the magnetic dipole-dipole interaction between the magnetic moments $\overline{\mu}_A$ and $\overline{\mu}_B$, of two spins A and B. For the purpose of this thesis, the focus will be on dipolar interactions between pairs of electrons, from which the distance between them can be resolved.

The energy of the dipolar interaction is a result of the change in Zeeman energy of the observed spin A, due to the presence of the second spin, B. In the high field approximation, the dipolar coupling to the external magnetic field dominates all other contributions. Hence the dipoles align parallel to $B_0$. The energy of the dipolar coupling interaction is simplified to that shown in equation 1.7:

$$E=\frac{\mu_0}{4\pi}\frac{1}{r^3}(1 - 3\cos^2\theta) \quad \text{(EQUATION 1.7)}$$

where $\mu_0$ is equal to the vacuum permeability constant, $r$ is the distance between the two unpaired electrons and $\theta$ is the angle formed by the applied field and the vector connecting the unpaired electrons.

The $\theta$ value can give information on the orientation, (Prisner *et al*, 2015, Gophane *et al*, 2014) however, this is not of interest for the work described in this thesis, where the focus is nitroxide spin labels at X and Q bands, which, due to conformational freedom,

and limited g-value splitting at X and Q band, do not give much away in terms of orientation.

### 1.3.1 Distance Measurements Using CW EPR Spectroscopy

CW EPR is most often applied to nitroxides whose powder spectra (*i.e.* static, *e.g.* frozen) are dominated by the (inhomogeneous) broadenings from nitrogen hyperfine interactions and *g* tensors (Bowen *et al*, 2013). The effect of dipolar coupling is a small broadening effect, which can be extracted using deconvolution methods, (Rabenstein & Shin, 1995) or multi-parameter fits (Hustedt *et al*, 1997) and from the dipolar coupling interaction, distances can be determined.

CW EPR is practical for the measurement of short distances of ~ 0.8nm up to a maximum of 1.5-2.0nm, with those distances less than 1.5nm being the most reliable (Banham *et al*, 2008). For protonated spin labels, the inhomogeneous line broadening (in part due to unresolved hyperfine coupling) obscures dipolar broadening for distances in the 1.5-1.7nm range resulting in lower distance resolution.

With increasing distance between paramagnetic centres, it becomes progressively more difficult to measure the dipolar interaction with CW EPR as an effect on line broadening, as the coupling strength becomes comparable to the linewidth. To extend this limit, with respect to CW EPR, the spin label can be deuterated (de Vira *et al*, 2015) or trityl spin labels may be used as an alternative to nitroxides. Due to the much narrower line-shape of trityl radicals (Kunjir *et al*, 2013) inter-spin distances can be accurately measured from the CW EPR spectrum up to ~2.4nm (Reginsson *et al*, 2012).

## 1.4 Pulsed Methods of EPR Spectroscopy

Unlike CW EPR spectroscopy, where the microwave frequency is kept constant throughout, pulsed methods of EPR spectroscopy apply a series of microwave pulses, in order to refocus the broadening that occurs in spectra obtained from CW EPR, and collect more high resolution data. The Hahn echo pulse set ($\pi/2$-$\tau$-$\pi$-$\tau$) is the basis for the majority of pulse sequences used in such techniques (Figure 1.4).

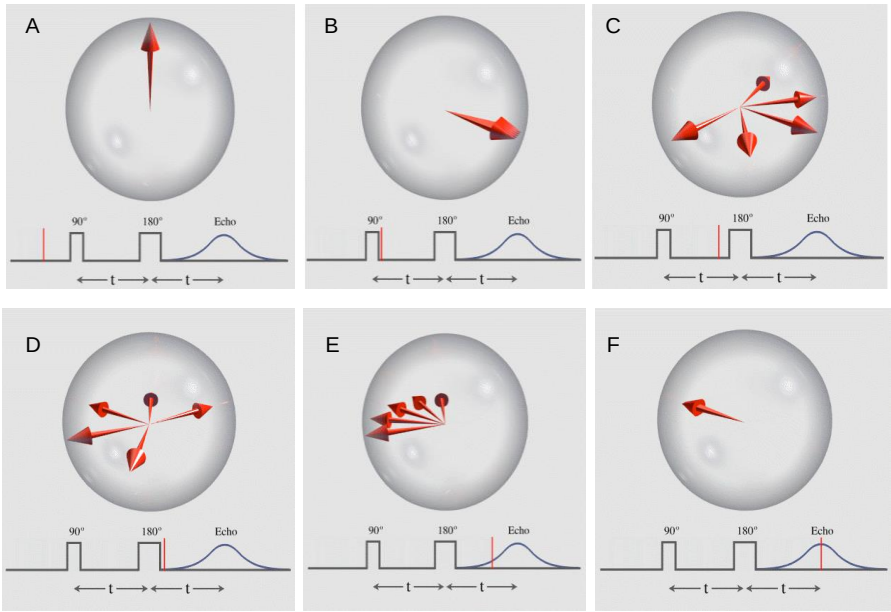



**Figure 1.4: The Principle of the Hahn Spin Echo.** (A) The vertical red arrow is the average magnetic moment of a group of spins. All are vertical in the vertical magnetic field and spinning on their long axis, but this illustration is in a rotating reference frame where the spins are stationary on average. (B) A 90° ($\pi/2$) pulse is applied that flips the arrow into the horizontal plane. (C) Due to local magnetic field inhomogeneity some spins slow due to lower local field strength while some speed up due to higher field strength. This makes the signal decay. (D) A 180° ($\pi$) pulse is applied so that the slower spins lead ahead of the main moment and the fast ones trail behind. (E) The fast moments catch up with the main moment and the slow moments drift back toward the main moment. (F) Complete refocusing has occurred and at this time, an accurate $T_2$ echo can be measured.

Pulsed methods of EPR spectroscopy come in many flavours. Electron Nuclear Double Resonance (ENDOR) and Hyperfine Sublevel Correlation Spectroscopies (HySCORE) are just two pulsed methods of EPR spectroscopy that can be used to probe the environment surrounding paramagnetic centres.

For more on other pulsed EPR methods see Borbat & Freed, 2013. However, the focus of this thesis is the use of DEER spectroscopy in obtaining nm scale distance measurements between paramagnetic centres, based on the dipolar interactions between them.

### 1.4.1 Distance Measurements Using Pulsed EPR Techniques - DEER

Whilst CW EPR is used for accurate distance determination of the shortest distances, it is the pulse techniques which allow access to long range distance distributions, more synonymous with large functional biological complexes. The focus of this thesis is the

use of DEER for studying the structure of biomolecules on the nanoscale, by measuring the dipolar coupling between paramagnetic centres.

Using DEER, the dipolar couplings are isolated by suppression of the hyperfine couplings, allowing longer range distance measurements to be obtained. Until recently the often quoted range for DEER distance measurements was 1.5-8nm (Tsvetkov & Grishin, 2009) (Gunnar Jeschke, 2012), however recent work by El Mkami (2014) and co-workers have pushed the limits to a theoretical maximum of 13nm.

#### 1.4.1.1 3-Pulse DEER

The constant time 3-Pulse DEER sequence (shown in Figure 1.5) was initially used to gain further insight into molecular distribution by Milov *et al*, 1981.

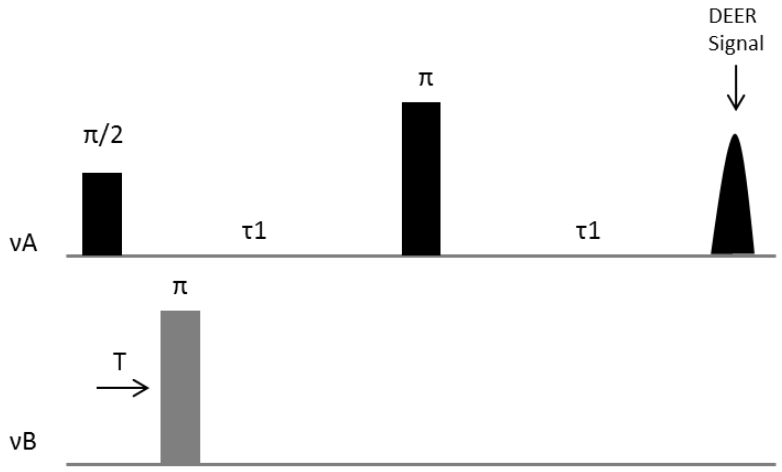



**Figure 1.5: Schematic representation of the 3-pulse DEER sequence.** Observer and pump pulses shown at frequencies A and B ($\nu_A$, $\nu_B$).

Application of the π pulse at $\nu_B$, between the Hahn Echo pulse sequence (π/2, π) causes a focusing of the electrons coupled to the observer spins. At time T, the local field at the A spins (in resonance with $\nu_A$) is altered. T is incremented, and the extent of coupling determines the modulation period.

However, the start of the DEER time trace is usually distorted due to application of the pump pulse, at $\nu_B$, at the same time as the first observer pulse, π/2, at $\nu_A$. Consequently, the complete shape of the dipolar spectrum is not recovered, leading to a 'dead-time', which proves detrimental if distance distributions are broad (Lovett *et al*, 2012).

4-Pulse DEER, although less sensitive than 3-pulse DEER, is commonly used in measuring distance distributions in structural biology investigations, where broad distance distributions are more commonplace.

#### 1.4.1.2 4-Pulse DEER

The four pulse DEER technique provides a dead-time-free method for measuring the frequency of the dipolar coupling interactions and is the chosen method used in this thesis.

For a two spin system containing spins A and spin B, both from nitroxide radicals, the spins are differentiated by exciting different parts of the nitroxide spectrum, using two very different microwave frequencies ($v_A$ and $v_B$), as illustrated in Figure 1.6B and 1.6C.

A two-pulse Hahn echo sequence is employed to selectively monitor the intensity of the echo (Figure 1.6C), $v$(T), of the paramagnetic species A, at frequency $v_A$, in resonance with spins A. This particular pulsed sequence generates a refocused echo after the last delay, τ2, which is the measurable DEER signal. The π pulse at microwave frequency $v_B$, applied during the time interval T, often referred to as the pump pulse, then excites spins B, in resonance with frequency B, $v_B$. This introduces a coupling between spins A and B.

Using the two microwave frequencies $\nu_A$ and $\nu_B$, and keeping the refocused echo at a fixed position in the time domain, strongly suppresses the hyperfine interactions. The time T, at which the pump pulse is applied, is incremented, resulting in an increase or decrease in the intensity of the refocused echo. The frequency of this oscillation, $\nu_{AB}$, can be used to obtain the dipolar coupling.

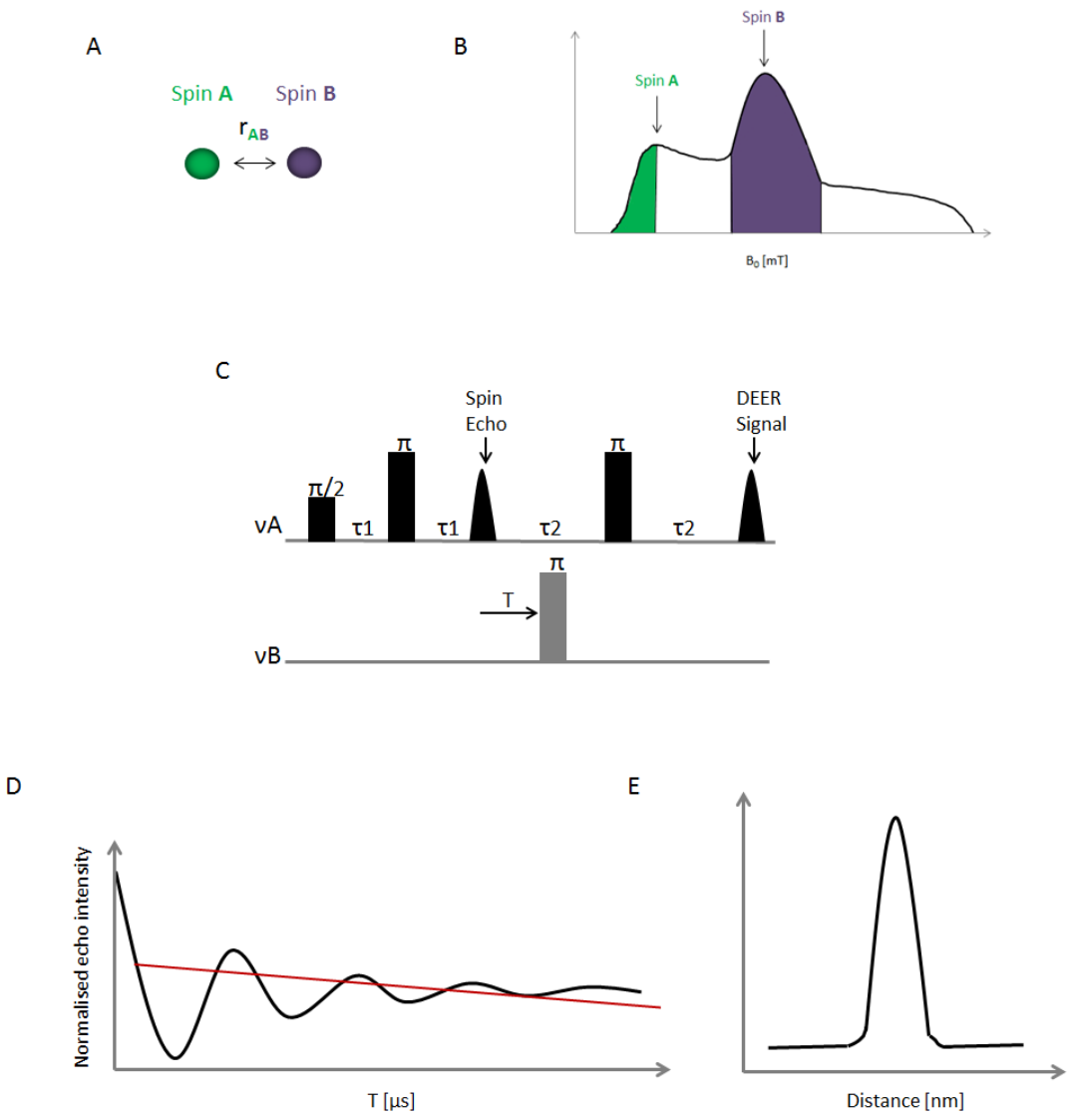



**Figure 1.6: Schematic representation of the 4-Pulse DEER Experiment.** (A) For a two spin system, the distance between Spin A (green sphere) and Spin B (purple sphere) is denoted $r_{AB}$. (B) Field sweep spectrum of a nitroxide at X-band, illustrating where in the spectrum Spins A and Spins B are excited (green and purple respectively). (C) The dead-time free 4 pulse DEER sequence. (D) The four pulse DEER time trace (black line) together with the fitted intermolecular background decay (red line). (E) Background corrected distance distributions $r_{AB}$ obtained following Tikhonov regularisation of the DEER time trace (Jeschke 2006).

Due to potential intermolecular spin-spin interactions, there may be an exponential decay of the DEER signal. These interactions are the background contribution (as shown in Figure 1.6D), and so the DEER trace obtained must first be background corrected, in order to obtain an accurate representation of the intramolecular spin-spin distance distribution (as shown in Figure 1.6E). A spin concentration of between 50$\mu$M

and 200$\mu$M will minimise this contribution, whilst ensuring sensitivity. For the experiments performed throughout, a final sample of ~80$\mu$L is required.

### 1.4.2 Enhancing DEER Sensitivity - Solvent, Temperature and Deuteration Effects

Typically, DEER experiments are conducted at temperatures in the range of 50K, with the protein solution forming a frozen glass. At 50K, one of the major factors contributing to the persistence of an echo, and therefore the sensitivity and measureable distances between spin labels, is the electron spin echo dephasing time, $T_M$.

Due to the cryogenic temperatures involved, cryo-protectants are required to prevent ice crystal formation and protein aggregation, which can leads to a dramatic decrease in the $T_M$, therefore seriously compromising the signal to noise ratio (S de Vera *et al*, 2015).

Due to relatively short $T_M$ (Ward *et al*, 2010) when measuring proteins, it is standard practice to deuterate the solvent in which the proteins are present, which slows relaxation and extends the range of distance measurements and level of sensitivity (Huber *et al*, 2001).

In a non-deuterated environment, short spin echo dephasing times – between 2-4$\mu$s- would only allow for distances in the 3-4nm range, with limited sensitivity (El Mkami *et al*, 2014), however, deuteration of the solvent can significantly increase the $T_M$ in the 5-6$\mu$s range (El Mkami *et al*, 2014). This was demonstrated by Banham *et al*, 2006, in the study of a von Willebrand factor  domain, where deuteration of the solvent matrix (60% $D_2O$, 40% deuterated glycerol), doubled the $T_M$, compared to the protonated sample, extending the dipolar evolution time to 6$\mu$s, allowing distances in the 6.8nm range.

It is quoted throughout the literature that the upper distance limit for measuring using EPR spectroscopy is 8nm (Gunnar Jeschke, 2012). However, full deuteration of not only the sample, but the protein itself (if expressed in deuterated conditions) has the potential to increase $T_M$s, therefore allowing access to longer distances. This was demonstrated beautifully by Bowman *et al*, 2014 in their work on the histone chaperone proteins Vps75 and NapI.

Using entirely deuterated samples, a distance of 10.2nm was obtained, a record for distances obtained using 4-pulse DEER. As stated above, DEER measurements are

typically taken at 50K, as at this temperature, $T_M$s are increased. It is generally accepted that at higher temperatures, a decreased $T_M$ limits the distance range that can be measured, however entire deuteration of the protein could circumvent this. As demonstrated by Ward *et al*., 2010, the $T_M$ of an entirely deuterated protein at 100K was comparable to a non-deuterated protein at 50K, highlighting the potential to measure at higher temperatures which previously were not possible.

### 1.4.3 DeerAnalysis

The initial time trace (Figure 1.6D) represents both inter and intra molecular contributions as a function of time. The desired intra-molecular signal is obtained following division of the original time trace (D) by the intermolecular contribution. Fourier transformation of the intra molecular interaction trace generates a Pake pattern (as reviewed in Jeschke, 2012).

Small distortions in the DEER signal such as noise or some orientation selection, as well as the measurement of a distribution of distances, will make reading distances from the Pake spectrum difficult. Tikhonov regularisation of the intramolecular time trace is carried out to yield distance distributions, using the software package DeerAnalysis2015 (G. Jeschke *et al*, 2006). DeerAnalysis will take the experimental DEER data, background fit and then use Tikhonov regularisation, or models such as Gaussian lineshapes, to extract distances and distance distributions.

## 1.5 Site-Directed Spin Labelling (SDSL) – Cysteine Substitution Mutagenesis

As stated before, EPR spectroscopy specifically detects unpaired electrons within a system. These unpaired electrons, or radicals, can be naturally occurring, or are more commonly specifically introduced. Consequently, this allows one to look at many different proteins, regardless of cofactors, precisely where desired within the protein structure.

When studying biological complexes, radicals are specifically introduced at specific sites of interest within recombinant proteins, in a process known as site-directed spin labelling (SDSL).

### 1.5.1 Cysteine Chemistry

The chemical versatility of the thiol group of a cysteine residue is commonly exploited for a wide range of chemical transformations of proteins. Illustrated in Figure 1.7 is the chemical structure of cysteine.

Cysteine is usually the most powerful nucleophile in a protein, and, as a result, is frequently the easiest to selectively modify with a variety of reagents, including the site specific incorporation of spectroscopic probes, allowing one to obtain precise structural detail on proteins of interest.

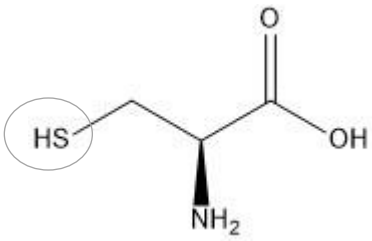



**Figure 1.7: Chemical structure of the amino acid cysteine.** Visualised using ChemDraw™. The highly reactive thiol group of the cysteine residue is circled.

Native cysteine residues can be used as a chemical handle for the incorporation of spin labels, or, alternatively, as demonstrated in pioneering work by Hubbell *et al*, 1994, reactive native cysteine residues can be replaced by a non-reactive amino acid, and further cysteine residues introduced at sites of interest, creating a specific labelling site.

### 1.5.2 Nitroxide-Radical Spin Labels

The first attempt to introduce a paramagnetic reporter group into an otherwise EPR silent macromolecule, was carried out by Stone *et al*, 1965, where a nitroxide radical was introduced into Bovine Serum Albumin (BSA), in order to obtain conformational information about the biomolecule *via* EPR spectra.

Since then, the technique has evolved, and is now used to gain insight into dynamics (Abdullin *et al*, 2014), orientation selection (Lovett *et al*, 2009), conformational transitions and protein folding in real time (Steinhoff *et al*, 1994), whilst not being limited to the size of the biological complex under study (Hubbell *et al*, 1996).

The process of SDSL incorporates a stable radical into the protein structure. Many commercially available spin labels contain thiol-reactive functionalities including methanethiosulfonate, maleimide, and iodoacetamide, which all form a covalent bond with the thiol of cysteine residues, however, the most commonly used spin label is MTSSL (1-Oxyl-2,2,5,5-tetramethylpyrroline-3- methyl) Methanethiosulfonate), the chemical structure for which is shown in Figure 1.8.

MTSSL contains a nitroxide radical, protected by bulky methyl groups, which function to sterically hinder the radical, preventing collisions, and therefore limiting its reactivity (Brown *et al*, 2002). Unlike the vast majority of free radicals that are highly unstable, the nitroxide radical is stable under a variety of physicochemical conditions, including different solvents, pH and temperatures (Schreier *et al*, 2012), which proves desirable when working with biological systems.

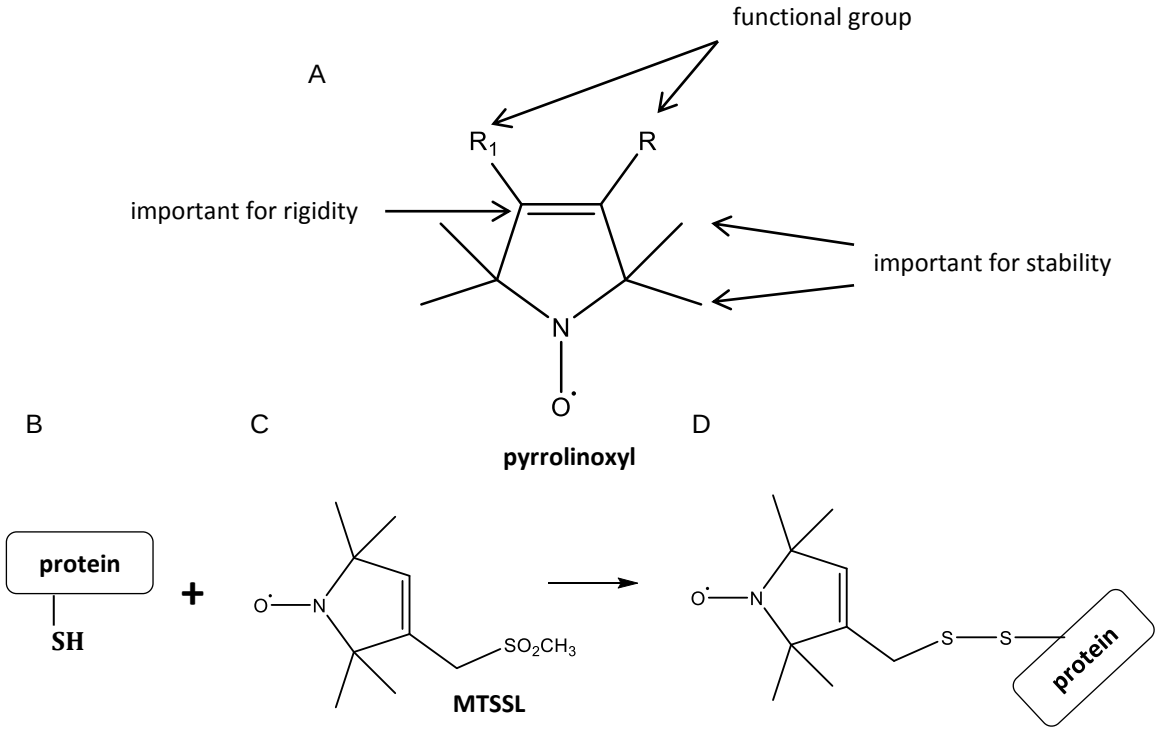



**Figure 1.8: Structural composition of Pyrrolinoxyls, and SDSL of Cysteine Residues with MTSSL.** (A) General structure of pyrrolinoxyls, highlighting key features which make them amenable, and therefore more commonly used for protein labelling. (B) Schematic representation of any given protein, with the thiol group (-SH) of a given cysteine residue (be it naturally occurring, or incorporated *via* site directed mutagenesis) shown. (C) Chemical structure of the commercially available MTSSL. (D) Illustrates the MTSSL covalently bound to the thiol group of the cysteine residue, forming a disulphide bond (Visualised using ChemDraw™).

MTSSL is highly selective in its reactivity for cysteine, and undergoes covalent attachment to the thiol group of the incorporated cysteine residue, forming a disulphide bond, and sidechain designated $R_1$ (Hubbell *et al*, 1996). Due to its notably smaller molecular weight compared to other commercially available spin labels, and the degree of flexibility it exhibits around single bonds, the MTSSL should not interfere with the native fold of the protein (Longhi *et al*, 2011), therefore resulting in minimal structural and functional perturbation.

## 1.6 SDSL – Incorporation of Unnatural Amino Acids (UAAs)

SDSL using cysteine substitution mutagenesis proves difficult when the protein of interest is cysteine rich and there is a driving force to be able to tag or label proteins in cell using unique chemistries. Both these aspects can be tackled by the site-specific incorporation of non-canonical UAAs with novel functional groups into peptides and recombinant proteins, with a subset containing chemically reactive functional groups for the selective incorporation of biophysical probes (see Lang & Chin, 2014, and references therein).

### 1.6.1 Solid Phase Peptide Synthesis (Merrifield Coupling) and Semi- Synthesis

Several methods have been established for UAA incorporation in both peptides and recombinant proteins. The spin label amino acid TOAC ( 2,2,6,6-tetramethyl-N-oxyl-4-amino-4-carboxylic acid) was the first spin label probe incorporated into the peptide backbone by means of a peptide bond (as reviewed in Schreier *et al*, 2012), allowing one to study backbone dynamics and peptide secondary structure. However, due to its rigidity, TOAC has a limited range of backbone dihedral angles, resulting in a significant distortion of the secondary structure of proteins. Consequently it currently can only be incorporated into small peptides and proteins by total solid phase synthesis (Fielding *et al*, 2014).

Work carried out by Becker *et al*, (2005) has tried to overcome such obstacles, by combining SPPS with recombinant techniques, providing the tool for introduction of UAAs at specific sites of interest in large proteins, and even membrane proteins. Using SPPS, paramagnetic UAAs can be incorporated into short peptides *via* protected lysine residues. Using chemical ligation, the peptide can then be incorporated into recombinant

proteins of interest, which in this case is the kinase, cRaf1. The combination of chemically synthesised peptides with recombinant polypeptides greatly increases the versatility and applicability of chemical synthesis of labelled proteins.

### 1.6.2 Expanding the Genetic Lexicon

Continuous Exchange Cell Free Systems allow incorporation of unnatural amino acids at multiple sites within the same protein. Using S30 extracts is an effective tool for the complete cell free synthesis of recombinant proteins (Loscha *et al*, 2012) (Kanda *et al*, 2000) (Cornish *et al*, 1994). S30 extracts are named as such, due to their sedimentation rate by ultracentrifugation (30,000 x g) (Meyers, 1995). The expense of the systems, the degradation of essential components, and the comparatively poor protein yields associated, mean that methods for UAA incorporation in recombinant proteins are continuously evolving.

Pioneering work by Peter Schultz (Scripps Institute) and Jason Chin (Medical Research Council Centre, Cambridge) aims at expanding the genetic lexicon by engineering the translational machinery of cells, for the *in vivo* incorporation of novel amino acids (Young & Schultz, 2010) (Chin, 2011).

#### 1.6.2.1 Reprogramming Translation Using Orthogonal Ribosomes

The engine of translation and where mRNA is ultimately decoded into protein is the ribosome. Engineering orthogonal ribosomes, uncoupled from the requirement to synthesise the cellular proteome, results in a non-essential cellular ribosome with the potential for the incorporation of specific UAAs, in response to unique codons. To exemplify this approach, Neumann *et al*, 2010, evolved an orthogonal ribosome, Ribo-Q, that was able to efficiently decode a series of quadruplet codons using extended tRNA anti-codons. These codons were not read by endogenous ribosomes, and so corresponding tRNAs can be loaded with new amino acids, and these incorporated into proteins in response to the quadruplet codon at the orthogonal ribosome.

#### 1.6.2.2 Orthogonal Amino-acyl tRNA Synthetase/ tRNA Pairs

Whilst there are no blank codons in the genetic code, Schultz took advantage of the amber stop codon, and its ability to be decoded, using amber suppressor tRNAs, for the incorporation of UAAs. Amber suppression provides a codon TAG, that can be used as an initial insertion signal for UAA incorporation into recombinant proteins in *Escherichia coli (E. coli).* By employing tRNA synthetase and tRNA pairs from other organisms, UAAs can be incorporated in response to the amber stop codon TAG, as long as no cross talk exists between the orthogonal tRNA synthetase/tRNA and exogenous pairs.

#### 1.6.2.3 Incorporation of the UAA p-acetyl Phenylalanine (pAcPhe) in response to the amber Stop Codon

The chemical structure of the UAA pAcPhe is shown in Figure 1.9. This UAA contains the chemically versatile keto functional group that is not present in any of the common amino acids, and readily reacts with hydroxylamines in aqueous solution, due to the accessibility of the ketone functionality to the nucleophile.

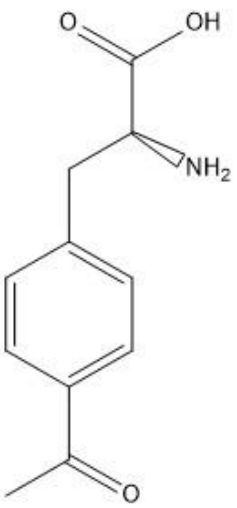



**Figure 1.9: Chemical Structure of the unnatural amino acid p-acetyl-phenylalanine (pAcPhe).** Visualised using ChemDraw™

Work carried out by Fleissner *et al,* 2009 incorporated the pAcPhe into T4 lysozyme, forming a keto linked spin label when reacted with hydroxylamine-nitroxide. In this

strategy, a non-native suppressor tRNA and aminoacyl tRNA synthetase pair is used to incorporate the UAA into the protein site specifically, in response to the unique amber stop codon TAG. The conditions of such reactions are perhaps not favourable for most protein systems (pH 4, 37$^{o}$C). Recent modification of protocols (Hahn *et al*, 2014) has shown that the pAcPhe can be labelled at physiological pH, which is more amenable to the study of bio-molecules, however now the caveat is that organic solvents are necessary.

## 1.7 Simulation of DEER Data

Having a fast and reliable approach for the prediction of spin label conformations in bio-macromolecular complexes facilitates the application of SDSL-EPR methods in the study of protein structure and function. In order to obtain global structural elements of a protein from DEER, one must first have an understanding of the environment in which the spin-label resides, so as to obtain distance distributions which can be critically interpreted.

Comparison of the simulations with the experimental DEER distance distribution can be used to validate structures in solution, or indeed differentiate between different conformational states of a protein. If two or more potential models for a particular structure exist, comparison of the experimental data obtained with the *in silico* can often help decide which model is in in best agreement.

Modelling programmes such as Multiscale Modelling of Macromolecules (MMM) (Polyhach *et al*, 2011) and MtsslWizard (Hagelueken *et al*, 2012) can be used to generate a model distance distribution between spin labels.

### 1.7.1 Multiscale Modeling of Macromolecules (MMM)

MMM is a freely available tool for predicting spin label orientations and the resultant DEER, and can be used within Matlab. MMM is based on a rotamer library of computationally-calculated likely conformations of the MTSSL in relation to a protein of interest, which then simulates the resulting DEER. The MMM simulation is calculated at 175K – the water/glass transition temperature.

### 1.7.2 Mtssl Wizard

The wizard operates as a plugin for the PyMOL molecular graphics system and works by estimating distances between spin labels on proteins quickly, and with user configurable options, searching for MTSSL conformations that do not clash with a static model of the protein of interest.

These particular programmes work by modelling the MTSSL onto user-specified positions with the protein structure, therefore allowing the spin label to sample different conformations in relation to the protein.

## 1.8 Advancing the Field of Structural Biology

Established biophysical techniques applicable for the study of biological systems, such as high resolution Nuclear Magnetic Resonance (NMR) Spectroscopy and X-Ray Crystallography (XRC), have contributed significantly to advancement in the field of structural biology (as reviewed by Maslennikov & Choe, 2013, Zheng *et al*, 2014, Malito *et al*, 2015), and have made it possible to visualise individual biological molecules, as well as their bio-macromolecular complexes at an atomic level.

Fluorescence based methods such as FRET, are also very powerful methods for exploring long range structure, and can be used to obtain nanometre length distance restraints in biological systems. FRET is a technique, that can be applied at the single molecule level providing real time dynamics over several time scales (Reginsson & Schiemann, 2011), however, it is not without its caveats. There is a lack of precision when it comes to quantifying distances and distance distributions, and the labels used for FRET are significantly bigger than those used for EPR, and may affect the structural integrity and conformation of the protein complexes.

Employing EPR spectroscopy in place of, or indeed in collaboration with such techniques, allows structural resolution at the level of the backbone fold, and allows one to monitor changes in equilibria, as well as check conflicting interpretations.

### 1.8.1 EPR Spectroscopy in the Study of Large and Flexible Soluble Proteins

Using EPR as a tool for measuring nm length distances in proteins, and complexes thereof, is a useful tool in the study of protein structure, and protein-protein interactions, and how such interactions contribute to overall function**.** The application of EPR

spectroscopy has resulted in the structures of many large and complex biologically significant proteins being elucidated.

The combined use of NMR and EPR proved to be a powerful approach in the study of the cytoskeletal complex, talin and vinculin, due to the large nature of both proteins involved in the complex, coupled with the high degree of flexibility exhibited in the extended talin rod (Gingras *et al*, 2006). These variables would deem this complex unsuitable to be studied by NMR alone, and so in collaboration with EPR, one can gain insight into such interactions, which otherwise would be inaccessible.

### 1.8.2 EPR in the Study of Membrane Proteins

Membrane proteins account for ~25% of all proteins within the cell. Structural determination of membrane proteins by X-Ray diffraction (Bai, Yan *et al*, 2015), solution NMR (Mineev *et al*, 2015) and more recently single-particle electron cryo-microscopy (cryo-EM) (Schmidt-Krey *et al*, 2011) has resulted in more than 2,000 structures deposited (Rodriguez *et al*, 2014).

However in the majority of cases, determination by such means does not reveal the structure of the protein in a membrane, but rather in a crystal lattice or in a micelle/bicelle. Using EPR to obtain some simple distance restraints, it is possible to check for potential crystallisation artefacts.

Work carried out by Pliotas *et al*, 2012 looks at the different conformational states of the membrane protein MscS, a mechanosensitive ion channel, composed of 7 monomers made up of 21 transmembrane helices. This particular multi-domain protein has been crystalized, although the crystal repeatedly challenged, with three mutually incompatible models hypothesised, which proves detrimental to the field.

Using SDSL in collaboration with XRC, Pliotas *et al*, 2012 demonstrated the wider utility of EPR, and showed it to be a reliable technique when studying such a complex system. Similarly, Endeward *et al*, 2009 and Padmavathi & Steinhoff, 2008 used SDSL to study the tetrameric potassium ion channel KcsA, and the closed channel of Colicin A, respectively.

Using EPR, it was possible to gain information on side chain mobility, solvent accessibility, and the orientation of specific regions, and follow dramatic

conformational changes (as a function of time), which had been disputed in the literature.

### 1.8.3 EPR as a Stand Alone Technique

Although using EPR in collaboration with other higher resolution structural techniques has proven successful in elucidating the structures of previously disputed complexes, EPR stands alone as a powerful structural technique. Many transporter and importer structures have been solved using EPR and SDSL (see Joseph *et al*, 2011, and Borbat *et al*, 2007) as many of these systems are important pharmaceutical targets.

Hilger *et al*, 2009, used *ab initio* protein modelling to build a three dimensional model of the $Na^+$/Proline transporter PutP of *E. coli* from scratch, solving the backbone structure by obtaining distances between spin labelled mutants at 2Å resolution, allowing insight into the structural and functional relationship. The crystal structure was later solved, verifying the original model.

## 1.9 Beyond *in vitro* EPR, and towards physiological conditions

The next level in elucidating protein structure and ultimately function, requires one to consider the natural cellular environment in which the protein of interest resides.

The use of DEER for obtaining nm distance restraints is typically done in frozen solutions at cryogenic temperatures, in order to reduce the rapid electron spin relaxation rates. Also, the matrix in which the proteins are present must be carefully considered, in order to stop tumbling of the protein averaging the dipolar interaction to zero. Measurements cannot be taken in an aqueous matrix, as the high dielectric constant of liquid water absorbs microwaves.

However, using EPR spectroscopy at ambient temperatures would allow for a more direct comparison of the experimental environment with how proteins interact in physiological environments. The process of SDSL and the pulsed EPR techniques themselves have evolved considerably from when the techniques were in their infancy, however, comparatively, the labelling process, and evolution of the labels themselves has lagged somewhat behind.

### 1.9.1 Trityl and Spiro Radical Spin Labels

In order to obtain measurements in liquid solution without the need for cryogenic temperatures, a number of groups have successfully incorporated trityl spin labels into proteins such as T4 lyzosyme (Yang *et al*, 2012) and into nucleic acids (Shevelev *et al*, 2014). Due to the relatively long relaxation times exhibited by these labels, using pulsed methods, namely Double Quantum Coherence (DQC), these bio-macromolecules could be measured in solution. Shown in Figure 1.10 is an example of a cysteine specific trityl radical such as that used in the study of T4 lyzosyme.

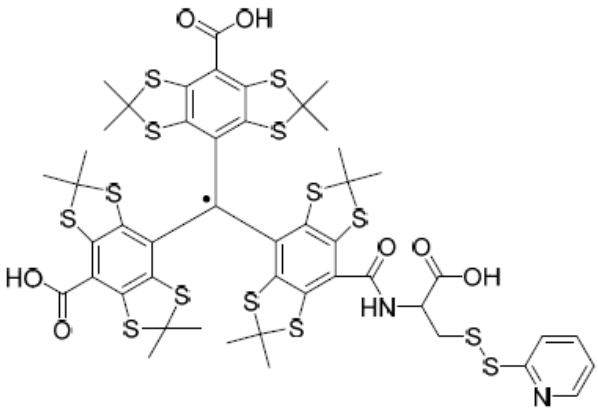



**Figure 1.10: Chemical structure of a cysteine-specific trityl radical.**

Using trityl radicals in these model systems, distances of up to approximately 4.6nm at 310K (37$^{o}$C) have been measured (Shevelev *et al*, 2014), demonstrating the potential, with significant improvement to the overall size and hydrophobicity of the label, for the future application of trityls as spin labels, at temperatures comparable to physiological.

Studies by Eaton *et al*, 2015, have shown that the incorporation of spiro-cyclohexyl radical labels using SDSL to sites on T4 lysozyme resulted in inter-spin distances of up to ~4nm being measured by DEER at 160K, however, when in a trehalose matrix, an extended $T_M$ allowed a distance of ~3.2nm to be measured at room temperature (r.t).

These distances could not be measured using the conventional MTSSL label, and so evolution of labelling and labels themselves, will allow long range distances to be obtained at temperatures and in environments which are closer to the physiological environment in which the proteins natively reside.

### 1.9.2 *in vivo* EPR Studies

Ultimately, in order to truly elucidate protein structure and function, one must be able to mimic, as closely as possible the natural cellular environment. The in-cell environment is characterised by thousands of cellular components that can interfere with the biomolecule under study. Due to the sensitivity of DEER for paramagnetic systems, background interference is negligible and so DEER could be adapted for *in vivo* structural studies.

Schmidt *et al*, (2014) have successfully encoded an UAA in *E. coli*, in response to the amber stop codon, TAG. Direct incorporation of the nitroxide radical with the 2,2,5,5-tetramethyl-pyrrolin-1-oxyl moiety enables intracellular biosynthesis of spin labelled proteins *in vivo*. The chemical structure of the radical is shown in Figure 1.11.

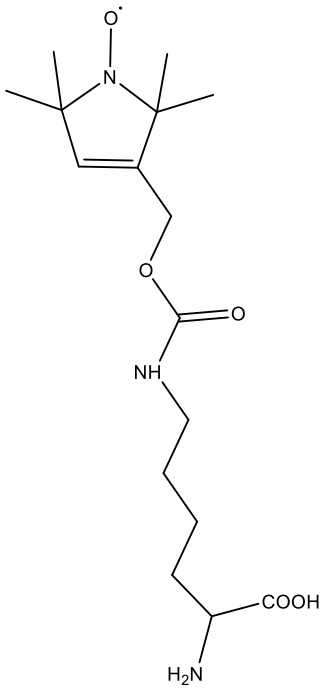



**Figure 1.11: Chemical structure of the UAA containing the 2,2,5,5-tetramethyl-pyrrolin-1-oxyl moiety.** (Visualised using Chemdraw ®) (Schmidt 2014).

The higher stability of the nitroxide radical in *E. coli* than other intracellular environments enabled cells expressing thioredoxin with two of the nitroxide containing UAAs to be frozen and a DEER measurement taken. Furthermore, using specific *E. coli* cell lines, such as Origami, which are deficient in specific reducing agents, it is possible to extend the lifetime of the nitroxide radical in cell (Dunkel *et al*, 2015).

In human cells, nitroxide radicals are reduced and converted into a diamagnetic hydroxylamine with a half-life of approximately 50min (Martorana *et al*, 2014) however, it is also possible to adapt the nitroxide spin label itself to one that is more resistant to reduction. Studies by Jagtap *et al*, 2015 and Kinoshita *et al*, 2010 have found that sterically shielded spin labels, such as tetraethyl substituted nitroxides (pyrrolidine and piperidine derivatives) are more resistant to reduction by ascorbic acid, and in the presence of cellular extracts and oocytes.

Furthermore, reduction resistant $Gd^{3+}$ chelates have been proposed as a potential alternative for in-cell DEER distance measurements in human cells (Martorana, 2014). Building on adapting nitroxide spin labelling *in vivo*, and labelling with $Gd^{3+}$ chelates will broaden the application of SDSL EPR, allowing one to further characterise proteins which before were not suited to study by DEER.

## 1.10 Application of EPR in this Thesis

Following thorough exploration of the literature, it is clear than EPR has advanced from a technique that was once used by few, in complement with high resolution structural biology techniques, to a powerful technique in its own right. Using a variety of methods of SDSL, this thesis aims to explore long range structure in an assortment of recombinant and native proteins, and complexes thereof.

Using various expression systems, different methods of SDSL will be explored in the study of various protein complexes, in hopes of confirming proposed structural hypotheses. This will include the SDSL of highly flexible, cysteine rich recombinant proteins, and the manipulation of native plasma proteins for SDSL purposes.

New methods for the specific labelling of cysteine rich proteins will be explored. The ability of *E. coli* to read through the amber stop codon will be exploited for the incorporation of UAAs for SDSL, and novel spin labels, specific for labelling cysteine pairs tested in several model systems. Furthermore, native paramagnetic centres of recombinant proteins will be explored as potential labelling sites.

The work described in this thesis covers experimental design, molecular biology and cloning, protein expression and purification, as well as functional characterisation, and finally EPR distance measurements, data analysis and interpretation. Challenges and pitfalls will also be addressed.

# Chapter 2: Materials & Methods

All chemicals, unless otherwise stated, were purchased from Sigma Aldrich (Gillingham, UK). All buffers, stock solutions and media recipes are included in Appendix A. Primers were ordered from Sigma Aldrich and stored at -20$^{o}$C as 100$\mu$M stocks, and a 10$\mu$M working stock concentration of each primer used for each PCR reaction. All primer sequences are included in Appendix A.

## 2.1 Mutagenesis, Purification, and Manipulation of DNA for Transformation into Chemically Competent *E. coli*

### 2.1.1 Site – Directed Mutagenesis (SDM)

*In vitro* SDM is a commonly used technique in molecular biology for generating point mutations in DNA coding genes of interest. It is possible to alter the primary amino acid sequence by replacing, inserting or deleting specific residues. All mutagenesis was carried out following the protocols set out in the Quikchange© site-directed mutagenesis manual (Stratagene, California, USA). Reactions were carried out (Eppendorf Mastercycler personal, Eppendorf, Hamburg, Germany) according to the pipetting scheme and PCR programme detailed in Tables 2.1 and 2.2, using the mutagenic primers, and thermo-stable PfuTurbo® DNA polymerase as it exhibits high fidelity 3'-5' proof reading capability, thus minimizing the number of PCR induced mutations.

Each PCR reaction mixture (Table 2.1) underwent the following cycle for mutagenesis.

**Table 2.1: PCR Reaction Mixture for SDM**

| | |
|---|---|
| 5.0$\mu$L | Pfu-buffer (10x) |
| 1.0$\mu$L | ddNTPs |
| 1.0$\mu$L | Primer 5' |
| 1.0$\mu$L | Primer 3' |
| 3.0$\mu$L | DMSO (dimethyl sulfoxide) |
| 1.0$\mu$L | Template (125ng) |
| 1.0$\mu$L | PfuTurbo® |
| 37.0$\mu$L | dd$H_2O$* ( to 50$\mu$L) |
| *double distilled $H_2O$ | |

**Table 2.2: Programme used for Site-Directed Mutagenesis PCR**

| Step | Temperature ($^{o}$C) | Time (s) | Cycles |
|---|---|---|---|
| INITIATION | 95 | 30 | 1 |
| DENATURATION | 95 | 30 | 12-18* |
| ANNEALING | 55 | 30 | |
| ELONGATION | 68 | 60/kb plasmid length | |
| HOLD | 4 | ∞ | |

* = 12 for point mutation, 16 for single amino acid change, 18 for multiple amino acid deletion/insertion

Immediately following the reaction, the PCR product was treated with 2$\mu$L of the restriction endonuclease *DpnI* which promotes the specific digestion of methylated template (parental) DNA (target sequence: 5'-GM6ATC-3'). The PCR product was incubated with the *DpnI* enzyme at 37$^{o}$C for 2hr, before 3$\mu$L of the digested mutated DNA-plasmid was transformed into XL1 Blue or Top 10 (homemade cells cultured from Life Technologies stocks – see Appendix B) chemically competent *E. coli* cells, as described in 2.1.2.

#### 2.1.1.1 Designing Mutants Suitable for EPR

Generally, to be considered a suitable candidate for SDSL, mutations are made in non-conserved, hydrophilic, surface exposed regions, usually alpha helices, to facilitate efficient incorporation of spin labels into proteins of interest, whilst maintaining stability.

When making mutations, one must take into consideration which amino acids are being substituted, and opt for substitutions which are conservative, with amino acids possessing similar characteristics. For example, serine, glutamate and asparagine are all uncharged residues, as is cysteine, which makes substitution to cysteine for SDSL less drastic.

#### 2.1.1.2 Primer Design

Following guidelines set out in the Stratagene Quikchange © SDM manual, primers were designed to facilitate directional cloning. Generally, both forward and reverse mutagenic primers were designed to contain the desired mutation, and anneal complementarily to the plasmid. Primers should be between 25 and 45 bases in length,

with a melting temperature ($T_{melt}$), higher than, or equal to 78°C. The $T_{melt}$ for each primer was determined using equation 2.1:

$$T_{melt} = 81.5 + 0.41\ (\%GC) - 675/N - \%\ \text{mismatch}\ _{(\text{EQUATION 2.1})},$$

where N is the primer length in bases and % GC and % mismatch are whole numbers. The mutation was designed to be in the centre of the mutagenic primer, flanked on either side by 10 – 15 bases homologous to the plasmid. The overall GC content of the primer did not exceed 40%, and the primer terminated in one or more C or G bases.

### 2.1.2 Transformation of Plasmid DNA into Chemically Competent *E. coli* Cells

For each transformation reaction, 50$\mu$L of XL1 blue or Top 10 (for high efficiency cloning) chemically competent cells were thawed on ice before being transferred to chilled, round bottomed 15mL Falcon© tubes. The cells were mixed with 3$\mu$L of *DpnI* digested DNA following the mutagenesis PCR. Following heat shock at 42°C in a circulating water-bath for 30s (Top 10) or 45s (XL1 Blue), cells were cooled on ice for 2min, before 300$\mu$L of pre-warmed (37°C) super optimal broth (SOB) with catabolite repression (SOC, Life Technologies) was added to each transformation reaction to allow cell recovery. Cells were then incubated with shaking at 200rpm at 37°C.

This outgrowth step allows the bacteria time to generate the antibiotic resistance proteins encoded on the plasmid backbone so that they are able to grow once plated on the antibiotic containing agar plates. Following outgrowth, cells were spun down at 13,000rpm, and pelleted cells re-suspended in 200$\mu$L SOC, before being plated onto LB or LSLB (Zeocin™ only) agar plates containing the appropriate antibiotics for selection (Table 2.3) (50 and 150$\mu$L).

**Table 2.3: Final concentration of antibiotics for selection following *E. coli* transformation**

| | |
|---|---|
| Kanamycin (Kan.) | 100$\mu$g/mL |
| Ampicillin (Amp.) | 100$\mu$g/mL |
| Tetracycline (Tet.) | 25$\mu$g/mL |
| Chloramphenicol (Cm.) | 100$\mu$g/mL |
| Zeocin$^{TM}$ (Zeo.) | 25$\mu$g/mL |
| Amp. & Kan. | 50 & 50$\mu$g/mL |
| Amp. & Cm. | 50 & 50$\mu$g/mL |
| Kan. & Cm. | 50 & 50$\mu$g/mL |
| Kan. & Tet. | 50 & 25$\mu$g/mL |

### 2.1.3 Isolation and Purification of Plasmid DNA

To prepare plasmid DNA from bacterial cultures, a QIAprep Spin MiniPrep kit was used for the preparation and purification of plasmid DNA on a small-scale (up to 20$\mu$g of high copy plasmid DNA from 5mL of *E. coli* culture grown overnight in LB broth) and the HiSpeed Plasmid Maxi (QIAgen, Hilden, Germany) kit for large-scale isolation (up to 500$\mu$g from up to 500mL of culture). The QIAquick Gel extraction kit (QIAgen) was used to recover and purify DNA from agarose gels. All buffers used in 2.1.3.1, 2.1.3.2, and 2.1.3.3 were provided with, and used according to the manufacturer's instructions (Qiagen, Hilden, Germany). Purified DNA was stored at -20°C.

#### 2.1.3.1 Mini-Prep

A single colony from a freshly streaked transformation plate was used to inoculate 5mL of LB/LSLB medium containing the appropriate antibiotic for selection. The cultures were incubated with shaking at 200rpm at 37$^{o}$C for 16hr before cells were harvested by centrifugation at 4,500rpm, for 10min at 4$^{o}$C. Pelleted bacterial cells were re-suspended in 250$\mu$L of Buffer P1, followed by the addition of 250$\mu$L of P2 to initiate cell lysis.

Buffer N3 was then added at a volume of 350$\mu$L to neutralise the lysis reaction, and the reaction mixture was spun at 13,000rpm at 4$^{o}$C for 10min. The resulting supernatant was then loaded onto the QIAprep® spin column. The column was spun at 13,000rpm for 1min, discarding the flow-through. The column was washed with 500$\mu$L of PB Buffer and 750$\mu$L PR Buffer, spinning between washes to remove residual buffer. DNA was eluted by the addition of 50$\mu$L of EB buffer to the column, followed by spinning at 13,000rpm for 1min, and the plasmid DNA concentration checked by UV spectrometry.

#### 2.1.3.2 Maxi-Prep

A single colony grown on selective media was used to inoculate 5mL of LB/LSLB containing the same antibiotic. The culture was incubated by shaking at 200rpm at 37$^{o}$C for 9hr, before being used to seed a fresh 250mL of LB/LSLB, again with the same antibiotic concentrations. This culture was incubated by shaking for 16hr at 37$^{o}$C before being harvested by centrifugation at 4,500rpm for 10min at 4$^{o}$C.

Pelleted cells were re-suspended in 10mL of Buffer P1 before the addition of 10mL of Buffer P2. Cells were mixed gently by inverting, and incubated at r.t for 5min. To the lysed cells, 10mL of chilled Buffer P3 was added, and the mixture was allowed to separate into two distinct layers. The aqueous layer was applied to a QIAgen-HiSpeed Maxi tip, equilibrated in 10mL QBT Buffer, and allowed to enter the resin by gravity flow. The tip was washed with 60mL of Buffer QC, and DNA eluted from the tip by applying 15mL of Buffer QF by syringe, and collecting the DNA in the flow through. DNA was precipitated by the addition of 10.5mL of r.t isopropanol (IPA) and incubated at r.t for 5min after mixing. The mix was applied to the QIAprecipitator. Using a syringe, 2mL of 70% ethanol (EtOH) was added to the precipitator.

The DNA (contained within the precipitator) was eluted by adding 1mL of Elution Buffer (EB) to the precipitator with a syringe. The concentration of the DNA in the flow-through was determined by UV spectrometry.

#### 2.1.3.3 Gel Extraction

For isolation of plasmid DNA from agarose gels following DNA agarose gel electrophoresis (described in 2.2.2), all extractions were carried out using the QIAquick Gel Extraction Kit (QIAgen). The desired DNA fragment was excised from the agarose

gel, with a clean, sharp scalpel, taking care to avoid the inclusion of excess agarose. The gel slices were then weighed, and 300$\mu$L of Buffer QG added per 100mg of gel slice, before incubation at 50$^{o}$C for 10min, or until gel slices dissolved. One gel volume (*e.g.* 100$\mu$L to 100mg of gel) of r.t IPA was added. The mixed sample was placed in the QIAquick spin column and the column placed in a 2mL collection tube. The sample was spun at 13,000rpm for 1min and flow-through discarded. Buffer QG (500$\mu$L) was then added to the column, followed by spinning at 13,000rpm for 1min, discarding the flow-through.

The bound DNA was washed by applying 750$\mu$L of PE Buffer to the column, spinning at 13,000rpm for 1min, discarding the flow-through. The column was spun once more to remove residual buffer, before the column was placed in a clean 2mL collection tube. DNA was eluted by the addition of 30$\mu$L of EB Buffer to the centre of the column membrane. The column was left at r.t for 3min, before being spun at 13,000rpm for 1min to elute the now unbound DNA. Concentrations of recovered DNA were determined by UV spectroscopy.

### 2.1.4 Sequencing of Plasmid DNA

In order to confirm whether mutagenesis or cloning have been successful, the plasmid DNA was screened using the automated service GenePool (University of Edinburgh). DNA sequencing for all fH constructs cloned into the pPICZα vector used primers targeting the alpha secretion factor, the sequences for which are shown in Table 2.4.

**Table 2.4: Sequencing primers for pPICZα**

| Primer | Sequence |
|---|---|
| 5’α-factor | 5’-TACTATTGCCAGCATTGCTGC-3’ |
| 3’α-factor | 5’-GCAAATGGCATTCTGACATCC-3’ |

DNA sequencing for all constructs cloned into vectors carrying the T7 promoter and terminator (common in IPTG inducible proteins) was carried out at Genepool, University of Edinburgh. Primers used for targeting the T7 promoter and terminator are shown in Table 2.5.

**Table 2.5: Sequencing primers for T7 promotor and terminator**

| Primer | Sequence |
|---|---|
| 5'α-factor | 5'-TACTATTGCCAGCATTGCTGC-3' |
| 3'α-factor | 5'-GCAAATGGCATTCTGACATCC-3' |

For each sequencing reaction, 250-300ng of template DNA was added to 1$\mu$L of forward or reverse primer, and 4$\mu$L of prism BigDye Terminator Mix (Applied Biosystems, California, USA), and the reaction mixture made up to 20$\mu$L with the addition of $ddH_2O$. The BigDye Reagent is necessary for the dideoxy method for sequencing as developed by Sanger *et al*, 1977, which utilizes the ability of DNA-polymerases to incorporate nucleotide analogues such as 2',3'-dideoxynucleotides (ddNTPs).

When a dideoxynucleotide is incorporated at the 3'-end of the growing chain, chain elongation is terminated selectively at A, C, G, or T because the ddNTP lacks a 3-hydroxyl group. The BigDye terminator mix contains four different fluorescently labelled ddNTPs, which are added sequentially to the primer through cycled sequencing reactions. The outcome is a mixture of fragments with different ddNTP residues at the 3'-end which can be separated by polyacrylamide gel electrophoresis, where a pattern of bands is obtained from which the sequence can be read off. Table 2.6 outlines the parameters for the PCR reaction for these sequencing reactions.

**Table 2.6: Programme used for Sequencing PCR**

| Step | Temperature ($^oC$) | Time (s) | Cycles |
|---|---|---|---|
| INITIATION | 95 | 30 | 1 |
| DENATURATION | 96 | 30 | 24 |
| ANNEALING | 50 | 30 | 24 |
| ELONGATION | 60 | 240 | 24 |
| HOLD | 4 | ∞ | - |

## 2.2 Mutagenesis, Purification, and Manipulation of DNA for Transformation into Electro-Competent *Pichia pastoris* (*P. pastoris*)

### 2.2.1 Amplification of specific CCPs from the full length codon optimised fH gene using PCR

The gene encoding full length fH, codon optimised for expression in *P. pastoris*, was purchased from GeneArt® (California, USA). This served as the template DNA from which both of the fH fragments used in this study were amplified. Primers were designed to match the first 25 bases of the desired CCP fragment for the 5' annealing, and the last 25 bases of the fragment for 3' annealing. Designing primers would result in amplification of just the desired fragment, either fH1-4 or fH19-20, from the full length gene. The following pipetting scheme and PCR programme were used to facilitate this (Tables 2.7 and 2.8).

**Table 2.7: PCR Reaction Mixture for CCP Amplification**

| | |
|---|---|
| 5.0 $\mu$L | Pfu-buffer (10x) |
| 1.0 $\mu$L | ddNTPs |
| 1.0 $\mu$L | Primer 5' |
| 1.0 $\mu$L | Primer 3' |
| 1.5 $\mu$L | DMSO |
| 1.0 $\mu$L | Template DNA (fH gene) |
| 1.0 $\mu$L | Pfu Turbo® polymerase |
| 38.5 $\mu$L | $ddH_2O$ (final volume 50$\mu$L) |

**Table 2.8: Programme used for fH CCP Amplification PCR**

| Stage | Temperature ($^oC$) | Time (s) | Cycles |
|---|---|---|---|
| INITIATION | 95 | 300 | 1 |
| DENATURATION | 95 | 30 | 35 |
| ANNEALING | $T_{melt}$ -5$^oC$ | 30 | 35 |
| ELONGATION | 68 | 120 (60/kb plasmid) | 35 |
| EXTENSION | 72 | 600 | 1 |
| HOLD | 4 | - | - |

### 2.2.2 DNA Agarose Gel Electrophoresis

Agarose gel electrophoresis is a well-established method for separation of DNA molecules according to their size. Following amplification of the fH fragments by PCR, the resulting PCR product was run on an agarose gel to confirm the size. Upon application of an external electrical field, DNA molecules migrate through the gel due to the nature of the negatively charged nucleic acids. DNA fragments of smaller size migrate readily and more quickly through the agarose gel matrix than those of larger size, and so are resolved further down the gel.

To prepare a 1% gel, 1g of agarose was added to 100mL of TBE Buffer, and the mixture heated in a microwave oven until the agarose was fully dissolved and the mixture molten. The mixture was allowed to cool to 55$^{o}$C, before either 10$\mu$L of EtBr (10mg/mL stock) or 10$\mu$L of SYBR®Safe (Life Technologies, California, USA) (1:10,000 dilution) was added. The gel was cast with combs to create wells, and allowed to set.

Dependent on the size of construct being resolved, either a 100bp or 1kbp DNA ladder (2$\mu$L) was used as a size standard (New England BioLabs (NEB), Massachusetts, USA). Samples were prepared by the addition of sample DNA (2$\mu$L) to (5x) DNA loading buffer (8$\mu$L) (New England BioLabs (NEB), Massachusetts, USA). The gel tank was filled with TBE buffer and samples run at 90V for 30min, or until bands were nicely separated. EtBr and SYBR®Safe are hydrophobic molecules that work by intercalating DNA between the stacked bases which can then be visualised using UV light and a trans-illuminator (Benchtop 3 UV transilluminator, UVP, Cambridge, UK).

### 2.2.3 Cloning of fH Fragments into pPICZαB Vector, Using Restriction Enzymes

Once fragment size was confirmed, SDM was carried out, as detailed in 2.1.1, to incorporate cysteine residues for SDSL purposes, and the restriction sites *PstI* and *XbaI* at the N' and C' termini of each fragment, respectively. Incorporation of these restriction sites allows for ligation of the DNA fragments into the Multiple Cloning Site (MCS) of the pPICZαB vector. The chosen mutations in each fragment were K247C and G1107C for the fH1-4 and fH19-20 fragments, respectively. Following unsuccessful expression tests of the fH1-4 fragment, the fH1-4 fragment was further mutated by the inclusion of a C' terminal hexa-histidine tag.

#### 2.2.3.1 Double Digest

The remainder of the PCR products of the amplified fH fragments were run on an agarose gel, alongside some of the empty pPICZαB vector, as detailed in 2.2.2, and gel extraction carried out, as detailed in 2.1.3, in order to purify the DNA. Once eluted from the gel extraction columns, the DNA was treated with 1$\mu$L each of the restriction enzymes *PstI* and *XbaI*, 1$\mu$L of Bovine Serum Albumin (BSA), and 60$\mu$L of Buffer 3.1 (all NEB). Each fragment, and the empty vector, was incubated with the restriction enzymes for 2hr at 37$^{o}$C before the enzymes were heat inactivated at 80$^{o}$C for 20min.

#### 2.2.3.2 Ligation of Digested Fragments into pPICZαB

Following digestion with the restriction enzymes, 50ng of the cut vector was combined with 150ng of the relevant fragment, 2$\mu$L of 10x ligation buffer (NEB) and the reaction left at r.t for 10min, before 1$\mu$L of T4 DNA ligase (NEB) was added, and the reaction mixture adjusted to a final volume of 20$\mu$L with the addition of $ddH_2O$. The reaction was left at r.t for 3 hr, before 5$\mu$L of the ligation reaction was transformed into Top 10 chemically competent *E. coli* cells, as described in 2.1.2, using Zeocin$^{TM}$ for selection.

#### 2.2.3.3 PCR Colony Screen

In order to confirm whether a ligation reaction has been successful following transformation, it is possible to screen any colonies that form and differentiate between those which contain the plasmid with insert, or those that contain the empty plasmid vector. From fresh ligation transformation plates, several colonies were re-streaked onto relevant agar media with appropriate antibiotic to give more working material. After a 20hr period, a scraping from each colony restreak was used to inoculate a separate 20$\mu$L of EB Buffer (QIA Prep MiniPrep Kit), until the reaction mixture was cloudy. The samples were heated at 95$^{o}$C for 5min to lyse the bacteria and solubilize the plasmid DNA. To 5$\mu$L of PCR MasterMix (Promega), 2$\mu$L of both the forward and reverse sequencing primers (Tables 2.4 and 2.5) were added, alongside 2$\mu$L of the lysed bacteria. The programme used for thermal cycling is outlined in Table 2.9.

**Table 2.9: Programme used for PCR Colony Screen**

| Step | Temperature ($^{o}$C) | Time (s) | Cycles |
|---|---|---|---|
| INITIATION | 95 | 120 | 1 |
| DENATURATION | 95 | 45 | 39 |
| ANNEALING | 55 | 45 | 39 |
| ELONGATION | 72 | 120 | 39 |
| EXTENSION | 72 | 300 | 39 |
| HOLD | 4 | ∞ | - |

Immediately following the PCR reaction, each colony was tested for successful ligation by analysing each PCR product by DNA agarose gel electrophoresis (2.2.2). This procedure distinguishes between those colonies carrying the relevant insert and those which contained empty plasmid vector, which still contains the relevant antibiotic resistance markers. Those clones which appear at the expected size for the fragment successfully ligated into the vector were confirmed by sequencing (2.1.4), before maxi-preps (2.1.3) were carried out to yield more DNA.

### 2.2.4 Preparation of DNA for Transformation into *P. pastoris*

Purified plasmid DNA that had been previously transformed into *E. coli* required linearization before it could be successfully incorporated into *P. pastoris via* homologous recombination. Therefore, 25$\mu$g of plasmid DNA (from maxiprep), was incubated with 5$\mu$L of the restriction enzyme *SacI* (plus 6$\mu$L BSA and 60$\mu$L Buffer I, NEB) which cuts at a specific recognition site in pPICZαB. Consequently, it is vital that the gene of interest does not contain this restriction site. Successful linearization was confirmed by agarose gel electrophoresis (2.2.2) and the enzyme heat-inactivated by incubating at 62$^{o}$C for 30min.

#### 2.2.4.1 Phenol-Chloroform Extraction

To clean the DNA and eliminate any residual enzyme, a phenol-chloroform extraction was performed using a phenol:chloroform:isoamyl alcohol mixture (25:24:1) saturated with 10mM Tris, pH 8.0, 1mM EDTA. Briefly, the DNA was washed with an equal volume of phenol:chloroform before being spun for 1min at 13,000rpm (4$^{o}$C). The aqueous layer was then washed in the same manner. Finally, an equal volume of

chloroform was added to the aqueous layer, the mixture spun, and the aqueous layer taken off. A 10% solution (v/v) of sodium acetate (NaAc) was added (3M, pH 5.2) as well as 2.5mL of ice cold 100% EtOH. The mixture was then kept at -20$^{o}$C overnight before undergoing EtOH precipitation.

#### 2.2.4.2 EtOH Precipitation

Following overnight incubation at -20$^{o}$C, the DNA sample was spun at 13,000rpm for 30min at 4$^{o}$C. The supernatant was discarded and 400$\mu$L of ice cold 70% EtOH added. Further spinning at 13,000rpm for 10min at 4$^{o}$C followed. The supernatant was removed and the pellet allowed to air dry, before being re-suspended in 11$\mu$L of dd$H_2O$. DNA was stored at -20$^{o}$C.

### 2.2.5 Preparation of electro-competent *P. pastoris* cells

From a stab culture (Invitrogen) of *P. pastoris* strain KM71H, YPD agar plates were streaked to obtain single colonies. A single colony was used to inoculate 5mL of YPD medium and this was left shaking overnight at 30$^{o}$C. The overnight culture was then used in volumes of 20,40 and 80$\mu$L to inoculate a fresh 100mL of YPD medium (in 3x1L baffled flasks), and left overnight at 30$^{o}$C until an $OD_{600}$ of between 1.3 and 1.5 was achieved. Three different inoculation volumes were used to ensure there were cells available should one culture overgrow.

Cells were harvested by centrifugation at 1,500 x g for 5min at 4$^{o}$C and the pellet re-suspended in 100mL of ice-cold dd$H_2O$ (sterilized by autoclaving). Spinning was repeated, and cells re-suspended in 50mL of ice-cold dd$H_2O$. Following another spin the cells were re-suspended in 4mL of ice cold IM sorbitol (sterilized by autoclaving) before the final spin and re-suspension in 200$\mu$L of ice-cold 1M sorbitol. Cells were kept on ice for use that day as they cannot be stored.

### 2.2.6 Electro-transformation of *P. pastoris*

Linearised DNA (~30$\mu$g in no more than 10$\mu$L) was added to the wall of a pre-chilled electroporation cuvette (Gene Pulser Cuvettes 0.2cm gap, BioRad, Hemel Hempstead, UK) and 80$\mu$L of the electro-competent *P. pastoris* cells allowed to wash over the DNA. The cuvette was incubated on ice for 5min then cells pulsed at 1.5V for 6s (resistance selection: low range 200Ω, high range 1500Ω, BioRad Genepulser © II,

BioRad). Immediately after, 1mL of ice cold YPDS media was added to the cuvette and the contents transferred into a sterile pre-chilled 15mL Falcon ™ conical tube.

The cells were incubated without shaking for 3hr at 30°C, before being plated in volumes of 150$\mu$L and 250$\mu$L on YPDS agar at varied Zeocin™ concentrations (100, 200, 300, and 500$\mu$g/mL). Plates were incubated for 5 days at 30°C until colonies formed. Colonies found on those plates with higher Zeocin™ concentration contain multiple plasmid inserts and are therefore likely to yield higher levels of protein due to higher expression levels.

## 2.3 Recombinant Protein Expression Using *E. coli* Expression Systems

Expression on all scales (unless otherwise stated) was carried out at 37°C in a shaking incubator (250rpm, New Brunswick Scientific Model G25, Innova 4430 incubator shaker, New Brunswick Scientific, St Albans, UK). For overexpression of recombinant proteins, plasmid DNA was transformed into either DH10β or BL21 (DE3) *E. coli* cells, as described in 2.1.2, using heat shock at 42°C for 45 and 30 s respectively. Otherwise, cells provided as glycerol stocks in Rosetta™2(DE3) cells were plated on LB agar containing the appropriate concentrations of antibiotic in order to isolate single colonies. All cells were harvested by centrifugation with the Sorvall© legend RT centrifuge, using the Sorvall© SH-3000 swinging bucket rotor and SS-34 fixed angle rotor (Thermo Scientific, UK).

### 2.3.1 Expression of RLC/MiniHMM Constructs

Original plasmids for expression of the MiniHMM fragment, human cardiac RLC and cardiac myosin binding protein C, domain C0, were provided by Dr Mark Pfuhl Kings College, London.

The antibiotics used for selection throughout were Amp and Kan at the concentrations outlined in Table 2.3. Following successful transformation of the plasmids containing the MiniHMM construct and the myosin RLC, a single colony was used to inoculate 10mL of LB broth, and the culture grown over the course of one day, before 600$\mu$L was used to inoculate 10mL of fresh LB media. From the new culture, 500$\mu$L was added to each litre of LB broth (for large scale expression), and cultures grown overnight at 30°C with shaking at 200rpm.

Following overnight incubation, the temperature was increased to 37$^{o}$C, and cells grown until an $OD_{600}$ of ~1.00 was achieved. At this point, cells were moved to 4$^{o}$C for 30min, before expression was induced overnight at 18$^{o}$C, with shaking at 200rpm, with the addition of 100mg of IPTG (dissolved in 1mL of $ddH_2O$) per 1L of cell culture. Cells were harvested by centrifugation at 6,000 rpm for 15min at 4$^{o}$C. Supernatants were discarded and cells lysed.

#### 2.3.1.1 Cell Lysis of RLC/MiniHMM Constructs

Following pelleting of cells, 100mL of lysis buffer was added, and cells re-suspended. Cells were kept on ice for 30min, and 10mg of lysozyme added to the mixture, with frequent stirring. To this, 1mL of Triton X-100 (25% stock) was added, giving a final detergent concentration of 0.25%, and cells kept on ice for 15min, again with frequent stirring.

Cells were frozen in liquid nitrogen for 5min before being removed and left at r.t for 10min. Cells were broken up, and 100mL of lysis buffer added with stirring at r.t for 15min, until all cells were thawed, and the mixture homogeneous. Cells were then transferred back to ice, and 500$\mu$L of DNase (2mg/mL stock) added, and cells stirred frequently on ice, until no longer viscous. Cells were diluted with a half volume of lysis buffer before being spun at 9,000rpm for 2hr at 4$^{o}$C.

### 2.3.2 Expression of Myoglobin Mutants and cMyBP-C Domain C0

Original plasmids for the expression of sperm whale myoglobin, as well as the plasmid containing the pyrrolysyl tRNA synthetase / $tRNA_{CUA(Pyl)}$ pair for Unnatural amino acid incorporation were provided by Prof Jason Chin, MRC, Cambridge.

The antibiotics used for selection were Kan and Tet at the concentration outlined in Table 2.3 for cMyBP-C and Myoglobin respectively. Following successful transformation of plasmid DNA (2.1.2) into DH10β *E. coli* cells, a single colony was used to inoculate 100mL of LB, supplemented with the appropriate antibiotic, and the culture grown overnight at 37$^{o}$C with shaking (200 rpm). Following overnight incubation, 10mL of the starter culture was used to sub-culture each litre of LB. Cells were grown until an $OD_{600}$ of between 0.6 -0.8 was achieved, before protein expression induced for 5hr at 37$^{o}$C with the addition of L-arabinose (0.2% final concentration) for

myoglobin, and 1mM (final concentration) IPTG for the cMyBP-C domain C0. Cells were pelleted by centrifugation, supernatants discarded, and cells kept at -20$^{o}$C until cell lysis by sonication.

#### 2.3.2.1 Cell Lysis of Myoglobin Mutants and cMyBP-C Domain C0 by Sonication

All cells were allowed to thaw on ice, before being re-suspended in lysis buffer. Sonication was carried out on ice, using a 30s on/off cycle for 15min at 15 amplitude microns. The suspension was clarified by centrifugation (16,000rpm, 4$^{o}$C) for 30min, to precipitate cellular debris and insoluble materials, and obtain a cell free extract.

### 2.3.3 Expression of 6xHis TEV Protease

A glycerol stock containing Rosetta$^{TM}$ 2(DE3) cells successfully transformed with the recombinant 6xHIS TEV protease was kindly donated by Dr Huanting Liu and Prof. James H. Naismith, University of St Andrews, Scotland, UK.

The antibiotics used for selection throughout were Cm and Amp at the concentrations outlined in Table 2.3. To 10mL of LB broth, a single colony was added, and cells grown overnight at 37$^{o}$C with shaking at 200rpm, before being used to inoculate 1L of LB broth. Cells were grown until an $OD_{600}$ of 0.6 was achieved, before protein expression was induced by the addition of IPTG at a final concentration of 0.4mM, and cells incubated with shaking overnight at 37$^{o}$C. Following overnight induction, cells were harvested by centrifugation at 6,000rpm for 10min at 4$^{o}$C, and the pelleted cells kept at -70$^{o}$C until ready to be lysed.

#### 2.3.3.1 Cell Lysis of 6xHis TEV Protease by Sonication

Pelleted cells were thawed on ice, and re-suspended in 10mL of Lysis Buffer . Once homogenous, cells were sonicated on ice for 3x45s intervals, with a 1min break in between, at 15 amplitude microns, before being centrifuged at 10,000 rpm for 20min at 4$^{o}$C. The supernatant was collected and further clarified by further centrifugation at 18,000 rpm for 30min at 4$^{o}$C.

### 2.3.4 Expression of Myoglobin with the UAA pK Incorporated

The antibiotics used for selection throughout were Tet and Kan at the concentrations outlined in Table 2.3. Following successful co-transformation of the plasmids

containing the myoglobin construct and the plasmid containing the relevant tRNA and tRNA synthetase for incorporation of the UAA pK, a single colony was used to inoculate 100mL of LB broth and this incubated with shaking (200rpm) overnight at 37$^{o}$C. Following overnight incubation, the starter culture was used to seed more LB broth. Cells were grown until an $OD_{600}$ of 0.6-0.8 was achieved.

At this point, the UAA pK was added at a final concentration of 2mM and 3mM for single and double mutants respectively. The UAA was dissolved in DMSO. Cells were left shaking for 30min at r.t, before L-arabinose was added at a final concentration of 0.2% to initiate induction. Cells were induced for 5hr at 37$^{o}$C, before being harvested by centrifugation at 4,500 rpm for 15min at 4$^{o}$C. Pelleted cells were kept at -20$^{o}$C until ready to be lysed.

#### 2.3.4.1 Cell Lysis of Myoglobin containing the UAA, pK,

Pelleted cells were allowed to thaw on ice, before 10mL of r.t BugBuster$^{TM}$ (MerkMillipore, Massachusetts , USA) was added, and cells re-suspended. Cells were shaken at r.t for 60min before lysed cells were pelleted by centrifugation at 10,000rpm for 20min at 4$^{o}$C.

### 2.3.5 Expression of 14-3-3ς and Vps75 mutants

Agar plates from the successful transformation of the recombinant 14-3-3 ς and Vps75 proteins into Rosetta$^{TM}$ 2(DE3) were kindly provided by Dr David Norman, University of Dundee, Scotland, UK. From these glycerol stocks were made for subsequent expression.

The antibiotics used for selection throughout were Amp and Cm for 14-3-3ς and Vps75, respectively, at the concentrations outlined in Table 2.3. To 10mL of LB broth, a single colony was added, and cells grown overnight at 37$^{o}$C with shaking at 200rpm, before being used to inoculate 1L of LB broth.

Cells were grown until an $OD_{600}$ of between 0.6 and 0.8 was achieved, before protein expression was induced by the addition of IPTG at a final concentration of 1mM, and cells incubated with shaking for 4hr at 37$^{o}$C. Following induction, cells were harvested by centrifugation at 6,000rpm for 10min at 4$^{o}$C, and the pelleted cells kept at -20$^{o}$C until ready to be lysed.

#### 2.3.5.1 Cell Lysis of 14-3-3ς and Vps75 Mutants by Sonication

All cells were allowed to thaw on ice, before being re-suspended in lysis buffer. Sonication was carried out on ice, using a 30s on/off cycle for 15min at 10 amplitude microns. The suspension was clarified by centrifugation (16,000rpm, 4$^{o}$C) for 30min, to precipitate cellular debris and insoluble materials, and obtain a cell free extract.

## 2.4 Recombinant Protein Expression Using the *P. pastoris* Expression System

Expression on all scales (unless otherwise stated) was carried out at 30$^{o}$C in a shaking incubator (250rpm, New Brunswick Scientific Model G25, Innova 4430 incubator shaker, New Brunswick Scientific, St Albans, UK). In order to optimise protein expression in *P. pastoris* there must be sufficient aeration, in particular during the methanol (MeOH) induction phase and therefore culture volumes did not exceed 30% of the total volume of the baffled flasks. All cells were harvested by centrifugation with the Sorvall© legend RT centrifuge, using the Sorvall© SH-3000 swinging bucket rotor, Thermo Scientific, UK). All media used for the expression of recombinant proteins in *P. pastoris* are included in Appendix A1 and A12.

### 2.4.1 The *P. pastoris* Expression System

*P. pastoris* is a eukaryotic, single celled microorganism, ideally suited for protein over-expression due to the ease with which it can be manipulated and cultured. *P. pastoris* can be cultured to high cell densities for high level production of recombinant proteins, with the added capability in the form of post translational modifications, such as disulphide bond formation, protein folding, and glycosylation (Zhang *et al*, 2009). The ability to fold proteins correctly, as well as the efficient secretion of proteins extra-cellularly puts *P. pastoris* at a major advantage over more traditional expression systems such as *E. coli*, especially when working with proteins with internal disulphide bonds and additional free cysteine residues.

*P. pastoris* is one of several expression systems capable of utilizing MeOH as both a carbon and energy source. This organism has a metabolic pathway for the consumption of MeOH based upon its oxidation to formaldehyde, catalysed by the enzyme alcohol

oxidase (AOX). There are numerous host strains of *P. pastoris*, each varying in their ability to utilise MeOH, due to deletions in one or both of the AOX genes. The KM71H strain used throughout has the $Mut^S$ phenotype, meaning that it is particularly slow in MeOH utilisation, due to a disrupted AOX I gene. In $Mut^S$ strains, transcription of the AOX genes is repressed when grown in the presence of sufficiently high concentrations of either glucose or glycerol, meaning that addition of MeOH serves primarily as an inducer of the disrupted AOX I promoter, whilst the alternative carbon source acts as the growth medium.

The pPICZα vector (Invitrogen) illustrated in Figure 2.1 contains the promoter for the AOX I region, one of the strongest known promoters in nature (Schmidt *et al*, 2011). The vector used throughout is the 3.6kb pPICZαB.

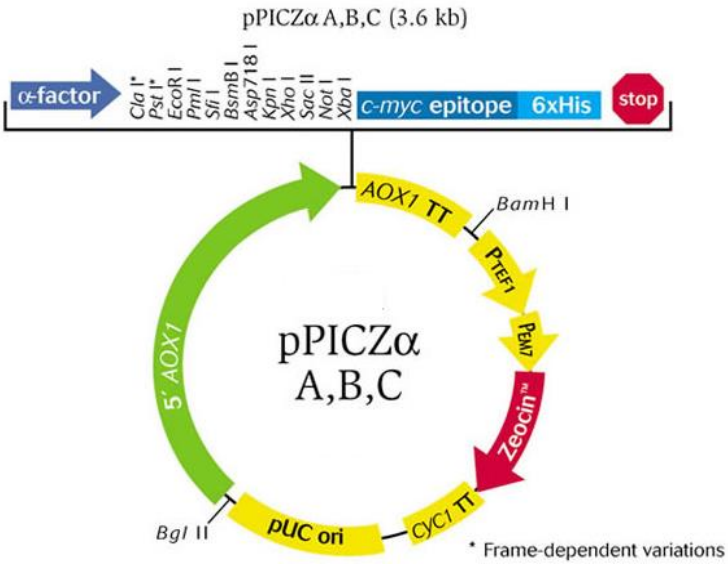



**Figure 2.1: The pPICZαB vector:** The vector map summarises the features of the pPICZαB vector, and its use for protein expression in *P. pastoris*. From Invitrogen.com

Upon induction with MeOH, the organism expresses the protein of interest, the gene of which is inserted directly behind the promoter. Insertion of the foreign gene is done so in the multi-cloning site which houses several unique restriction sites for the insertion of the desired gene.

The transcriptional termination sequence of the AOX I gene follows. *Pichia* secretes only a small number of proteins, and so insertion of the gene encoding the yeast mating

α factor peptide downstream from the AOX I targets the secretory pathway, and allows most proteins to be secreted directly into the media. *P. pastoris* has favourable properties for the secretion of high molecular weight proteins which are otherwise inefficiently secreted in most other yeast species, due to the retention of the recombinant protein in the periplasmic space (Schmidt *et al*, 2011). Lastly a Zeocin™ resistance cassette is incorporated to allow for selection in both *E. coli* for plasmid manipulation/cloning purposes and in *P. pastoris* for recombinant protein expression.

### 2.4.2 Small-scale protein production

Due to the slow growing nature of this particular strain of *P. pastoris*, it is necessary to identify those clones which show higher levels of recombinant protein expression. Following transformation, small – scale expression tests were carried out for this purpose, allowing the best clones to be selected for larger scale flask expression or for fermentation purposes. A single colony was used to inoculate 5mL of BMGY media (for test expressions from glycerol stocks, the stock was first streaked onto the appropriate concentration of Zeocin™ on YPDS agar, to achieve single colonies) and grown overnight with shaking at 30°C. This 5mL starter culture was then used to inoculate a further 95mL of fresh BMGY media and shaken at 30°C for 72hr, allowing the cell density to increase with the addition of a carbon source, whilst inhibiting induction. Each 100mL culture was spun down at 3,000 x g for 15min to pellet cells, which were then re-suspended in 25mL of BMMY media.

Cells were incubated at 18°C for 96hr with shaking, supplementing cultures with 0.5% (v/v) MeOH once every 24hr, and twice (AM and PM) in the final 24hr period. Cell cultures were harvested by centrifugation at 3000 x g for 15min, and supernatants filter sterilized (using Millex® Syringe Filter Units, 33mm, 0.22$\mu$M PES, Merck Millipore, Nottingham, UK). Supernatants were either concentrated by centrifugation at 3,000rpm using Vivaspin® 20 10,000 Dalton MWCO spin concentrators (Merck Millipore, Nottingham, UK) to a final volume of 1mL, or samples were taken from un-concentrated supernatants. The supernatants were analysed for protein expression by sodium dodecylsulfate (SDS) polyacrylamide gel electrophoresis (Section 2.7.1). Those clones which were found to express best were used to make glycerol stocks, from which all future expression and purification were performed.

### 2.4.3 Large-scale protein production

Once small-scale expression tests had established which clones were better expressers of the protein of interest, large scale flask expression could then be carried out to increase yield. In a similar fashion, a single colony was used to inoculate 5mL of BMGY media. Following overnight incubation at 30$^{o}$C, the 5mL starter culture was used to inoculate 95mL of fresh BMGY media. Cells were left shaking overnight at 30$^{o}$C before the 100mL was used to inoculate 400mL of fresh BMGY media, giving a final culture volume of 500mL per 2L baffle flask. Cells were left shaking at 30$^{o}$C for 48hr before being spun down at 3000 x g for 15min at 4$^{o}$C and cells re-suspended in 125mL BMMY media. Cells were grown for a further 96hr at 18$^{o}$C with shaking, supplementing the cells with 0.5% MeOH (v/v) every 24 hr for the first 48 hr, and subsequently 1% MeOH every 24hr for the last 48hr (AM and PM).

Cell cultures were then harvested by centrifugation at 3000 x g for 15min, and supernatants filter sterilized. Supernatants were either concentrated by centrifugation at 3,000rpm to a final volume of 1mL, or samples were taken from un-concentrated supernatants. The supernatants were analysed for protein expression by sodium dodecylsulfate (SDS) polyacrylamide gel and kept at 4$^{o}$C until required for purification.

### 2.4.4 Fermentation

All large-scale (ten-litre) fermentations were performed by Dr. John White at the University of Edinburgh. Fermentation, conducted on a Bioflow 3000 (New Brunswick Scientific, New Jersey, USA), allowed the temperature, pH, agitation, and air flow to be monitored and adjusted as required to optimise protein expression. Following successful transformation of fH fragments into *P. pastoris*, the clone which was found to express best was used to inoculate 2x 5mL of BMGY. The starter cultures were grown overnight at 30$^{o}$C, 200rpm before being used to inoculate 2x95mL of fresh BMGY media. This was then grown overnight at 30$^{o}$C, 200rpm, before the two cultures were used to inoculate 2x400mL of fresh BMGY media, giving a final culture volume of 1L (2x500mL).

The final 2x500mL cultures were grown overnight for 48hr at 30°C, 200 rpm and were used to inoculate the fermenter. The fermenter was previously supplemented with autoclaved fermentation media. To induce gene transcription, 0.5 % (v/v) MeOH and

1.5% $PTM_1$ salts were added. After the initial MeOH had been consumed, further feeds of MeOH – to 1.5% (v/v) - and 1% tryptone (w/v) were given for three days; the timing of which was based on the agitation rate and the amount of dissolved oxygen present in the vessel, so as to avoid overfeeding or poisoning of the cells. Following the fermentation procedure, cells were spun at 4,000 rpm for 60min.

The supernatant was then transferred to sterile centrifuge bottles and spun again for 60min at 7,000rpm to ensure all cell mass was removed. The spinning was all carried out at $4^oC$. Supernatants were then carefully decanted and a final concentration of 5mM EDTA and 0.5mM PMSF were added to reduce protease activity. The supernatant was then stored at $4^oC$ for immediate purification, or at $-80^oC$ for long term storage before purification.

## 2.5 Purification of Recombinant Proteins

All columns and resins, with the exception of the Ni-NTA Superflow cartridges (5mL) (QIAgen), were purchased from GE healthcare Bio-Sciences (Amersham, UK). Dialysis was carried out using SnakeSkin™ Dialysis Tubing, with a molecular weight cut off (MWCO) suitable for the protein of interest, at $4^oC$ with stirring. In between each purification step, samples were concentrated using Vivaspin® Centrifugal concentrators, with MWCO suitable for the protein of interest.

With the exception of purification steps being performed using a peristaltic pump (P-1, GE Healthcare, Amersham, UK), all chromatography was performed using the fast protein liquid protein chromatography (FPLC) systems, AKTA FPLC (GE Healthcare Bio-Sciences, Amersham, UK) and AKTAxpress (Amersham Bio-Sciences, Amersham, UK).

For Immobilised Metal Affinity Chromatography (IMAC), the column was stripped with 5CV of 100mM EDTA, washed with 5CV of $ddH_2O$, recharged with 5CV of 100mM $NiS0_4$ and finally washed with 5CV of $ddH_2O$ to remove residual $NiSO_4$, before purification. Before being loaded onto the Ni-NTA columns, all supernatants were filter sterilised using Millex® Syringe Filter Units, 33mm, 0.22$\mu$M PES, Merck Millipore, Nottingham, UK).

### 2.5.1 Purification of 6xHIS TEV Protease Using IMAC

All buffers used for the purification of the TEV protease are included in Appendix A8. Following centrifugation to obtain a cell-free extract, the supernatants were applied to a 5mL Ni-NTA Superflow Cartridge, pre-equilibrated in 5CV of equilibration buffer, using a peristaltic pump at a flow rate of 0.5mL/min. The column was washed with 5 CV, and the TEV protease eluted over 5CV of elution buffer in 3 mL fractions. Following SDS-PAGE analysis those fractions containing TEV protease were pooled, and the pool dialysed for 2hr at 4$^{o}$C (Buffer 1). This was repeated, before a final 2hr dialysis at 4$^{o}$C (Buffer 2). The protein was then centrifuged at 14,000rpm for 10min at 4$^{o}$C to remove any precipitate, and the TEV protease stored at -80$^{o}$C at a final concentration of ~ 6mg/mL.

### 2.5.2 Purification of RLC/MiniHMM Constructs Using IMAC

All buffers used for the purification of the RLC/MiniHMM are included in Appendix A7. Following centrifugation to obtain a cell-free extract, the supernatants were applied to a pre-equilibrated (5CV) 5mL Ni-NTA Superflow Cartridge, using a peristaltic pump at a flow rate of 5mL/min. The column was washed with 10CV of equilibration buffer, before a second wash of 5CV was performed. The protein was eluted in 5CV of elution buffer in 3mL fractions, and protein containing fractions analysed by SDS-PAGE.

#### 2.5.2.1 Cleavage of the Histidine Tag

Fractions found to contain the myosin RLC/MiniHMM complex were pooled, and to this, recombinant TEV protease was added in the ratio of 1 unit of TEV protease to 100 units of target protein (weight : weight). The cleavage reaction was allowed to persist for 3hr at r.t, before the mixture was dialysed overnight at 4$^{o}$C into non-imidazole buffer. Following overnight dialysis, the pooled protein fractions were flowed back over the recharged pre-equilibrated (5CV) Ni-NTA cartridge at a flow rate of 2mL/min. The flow through containing the cleaved RLC/MiniHMM complex was collected from the column, and the column washed with 5CV of wash buffer, followed by 5CV of elution buffer, which acts to remove un-cleaved protein, the cleaved tag and residual TEV protease from the column. The elution fractions were confirmed by SDS-PAGE

analysis, and the cysteine mutants treated with a final concentration of 10mM DTT to ensure the cysteine residue incorporated for SDSL purposes is free for labelling.

### 2.5.3 Purification of Wild Type, Cysteine and pK Containing Myoglobin Mutants, and cMyBP-C Domain C0 Using IMAC

All buffers used for the purification of domain C0 and the myoglobin mutants are included in Appendix A7 and A9, respectively. Following centrifugation to obtain a cell-free extract, the supernatants were applied to a pre-equilibrated 5mL Ni-NTA Superflow Cartridge, using a peristaltic pump at a flow rate of 2mL/min, or manual injection using a syringe for the UAA pK containing mutants. The column was then washed with 20CV, and the protein eluted in 6CV, in 1.5mL fractions, and purity determined by SDS-PAGE analysis. Concentrated protein was stored in 50% glycerol at -20$^{o}$C before the labelling reaction was carried out.

Purification of low spin myoglobin for data collection on HiPER differs slightly to that detailed in 2.5.3, and as such is described separately in detail in 5.13.

#### 2.5.3.1 Size Exclusion Chromatography

For cysteine mutants which would be subsequently spin-labelled (2.9.1.), the sample was treated with a final concentration of 10mM DTT at r.t for 1 hr. A pre-equilibrated HiLoad$^{TM}$ 16/600 Superdex$^{TM}$ S-75 size exclusion column was loaded with the 7.5mL concentrated sample, which was eluted over 1CV of elution buffer, collecting 2mL fractions.

### 2.5.4 Purification of 14-3-3 ς Using IMAC

All buffers used for the purification of 14-3-3ς are included in Appendix A10. Following centrifugation to obtain a cell-free extract, the supernatants were applied to a pre-equilibrated (5CV) 5mL Ni-NTA Superflow Cartridge, using a peristaltic pump at a flow rate of 5mL/min.

The column was washed with 10CV of equilibration buffer, before a second wash of 5CV was performed. The protein was eluted in 5CV of elution buffer in 3mL fractions, and protein containing fractions analysed by SDS-PAGE. Samples were concentrated using Vivaspin® centrifugal concentrators.

#### 2.5.4.1 Size Exclusion Chromatography

A HiLoad™ 16/600 Superdex™ S-75 size exclusion column was pre-equilibrated with 2CV, and the concentrated 14-3-3ς sample loaded onto the column, and eluted over 1CV with elution buffer (AKTAxpress, Amersham Biosciences, UK), collecting 2mL fractions.

### 2.5.5 Purification of Vps75 Using IMAC

All buffers used for the purification of Vps75 are included in Appendix A11. Following centrifugation to obtain a cell-free extract, the supernatants were applied to a pre-equilibrated (5CV) 5mL Ni-NTA Superflow Cartridge, using a peristaltic pump at a flow rate of 5mL/min. The column was washed with 10CV of equilibration buffer, before a second wash of 5CV was performed. The protein was eluted in 5CV of elution buffer in 3mL fractions, and protein containing fractions analysed by SDS-PAGE. Samples were concentrated using Vivaspin® centrifugal concentrators.

#### 2.5.5.1 Size Exclusion Chromatography

A HiLoad™ 16/600 Superdex™ S-75 size exclusion column was pre-equilibrated with 2CV, and the concentrated Vps75 sample loaded onto the column, and eluted over 1CV with elution buffer (AKTAxpress, Amersham Biosciences, UK), collecting 2mL fractions.

### 2.5.6 Purification of fH 1-4 K247C Fragment Using IMAC

All buffers used for the purification of fH1-4 K247C are included in Appendix A3. The inclusion of a hexa-histidine tag at the C' terminus of the fH1-4 facilitates the use of IMAC as the first catchment step in the purification of fH1-4 K247C. Due to the addition of EDTA (5mM) following harvesting of fermentation, the supernatant was diluted with wash buffer to give a final concentration of <1mM EDTA, so as not to strip the charged Ni-NTA cartridge. The diluted supernatant was loaded onto the pre-equilibrated column (5CV), using a peristaltic pump at a flow rate of 2mL/min. The column was washed with 10CV, before the bound protein was eluted using an imidazole gradient of 50-500mM, over 10CV, collecting 5mL fractions. Fractions were analysed by SDS-PAGE.

#### 2.5.6.1 Anion Exchange Chromatography

The concentrated samples were exchanged into wash buffer to eliminate NaCl and facilitate binding of the fH1-4 K247C mutant to the TricornMonoQ 4.6/100 (1.7 mL CV) Mono Q anion exchange column. Elution of the desired protein was achieved by applying a linear NaCl-gradient from 100% wash buffer to 100% elution buffer, over 20CV, at a flow rate of 2mL/min, collecting 2mL fractions. Fractions were then analysed using SDS-PAGE and relevant fractions pooled and concentrated before being treated with a final concentration of 250µM TCEP-Hydrochloride, which acts to keep the incorporated cysteine free for SDSL purposes, whilst maintaining the integrity of the internal disulphide bonds of the fH1-4 fragment (Burns *et al*, 1991).

#### 2.5.6.2 Size Exclusion Chromatography

A HiLoad™ 16/600 Superdex™ S-75 size exclusion column was pre-equilibrated with 2CV, and the concentrated fH1-4 sample loaded onto the column, and eluted over 1CV with elution buffer (AKTAxpress, Amersham Biosciences, UK), collecting 2mL fractions.

### 2.5.7 Purification of fH19-20 G1107C Fragment Using SP Sepharose FastFlow Cation Exchange Chromatography Resin

All buffers used for the purification of fH19-20 G1107C are included in Appendix A4. SP Sepharose FastFlow resin was used as the first catchment step in the purification of the fH19-20 G1107C fragment. To an empty XK 26/20 column (GE Healthcare), 30mL of the resin was added, and the column packed using a peristaltic pump. Following expression, the supernatants were diluted by a factor of 10 using $ddH_2O$, in order to reduce the conductivity, which rises as a result of the addition to $PTM_1$ salts, and aids binding of the fH19-20 fragment to the ion exchange resin.

The diluted supernatant ranges from a final volume of between 10 and 80 litres, and so the supernatant was applied to the packed XK 26/20 column using a peristaltic pump, overnight at $4^oC$ at the maximum flowrate. Elution of the desired protein was achieved by applying a linear NaCl-gradient from 100% wash buffer to 100% elution buffer, over 20CV, at a flow rate of 2mL/min, collecting 2mL fractions.

#### 2.5.7.1 Cation Exchange Chromatography

The concentrated samples were exchanged into wash buffer to eliminate NaCl and facilitate binding of the fH19-20 G1107C mutant to the Resource™ 15S cation exchange column. Elution of the desired protein was achieved by applying a linear NaCl-gradient from 100% wash buffer to 100% elution buffer, over 20CV, at a flow rate of 2mL/min, collecting 2mL fractions. Fractions were then analysed using SDS-PAGE and relevant fractions pooled and concentrated before being treated with a final concentration of 250µM TCEP-Hydrochloride, which acts to keep the incorporated cysteine free for SDSL purposes, whilst maintaining the integrity if the internal disulphide bonds of the fH19-20 fragment.

#### 2.5.7.2 Size Exclusion Chromatography

A HiLoad™ 16/600 Superdex™ S-75 size exclusion column was pre-equilibrated with 2CV, and the concentrated fH19-20 sample loaded onto the column, and eluted over 1CV with elution buffer (AKTAxpress, Amersham Biosciences, UK), collecting 2mL fractions.

## 2.6 Isolation and Purification of Complement Proteins From Human Blood/Plasma

Experiments looking at the interactions of Complement Factor H and C3b began at the University of Edinburgh in collaboration with Prof Paul N. Barlow (2011-2013). Following relocation of our research group to the University of St Andrews, work carried out was done so in compliance with the Ethical Application Submitted to The University of St Andrews, School of Physics and Astronomy (**Ethics Reference No: PA10837**, 20th February 2014). Copies of the application form, participant information forms, statement of approval, and consent forms are included as Appendices.

### 2.6.1 Small-scale isolation of complement C3

All buffers used for the purification of C3 are included in Appendix A5.

#### 2.6.1.1 Obtaining the Plasma Fraction from Whole Blood

Complement C3 is one of the most abundant plasma proteins, present in the plasma at a concentration of ~ 1.2 mg/mL (Sahu & Lambris, 2001), and so was purified from fresh human plasma using methods adapted from Dodds, 1993. Plasma was prepared by collecting freshly drawn blood (typically 20mL) from healthy volunteers into 50mL falcon tubes containing sodium EDTA at a final concentration of 10mM. Tubes were inverted to ensure thorough mixing and samples left on ice for 15min before centrifugation at 2,000g, for 10min at $4^{o}$C to obtain the plasma fraction.

#### 2.6.1.2 Precipitation of Major Blood Proteins

The plasma was made 5% (w/v) by the addition of polyethylene $glycol_{3350}$ ($PEG_{3350}$). To the plasma, 5mL of a 15% $PEG_{3350}$ solution (in wash buffer) was added to precipitate major blood proteins, such as antibodies (*e.g.* IgG and IgM), and rolled for 30min at $4^{o}$C, before being spun at 10,000rpm for 20min to pellet the resulting precipitate.

#### 2.6.1.3 Anion Exchange Chromatography – Q-Sepharose Fast Flow Resin

The supernatant was then loaded at a flow rate of 2mL/min at $4^{o}$C onto a 20cm x 1.6cm diameter column (GE Healthcare) packed with Q-Sepharose® FastFlow resin, pre-equilibrated in 95% wash buffer and 5% elution buffer. C3 was eluted with a linear NaCl gradient (10CV, 5-50% elution buffer at a flow rate of 1mL/min) to a final concentration of 500mM NaCl, collecting elution fractions (2mL) on ice. The C3 pool then underwent three rounds of dialysis in a final volume of 5L of wash buffer to eliminate NaCl and ultimately lower the ionic strength and allow binding to the higher resolution Mono Q column for subsequent purification steps. (Zhang *et al*, 2003)

#### 2.6.1.4 Anion Exchange Chromatography – Mono Q

The 5/50 Mono Q column was pre-equilibrated with 90% wash buffer, 10% elution buffer. The pooled C3 fractions were eluted over several rounds of anion exchange chromatography using a 1M NaCl gradient (20mL, 10-30% elution buffer, at a flow rate of 1mL/min). Following purification on the Mono Q column, C3 fractions were pooled and again dialysed three times against 5L of wash buffer to again eliminate NaCl and

lower ionic strength. The concentration of the purified C3 was determined to be 0.3mg/ml using a UV spectrophotometer (Eppendorf) using C3 molar Extinction Coefficient at 280nm of 180,000$M^{-1}cm^{-1}$.

#### 2.6.1.5 Size Exclusion Chromatography

A HiLoad™ 16/600 Superdex™ S-75 size exclusion column was pre-equilibrated with 2CV, and the concentrated C3 sample loaded onto the column, and eluted over 1 CV, collecting 2mL fractions.

### 2.6.2 Large Scale Isolation of Complement C3

All buffers used for the purification of C3 are included in Appendix A5. The process was scaled up, from a starting material of 20mL of blood (giving 10 mL plasma) to 150mL of blood (75mL plasma). The purification process was carried out as detailed above, however, a Resource™ 15 Q column (20ml) was used between the QSFF and Mono Q purification steps. The concentrated samples were exchanged into wash buffer to eliminate NaCl and facilitate binding of the protein to the Resource™ 15Q anion exchange column. Elution of the desired protein was achieved by applying a linear NaCl-gradient from 100% wash buffer to 100% elution buffer, over 20 CV, at a flow rate of 2mL/min, collecting 2mL fractions. The remainder of the purification proceeded as detailed above.

### 2.6.3 Conversion of C3 to C3b using Trypsin

All buffers used for the purification of C3b are included in Appendix A6.

Purified C3 was then converted to C3b using methods adapted from (Wu *et al*, 2009). To successfully convert C3 to the activated C3b and release the cysteine residue in the thioester for spin labelling purposes, C3 must undergo a limited trypsin proteolysis at 37$^{o}$C to release the C3a fragment, leaving activated C3b. Purified C3 at a concentration of 1mg/ml was treated with a 1% trypsin solution (w/w, enzyme/protein) for 2min at 37$^{o}$C in PBS, to simulate physiological pH. Immediately after incubation, 5% (w/w inhibitor/enzyme) soybean trypsin inhibitor was added to stop the reaction.

For the proposed DEER experiments, C3b is required in both a labelled and non-labelled form. For the unlabelled reaction, following trypsin proteolysis and trypsin

inhibition, the sample was immediately transferred to ice, followed by treatment with a final concentration of 20mM iodoacetamide for 30min. Iodoacetamide binds covalently with the newly released cysteine residue on the activated thioester on C3b, therefore blocking this binding site.

The C3b was then diluted at a ratio of 1:1 with wash buffer, and eluted over 30 CV with a linear salt gradient of 100-300mM NaCl from the pre-equilibrated Mono Q column. The fractions containing C3b in its pure form were pooled and dialysed against PBS, before being stored at -80$^{o}$C. The purity of C3b was verified using SDS-PAGE.

### 2.6.4 Conversion of C3 to 'C3 (N)' – An Alternative to C3b

All buffers used for the purification of C3(N) are included in Appendix A6. Conversion of C3 to C3(N) was done so using protocols adapted from Holm *et al*, 2012). Purified C3 was treated with a final concentration of 200mM methylamine in wash buffer, at 37$^{o}$C for 3hr. The pH of the buffer was adjusted following the addition of methylamine (pH7.5). The reaction mixture was incubated in the presence of either iodoacetamide (20mM final concentration) or with a 10x molar excess of MTSSL spin label, depending on whether the protein was required in its labelled or unlabelled form.

Following the 3hr incubation period, the C3 pool was then purified in the same way detailed above, as for C3b. Following the final SEC step, purity of the methyl amine treated C3, or C3 (N), was determined by SDS-PAGE analysis, and proteins concentrated to ~1mg/mL before being stored at -80$^{o}$C.

## 2.7 Protein Quantification

### 2.7.1 Sodium-Dodecyl Sulphate Polyacrylamide Gel Electrophoresis(SDS-PAGE)

All gels used for protein identification were 4-12% NuPage® Bis-Tris gels (Life Technologies) and were run in 1X in NuPage® MES or MOPS Buffers. For improved resolution of high molecular weight proteins, *e.g.* C3, C3b and C3(N), MOPS was the preferred buffer. Samples were prepared, and 4x NuPage® LDS sample buffer was added to a final concentration of 1x. If samples were being run in their reduced form, then 10x NuPage® sample-reducing agent (Life Technologies) was added to a final concentration of 1x.

Samples were heated at 95 – 100 °C for 5min. The samples were run alongside a protein ladder (10 – 250 kDa range) to provide a size standard, at 250V until good separation of the ladder was achieved (~ 25min). Prior to staining, the gel was washed in water followed by heating, using a microwave oven, for 30s; this was repeated 3 times to remove the SDS from the gel.

Staining was performed using traditional Coomassie Blue staining for 15min with shaking. The gel was then destained by the addition of Coomassie Destain and heated for 1min before shaking, until the Coomassie was removed from the gel, and protein bands could be visualised.

### 2.7.2 Western Blotting

Western Blotting was carried out for more specific detection of particular constructs if, following Coomassie staining, the SDS-PAGE gels was inconclusive. All antibodies (Table 2.10) were diluted as per the manufacturer's instructions.

**Table 2.10: Antibodies used for Blotting, company purchased from, and working dilution**

| | | |
|---|---|---|
| Anti-Factor H polyclonal antibody (raised in goat) | Calbiochem | 1:200 |
| Anti-Goat IgG (whole molecule)–Peroxidase antibody (raised in rabbit) | Sigma Aldrich | 1:50,000 |
| Histidine Tag (6xHis) Monoclonal Antibody (raised in mouse) HRP Conjugate | Life Technologies | 1:5,000 |

In this instance, a gel was run as detailed in 2.7.1, however, the Coomassie staining procedure was omitted. Western blotting was carried out using standard protocol and PVDF membrane. Blotting pads and filter paper were soaked in Towbin Transfer Buffer, before being layered in the blotting module with the SDS-PAGE gel, and PVDF membrane, pre-soaked in 100% MeOH. The transfer was carried out using XCell II.Blot Module (Life Technologies) at a constant voltage of 30V for 1hr in Towbin Transfer buffer. The transfer was monitored by staining the gel afterwards and using pre-stained molecular-weight markers.

After the transfer, the membrane was blocked overnight on an orbital rocking platform at $4^{o}C$ using PBS containing 5 % non-fat dried milk (w/v) (blocking solution, supermarket brand). The membrane was then placed in a 5% (w/v) non-fat milk solution (in PBS) with the recommended dilution (Table 2.10) of the primary antibody, and incubated at $4^{o}C$ for 3hr.

The membrane was rinsed with PBS and washed three times with PBS containing 0.05 % TWEEN®20 for 30min, and washed with PBS for another 10min. The membrane was then probed with the secondary antibody at the desired dilution in a 5% (w/v) solution of non-fat milk in PBS. The secondary antibody is conjugated to horse radish peroxidase. This was incubated for 2hr at $4^{o}C$. The blot was then washed as described above with PBS and PBS-TWEEN® before being developed.

The blot was developed in the developing buffer, containing 1 diamino benzidine (DAB) tablet and 1 urea hydrogen peroxide tablet in 15mL of $ddH_2O$. The HRP conjugated to the secondary antibody oxidizes the DAB in the presence of hydrogen peroxide, giving the characteristic brown staining on the membrane, indicative of the secondary antibody bound to the primary. The membrane was shaken for 15min to develop and washed with water to remove excess stain, and dried overnight.

### 2.7.3 Determination of Protein Concentration Using Absorbance $A_{280}$

Protein concentrations can be determined using the Beer-Lambert Law, shown in equation 2.2, where (A) = the absorbance at a given wavelength, (ε) = the molar extinction coefficient, (c) = Protein concentration, and (l) is the path length.

$$A = \varepsilon c l \quad \text{(EQUATION 2.2)}$$

The molar extinction coefficients can be determined by entering the amino acid sequence of the protein of interest into the ExPASy ProtParam Tool. Proteins absorb light at 280nm due to the presence of chromophores such as tryptophan, tyrosine and phenylalanine residues.

### 2.7.4 Determination of Protein Concentration Using the Micro-Bradford Assay

The concentration of purified proteins could be determined using a modified ‘micro’ version of the Bradford assay (Bradford, 1976)With the micro-Bradford assay one can

determine protein concentrations in the range of 1 - 10 $\mu$g/mL. The assay is a colourimetric technique, based on a shift in absorbance from 465 to 595nm, upon binding of Coomassie brilliant blue to proteins. For the purposes of this assay, the Bradford reagent (BioRad) was diluted by a factor of 5, and a standard series of BSA at concentrations 0,1,3,5,7, and 10$\mu$g/mL used for calibration. The standard and samples were incubated with the Bradford reagent for 10min at r.t and the absorbance was read at 595 nm.

## 2.8 Protein Characterisation

### 2.8.1 Mass Spectrometry (MS) Analysis

Samples were submitted to the Scottish Instrumentation and Resource Centre for Advanced Mass Spectrometry (SIRCAMS) at the University of Edinburgh for high resolution liquid chromatography (LC) – MS, with the aim of acquiring an approximate mass. The samples were supplied in 2–5$\mu$M concentrations in 20$\mu$l of either 20mM potassium phosphate, pH 6, or PBS. Data was collected by passage through an LC-MS U300HPLC 500$\mu$M PSDVB monolyth column (Dionex) followed by Fourier-transform ion-cyclotron resonance (FT–ICR) mass spectrometry (Bruker Daltonics 12T SolariX Fourier Transform Ion Cyclotron Resonance Mass Spectrometer).

MS was carried out at the University of Edinburgh SIRCAMS facility by Dr David Clarke and Dr Logan Mackay.

At the University of St Andrews MS facility, analysis was carried out by Dr Catherine Botting, and Dr Sally Shirran. Proteins submitted for intact mass determination by Electro-Spray Ionisation-Time of Flight MS were run at a concentration of 10pmoles/$\mu$L.

### 2.8.2 Ellman's Assay - For the Quantification of Free Cysteine

This assay for the quantification of free thiols was developed in 1959 by Ellman et al, and further modified in 1968 by Sedlak & Lindsay. When in its reduced form, one can use the assay to determine the concentration of free cysteine (thiol) in a given sample, which proves useful for this work, when quantifying free cysteine levels before SDSL is carried out. The Ellman's reagent 5,5'-dithiobis(2-nitrobenzoic acid) (DTNB), as shown

in Figure 2.2A, reacts quantitatively with free thiol groups, forming a mixed disulphide. (SH group of free cysteine, and 2-nitro-5-thiobenzoate (NTB). The NTB component is shown in Figure 2.2B.

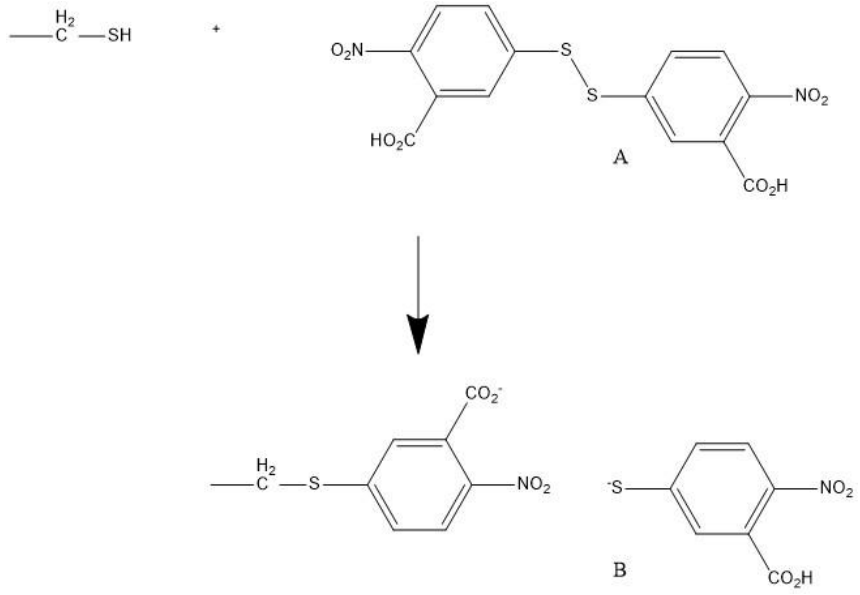



**Figure 2.2: Ellman's Assay Reaction.** DTNB **(A)** reacts with a free thiol to form NTB **(B)**, the reactive reagent of the Ellman's assay.

During the reaction of the Ellman's Reagent with proteins containing free thiols, the 2-nitrobenzoate-thio-anion, NTB, is stoichiometrically released, and can be detected by UV at 412nm. By measuring this, one can calculate the concentration of the NTB anion and thus the concentration of sulfhydryl groups. To ensure all proteins were in their reduced form prior to performing the assay, they were treated with a final concentration of 1mM DTT or 250$\mu$M TCEP, for 1h at r.t. The reducing agent was then removed by either dialysis or buffer exchange, and the Ellman's assay was performed.

The DTNB working stock was made by dissolving 40mg in 10mL of DMSO, before diluting this 100 fold with 0.1M Tris-HCl, pH7.5, giving a 0.1mM DTNB working solution. Samples were prepared by adding 50$\mu$L of standard or protein to 950$\mu$L of DTNB working solution, incubating at r.t for 15min before measuring the absorbance of samples at 412nm. The assay was calibrated using a solution of L-Cysteine at various

known concentrations, and utilising the known extinction coefficient of the reaction product, NTB (13,800 $M^{-1}cm^{-1}$). The L-cysteine calibration curve is shown in Figure 2.3.

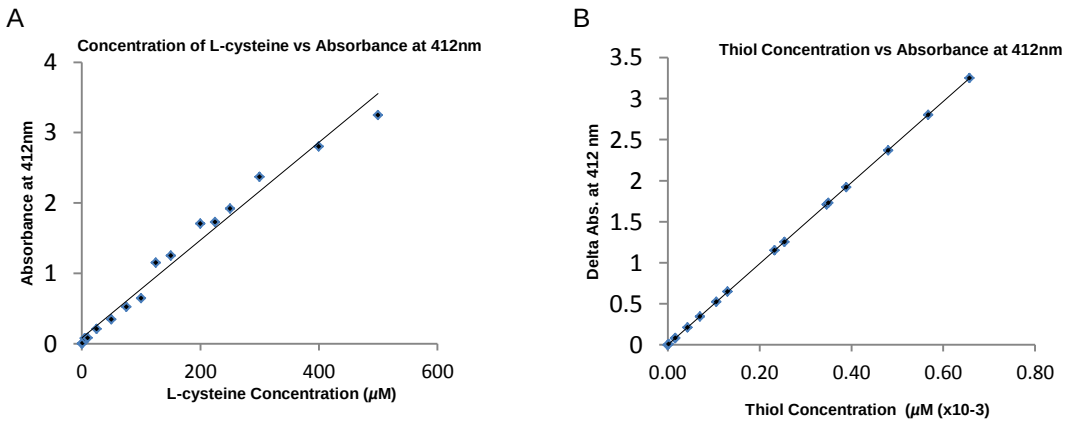



**Figure 2.3: L-cysteine calibration curves for determining free cysteine concentration.** (A) L-cysteine standards measured at 412nm upon reaction with DTNB. (B) Determination of thiol concentration from each standard, using the extinction coefficient of the NTB component at 412nm.

## 2.9 Protein Chemistry

### 2.9.1 SDSL of Incorporated Cysteine Residues

As stated previously, each protein was treated with a reducing agent prior to labelling, to ensure that the incorporated cysteine residue was free for spin labelling purposes. Following quantification of free thiol levels (2.8.3), each protein could then be spin-labelled. Typically, the MTSSL spin label (Toronto Research Chemicals, Ontario, Canada) was added in a 10x molar excess.

The spin label was first dissolved in 100$\mu$L of DMSO, before the dissolved label was added to ~1mL of r.t 50mM Tris-HCl (pH7.5). The label was then added directly to pooled protein fractions, and left overnight at 4$^{o}$C, before excess spin label was removed by successive rounds of dialysis or by buffer exchange using Vivaspin® concentrators.

## 2.10 Functional Characterisation

### 2.10.1 Assessing Cofactor Activity of C3b, C3 (N) and fH1-4 K247C

FH regulates the activation of the complement system by acting as a cofactor in the cleavage of complement C3b to its inactive form, iC3b, by fI. A fluid-phase cofactor assay can be used to definitively say whether plasma purified C3, converted to C3b/C3(N), functions as such.

In this particular assay, the 110kDa alpha chain of C3b is cleaved into two smaller components (67 and 40 kD). Figure 2.4 shows a schematic representation of cleavage of the alpha chain, in the presence of fH and fI.

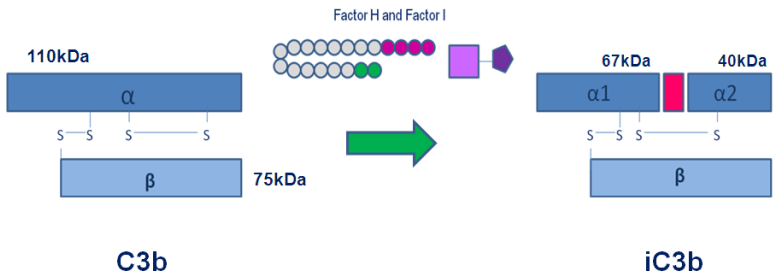



**Figure 2.4: Schematic Representation of the Fluid-Phase Cofactor Assay.** C3b is composed of two disulphide linked chains, α and β. Following incubation with both fH and fI, cleavage of the 110kDa alpha chain results in two smaller 67 and 40kDa components. C3b is converted to iC3b, which is visualised using SDS-PAGE.

Upon cleavage, C3b is converted to iC3b, and no longer functions in complement activation, ultimately limiting the production of the convertase C3bBb, therefore stopping the positive feedback loop of complement activation, and limiting activation. FH, fI, and C3b were purchased from Complement Technology (CompTech, Texas, USA). For the control reactions, all three components of the assay were from CompTech stocks, with the negative control being carried out in the absence of fH.

When assessing the function of the C3b, C3 (N) and fH1-4 K247C expressed and purified in this work, these proteins were used in place of the CompTech standards. For each reaction 45ng of fI, 120ng of fH and 3$\mu$g of C3b/C3 (N) were combined in a final assay volume of 20$\mu$L in PBS, pH7.4. The reaction mix was vortexed and incubated at

37$^{o}$C for 2 hr in a circulating water bath. An aliquot of SDS loading dye was added to stop the reaction, and samples were analysed by SDS-PAGE.

## 2.11 EPR Spectroscopy

R.t. CW EPR Spectra were recorded at the University of Edinburgh on a Bruker ESP300 X-band EPR spectrometer. At the University of St Andrews CW EPR spectra were recorded on a Bruker EMX X-band EPR spectrometer, using either a rectangular low sensitivity cavity or high sensitivity SHQE resonator, and using the AffirmoEx benchtop spectrometer, by ActiveSpectrum.

DEER experiments were performed on a Bruker Elexsys 580 X-band spectrometer, until 2014 when high powered Q-band was introduced.

For DEER measurements performed at X-band the 12ns pump pulse was set to the maximum of the echo detected spectrum. The 32ns observer pulse was set to 65MHz above.

For DEER measurements performed at Q-band the 14ns pump pulse was set to the maximum of the echo detected spectrum. The 24ns observer pulse was set to 80MHz below.

Typically, CW measurements were taken in non-deuterated buffer. Where samples had to be frozen for DEER, a cryo-protectant was added, either glycerol or deuterated glycerol to a final concentration of 50%. The glycerol helps in glass formation during freezing, which helps with relaxation. Deuterated buffers also help prolong spin coherence.

All experiments conducted at W-band using HiPER (Cruickshank *et al*, 2009) were performed by Miss Claire Motion, University of St Andrews, Scotland, UK.

# Chapter 3: The Complement Factor H (fH) – C3b Complex: An EPR Study

**AIMS:** This body of work uses EPR spectroscopy to probe the interactions of complement proteins C3b and Factor H (fH). Complement is a complex cascade of enzymatic cleavages, cumulating in cleavage of C3 to C3b, marking cells for destruction by the immune system. This is tightly regulated by Factor H (fH). FH domains 1-4 and 19-20 are known binding sites for C3b. DEER will be used to gain further insight into the structure of fH in the context of C3b. FH fragments are expressed recombinantly in *Pichia pastoris* yeast, and cysteine residues incorporated using mutagenesis for SDSL. C3b cannot be made recombinantly, and in this work is generated from cleavage of the purified plasma protein C3. A hidden thioester is then utilised as a spin-labelling site.

## 3.1 The Role of the Complement System in Innate Immunity

First identified in human serum in the late 19$^{th}$ century as a complement to antibodies in mediating bacterial lysis, the complement system emerged more than one billion years ago, probably as the first humoral immune system (Liszewski & Atkinson, 2015). The complement system is a major primary defence and clearance mechanism which straddles both the innate and adaptive immune systems, providing a non-specific and potent first-line in defence against infection (Walport, 2001). Complement activity is established by the orchestration of between 35 and 40 soluble plasma and cell surface proteins circulating as inactive precursors, which function to mark pathogenic surfaces and host apoptotic cells for cell lysis and clearance, whilst stimulating the adaptive immune response (M. C. Carroll, 2004) (Carroll & Sim, 2011).

Initial protection against infection occurs as a result of generation of large numbers of active complement proteins that bind and target cells for phagocytosis. Cleavage of specific complement proteins follows, which results in the release of inflammatory anaphylatoxins, recruitment of immune cells to the vicinity, and the initiation of an amplifying cascade of further cleavages, cumulating in assembly of the membrane attack complex (MAC), and ultimately cellular destruction (Ricklin *et al*, 2010). Regulation of these amplification cascades is of critical importance for maintaining

homeostasis, as well as preventing rapid depletion of complement proteins, whilst limiting damage to the host (Markiewski & Lambris, 2007).

### 3.1.1 Mechanisms of Complement Activation

Dependent upon the trigger, complement can be activated *via* three very distinct cascades, namely the classical, lectin and alternative pathways of complement activation. Initiation of each of these biochemical cascades occurs *via* the proteolytic cleavage of inactive complement pre-cursors upstream, producing the active enzymatic form, enabling activation of complement proteins further downstream.

Although each of the three pathways is distinct in its activation, they all converge at the production of C3 and its cleavage to C3a and C3b, resulting in the generation of anaphylatoxin C5a, activation of the common terminal pathway, and assembly of the MAC (as reviewed by (Merle, Church *et al*, 2015) and (Merle, Noe *et al*, 2015)).

Activation of complement *via* the alternative pathway (AP) differs somewhat from other means of complement activation, and it is the interactions, and protein complexes associated with the AP of complement activation on which this work will focus.

### 3.1.2 The Alternative Pathway of Complement Activation

Complement activation *via* the AP contrasts significantly with other means of complement activation, in that it is continuously stimulated due to an inherent 'tick-over' mechanism (Janssen & Gros, 2007), whereby a buried thioester (TE) moiety in the thioester domain (TED) on component C3, is spontaneously hydrolysed at a slow but constant rate, by nucleophilic attack by water, leading to the generation of C3($H_2O$). The tick-over mechanism of the AP of complement activation is illustrated in Figure 3.1.

Binding of Factor B (fB) to the C3($H_2O$) complex, makes it susceptible to cleavage by Factor D (fD), a serine protease, generating a C3 convertase, C3($H_2O$)Bb, which initiates the proteolytic cleavage of C3, releasing anaphylatoxin C3a, and main fragment C3b. C3b is structurally similar to C3($H_2O$), and so in a similar fashion, it binds fB, forming C3bB – the 'pro-convertase' complex. This encourages cleavage by fD, forming the fluid phase AP convertase, C3b.Bb.

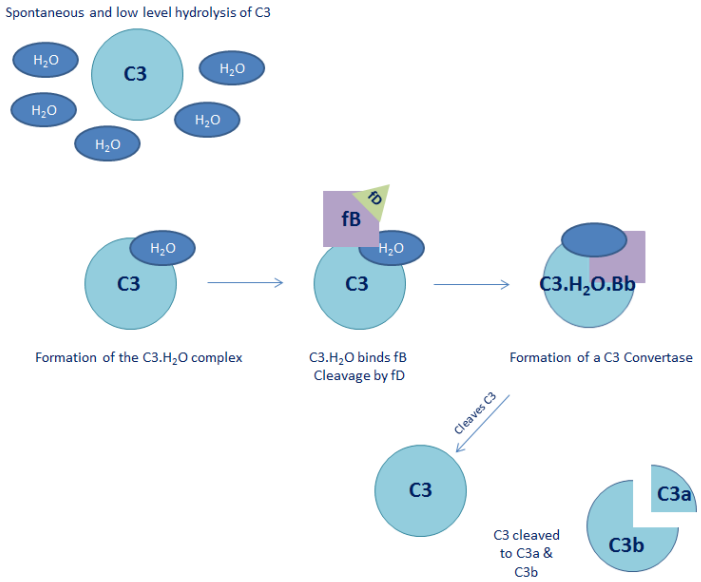



**Figure 3.1: 'Tick-Over' Mechanism of the AP of Complement Activation.** Spontaneous hydrolysis of C3 at low level, leads to the formation of C3($H_2O$). C3($H_2O$) binds fB, enabling cleavage of this complex by fD, resulting in formation of a soluble C3 convertase, C3.$H_2O$.Bb. Low levels of the C3 convertase C3.$H_2O$.Bb are constantly generated resulting in cleavage of C3 to C3b, which in turn, forms the AP convertase C3b.Bb.

Formation of this AP convertase amplifies the response, generating a positive feedback loop of C3 conversion to C3b, allowing rapid deposition of large numbers of C3b molecules (in excess of $10^8$) onto individual cellular surfaces in a process known as opsonisation. This amplification loop is believed to account for up to 80% of total complement activation (Pouw *et al*, 2015).

Some C3b re-associate with C3b.Bb complexes, resulting in the formation of C3b.Bb.C3b, the C5 convertase, switching the substrate to C5, and initiating the terminal pathway of complement activation – formation of the MAC – and targeted cellular destruction (Gros, 2011).

### 3.1.3 Complement Activation – Driven by Conformational Change

Complement component C3 is a member of a family of $\alpha_2$-macroglobulins ($\alpha_2$-M), also containing homologous proteins C4 and C5 (Sottrup-Jensen *et al*, 1985), which are characterised by homologous sequence features including a unique and eponymous TE motif (Gros, Milder, & Janssen, 2008).

Human C3 is synthesised in the body as a large, 185kDa, single chain polypeptide (Janssen & Gros, 2007), which is processed by the selective removal of four consecutive arginine residues (Arg 646-Arg 649), resulting in a 110kDa alpha chain (residues 650-1641) and a 75kDa beta chain (residues 1-645), linked by a single disulphide bond. Figure 3.2 shows the XRC structures of both C3 and C3b, as well as the TE in both its open and closed formation.

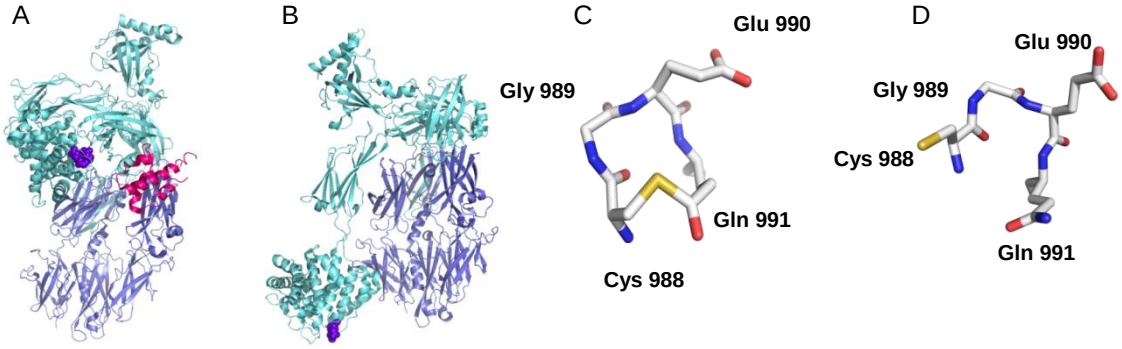



**Figure 3.2: Conversion of Complement C3 to its activated form C3b.** Upon activation, a previously hidden TE, responsible for binding of C3b to target cells in the opsonisation process, is exposed**.** (A) Cartoon representation of C3 with alpha chain shown in Cyan and beta chain in Slate. The TE domain is shown as purple spheres. The ANA domain (C3a) is shown in magenta. PDB_ID: 2A73 (Janssen *et al*, 2006). Upon cleavage to C3b (B), the ANA domain is lost, and the remainder of the protein (now C3b) undergoes conformational change, exposing the previously hidden TE. PDB_ID: from 2WII (Wu *et al*, 2009). (C) Stick representation of the TE formed between Cys 988 and Gln 991, which is in the closed state in C3. Upon conversion to C3b from C3, the TE bond is broken (D), and C3b uses the newly exposed and activated TE for cell binding.

Activation of component C3 to C3b is a crucial step in each of the distinct pathways of the complement cascade, in which a previously unexposed TE domain becomes exposed, as well as multiple cryptic binding sites on C3b for interacting complement proteins (Janssen *et al*, 2005).

C3 is composed of 13 domains, with the eponymous TE domain housing the iso-glutamyl cysteine TE bond (formed between the sulphur of Cys 988 and the carbonyl group of Gln 991). Ultimately, localization of immune response is determined by the reactive TE moiety of C3 (Lambris, 1988), which is buried in a hydrophobic pocket (residues Met 1378, Tyr 1425, Tyr1460, Phe 1047, which are conserved across the $\alpha_2$-M family), but becomes exposed in C3b after the C3a anaphylatoxin (protective ANA domain) is lost from C3 (following cleavage at Ser726, Arg 727). This results in a truncated alpha chain (residues 727 – 1641), whilst the beta chain remains unaffected.

Subsequent reshuffling of the domains *via* conformational change, rotates the TED by $85^{o}$, rendering the TE solvent-accessible (Gros *et al*, 2008). Rearrangement of the TED cleaves the newly exposed TE into a highly reactive thiolate anion on Cys 988 and an acyl-imidazole intermediate by Gln 991, which will react with any nucleophile, particularly carbohydrates and hydroxyl groups on target cellular surfaces.

It is *via* this TE moiety that nascent C3b molecules bind covalently to cellular surfaces in a process known as opsonisation. Amplification of complement activity is achieved by association of surface bound C3b and pro-enzyme fB, yielding the short-lived C3 convertase C3bBb of the AP (Half Life ($T_{1/2}$) = 90s). Conversion of subsequent C3 molecules to C3a and C3b stimulates the relevant immune response whilst amplifying the positive feedback loop of complement activation (Law & Dodds, 1997a), enabling elimination of both self and non-self target cells. It is the covalent coupling of C3b *via* this TE to target particles which is critical in the generation of local complement activation and amplification (Gros *et al*, 2008).

### 3.1.4 Derivatization of Complement C3 – Labelling with Bio-Reporters

The highly labile and conserved TE motif of the $\alpha_2$-M class of proteins is an extremely rare and distinctive post –translational modification, making these proteins theoretically amenable to selective, site specific, labelling (Cole *et al*, 2009).

Upon activation from C3 to C3b, although the TE becomes more exposed, and extremely reactive to many nucleophiles, this reactivity is very short lived, and the coupling efficiency of C3b to targets is greatly reduced, due to the shortened life span of the TE (Holm *et al*, 2012). The fast hydrolysis of the TE in C3b provides means to contain the potentially damaging reaction to the immediate proximity of the site of activation, however, this greatly reduces the potential of labelling the free cysteine of the newly cleaved TE with specific probes *e.g.* with nitroxide spin labels for EPR studies.

The previously discussed inherent ‘tick over’ mechanism of the AP of complement activation, whereby intact C3 is slowly hydrolysed at low level to C3($H_2O$), demonstrates that the TE bond of C3, although buried in the inactive form, is solvent exposed. The intact TE protein behaves as a typical protein TE, with an estimated half-life of hydrolysis of ~160 hours (Holm *et al*, 2012), and so it seems that this particular part of the pathway would be an ideal point for the incorporation of such probes.

This hydrolysis of C3 to C3($H_2O$) demonstrates that protection of the TE before activation from C3 to C3b is not absolute, with small nucleophiles capable of making similar conformational rearrangements in the native C3 (Cole et al, 2009).

C3 can be treated with small nucleophiles, such as methylamine, which are covalently incorporated into the protein, in this case forming C3(N). The TE is amino-lysed, and the metastable cysteine side chain of the TE bond becomes transiently available for chemical derivatization. This is demonstrated in Figure 3.3.

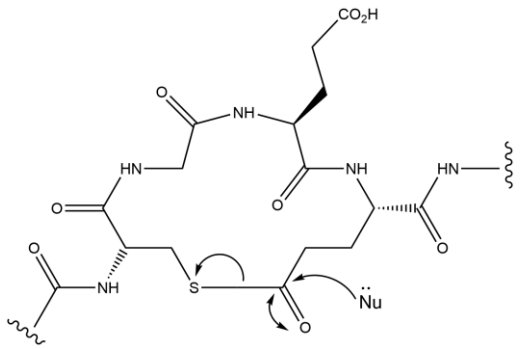



**Figure 3.3: Generalised Reaction Mechanism for the Reaction of a Nucleophile with the TE of C3.** This gives route for the modification of the intact C3 protein without need for activation.

This was demonstrated by (Cole et al, 2009) and their incorporation of nucleophiles with different functionalities (biotin, various small peptides, tags (V5 and FLAG) and fluorescein) into C3 *via* the TE. Breaking of the TE bond using small nucleophiles generates C3(N), which adopts a 'C3b-like' structure (Law & Dodds, 1997), with 'C3b-like' function (Isenman, *et al*, 1981), whilst the ANA domain remains bound, with interaction sites for fH and fI remaining exposed.

The TE of C3b has a $t_{1/2}$ of ~60$\mu$s, making derivitization with EPR spin labels particularly challenging. However, using C3(N) as an alternative, it is still possible to incorporate spin labels *via* the cysteine residue that becomes accessible following treatment with a nucleophile. Although the cysteine has to be derivatized relatively quickly, to prevent oxidation or disulphide bond formation, the potential labelling time is greatly increased, and should facilitate more efficient labelling of C3(N).

### 3.1.5 Regulators of Complement – fH

The efficacy of complement mediated immunity is dependent on maintaining a delicate balance between activation and regulation (Wu, *et al* 2010). Labelling of cells with complement component C3b is indiscriminate between pathogenic invaders and host cells (as reviewed by Pangburn, 2000) and so the system is tightly regulated by a number of proteins, which ensure C3 activation in the fluid phase is kept to a minimum, whilst deposition of C3b on cellular surfaces, and therefore further amplification of complement activation, is limited to pathogenic surfaces (as reviewd by De Córdoba & De Jorge, 2008)

FH is the chief regulator of complement C3, and circulates abundantly as a 155kDa single chain glycoprotein in the serum, at a concentration of ~300-800$\mu$g/mL (Sim & DiScipio, 1982), however recent estimates have been lowered to ~150-300$\mu$g/mL (Hakobyan *et al*, 2008). Upon binding to cellular membranes *via* the recognition of specific poly-anions (Pangburn, 2000), and to C3b, fH can modulate surface-associated and fluid-phase amplification of complement (Pangburn, *et al* 2000) (Sharma & Pangburn, 1996).

FH is composed of twenty homologous repeating domains, or complement control protein domains (CCPs), arranged in a 'beads on a string' like motif, each containing approximately 60 amino acid residues, and separated by short linkers of between 3 and 8 amino acids (as reviewed by Makou *et al*, 2013) (Pouw, *et al* 2015). Figure 3.4 shows a schematic representation of fH, showing the number of residues per CCP, and the number of linking residues.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| NUMBER OF RESIDUES PER CCP | 60 | 57 | 60 | 53 | 54 | 61 | 54 | 58 | 56 | 55 | 55 | 54 | 51 | 54 | 57 | 54 | 55 | 55 | 55 | 62 |
| NUMBER OF LINKING RESIDUES | | 4 | 4 | 4 | 4 | 4 | 3 | 5 | 3 | 4 | 6 | 6 | 8 | 7 | 5 | 4 | 4 | 4 | 6 | 3 |

**Figure 3.4: Linker lengths and residues in Complement fH**. Each CCP domain is labelled 1-20 on complement fH, with the number of residues found in each CCP stated underneath, as well as the number of linking residues between neighbouring CCP domains. Figure generated from data obtained from: **http://www.bionmr.chem.ed.ac.uk/bionmr/public_html/Residue_lengths.pdf**

It is these short linker sequences which introduce a degree of flexibility in the protein, and it is proposed that these linkers allow the protein to adopt a 'bent-back' arrangement, bringing the N' and C' termini in close proximity.

With the exception of CCPs 9, 14, and 17, which have not yet been solved, it has been found that all CCPs share a similar globular structure (Makou *et al*, 2012, Morgan *et al*, 2012), stabilized by two disulphide bonds per CCP module, formed between four conserved cysteine residues present in each CCP (between Cys I and Cys III, and between Cys II and Cys IV) (Barlow *et al*, 1991). Figure 3.5 shows the solution NMR structure of one CCP module (CCP 16), with the internal disulphides shown as sticks.

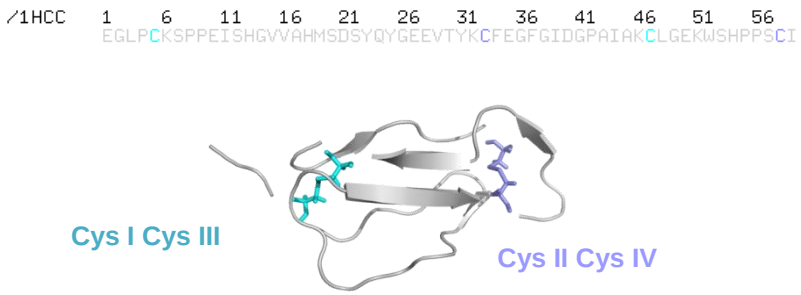



**Figure 3.5: NMR derived three- dimensional cartoon representation of CCP 16 of fH in solution.** The amino acid sequence of CCP 16 is shown, with the four invariant cysteine residues highlighted. In the structure the four invariant cysteine residues are represented as sticks, in the same colours as highlighted in the protein sequence. The structure was visualized using PyMOL (PDB_ID: 1HCC).

Presently, fH has not been crystallised in its entirety, this being accredited to its relatively large size, its ability to readily self-associate, forming dimers and higher order oligomers, the extent of glycosylation throughout the intact protein, and the potential flexibility introduced *via* the inter-domain linkers between neighbouring CCPs (Rodriguez *et al*, 2014).

Efforts have been made to obtain some degree of information on the entire fH structure, including solution scattering, analytical centrifugation and transmission electron microscopy (DiScipio, 1992) (Perkins *et al*, 1991) (Okemefuna *et al*, 2009) which point to an overall flexible structure that is, nonetheless, more compact, and predicted to be approximately half as long than would be the case if the protein existed in a fully extended conformation (Aslam & Perkins, 2001) (Makou *et al*, 2010). The

bent-back structure of fH is shown schematically in Figure 3.6, and the binding sites for C3b highlighted.

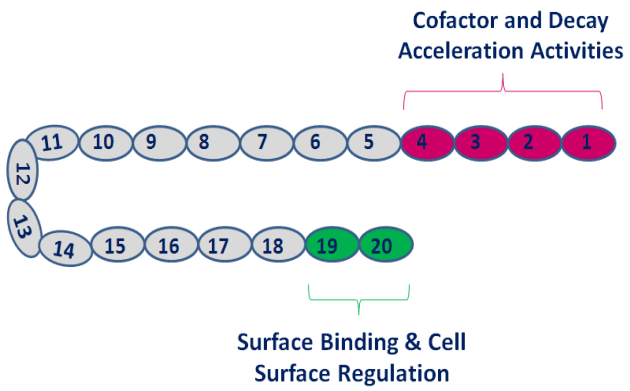



**Figure 3.6: Schematic representation of a 'bent back' fH, showing distinct interaction sites for C3b.** (N' terminal CCPS 1-4, responsible for cofactor and decay acceleration activities shown in pink. C' terminal CCPs 19-20, responsible for cell surface regulation shown in green). In this conformation the N and C' termini of fH are in close proximity.

Full length fH possesses interaction sites for C3b allowing for control of C3b deposition, with distinct function attributed to N' terminal CCPs 1-4 and C' terminal CCPs 19-20.

Structural and functional analyses have shown that domains 1-4 assert their function *via* down-regulation of the feedback loop of activation, by accelerating the decay of the C3 convertase (Gigli *et al*, 1979) (Mullick *et al*, 2005*)* as well as acting as a cofactor for the fI mediated cleavage of C3b to its inactivated form, iC3b (Gordon *et al*, 1995). Furthermore fH1-4 accelerates the decay of the fluid phase convertase C3bBb, limiting the conversion of C3b from C3 (Bhattacharjee *et al*, 2010).

Sugar molecules, such as glycosaminoglycans (GAGs)and sialic acids, provide a diverse and complex mechanism by which the complement system can not only identify bacteria and other pathogens, but also identify host cell surfaces that require protection (Langford-Smith *et al*, 2015). The function of host discrimination, assigned to the C' terminal CCPs 19 and 20 is made possible due to the specific binding of this region to the C3d portion of C3b, and to poly-anionic surfaces, such as GAGs (Clark *et al*, 2013) and sialic acids (Kajander *et al*, 2011) (Blaum *et al*, 2014).

FH engages most effectively with C3b or C3bBb convertases, when bound to surfaces carrying such poly-anion markers (Ferreira *et al*, 2013)**.** A High level of poly-anionic markers on self surfaces results in a high level of bound fH in the vicinity, and low levels on pathogenic surfaces, which lack these distinct markers. It is this characteristic

which allows discrimination by the innate immune system and activation of a relevant immune response (Morgan *et al*, 2011).

The importance of fH in maintaining a well-balanced immune response is reflected in the increasing number of diseases found to have strong association with mutations and polymorphisms in the gene encoding fH (Wu *et al*, 2009).

Irregular complement activation has association with a number of inflammatory conditions and autoimmune diseases (De Córdoba & De Jorge, 2008). Tight regulation of the complement response at a cellular level and in the fluid phase is of critical importance in providing selectivity against foreign bodies, and preventing complement mediated tissue damage. This is reflected in the number of medical conditions, such as age related macular degeneration (AMD) and atypical haemolytic uremic syndrome (aHUS), which occur as a result of abnormal or incomplete complement activation or consumption (Pechtl *et al*, 2011). Therapeutic targeting of fH is considered of vital importance for the treatment of such conditions, associated with loss of complement control (Wu *et al*, 2009).

### 3.1.6 The fH-C3b Interaction

The fH1-4 interaction with C3b has been well characterised, and the crystal structure, shown in Figure 3.7A solved (Wu *et al*, 2009), however, the interaction of C3b with fH19-20 is not as well defined, despite a small cluster of disease associated mutations being present in these modules, together with a polyanion-binding site that is necessary for self-surface recognition by fH (Morgan *et al*, 2011).

Work by Morgan *et al*, has shown that a complex formed between the C3d segment of C3b (TED) and fH19-20 emulates the interaction between fH19-20 and C3b in its entirety. Crystallisation of the fH19-20:TED complex resulted in three distinct structures within the crystal. In Figure 3.7 (BI, BII & BIII) the three potential conformations are shown.

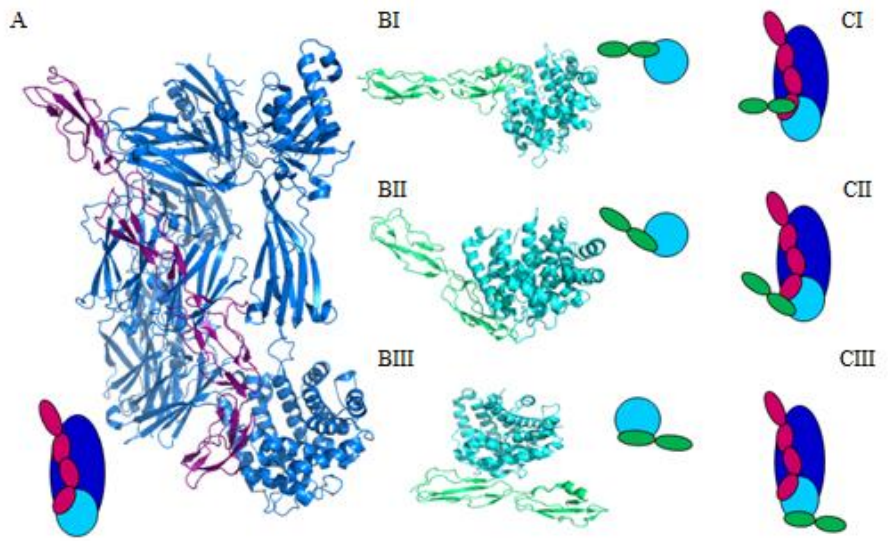



**Figure 3.7: Crystal Structures of fH bound to C3b, shown as cartoons, with accompanying schematics.** (Crystal structures visualised using PyMOL). (A) Complement C3b (blue) bound to fH1-4 (purple). (Wu *et al*, 2009) (PDB_ID: 2WII). (BI, BII, BIII) Three conformations of the TED of complement C3b (cyan), bound to fH19-20 (green). (Morgan *et al*, 2011) (PDB_ID: 3OXU). (CI, CII, CIII) Schematic representations of structure (A) superimposed onto each of the three fH19-20:TED complexes (BI, BII, BIII).

Superposition of each of the three potential fH19-20:TED structures onto the crystal structure of the fH1-4-C3b complex, suggests that orientation BIII corresponds to the physiologically crucial interface between the C' terminal CCPs 19-20 of fH and C3b. This is shown in Figure 3.8.

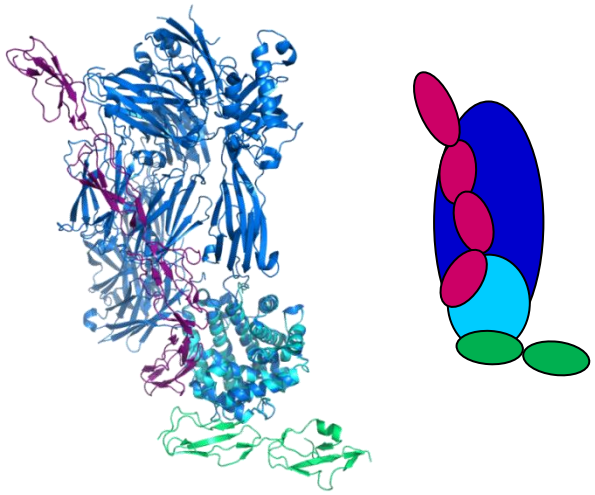


**Figure 3.8: Superposition of the X-Ray Crystal structure of fH1-4 bound to C3b, with fH19-20 bound to the TED of C3b.** Visualised using PyMOL and adapted from Morgan *et al*, 2011.

In such a position, fH does not experience any steric clashes when this interface is extrapolated onto the corresponding TED of C3b, nor does it impede or sterically hinder the functionally crucial TED interface, *via* which C3b is deposited on cellular surfaces. The alternative complexes are not possible, as these complexes would involve substantial steric clashes between the fH 1-4 and fH19-20 binding sites on C3b.

Superposition of the ternary complex places fH19-20 immediately adjacent to CCP 4 of fH1-4, with both fragments occupying distinct regions on C3b, consistent with simultaneous binding, leaving CCP 20 exposed and well positioned to interact with poly-anionic surface markers in its role in cellular discrimination. In the modelled ternary complex of fH1-4:fH19-20:C3b, the C' terminus of CCP 4 and the N' terminus of CCP 19 lay in close proximity, on the same side of C3b – but most importantly they do not clash. This conformation is consistent with a 1:1 complex of C3b:fH, in which intervening CCPs form a compact bent-back structure, or with a 1:2 fH:C3b complex, whereby fH extends and bridges between two C3b molecules.

The mid-region of fH, on average, contains longer amino acid linkers between CCPs (Schmidt *et al*, 2010) and so it is proposed that it is this region, in particular between CCPs 12 and 13 (Schmidt *et al* , 2010) that acts as a flexible connection between the two ends of the protein, introducing a bend in the middle of the single chain polypeptide, and allowing it to bend back on itself. Although this hypothesis is desirable, it offers incomplete experimental evidence on the architecture of fH, particularly in the context of C3b. The above superposition of both N' and C' termini of fH with C3b can be coupled with other structural data obtained on more CCPs of fH using SAXS modelling, NMR spectroscopy and X-Ray Crystallography (PDB_ID: 2WII (Wu *et al*, 2009), 3OXU (Morgan *et al*, 2011), 3SW0 (Morgan *et al*, 2012), 1HFH (Barlow *et al*, 1993), 2UWN (Prosser *et al*, 2007), 4B2R, 4B2S (Makou *et al*, 2012), 2KMS (Schmidt *et al*, 2010). Figure 3.9 illustrates a potential model for the fH:C3b structure, using the structures of already defined fH fragments.

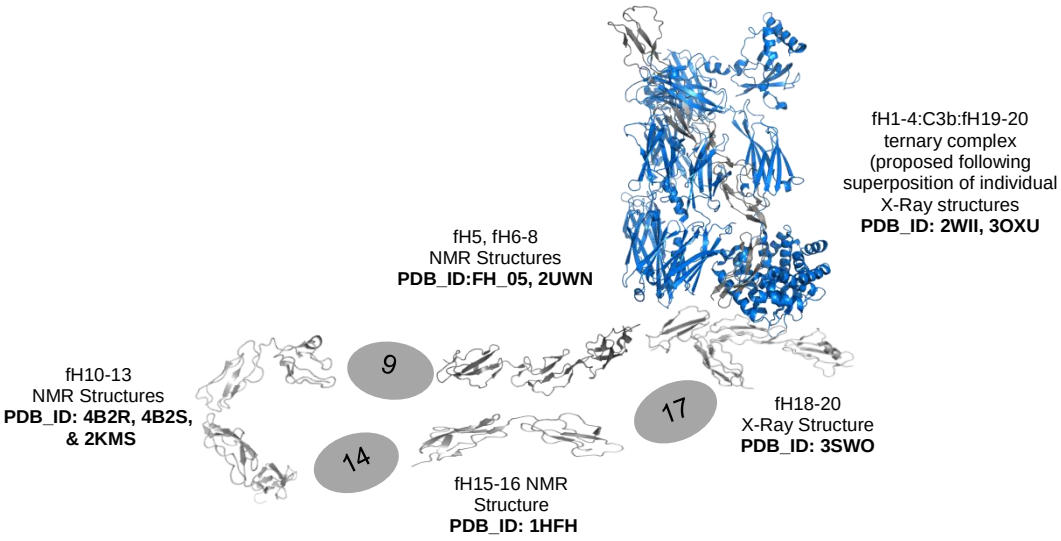



**Figure 3.9: Model for the interaction of full length fH with C3b.** The model is based on structures of individual CCPs, multiple CCPs, and multiple CCPs bound to C3b, as determined by X-Ray Crystallography and NMR Spectroscopy**.** In the model complex, the structures of fH CCP5, fH CCPs 6-8, fH CCPs 10-13 and fH CCPs 15-16 were solved by NMR Spectroscopy. The structure of fH CCPs 18-20 was solved by X-Ray Crystallography, as is the ternary fH1-4:fH19-20:C3b**.**

Combining this structural data, it is proposed that fH binds C3b in a 1:1 ratio, bending back on itself, so that the N' terminus is perfectly placed to carry out its cofactor and decay acceleration activities, whilst the C' terminus is primed for recognition of the TED of C3b, and poly-anionic carbohydrates. It is these functions which stop the unwarranted amplification of complement activity, whilst ensuring self surfaces can be distinguished from bacterial. Determining whether or not the isolated domains of fH can simultaneously bind C3b will help shed some light onto this complex problem, and perhaps provide further insight into the mechanisms of fH, and consequently C3b.

### 3.1.7 Aims of this work

To prove the hypothesis that at a given moment in time, fH can adopt a 'bent-back' structure with both N' and C' termini engaged simultaneously with a single molecule of C3b, this must be investigated at a structural level. Adopting higher resolution structural techniques, such as X-ray crystallography or NMR spectroscopy, would be the preferred strategy for probing such interactions, however, the nature of this complex is outwith the limits of these particular techniques.

Therefore EPR spectroscopy can be employed to measure nm distances to investigate this large bio-macromolecular complex, as the flexibility, as well as size of the interacting proteins, is not limited by this technique.

This work is done in collaboration with Professor Paul N. Barlow, University of Edinburgh. Using FRET, Barlow's group (Pechtl *et al*, 2010) positioned fH1-4 bound to C3b with CCP module 1 furthest from, and CCP4 closest to the TE domain, placing subsequent fH modules nearest any other surface to which C3b is bound. This data validates that of the proposed fH1-4/C3b crystal structure (Wu *et al*, 2009).

Consequently, it is the aim of this study to use EPR in a similar manner, using both regions of fH, (1-4 and 19-20) which bind C3b to address the problem which other structural methods cannot answer. Those fH fragments which bind C3b, namely fH1-4 and fH19-20 will be expressed recombinantly using the *P. pastoris* expression system, following cysteine substitution mutagenesis to incorporate cysteine residues at sites of interest for EPR studies. Two different fH19-20 mutants, G1107C and R1210C, will be used for this study, alongside the fH1-4 K247C mutant. The R1210C mutant was provided as purified protein by Dr Andy Herbert, University of Edinburgh, which was then spin-labelled, whilst the fH19-20 G1107C fragment was cloned and expressed specifically for this work.

C3b cannot be made recombinantly, and so will be purified directly from human plasma, before the eponymous TED is utilized as a spin labelling site.

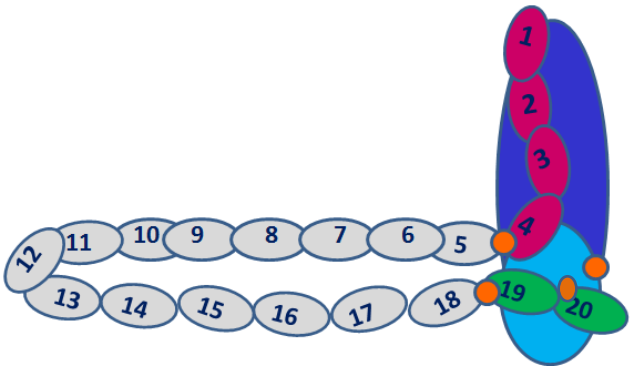
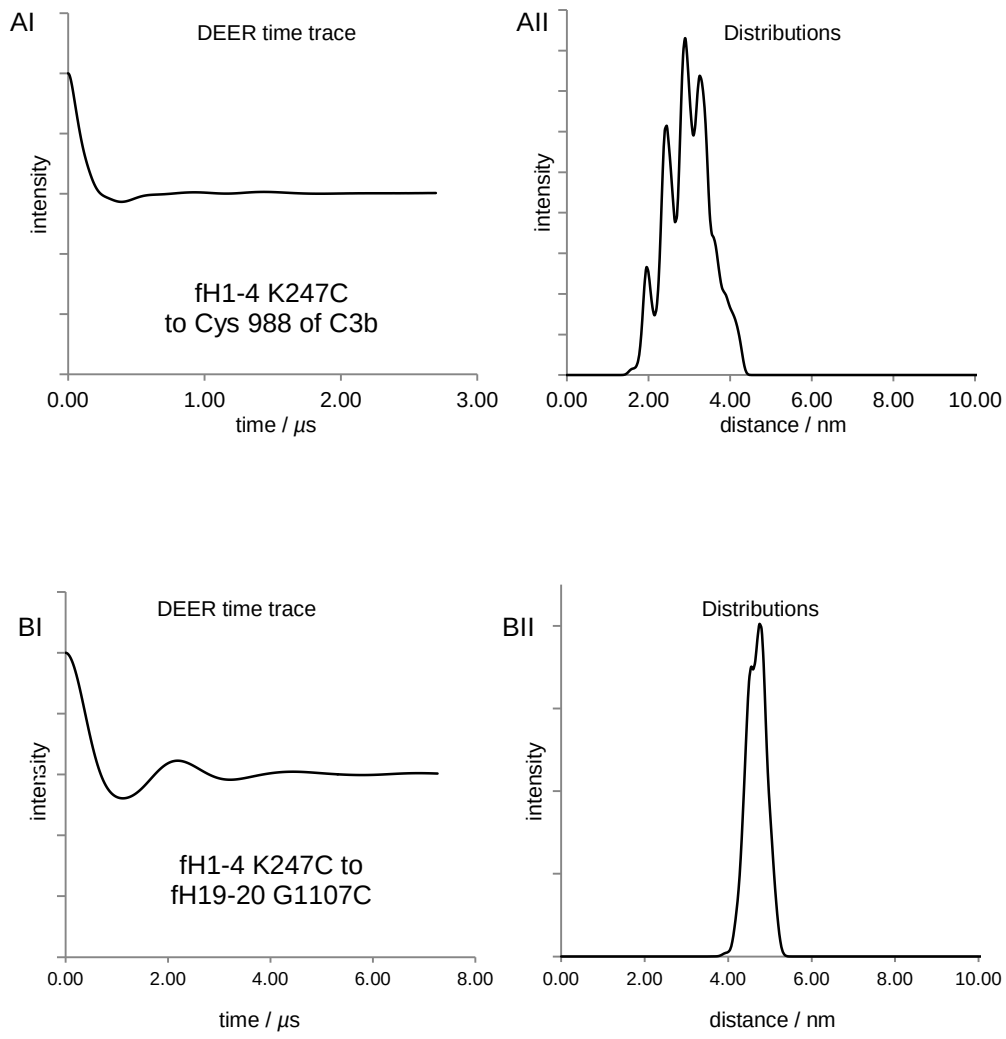



**Figure 3.10: Schematic representation of fH bound to Complement C3 – as shown in Figure 3.9.** Potential spin labelling sites are represented as orange circles. Of the 20 fH CCPs, the fragments highlighted in red and green represent the fH1-4 and fH19-20 fragments, respectively. C3b is shown in blue, with the TED domain coloured light blue. In this conformation, one molecule of fH is bound to one molecule of C3b at any given point in time, and using SDSL, EPR will be implemented to prove this hypothesis.

Figure 3.10 shows a schematic representation of the fH:C3b interaction, with spin labelling sites indicated. These aims will prove particularly challenging due to the large nature of the complex, and the presence of cysteine residues throughout the fH fragments.

## 3.2 Simulation of the DEER Experiment for the fH:C3b Complex Using MMM

Using MMM, the DEER time traces and distance distributions were obtained between each of sites for SDSL on fH, and the TE of C3b (Figure 3.11). All give distances that should be measurable by DEER, although distances to the fH19-20 R1210C may be on the long side for precise measurements, but therefore act as a good gauge of the correctness of the model.



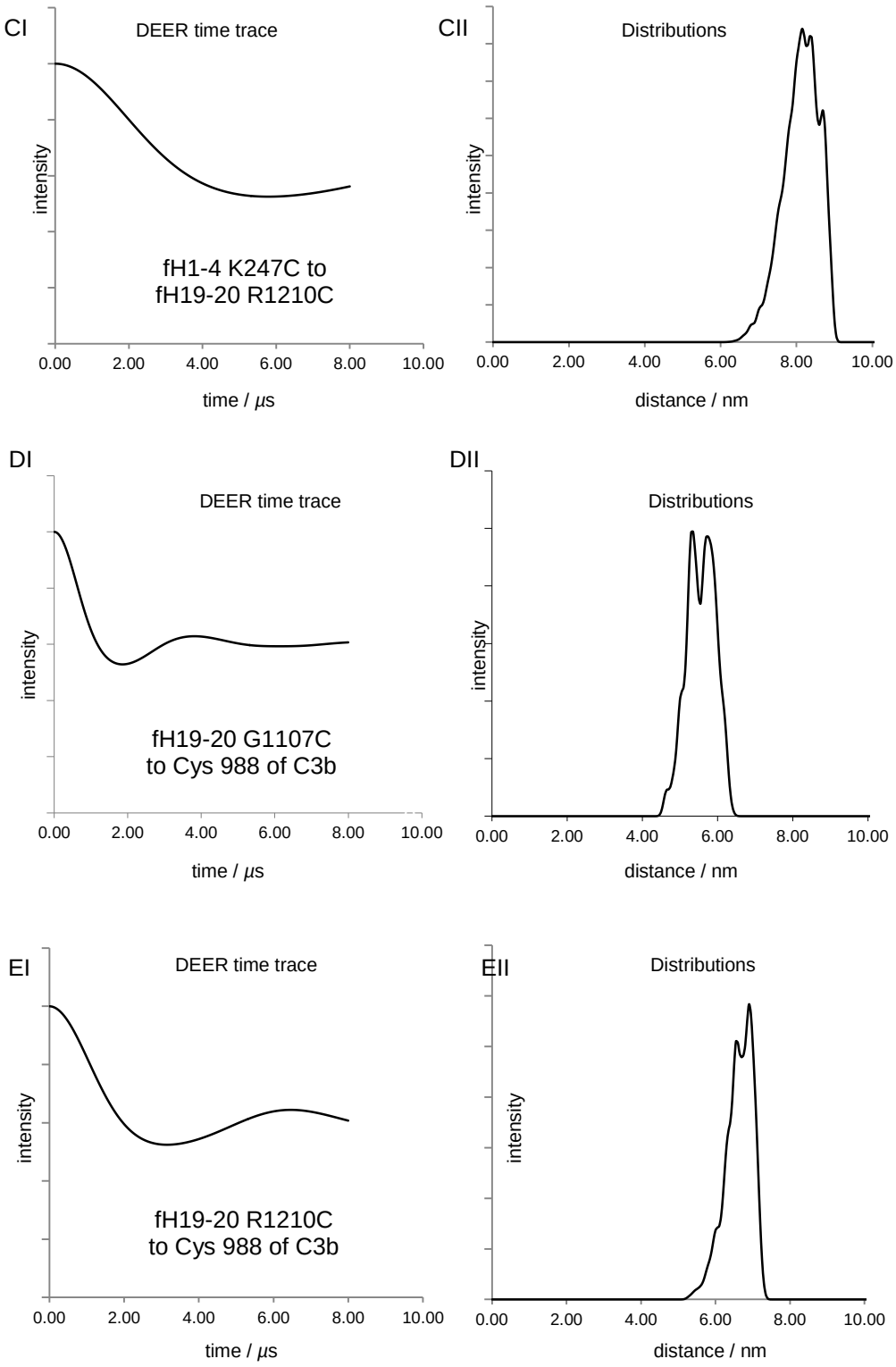



**Figure 3.11: MMM Simulations of the DEER Time Traces (I) and Distance Distributions (II) for fH/C3b complexes.** (A) fH1-4 K247C to the Cys988 of C3b, (B) fH1-4 K247C to fH19-20 G1107C (C) fH1-4 K247C to fH19-20 R1210C, (D) fH19-20 G1107C to the Cys988 of C3b (E) fH19-20 R1210C to the Cys988 of C3b.

## 3.3 Generation of Spin-Labelled fH1-4 K247C Fragment for EPR Studies, Using the *P. pastoris* Expression System

The first step in forming the complex was the generation of a fragment of fH containing just CCPs 1-4, however following cloning and expression tests this construct, which shall be known as fH1-4 K247C(A) showed no expression.

### 3.3.1 Mutagenesis

Following SDM (2.1.1) to incorporate a C' terminal cysteine into the fH1-4 K247C construct for SDSL purposes, the mutated fH DNA fragment was ligated into the pPICZαB vector, before the plasmid was transformed into *P. pastoris* for protein expression. Figure 3.12 shows the sequence chromatogram of the cysteine mutation, which can be compared to the wt sequence shown in Appendix A2.

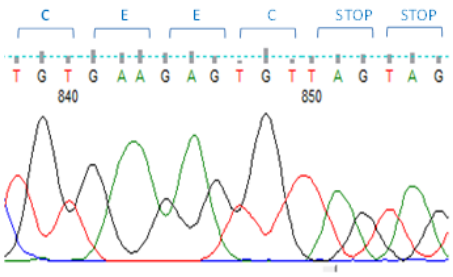



**Figure 3.12: Sequencing Chromatogram confirming the presence of a cysteine residue in the fH1-4 construct – K247C.**

### 3.3.2 Small Scale Test Expression of fH K247C

Following successful transformation of the K247C fragment into *P. pastoris*, multiple colonies were screened in small-scale expression volumes (initial volume 100mL) in order to identify the most highly expressing clone. At the transformation stage, cells were plated onto increasing concentrations of Zeocin™, as those which successfully grow on higher concentrations during selection contain a higher copy number of plasmid, and are therefore most likely to give better yield. Two clones A and B, from a 500$\mu$g/mL Zeocin™ plate underwent expression tests using two different induction media, namely BMM and BMMY, in order to ascertain which gives better protein expression. After harvesting, the cell pellet was discarded and the supernatant, which

contained the secreted protein, filtered (0.22$\mu$m) and the pH adjusted to 7.5. Concentrated supernatants were analysed by SDS-PAGE, as shown in Figure 3.13.

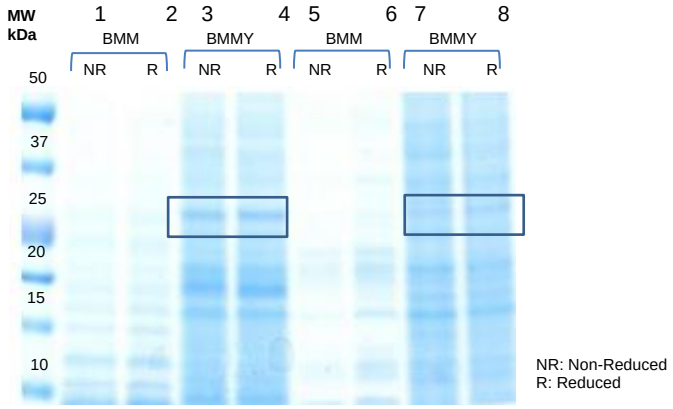



**Figure 3.13: SDS-PAGE Analysis of small scale test expression of the fH1-4 K247C mutant.** Sample lanes 1-4 and 5-8 show expression of clones A & B respectively, induced in both BMM and BMMY media.

The fH1-4 K247C should run at approximately 30kDa following SDS-PAGE analysis, and in sample lanes 3 and 4, and 7 and 8 there is indication of a band (circled) corresponding to this expected molecular weight. As stated prior, the samples were run in both their reduced and non-reduced forms, and, as so, there should be a notable difference between the observed molecular weights of these samples. When reduced, the fH1-4 K247C runs at a slightly higher molecular weight, due to a reduction of the internal disulphide bonds in fH, forcing it to migrate more slowly through the gel. It can therefore be assumed that any bands corresponding to the expected molecular weight are most likely yeast proteins found in the concentrated, un-purified supernatants, as the expected banding pattern is not observed.

### 3.3.3 The need for an Affinity Tag

fH1-4 is susceptible to proteolysis and degradation over time. Problems with its stability mean its expression and purification in high yield proves problematic (Hocking, 2008). Codon optimisation of the fH gene allows for optimal expression using the *P. pastoris* expression system, however, further methods must be employed to optimise expression and yield.

Previous work carried out by Dr. David Kavanagh, University of Edinburgh, found that inclusion of a C' terminal hexa-histidine affinity tag as well as an N' terminal c-

Myc-epitope (EQKLISEEDL) onto a non-codon-optimised fH1-4 fragment helped stabilise the construct, with evidence of reduced proteolytic degradation, as well as the additional advantage of ease of purification using IMAC, thus minimising sample handling (Pechtl *et al*, 2011).

### 3.3.4 Mutagenesis – Incorporation of a C' Hexa-Histidine Tag onto fH1-4 K247C

A C' terminal histidine tag was incorporated onto the fH1-4 K247C construct using SDM (Figure 3.14) in hopes of increasing yields and facilitating easier purification.

From this point onwards, fH1-4 K247C refers to the K247C mutant with the C' terminal hexa-histidine tag. Following successful transformation into *P. pastoris*, as before, small scale expression tests were carried out in order to identity the clones, if any, which gave best expression.

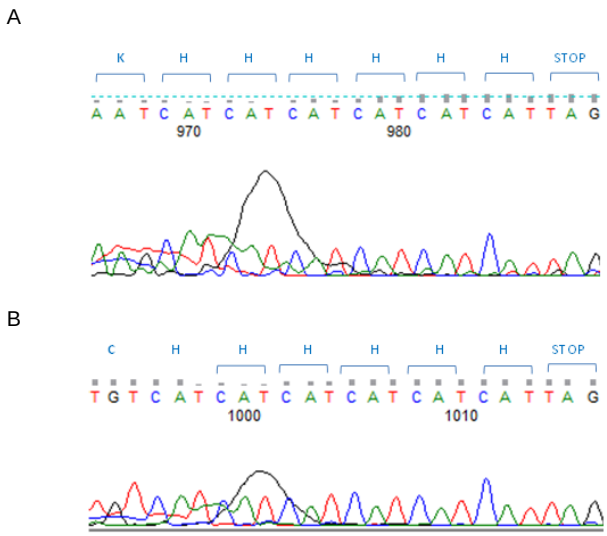



**Figure 3.14: Confirmation of the presence of a hexa-Histidine tag onto fH1-4 K247C.** (A) shows sequencing chromatogram for the wt fH1-4 protein, whilst (B) shows the sequencing chromatogram for the modified fH1-4 K247C mutant, with the inclusion of the affinity tag.

### 3.3.5 Small Scale Test Expression of fH1-4 K247C

Small scale test expressions were used to establish which clones expressed best, using three colonies from each of the three selection plates (200, 300 and 500$\mu$g/mL Zeocin$^{TM}$). From the test expressions of fH K247C, before the addition of the hexa-histidine tag, although there was no expression of fH, SDS-PAGE analysis shows that endogenous *P. pastoris* proteins showed better expression levels.

Consequently, the enriched media BMGY was used, with BMMY used for induction, in order to optimise protein expression and improve yield. After harvesting, the cell pellet was discarded and the supernatant, which contained the secreted protein, was filtered (0.22 $\mu$m) and the pH adjusted to 7.5. SDS-PAGE analysis (Figure 3.15) was carried out on the un-concentrated supernatants.

From SDS-PAGE analysis, it is clear that the incorporation of a hexa-histidine tag onto the C' terminus of the fH1-4 K247C construct is sufficient to induce expression of the fragment.

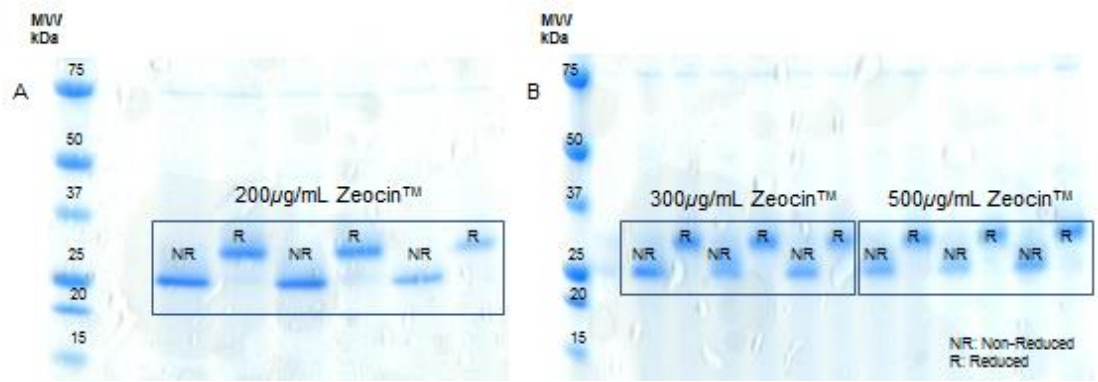



**Figure 3.15: SDS-PAGE Analysis of fH1-4 K247C Expression Tests.** (A) supernatants from the expression of three different fH1-4 K247C clones, selected on 200$\mu$g/mL Zeocin$^{TM}$, run both non-reduced and reduced (B) supernatants from the expression of three different fH1-4 K247C clones, selected on 300 and 500$\mu$g/mL Zeocin$^{TM}$.

### 3.3.6 Large Scale Expression of fH1-4 K247C

As described in 2.4.3, the fH1-4 K247C fragment was expressed on a larger scale of 8 x 500mL (initial culture volume). Following culturing, cells were harvested, and a pool was made of all filtered supernatants and SDS-PAGE analysis carried out, as shown in Figure 3.16A. To confirm that the expression bands visualised by SDS-PAGE analysis were indeed fH1-4 K247C, western blotting was carried out as described in 2.7.2 using a primary polyclonal antibody, specific for the detection of purified human fH. Results are shown in Figure 3.16B.

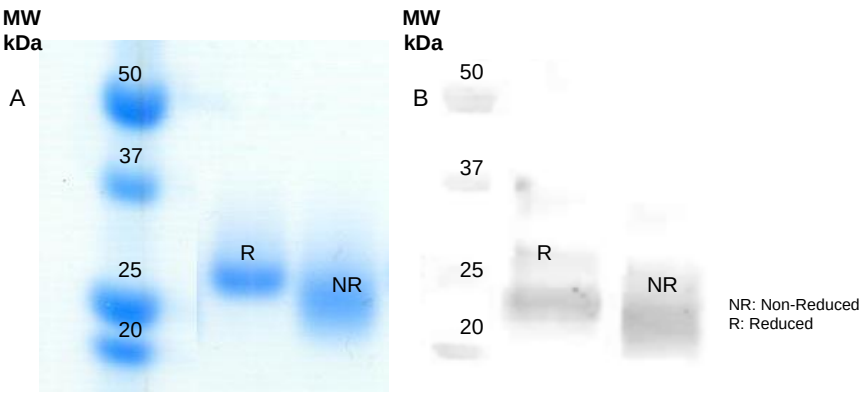



**Figure 3.16: SDS-PAGE and Western Blotting Analysis of Expression from a single fH1-4 K247C clone.** (A) SDS-PAGE analysis of the pooled supernatants. (B) Western blotting analysis of the pooled supernatants, probed with a polyclonal antibody specific for the detection of purified human fH.

From both SDS-PAGE analysis and western blotting, it was confirmed that that believed to be fH1-4 K247C is so.

### 3.3.7 Purification of fH1-4 K247C

#### 3.3.7.1 IMAC

IMAC was used as a first catchment step in the purification of fH1-4 K247C due to the incorporation of the hexa-histidine tag. The application of the protein to, and elution from the Ni-NTA Superflow Cartridge was performed using peristaltic pump. Purification was monitored by SDS-PAGE analysis of the flow-through (FT) from the column, the wash fraction, and elution fractions (E1-8), as shown in Figure 3.17.

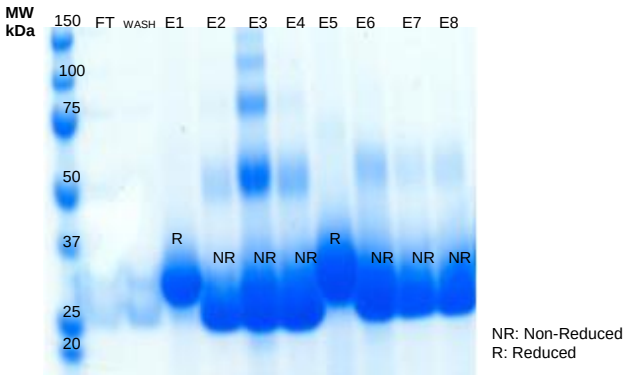



**Figure 3.17: SDS-PAGE Analysis of the Elution of fH1-4 K247C from the Ni-NTA Superflow IMAC Cartridge.** A sample was taken of the material that flowed through the column. A washing step was carried out before the protein was eluted in eight 1mL fractions (E1-E8).

### 3.3.7.2 Anion Exchange Chromatography (AEC) Tricorn™ Mono Q 4.6/100 PE

A common step in the purification of proteins which have been expressed under acidic conditions, such as those expressed in *P. pastoris*, is cation exchange chromatography (CEC). Usually, CEC is performed at least one pH unit below the isoelectric point (pI) of the protein of interest. The pI of fH1-4 K247C is ~ 5.2, meaning CEC would need to be performed at ~ pH4, which nears the limit of the resin (Sepharose is stable in the pH range 4-9, Turkova *et al*, 1978). AEC was carried out instead, at a pH of 9. Figure 3.18 shows the FPLC elution profile for the elution of fH1-4 K247C from the Mono Q column and the corresponding SDS-PAGE analysis.

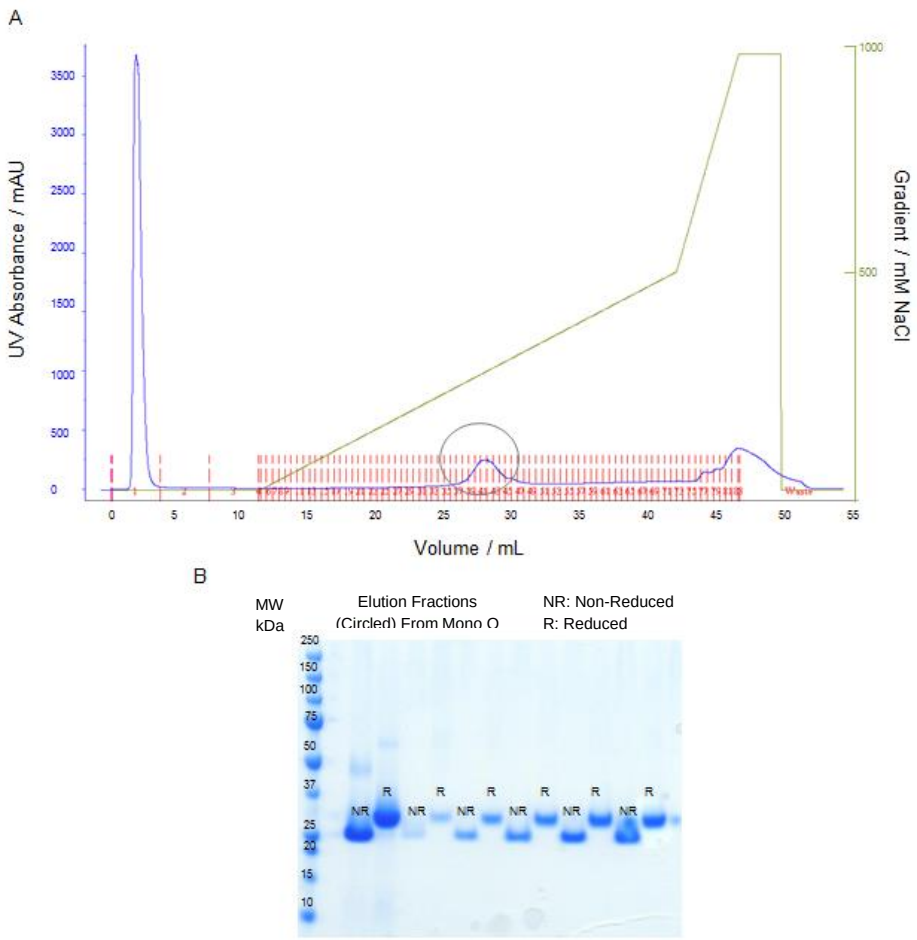



**Figure 3.18: Anion Exchange chromatography of fH1-4 K247C.** (A) Elution profile of fH1-4 K247C from the MonoQ column and (B) resulting SDS-PAGE analysis of fractions run within the circled peak.

#### 3.3.7.3 Size Exclusion Chromatography

From the SDS-PAGE analysis of the elution from the Mono Q, there appears to be some dimer in several of the non-reduced fractions, therefore the pooled fractions were treated with a final concentration of 250$\mu$M TCEP, before SEC was performed as a final step. Shown in Figure 3.19 is one SEC run from the pooled fractions, with resulting SDS-PAGE analysis

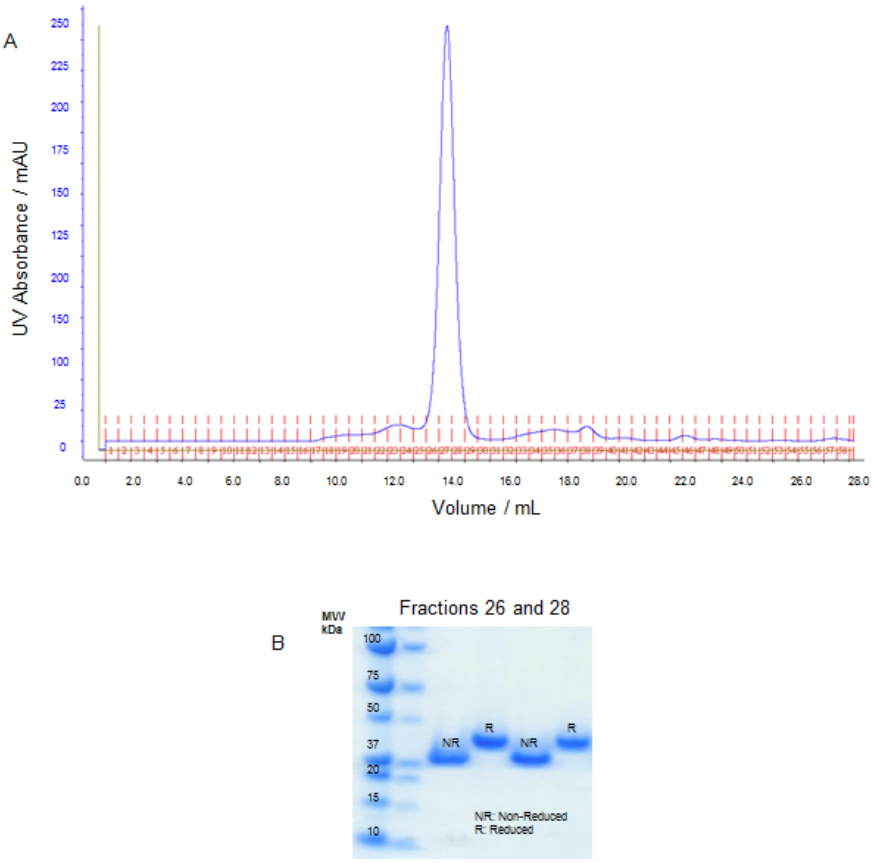



**Figure 3.19: Size Exclusion Chromatography of the fH1-4 K247C Fragment.** (A) Elution profile of the fH1-4 K247C construct and (B) resulting SDS-PAGE analysis of relevant fractions.

## 3.4 Generation of a Spin-Labelled fH19-20 Fragment for EPR Studies Using the *P. pastoris* Expression System

### 3.4.1 Mutagenesis

Following SDM to incorporate an N' terminal cysteine into the fH19-20 construct for SDSL purposes, the mutated fH DNA fragment was ligated into the pPICZαB vector, before the plasmid was transformed into *P. pastoris* for protein expression. Figure 3.20

shows the sequence chromatogram of the cysteine mutation, alongside that of the wt construct.

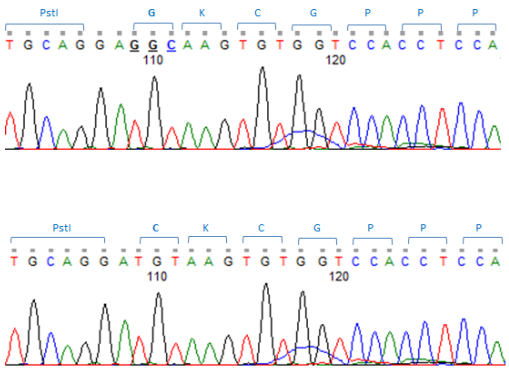



**Figure 3.20: Confirmation of the presence of a cysteine residue in the fH19-20 construct, G1107C.**

### 3.4.2 Small Scale Test Expression of fH19-20 G1107C

Nine different clones were selected which grew successfully on a range of different antibiotic concentrations, and screened in small-scale expression volumes in order to identify the most highly expressing clone. After harvesting, the cell pellet was discarded and the supernatant, which contained the secreted protein, was filtered (0.22$\mu$m) and the pH adjusted to 7.5. Concentrated supernatants were analysed by SDS-PAGE, as shown in Figure 3.21.

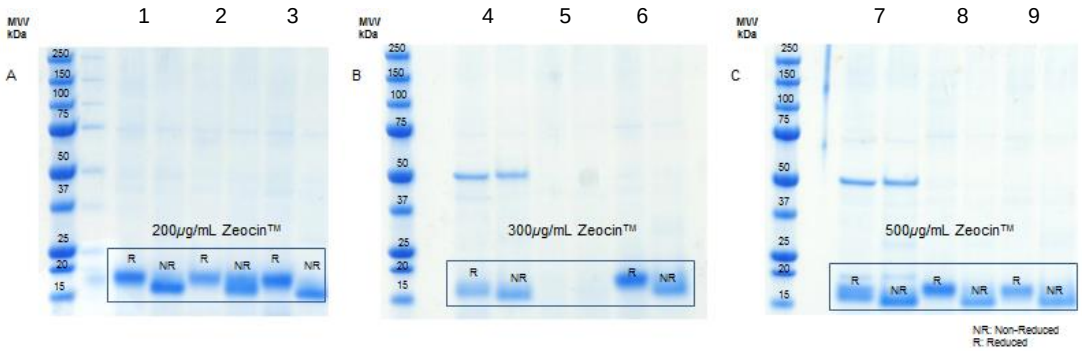



**Figure 3.21: SDS-PAGE Analysis of fH19-20 G1107C Expression Tests.** (A) supernatants from the expression of three different fH19-20 G1107C clones, selected on 200$\mu$g/mL Zeocin$^{TM}$ (B) supernatants from the expression of three different fH19-20 G1107C clones, selected on 300$\mu$g/mL Zeocin$^{TM}$ (C) supernatants from the expression of three different fH19-20 G1107C clones, selected on 500$\mu$g/mL Zeocin$^{TM}$.

Unlike the fH1-4 construct, there is no need for the addition of tags to fH19-20 for stability. However, in Figure 3.21B, it is clear that of the three clones which were selected with 300$\mu$g/mL Zeocin$^{TM}$, only two actually showed any level of protein expression. This result reinforces the necessity of screening multiple clones for expression, before large scale expressions/fermentations were performed. Of those clones which expressed, the levels of expression seemed consistent across all selection conditions. Protein expression appears consistent across all antibiotic concentrations. Those clones which express following selection with higher antibiotic concentrations generally produce more protein (Invitrogen EasySelect Manual), therefore clones 7, 8 and 9 would appear to be the best candidates for large scale expressions. Due to the high molecular weight protein that appears following expression from clone 7, this clone was rejected. This protein is likely endogenous yeast proteins. Clones 8 and 9 were used to make glycerol stocks, which were subsequently used for large scale expressions.

### 3.4.3 Fermentation of fH19-20 G1107C

As the fH19-20 G1107C construct shows good levels of expression on a small scale, it was decided to perform a 10L fermentation, to yield enough protein for all experiments required for this body of work.

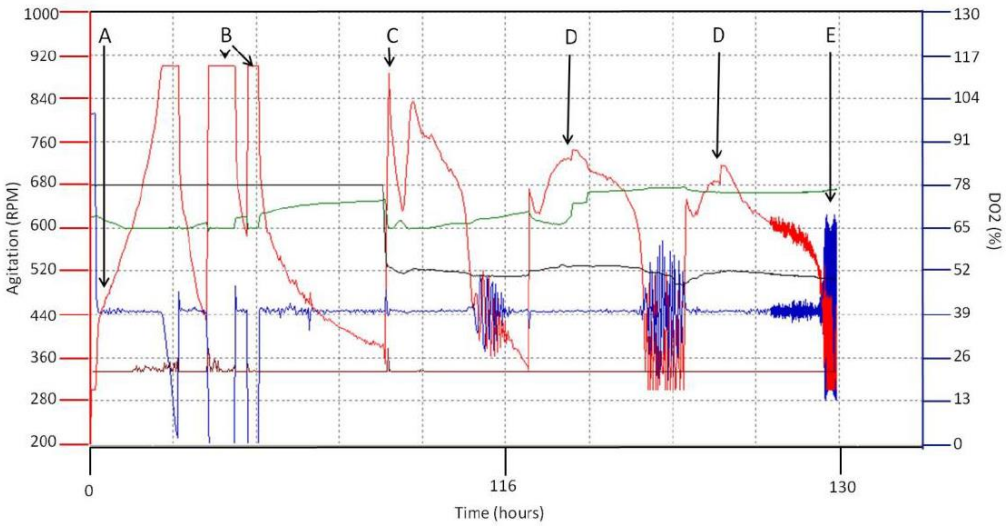



**Figure 3.22: Example Fermentation Log from the Bioflo4500 Vessel.** Throughout the fermentation process the agitation (red), pH (green), temperature (black) and percentage dissolved oxygen are displayed on the vessel, indicating when glycerol and MeOH feeds should be carried out.

During the course of the fermentation procedure, the agitation rates, levels of dissolved oxygen, temperature, and pH were monitored. Shown in Figure 3.22, is an example of the Fermentation Log obtained from the BioFlo4500 fermentation vessel.

Upon addition of the cells (A) to the fermentation vessel, a spike in agitation occurred. The yeast have begun consumption of the initial glycerol available to them, and the spike in agitation occurs as a result of maintaining oxygen levels in the vessel at 40%. Once all of the available glycerol was consumed, additional glycerol feeds were performed (B) in order to bulk cell mass.

At this point the temperature in the vessel was reduced from 30$^{o}$C to 15$^{o}$C before MeOH induction, which acts to decrease proteolysis. MeOH was then added (C) and a spike in agitation confirms consumption of the MeOH by the *P. pastoris*. Additional MeOH feeds followed (D) over a 72 hr period, before cells were harvested (E) by centrifugation.

### 3.4.4 Purification of fH19-20 G1107C Following Fermentation

#### 3.4.4.1 SP-Sepharose FastFlow$^{TM}$ CEC Resin

Following harvesting of cells, the crude supernatant was diluted by a factor of 10, in order to reduce the conductivity of the supernatant. The addition of PTM1 salts during the fermentation process increases the conductivity, and so diluting the supernatant accordingly, promotes binding of the target protein to ion exchange chromatography (IEC) resins, which are used throughout the purification process. EDTA and PMSF were added to the diluted supernatant to a final concentration of 1mM and 0.5mM respectively, in order to reduce protease activity, and the pH of the supernatant was adjusted to ~ pH5.5.

The supernatant (~80 L) was flowed over a XK 26/20 column, packed with 50mL of SP-Sepharose Fastflow$^{TM}$ resin, pre-equilibrated in non-salt buffer, using a peristaltic pump and maximum flowrate. Following application of the diluted supernatant to the column, the protein is eluted using a linear 1M NaCl gradient. Figure 3.23 shows the resulting elution profile and corresponding SDS-PAGE analysis of relevant fractions.

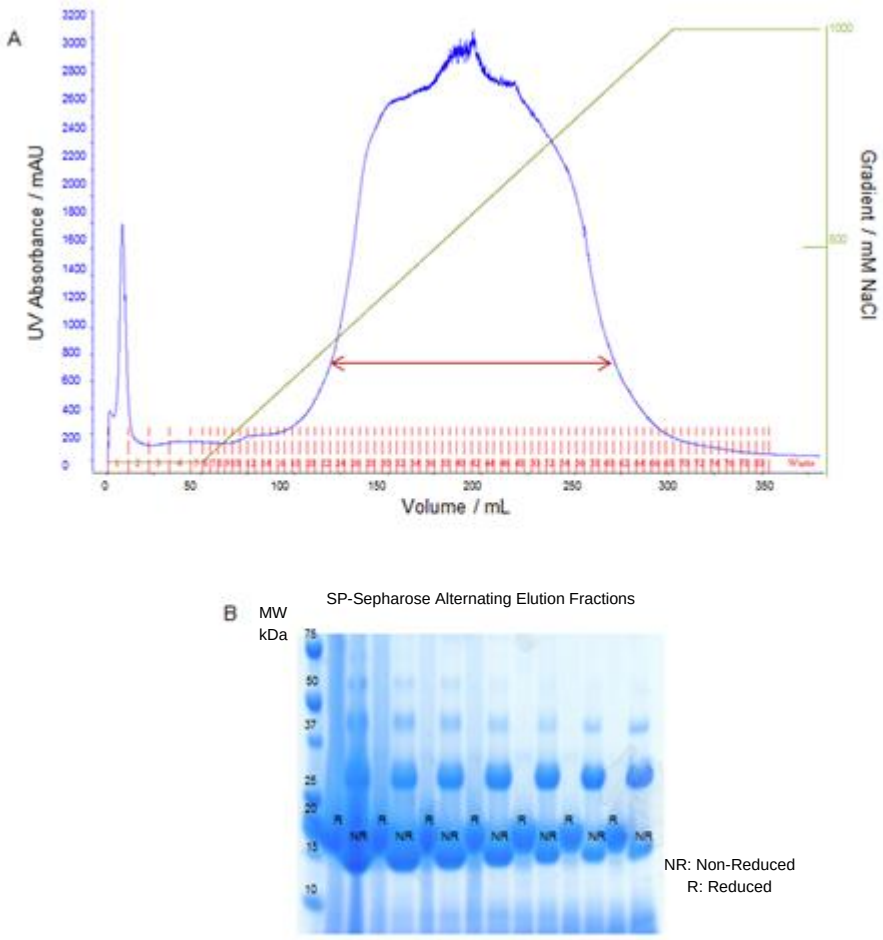



**Figure 3.23: Cation Exchange Chromatography of fH19-20 G1107C from 10 Litre Fermentation.** (A) Elution profile and (B) resulting SDS-PAGE analysis. Alternating fractions were run under the large peak, indicated by the red arrow in (A), with fractions run in both their reduced and non-reduced form.

Following SDS-PAGE analysis, those fractions which were found to contain fH19-20 G1107C were pooled. It was estimated that ~3g of the recombinant protein had been expressed.

### 3.4.5 Selection of an Appropriate Reducing Agent

In order to spin label the cysteine residue incorporated into fH19-20 G1107C, the protein must be in its monomeric form, and the cysteine residue free and accessible. Typically, following the final purification step, recombinant proteins are treated with a reducing agent, usually Dithiothreitol (DTT) in order to reduce any cystine which may form as a result of cysteine oxidation in the fermentation vessel, as well as reduce any

intra and intermolecular disulphide bonds formed between cysteine residues of proteins. Figure 3.24 shows the reversible oxidation of cysteine to cystine.

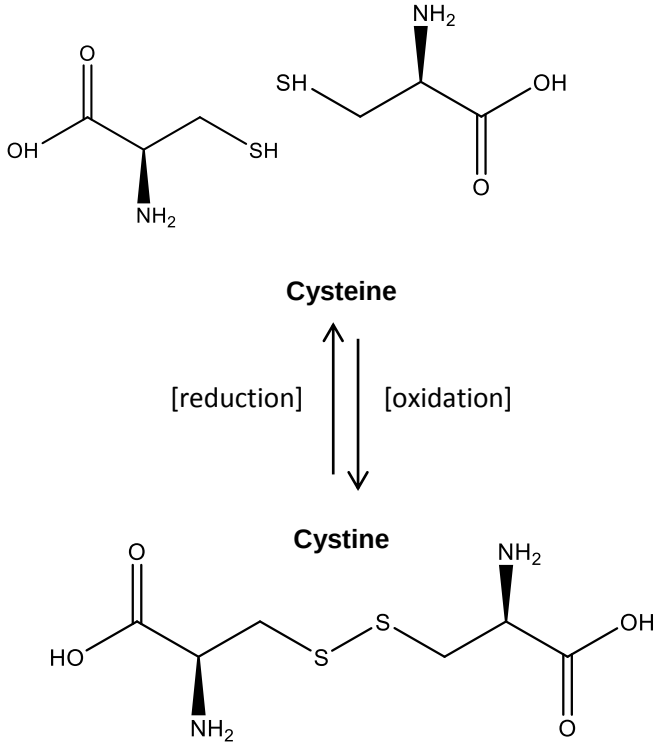



**Figure 3.24: Reversible Oxidation of Cysteine to Cystine, visualised using Chemdraw ®.**

The SP-Sepharose resin is a low resolution cation exchanger, and simply acts as a first catchment step in the purification of the fH19-20 G1107C from the crude supernatant. Immediately following the initial catchment step, the fH19-20 G1107C was treated with a final concentration of 5mM DTT, however DTT at this concentration was not sufficient to efficiently reduce the dimeric protein to an entirely monomeric species. A second sample was treated with 10mM DTT, which successfully reduced the dimeric fH19-20 to an entirely monomeric species, and following further purification using higher resolution CEC, the fH19-20 G1107C appeared clean under SDS-PAGE analysis (not shown).

Following unsuccessful attempts to spin label the purified proteins, it was proposed that at such high concentrations of DTT, the intermolecular disulphide bonds of the fH fragment had been compromised, resulting in a potential mis-folding of the protein.

Consequently, the stability fH19-20 G1107C in the presence of various concentrations of DTT, and in the presence of alternative reducing agents was tested. DTT is the most

commonly used thiol reductant (Netto & Stadtman, 1996), forming two products from one reactant, with the DTT being converted to a stable and cyclic disulphide through an intramolecular reaction (Getz *et al*, 1999), as shown in Figure 3.25.

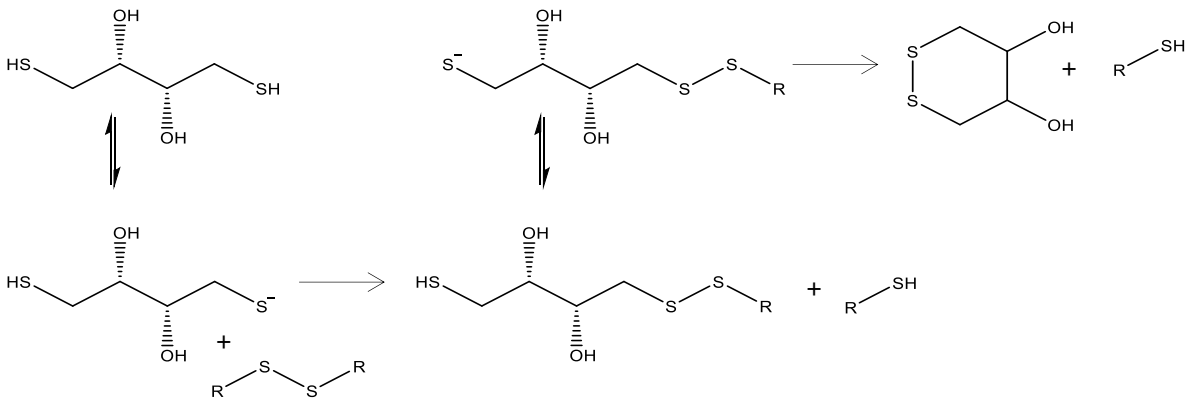



**Figure 3.25: Mechanism for the reduction of a disulphide with DTT, visualised using Chemdraw®.**

However, trialkylphosphines are also powerful reductants, selective in the reduction of disulphides (Cline *et al*, 2004). Tris (2-carboxyethyl) phosphine (TCEP) is one such reductant, which is able to reduce disulphides, driven by the formation of a strong phosphorus-oxygen bond, as shown in Figure 3.26.

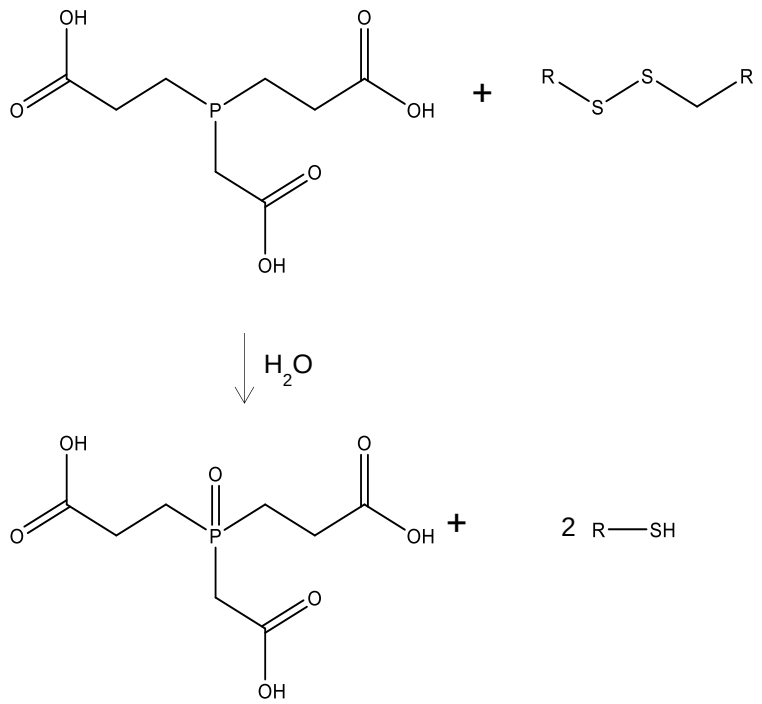



**Figure 3.26: Mechanism for the reduction of a disulphide bond with TCEP, visualised using Chemdraw ®.**

To assess which reducing agent was optimal for keeping the fH19-20 G1107C monomeric without disruption to the internal disulphide bonds, the purified protein was treated with varying concentrations of TCEP and DTT. The proteins were then analysed by SDS-PAGE to ensure complete reduction, before the amount of free thiol in each sample was quantified by the Ellman's reagent, DTNB. Results are presented in Figure 3.27 and Table 3.1.

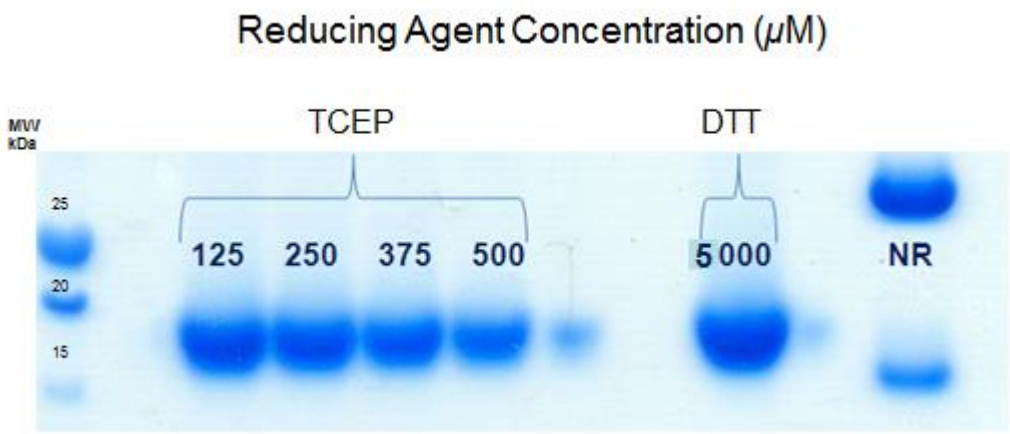



**Figure 3.27: SDS-PAGE Analysis of the reduction of fH19-20 G1107C with varying concentrations of DTT and TCEP.**

**Table 3.1: Concentration of free cysteine in the fH19-20 G1107C samples following reduction, as determined by the Ellman's assay (DTNB).**

| Reducing Agent (final conc.) | Conc. Free Cysteine* |
|---|---|
| DTT (5mM) | Significantly exceeds 45$\mu$M |
| TCEP (500$\mu$M) | 54$\mu$M |
| TCEP (375$\mu$M) | 48$\mu$M |
| TCEP (250$\mu$M) | 43$\mu$M |
| TCEP (125$\mu$M) | 27$\mu$M |

* Initial protein concentration - 45$\mu$M

The initial concentration of the fH19-20 G1107C sample was calculated to be approximately 45$\mu$M (as tested by the Bradford Assay, as described in 2.7.4) and confirmed by checking the absorbance at 280nm. In the fH19-20 construct, there should be only one free cysteine residue, as the remainder should be engaged in the intermolecular disulphide bonds found in the CCP residues.

Following reduction, the amount of free thiol (cysteine) in each sample can be quantified when the sample is reacted with DTNB. Previously it was found that a final

concentration of 1mM DTT was not sufficient to efficiently reduce the fH and keep it in its monomeric form. With this is mind, the ability of higher concentrations of DTT as well as varying concentrations of an alternative reducing agent were assessed in their ability to efficiently reduce fH. Upon reduction of the fH19-20 with a final concentration of 5mM DTT and a final concentration of over 250$\mu$M TCEP, the amount of free thiol in each sample significantly exceeded the protein concentration (45$\mu$M). This suggests that at these concentrations the reducing agent is present in high enough concentrations to disrupt and initiate the reduction of the internal disulphide bonds of the fH19-20, and potentially explains the problems with spin-labelling experienced.

There are a number of advantages in using TCEP as an alternative to DTT. Besides being odourless compared to DTT, TCEP has been shown to be more stable than DTT over a wider pH range (pH1.5-8.5), with increased stability at higher pH (7.5), and more efficient reduction at pH values below pH8 (Getz *et al*, 1999). A final TCEP concentration of 250$\mu$M was used for the reduction of any fH constructs due to disruption of the internal disulphide bonds as well as reduced stability at higher reducing concentrations.

Following quantification of free cysteine levels, it was concluded that purifying the fH fragments in the presence of such high levels of DTT is detrimental to the protein fold. Consequently, the fH fragment was prepared again, only large scale flask expression was performed instead of fermentation. The fH19-20 fragment expresses well, and therefore flask expression should provide enough material, whilst reducing the levels observed in the fermentation process. This should facilitate a quicker purification process, with less sample handling, and consequently less sample degradation.

### 3.4.6 Purification of fH19-20 G1107C Following Large Scale Flask Expression

In order to overcome the problems encountered during the fermentation process, large scale flask expression was carried out (8 x 500mL initial culture volume) for the fH19-20 G1107C construct.

### 3.4.6.1 SP-Sepharose FastFlow™ CEC Resin

The first purification step was carried out as for the fermentation. The elution chromatogram and resulting SDS-PAGE analysis of relevant fractions (as indicated by the red arrow) are shown in Figure 3.28.

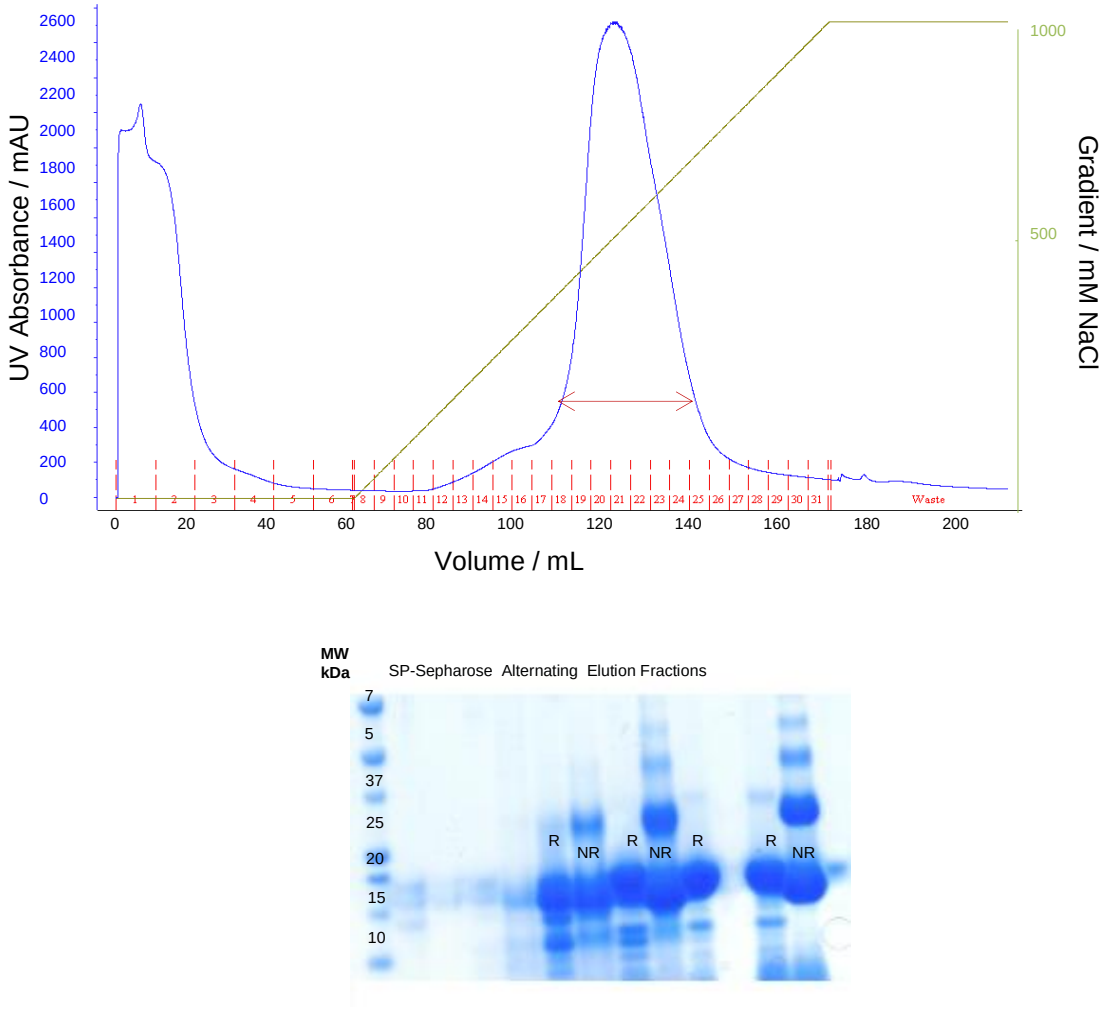



**Figure 3.28: Cation Exchange Chromatography of fH19-20 G1107C from Large Scale Flask Expression (SP Sepharose).** (A) Elution profile and (B) resulting SDS-PAGE analysis. Alternating fractions were run under the large peak, indicated by the red arrow in (A), with fractions run in both their reduced and non-reduced form.

### 3.4.6.2 Resource 15S™ CEC

Following the initial catchment step, the fH19-20 was subsequently purified by multiple rounds of higher resolution cation exchange chromatography, using the Resource 15 S column. The protein was eluted using a linear NaCl gradient up to 1M

NaCl. From previous purification steps, it was found that fH19-20 elutes at ~230mM NaCl, and so the fractions in the first major peak were analysed by SDS-PAGE. The elution chromatogram from one of the runs is shown alongside the resulting SDS page analysis in Figure 3.29.

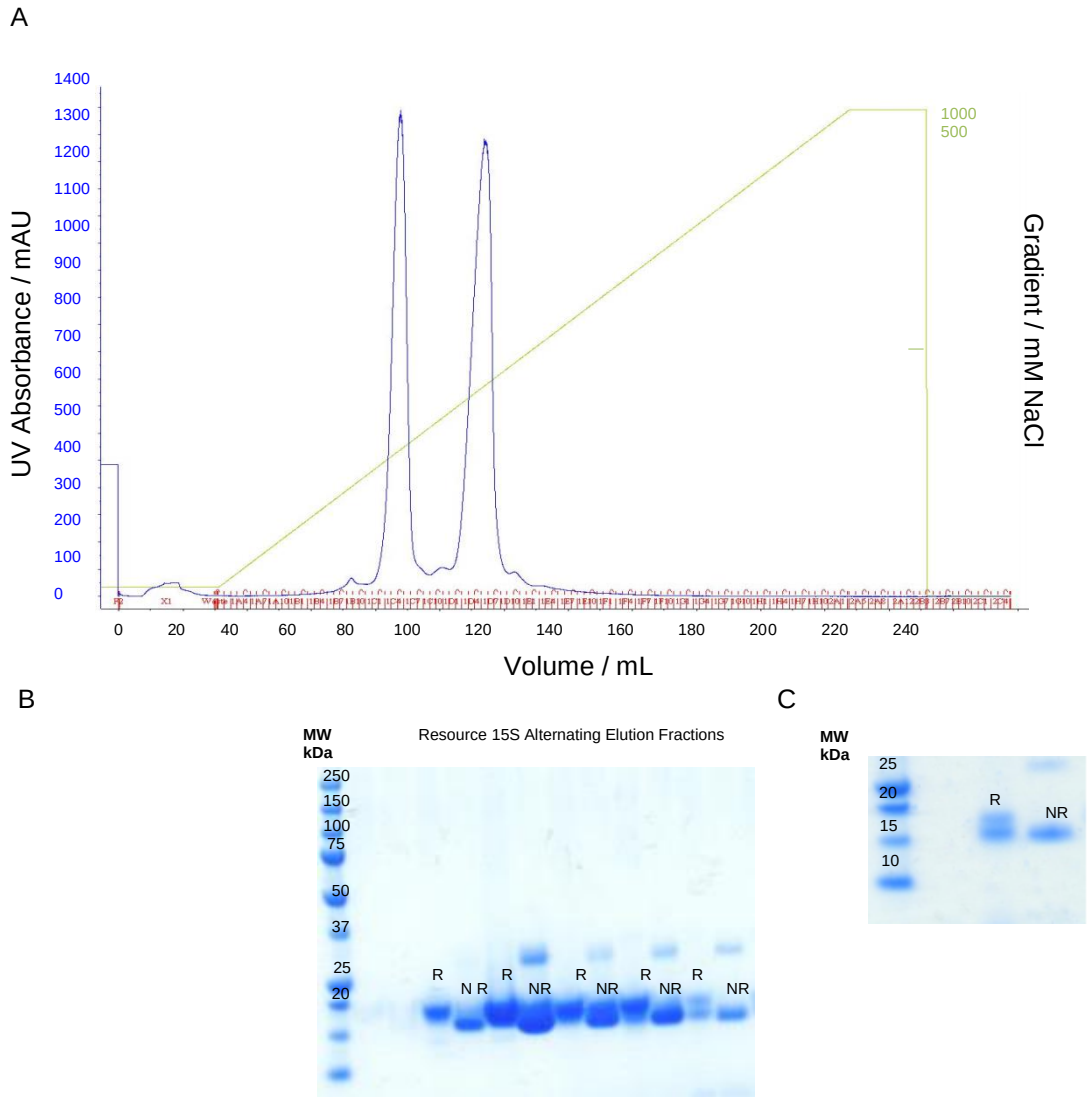



**Figure 3.29: Cation Exchange Chromatography of fH19-20 G1107C from Large Scale Flask Expression (Source 15 S).** (A) Elution profile and (B) resulting SDS-PAGE analysis of alternating fractions run under the second major peak (C) SDS-PAGE analysis of the central fraction from the first major peak.

Alternating fractions were analysed by SDS-PAGE under both major peaks. Under reducing conditions, two main bands are detected from those fractions under peak one, with the lower band likely corresponding to fH19-20 that had been proteolytically clipped (but remained intact until the disulphide bonds had been reduced. Those

fractions from the second major did not appear to have any clipped protein present, and so these were the fractions which were pooled for subsequent purification steps.

### 3.4.6.3 Size Exclusion Chromatography

From the SDS-PAGE analysis of the elution from the Resource 15S, there appears to be a little dimer in the non-reduced fractions, therefore the pooled fractions were treated with a final concentration of 250$\mu$M TCEP, before SEC was performed as a final purification step. Shown in Figure 3.30 is one run from the pooled fractions, with resulting SDS-PAGE analysis.

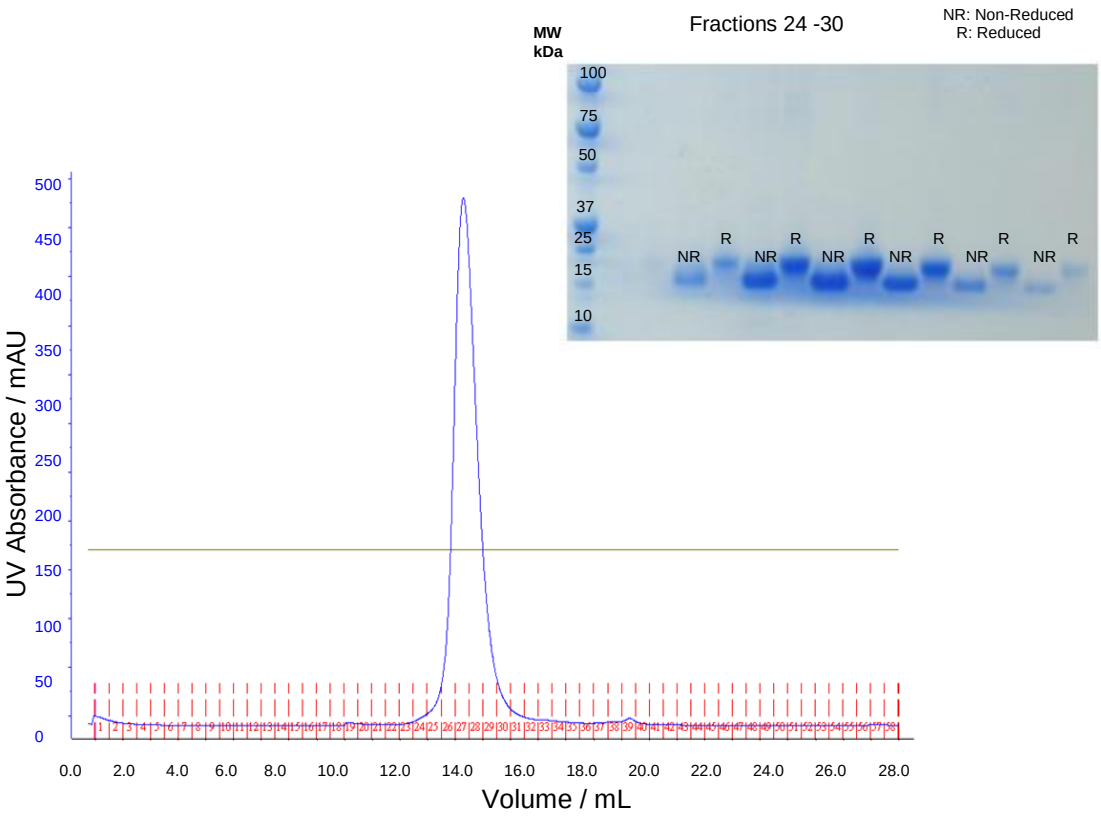



**Figure 3.30: Size Exclusion Chromatography of the fH19-20 G1107C Fragment.** (A) Elution profile of the fH19-20 G1107C construct and (B) resulting SDS-PAGE analysis of relevant fractions.

## 3.5 Spin labelling of fH Fragments 1-4 K247C and 19-20 G1107C

Following reduction of the purified fH proteins with a final concentration of 250$\mu$M TCEP, and quantification of free cysteine levels by the Ellman's Reagent DTNB, the fH fragments, K247C and G1107C were labelled with a 10x molar excess of the spin label

MTSSL, as described in 2.9.1. The excess spin label was removed, and r.t CW EPR spectra of the labelled fH samples collected. The CW EPR spectra are shown in Figure 3.31.

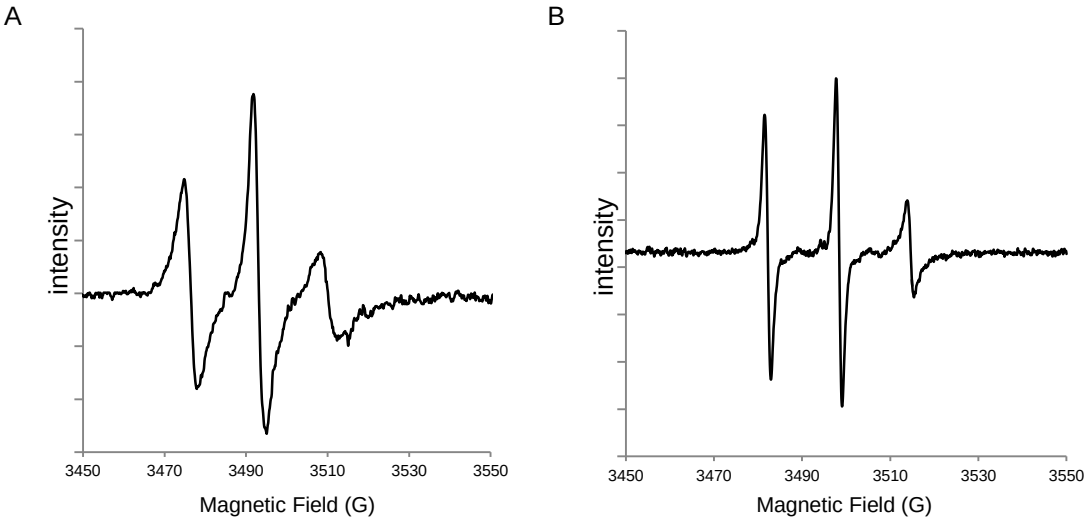



**Figure 3.31: X-band CW EPR of fH1-4 K247C and fH19-20 G1107C constructs.** Spectra taken at r.t (non-saturating conditions), using a Bruker EMX X-band EPR spectrometer, using a rectangular low sensitivity cavity.

The solution EPR results are important for two reasons: the degree of spin labelling can be determined and the rigidity of the label bound to the protein can be assessed. Solutions (40$\mu$L) of 4-hydroxy-TEMPO were used as calibration standards, and the spin labelling efficiency was deemed to be between 90 and 95% for the fH1-4 and fH19-20 samples respectively. Labelling efficiency can be approximated by quantifying the absolute number of spins in the sample through double integration of the CW-EPR spectrum, and comparing this value to the concentration

The spin-labelled fH fragments were frozen in 50% glycerol in PBS, and kept at -20$^{o}$C, however, upon thawing of fH fragments the CW EPR spectra for both fragments had changed considerably. Figure 3.32 shows the CW EPR spectra taken for the fH1-4 fragment following freeze-thaw.

It is evident that the spectrum obtained is not the characteristic three –line spectrum of a typical nitroxide radical. The same was observed for the fH19-20 G1107C mutant. It is proposed that upon freezing, due to the extensive disulphide bonds present in the

fragments, there has been a potential reshuffling of the disulphide bonds, forcing the incorporated spin labels from the residues used for incorporation.

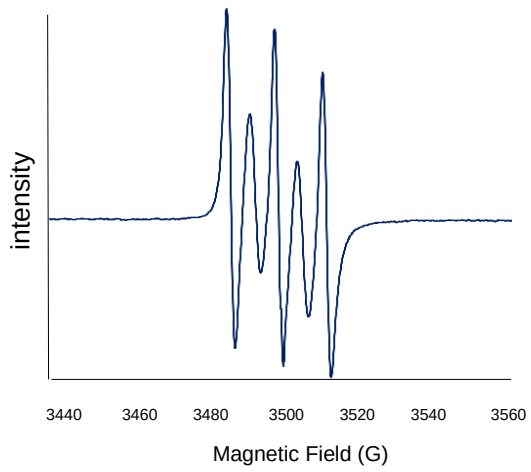



**Figure 3.32: X-Band CW EPR Spectrum of the fH1-4 K247C construct following freezing.** Spectrum taken at r.t under non-saturating conditions, using the AffirmoEx benchtop spectrometer, by ActiveSpectrum.

Typically, a five-line spectrum like the one shown above is indicative of biradical. In this instance that equates to the label which has been removed 'talking' to any label which remains bound to the protein of interest. At this stage, both fragments were extensively washed, in order to remove any now unbound label, however the bi-radical effect, although improved, was impossible to completely remove.

Going ahead with fH fragments in which labelling efficiencies could now not be determined, and the state of the label unsure was not possible, and so unfortunately, at this stage attempts were abandoned to salvage the labelled fragments, and they were both expressed again.

Halfway through this project, the lab was relocated, and so expression from the original glycerol stocks was tested to ensure expression was as it had been. Unfortunately these had not survived in transit, and so for each fragment, the entire transformation/test expression and large scale expression steps had to be repeated, and the proteins purified, assayed and finally spin labelled. These steps were repeated exactly as before, using large-scale flask expressions (8x500mL starting volume). Figure 3.33 shows the SDS-PAGE analysis and CW EPR of the new spin-labelled fH fragments.

The fH19-20 R1210C (purified by Dr Andy Herbert, University of Edinburgh, Scotland, UK) mutant was treated with a final concentration of 250$\mu$M TCEP, and spin labelled. The CW EPR spectrum and SDS-PAGE analysis of this mutant is shown alongside the SDS-PAGE analysis of the new fH1-4 K247C and fH19-20 G1107C fragments, as well as their CW EPR spectra in Figure 3.33.

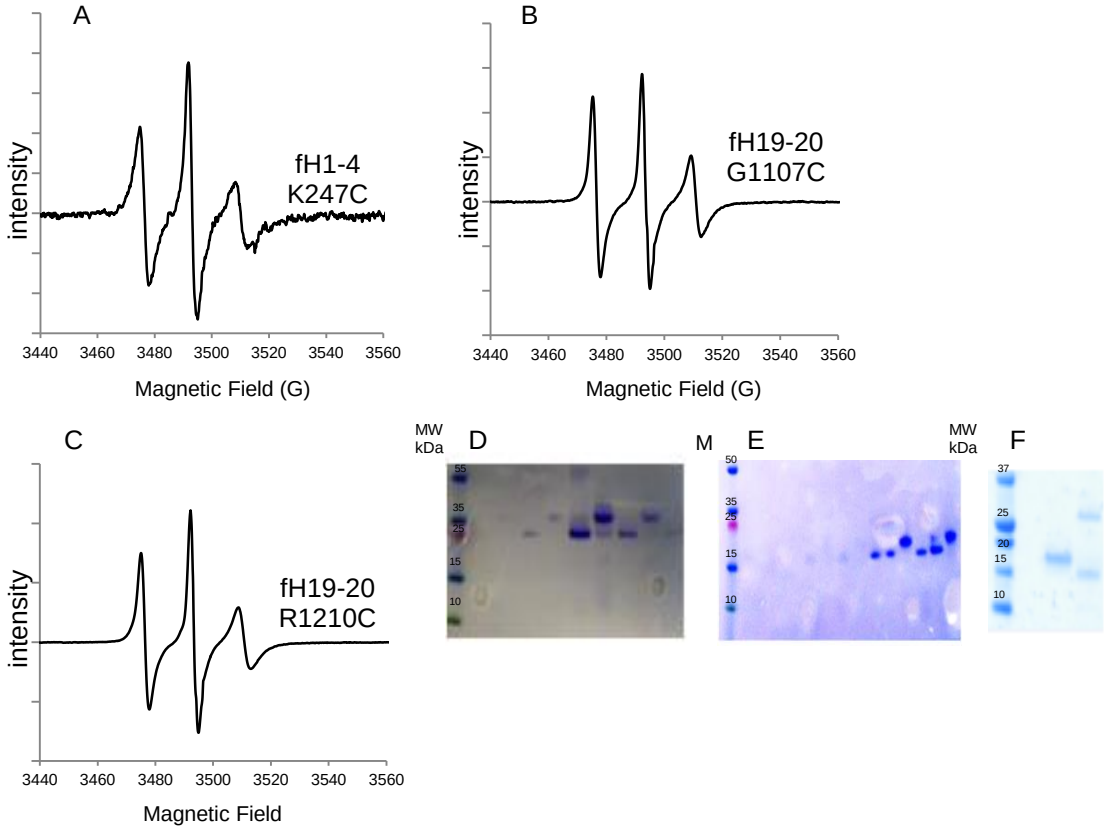



**Figure 3.33: CW EPR spectra and SDS-PAGE analysis of fH1-4K247C (A&D), fH19-20 G1107C (B&E) and fH19-20 R1210C (C&F).** CW EPR spectra were taken at r.t under non saturating conditions using the AffirmoEx benchtop spectrometer, by ActiveSpectrum, whilst SDS-PAGE analysis shows samples run in both their educed and non-reduced forms.

Final yields from the final fH preps used for EPR experiments are shown in Table 3.2**.**

**Table 3.2: FH fragments yields following expression in *P. pastoris*, and purification**

| fH Fragment | Expression Volume | Final Yield |
|---|---|---|
| fH1-4 K247C | 8 x 500mL (starting volume) | ~1.7mg |
| fH19-20 G1107C | 8 x 500mL (starting volume) | ~6mg |
| fH19-20 R1210C | 8 x 500mL (starting volume) | ~3.4mg |

## 3.6 Generation of Spin-Labelled C3b, Following Isolation and Purification of C3

Unlike the fH fragments which can be made recombinantly, C3b cannot be made using recombinant methods and so there is need for a different strategy for purification. C3b for the EPR study was derived from complement component C3, which was first isolated and purified from human plasma. In native C3, the highly reactive TE is shielded from reacting nucleophiles, however, upon activation to C3b, the TE domain becomes exposed, and can be exploited for SDSL of complement C3b.

### 3.6.1 Small Scale Isolation of C3/C3b

#### 3.6.1.1 Q-Sepharose FastFlow™ AEC Resin

Complement component C3 is one of the most abundant plasma proteins, and, as so, was purified from human plasma, following methods adapted from (Dodds, 1993). Following precipitation of the plasma fraction with $PEG_{3350}$, the supernatant containing C3 and C4 was loaded onto a XK 26/20 column packed with 10mL of pre-equilibrated QSFF resin following centrifugation. The protein was eluted at a flow rate of 2mL/min using a linear 1M NaCl gradient.

C3 was eluted from the column approximately halfway through the NaCl gradient at ~300mM NaCl, as indicated on the chromatogram. Its elution is directly preceded by the elution of the copper containing plasma protein ceruloplasmin, which appears blue in colour. Under reducing conditions, the two chains of C3 (a 70 and 110 kDa chain) run separately under SDS-PAGE analysis, and so can be distinguished from the three-chain C4 structure on the reducing gel, which elutes immediately after C3.

Care was taken to pool those fractions containing just C3, whilst avoiding contaminants such as C4 and ceruloplasmin. Avoiding these proteins resulted in a substantial loss of the original quantities, however, ceruloplasmin and C4 are abundant plasma proteins which prove difficult to eliminate in subsequent purification steps. The resulting elution chromatogram is shown alongside SDS-PAGE analysis of relevant elution fractions in Figure 3.34.

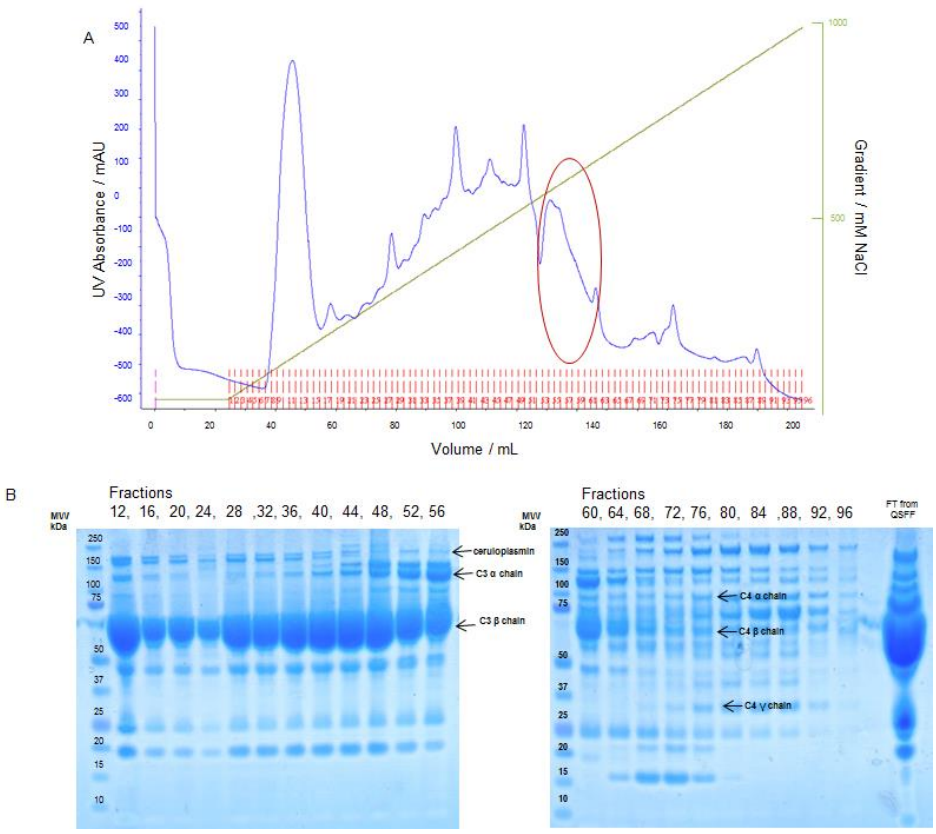



**Figure 3.34: Anion Exchange Chromatography of Plasma C3 following $PEG_{3350}$ Precipitation.** (A) Elution Chromatogram and (B) resulting SDS-PAGE analysis of the separation of complement components from human plasma. Every fourth fraction eluted between 210mM and 500mM NaCl was run in their reduced form.

### 3.6.1.2 Anion Exchange Chromatography (AEC) - Tricorn™ Mono Q 4.6/100 PE

The pooled C3 fractions were diluted with a half volume of $ddH_2O$ to lower the ionic strength and facilitate binding to the MonoQ column for subsequent purification steps. The mono Q column was equilibrated in 90% equilibration buffer and 10% elution buffer, and the pooled C3 fractions loaded in several batches, and eluted within a 20mL linear NaCl gradient to a final concentration of 500mM NaCl, at a flow rate of 1ml/min. Figure 3.35 shows the elution profile obtained for one of these runs, and SDS-PAGE analysis of the relevant fractions.

Similarly to the initial catchment step, the C3 protein elutes at ~300mM NaCl. The fractions indicated by the red arrow in Figure 3.35 were analysed by SDS-PAGE, with both the alpha and beta chains of C3 visible. All C3 containing fractions from the multiple Mono Q purifications were pooled and dialysed into non-NaCl buffer. The concentration of C3 was approximated using UV absorbance at 280nm and a molar extinction coefficient of $180{,}000M^{-1}cm^{-1}$ of C3.

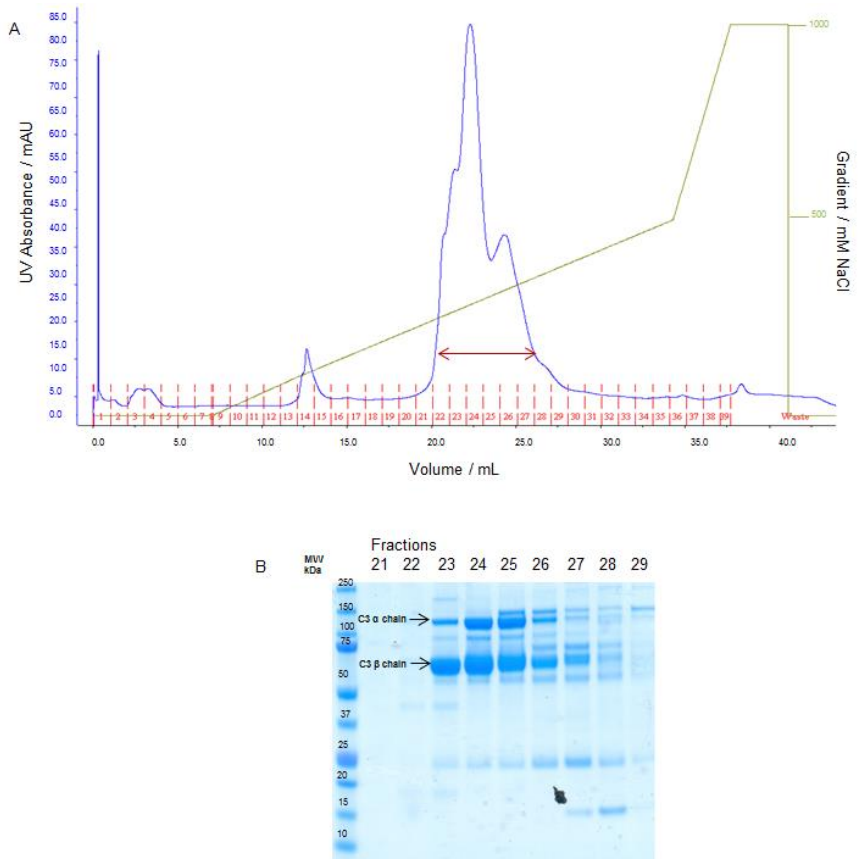



**Figure 3.35: Anion Exchange Chromatography (Mono Q) of pooled C3 Fractions Following the QSFF Catchment Step.** (A) Elution profile from multiple pooled fractions of C3 and (B) resulting SDS-PAGE analysis. C3 eluted form the Mono Q column at ~300mM NaCl. The fractions indicated in the major peak were all run in their reduced form.

### 3.6.2 Conversion of C3 to C3b Using Limited Trypsin Digestion

In nature, conversion of C3 to C3b arises *via* cleavage of C3 by the C3bBb convertase, yielding anaphylatoxin C3a and main fragment C3b. Production of C3bBb in the body requires the presence of C3b first, due to the cyclic nature of the complement cascade, and so use of the actual convertase is not possible.

CVFBb, a structural mimic of the C3bBb convertase (Fritzinger *et al*, 2009) can be used as an alternative for the conversion of C3 to C3b, however, due to species protection, CVF cannot be imported into the U.K. Therefore, to successfully convert C3 to the activated C3b and release the cysteine residue in the TE for spin labelling purposes, C3 must undergo a limited trypsin proteolysis at 37$^{o}$C to release the C3a fragment, leaving activated C3b.

Purified C3 at a concentration of 1mg/mL was treated with a 1% trypsin solution (w/w, enzyme/protein) for 2min at 37$^{o}$C in PBS, to simulate physiological pH. Immediately after incubation, 5% (w/w inhibitor/enzyme) soybean trypsin inhibitor was added to halt the reaction.

Following trypsin proteolysis and trypsin inhibition, the sample was immediately transferred to ice, followed by treatment with a 10 x molar excess of MTSSL or a final concentration of iodoacetamide. The labelled C3b is then diluted at a ratio of 1:1 with the non-salt buffer, and eluted over 30 column volumes with a linear salt gradient of 100-300mM NaCl from the same pre-equilibrated Mono Q column. The Mono Q AEC purification step acts to purify the C3b from unconverted C3, as well as removing any excess MTSSL/iodoacetamide. The fractions containing C3b in its pure form were pooled and dialysed against PBS, before being stored at -80$^{o}$C. The purity of C3b was verified using SDS-PAGE, as shown in Figure 3.36.

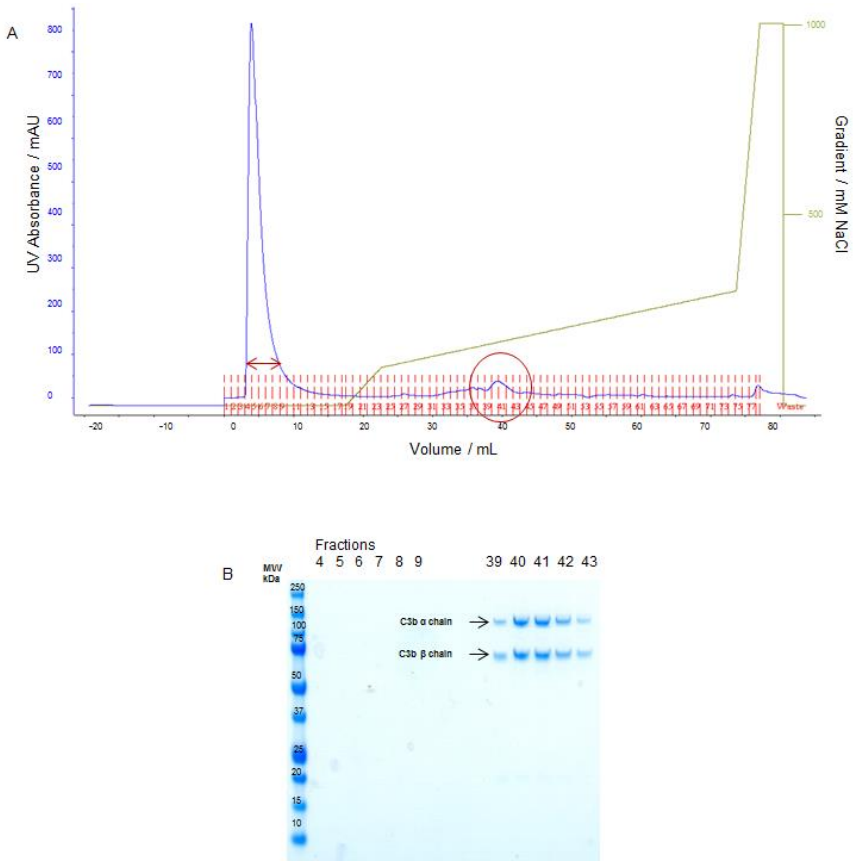



**Figure 3.36: Anion Exchange Chromatography (Mono Q) of Digested Plasma Purified C3, converted to C3b.** (A) Elution profile of C3b and (B) resulting SDS-PAGE analysis of fractions which fall under the major peak to the left, and the circled central peak.

### 3.6.3 Small Scale Isolation of C3/C3b – Troubleshooting

Due to the unspecific nature of trypsin proteolysis, final yields of C3b were drastically reduced. The theoretical yield from 10mL of human plasma is between 8 and 12 mg of C3, however, upon conversion to the activated C3b, the yield of C3b is approximately 1% of the starting material. The process was scaled up, from a starting material of 20mL of blood (giving 10 mL plasma) to 100mL of blood (50mL of plasma). The purification process was carried out in the same way as detailed above, however, a Resource 15 Q (20mL) anion exchange column was used as an intermediate purification step between the QSFF and Mono Q purification steps.

Resource 15 Q resin is another strong anion exchanger, with higher resolution than the initial QSFF step. It was hoped that further cleaning before the final highest resolution Mono Q purification step would increase purification efficiency and reduce sample handling; however it was impossible to isolate C3b in its pure form following digest of the purified C3. The scaled up procedure was repeated, with slight modification to protocol, with slight modification to flow rates and pH, as well as the inclusion of a CEC step in between in hopes of eliminating contaminants before conversion to C3b. The resulting SDS-PAGE analyses are shown in Figure 3.37.

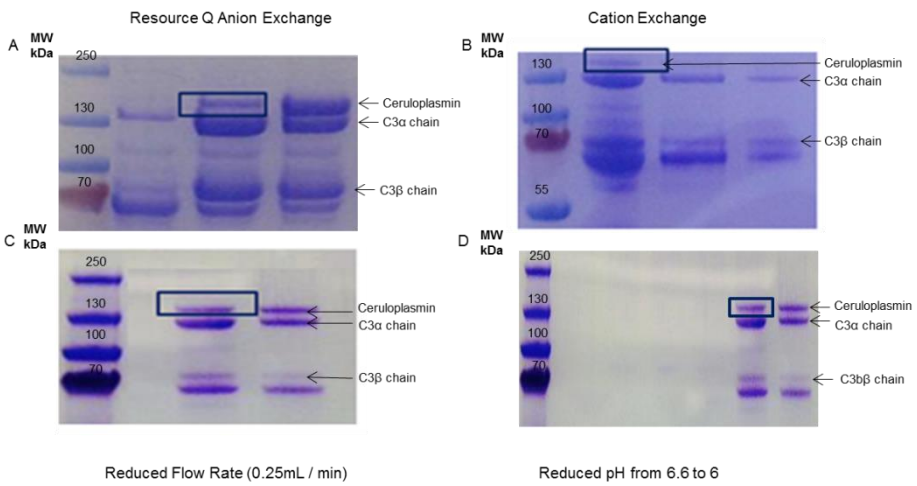



**Figure 3.37: SDS-PAGE analysis of C3 Purification Troubleshooting.** (A) Inclusion of a Resource Q purification step following the initial QSFF catchment step (B) Inclusion of a CEC (Resource S) purification step following MonoQ purification of the C3 pool (C) A reduction in flow rate from 1mL/min to 0.25mL/min during elution from the Mono Q (D) change in pH from 6.6 to 6.

Following purification of C3, it is evident that although care has been taken to avoid those fractions containing the protein ceruloplasmin, C3 co-purifies alongside it, as confirmed by in-gel digest of the relevant band and mass spectrometry of the impurity using MALDI MS and MSMS at the University of St Andrews, as shown in Figure 3.38. Following conversion from C3 to C3b, this is not eliminated

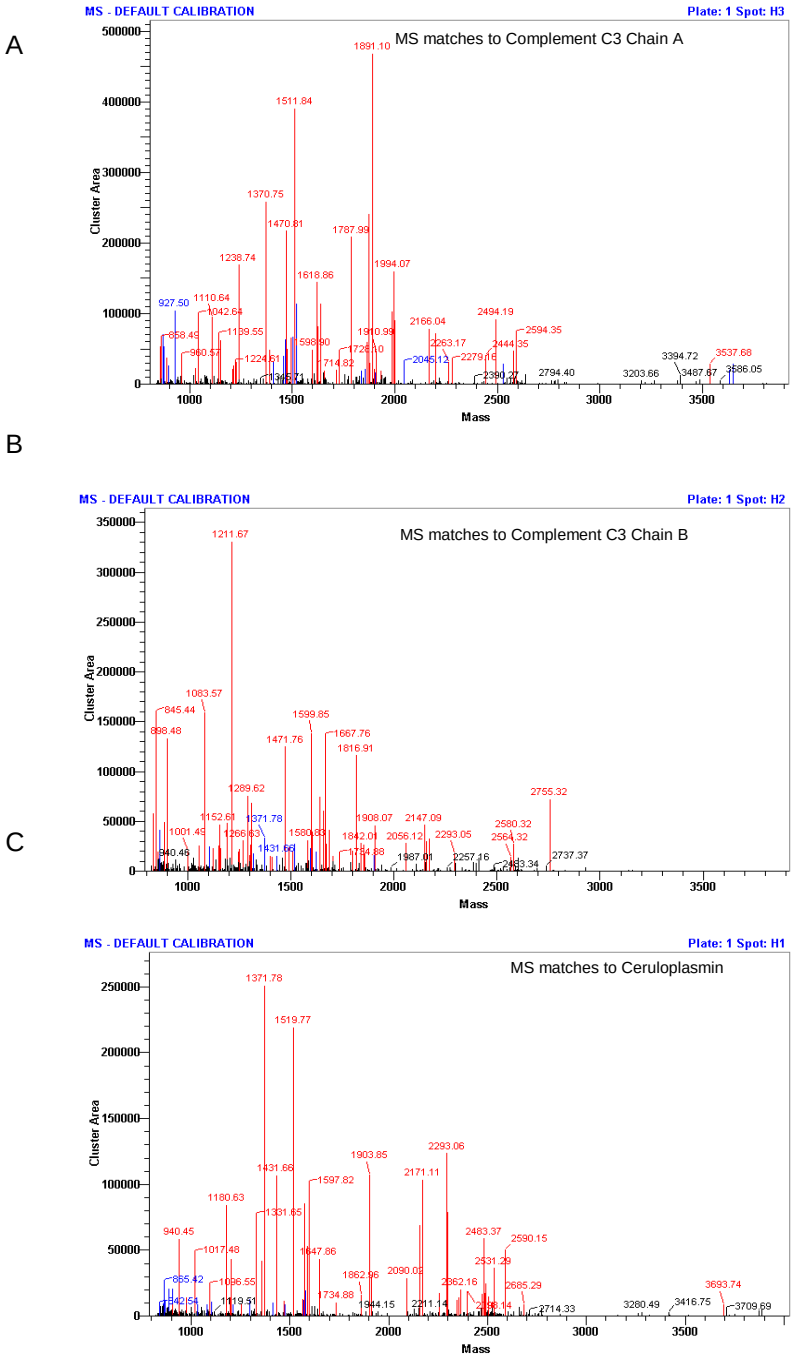



**Figure 3.38: Peptide Fingerprinting of the C3 α and β chains, as well as ceruloplasmin, following in-gel digest.**

Following slight modification of the purification procedures, as shown in the SDS-PAGE analysis in Figure 3.37, the ceruloplasmin impurity could not be separated from C3b. A small degree of impurity may not prove problematic when using the C3b samples for DEER, however, the ceruloplasmin is present in large quantities. Ceruloplasmin is a cysteine rich protein, which means specific labelling of just the cysteine in the exposed TE of C3b would prove impossible. Furthermore, ceruloplasmin has an EPR signal from some of its copper centres (Cannistraro *et al*, 1990) which could interfere with proposed DEER experiments. Figure 3.39 shows the XRC structure of ceruloplasmin.

**Figure 3.39: Structure of Plasma Protein Ceruloplasmin showing the copper centres (orange spheres).** (PDB_ID: 4ENZ,(Samygina *et al*, 2013)).

The protein ceruloplasmin has a theoretical pI of 5.44, and C3b 5.66, which highlights the difficulty in separating them using I.E.C., as based on their theoretical pIs, they should elute simultaneously from an anion exchange column at similar places along the NaCl concentration gradient. Following extensive troubleshooting it was decided new methods must be adopted to address these recurring issues.

## 3.7 Generation of Spin-Labelled C3(N), an Alternative to C3b, Following Isolation and Purification of C3

Although conversion of C3 to C3b drastically reduces the overall yield of C3b when using limited trypsin digest, isolation and purification of component C3 from human plasma gives a yield of between 85-92%. Methods were established which eliminate the need for proteolytic digest of C3 to C3b, and instead exploit the high yields of C3. The TE of C3, although buried, is accessible to small nucleophiles. With this in mind, C3

isolated and purified from 50mL of human plasma was treated with a small nucleophile, namely methylamine, activating the TE in C3, without eliminating the C3a portion.

C3(N) has C3b like properties, and the problem with unspecific digest of C3 eliminated. The C3 was isolated from human plasma exactly as described above, by the QSFF anion exchange catchment step, followed by a higher resolution Mono Q anion exchange chromatography step of the pooled C3 fractions from QSFF. Figure 3.40 shows the elution profile for one batch of pooled C3 fractions from QSFF after they had undergone the Mono Q purification step, as well as resulting SDS-PAGE analysis.

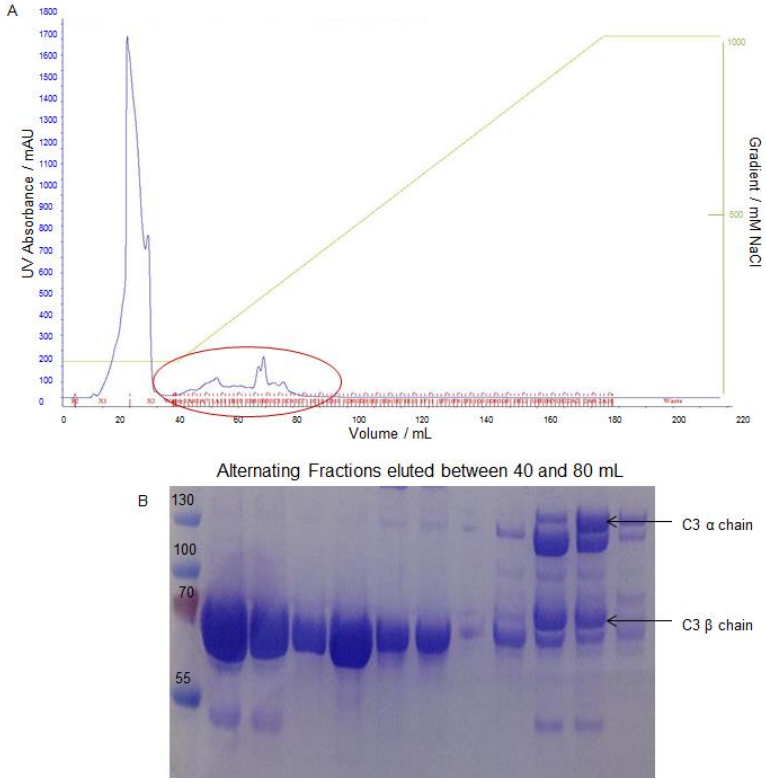



**Figure 3.40: Anion Exchange Chromatography of C3 pool, following QSFF, and before treatment with the nucleophile methylamine.** (A) Elution profile from the Mono Q Anion exchanger, and (B) resulting SDS-PAGE analysis of alternating fractions, eluted between 40 and 80mL, all run in their reduced form.

Purified C3 was treated with a final concentration of 200mM methylamine (pH7.5), at 37$^{o}$C for 3hr. The reaction mixture was incubated in the presence of either iodoacetamide (20mM final concentration) or with a 10x molar excess of MTSSL spin

label, depending on whether the protein was required in its labelled or unlabelled form. Treatment with the methylamine allows cleavage of the internal thiol ester of C3, releasing a free thiol group for spin labelling purposes or alternatively allows for blocking of the cysteine residue with iodoacetamide.

Following the 3 hr incubation period, the C3 pool was then purified in a similar fashion as described for C3b activated using trypsin proteolysis. Following purification on Mono Q, purity of the C3(N), was determined by SDS-PAGE analysis, as shown in Figure 3.41, and proteins concentrated to ~1mg/mL before being stored at -80$^{o}$C.

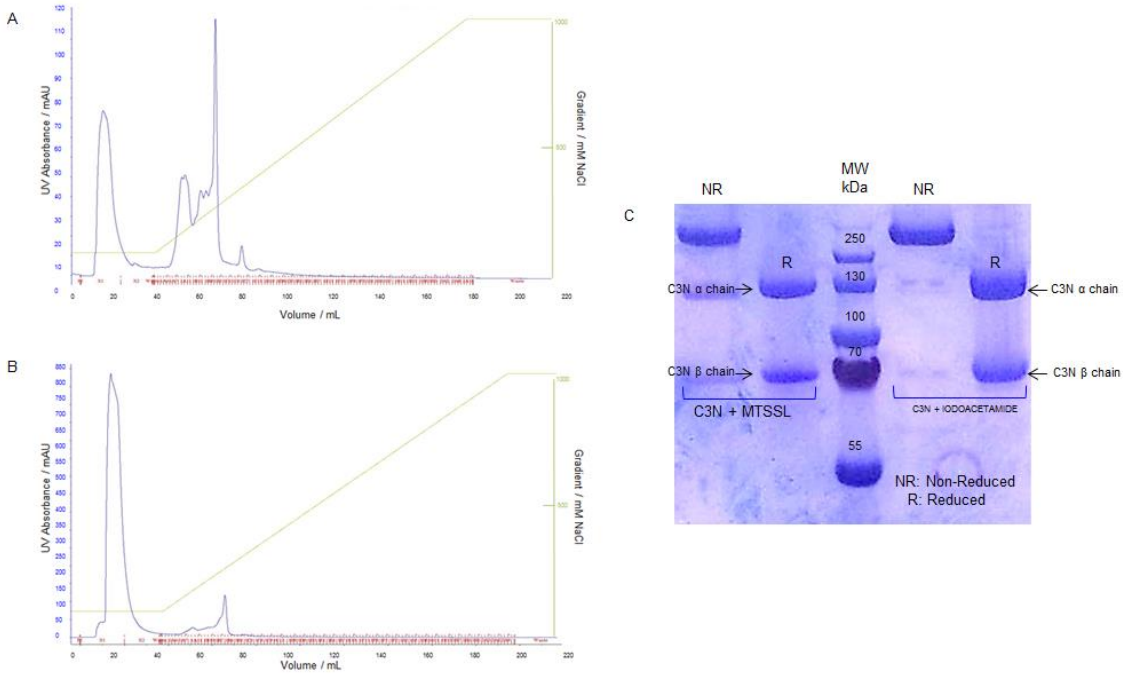



**Figure 3.41: Anion Exchange Chromatography of Methylamine treated purified C3 and its conversion to C3(N).** (A) Elution profile from the Mono Q Anion exchanger of the MTSSL labelled C3(N) and (B) Elution profile from the Mono Q Anion exchanger of the iodoacetamide treated C3(N) (C) Resulting SDS-PAGE analysis of both the labelled and unlabelled C3(N).

From the SDS-PAGE analysis of the labelled and unlabelled C3(N), it is evident that the ceruloplasmin impurity has been eliminated, as just C3(N) in both its reduced and non-reduced forms is detectable. Table 3.3 summarises approximate yields of C3, and the absolute yields of C3b and C3(N) following the standard purifications detailed above, as well as those carried out with slight modification to protocol.

**Table 3.3: Summary of C3/C3b/C3(N) Yields following Isolation and purification From Plasma**

| Volume of Plasma (mL) | Purification Procedure | C3 Yield (mg) | C3 Yield (%)* | C3b Yield following digest | C3(N) yield following methylamine treatment of C3 (mg) |
|---|---|---|---|---|---|
| 10 | Standard | 9 | 90 | 0.1 | n/a |
| 50 | Standard | 38 | 76 | 0.7 | n/a |
| 50 | + Resource Q | 42 | 84 | 1.1 | n/a |
| 50 | + Resource S | 33 | 66 | 0.4 | n/a |
| 50 | Reduced Flow | 41 | 82 | 0.9 | n/a |
| 50 | Reduced pH | 43 | 83 | 0.5 | n/a |
| 50 | Standard | 46 | 92 | n/a | 25 |

* Theoretical Yield of C3 = ~1mg/mL of fresh human plasma

From each individual C3 prep, the approximate yield of C3 following its purification from human plasma has been, on average, over 80%, based on theoretical yields. However, upon cleavage of C3 to C3b using limited trypsin digest the final yield of C3b is dramatically reduced, on average, to less than 0.1mg of C3b/10mL plasma.

Consequently, using C3(N) as an alternative, which is functional as C3b with the addition of the un-cleaved C3a segment, increases yield from 0.1mg/10mL of plasma to ~5mg/10mL of plasma, a 50 fold increase. This, coupled with the ability to remove impurities in the purification process highlight the benefits of using C3(N) as an alternative to C3b, in both its labelled and unlabelled forms.

## 3.8 Spin labelling of C3b and C3(N)

For the C3b sample, following activation from C3 by limited trypsin digest, and for C3(N), the samples were labelled, and excess spin label removed. The r.t CW EPR spectra of the labelled C3b and C3(N) samples were collected, as shown in Figure 3.42.

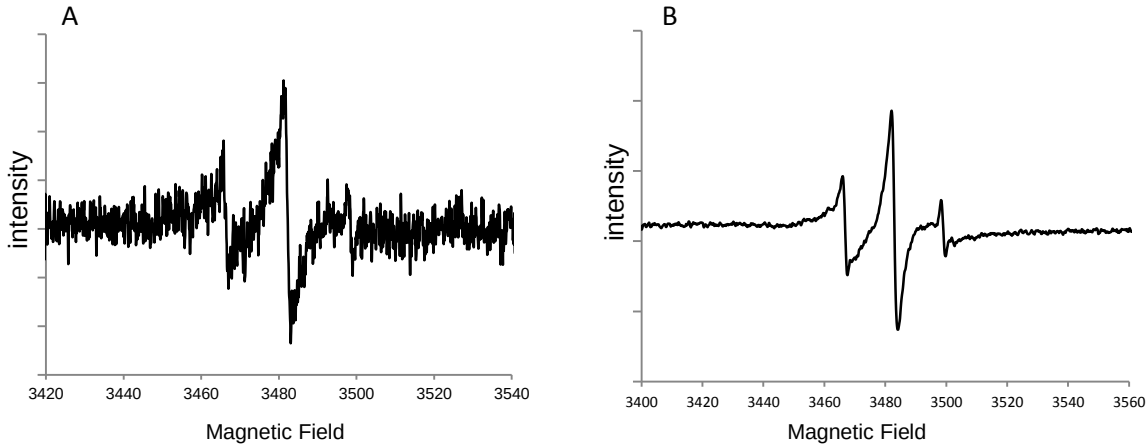



**Figure 3.42: X-band CW EPR of spin-labelled (A) C3b and (B) C3(N**). Spectra taken at r.t (non-saturating conditions), using a Bruker EMX X-band EPR spectrometer, using the high sensitive SHQE resonator.

From the CW EPR spectra, the labelling efficiencies of the C3b and C3(N) were determined to be between 85 and 90% respectively. There is a great deal of noise in the C3b CW spectrum due to the low concentration of the sample. Sample concentration could not be increased, as for the C3b prep, very little C3b was produced, due to unspecific cleavage of the purified C3 by trypsin. The C3(N) has a much better signal to noise, showing a nice broad spectrum, indicative of slow tumbling of the large C3(N) protein. There is indication of some free spin label in this sample.

## 3.9 Functional Characterisation – fH mediated cleavage of C3b and C3(N) by fI

As described in 2.10.1, a fluid phase cofactor assay can be carried out to visualise the fI mediated cleavage of complement component C3b to inactivated C3b (iC3b), for which fH is a cofactor. For the C3b and C3(N) proteins isolated and purified from plasma, their cleavage by fI can be monitored, to assess whether both the labelled and non-labelled forms give the banding pattern expected upon conversion to iC3b.

In the positive control reactions, C3b generated from plasma purified C3 was incubated with full length plasma purified fI and fH (all Complement Technology Inc, Texas, USA). The negative control contained no fH.

The C3b and C3(N) made in this body of work were assayed in both their spin-labelled and unlabelled forms. The ability of the fH1-4 K247C fragment, in both its

spin-labelled and unlabelled forms, to act as a cofactor in the cleavage of C3b to iC3b was also assessed. The two disulphide linked chains of C3b can be visualised using SDS-PAGE. Upon cleavage to iC3b by fI, the 110kDa alpha chain of C3b is cleaved into two smaller chains (67 & 40kDa). Reduction of the alpha chain can be visualised as well as the formation of the two smaller alpha chain fragments. All assays are shown in Figure 3.43.

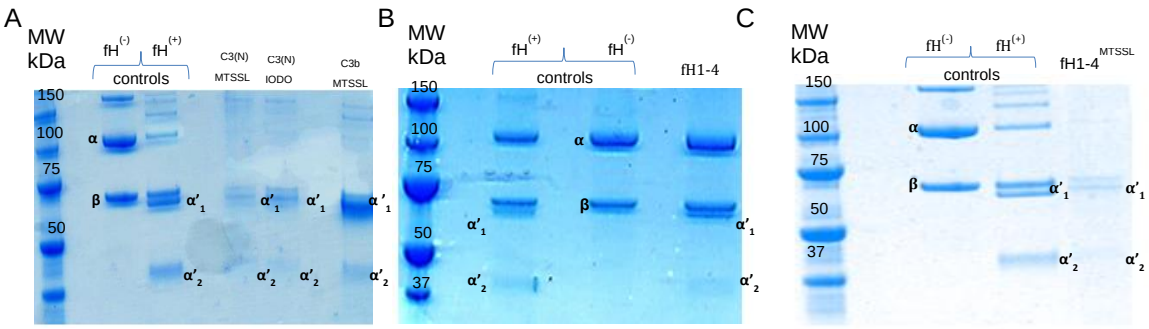



**Figure 3.43: Fluid-Phase Cofactor Assay.** This assay monitors the cleavage of C3b and C3(N) to iC3b by fI in the presence of fH. The assay also assesses the ability of fH1-4 to act as a cofactor in the cleavage of C3b to its inactive form. In all controls, full length native plasma purified proteins (Comptech) are used. FH is omitted from the negative control, whilst the positive control includes fH,fI, and C3b. (A) Shows both positive and negative CompTech controls, with the cleavage of C3(N) (labelled and unlabelled) and C3b (labelled) by fI being monitored, in the presence of fH (fI and fH both CompTech). (B) and (C) Show both positive and negative CompTech controls, whilst monitoring the ability of fH1-4 (unlabelled and labelled, respectively) to act as a cofactor for the cleavage of C3b to iC3b. (C3b, and fI both Comptech).

From SDS-PAGE analysis of the cofactor assays, it is clear that C3b and C3(N) display the same banding pattern as shown in the positive control, showing that in the presence of fI and fH, C3b and C3(N) are cleaved to the inactive iC3b. This suggests that the C3b and C3(N) made for these experiments bind fH and fI in the same manner as native plasma C3b. Similarly, the fH1-4 K247C, in both its labelled and unlabelled form is sufficient to act as a cofactor for the cleavage of plasma C3 to iC3b, however it is clear that the fH fragment is not as functionally active as the full length fH purified from plasma.

## 3.10 The fH1-4. C3(N). fH19-20 Complex – An EPR Study

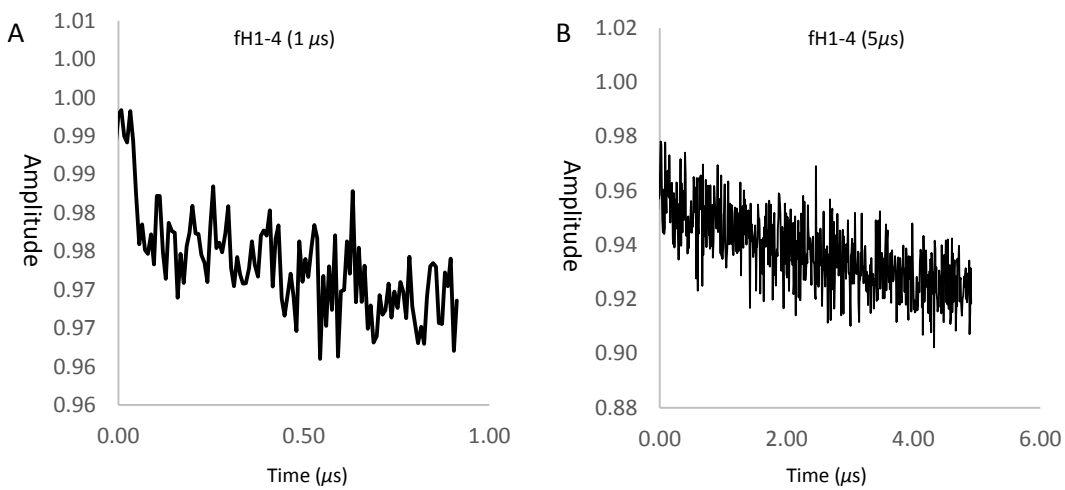
Following confirmation of spin-labelling of the fH fragments, the complex was ready to be formed for DEER measurements. In order to prepare samples, the binding dissociation constants ($K_d$s) were carefully considered. The $K_d$ of the fH1-4-C3b

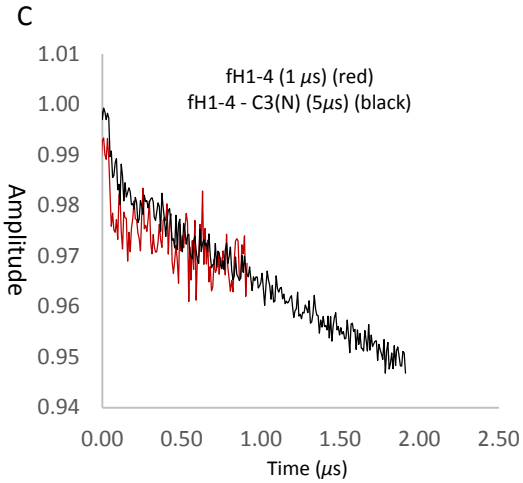
interaction is ~ 10$\mu$M (Schmidt *et al*, 2008) (Wu *et al*, 2009) whilst the $K_d$ of the fH19-20-C3b interaction is 1.4$\mu$M (Kajander *et al*, 2011). At the end of the preparations there was 55$\mu$l of 87$\mu$M fH1-4 K247C, 120$\mu$l of 777$\mu$M fH19-20 G1107C (which was observed to be slightly pink), 90$\mu$l of 548$\mu$M fH19-20 R1210C, 30$\mu$l of 50$\mu$M C3(N) iodo and 81$\mu$l of 55$\mu$M of C3(N)-SL.

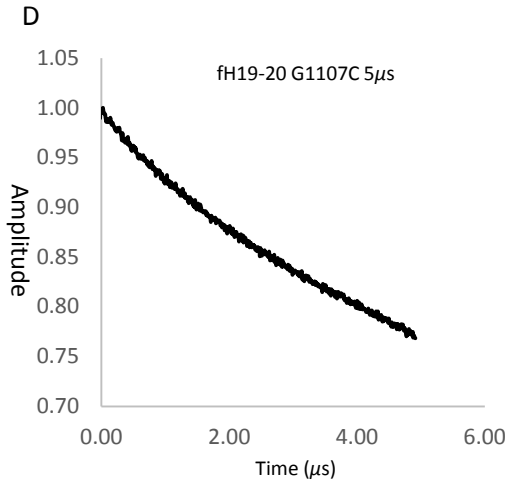
Both C3(N) samples had been concentrated from frozen stocks and all proteins were washed into deuterated PBS. DEER samples were made with 40$\mu$l protein solutions or buffer and 20$\mu$l deuterated glycerol. Sample compositions are summarised in Table 3.4.

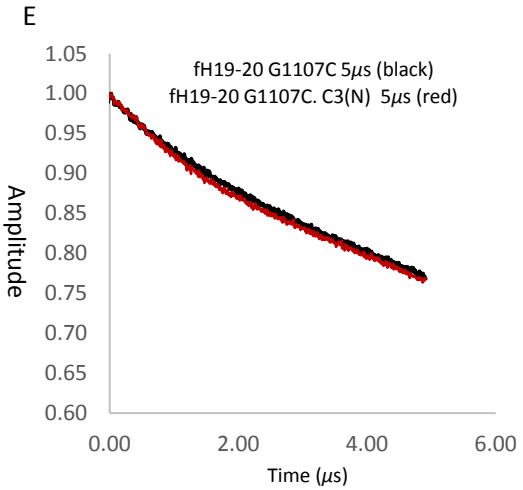
**Table 3.4: Compositions of Samples analysed by DEER**

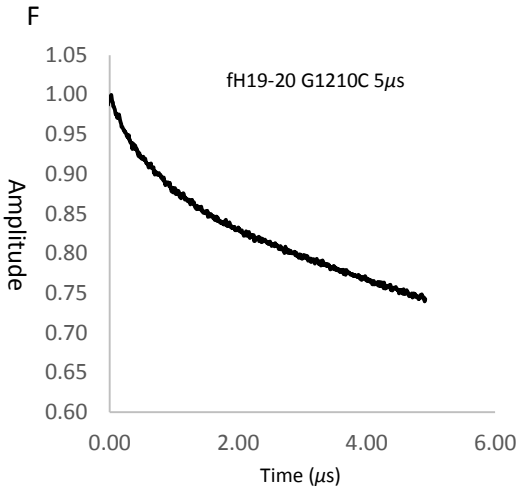
| Measurement | fH1-4 K247C (87$\mu$M) | fH19-20 G1107C (777$\mu$M) | fH19-20 R1210C (548$\mu$M) | C3(N) MTSSL (55$\mu$M) | C3(N) IODO (50$\mu$M) | Deuterated Glycerol |
|---|---|---|---|---|---|---|
| fH1-4 K247C – C3(N) (Cys988) | 13$\mu$L | - | - | 27$\mu$L | - | 20$\mu$L |
| fH1-4 K247C – fH19-20 G1107C | 13$\mu$L | 13$\mu$L | - | - | 14$\mu$L | 20$\mu$L |
| fH1-4 K247C – fH19-20 R1210C | 13$\mu$L | - | 13$\mu$L | - | 14$\mu$L | 20$\mu$L |
| fH19-20 G1107C – C3(N) (Cys988) | - | 13$\mu$L | - | 27$\mu$L | - | 20$\mu$L |
| fH19-20 R1210C – C3(N) (Cys988) | - | - | 13$\mu$L | 27$\mu$L | - | 20$\mu$L |

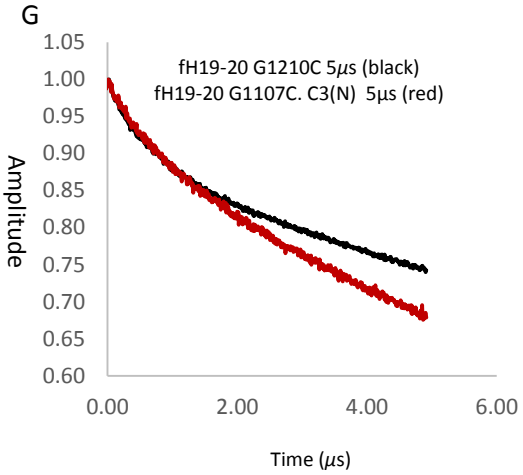


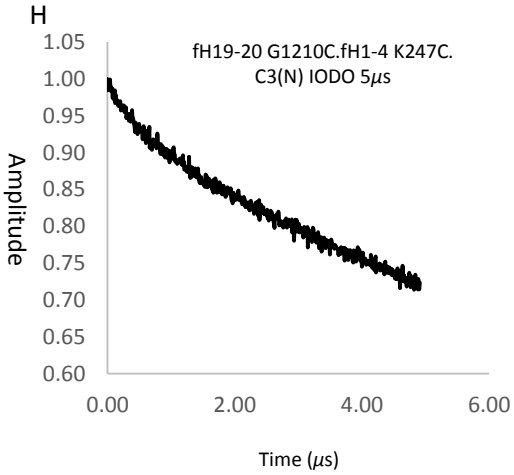

C
fH1-4 (1 μs) (red)
fH1-4 - C3(N) (5μs) (black)
Amplitude
Time (μs)


D
fH19-20 G1107C 5μs
Amplitude
Time (μs)


E
fH19-20 G1107C 5μs (black)
fH19-20 G1107C. C3(N) 5μs (red)
Amplitude
Time (μs)


F
fH19-20 G1210C 5μs
Amplitude
Time (μs)


G
fH19-20 G1210C 5μs (black)
fH19-20 G1107C. C3(N) 5μs (red)
Amplitude
Time (μs)


H
fH19-20 G1210C.fH1-4 K247C.
C3(N) IODO 5μs
Amplitude
Time (μs)

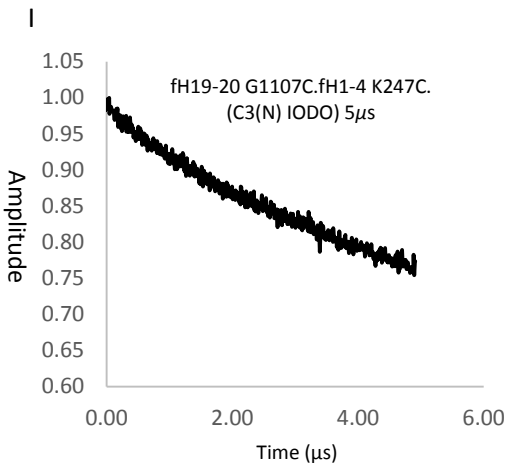



**Figure 3.44: Raw data from the EPR experiments taken for the spin labelled fH fragments in complex with either labelled or unlabelled C3(N).** Sample compositions shown in Table 3.4.

Following assessment of the cofactor ability of the fH1-4 fragment in both its labelled and unlabelled forms, with C3b and C3N, it is clear that the fH1-4 fragment is binding, as evidenced by the banding patterns following SDS-PAGE analysis.

In the EPR experiments, there is some indication of interactions of the individual fH fragments when mixed with C3(N), however this interaction is also observed with the labelled fH fragments without C3(N) being present (Figures 3.44 C & E). On their own, there should be no dipolar signal, and so this could perhaps be an indication of protein aggregation, or a multiple labelling effect.

When considering complex formation, the final concentration of each protein in the sample was determined based on $K_d$s obtained from surface plasmon resonance (SPR) of the intact fH:C3b complex.

Work carried out by Barlow (Pechtl *et al*, 2011) used FRET as an orthogonal technique to validate the already proposed crystal structure of the fH1-4 complex. As discussed prior, EPR and FRET are similar techniques used to extract long range distances between probes introduced specifically in samples of interest. Final concentrations of fH1-4 and C3b for the FRET study were 33$\mu$M and 400nm or 400nm and 33$\mu$M, respectively. FRET is more sensitive to concentration than EPR, however in this study the concentrations of fH used exceeds that used in our work. Taking this into

consideration, as well as the excess of C3(N) used, this could somewhat explain the lack of complex formation, as well as the aggregation observed.

## 3.11 Outlook and Future Work

The application of EPR, where cysteine residues must be incorporated and specifically modified, to a system which is protein rich proved far more complex than was first anticipated. Although the cysteine residues in the fH fragments are orchestrated in disulphide bonds, protein aggregation, as well as problems with reducing agents, and spin labelling would suggest that other avenues for the SDSL of fH be explored. Chapter 5 of this thesis looks at the development of new spin labels and spin labelling techniques.

The complement system is tightly orchestrated so as to effectively eliminate pathogens, whilst limiting damage to the host. This thesis has explored the interactions of C3b (or C3(N)) with complement fH, a key regulator of the AP of complement activation. However, C3b also associates with other complement proteins, *e.g.* the Decay Accelerating Factor (DAF). DAF recognises C3b, allowing association of DAF with cell associated C3b, thus preventing the formation of the C3 convertase C3bBb. Preventing convertase formation has an important regulatory role in preventing MAC assembly. Our group has experience with the expression and purification of DAF for spin labelling studies (Lovett *et al*, 2013), and so the interaction of DAF with C3b (C3(N)) could be further explored, to give further insight into its interactions and regulatory role.

The highly reactive TE motif of complement C3b is a rare post-translational modification. As stated previously the relatively short lifespan of the TE bond as it becomes exposed upon conversion from C3 makes this thiol specific labelling site difficult to exploit. However, protection of the TE in C3 is not absolute, and so we have shown that treatment with small nucleophiles facilitates the manipulation of the TE for spin labelling purposes.

Access of the TE *via* chemical modification by nucleophiles is limited to the size of the attacking nucleophile. The MTSSL spin label is relatively small, and able to efficiently label the TE upon freeing of the cysteine residue with methylamine. This highlights the potential for studying TE proteins *in vivo* following chemical

modification with thiol specific probes. TE proteins have major roles in immunity and in host defence, and consequently, advancing our understanding of their interactions *in vivo* could pave the way for the development of and application of novel immunotherapies and in vaccine design (Cole et al., 2009).

Furthermore, TE domains have found to be particularly prevalent in Gram positive bacteria, proving essential for binding to host cells, highlighting their role in pathogenesis.

This opens the avenue of chemical modification of Gram positive bacteria to study their interactions on the cell surface (Walden *et al*, 2015) as well as highlighting the attraction of targeting TEs with small molecules with the potential to inhibit infection.

# **Chapter 4:** Probing the Interaction of the Cardiac Myosin Regulatory Light Chain with Myosin Binding Protein C (cMyBP-C) Using EPR Spectroscopy

**AIMS:** This body of work uses EPR spectroscopy to probe the interactions of the Regulatory Light Chain (RLC) of human cardiac myosin, with the cardiac specific domain, C0, of the regulatory protein cMyBP-C. There are two potential models for the C0-RLC interaction, which potentially influence the relative orientation of the S1 myosin heads. The cardiac specific cMyBP-C domain C0 will be expressed in *E. coli.* The RLC construct will be expressed alongside the MiniHMM (myosin fragment) construct. Mutations in the RLC will then be introduced at sites of interest for EPR studies. If the proposed mechanism stands true, and C0 alters the positions of the S1 myosin heads, there should be a notable difference in distance measurements with and without C0, as confirmed, or otherwise by EPR spectroscopy.

## 4.1 Muscle Contraction - The 'Dance' of Actin and Myosin, and regulation in the sarcomere

The heart is a muscular organ in humans and other vertebrates, which undergoes calcium dependent contraction in order to pump blood through the blood vessels of the circulatory system, providing the body with the oxygen and nutrients essential for survival. The vertebrate heart is principally made of connective tissue and cardiac muscle which belongs to the category of striated muscle, so called due to the striated pattern formed by a series of basic muscle units called sarcomeres (Ford, 2000)

The main protein components of striated muscle are myosin and actin, originally discovered by Kuhne (1859) and Straub (1942) respectively (Szent-Györgyi, 2004), which localise to different positions within the sarcomere, into thick myosin and thin actin filaments (Hanson & Huxley, 1953), as illustrated in Figure 4.1. Muscle contraction occurs as a result of an interaction between actin and myosin, together with many accessory sarcomeric proteins.

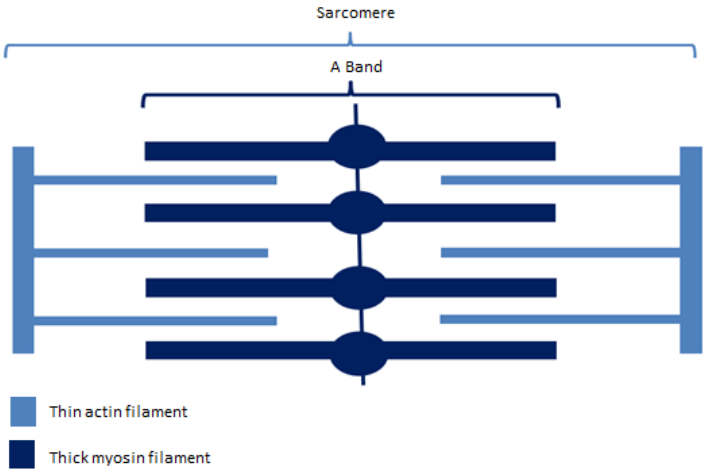



**Figure 4.1: Schematic Representation of Actin and Myosin Positioned in the Sarcomere.** Repetition of the basic unit of muscle – the sarcomere – results in the distinctive pattern associated with striated muscle. The thin filament (composed of the actin monomer and other regulatory proteins) is shown in light blue, whilst the thick myosin filament is shown in dark blue. Figure adapted from Krans *et al*, 2010.

The main protein components of striated muscle are myosin and actin, originally discovered by Kuhne (1859) and Straub (1942) respectively (Szent-Györgyi, 2004), which localise to different positions within the sarcomere, into thick myosin and thin actin filaments (Hanson & Huxley, 1953). Muscle contraction occurs as a result of an interaction between actin and myosin, together with many accessory sarcomeric proteins.

The contraction and relaxation of this striated muscle is the function of a complex macromolecular machine, in which force is generated by filamentous actin and thicker myosin filaments sliding past one another towards the centre of the A-band (Luther & Squire, 2014) (Figure 4.1) in an Adenosine Tri-Phosphate (ATP) dependent manner (Gruen & Gautel, 1999), thus shortening the length of the sarcomere.

Actin and myosin interact in a cyclic series, linked to the hydrolysis of ATP (Hanson & Huxley, 1954). Hydrolysis of ATP generates ADP and a free phosphate (Pi) (Geeves & Holmes, 2000). This de-phosphorylation reaction releases energy, which is then harnessed to drive further chemical reactions, which in this case would be the interaction of actin and myosin, and ultimately actomyosin contraction (Root, 2002).

### 4.1.1 Cardiac Myosin Structure

Myosin exists in the sarcomere as a hexameric structure, composed of two heavy chains and two pairs of light chains, the Essential Light Chain (ELC) and Regulatory Light Chain (RLC). Myosin can be further divided into light - mero - myosin (LMM) and heavy mero – myosin (HMM) (Figure 4.2A), the latter of which can be further divided into the globular N' terminal S1 head group containing the myosin motor domain and light chains, whilst the C' terminal tails (S2) take on a coiled coil structure, holding the head groups in place (Figure 4.2B) (Al-khayat, 2013). The light chains bind the heavy chains in the 'neck' region, the interface between the S1 heads and the S2 tail (Ratti *et al*, 2011).

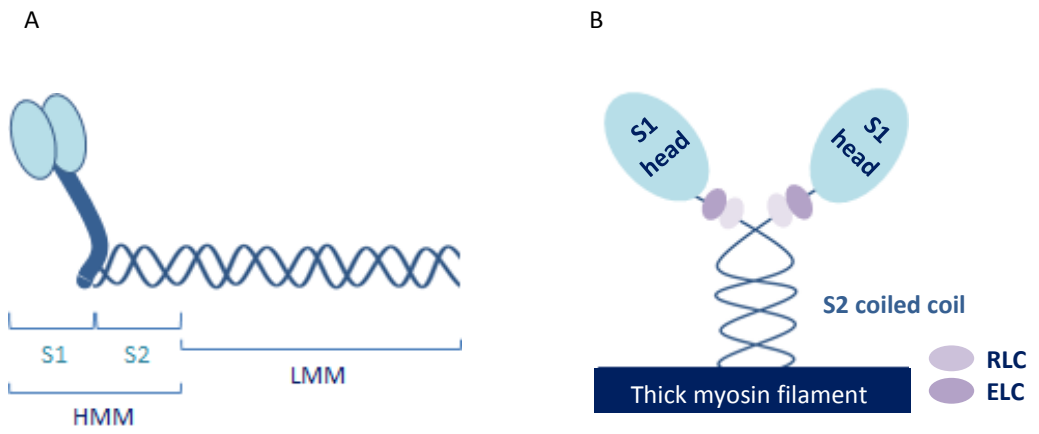



**Figure 4.2: Schematic Representation of the Thick Myosin Filament Structure**. (A) Shows how myosin can be further divided into both heavy and light mero-myosin (HMM & LMM). (B) The heavy chains, composed of the S1 heads and S2 coiled coil, are shown alongside the Essential and Regulatory Light Chains. Figures adapted from (Sadayappan *et al*, 2009).

The globular S1 head of the myosin molecule, also known as the myosin cross - bridge, is the molecular motor that hydrolyzes the ATP necessary for muscle contraction. The formation of these cross - bridges in striated muscle (Huxley, 1957) is regulated by $Ca^{2+}$ (Figure 4.3).

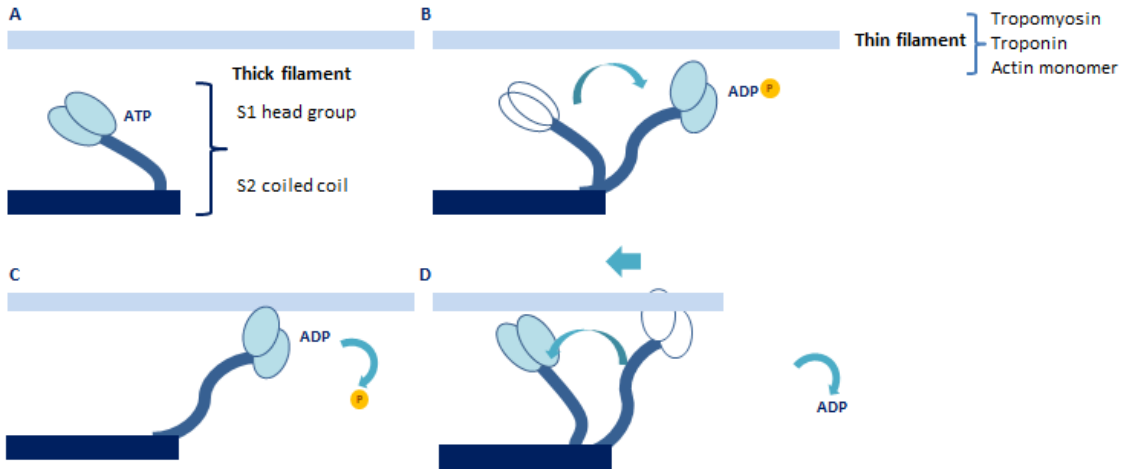



**Figure 4.3: Schematic Representation of the Actin-Myosin Interaction in Striated Muscle.** (A) ATP binding to the myosin head results in a conformation that cannot bind actin. (B) Hydrolysis of ATP to ADP and Pi by the globular S1 heads release energy which is then harnessed to initiate the cross bridge, linking the thin and thick filaments. (C)(D) The cross bridge pulls the actin filament (the 'power stroke'), releasing Pi and ADP from myosin. The myosin heads then bind the actin filament and the power stoke is released.

Binding of ATP in a cleft at the back of the myosin head region causes a conformation of myosin that is unable to bind actin. As the ATP is hydrolysed, the head swings back, approximately 5nm, to a 'cocked' position, and ADP and Pi remain bound. The next stages of the cycle are the force generating steps when the Pi is released from myosin. The head group is free to bind myosin and the 'power stroke' is released. At this stage, the myosin heads are bound to actin in the rigor state (tightly bound). The mechanism of actin-myosin binding allows the 'cross-bridge' formation between the globular myosin head groups and the adjacent actin filaments (Baker & Voth, 2013). This event, which is part of a cycle driven by the hydrolysis of 1 ATP molecule, is known as the 'Power Stroke' (Holmes & Geeves, 2000). The release of ADP releases myosin from the actin filaments. Further hydrolysis of another ATP molecule results in further actin-myosin binding, thus repeating the cycle.

The combined effect of this myriad of power strokes causes muscle contraction, and the motion can be described as myosin 'walking along' the actin filaments (Root, 2002). Although actin and myosin alone can generate force via contraction, they cannot reproduce all of the properties of the contractile system of striated muscle, and so regulatory proteins, such as the RLC control the calcium dependent transitions between

resting and force generation, by blocking actin-myosin interactions in the absence of calcium (Morita *et al*, 1985).

### 4.1.2 The Light Chain Domains – ELC & RLC

The molecular mechanism, whereby actin and myosin are regulated in cardiac muscle contraction is still not fully understood. The light chain domain of myosin, often referred to as the 'lever arm' or 'neck' region, consists of both ELC and RLC, wrapped round the heavy chain through a series of hydrophobic and polar interactions (Ho & Chisholm, 1997) providing stability to the lever arm.

The function of the ELC has not as yet been defined, however, a cluster of mutations within the RLC found to have strong association with heart conditions such as Familial Hypertrophic Cardiomyopathy, linked to high risks of cardiac failure and sudden cardiac death (Longhi et al., 2011), suggest that the RLC perhaps has more purpose than a regulatory role. Selective removal of the RLC causes a change in the structure of the cardiac myosin, leading to myosin disorder and weakens binding to other regulatory proteins (Pant *et al*, 2009) (Garrigos *et al*, 1992). Such links with disease, as well as the impact of complete removal, suggest RLC actively engages in muscle contraction at both a structural and functional level.

### 4.1.3 Myosin Binding Protein-C (cMyBP-C)

In addition to the principal components of the sarcomere; the thick and thin filaments, the sarcomere also contains several accessory proteins that are essential for assembly, maintenance of structural integrity, and regulation of contractile activity. One such regulatory protein identified in recent years is the multi-domain myosin binding protein-C (cMyBP-C), which is involved in both sarcomere formation and in contraction regulation.

cMyBP-C was first detected as an impurity in skeletal muscle myosin preparations, co-purifying with myosin, suggesting a role in myosin-binding (Craig *et al*, 2014). There are three different isoforms of cMyBP-C associated with adult muscle, namely the fast skeletal, slow skeletal and cardiac isoforms, with unique genes encoding each isoform. Cardiac MyBP-C (cMyBP-C) is a137kDa multi-domain protein of the immunoglobulin superfamily of proteins, associated with the thick filament, and contributing 1-2% of the

myofibriallar mass (Winegrad, 1999). The protein is localised to the cross-bridge containing C-zones of striated muscle sarcomeres (Oakley *et al*, 2004) (Figure 4.4).

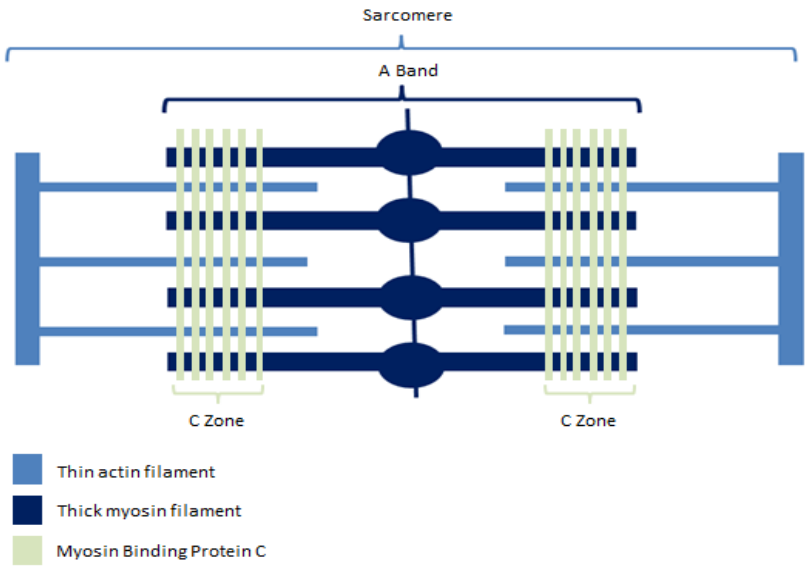



**Figure 4.4: Schematic representation of cMyBP-C positioning in the C zone of the sarcomere, in relation to the thick and thin filaments.** As before, the thin filament (composed of the actin monomer and other regulatory proteins) is shown in light blue, the thick myosin filament is shown in dark blue, and cMyBP-C in green. Figure adapted from (Oakley *et al*, 2004).

The role of cMyBP-C in the sarcomere is complex and not as yet fully understood, however, it has been proposed that as well as playing a structural role, cMyBP-C may have a possible role in force regulation (Witt *et al*, 2001). Of the three isoforms occurring in nature, the cardiac isoform presents some unique features.

The cardiac isoform is exclusively expressed in the heart of mammals and is composed of eleven domains, eight of which are immunoglobulin I – like (IgI), and the remaining three fibronectin type III-like domains. Consisting of 11 modules labelled C0 to C10, from the N' to the C' terminus, the cardiac isoform differs from the skeletal isoforms with three additional characteristics: an additional IgI domain, namely C0, three additional phosphorylation sites between domain C1 and C2 at the N' terminus, and an additional 30 amino acids within the C5 domain (Howarth *et al*, 2012) (Idowu *et al*, 2003) (Figure 4.5).

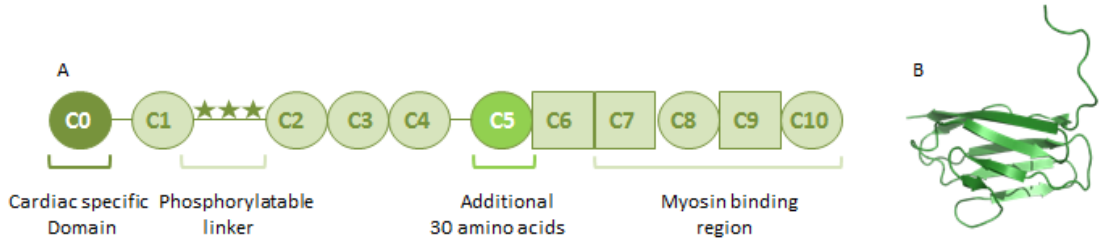



**Figure 4.5: Schematic representation of cMyBP-C showing features unique to the cardiac isoform, including the 3D NMR Structure of domain C0.** (A) Immunoglobulin I type and Fibronectin Type III domains are shown as circles and squares respectively. The additional phosphorylation sites in the C1-C2 linker are represented as stars and the additional thirty amino acids in domain C5 are highlighted. (B) 3D NMR structure of domain C0 of cMyBP-C. (PDB_ID: 2K1M, (Ratti *et al*., 2011)). This structure conforms to the IgI fold typical of the majority of cMyBP-C IgI domains.

cMyBP-C contributes to thick filament structure via interactions at the C' terminus (domains C7-C10) with the LMM section of the myosin rod, and also has a role in contraction regulation due to interactions at its N' terminus with the S2 region of myosin, close to the myosin motor (Ababou *et al*, 2007). cMyBP-C was initially suggested to simply act as a tether, holding the myosin heads close to the neck region of the myosin molecule via its N' terminus (Harris *et al*, 2004), and subsequently preventing cross-bridge formation and actomyosin contraction (Ababou *et al*, 2008) (Figure 4.6).

It is suggested that binding of the N' terminal portion of cMyBP-C to myosin in the region of the S1 heads holds the myosin heads back, reducing their ability for actin interaction and force generation. This returns to normal upon phosphorylation of cMyBP-C, which results in the release of the head groups (Figure 4.6A). Phosphorylation occurs at three sites in cMyBP-C (between domains C1 and C2) (Figure 4.5). This 100-residue region, the cMyBP-C motif, is highly conserved between all isoforms of cMyBP-C and between species (Kunst *et al*, 2000).

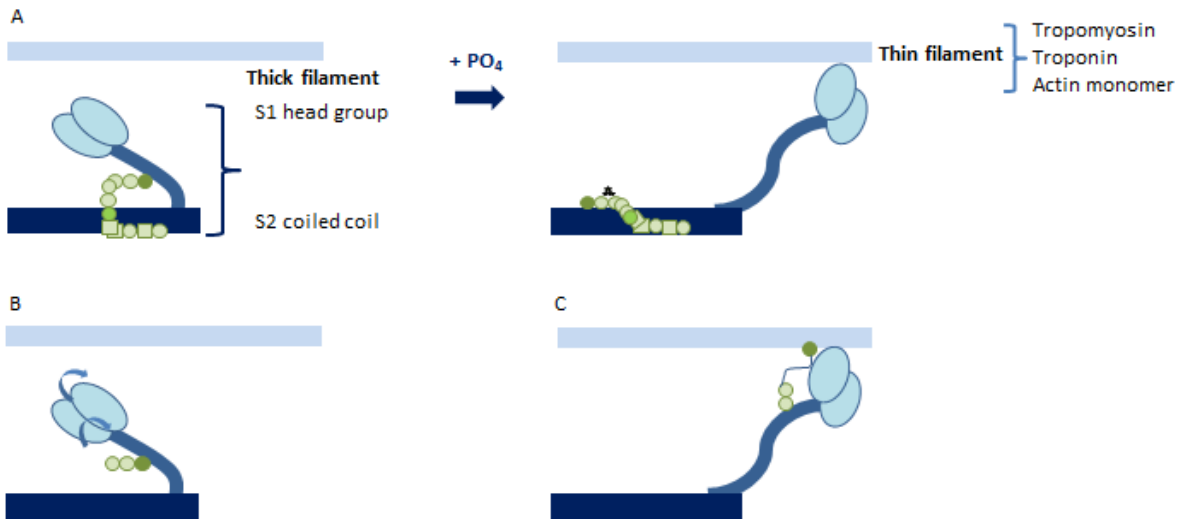



**Figure 4.6: Cartoon depicting the current understanding of cMyBP-C Function.** (A) Effect of phosphorylation of the N' terminal myosin binding site. cMyBP-C is dislodged, releasing the S2 coiled coil and promoting cross-bridge formation and ultimately muscle contraction, via interaction of the S1 cross-bridge with the thin filament. (B) A possible interpretation of the observation that cMyBP-C constructs too short to act as a tether can influence muscle contraction. In this interpretation, the N'-terminus directly affects the orientation of the myosin head group, indicated by the arrows. (C) Alternative interpretation whereby the cross-bridge formation occurs as a direct result of the N' terminal domains on cMyBP-C interacting with the thin actin filaments. Figure adapted from Ratti *et al*, 2011.

It is suggested that binding of the N' terminal portion of cMyBP-C to myosin in the region of the S1 heads holds the myosin heads back, reducing their ability for actin interaction and force generation. This returns to normal upon phosphorylation of cMyBP-C, which results in the release of the head groups (Figure 4.6A). Phosphorylation occurs at three sites in cMyBP-C (between domains C1 and C2) (Figure 4.5). This 100-residue region, the cMyBP-C motif, is highly conserved between all isoforms of cMyBP-C and between species (Kunst *et al*, 2000).

It is suggested that the function of cMyBP-C extends beyond that of a mere tether, as cMyBP-C fragments which are too short to exhibit this tethering role can still influence muscle contraction (Figure 4.6B). Knockout mouse studies with altered cMyBP-C show that domains C0-C2 are sufficient for cross-bridge formation, showing cMyBP-C fragments too short to perform the tether role could still bridge the gap between the S1 heads and the thin filaments (Harris *et al*, 2002).

Furthermore, using knock-in mouse models (Witt *et al*, 2001) carrying a gene for shortened cMyBP-C without the cardiac specific C0 domain, all animals with the deletion were still viable with no significant ultrastructural changes to the heart, or

impaired lifespan. The mutant cMyBP-C was still phosphorylated, however, there appeared to be a decrease in the levels of force generated, suggesting the additional domain in the cardiac isoform could be an evolutionary change, introduced by nature to aid in force regulation at the cross-bridge level (Witt *et al*, 2001).

The presence of the extra C0 domain as well as the phosphorylatable linker between domains C1 and C2 suggest this may be vital for effective regulation of cardiac muscle contraction at the cross bridge level, and how extensions of these finding could suggest a mechanism whereby N' terminal mutations in the cMyBP-C could cause Familial Hypertrophic Cardiomyopathy.

### 4.1.4 The RLC-C0 interaction

It is clear from the above that the C0 domain is crucial for cMyBP-C participation in muscle contraction regulation, and that this is perhaps by more direct means than simply acting as a mere tether (Granzier & Campbell, 2006). An interaction between the cardiac isoform of cMyBP-C and the RLC has been proposed as early as 1985 (Margossian, 1985) and most likely occurs via the cardiac specific C0 domain.

NMR, ITC, and *in vivo* immuno-fluorescent localisation of C0 (Ratti *et al*, 2011) places the N' terminus of cMyBP-C right on top of the S1-S2 hinge interface, and therefore in the immediate vicinity of the light chains. In this position, cMyBP-C is ideally situated to directly interact with the hinge region, potentially influencing the position of the S1 heads, an interaction which is attributed to the cardiac specific domain C0. The ability to manipulate the S1-S2 junction would give cMyBP-C the potential to directly adjust the position of the S1 heads and therefore subtly influence muscle contraction.

Confirmation of an interaction between C0 and RLC drives this study, which aims to use EPR spectroscopy to further characterise this interaction, and its contribution to muscle contraction.

### 4.1.5 Aims of this work

There are two potential models for C0-RLC interaction, suggesting that C0 could bind both RLCs on one myosin molecule simultaneously (Figure 4.7A), and in this position could influence the relative orientation of the S1 heads. Wedging C0 between the RLCs

could push the S1 heads apart, or pull them together. Otherwise it is proposed that C0 binds only one RLC leading to an asymmetry of the S1 heads (Figure 4.7B). It is most likely that the former is the case, reinforced by data accumulated from both structural and binding studies.

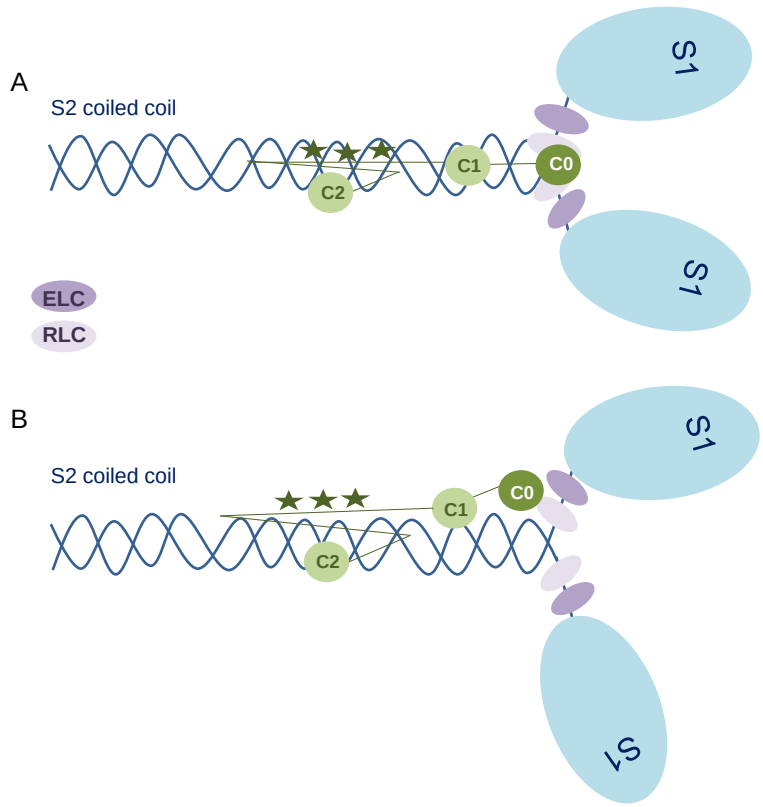



**Figure 4.7: Model for the arrangement of the N' terminus of cMyBP-C around the myosin S1-S2 junction and the light chains.** cMyBP-C domains C1 and C2 bind to the S2 chain very close to the S1/S2 junction, whereas C0 binds to the RLC. Illustrated here are the two proposed mechanisms whereby domain C0 can interact with the RLCs. In (A), C0 could be placed between the RLC, and in such a position could change the position of the S1 heads by pulling them closer together, or by pushing them further apart. Alternatively in (B), domain C0 could interact with only one RLC, leading to an asymmetry of the S1 heads.

It is the aim of this project to first express and purify cMyBP-C domain C0 in *E. coli*. The RLC construct will be expressed alongside the MiniHMM (myosin fragment) construct. Mutations in the RLC will then be introduced at sites of interest for EPR studies.

EPR measurements will then be made, with and without C0. If the proposed mechanism stands true, and C0 alters the positions of the S1 heads by bringing them

closer together, or by pushing them apart, there should be a notable difference in distance measurements with and without C0, as confirmed, or otherwise by EPR spectroscopy. No matter which model is preferred following EPR, we hope to learn more about the positioning of the RLC in myosin, in relation to regulatory proteins such as cMyBP-C, with these preliminary results opening avenues for further research.

## 4.2 Small Scale Test Expressions of cMyBP-C C0 and RLC/MiniHMM

Before expression of the recombinant proteins was carried out on a large scale, expression was first tested on a small scale (100mL). Figure 4.8 shows the resulting SDS-PAGE analysis of samples taken from the crude cultures before and after induction for the cMyBP-C C0 domain and the RLC/MiniHMM Complex.

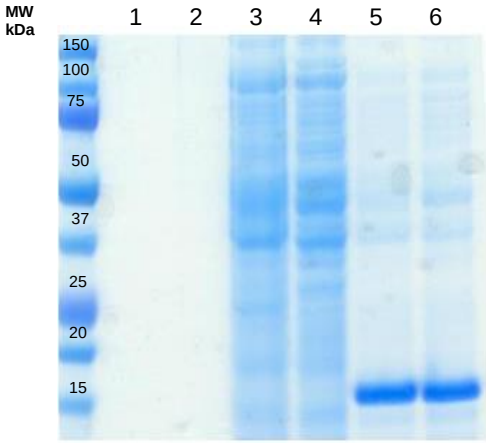



**Figure 4.8: SDS-PAGE Analysis Following Small-Scale Test Expression (100mL) of the C0 and RLC/MiniHMM Constructs**. Sample lanes 1 and 2 are samples taken from the crude cultures before induction. Sample lanes 3 and 4 are samples taken from the crude RLC/MiniHMM cultures, run both reduced and non-reduced, respectively. Sample lanes 5 and 6 are samples taken from the crude C0 culture, run both reduced and non-reduced, respectively.

Following small scale test expression, expression of the C0 domain of cMyBP-C is clear, however, it appears that the RLC/MiniHMM complex is not expressing in the soluble fraction. cMyBP-C domain C0 was then expressed and purified on a larger scale before problems with the RLC/MiniHMM were addressed.

## 4.3 Expression and Purification of cMyBP-C Domain C0

NMR and binding studies have confirmed an interaction between the cardiac specific C0 domain of cMyBP-C and the cardiac RLC of myosin. The cardiac specific domain C0 was expressed and purified using an *E. coli* expression system, as described in 2.3.2. The C0 construct contains a C' terminal hexa-histidine tag, and therefore IMAC can be used as the first catchment step. Figure 4.9 shows the elution chromatogram following IMAC and corresponding SDS-PAGE analysis, following purification of the construct.

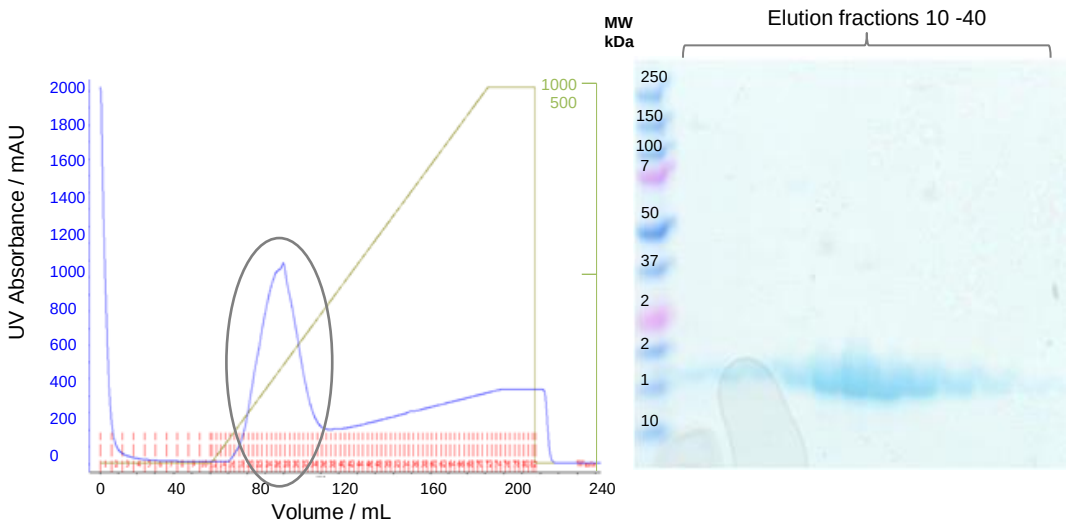



**Figure 4.9: Elution profile and corresponding SDS-PAGE analysis for the cMyBP-C Domain C0 Following IMAC.** Every third fraction from 10-40 beneath the circled peak was analysed.

Following determination of purity by SDS-PAGE analysis, fractions 10 - 40 were pooled and concentrated before being dialysed overnight at 4$^{o}$C into non-imidazole buffer. The final yield was determined as ~20mg from 1Litre of culture. The protein was diluted with a half volume of glycerol before being stored at -20$^{o}$C. A sample was submitted for Mass Spectrometry to the SIRCAMS facility, University of Edinburgh.

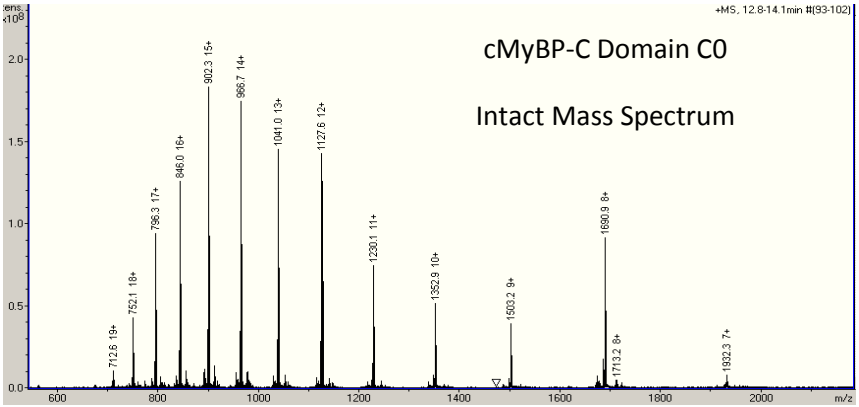



**Figure 4.10: Mass Spectrum of cMyBP-C, domain C0, performed at SIRCAMS, using FT–ICR mass spectrometry.**

MS was performed in order to obtain an intact and accurate mass for the cMyBP-C domain C0. The mass obtained (13,519.2 Da) was as predicted using the ExPASy Molecular weight calculator (13,520.05 Da).

## 4.4 Optimisation of Expression of the WT RLC/MiniHMM Complex

In order to characterise the interaction of the RLC with C0, the RLC was co-expressed with a myosin fragment – 'MiniHMM' containing residues 806-963 (806-835 = RLC binding site) of the S2 coiled coil. Co-expression of the RLC with the MiniHMM ensures proper folding of the RLC, and provides stability to the binding site, promoting formation of the complex (Ratti *et al*, 2011) shown in Figure 4.11.

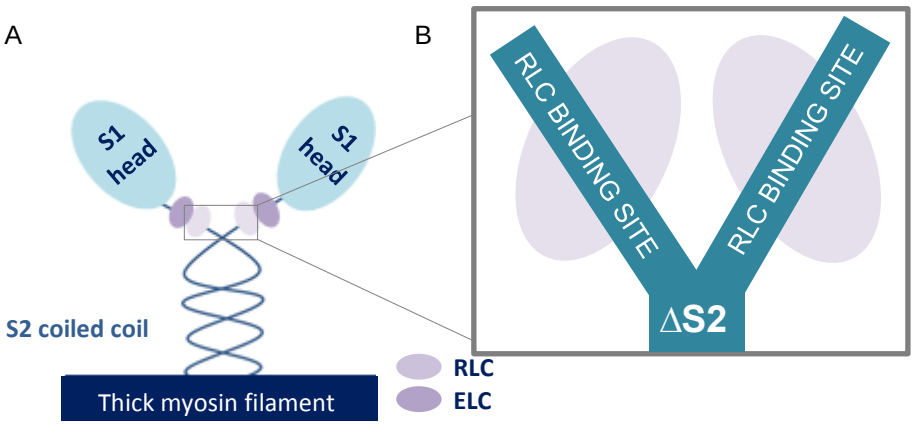



**Figure 4.11: Schematic Representations of the Myosin Heavy and Light chains (A), zoomed in showing the MiniHMM Fragment, with bound RLC (B).**

Expression of the WT RLC/MiniHMM complex was optimised on a small scale, before cysteine mutagenesis was carried out, and mutants expressed. Previous work using the RLC/MiniHMM complex has shown that the protein expression is partly insoluble, and expression levels irregular (Mark Pfuhl, personal communication), and so optimisation of expression should facilitate an easier transfer of protocols to cysteine mutants. Following adjustments to induction time, temperature and IPTG concentrations the majority of the protein was expressed in the insoluble fraction.

### 4.4.1 Modification of the MiniHMM Construct

The original MiniHMM construct contains two cysteine residues. Shown in Figure 4.12 is the structure of the coiled coil, highlighting cysteine positioning.

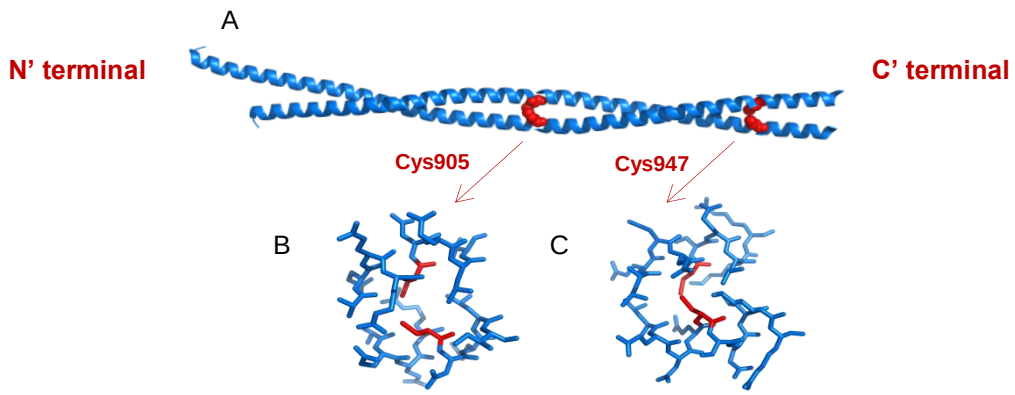



**Figure 4.12: Crystal structure of human cardiac beta-myosin II S2Δ** (PDB_ID: 2FXM, Blankenfeldt *et al*, 2006).

From Figure 4.12, it is clear that the cysteine residues are in such close proximity, that it is likely that they form disulphide bonds. This was confirmed using mutagenesis studies, however the presence of disulphide bonds in the S2Δ, led to an instability, and precipitation of the protein at temperatures above 25$^{o}$C (Mark Pfuhl, personal communication).

Typically, coiled coil structures are based on the hydrophobic interactions of two helices, rather than an interaction based on disulphide bond formation. To optimise packing, the two helices wrap around one another, hence the coiled coil structure, which are often based on a sequence motif known as the leucine zipper, as illustrated in Figure 4.13.

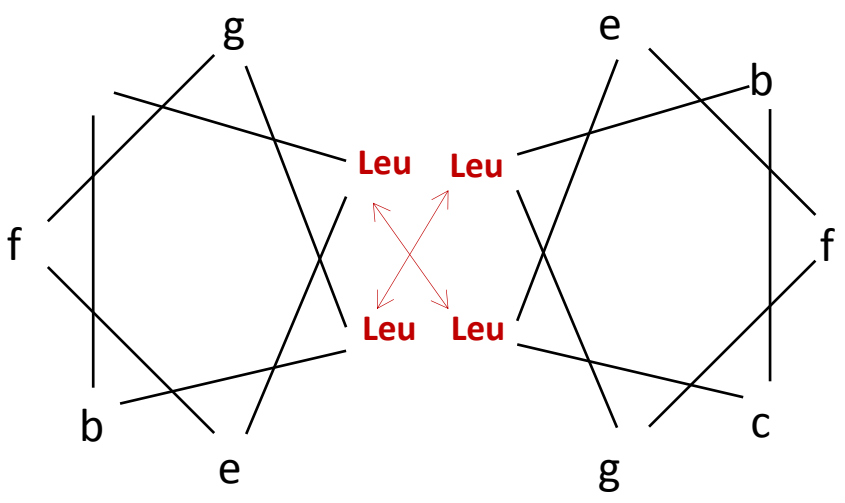



**Figure 4.13: Schematic representation of the Leucine Zipper motif.**

The leucine residues form a hydrophobic core, stabilising the coiled coil structure, holding it in place. Therefore substituting the cysteine residues for leucine residues, as shown in Figure 4.14, promotes the leucine zipper formation, by eliminating the potential to form disulphide bonds, whilst improving stability.

```
Query  1    MHHHHHHSTENLYFQGSSERRDSLLVIQWNIRAFMGVKNWPWMKLYFKIKPLLKSAEREK  60
Sbjct  1    MHHHHHHSTENLYFQGSSERRDSLLVIQWNIRAFMGVKNWPWMKLYFKIKPLLKSAEREK  60
Query  61   EMASMKEEFTRLKEALEKSEARRELEEKMVSLLQEKNDLQLQVQAEQDNLADAEERCDQL  120
Sbjct  61   EMASMKEEFTRLKEALEKSEARRELEEKMVSLLQEKNDLQLQVQAEQDNLADAEERLDQL  120
Query  121  IKNKIQLEAKVKEMNERLEDEEEMNAELTAKKRKLEDECSELKRDIDDLELTLAK  175
Sbjct  121  IKNKIQLEAKVKEMNERLEDEEEMNAELTAKKRKLEDELSELKRDIDDLELTLAK  175
```

**Figure 4.14: Sequence of the original MiniHMM construct aligned with the modified MiniHMM.** In the new construct the Cys residues at positions 117 (905) and 160 (947) are substituted for Leu residues.

From this point the RLC/MiniHMM complex refers to the RLC bound to this modified cysteine-free MiniHMM complex.

## 4.5 Large Scale Expression of wt RLC/MiniHMM

In order to avoid the RLC/MiniHMM complex being expressed in the insoluble fraction, both the expression and cell lysis conditions were altered from standard protocol (Growth at 37$^{o}$C, induction at 37$^{o}$C; 5hr with 1mM IPTG, lysis by sonication)

to the methods outlined in Chapter 2 which were developed and optimised on a visit to Dr Lena Rostkova from Dr Mark Pfuhl's group at King's College London.

In cooling induction temperatures and reducing the final concentration of IPTG used for induction, this encourages a reduced rate of expression, which promotes expression of more soluble protein (Graslund, 2008), whilst lysis using repeated freeze thaw techniques ensures all cells are fully lysed, and the protein not exposed to high temperatures unnecessarily during the sonciation procedure.

### 4.5.1 IMAC of wt RLC/MiniHMM

The MiniHMM construct contains an N' terminal hexa-histidine tag, and therefore IMAC can be used as the first catchment step. The histidine tag can then be removed using TEV protease, as described in 2.3.1, 2.5.2, and 2.5.2.1.

Figure 4.15 shows the resulting SDS-PAGE analysis from large scale expression of the WT RLC/MiniHMM complex following modification to protocols and expression using IMAC.

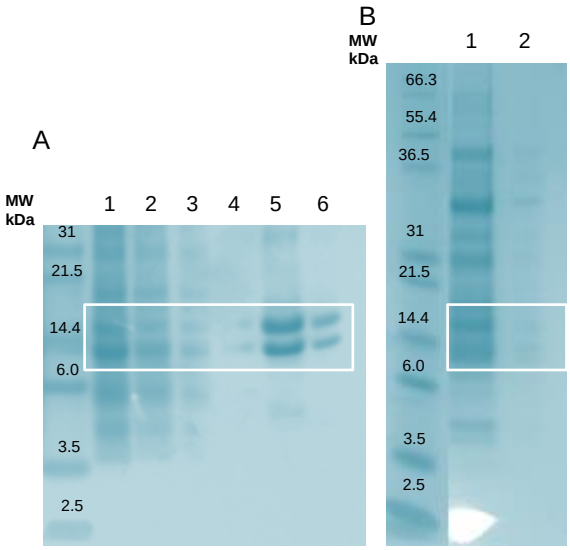



**Figure 4:15 SDS-PAGE Analysis Following Large Scale Expression and Purification of WT RLC/MiniHMM Using IMAC.** (A) Samples taken throughout the purification process. 1. Crude cell-free extract before IMAC. 2. Flow-though from IMAC. 3. Wash. 4. Wash 5. Elution Fraction. 6. Elution Fraction. (B) 1. Flow-through from IMAC following TEV (cleaved protein) 2. 500mM imidazole wash.

Following purification of the wt RLC/MiniHMM complex, the protein has been solubilised, and final yields were ~5mg/1L of culture. It is evident that modification to

both the expression and lysis protocols promotes expression of a soluble RLC/MiniHMM complex, as the MiniHMM can be seen running at ~15kDa, and the RLC at ~20kDa, as indicated by the box. Although TEV protease was used for cleavage of the hexa-histidine tag, (Figure 4.15B), the cleavage reaction was not complete, as there appears to be some uncleaved material in elution fraction 2 (Figure 4.15B), as assessed by SDS-PAGE.

### 4.5.2 Size Exclusion Chromatography of wt RLC/MiniHMM

Following the cleavage reaction, the wt RLC/MiniHMM was further purified using SEC.

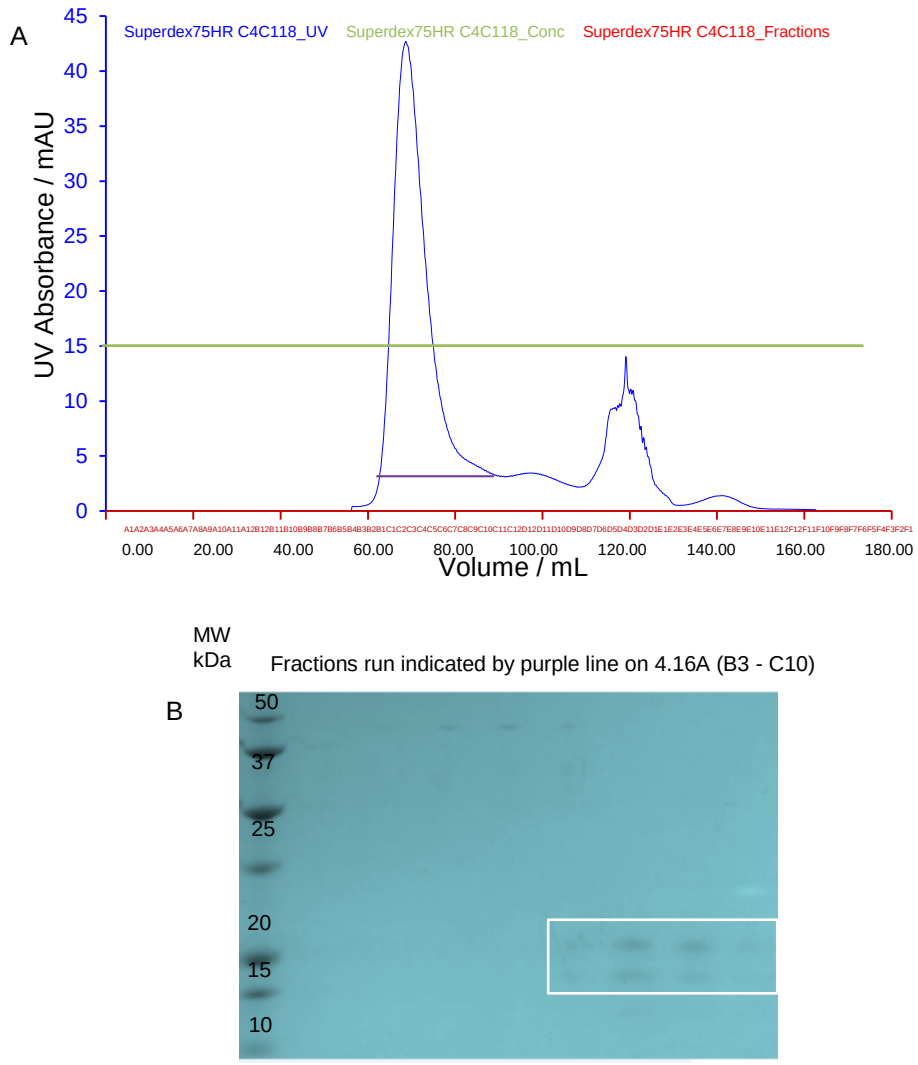



**Figure 4.16: SEC elution profile (A) and SDS-PAGE analysis (B) for wt RLC/MiniHMM**

Figure 4.16 shows the SDS-PAGE analysis and elution profile for the wt RLC/MiniHMM construct. The protein was treated with a final concentration of 10mM DTT before SEC was performed, to mimic the treatment for cysteine mutants, and allow comparison. From the SDS-PAGE analysis, the wt RLC/MiniHMM complex has eluted from the SEC column in the first major peak, whilst the DTT elutes in the second. The protein does not elute in the centre of the main peak, but instead towards the shoulder (fractions C7-C10).

Following purification of the wt RLC/MiniHMM complex, the protein has been solubilised, and final yields were ~5mg/1L of culture. It is evident that modification to both the expression and lysis protocols promotes expression of a soluble RLC/MiniHMM complex, as the MiniHMM can be seen running at ~15kDa, and the RLC at ~20kDa, as indicated by the box. Although TEV protease was used for cleavage of the hexa-histidine tag, the cleavage reaction was not complete.

## 4.6 Generation of Cysteine Mutants of the myosin RLC for SDSL Purposes

Mutations have been made in the chicken isoform of the myosin RLC for the purposes of fluorescence studies (Hopkins *et al*, 1998) and exchanged into muscle fibres to check the effect of the mutation on force. Consequently, the mutations from the chicken isoform could be transferred onto the human isoform due to significant sequence homology, and SDM performed to generate the cysteine mutants.

```
Chicken Query  1    MAPKKAKRRAAEGSSNVFSMFDQTQIQEFKEAFTVIDQNRDGIIDKDDLRETFAAMGRLN  60
Human   Sbjct  1    MAPKKAKKRAGGANSNVFSMFEQTQIQEFKEAFTIMDQNRDGFIDKNDLRDTFAALGRVN  60
Chicken Query  61   VKNEELDAMIKEASGPINFTVFLTMFGEKLKGADPEDVIMGAFKVLDPDGKGSIKKSFLE  120
Human   Sbjct  61   VKNEEIDEMIKEAPGPINFTVFLTMFGEKLKGADPEETILNAFKVFDPEGKGVLKADYVR  120
Chicken Query  121  ELLTTQCDRFTPEEIKNMWAAFPPDVAGNVDYKNICYVITHGEDKE  166
Human   Sbjct  121  EMLTTQAERFSKEEVDQMFAAFPPDVTGNLDYKNLVHIITHGEEKD  166
```

**Figure 4.17: Sequence alignment of the Chicken RLC with the Human RLC.** Mutations chosen in the human protein were selected based on functional cysteine mutants in the chicken isoform. Mutations are highlighted underneath the residue.

Figure 4.17 shows the sequence of the chicken RLC aligned with the human RLC. The residues selected in the chicken sequence have been successfully mutated to cysteine residues, and so the corresponding residues in the human sequence (K30, E88, E97, D117, and T125) were selected for SDSL purposes. Mutations were selected in both N' and C' terminal regions to optimise potential for observing changes in distance measurements with and without C0.

## 4.7 MMM Simulation of the DEER Experiment for RLC/MiniHMM Cysteine Mutants

As stated above, cardiac myosin is composed of 4 chains, two heavy and two light. The two light chains, the RLC and ELC, bind the heavy S2 coiled coil chain, in the vicinity of the heavy S1 head group, as sown in Figure 4.18.

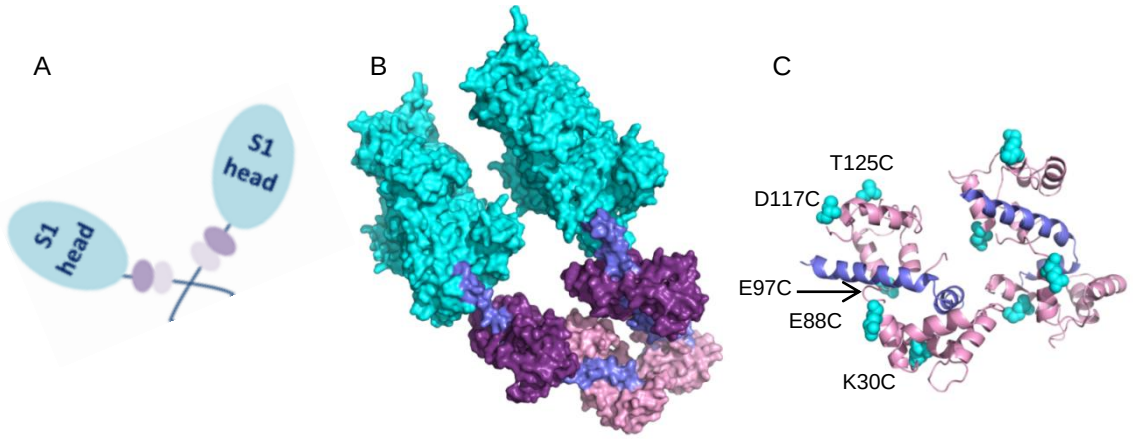
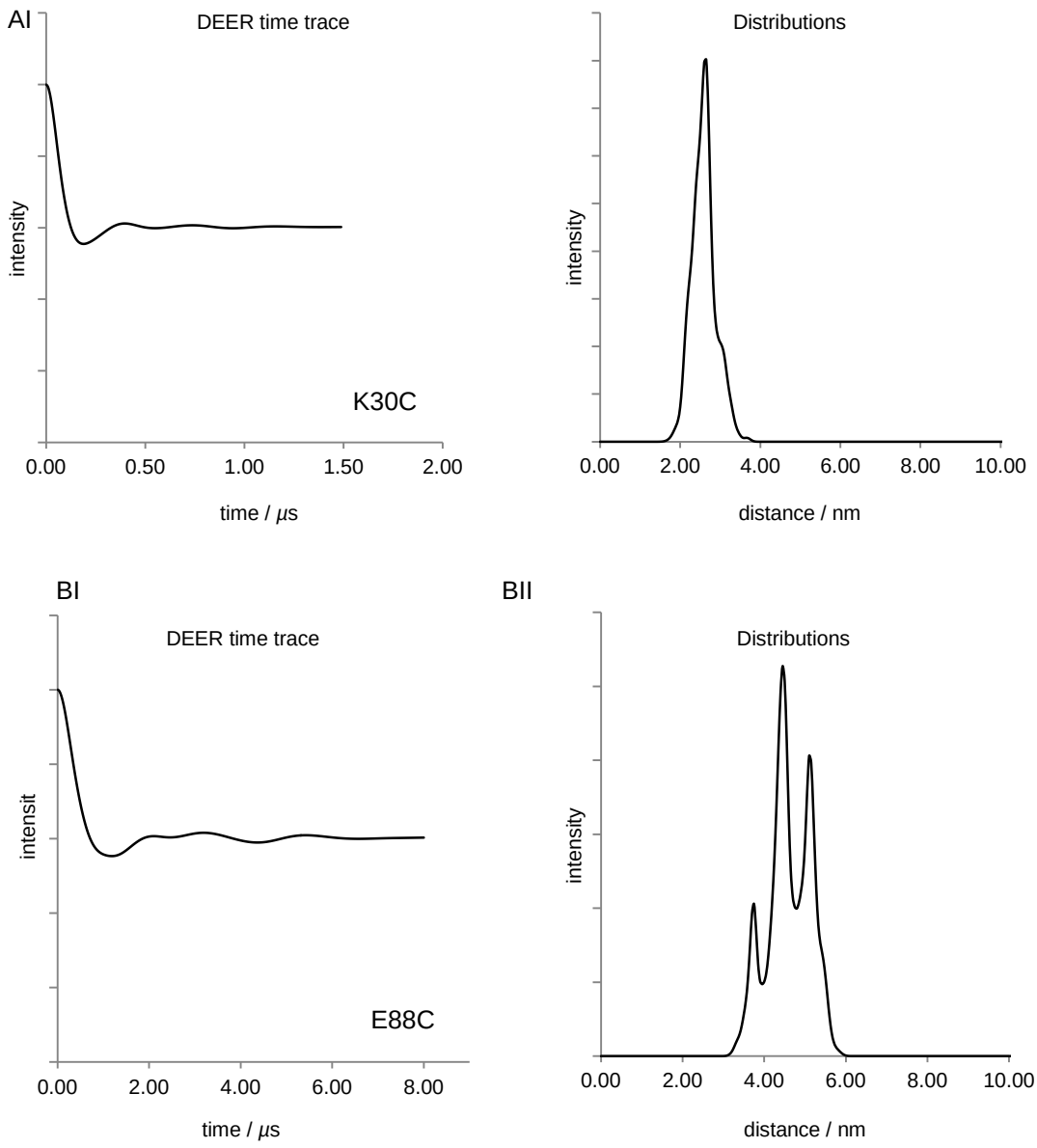



**Figure 4.18: Structures illustrating the positioning of the RLC in the S2 junction, in the vicinity of the S1 heads.** (A) Schematic representation and (B) Surface representation of the S1 heads (cyan), with the RLC (pink) and ELC (purple) bound to the S2 junction (Modelled using chicken proteins, from which our mutations are based) (C) RLC (pink) bound to part of the S2 (purple) with cysteine mutations represented as cyan spheres. (B and C visualised using PDB_ID:1M8Q, Chen *et al*, 2002).

Figure 4.18 shows the structure of the chicken proteins, from which our mutations have been based. Using MMM, the DEER time traces and distance distributions were modelled between pairs of RLC bound to the MiniHMM complex, in the off state, *i.e.* without C0 (Figure 4.19). All show distances that should be measurable by DEER. From Figure 4.18C, and from models generated by Dr Mark Pfuhl based on the

sequence of the human cardiac RLC, we hypothesise that if the C0 wedges between the RLC proteins and moves them apart, the K30C distance should provide the measurement that is both easiest to measure with DEER and will change upon C0 binding.



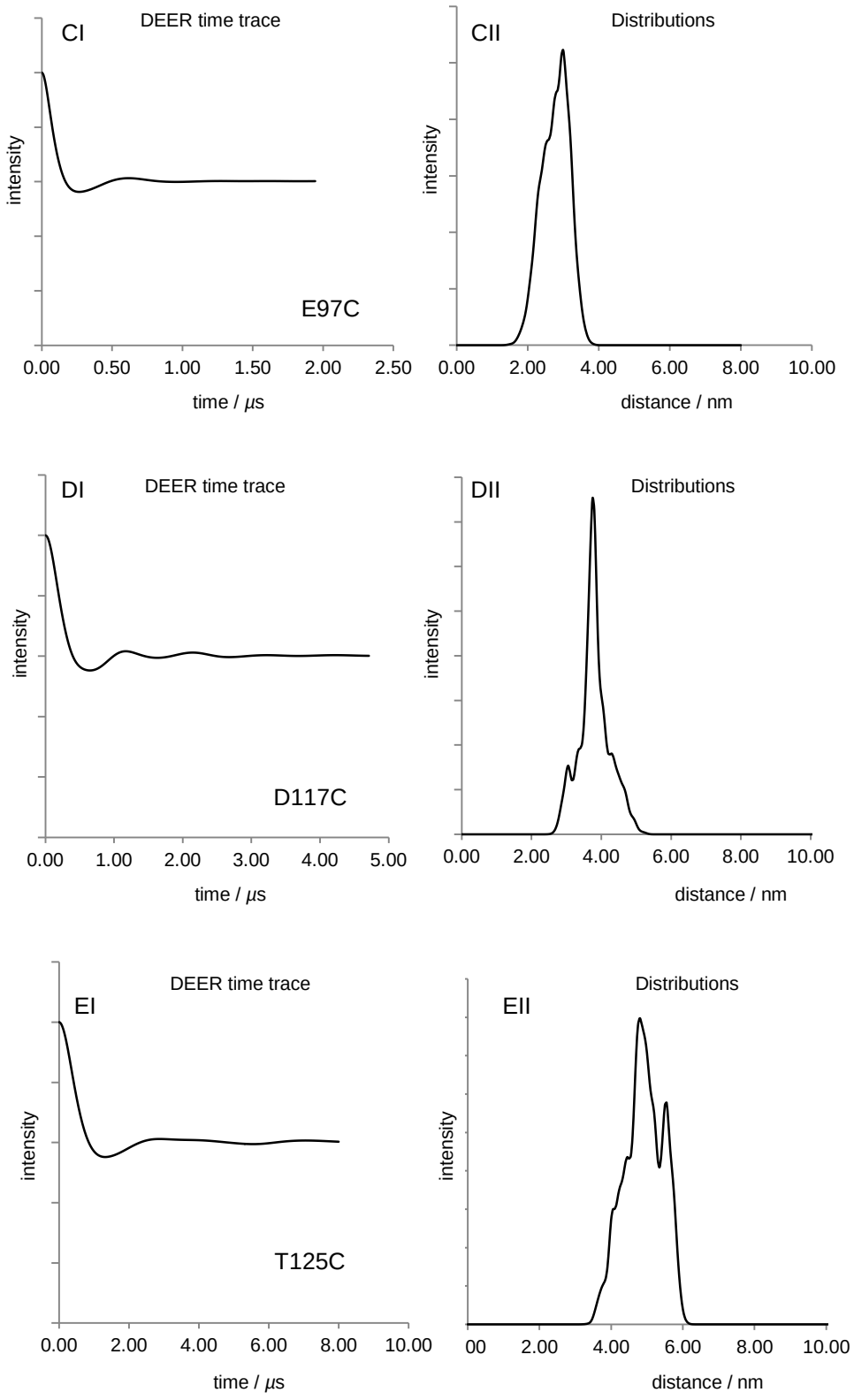



**Figure 4.19: MMM Simulations of the DEER Time Traces (I) and Distance Distributions (II) shown alongside PyMOL cartoons illustrating the site for cysteine mutagenesis**. These calculations were performed on the chicken structure (PDB_ID:1M8Q, Chen *et al*, 2002) **(A)** K30C **(B)** E88C **(C)** E97C **(D)** D117C **(E)** T125C.

## 4.8 Generation of Cysteine Mutants of the myosin RLC for SDSL Purposes

Mutations have been made in the chicken isoform of the myosin RLC for the purposes of fluorescence studies (Hopkins *et al*, 1998) and exchanged into muscle fibres to check the effect of the mutation on force. Consequently, the mutations from the chicken isoform could be transferred onto the human isoform due to significant sequence homology, and SDM performed to generate the cysteine mutants

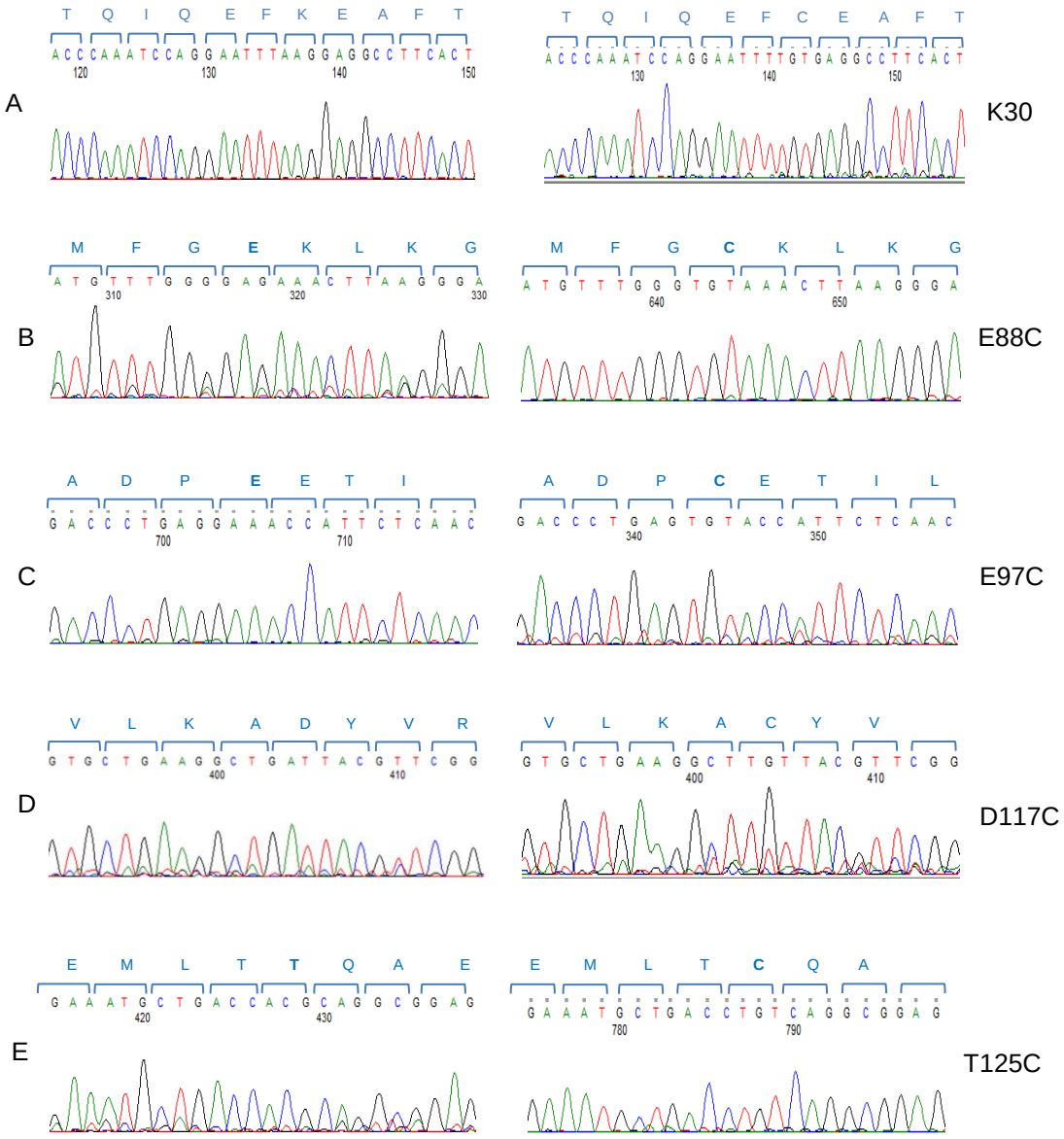



**Figure 4.20: Sequencing chromatograms, comparing the wt sequence to the sequence following SDM, showing successful substitution for cysteine in all cases.** (A) K30C (B) E88C (C) E97C (D) D117C and (E) T125C.

Figure 4.20 shows sequencing chromatograms for the five of the single cysteine mutants chosen for this study, compared to the wt.

## 4.9 Large Scale Expression and Purification of cardiac Myosin RLC Cysteine Mutants

Utilising the histidine tag on the MiniHMM fragment, each of the cysteine mutants were expressed using the modified expression and lysis protocols, before being purified using IMAC, with manual elution using a peristaltic pump. Figure 4.21 shows the SDS-PAGE analysis from the purification of each of the four cysteine mutants.

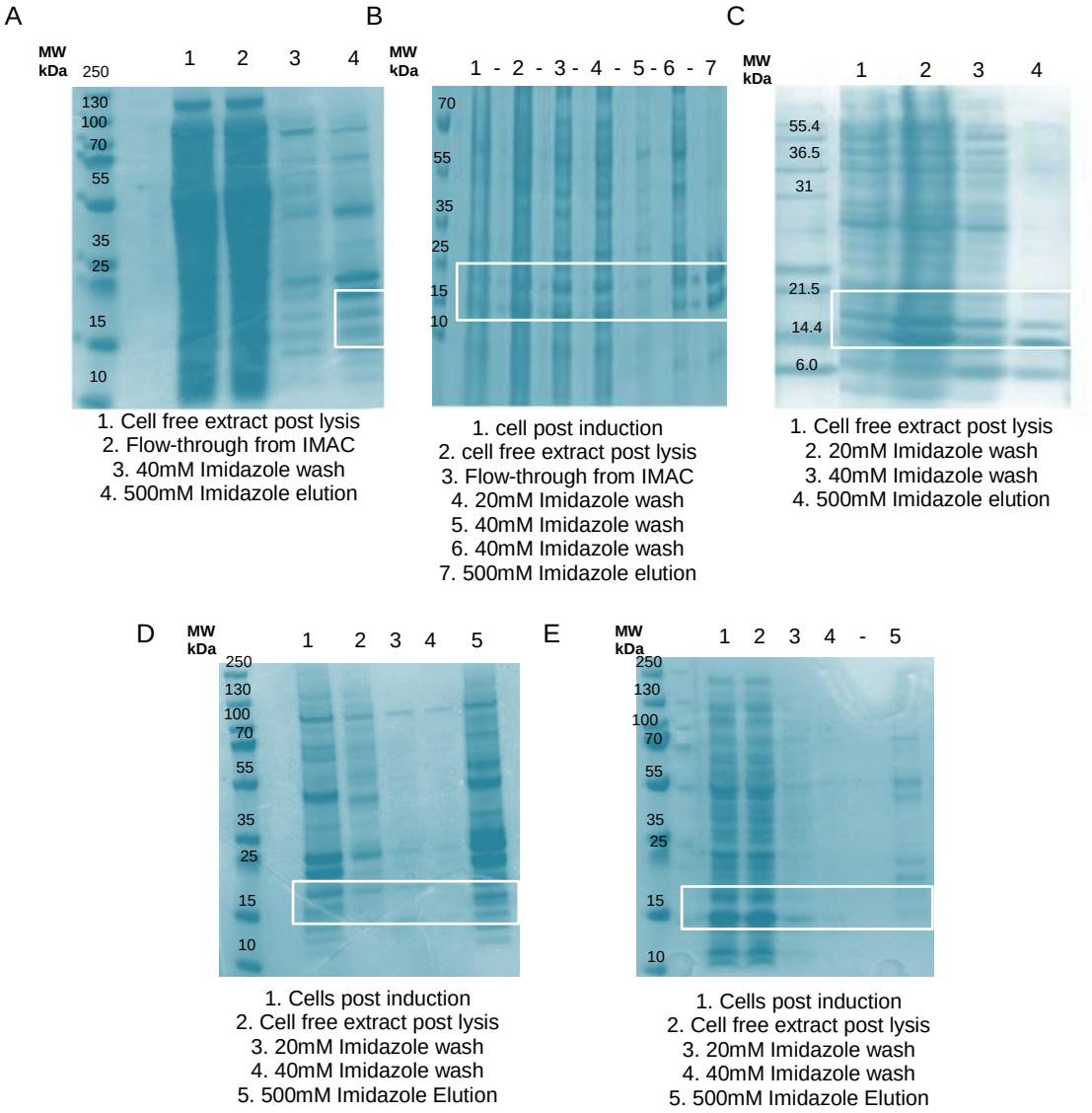
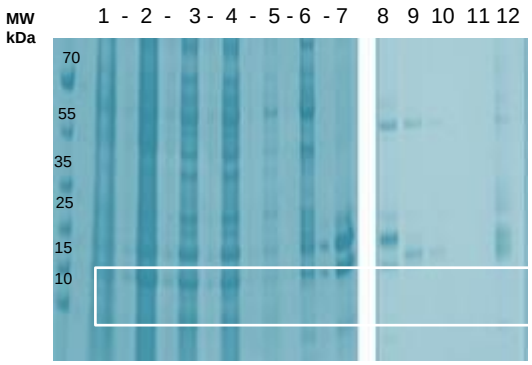



**Figure 4.21 SDS-PAGE analysis following purification of each of the Cysteine RLC/MiniHMM Mutants following IMAC.** (A) purification of K30C (B) purification of E88C (C) purification of E97C (D) purification of D117C and (E) purification of T125C.

As for the wt RLC/MiniHMM complex, complete cleavage of the hexa-histidine tag was not achieved for the cysteine mutants. Following passage of the 500mM elution fraction over the Ni-NTA column (post –TEV cleavage), the RLC/MiniHMM complex did not bind to the column, indicative of cleaved histag. The column was washed with buffer containing 500mM imidazole which serves to remove bound cleaved histag and the TEV protease. Figure 4.22 shows the SDS-PAGE analysis of elution from the Ni-NTA column for the E88C mutant following cleavage of the Histag by TEV protease.




1. cell post induction
2. cell free extract post lysis
3. Flow-through from IMAC
4. 20mM Imidazole wash
5. 40mM Imidazole wash
6. 40mM Imidazole wash
7. 500mM Imidazole elution
8. Flow-through (cleaved Protein)
9. 20mM Imidazole wash
10. 20mM Imidazole wash
11. 40mM Imidazole wash
12. 500mM Imidazole elution

**Figure 4.22: SDS-PAGE analysis of the E88C RLC/MiniHMM Mutant, showing purification using IMAC and elution from the Ni-NTA following Histag Cleavage.**

Following the cleavage reaction the sample was flowed over the Ni-NTA cartridge. The cleaved protein now contains no histidine tag and so does not bind (sample lane 8). Following washing with a final concentration of 500mM imidazole there appears to be some RLC/MiniHMM complex in this fraction. This is indicative of an incomplete cleavage reaction, as some complex remains bound to the column *via* the histidine tag. At this stage, only the cleaved fraction was used for subsequent experiments for each mutant.

Expression levels varied considerably between mutants. Illustrated in Table 4.1 are final yields obtained for the wt protein, compared to the single cysteine mutants.

**Table 4.1: Final Yields of RLC/MiniHMM Mutants per Litre of Cell Culture**

| RLC/MiniHMM Mutant | Yield / L. of Culture |
|---|---|
| wt | 5.2mg |
| K30C | 4.75mg |
| E88C | 1.2mg |
| E97C | 0.4mg |
| D117C | 0.8mg |
| T125C | 0.2mg |

It is clear looking at Table 4.1 that the final yields for all cysteine mutants vary considerable from the wt. For each mutant, typically 4L of *E. coli* culture was grown, and even when scaled up to 12L there was little effect on final yields. Of the five mutants selected, only the K30C and E88C mutants showed expression levels that would yield enough protein for future experiments.

Expression was repeated numerous times for the remaining mutants, however expression and purity at this stage could not be improved. K30C and E88C were taken on since they would give a good preliminary idea of whether the inter-RLC distance changes upon C0 binding.

### 4.9.1 Further Purification of K30C RLC/MiniHMM

Although Table 4.1 gives the final yield of the K30C mutant, this is not a true representation of the final yield, as the yield was calculated based on $A_{280}$, and it is clear from the SDS-PAGE analysis in Figure 4.21A, that there are a lot of contaminating proteins that will be contributing towards the $A_{280}$ value given. Figure 4.23 shows the SDS-PAGE analysis of the K30C mutant, following cleavage of the histidine tag.

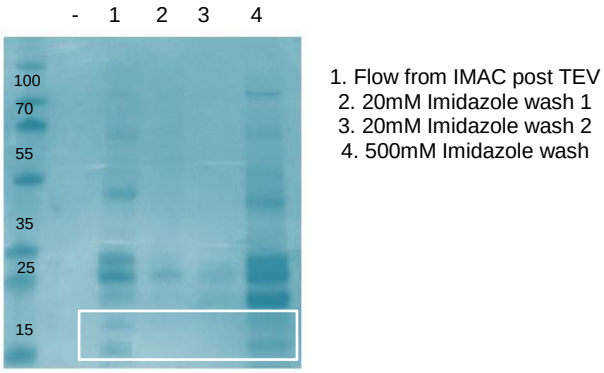



**Figure 4.23: SDS-PAGE analysis of the TEV-treated K30C mutant following passage over the IMAC column.**

Following the cleavage reaction, the mutant is still contaminated, and like the wt protein, the K30C underwent SEC in hopes of eliminating the contaminants. The elution profile and SDS-PAGE analysis for the K30C mutant following SEC is shown in Figure 4.24.

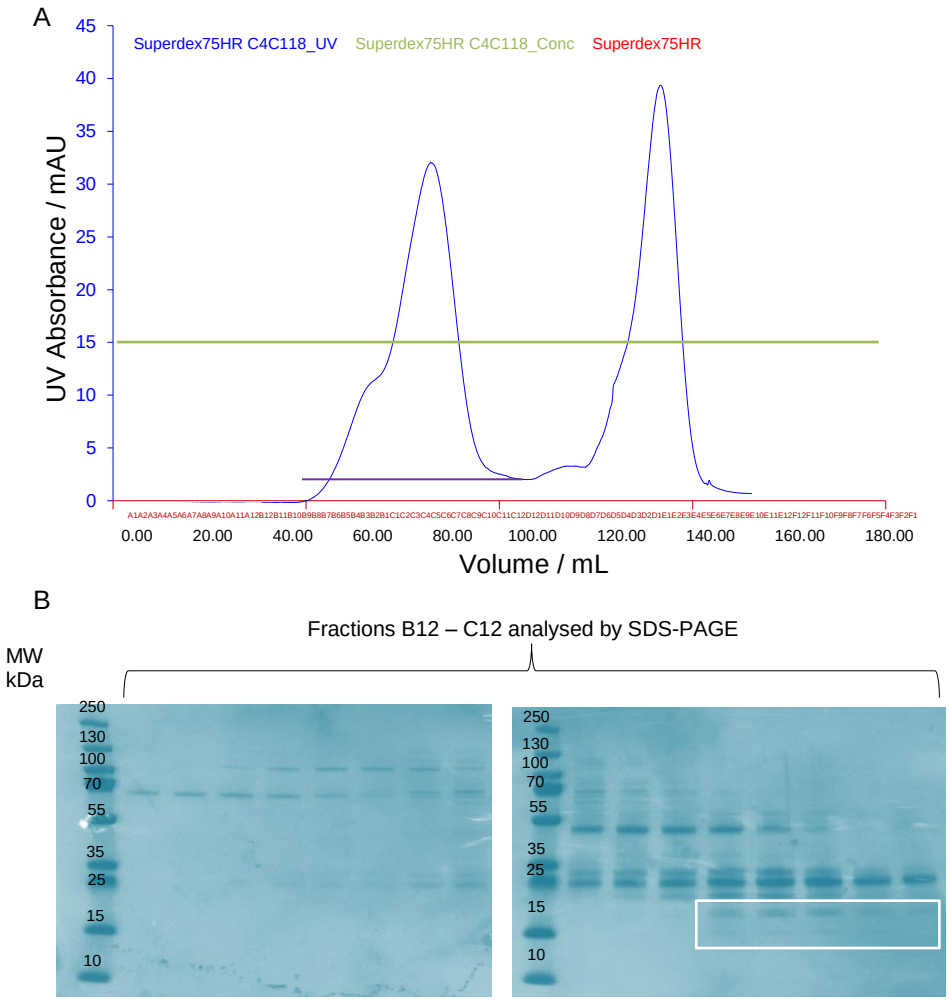



**Figure 4.24: Elution profile for the K30C RLC/MiniHMM mutant following SEC and resulting SDS-PAGE analysis.**

Looking at the elution profile for the K30C mutant, it elutes in the same place as the wt protein, with a slight shoulder to the left of the first peak. Following SDS-PAGE analysis, some of this shoulder is separated from the main peak, however not all contaminants are removed. The RLC/MiniHMM complex is shown in the circled fractions. Unlike the wt RLC/MiniHMM, SEC has not improved the purity of this mutant, and so at this stage, the E88C mutant was the only one selected for DEER

measurements, as this gave reasonable protein yield with which to work, as well as appearing clean following SDS-PAGE analysis (Figure 4.21). Optimisation of one mutant, with high purity, good spin labelling efficiencies and overall good yield would ensure that one could be confident that any changes observed during the DEER experiments were true.

## 4.10 Spin labelling of the E88C RLC/MiniHMM Mutant

Following cleavage of the hexa-histidine tag, the E88C mutant was treated with a final concentration of 10mM DTT, and free cysteine levels quantified using the Ellman's reagent DTNB. From SDS-PAGE analysis (Figure 4.21) it is clear that the sample is not entirely homogenous in regards to purity, and so Ellman's testing can only give a rough estimate as to the quantification of free cysteine in each sample.

DTT was eliminated by extensive dialysis, and mutants spin-labelled with MTSSL (10x molar excess, relative to free cysteine concentration). Excess spin label was removed using Vivaspin® centrifugal concentrators, before a CW EPR spectrum was recorded which showed some level of free spin label contaminant, shown in Figure 4.25.

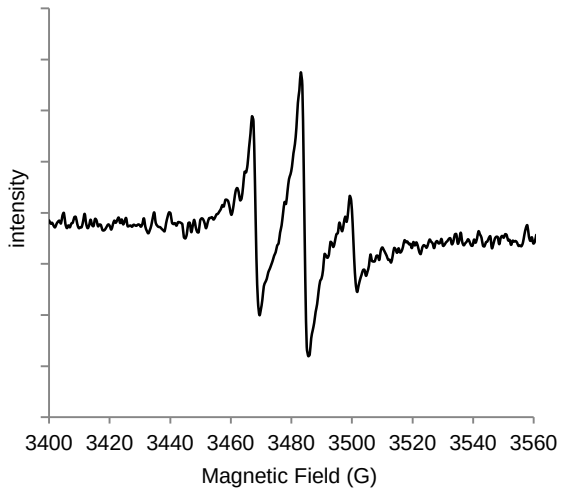



**Figure 4.25: CW EPR spectrum for the E88C mutant.** Taken at r.t under non-saturating conditions, using the AffirmoEx benchtop spectrometer, by ActiveSpectrum**.**

## 4.11 DEER Measurements of E88C RLC/MiniHMM, With and Without cMyBP-C domain C0.

Before DEER measurements were taken, the sample was dialysed into the deuterated form of the buffer used throughout (minus imidazole). When studying the binding interaction of C0 with the RLC using ITC (Ratti *et al*, 2011), an excess of C0 over MiniHMM of ~ 1.5-2 confirms an interaction (dissociation constant of 3.2 ± 1.7$\mu$m). Consequently, concentrated C0 (1mM) was added to the E88C mutant in a 6 times excess, to ensure an interaction, in a final volume of 50$\mu$L, before the addition of a further 50$\mu$L of deuterated glycerol.

The data we measured did not highlight a large change upon C0 addition but also was very non-specific with respect to any particularly distance. The modulation depth is low for nitroxide-nitroxide DEER, which is likely a result of low spin-labelling efficiencies of the E88C mutant. It was estimated that labelling efficiency was in the range of 25% by protein concentration, when compared to standard nitroxide solution (TEMPO). Further, long after the sample was disposed of we found a fault with the video amplifier which was leading to distorted data. This has cast more doubt on the results since we don't know whether slight changes are true or if we masked larger changes.

At this stage we decided that a new sample would have to be made, and measurements taken again.

At this time Mark Pfuhl also found that the MiniHMM complex can form dimers and that these could potentially block the interaction site for C0. An increase in NaCl destabilises this dimer, although may also abrogate C0 binding. Our plan therefore was to repeat E88C DEER measurement in high and low salt conditions.

## 4.12 DEER Measurements of E88C in High and Low NaCl

Following IMAC, the wash and elution fractions were pooled and concentrated using Vivaspin® centrifugal concentrators. The free cysteine levels were quantified using the Ellman's reagent, before the E88C mutant was spin-labelled. Figure 4.26A shows the final SDS-PAGE analysis of the new E88C sample following elution from IMAC. The sample was treated with a final concentration of 10mM DTT, and dialysis used to remove this. Based on absorbance at 280nm, the protein concentration was determined to be approximately 50 $\mu$M. The quantity of free cysteine was determined to be

approximately 73 $\mu$M by Ellman's testing. It should be noted at this point that the E88C mutant is not entirely pure (~60%), and so quantification of free cysteine also accounts for any free cysteine in contaminating proteins. The E88C was spin labelled and a CW EPR spectrum was taken, as shown in Figure 4.26B.

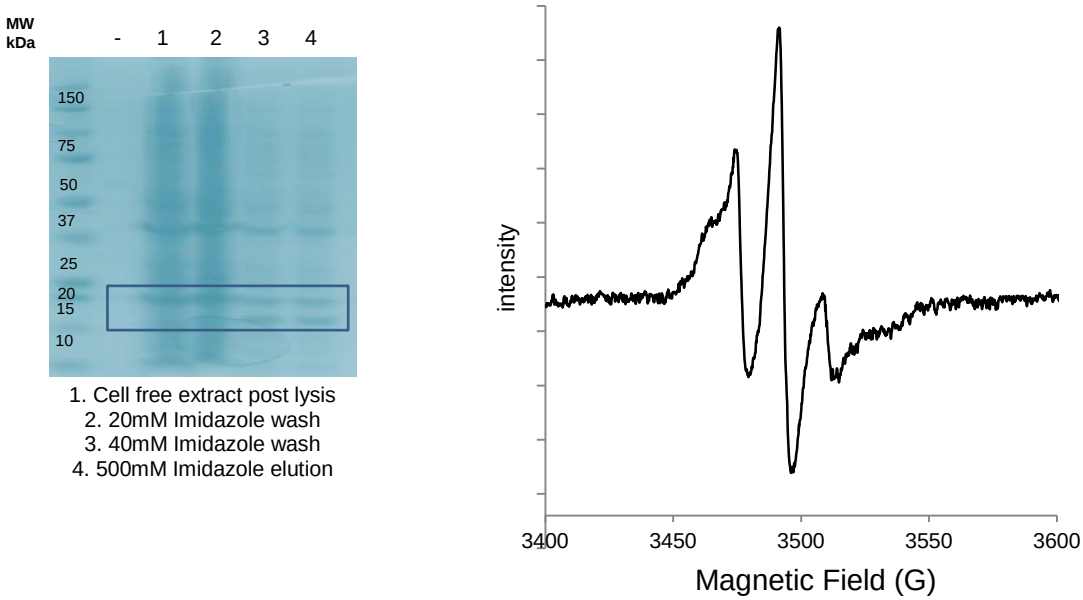


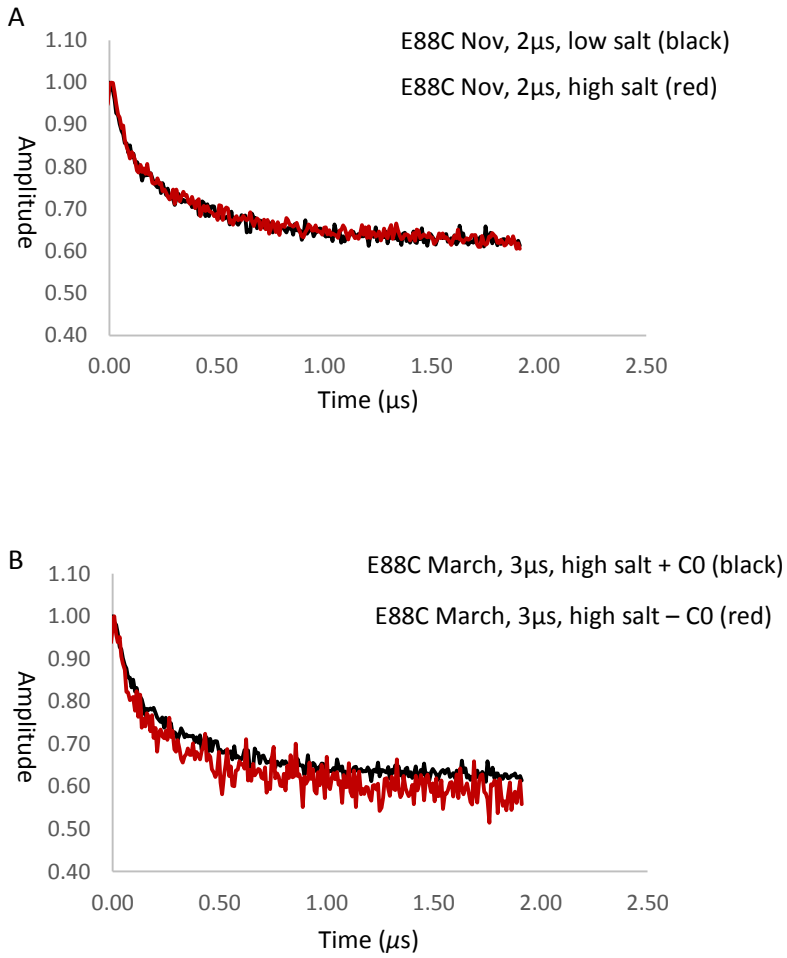

**Figure 4.26: (A) SDS-PAGE analysis and (B) CW EPR spectrum for the E88C mutant. of the elution of the new E88C sample following IMAC.** The elution from the IMAC column was performed manually using a syringe, and the wash and elution fractions pooled. The CW EPR spectrum was taken at r.t under non-saturating conditions, using the AffirmoEx benchtop spectrometer, by ActiveSpectrum.

Comparing the new labelled E88C mutant with standards of known concentration, it was estimated that labelling efficiency was improved from ~25% to somewhere in the region of 40-50%. Following purification of the RLC mutants, the proteins were stored in the same buffer used throughout the purification process, as detailed in Appendix A7 (minus imidazole). However, in retrospect, for future preparations, it may be worthwhile considering spin-labelling the RLC/MiniHMM constructs in the presence of high NaCl, as dimer formation of the MiniHMM could be affecting labelling efficiencies.

Samples were prepared washing E88C and E88C/C0 samples into either 50mM Hepes, 250mM NaCl or 100mM Hepes, 500mM NaCl with deuterated water. Finally

40% deuterated glycerol was added. Protein concentrations are 54µM (E88C) and 171µM for C0.

Following communication with collaborators, it was suggested that the MiniHMM complex has the potential to form dimers, which could explain why there is no indication of complex formation. This dimerization of the MiniHMM construct could be inhibiting the binding site for C0. Measurements were therefore taken in both low and high salt conditions (Figure 4.27).



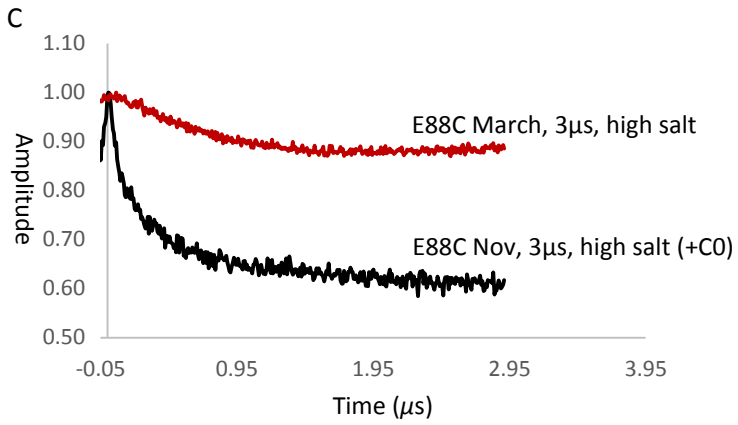



**Figure 4.27: Raw DEER data obtained from the RLC E88C mutant.** (A) Compares E88C mutant in both high and low salt (B) Compares E88C mutant with and without C0 (both measured in high salt) (C) Compares two different E88C mutants prepared under the same conditions.

When looking at the raw EPR data, it seems that there is no difference observed in both high and low salt for the E88C mutant. When looking at the mutant with and without C0, again no change is observed in the data collected.

This suggests that in the presence of C0, there is no change in the position of the RLC, however, when comparing the data obtained here, to that obtained in previous experiments for the E88C mutant, there is a difference. Although both protein preparations followed the same protocol, and were of similar concentrations and levels of purity, the EPR is not reproducible. This brings into question the state of the proteins under study

During cell lysis, cells were treated with Triton detergent to promote lysis and increase solubility, however it is possible that the use of detergents in the preparation are potentially affecting the fold of the protein, which could explain low spin-labelling efficiencies, even in the presence of a high concentration of free cysteine (as determined by the Ellman's Assay).

Furthermore, the inclusion of detergents is possibly solubilising endogenous E coli proteins, which are being brought through with the protein of interest, and approve difficult to eliminate with subsequent purification steps. Labelling should therefore be done in the presence of high and low salt, rather than transferring to high/low salt before the EPR experiment is run. If no signal is observed in this instance than a potential protein misfold as a result of that discussed above, should be considered.

## 4.13 Outlook and Future Work

Unfortunately, from the data obtained thus far for the RLC/MiniHMM/C0 interaction, no solid conclusions can be drawn. From the outset, this particular project appeared relatively straightforward, in that the constructs had been expressed before and the interaction between RLC and C0 confirmed.

E88C, E97C, D117C and T125C were chosen based on mutations previously made with the chicken RLC protein, with which there is a high degree of sequence homology, and corresponding mutants in chicken appear fully functional in force analysis when reconstituted into muscle fibres. This may not be true for the same mutants in the human protein. The K30C mutant was made following modelling from our collaborators which indicated more flexibility in the N' terminus of the human MiniHMM complex, with the RLC folded more towards the S2 junction, and therefore we hoped that the K30C mutant would help differentiate between the chicken complex and the human. Unfortunately it is difficult to predict how different mutants will behave before they are expressed, and so using models from other species is far from ideal, but otherwise would be done blindly.

Significant problems also arose from solubility issues, in that even the wt protein was not initially expressed in its soluble form. It was only after extensive optimisation and significant changes to the protocols, that expression of the wt protein could be visualised by SDS-PAGE. Frequently mutant expression levels are less than for wt but we saw a significant loss, and more problematic was our finding that the RLC/miniHMM complex could not be purified adequately for most mutations. In fact, while we repeated expression multiple times (not discussed here in detail) we only ever had adequate protein from the E88C mutant.

Our first DEER experiments with E88C did not show promising results but later we found that there was a fault which may have distorted our measurements and therefore these results needed to be abandoned. We did however find from these DEER experiments and the CW EPR scans that the labelling efficiency was quite low. We planned to repeat the experiment and additionally use both high and low salt conditions in case the MiniHMM dimerises at higher concentrations in low salt and this abrogates the C0 binding. However, different spectrometer problems have led to a delay in measuring these samples.

The C0 expressed beautifully and in high yield, and so there is a possibility of introducing a cysteine residue to C0. This would lead to triangulation which, if the distributions are somewhat broad, can cause inaccuracies in analysing DEER data for distances. However, if the label on the C0 had an EPR spectrum that does not fully overlap with the nitroxide, then selective DEER measurements should be possible. This is the basis of orthogonal spin labelling using *e.g.* $Gd^{3+}$ labels (Garbuio *et al*, 2013).

Thoughts on how one could explore the interaction further: Preliminary work done by our collaborators, has shown that upon phosphorylation of the RLC, the MiniHMM dissociates (Figure 4.28) from a heterotetramer composed of 2RLCs (red) and 2 heavy chains which form a coiled coil (blue), to 2 heterodimers composed of 1 heavy chain and 1 RLC. This is a result of dissociation of the S2 coiled coil.

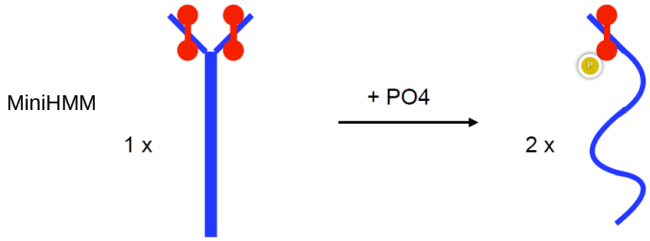



**Figure 4.28 Schematic representation of the effect phosphorylation of the RLC has on the RLC/MiniHMM complex.**

The MiniHMM construct used in this body of work, contains only 25% (Figure 4.29A) of the entire myosin S2 coiled coil (Figure 4.29B), and therefore dissociation of the entire coil *in vivo* is highly unlikely. It is proposed that upon phosphorylation of the RLC, the N' terminal portion of the S2 coil, in the immediate proximity to the RLC dissociates.

The spin-labelled RLC mutants (spin-labelling sites represented as yellow triangles) could then be phosphorylated, and EPR used to monitor the effect of phosphorylation of the RLC, on the myosin fragment (Figure 4.29C).

The MiniHMM fragment would be lengthened, so it remains a tetramer upon phosphorylation. Spin labels could also then be placed along the coil itself, to measure by EPR the dissociation that occurs as a result of phosphorylation of the RLC, and assess how far the coiled coil opens. The same can be done in the presence of C0, and longer cMyBP-C constructs (Figure 4.29D).

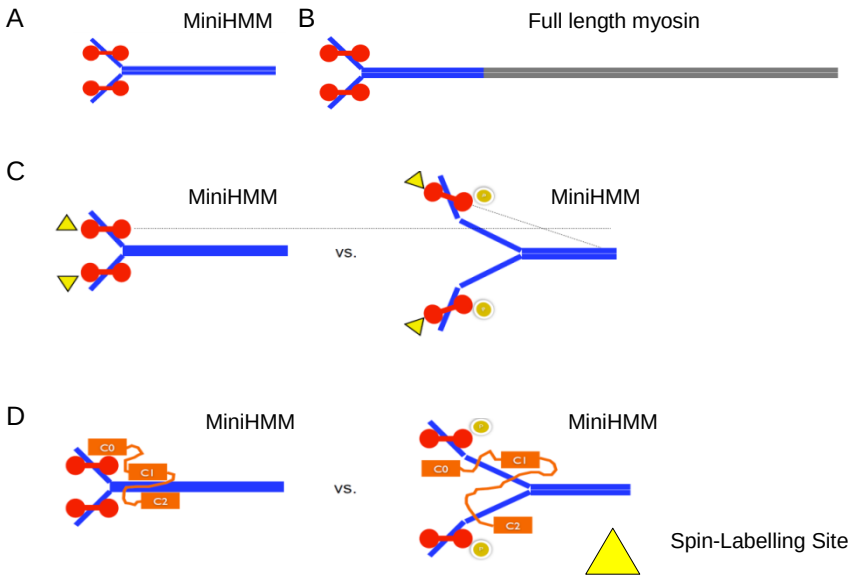



**Figure: 4.29 Schematic representations of the RLC/MiniHMM complex, showing how SDSL could be used to monitor dissociation of the S2 coiled coil upon phosphorylation of the RLC, both in the absence and presence of C0.**

Further elucidation of the Myosin/cMyBP-C interaction is essential in understanding the regulation of cardiac muscle contraction in health and disease. Although less vital in triggering contraction in cardiac muscle, phosphorylation of the RLC is vital for the control and power development and for precise mechanical operation of cardiac muscle.

Despite its value and significance in muscle contraction, as far as the human cardiac isoform is concerned there are few functional data, and no structural information regarding its function. If the problems with yield can be addressed, and better data obtained using EPR for more N' terminal mutants , this could provide for the first time a detailed insight into the Myosin/cMyBP-C interaction, and the effects of phosphorylation on heart muscle contraction.

These experiments could then be extended into mutations associated with disease, providing further understanding of the origin of cardiomyopathies, with the ultimate goal of advancement in medical treatments for conditions associated with mutations in these genes.

# Chapter 5: Development of New Strategies for the Site-Directed Spin–Labelling of Cysteine Rich Proteins

**AIMS:** This body of work aims to develop a range of new methods for the site-directed spin-labelling of cysteine rich proteins. The ability of *E. coli* to read through the amber stop codon will be exploited for the incorporation of unnatural amino acids for spin-labelling purposes, and novel spin labels, specific for labelling cysteine pairs will be tested in several model systems. Furthermore, native paramagnetic centres in recombinant proteins will be explored as potential labelling sites.

## 5.1 The Limits of SDSL by Cysteine Substitution Mutagenesis

Traditional methods of SDSL involving cysteine substitution mutagenesis are effective but what happens when this is not feasible? As discussed in detail in 1.4.2 the site specificity of EPR spectroscopy is typically introduced *via* the incorporation of spin labels into proteins of interest, *via* the thiol group of cysteine residues, due to its high reactivity at physiological pH. Cysteine residues are rarely present as free thiols in extracellular proteins, and are usually found in disulphide bonds, if at all. For this reason a great number of extracellular proteins have been engineered at specific sites of interest to incorporate cysteine residues for SDSL.

However, a huge caveat to using EPR to obtain distance measurements in biological systems is that the technique is at its most accurate in the simplest of cases; when looking at the interaction between a pair of radicals. When considering the application of EPR to the study of intracellular proteins, which are typically cysteine rich, the issue of specificity is raised. In order to specifically study the interaction between any two naturally occurring cysteine residues within a cysteine rich protein, any additional cysteine residues would need to be selectively removed, which could prove detrimental to both protein fold and function.

Structure underpins function, even at the simplest level, and so the field of structural biology is constantly evolving, with EPR spectroscopy expanding into *in vivo* studies, in order to elucidate how protein-protein interactions manifest in the natural environment of the cell. These principles and problems extend to all sorts of techniques that require specific labelling of isolated proteins or proteins within a living organism. The development of new labelling techniques for the specific labelling of cysteine rich

proteins, as well as the development of new spin labels which can withstand the reducing environment of the cell, is paramount for the study of proteins which are otherwise out with the realms of this technique, and for the study of protein-protein interactions *in vivo.*

## 5.2 Site Directed Site Labelling *via* Reassignment of the Amber Stop Codon

Over fifty years ago, the revolutionary work of Nirenberg and Matthaei,1961 initiated the cracking of the genetic code, by discovering that the amino acid (AA) phenylalanine was specifically incorporated into proteins in response to the codon UUU. This pioneering work paved the way in uncovering the rules that govern the translation of an organism's genetic lexicon for the ribosomal biosynthesis of proteins, with each of the 64 codons assigned to the 20 proteinogenic AAs, and the three termination signals.

With few exceptions, all living organisms are confined to the use of only twenty canonical AAs for protein biosynthesis. It is therefore somewhat remarkable that polypeptide synthesis from these 20 AAs is sufficient for performing all complex processes of life. It is clear that many proteins require more chemical complexity to function, as illustrated by the frequent use of post translational modification, and the dependence of many enzymes on cofactors (Young & Schultz, 2010). Therefore, the addition of new AAs to the genetic code, with novel chemical, biological and physical properties, could further expand the range of functions available to proteins, and provide powerful means for probing protein structure, function, and dynamics, both *in vitro* and *in vivo* (Loscha *et al*, 2012).

### 5.2.1 Translation of mRNA into a Polypeptide at the Ribosome

Translation is the process by which genetic information in the form of messenger RNA (mRNA) is converted into a sequence of corresponding AAs following ribosomal processing, forming a peptide chain, which is later folded into an active protein . This is illustrated in Figure 5.1.

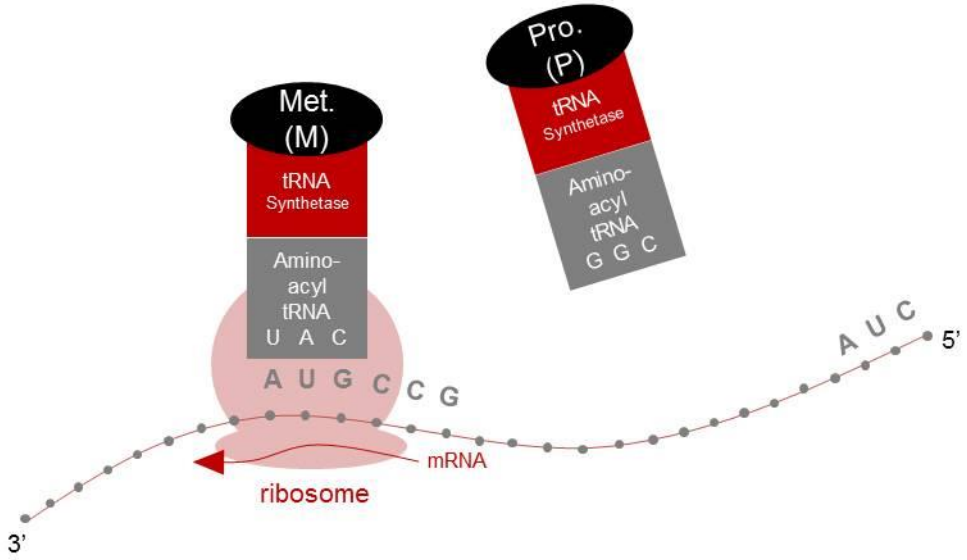



**Figure 5.1: Schematic representation of mRNA processing at the ribosome, and resulting polypeptide synthesis.**

The process of protein translation from mRNA occurs *via* four distinct steps.

**INITIATION:** In the first instance the ribosome and accessory proteins assemble on the target mRNA strand. The initiator tRNA binds *via* complementary base pairing, following recognition of the start codon AUG (Met) on the mRNA strand. **ELONGATION**: In response to the relevant anti-codon on the tRNA, specific AAs are delivered to the tRNA molecule by a tRNA synthetase, specific for that AA. **TRANSLOCATION**: The ribosome then moves in a 5' to 3' direction, to the next mRNA codon, continuing the process, and creating an extended AA chain. **TERMINATION**: Upon recognition of a stop codon on the polypeptide strand, the ribosome releases the protein.

Although the process of cellular translation at the ribosome functions to produce proteins composed entirely of the twenty canonical amino acids, it also represents the ultimate paradigm for the encoded synthesis of proteins containing additional amino acids, with novel function, with the potential to the evolve into the synthesis of completely unnatural polymers.

### 5.2.2 Incorporation of Unnatural Amino Acids (UAAs) into Recombinant Proteins

Although recent advances in synthetic and semi-synthetic methods have proven useful for the incorporation of UAAs into proteins (as discussed in 1.5.1), they are generally limited by low yields and are technically challenging in the production of proteins of large molecular weight. The use of cellular biosynthetic machinery to introduce novel AAs abrogates issues relating to scalability and protein size, whilst simplifying the study of modified proteins in living cells (Young & Schultz, 2010). This body of work looks at exploiting the ability for *E. coli* cells to read through the amber stop codon.

In order to expand the genetic repertoire beyond that of the 20 canonical AAs, a codon unique for the incorporation of a novel (or unnatural) AA must be assigned. The twenty AAs are encoded by 61 degenerate triplet codons, leaving the remaining three codons (TAG, amber; TAA, ochre; TGA, opal) to serve as translational termination signals. Using the redundancy of what is essentially a blank codon, together with the ribosomal machinery of the cell, many UAAs have been site-specifically incorporated into proteins in response to the amber stop codon in *E. coli* (Chin *et al*, 2002a), yeasts *Saccharomyces cerevisiae* (Hancock *et al*, 2010) and *Pichia pastoris* (Young *et al*, 2009), mammalian cell cultures (Gautier *et al*, 2010), and most recently a full animal model (Greiss & Chin, 2011).

## 5.3 Project Aims

Inspired by the pioneering work of Peter Schultz and Jason Chin, this thesis aims to exploit the ability of *E. coli* to read-through the amber stop codon, TAG, for the specific incorporation of UAAs with novel functional groups. Specifically; sperm whale myoglobin will be expressed and purified with the UAA propargyl lysine (pK) (Lysine modified with alkyne function group) incorporated at multiple sites.

Myoglobin is where the science of protein structure really began, with John Kendrew laying the foundation for an era of biological understanding in solving the first protein structure - sperm whale myoglobin – by X-Ray analysis (Kendrew *et al*, 1958). The XRC structure is shown in Figure 5.2.

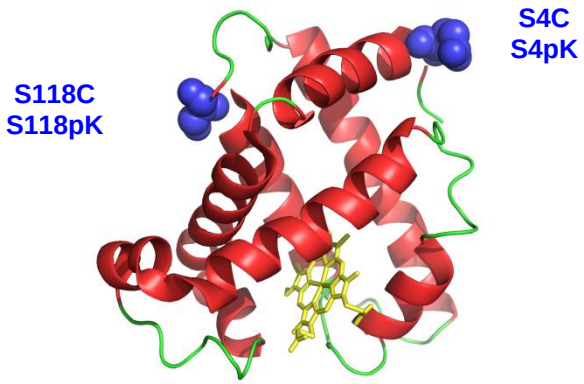



**Figure 5.2: X-Ray Crystal Structure of Sperm Whale Myoglobin**. The heme group of the myoglobin is represented as yellow sticks, and the blue spheres represent potential spin labelling sites, *via* both cysteine substitution mutagenesis, and pK incorporation. (Visualised using PyMOL & PDB_ID: 1MBN, H.C. Watson, 1969).

Consequently, myoglobin is very well characterised, with well-defined structure and function, and therefore a good model system for this work.

The UAA pK links an alkyne functional group to a lysine residue *via* a carbamate bond forming an aliphatic alkyne, as shown in Figure 5.3. The flexibility of the carbamate linkage facilitates more efficient incorporation of the pK into cells (Nguyen *et al*, 2009), and, unlike aromatic azides previously reported in the literature, is photo-stable, and consequently easier to handle. Copper-catalysed azide-alkyne cycloaddition (or 'click') reactions (Kolb *et al*, 2001) will then be performed to conjugate the pK containing protein to azide spin labels, in a manner that is comparable and orthogonal to cysteine labelling (Nguyen *et al*, 2009).

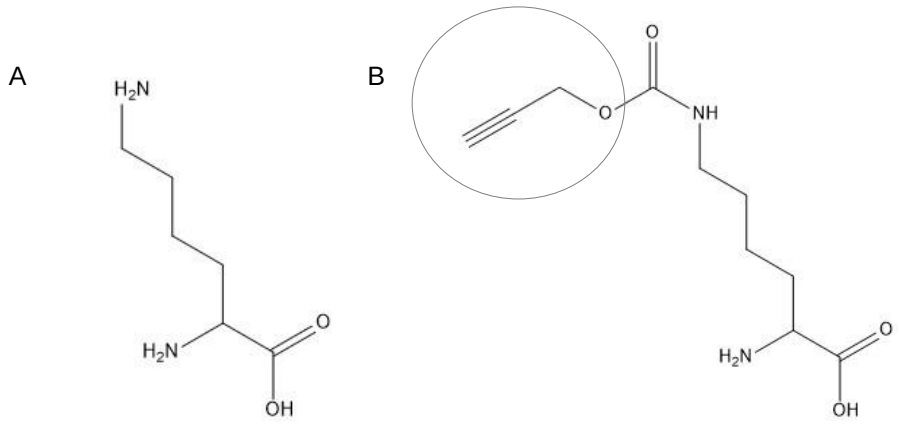



**Figure 5.3: Chemical Structures of (A) Lysine and (B) the UAA propargyl Lysine (pK), visualised using ChemDraw ®.** The structure of pK shows a Lysine residue modified by the addition of an alkyne functional group (circled).

EPR will be performed to assess whether this method of spin labelling is a surrogate for traditional methods, using cysteine substitution mutagenesis. The pK will be incorporated using the pyrrolysyl tRNA synthetase /$tRNA_{CUA(Pyl)}$ pair from *Methanosarcina barkeri*, which is orthogonal in *E. coli*, ensuring no cross talk between endogenous sets. A lack of cross-reactivity should ensure that pK, is incorporated with high fidelity, only in response to the non-sense codon TAG (Young *et al*, 2010).

Even in the wake of recent improvements in protein synthesis techniques, it is challenging to predict the facility in incorporating a specific UAA *in vitro*, at a given site, prior to performing the experiment (Short *et al*, 1999). To date, phenylalanine derivatives bearing alkynyl (Deiters & Schultz, 2005), azido (Chin *et al*, 2002b) and keto (Fleissner *et al*, 2009) groups that are bio orthogonal in their chemical reactivity have been incorporated into recombinant proteins in response to the amber stop codon.

Generally, the efficiency of incorporation of an UAA can be predicted to some extent, based on the chemical nature and structure of that being incorporated, with hydrophobic AAs being incorporated with high efficiency, and dipolar AAs proving troublesome (Short *et al*, 1999). However, the incorporation of aromatic UAAs into proteins where aliphatic AAs are naturally found may cause the protein to become misfolded, or result in an entire loss of protein function (Nguyen *et al*, 2009). Consequently, there is great need for the incorporation of aliphatic AAs, that contain bio-orthogonal chemical

handles, which can subsequently be used for the specific labelling of recombinant proteins.

UAAs are incorporated with low efficiencies in comparison to natural AAs, possibly due to competition between the tRNA and endogenous cellular release factors, which function to recognise stop codons and terminate translation (Loscha *et al*, 2012). A significant decrease in incorporation efficiency, as well as truncated proteins being expressed may influence the target protein, or prove deleterious to the host. Furthermore, yields drop precipitously with the addition of even a second stop codon (Johnson *et al*, 2011).

Consequently, the site of interest, the cell line required for expression, as well as the effect on incorporation efficiency as multiple stop codons are introduced, must all be considered in order to optimise expression of the target protein. Proving the latter can be overcome, and DEER results show that SDSL *via* genetic code expansion is a viable alternative to cysteine substitution mutagenesis, there is potential to extend these methodologies into cysteine rich proteins, which otherwise are out-with the realms of the technique.

## 5.4 Generation of Single (SC4 &S C118) and Double Cysteine (S4CS118C (or C4C118)) Myoglobin Mutants

The plasmids pBk-Pyls and pMyo4-TAG-PylT used throughout this work were kindly supplied by Prof. Jason Chin. The pBk-Pyls plasmid contains the pyrrolysyl tRNA synthetase / $tRNA_{CUA(Pyl)}$ pair whilst the pMyo4-TAG-PylT contains the gene for recombinant myoglobin expression.

### 5.4.1 Mutagenesis

Before myoglobin could be expressed with the UAA, pK, it was first expressed as the single cysteine mutants S4C and S118C and the double cysteine mutant S4CS118C, with the native serine residues at position 4 and 118 mutated to cysteine residues. Figure 5.4 shows the sequence chromatogram of the cysteine substitution mutagenesis, which can be compared to the wt. To generate the double mutant, the confirmed C4 mutant was used as template DNA.

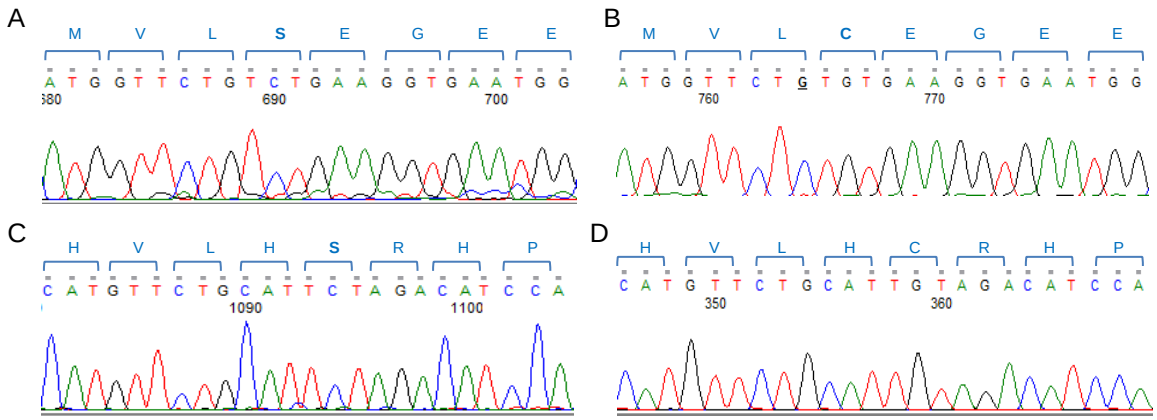



**Figure 5.4: Confirmation of mutagenesis of S4C and S118C in sperm whale myoglobin.**

### 5.4.2 Expression and Purification of the S4CS118C Myoglobin Mutant

Following transformation of the template DNA into DH10β *E. coli* cells, large scale flask expression was carried out, and the proteins purified by IMAC and SEC as described in 2.3.4 & 2.5.3. For each of the myoglobin mutants (Cys and pK mutants), due to the presence of a C' terminal hexa-histidine tag, IMAC can be used for the initial catchment step. The SDS-PAGE analysis from all stages of purification, as well as the elution profile following the final purification step are shown in Figure 5.5.

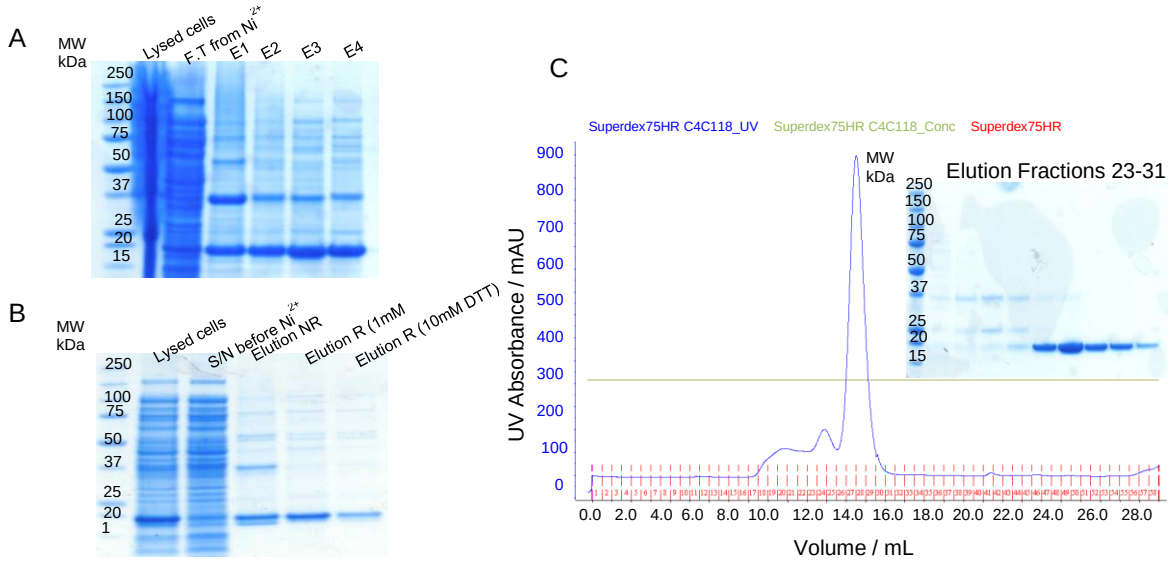



**Figure 5.5: Expression and purification of the Sperm Whale Myoglobin, S4CS118C mutant.** (A) SDS-PAGE analysis of the first catchment step; manual elution following IMAC. (B) SDS-PAGE analysis from small scale expression (10mL), testing reducing conditions. (C) Elution from the S75 SEC column, and resulting SDS-PAGE analysis.

Following the initial catchment step, using IMAC, from the S4CS118C elution fractions, it is clear (Figure 5.5A) that although the majority of the myoglobin S4CS118C mutant is present as a monomeric species, there is a significant amount of dimer present. Typically, myoglobin does not contain any free thiols, and so it can be assumed that the dimerization is a result of disulphide bond formation between the introduced cysteine residues.

Therefore, the C4C118 mutant was treated with 1mM and 10mM DTT in order to reduce the disulphide bonds, and free the newly incorporated cysteine residues for spin labelling purposes. From the SDS-PAGE analysis in (Figure 5.5B), treatment with 1mM DTT is enough to reduce the disulphide bonds, although a little dimer remains, therefore, following IMAC, the pooled elution fractions were treated with a final concentration of 10mM DTT, before undergoing SEC.

### 5.4.3 Expression and Purification of the Single Cysteine Mutants S4C and S118C

Following transformation of the template DNA into DH10β *E. coli* cells, large scale flask expression was carried out, and the proteins purified by IMAC as described in 2.3.4 & 2.5.3. Figures 5.6 and 5.7 show the SDS-PAGE analysis for all stages of purification, as well as the elution profile following the final purification step, for both the S4C and S118C mutants, respectively.

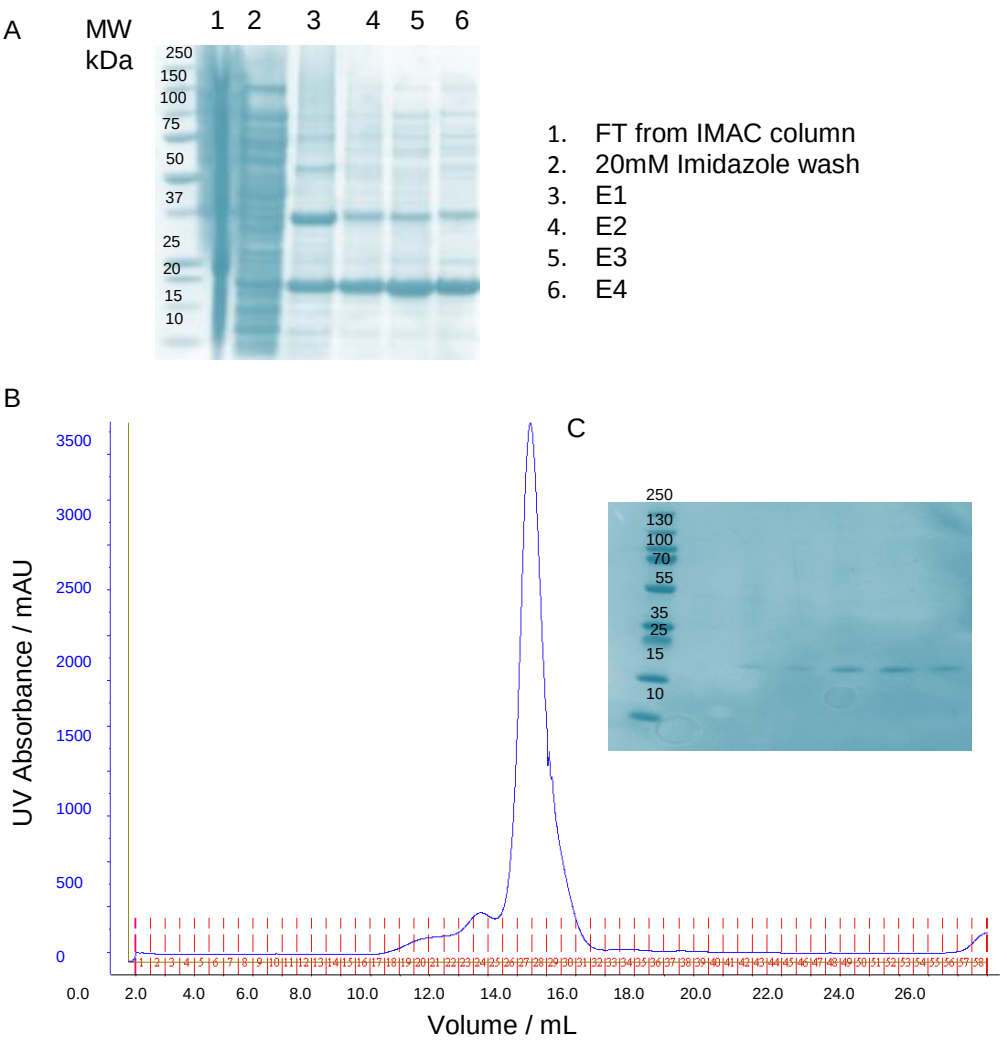



**Figure 5.6: SDS-PAGE analysis following purification of the S4C Mutant following IMAC (A), shown alongside the elution profile (B) and resulting SDS-PAGE analysis following SEC (post-DTT treatment) (C).**

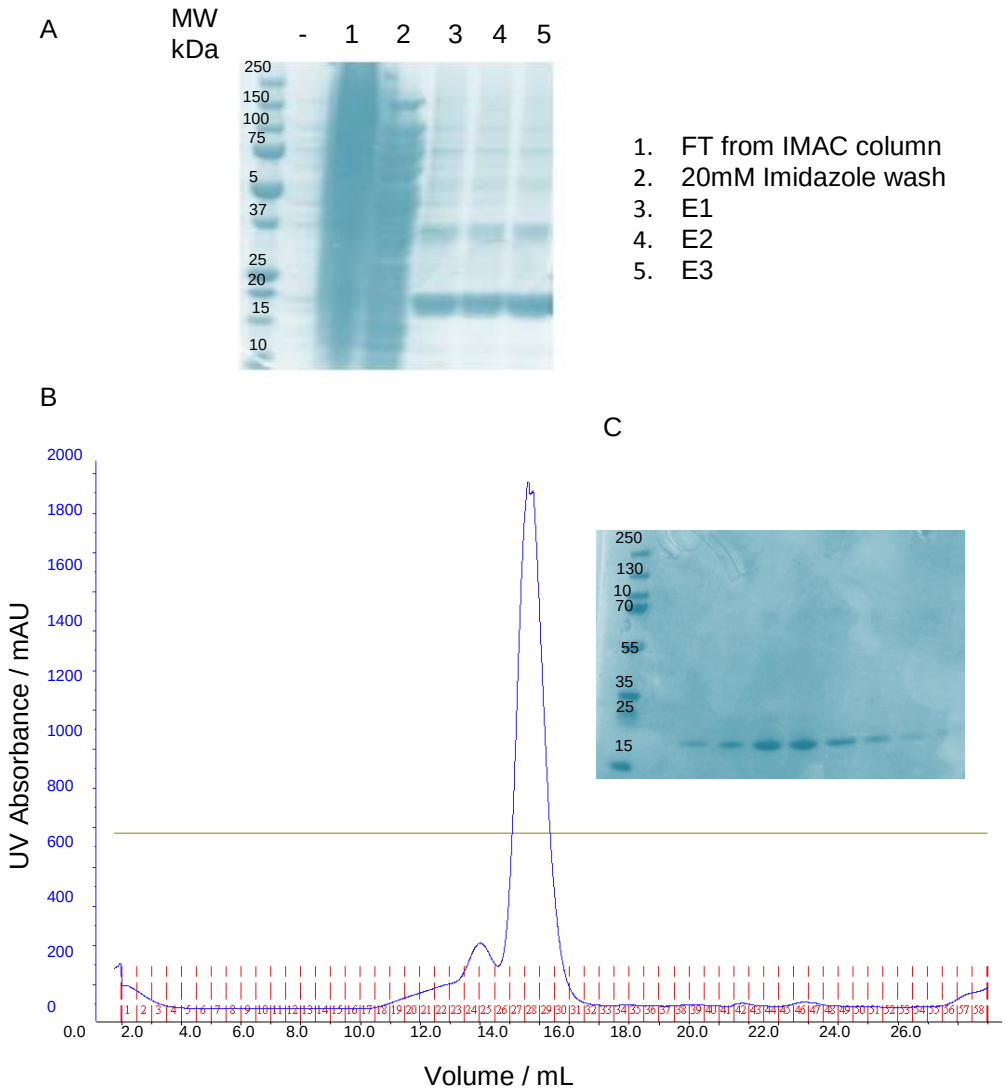



**Figure 5.7: SDS-PAGE analysis following purification of the S118C Mutant following IMAC (A), shown alongside the elution profile (B) and resulting SDS-PAGE analysis following SEC (post-DTT treatment) (C).**

### 5.4.4 Spin Labelling of Cysteine Myoglobin Mutants

Ideally, following treatment with DTT, an Ellman's assay would be performed as described in 2.8.2 in order to quantify the amount of free cysteine before spin labelling, however, due to the characteristic red colour of myoglobin, it is impossible to monitor the release of the NTB component at 412nm, and so it must be assumed that any cysteine/cystine is reduced by the addition of 10mM DTT.

Immediately following SEC, relevant fractions were pooled and treated with a 10x molar excess of MTSSL. The excess MTSSL was removed using extensive dialysis of the samples into PBS. The CW EPR spectrum of the double mutant, S4CS118C, is shown in Figure 5.8. The degree of spin labelling is close to 100%.

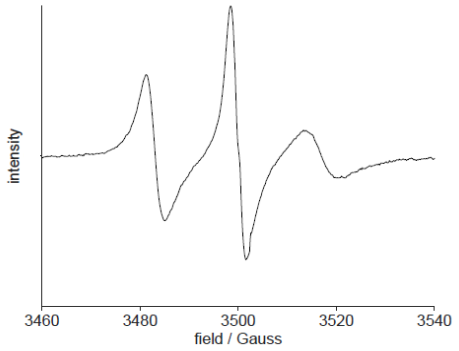



**Figure 5.8: X-band CW EPR of myoglobin S4CS118C SL.** Taken on the AffirmoEx at r.t. in aqueous buffer.

## 5.5 Expression and Purification of Sperm Whale Myoglobin Mutants with the UAA, pK.

### 5.5.1 Mutagenesis

For incorporation of the pK UAA into myoglobin, the serine residues at position 4 and 118 were changed to the amber stop codon, TAG. Figure 5.9 shows the sequence chromatogram of the mutagenesis, compared to wt.

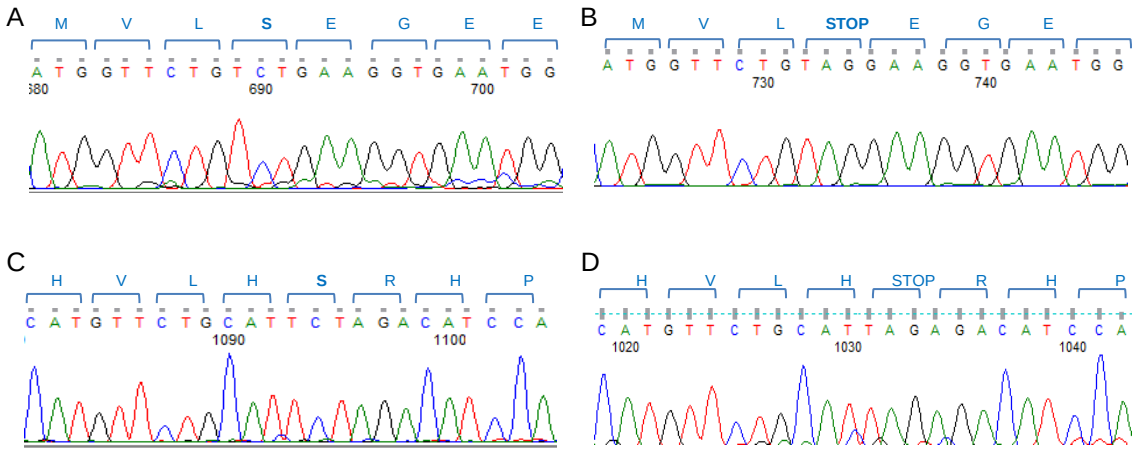



**Figure 5.9: Confirmation of mutagenesis of S4pK and S118pK in sperm whale myoglobin.**

### 5.5.2 Small-Scale Expression of Myoglobin pK Single Mutants

Following transformation of the template DNA (S4pK and S118pK) into DH10β *E. coli* cells, alongside the plasmid containing the pyrrolysyl tRNA synthetase /$tRNA_{CUA(Pyl)}$ pair, the single pK mutants were expressed on a 10mL scale, with the media supplemented with a final concentration of 2mM pK.

The pK used for the small scale test expressions was provided as the TFA salt, by Dr Alison Hulme, University of Edinburgh, Scotland, UK. It is suggested that supplementing the media with the UAA from inoculation yields higher expression levels, and reduces leaky expression, however, supplementing the media with the UAA pK at different points in the growth stage, (start, $OD_{600}$ 0.2, $OD_{600}$ 0.6, $OD_{600}$ 1.0) shows that when the UAA acid is added to the growth media has no effect on final yield.

Supplementing the growth media with the UAA significantly reduces the growth rate of the *E. coli* cells, when compared to expression of wt myoglobin, or cysteine mutants. Adding pK later in the expression proves more economical, as if cells do not reach their exponential growth phase, the pK, which must be synthesised, is not added unnecessarily. Figure 5.10 shows the SDS-PAGE analysis of the small scale expression tests of both the S4pK and S118pK single mutants.

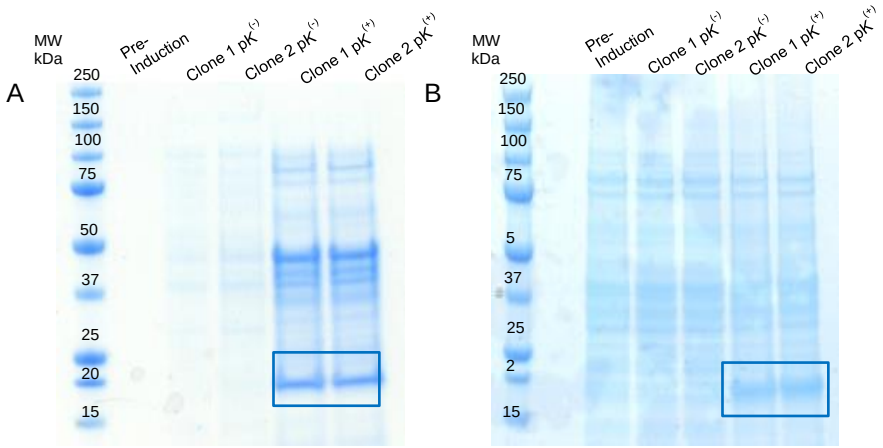



**Figure 5.10: SDS-PAGE Analysis of Small Scale Test Expressions OF the single pK mutants (A) S4pK and (B) S118pK.** For each mutant two different clones were tested for myoglobin expression. All clones were tested with and without the UAA pK in the growth media, as well as a sample taken before induction.

In both Figures 5.10A and B, before induction with L-arabinose (final concentration 0.2%) there is no indication of protein expression in cultures containing pK. In those

clones which were grown in the absence of pK, there is no protein expression post-induction. When looking at the clones expressed in the presence of pK, for both single mutants, as highlighted by the circled bands, there is a definite over-expression of protein, which appears at the expected MW of myoglobin. At this stage, both mutants must be expressed on a larger scale, and the over-expressed protein further characterised, to determine whether it is myoglobin containing the UAA, pK.

### 5.5.3 Large Scale Expression and Purification of Myoglobin Single pK Mutants

Small scale expression of both single pK myoglobin mutants suggest that pK is being incorporated in response to the amber stop codon and so expression must be scaled up, so the proteins can be purified and further characterised. Figure 5.11 shows the SDS-PAGE analysis from large scale expression (1 litre) of each of the single pK mutants, as well as Western Blotting Analysis, of samples probed with an antibody that specifically detects a C' terminal histidine tag.

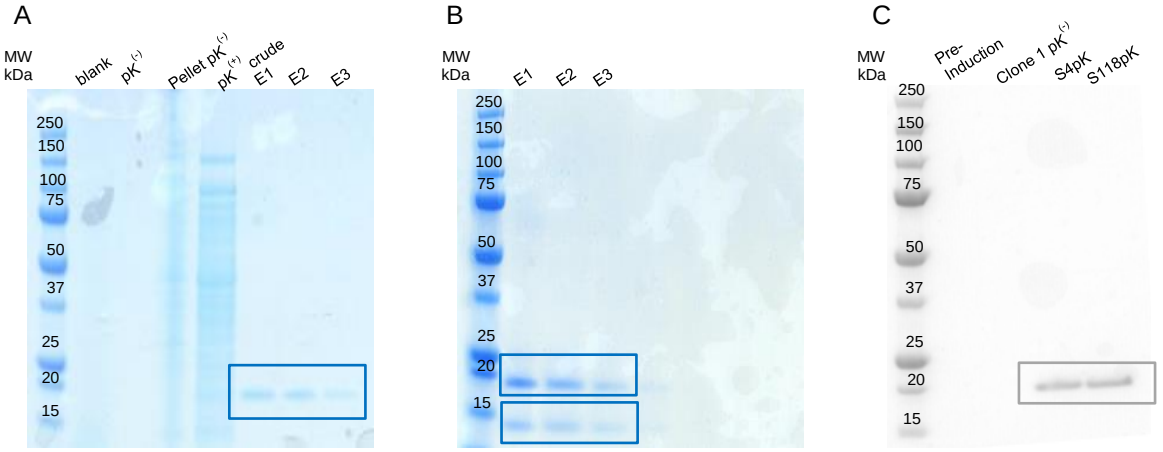



**Figure 5.11: SDS-PAGE analysis from large scale expression of the single mutants.** (A) S4pK and (B) S118pK, and corresponding (C) western blotting analysis, of samples probes with an antibody specific for the detection of a C' terminal hexa-histidine tag.

From the SDS-PAGE analysis of the S4pK mutant (A), the circled area represents elution fractions containing protein corresponding to the MW of myoglobin. Similarly in the SDS-PAGE analysis of the S118pK mutant, the elution fractions thought to contain myoglobin are circled. Following western blotting analysis of both mutants with an antibody conjugated to HRP, and specific for the detection of C' terminal histidine tags, the bands circled in (C) indicate that a C' terminal histidine tag has been detected. The bands on the western blot suggest that the *E. coli* cells have read through the stop

codon for both mutants, as the histidine tag follows the myoglobin gene. However, looking at the SDS-PAGE analysis from the S118pK mutant, there is indication of a lower molecular weight band. It can be assumed that the circled bands appearing at ~18kDa in (A) and (B) are pK containing myoglobin. The lower molecular weight band in (B) was excised from the SDS-PAGE gel, and sent for tryptic digest mass spectrometry (MS). MS confirmed that the circled lower molecular weight band in (B) had 52% identity to myoglobin, with no peptides identified beyond pK. This suggests at the 118 position, there is early termination of protein synthesis due to latent reading of the amber stop codon.

### 5.5.4 Expression of the double mutant S4pKS118pK

Small scale expression tests were carried out for several clones of the myoglobin containing pK at both positions 4 and 118, however, following SDS-PAGE analysis there appeared to be no expression of the recombinant protein. This is not particularly surprising due to the reduced yields of single mutants when compared to wt, coupled with the drop off in expression that results from an increase in stop codons.

Following successful transformation of the template DNA, as well as the plasmids encoding the pyrrolysyl tRNA synthetase /$tRNA_{CUA(Pyl)}$ pair, 2L of LB was supplemented with a final concentration of 3mM pK. The double mutant was purified using IMAC, utilising the histidine tag at the C' terminal of the myoglobin gene in the first catchment step, and fractions analysed by SDS-PAGE. Figure 5.12 shows the SDS-PAGE analysis for the double pK mutant. For the cysteine mutants, at this point SEC would be performed to further purify the myoglobin, however, due to the small sample volumes, this was avoided. From the SDS-PAGE analysis, purity was estimated to be approximately ~60%, and final yields calculated based on absorbance of the soret band (409nm), rather than at 280nm.

For large-scale expression of the double mutant, the sodium salt of pK was used, rather than the TFA salt to aid solubility of the UAA in water. (Synthesised by Dr. Bouchra Hajjaj, University of St Andrews, Scotland, UK). From the SDS-PAGE analysis, the fractions in the circled region were pooled, ready for spin labelling via 'click' chemistry of the azide spin label to the pK containing myoglobin.

Elution fractions 1-14 (1.5mL)

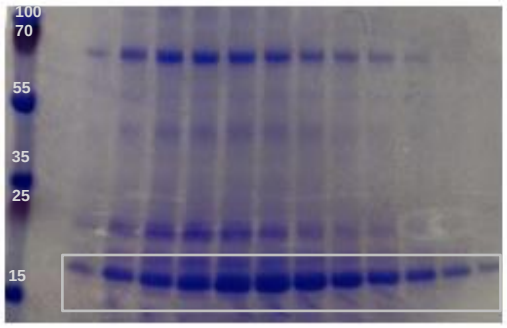



**Figure 5.12: SDS-PAGE analysis from large scale expression of double pK mutant S4pKS118pK.**

**Table 5.1 Protein Yields for Myoglobin Mutants Per Litre of Cell Culture**

| Myoglobin Mutant | Yield/Litre of Culture) |
|---|---|
| wt | 18mg |
| S4C118 | 9mg |
| C4S118 | 12mg |
| C4C118 | 8mg |
| S4pK | 1.6mg |
| S118pK | 0.871mg |
| pK4pK118 | 1.2 & 1.8mg |

## 5.6 Coupling of a Peptide Containing pK to an Azide Spin Label Using 'Click' Chemistry

The reported work described in 5.1.3 and 5.1.4 was undertaken by a PDRA in our lab, Dr Bouchra Hajjaj.

Before the pK containing protein was conjugated to an azide spin label using 'click' chemistry, the reaction was tested with a small peptide containing pK, flanked either side by alanine residues.

The synthesized pK peptide, **1** (see Figure 5.13) (6.3mg, 5mM final concentration) and the spin label **2**[1] (3.0mg, 5mM final concentration) were solubilized in freshly degassed water (3.1mL). A r.t CW EPR spectrum and HPLC analysis were taken of this solution to ensure the radical was stable before the addition of the ascorbic acid, which can reduce the nitroxide. The reaction mechanism is shown in Figure 5.13.

[1] Azide spin label was synthesized by Marius Haugland - University of Oxford

A 10mM solution of Copper (II) sulphate, TBTA in 55% (v) DMSO (61.4$\mu$M; 0.2mM) was then added followed by ascorbic acid (0.1mg; final concentration 0.2mM). The reaction mixture was stirred at r.t under Argon, before an HPLC of the crude mixture performed, after 30min and 1h30min. The HPLC traces are shown in Figure 5.14.

HPLC analyses show that the coupling reaction is complete after 30min at r.t. A r.t CW EPR was recorded for the uncoupled reagents and the coupled product after 2h reaction time at r.t, and formation of the desired product, **3**.

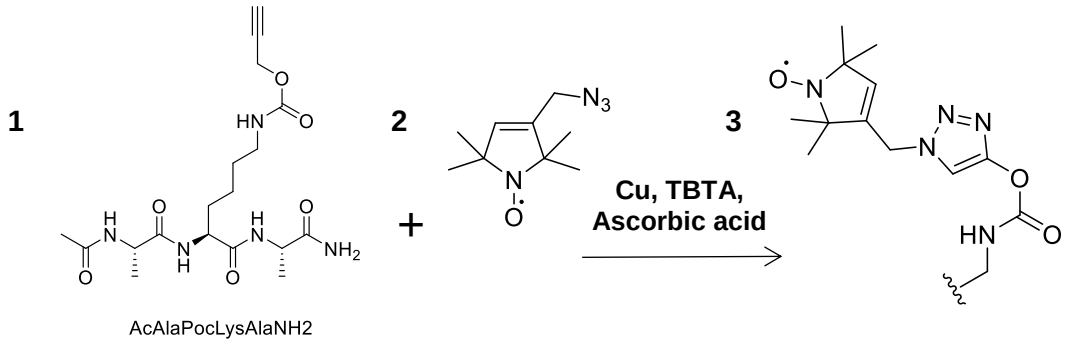



**Figure 5.13: Reaction Mechanism for the copper catalyzed azide-alkyne cycloaddition of the pK containing peptide to spin label.**

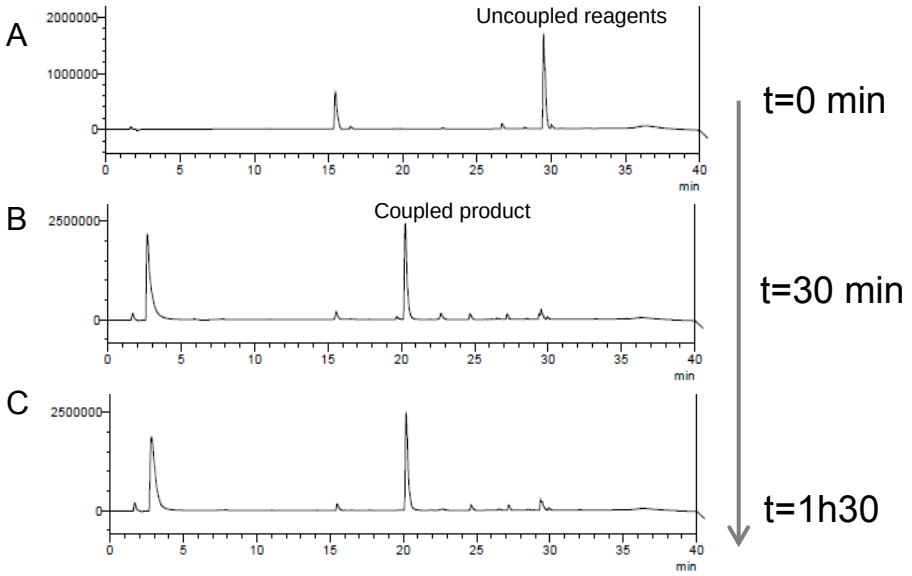



**Figure 5.14: HPLC (Shimadzu) of the reaction mixture at t=0, 30, 90min using Phenomenex $C_{18}$ column 150*4.60mM.** (A) the pK peptide and azide spin label before coupling (B) after coupling (30min) and (C) after coupling (1h30min). A shift in retention time was recorded from 30min for the uncoupled reagents to 20 min for the coupled product.

## 5.7 Coupling of pK containing myoglobin to an azide spin label

### 5.7.1 From Peptide to Protein - Optimisation of Coupling

Although coupling of the peptide to the azide spin label proves successful, the conditions must be optimised, as the coupling reaction must be performed in aqueous buffers, in conditions which would not result in denaturation of the protein. The initial coupling reaction was performed using 5mM peptide and 5mM label, as highlighted in Test 1 (Table 5.2).

Obtaining final yields in the milli-molar range for recombinant proteins expressed with unnatural amino acids is unheard of and so the coupling reaction must be adjusted accordingly. Shown in Table 5.2 are a variety of coupling conditions tested for the coupling of the peptide to the azide spin label, which subsequently could be transferred to coupling of the protein, which will likely be in the micro-molar concentration range.

**Table 5.2 Test Coupling Conditions of the Peptide to the Spin Label**

| Test | Peptide | Label | Cu TBTA | Ascorbic acid | Water | 50mM Hepes, 125NaCl, pH7.0 | time |
|---|---|---|---|---|---|---|---|
| 1 | 5mM | 5mM | 0.2mM | 0.2mM | - | Yes | 1h |
| 2 | 34$\mu$M | 34$\mu$M | 1.4$\mu$M | 1.4$\mu$M | Yes | Yes | 1h & overnight |
| 3 | 34$\mu$M | 34$\mu$M | 0.2mM | 0.2mM | Yes | Yes | 1h |
| 4 | 34$\mu$M | 68$\mu$M | 11$\mu$M | 11$\mu$M | Yes | Yes | 1h & overnight |
| 5 | 34$\mu$M | 5mM | 0.2mM | 0.2mM | yes | Yes | 1h |

Of all of the conditions tested, it was found that only test 5, using a final concentration of 34$\mu$M peptide, 5mM label, 0.2mM CuTBTA and 0.2mM ascorbic acid resulted in a successful coupling reaction, whilst maintaining the nitroxide radical.

These conditions were then used for the coupling of the proteins to the azide spin label.

### 5.7.2 Spin Labelling of S4pK Using Copper Catalysed 'Click' chemistry

The buffer used throughout the purification process of the myoglobin pK mutants was 50mM Hepes, 125mM NaCl, pH7.0 (see appendix A9), and so the conditions which were optimized for the peptide, at a final concentration of 34$\mu$M, were transferred onto the S4pK myoglobin at the same concentration. Shown in Figure 5.15 is the r.t CW EPR spectrum taken, of the single pK mutant S4pK.

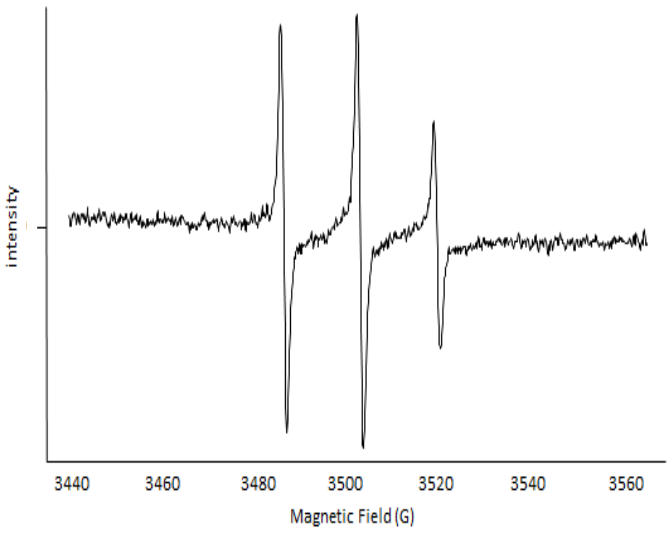



**Figure 5.15: X-band CW EPR of the S4pK myoglobin Mutant.** Spectrum taken under non-saturating conditions.

The CW spectrum was taken in an aqueous buffer/ glycerol mixture which should increase correlation time, giving a broader EPR spectrum, however, when looking at the EPR line-shape it is very narrow, and indicative of a mobile, but not free, spin label.

Although the label is mobile, the CW EPR spectrum indicates that the pK containing myoglobin is spin labelled. The UV-Vis absorption spectrum was recorded for the pK myoglobin following the labelling reaction, confirming the protein to be oxy-myoglobin. The double pK myoglobin mutant was then spin labelled in the same manner as the single mutant.

### 5.7.3 Spin Labelling of S4pKS118pK Using Copper Catalysed 'Click' chemistry

The purified S4pKS118pK was labelled under the same conditions as for the singly labelled pK myoglobin. There was no need to adjust the concentrations of label or the

other reagents, as these were already present in excess, when compared to protein concentration. Figure 5.16 shows the r.t CW EPR spectrum of the doubly labelled pK mutant.

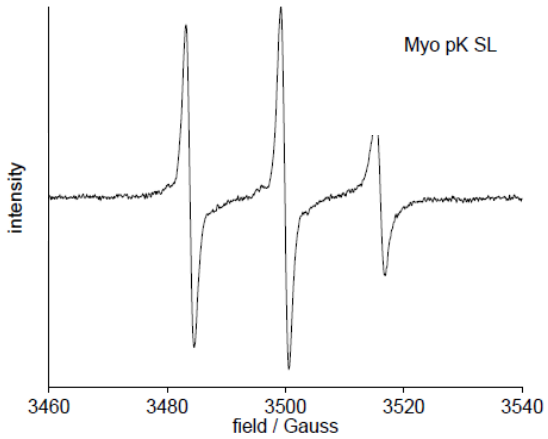



**Figure 5.16: X-band CW EPR of the S4pKS118pK myoglobin Mutant.** Spectrum taken under non-saturating conditions.

Although CW EPR will give some information as to whether or not the protein is labelled, mass spectrometry was used to confirm if one or both of the pK amino acids were labelled, before the DEER experiments could be performed.

## 5.8 Confirmation of Spin Labelling of the S4pKS118pK mutant using Mass Spectrometry

A sample of the double pK myoglobin mutant was taken before labelling and after, and mass spectrometry (LCT LC-MS) carried out to estimate the accurate mass of both the labelled and unlabelled samples. Figure 5.17 shows the LCT LC-MS analyses of both samples, processed over a narrow mass range to 0.1 Da (Performed at the University of St Andrews). For the unlabelled the expected mass is 18601Da, measured 18602.4Da and for the labelled the expected mass is 18991, measured 18992.5Da. This high degree of correlation seemed to confirm successful incorporation of two pK UAAs and their labelling.

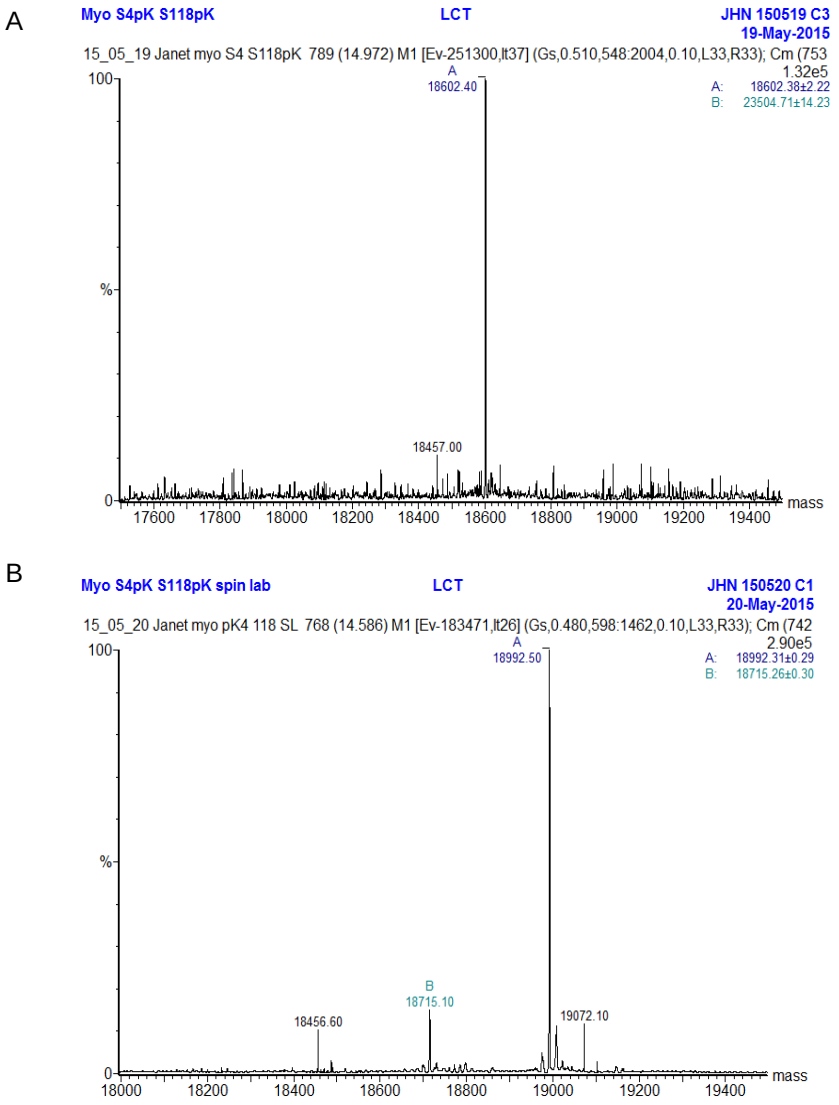



**Figure 5.17: LCT LC Mass Spectrometry analysis of the double pK myoglobin mutant S4pKS118pK.** (A) analysis of the unlabeled sampled and (B), analysis of the doubly spin labelled sample.

## 5.9 DEER Experiments

DEER was performed at Q-band for both the doubly labelled cysteine myoglobin mutant and the doubly labelled pK mutant. The results of the DEER and the MMM modelling for the MTS label are shown in Figure 5.18 which also shows the final SDS-PAGE analysis of the double pK labelled mutant.

Looking at the distance distributions obtained for the S4CS118C mutant (MTSSL labelled), these overlap well with the distributions simulated using MMM. However, the

S4pKS118pK result is much broader and shows a shorter distance. This is not entirely unexpected given the much longer amino acid side chain the label is attached to and the much larger degree of flexibility the label has at room temperature according to the CW EPR. This conformational freedom is likely to have been frozen in and results in a distribution of distances. More of a concern is the smaller modulation depth of the DEER time trace which brings the labelling efficiency of the click reaction into question since modulation depth is related to the number of pairs of interacting labels.

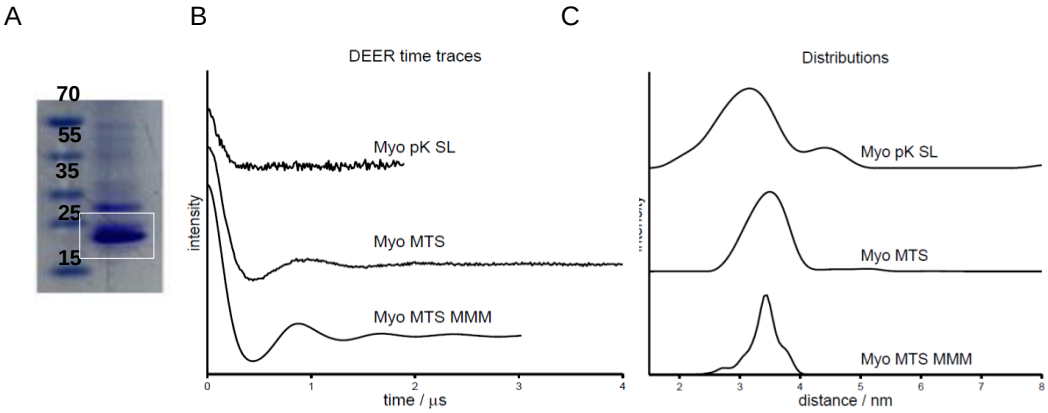



**Figure 5.18: Comparisons of the Background Corrected EPR Time-Traces and Distance Distributions, of the S4CS118C and S4pKS118pK doubly spin-labelled mutants, shown alongside MMM simulation.** (A) sample of the doubly labelled S4pKS118pK mutant analysed by SDS-PAGE following coupling to the azide spin label. (B) DEER time traces for the doubly labelled cysteine and pK mutants, overlaid with the MMM simulation. (C) DEER distance distributions for the doubly labelled cysteine and pK mutants, overlaid with the MMM simulation.

Although Mass Spectrometry analyses of both the labelled and unlabelled samples suggest thorough labelling of the sample, it is implied that although the labels themselves have been successfully conjugated to the pK UAAs within the myoglobin, it is possible that the nitroxide itself is not intact. It may have been reduced to the hydroxylamine during the click reaction and this would only make a difference of 2Da. We compared the CW EPR spectrum to standards and estimate that only 60% of the radical is intact.

At this stage, further large scale expression was carried out for the double pK mutant, and the experiment repeated, only a desalting column was used for the removal of excess ascorbic acid/copper with the hope that this more efficient removal would reduce the chances of reduction. Expression this time of the double pK mutant was high,

resulting in a final yield of ~1.8mg from 1L of culture. (Yield comparable to that suggested by Nguyen *et al*, 2009 for a single UAA mutant).

The click reaction was then performed on the new S4pKS118pK sample, however CW EPR and MS suggest this was unsuccessful, and so was repeated. MS again indicates that this has not been successful, whilst the CW EPR implies otherwise, and so the sample was prepared for DEER, as there was not enough sample to repeat the click reaction.

The final concentration of the sample was determined to be ~100$\mu$M, in 50mM Hepes, 125mM NaCl pH7, in 50% deuterated glycerol. Following the DEER experiment, unlike the original sample, there was no signal observed. This contradicts that observed from the CW EPR, and so it is possible that the label is unspecifically binding to the protein, and is not conjugated via the 'click' mechanism.

## 5.10 Outlook and Future Work

Contrary to what is stated in the literature (Loscha *et al*, 2012) (Johnson *et al*, 2011), we have successfully managed to incorporate two UAAs, namely the UAA, pK, into recombinant sperm whale myoglobin in response to the amber stop codon, giving final yields in the hundred micro-molar range, without the need for modified cell lines.

A doubly spin-labelled mutant has given EPR distance distributions comparable not only to a doubly spin-labelled cysteine mutant, but also to simulated distance distributions obtained from the published X-Ray Crystal structure. This demonstrates the potential for SDSL using genetic code expansion as an alternative to cysteine substitution mutagenesis for the purpose of DEER spectroscopy.

Improving final yields, incorporation efficiencies, conformational freedom of incorporated spin labels and the labels themselves would prove advantageous in improving data quality. The incorporation of the UAA, pK, into recombinant sperm whale myoglobin has proven to be a favourable model system, and proof of principle, that EPR can be performed between two spin-labelled UAAs. With slight adjustments to the protocols described above, the system can be optimized, to facilitate transfer from a model system, to a protein system where EPR would be used for structural characterization.

Assignment of the amber stop codon to the UAA, pK, is ambiguous, as it simultaneously codes for the incorporation of the UAA, as well as the termination of protein synthesis. This is due to a direct competition between the exogenous $tRNA_{CUA}$ and the endogenous release factors.

Eukaryotic cells typically express two release factors, RF1 and RF2, with the RF1 terminating protein synthesis in response to the codons UAA and UAG, and RF2 terminating at UAA and UGA. Johnson *et al*, 2011have developed the *E. coli* cell-line, JX33, with the gene encoding RF1 knocked out, which allows for more efficient incorporation of UAAs at multiple sites, with up to ten unnatural amino acids incorporated into a single recombinant protein, with little effect on protein yield. By limiting termination of the stop codon UAA to RF2 only, RF1 is completely reassigned to the amber stop codon.

It is proposed that sperm whale myoglobin could be expressed using this modified cell line or similar, to optimise expression and incorporation efficiency. It should be noted however, that switching to the specialized cell line has no guarantee of increased yields, as the efficiency of UAG as a sense codon depends on multiple factors.

The orthogonal synthetases and tRNAs, although effective, have been artificially mutated for the incorporation of UAAs, unlike endogenous sets that have been evolutionarily tuned for optimal decoding. Furthermore, local protein structure, as well as mRNA context can possibly affect incorporation efficiency, despite expression using the modified cell line (Loscha *et al*, 2012). Mottagui-Tabar & Isaksson (1997) found that the presence of a lysine residue, three positions N' terminal to the stop codon, resulted in increased read-through. With this is mind, positons for UAA incorporation could be changed to reflect this observation, as well as adopting the modified JX33 cell-line, or equivalent. In doing so expression and incorporation of a given UAA can be further optimised.

The UAA being incorporated is another variable which should be considered. As stated prior, the incorporation of aliphatic amino acids is something which should be further explored. The modified lysine residue in this study is efficiently incorporated, however, from CW EPR spectra taken at r.t for the singly spin labelled pK4 mutant, there is a high degree of flexibility introduced by the extended lysine residue. With this in mind, one could take inspiration from the bis-legged Rx spin label (Fleissner *et al*,

2011), in synthesising a bis-legged spin label, with azide and MTS functional groups, for spin labelling of pK, anchored by labelling of the MTS functional group to a nearby cysteine residue. This could potentially produce a more rigid spin label, with less flexibility and conformational freedom than that demonstrated for the spin-labelled S4pK mutant.

By considering different spin labels, with various functional groups, as well as the structure of the amino acid being incorporated, and the cell line being used for expression, a 'Tool-Box' approach can be developed for the incorporation of different spin labels and UAAs into recombinant proteins.

Optimisation of current coupling conditions, as well as testing new labels and UAAs in the myoglobin model system will facilitate a smoother transition from the test system to systems which we would like to study, but would prove difficult by EPR.

As stated above, there are many cysteine rich proteins which can be recombinantly expressed using *E. coli* expression systems, and so perhaps application of such methods to the study of these proteins, using modified cell lines, could open up a new avenue for the study of cysteine rich proteins, using EPR spectroscopy.

Furthermore, such methods using orthogonal pyrrolysyl tRNA synthetase /$tRNA_{CUA(Pyl)}$ pairs are transferable to more complex expression systems, such as *Pichia pastoris*. As described in Chapter 3, when looking at the interactions of complement proteins fH and C3b, expression of the recombinant fH fragments proved challenging.

The presence of intermolecular disulphide bonds, making the proteins cysteine rich, significantly increased protein aggregation, and complicated the purification procedures. Keeping the protein in its reduced form for labelling purposes whilst ensuring the integrity of the disulphide bonds, and consequently protein fold, proved particularly challenging. Consequently, there is the potential for the incorporation of UAAs, into recombinant fH, using the *Pichia pastoris* system, which is well known for yielding large amounts of recombinant proteins from fermentation (often in gram quantities) (Macauley-Patrick *et al*, 2005).

By considering all variables mentioned above, the concept of SDSL opens up many avenues for exploring long range structure in biological systems which were otherwise out with the limits of the technique. Not only is this an exciting prospect for the EPR field, but for structural biology as a whole.

## 5.11 A New Generation of Spin Label

We are currently developing new strategies using several model protein systems for the incorporation of novel nitroxide spin labels for the SDSL of cysteine rich proteins. The new labels are specific for addition to only vicinal cysteine pairs, allowing specific labelling in the presence of single cysteine residues. This work is inspired by Huang *et al*, 2011 and their work on fluorescence labelling with $As^{III}$ binding and Smith *et al*, 2010 and their work on bromo-malemides and disulphide bridging.

Figure 5.19 shows the chemical structure of three spin labels (synthesised by Dr. Bouchra Hajjaj, University of St Andrews), BMSL, POMSL and SLAsH, designed for the specific labelling of cysteine pairs, in the presence of single cysteines.

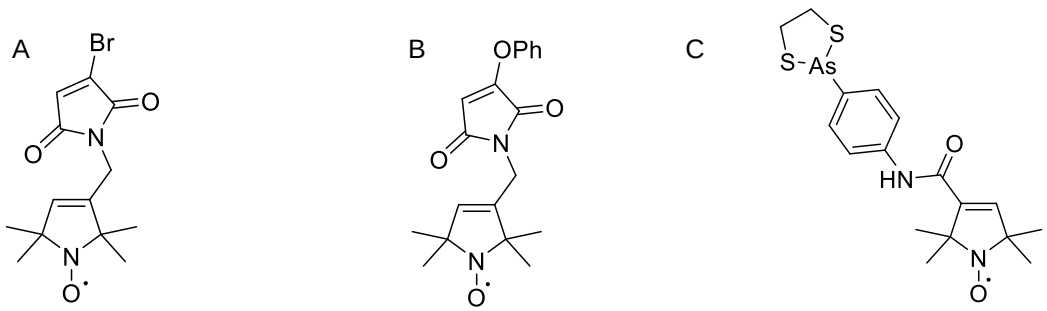



**Figure 5.19: Chemical Structures of novel nitroxide spin labels (A) BMSL, (B) POMSL, and (C) SLAsH. (visualised using ChemDraw™).**

### 5.11.1 Expression and Purification of the 14-3-3 ς and Vps75 Proteins

The recombinant proteins Vps75 and 14-3-3ς have been expressed with vicinal cysteine residues as test systems.

14-3-3ς is a dimer with each monomer containing a pair of *i*, *i*+1 (Q150C E151C) vicinal cysteine residues joined in a disulfide bond, whilst the Vps75 construct has vicinal cysteine residues incorporated at *i*, *i*+3 (L16C A19C) and *i*, *i*+4 (L16C K20C).

Figure 5.20 shows the XRC structure of 14-3-3ς, with the cysteine residues shown as spheres, alongside the XRC structure of Vps75 (*i* to *i*+3).

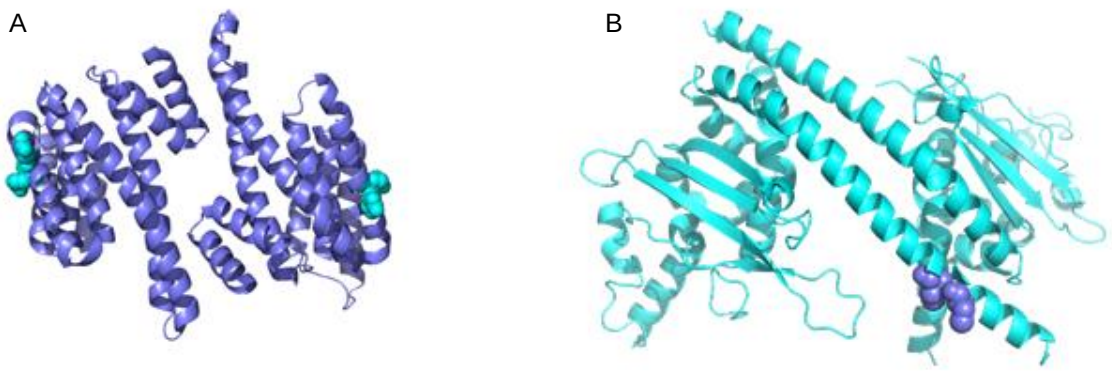



**Figure 5.20: XRC Structures of the (A) 14-3-3ς (PDB_ID 2BTP** (Yang *et al*, 2006)**) and (B) Vps75 (PDB_ID 2ZD7** (Park *et al*, 2008)**).** Mutants, with vicinal cysteines shown as spheres.

Figure 5.21 and 5.22 show the elution profiles and SDS-PAGE analysis of the model proteins following purification.

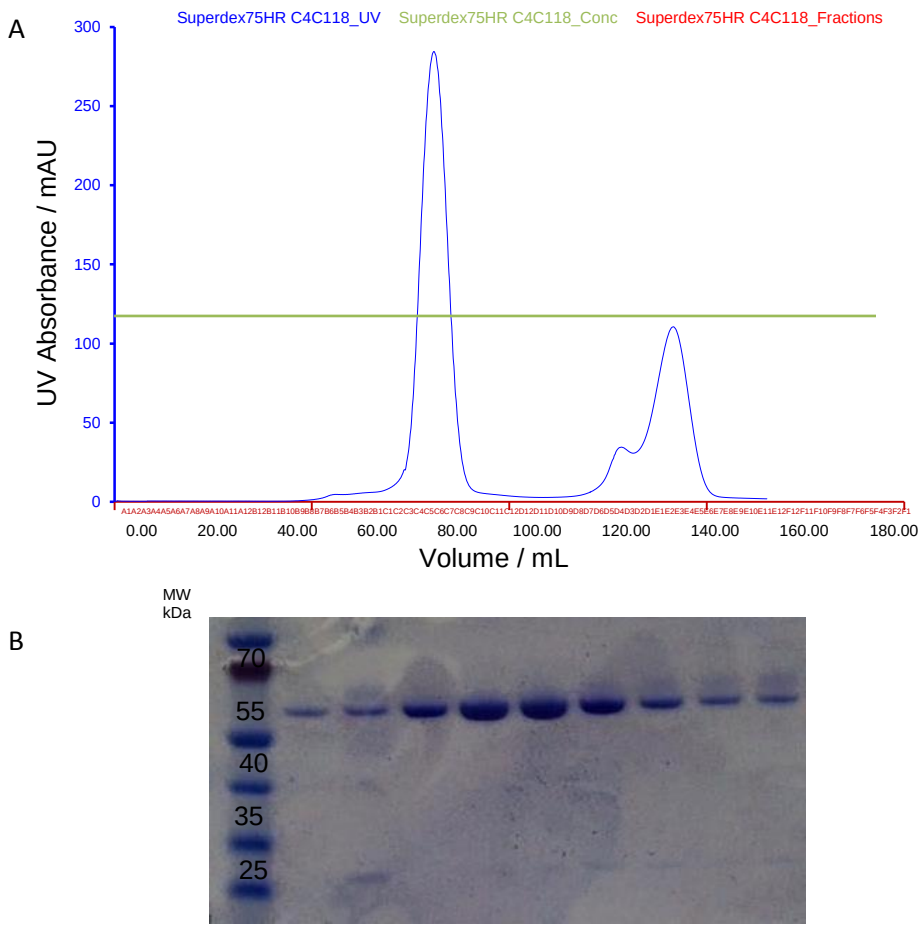



**Figure 5.21: Elution profile (A) and SDS-PAGE (B) analysis of 14-3-3ς following the final purification step (SEC).**

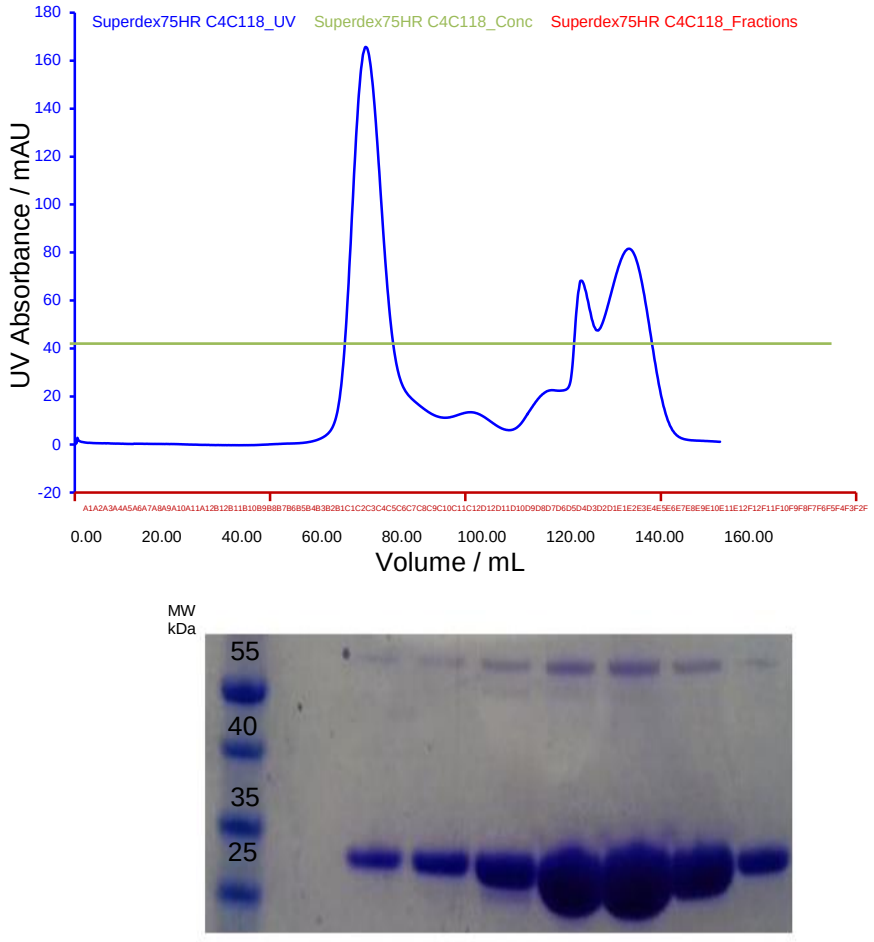



**Figure 5.22: Elution profile (A) and SDS-PAGE (B) analysis of Vps75 following the final purification step (SEC).**

### 5.11.2 Spin Labelling of 14-3-3ς

The 14-3-3ς Q150C E151C mutant was labelled with MTSSL, SLAsH and BMSL labels. Figure 5.23 shows the CW EPR spectra obtained. The SLAsH labelling was very poor and addition of it to the protein caused precipitation. Other methods for adding it will be investigated in the future or methods for making it more water soluble. Tests using varying levels of spin label together with Ellman's and mass spectrometry indicated that the BMSL cannot bridge these cysteines. Preliminary tests with the Vps75 (L16C A19C mutant) indicate that BMSL or POMSL can bridge the cysteines. Ongoing work will be to find conditions where the maleimide attached to one cysteine is selectively removed in the presence of the succinimide bridged pair of cysteines.

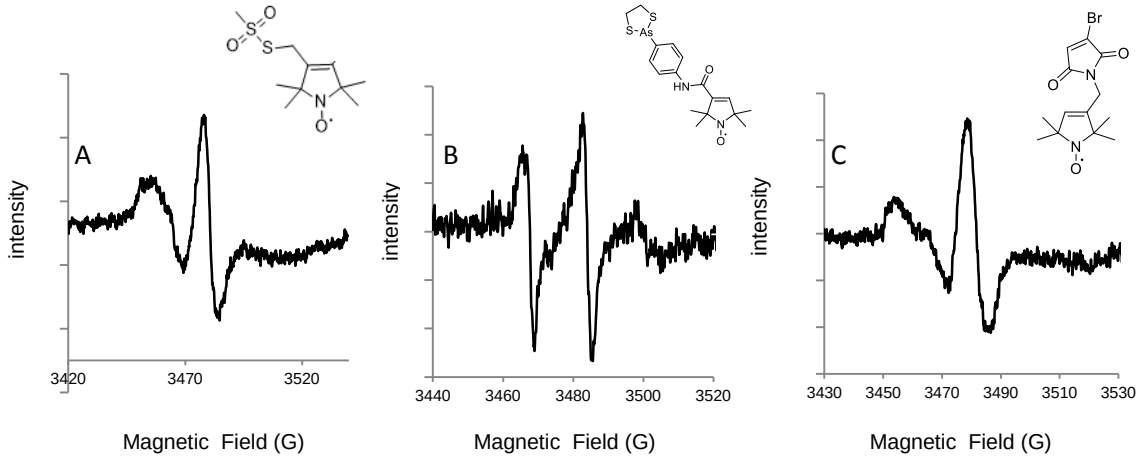



**Figure 5.23: CW EPR of labelled 14-3-3 ς with (A) MTSSL, (B) SLAsH and (C) BMSL.** Spectra taken at rt under non-saturating conditions.

## 5.12 Obtaining Nanometre Scale Distances Between the Ferric Heme of Sperm Whale Myoglobin and Nitroxide Spin Labels Using DEER

Many proteins contain naturally occurring paramagnetic metal centres or intrinsic organic radicals, however, in spite of this, the number of systems with intrinsic paramagnetism studied using DEER is limited. By taking advantage of these intrinsic EPR active sites, there is potential to study proteins and complexes thereof, which otherwise cannot be studied using traditional methods of EPR spectroscopy, such as in the study of cysteine rich proteins. However, the sensitivity of pulsed EPR measurements on broad-line metal centres is often restricted by the available excitation bandwidth, and to metal centres which exhibit moderate g-anisotropy and fast relaxation, e.g. $Cu^{II}$ (Narr *et al*, 2002, Yang *et al*, 2007).

For this reason, the field is so far restricted to the use of nitroxide spin labels, as these typically have only a small effect on the spectral line shape in DEER (Amsterdam *et al*, 2003). Many proteins contain ferric heme groups, in which the $Fe^{III}$ is paramagnetic, however, mainly due to large g-anisotropy and the fast electronic relaxation associated with low spin heme centres (even at low temperatures), obtaining high quality measurements, comparable to published crystal structures is difficult. There are two examples in the literature of measurements between a low spin ferric heme and a nitroxide. Schiemann (Abdullin et al, 2014) showed a very nice measurement using the

DEER experiment and another pulsed EPR set-up called RIDME (which takes advantage of the very different relaxation times of the centres). However, the P450 system is not as broad and is easier to manipulate into low spin than typical heme centres. Van Doorslaer (Ezhevskaya *et al*, 2013) measured a neuroglobin heme to nitroxide distance at X and Q-bands but the broadband and fast relaxing nature of the iron meant that the results had a poor signal to noise and an upper limit of 3nm distances seemed to be imposed.

### 5.12.1 Aims of this work

High field EPR is becoming an increasingly useful tool for the exploration of spin-labelled proteins, due to lifting of the degeneracy of the g-tensor at higher fields. The use of DEER at W band in the study of nitroxide-nitroxide distance measurements provides concentration sensitivity gains in the range of 20-30 times on commercial X-band spectrometers, however, similar gains may not be achieved with ferric heme proteins, as the large g anisotropy in such systems will lead to a significant broadening of the linewidth, meaning few spins may be detected. By applying a set of non-standard pulses during the DEER experiment at W band, such imperfections can be compensated for, provide broadband excitation whilst compensating for substantial applied field inhomogeneities.

This work looks at incorporating cysteine residues into recombinant sperm whale myoglobin, generating the single mutants S4C, S118C, and the double mutant S4CS118C. In each case, the intrinsic ferric heme ($Fe^{III}$) will be coordinated to imidazole, to induce low spin state, S=½, in the iron. Using broadband composite pulses applied to the single cysteine mutants, it should be possible to extract long range distance measurements between the ferric heme and the nitroxide spin label. With the double mutant, the nitroxide-nitroxide interaction can be independently determined from the nitroxide-heme interaction by appropriate spectral selection, highlighting the possibility for triangulation for studies involving protein-protein docking.

## 5.13 Expression and Purification of Low Spin (s= ½) Myoglobin Mutants

In order to yield low spin iron, the buffers used throughout were slightly different to those used in Appendix A9. Unless stated below, the purification was performed in the same manner, with washes and elution performed in the same volumes stated in 2.5.3.

Immediately following cell lysis, DTT was added to the non-cell extract at a final concentration of 50mM, before this was diluted to allow passage over the IMAC column. The protein was eluted from the column in 50mM Hepes, 125mM NaCl and 0.5M imidazole, pH 7.0. This was treated with 10mM DTT and further purified using SEC with a HiLoad 16/600 Superdex S-75 column (GEHealthcare) into 50mM Hepes, 125mM NaCl, pH 7.0. An SDS-PAGE gel with coomassie staining was used to confirm the purity of the monomeric species. A greater than 5 times excess of MTS spin label (Toronto Research Chemicals) was then added and the protein left at room temperature for 1 h. The protein was transferred into 50mM HEPES, 125mM NaCl, 200mM imidazole, pH 7.0 buffer in deuterium oxide, before 50% deuterated glycerol was added.

The presence of DTT throughout the purification process ensures that the myoglobin is maintained in its oxy-form. In a non-living system, oxymyoglobin ($Fe^{II}$) is slowly converted to metmyoglobin ($Fe^{III}$), and so the presence of a strong reducing agent prevents the formation of a mixed species. Full conversion of oxymyoglobin to metmyoglobin was monitored by the UV absorption spectra, by monitoring the position of the soret band and peaks in the 500-650 nm region, whilst the imidazole ligand coordinates to the heme, inducing a low spin state of the iron (S= ½) (Gurd *et al*, 1967).

## 5.14 DEER Experiments

HiPER is a high power home-built W-band spectrometer which is capable of reducing nitroxide-nitroxide DEER measurement times from 24h to as little as 20min (Cruickshank *et al*, 2009). The use of composite pulses, or wideband optimal control sequences, is particularly attractive when applied to high field EPR, as sensitivity is often constrained by a limited excitation bandwidth and B1 inhomogeneity. A composite pulse emulates the effect of traditional pulses, but includes an inbuilt compensatory mechanism rendering it less sensitive to common experimental imperfections (Levitt, 2011).

All measurements were carried out by Miss. Claire Motion, University of St Andrews, using the home-built HIPER 94 GHz (W band) high power pulsed electron spin resonance spectrometer. The samples were tested using the standard 4 pulse DEER experiment and with broadband composite pulses used when measuring between the ferric heme and a nitroxide spin label, and standard pulses for the doubly labelled cysteine mutant.

### 5.14.1 Nitroxide-Nitroxide DEER Measurements (S4CS118C) at W-Band

The ability to extract distances from a system that shows both nitroxide-nitroxide interactions as well as nitroxide-heme interactions shows the potential for triangulation and protein docking using EPR spectroscopy. Figure 5.24 shows the DEER time traces and distance distributions obtained using 4-pulse DEER for the double cysteine mutant at 50K. These are shown alongside time traces and distributions simulated using MMM. They show good agreement.

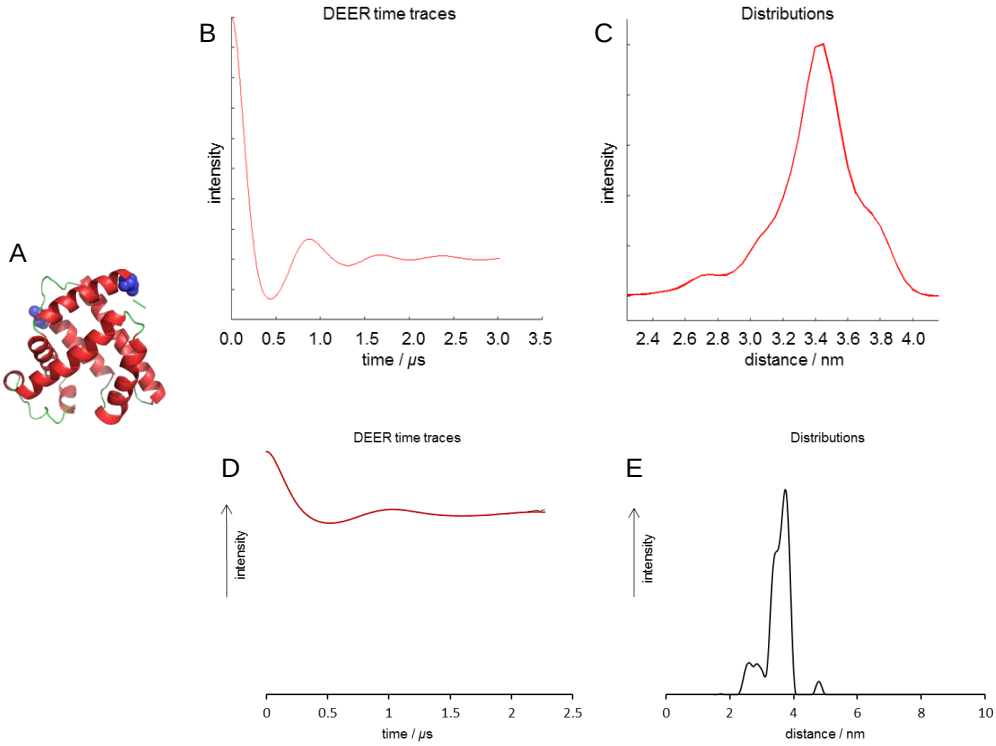



**Figure 5.24: DEER time traces and associated distance distributions for the S4CS118C doubly labelled Met-myoglobin mutant, shown alongside MMM simulation.** (A) Structure of sperm whale myoglobin (PDB_ID: 1mBN), with cysteine residues represented as blue spheres. (B) MMM modelled DEER time trace and (C) Distance distributions for the S4CS118C mutant; (D) Background corrected DEER time trace (following DeerAnalysis) for the S4CS118C mutant measured on HiPER and (E) Distance distributions from DeerAnalysis for the S4CS118C mutant.

### 5.14.2 Nitroxide-Heme ($Fe^{III}$) DEER Measurements at W-Band

The DEER between the single cysteine mutant spin labels and the ferric heme was taken using composite pulses to compensate for the increased inhomogeneity introduced by the metal centre at high frequency.

For these experiments it is highly advantageous to probe on the $Fe^{III}$, and pump on the nitroxide. In doing so, $Fe^{III}$- nitroxide DEER measurements are experimentally viable, despite the spectral width of the $Fe^{III}$ signal being 10x greater at W-band than at X-band, and 200x greater than the nitroxide spectral width at W band (Cruickshank *et al*, 2009). Figure 5.25 and 5.26 shows DEER time traces and distance distributions obtaining using composite pulse DEER (based on the 4-pulse sequence) for the S4C and S118C nitroxides, to the ferric heme, respectively. Measurements were taken at 6K. These are shown alongside time traces and distributions simulated using MMM

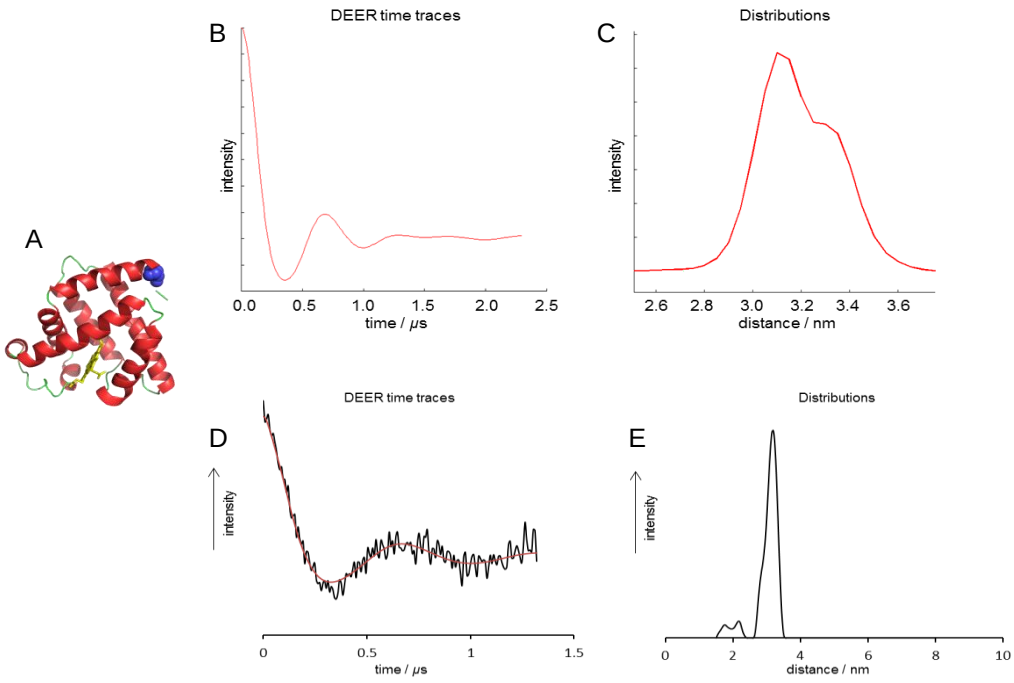



**Figure 5.25: DEER time traces and associated distance distributions for the S4C singly labelled Met-myoglobin mutant to the ferric heme, shown alongside MMM simulation.** (A) Structure of sperm whale myoglobin (PDB_ID: 1mBN), with cysteine residue represented as blue spheres. (B) DEER time trace and (C) Distance distributions for the S4C mutant, simulated using MMM. (D) Background corrected DEER time trace (following DeerAnalysis) for the S4C mutant. (E) Distance distributions from DeerAnalysis for the S4C mutant.

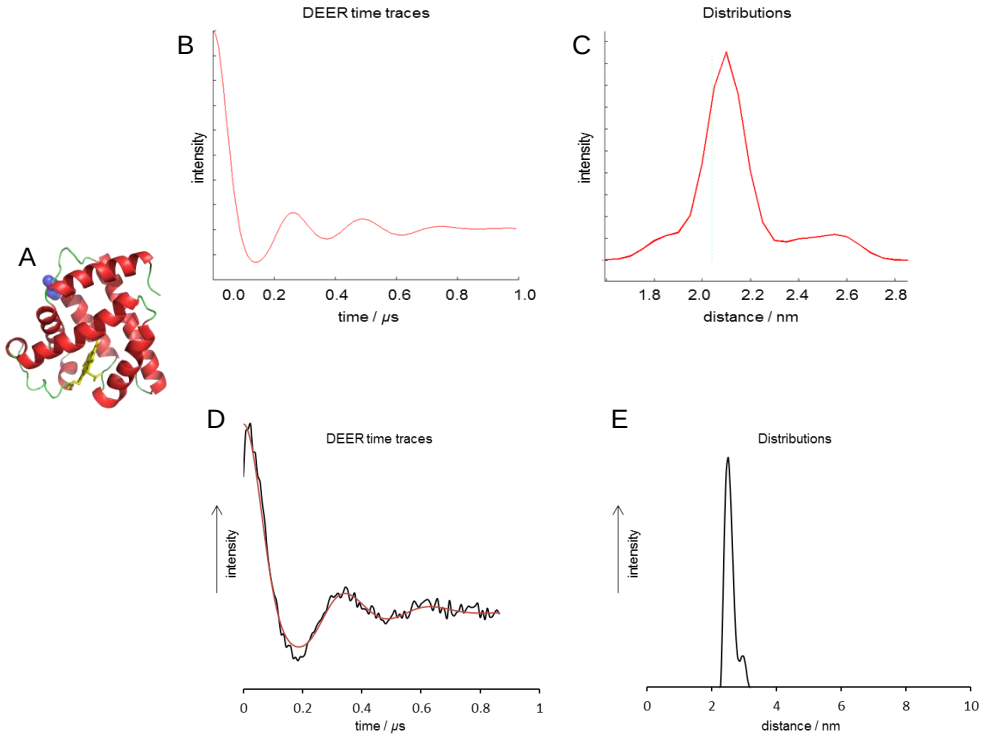



**Figure 5.26: DEER time traces and associated distance distributions for the S118C singly labelled Met-myoglobin mutant to the ferric heme, shown alongside MMM simulation.** (A) Structure of sperm whale myoglobin (PDB_ID: 1mBN), with cysteine residue represented as blue spheres. (B) DEER time trace and (C) Distance distributions for the S4C mutant, simulated using MMM. (D)Background corrected DEER time trace (following DeerAnalysis) for the S4C mutant. (E) Distance distributions from DeerAnalysis for the S4C mutant.

Looking at the experimental data for each of the singly labelled single cysteine mutants measured to the ferric heme centre of the met-myoglobin, and comparing this to simulation, the distance distributions obtained are in good agreement. The results reported here are very exciting since they open up the possibility of measuring many otherwise unmeasurable systems. Measurements are underway for the S4C S118C doubly labelled mutant with ferric heme to investigate whether triangulation can be directly applied in this system. Preliminary results are promising.

## 5.15 Beyond MTSSL - Outlook and Future Work

The diversity of protein function is made possible because almost half of all enzymes require the incorporation of cofactors, such as metal atoms, to function (Finkelstein, 2009), whilst metal binding sites on proteins are responsible for catalyzing some of the most physiologically relevant reactions (Yeung *et al*, 2009). This development opens up the potential to study a full range of metalloproteins, or proteins with native accessible paramagnetic centres, that otherwise would be out with the remit of DEER.

The MTSSL spin label is effective for the study of many biologically relevant systems using EPR spectroscopy however, it is not without limitation. EPR is a highly specific technique which can be used to explore long range structure in systems which cannot be explored by other high resolution structural techniques. However, the main caveat of EPR is that it works best in its simple form, when looking at the interactions of two radicals.

This is simple enough when specifically introducing cysteine residues into proteins which do not have any or have few native cysteine residues. The problem of specificity arises when studying proteins which have many free cysteine residues, which, in order to maintain structural integrity, cannot be exchanged. In such instances, the development of new methods of spin-labelling and new labels themselves can only progress the field of SDSL, and open avenues of structural exploration that otherwise would not be possible.

The inspiration behind investigating new labelling strategies was the protein Nitric Oxide Synthase (NOS) (Structure and Mechanisms reviewed by (Daff, 2010). NOS is found in the human body in several different isoforms and is a major signalling molecule, and is consequently tightly regulated. It exists as a 160kDa homodimer which binds Calmodulin to exert its function. Until 2014 the structure of the homodimer was not known (Campbell *et al*, 2014) and while many aspects of the system were suitable for investigation by EPR distance measurements the presence of 30 free cysteine residues made site-directed spin labelling impossible.

Consequently it was proposed that NOS structure was explored by utilising its binding partner calmodulin (CaM) for SDSL. Figure 5.27 is a schematic representation of the NOS domains, with calcium bound.

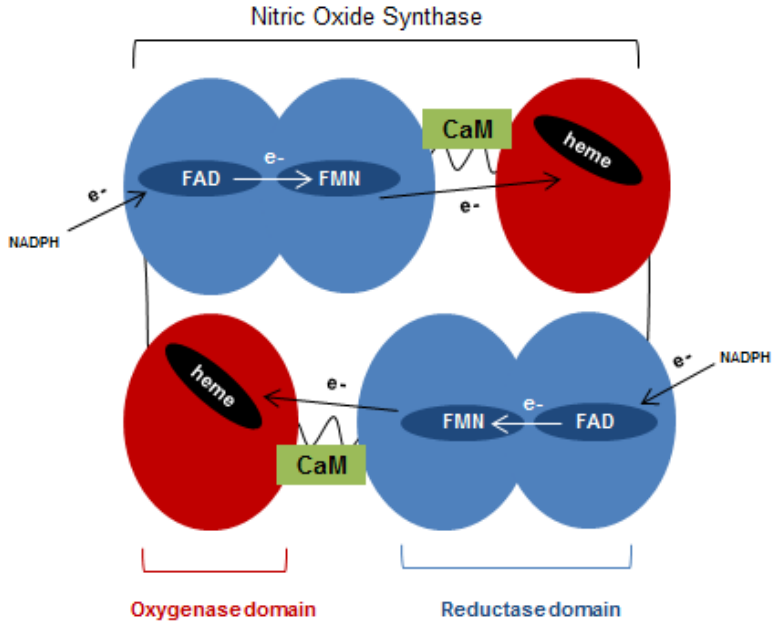



**Figure 5.27: Schematic representation of the NOS homodimer with NOS CaM bound.**

By spin-labelling the calmodulin, one could use EPR to extract a range of distances between mutants on each monomer, which could give further insight into the entire heterodimer. This, however, does little to shed insight into structural changes that occur as a result of activity. Ideally one would like to explore NOS structure when CaM is bound, however, the cysteine residues prove problematic.

NOS contains within its domains up to four paramagnetic centres, a FMN, FAD, heme and proteins (Rakowsky *et al*, 1998). Measuring distances from these to spin labels would be valuable for determining structure. Following the work done in 5.3, it was possible to measure accurately distances between spin labelled cysteine residues and the ferric heme of sperm whale myoglobin, using high field EPR, and composite pulses. Consequently, the same methodologies could be employed, to measure distances between the native paramagnetic centres of NOS (heme and flavin radicals), and local affects in NOS structure by the addition of CaM.

# Appendices

## Appendix A: Stock Solutions, Buffers, and Media Compositions

### *A1: Stock Solutions*

**10x YNB:** 34g YNB, 100g $(NH_4)_2SO_4$, adjusted to 1000mL with $ddH_2O$, filter sterilised

**500x B**: 20mg biotin adjusted to 100mL with $ddH_2O$, filter sterilised

**10x M:** 5mL MeOH, 95mL $ddH_2O$, filter sterilised

**10x GY:** 100mL glycerol, 900mL $ddH_2O$, autoclave for 20min

**10x D:** 200g D-glucose, adjusted to 1000mL with $ddH_2O$, autoclave 20min

**0.5m EDTA:** 76.06g EDTA, pH8.2, adjusted to 500mL with $ddH_2O$

**1M $K_2HPO_4$**: 87.09g $K_2HPO_4$, adjusted to 500mL with $ddH_2O$

**1M $KH_2PO_4$:** 136.09g $KH_2PO_4$, adjusted to 500mL with $ddH_2O$

**1M Potassium Phospahte Buffer:** 132mL of 1M$K_2HPO_4$, 868mL of $KH_2PO_4$, pH6.0

**4M NaCl:** 233.76g NaCl, adjusted to 1000mL with $ddH_2O$

**1M DTT:** 0.154g DTT, adjusted to 1000$\mu$L with $ddH_2O$

**1M TCEP**: 0.286g TCEP, adjusted to 1000$\mu$L with $ddH_20$

**Coomassie Stain:** 1.00g Coomassie $_{R250}$, 100mL glacial acetic acid, 400mL MeOH, adjusted to 1000mL with $ddH_2O$

**Coomassie Destain Solution:** 200mL MeOH, 100mL glacial acetic acid, adjusted to 1000mL with $ddH_2O$

### *A2: Buffers*

**10x PBS:** 80g NaCl, 2.0g KCl, 14.4g $Na_2HPO_4$, 2.4g $KH_2PO_4$, adjusted to 1000mL with $ddH_2O$ (pH7.4)

**10X TBE:** 108g Trizma® base, 55g boric acid, 7.5g EDTA, disodium salt, adjust to 1000ml with $ddH_2O$

## *A3: fH1-4 Preparations*

The gene for full length fH was purchased from GeneArt, and fH fragments cloned into the pPICZαB vector (both provided by Prof Paul N. Barlow, University of Edinburgh, Scotland, UK).

IMAC – Ni-NTA Superflow Cartridge 5mL

**Equilibration Buffer:** 20mM Potassium Phosphate, 0.5M NaCl, pH7.0

**Wash Buffer:** 20mM Potassium Phosphate, 0.5M NaCl, 20mM Imidazole, pH7.0

**Elution Buffer:** 20mM Potassium Phosphate, 0.5M NaCl, 500mM Imidazole, pH7.0

Anion Exchange Chromatography – Tricorn™ Mono Q 4.6/100 PE

**Equilibration Buffer:** 20mM Sodium Carbonate, 1mM EDTA, pH9.0

**Elution Buffer:** 20mM Sodium Carbonate, 1mM EDTA, 1M NaCl, pH9.0

SEC: HiLoad™ 16/600 Superdex™ S-75

**Equilibration/Elution Buffer:** 1X PBS, pH7.4

## *A4: fH19-20 Preparations*

The gene for full length fH was purchased from GeneArt, and fH fragments cloned into the pPICZαB vector (both provided by Prof Paul N. Barlow, University of Edinburgh, Scotland, UK).

Cation Exchange Chromatography - SP-Sepharose FastFlow™ Resin

**Equilibration/Wash Buffer:** 20mM Potassium Phosphate, pH6.6

**Elution Buffer:** 20mM Potassium Phosphate, 1M NaCl, pH6.6

Cation Exchange Chromatography – Resource 15S™

**Equilibration/Wash Buffer:** 20mM Potassium Phosphate, pH6.6

**Elution Buffer:** 20mM Potassium Phosphate, 1M NaCl, pH6.6

SEC: HiLoad™ 16/600 Superdex™ S-75

**Equilibration/Elution Buffer:** 1X PBS, pH7.4

## *A5: C3 Preparations*

Anion Exchange Chromatography - Q-Sepharose FastFlow™ Resin

**Equilibration/Wash Buffer:** 90% 20mM Tris, 50mM εACA, 5mM EDTA, 0.2mM PMSF, 10% Elution Buffer

**Elution Buffer:** 20mM Tris, 50mM εACA, 5mM EDTA, 0.2mM PMSF, 1M NaCl

Anion Exchange Chromatography – Tricorn™ Mono Q 4.6/100 PE

**Equilibration/Wash Buffer:** 95% 20mM Tris, 50mM εACA, 5mM EDTA, 0.2mM PMSF, 5% Elution Buffer

**Elution Buffer:** 20mM Tris, 50mM εACA, 5mM EDTA, 0.2mM PMSF, 1M NaCl

## *A6: C3b/C3(N) Preparations*

Anion Exchange Chromatography – Tricorn™ Mono Q 4.6/100 PE

**Equilibration/Wash Buffer:** 20mM Tris, 50mM εACA, 5mM EDTA, 0.2mM PMSF

**Elution Buffer:** 20mM Tris, 50mM εACA, 5mM EDTA, 0.2mM PMSF, 1M NaCl

SEC: HiLoad™ 16/600 Superdex™ S-75

**Equilibration/Elution Buffer:** 1X PBS, pH7.4

## *A7: MiniHMM/RLC and cMyBP-C Domain C0 Preparations*

The plasmids for expression of the MiniHMM fragment, cardiac RLC and cMyBP-C Domain C0 were provided by Dr Mark Pfuhl, Kings College, London, UK.

IMAC – Ni-NTA Superflow Cartridge 5mL

**Lysis Buffer:** 25mM Hepes, 125mM NaCl, 10% Glycerol, 1mM PMSF, 2 SigmaFAST™ Protease Inhibitor Cocktail Tablets, EDTA-Free (per 100mL of Lysis Buffer),pH6.8

**Equilibration Buffer:** 25mM Hepes, 125mM NaCl, 20mM Imidazole, 10% Glycerol, pH6.8

**Wash Buffer:** 25mM Hepes, 125mM NaCl, 40mM Imidazole, 10% Glycerol, pH6.8

**Elution Buffer:** 25mM Hepes, 125mM NaCl, 500mM Imidazole, 10% Glycerol, pH6.8

**SEC: HiLoad™ 16/600 Superdex™ S-75**

**Equilibration/Elution Buffer:** 25mM Hepes, 125mM NaCl, 10% Glycerol, pH6.8

## *A8: 6xhis TEV Protease Preparations*

A glycerol stock containing Rosetta™ 2(DE3) cells successfully transformed with the recombinant 6xHIS TEV protease was kindly donated by Dr Huanting Liu and Prof. James H. Naismith, University of St Andrews, Scotland, UK.

**Lysis Buffer:** 1xPBS, 0.3M NaCl, 10mM Imidazole, 1mM PMSF, 1mM Benzamidine, pH7.4

**Equilibration/Wash Buffer:** 1xPBS, 0.3M NaCl, 30mM Imidazole, 1mM PMSF, 1mM Benzamidine, pH7.4

**Elution Buffer:** 1xPBS, 0.3M NaCl, 250mM Imidazole, 1mM PMSF, 1mM Benzamidine, pH7.4

**Dialysis Buffer 1:** 50mM Tris-HCl, 0.3M NaCl, 1mM PMSF, pH7.4

**Dialysis Buffer 2:** 50mM Tris-HCl, 0.3M NaCl, 50% Glycerol, pH7.4

## *A9: Myoglobin Preparations*

The plasmids pBk-Pyls and pMyo4-TAG-PylT used throughout this work were kindly supplied by Prof. Jason Chin. The pBk-Pyls plasmid contains the pyrrolysyl tRNA synthetase / $tRNA_{CUA(Pyl)}$ pair whilst the pMyo4-TAG-PylT contains the gene for recombinant myoglobin expression.

IMAC – Ni-NTA Superflow Cartridge 5mL

**Lysis Buffer:** 50mM Hepes, 125mM NaCl, 10mM Imidazole, 1mM PMSF, 2 SigmaFAST™ Protease Inhibitor Cocktail Tablets, EDTA-Free (per 100mL Lysis Buffer), pH7.0

**Equilibration/Wash Buffer:** 50mM Hepes, 125mM NaCl, 25mM Imidazole, pH7.0

**Elution Buffer:** 50mM Hepes, 125mM NaCl, 500mM Imidazole, pH7.0

SEC: HiLoad™ 16/600 Superdex™ S-75

**Equilibration/Elution Buffer:** 50mM Hepes, 125mM NaCl, pH7.0

## *A10: 14-3-3 ς Preparations*

Agar plates from the successful transformation of the recombinant 14-3-3 ς proteins into Rosetta™ 2(DE3) were kindly provided by Dr David Norman, University of Dundee, Scotland, UK. From these glycerol stocks were made for subsequent expression.

IMAC: Ni-NTA Superflow Cartridge 5mL

**Lysis Buffer:** 50mM Tris-HCl, 100mM NaCl, 1mM PMSF, 2 SigmaFAST™ Protease Inhibitor Cocktail Tablets, EDTA-Free (per 100mL of Lysis Buffer), pH7.5

**Equilibration/Wash Buffer:** 50mM Tris-HCl, 100mM NaCl, 20mM Imidazole, pH7.5

**Elution Buffer:** 50mM Tris-HCl, 100mM NaCl, 500mM Imidazole, pH7.5

SEC: HiLoad™ 16/600 Superdex™ S-75

**Equilibration/Elution Buffer:** 50mM Tris-HCl, 100mM NaCl, pH7.5

## *A11: Vps75 Preparations*

Agar plates from the successful transformation of the recombinant Vps75 proteins into Rosetta™ 2(DE3) were kindly provided by Dr David Norman, University of Dundee, Scotland, UK. From these glycerol stocks were made for subsequent expression.

IMAC: Ni-NTA Superflow Cartridge 5mL

**Lysis Buffer:** 20mM Tris, 500mM NaCl, 1mM PMSF, 2 SigmaFAST™ Protease Inhibitor Cocktail Tablets, EDTA-Free (per 100mL of Lysis Buffer), pH7.5

**Equilibration Buffer:** 20mM Tris, 500mM NaCl, pH7.5

**Wash Buffer:** 20mM Tris, 500mM NaCl, 5mM Imidazole, pH7.5

**Elution Buffer**: 20mM Tris, 500mM NaCl, 400mM Imidazole, pH7.5

SEC: HiLoad™ 16/600 Superdex™ S-75

**Equilibration/Elution Buffer:** 20mM Tris, 500mM NaCl, pH7.5

## A12: Media

**LB Broth:** 10g Bacto-tryptone, 5g Yeast Extract, 10g NaCl, adjusted to 1000mL with $ddH_2O$, pH7.5

**LB Agar:** 10g Bacto-tryptone, 5g Yeast Extract, 10g NaCl, 15g agar, adjusted to 1000mL with $ddH_2O$, pH7.5

**LSLB Broth:** 10g Bacto-tryptone, 5g Yeast Extract, 5g NaCl, adjusted to 1000mL with $ddH_2O$, pH7.5

**LSLB Agar**: 10g Bacto-tryptone, 5g Yeast Extract, 5g NaCl, 15g agar, adjusted to 1000mL with $ddH_2O$, pH7.5

**BMG Media:** Autoclave 700mL $ddH_2O$, cool to r.t before the addition of 100mL 1M potassium phosphate buffer, pH6.0, 100mL 10X YNB, 2mL 500X B and 100mL of 10X GY.

**BMGY Media:** As BMG, but supplemented with 10g Yeast Extract and 20g Bacto-peptone before autoclaving (20min).

**BMM Media:** Autoclave 700mL $ddH_2O$, cool to r.t before the addition of 100mL 1M potassium phosphate buffer, pH6.0, 100mL 10X YNB, 2mL 500X B and 100mL of 10X M.

**BMMY Media:** As BMM, but supplemented with 10g Yeast Extract and 20g Bacto-peptone before autoclaving (20min).

**YPD Broth:** Dissolve 10g Yeast Extract and 20g of Bacto-peptone in 900mL $ddH_2O$. Autoclave for 20 min before the addition of 100mL 10X D.

**YPD Agar:** As YPD Broth, but supplemented with 20g of agar before autoclaving (20min).

**YPDS Broth:** Dissolve 10g Yeast Extract, 20g of Bacto-peptone and 182.2g sorbitol in 900mL $ddH_2O$. Autoclave for 20 min before the addition of 100mL 10X D.

**YPDS Agar:** As YPDS Broth, but supplemented with 20g of agar before autoclaving (20min).

## A13: Primers

Amplification of fH1-4 from full length fH gene

FOR: 5’ gaggattgtaacgagttgccaccaag 3’

REV: 5’ cttctcttcacaggatggcaatggtc 3’

K247C FOR: 5’ catgatctgcaggagaggattgtaacgagttgccaccaag 3’

K247C REV: 5’ taataatctagactactaacactcttcacaggatggcaatggtc 3’

K247CHistag FOR: 5’ catgatctgcaggagaggattgtaacgagttgccaccaag 3’

K247CHistag REV: 5’ taataatctagactactaatgatgatgatgatgatgacactcttcacaggatggcaatggtc 3’

Amplification of fH1920 from full length fH gene

FOR: 5’ ggggtccactggtaagtgtggtccacctcc 3’

REV: 5’ cccctctcttagcacaagttgggtactcc 3’

RLC Mutagenesis Primers

K30C FOR: 5’ ccaggaattttgtgaggccttcactatcatgg 3’

K30C REV: 5’ ccatgatagtgaaggcctcacaaaattcctgg 3’

E88C FOR: 5’ cctcacaatgtttgggtgtaaacttaagggagcgg 3’

E88C REV: 5’ ccgctcccttaagtttacacccaaacattgtgagg 3’

E97C FOR: 5’ gcggaccctgaggaatgtattctcaacgcattc 3’

E97C REV: 5’ gaatgcgttgagaatacattcctcagggtccgt 3’

D117C FOR: 5’ ggggtgctgaaggcttgttacgttcgggaaatgc 3’

D117C REV: 5’ gcatttcccgaacgtaacaagccttcagcacccc 3’

T125C FOR: 5’ cgggaaatgctgacctgtcaggcggagagg 3’

T125C REV: 5’ cctctccgcctgacaggtcagcatttcccg 3’

Sperm Whale Myoglobin Mutagenesis Primers

S118C FOR: 5’ catgttctgcattgtagacatccagg 3’

S118C REV: 5’ cctggatgtctacaatgcagaacatg 3’

S4C FOR: 5’ accatggttctgtgtgaaggtgaatgg

S4C REV: 5’ ccattcaccttcacacagaaccatggt 3’

TAG4S FOR: 5’ accatggttctgtctgaaggtgaatgg

TAG4S REV: 5’ ccattcaccttcagacagaaccatggt 3’

S118TAG FOR: 5’ catgttctgcattagagacatccaggt

S118TAG REV: 5’acctggatgtctctaatgcagaacatg 3’

# Appendix B: Protein Sequences (Wild Type)

## B1: Full Length fH, Codon Optimised for Expression in *Pichia pastoris*

```
        10         20         30         40         50         60
MRLLAKIICL MLWAICVAED CNELPPRRNT EILTGSWSDQ TYPEGTQAIY KCRPGYRSLG
        70         80         90        100        110        120
NVIMVCRKGE WVALNPLRKC QKRPCGHPGD TPFGTFTLTG GNVFEYGVKA VYTCNEGYQL
       130        140        150        160        170        180
LGEINYRECD TDGWTNDIPI CEVVKCLPVT APENGKIVSS AMEPDREYHF GQAVRFVCNS
       190        200        210        220        230        240
GYKIEGDEEM HCSDDGFWSK EKPKCVEISC KSPDVINGSP ISQKIIYKEN ERFQYKCNMG
       250        260        270        280        290        300
YEYSERGDAV CTESGWRPLP SCEEKSCDNP YIPNGDYSPL RIKHRTGDEI TYQCRNGFYP
       310        320        330        340        350        360
ATRGNTAKCT STGWIPAPRC TLKPCDYPDI KHGGLYHENM RRPYFPVAVG KYYSYYCDEH
       370        380        390        400        410        420
FETPSGSYWD HIHCTQDGWS PAVPCLRKCY FPYLENGYNQ NYGRKFVQGK SIDVACHPGY
       430        440        450        460        470        480
ALPKAQTTVT CMENGWSPTP RCIRVKTCSK SSIDIENGFI SESQYTYALK EKAKYQCKLG
       490        500        510        520        530        540
YVTADGETSG  SITCGKDGWS  AQPTCIKSCD  IPVFMNARTK  NDFTWFKLND  TLDYECHDGY
       550        560        570        580        590        600
ESNTGSTTGS  IVCGYNGWSD  LPICYERECE  LPKIDVHLVP  DRKKDQYKVG  EVLKFSCKPG
       610        620        630        640        650        660
FTIVGPNSVQ  CYHFGLSPDL  PICKEQVQSC  GPPPELLNGN  VKEKTKEEYG  HSEVVEYYCN
       670        680        690        700        710        720
PRFLMKGPNK  IQCVDGEWTT  LPVCIVEEST  CGDIPELEHG  WAQLSSPPYY  YGDSVEFNCS
       730        740        750        760        770        780
ESFTMIGHRS  ITCIHGVWTQ  LPQCVAIDKL  KKCKSSNLII  LEEHLKNKKE  FDHNSNIRYR
       790        800        810        820        830        840
CRGKEGWIHT  VCINGRWDPE  VNCSMAQIQL  CPPPPQIPNS  HNMTTTLNYR  DGEKVSVLCQ
       850        860        870        880        890        900
ENYLIQEGEE  ITCKDGRWQS  IPLCVEKIPC  SQPPQIEHGT  INSSRSSQES  YAHGTKLSYT
       910        920        930        940        950        960
CEGGFRISEE  NETTCYMGKW  SSPPQCEGLP  CKSPPEISHG  VVAHMSDSYQ  YGEEVTYKCF
       970        980        990       1000       1010       1020
EGFGIDGPAI  AKCLGEKWSH  PPSCIKTDCL  SLPSFENAIP  MGEKKDVYKA  GEQVTYTCAT
      1030       1040       1050       1060       1070       1080
YYKMDGASNV  TCINSRWTGR  PTCRDTSCVN  PPTVQNAYIV  SRQMSKYPSG  ERVRYQCRSP
      1090       1100       1110       1120       1130       1140
YEMFGDEEVM  CLNGNWTEPP  QCKDSTGKCG  PPPPIDNGDI  TSFPLSVYAP  ASSVEYQCQN
      1150       1160       1170       1180       1190       1200
LYQLEGNKRI  TCRNGQWSEP  PKCLHPCVIS  REIMENYNIA  LRWTAKQKLY  SRTGESVEFV
      1210                          1220                          1230
CKRGYRLSSR SHTLRTTCWD GKLEYPTCAK R
```

**Molecular weight:** 13,9096.3 (actual molecular weight ~155kDa due to glycosylation)
**Theoretical pI:** 6.2
**Extinction Coefficient ($A_{280}$):** 247,300

## B2: fH 1-4 Fragment, Codon Optimised for Expression in *Pichia pastoris*

```
        10             20             30             40             50
MRLLAKIICL     MLWAICVAED     CNELPPRRNT     EILTGSWSDQ     TYPEGTQAIY
        60             70             80             90            100
KCRPGYRSLG     NVIMVCRKGE     WVALNPLRKC     QKRPCGHPGD     TPFGTFTLTG
       110            120            130            140            150
GNVFEYGVKA     VYTCNEGYQL     LGEINYRECD     TDGWTNDIPI     CEVVKCLPVT
       160            170            180            190            200
APENGKIVSS     AMEPDREYHF     GQAVRFVCNS     GYKIEGDEEM     HCSDDGFWSK
       210            220            230            240            250
EKPKCVEISC     KSPDVINGSP     ISQKIIYKEN     ERFQYKCNMG     YEYSERGDAV
       260
CTESGWRPLP SCEEK
```

**Molecular weight:** 29,972.1
**Theoretical pI:** 5.32
**Extinction Coefficient ($A_{280}$):** 52,370

## B3: fH19-20 Fragment, Codon Optimised for Expression in *Pichia pastoris*

```
      1110       1120       1130       1140       1150
GKCG       PPPPIDNGDI       TSFPLSVYAP       ASSVEYQCQN       LYQLEGNKRI
      1160            1170            1180            1190            1200
TCRNGQWSEP     PKCLHPCVIS     REIMENYNIA     LRWTAKQKLY     SRTGESVEFV
      1210                           1220                           1230
CKRGYRLSSR SHTLRTTCWD GKLEYPTCAK R
```

**Molecular weight:** 14,277.2
**Theoretical pI:** 9.08
**Extinction Coefficient ($A_{280}$):** 26,930

## B4: Complement C3

```
   10         20         30         40         50         60
SPMYSIITPN ILRLESEETM VLEAHDAQGD VPVTVTVHDF PGKKLVLSSE KTVLTPATNH

         70         80         90        100        110        120
MGNVTFTIPA NREFKSEKGR NKFVTVQATF GTQVVEKVVL VSLQSGYLFI QTDKTIYTPG

       130        140        150        160        170        180
STVLYRIFTV NHKLLPVGRT VMVNIENPEG IPVKQDSLSS QNQLGVLPLS WDIPELVNMG

       190        200        210        220        230        240
QWKIRAYYEN SPQQVFSTEF EVKEYVLPSF EVIVEPTEKF YYIYNEKGLE VTITARFLYG

       250        260        270        280        290        300
KKVEGTAFVI FGIQDGEQRI SLPESLKRIP IEDGSGEVVL SRKVLLDGVQ NPRAEDLVGK

       310        320        330        340        350        360
SLYVSATVIL HSGSDMVQAE RSGIPIVTSP YQIHFTKTPK YFKPGMPFDL MVFVTNPDGS
```

```
       370        380        390        400        410        420
PAYRVPVAVQ GEDTVQSLTQ GDGVAKLSIN THPSQKPLSI TVRTKKQELS EAEQATRTMQ

       430        440        450        460        470        480
ALPYSTVGNS NNYLHLSVLR TELRPGETLN VNFLLRMDRA HEAKIRYYTY LIMNKGRLLK

       490        500        510        520        530        540
AGRQVREPGQ DLVVLPLSIT TDFIPSFRLV AYYTLIGASG QREVVADSVW VDVKDSCVGS

       550        560        570        580        590        600
LVVKSGQSED RQPVPGQQMT LKIEGDHGAR VVLVAVDKGV FVLNKKNKLT QSKIWDVVEK

       610        620        630        640        650        660
ADIGCTPGSG KDYAGVFSDA GLTFTSSSGQ QTAQRAELQC PQPAASVQLT EKRMDKVGKY

       670        680        690        700        710        720
PKELRKCCED GMRENPMRFS CQRRTRFISL GEACKKVFLD CCNYITELRR QHARASHLGL

       730        740        750        760        770        780
ARSNLDEDII AEENIVSRSE FPESWLWNVE DLKEPPKNGI STKLMNIFLK DSITTWEILA

       790        800        810        820        830        840
VSMSDKKGIC VADPFEVTVM QDFFIDLRLP YSVVRNEQVE IRAVLYNYRQ NQELKVRVEL

       850        860        870        880        890        900
LHNPAFCSLA TTKRRHQQTV TIPPKSSLSV PYVIVPLKTG LQEVEVKAAV YHHFISDGVR

       910        920        930        940        950        960
KSLKVVPEGI RMNKTVAVRT LDPERLGREG VQKEDIPPAD LSDQVPDTES ETRILLQGTP

       970        980        990       1000       1010       1020
VAQMTEDAVD AERLKHLIVT PSGCGEQNMI GMTPTVIAVH YLDETEQWEK FGLEKRQGAL

      1030       1040       1050       1060       1070       1080
ELIKKGYTQQ LAFRQPSSAF AAFVKRAPST WLTAYVVKVF SLAVNLIAID SQVLCGAVKW

      1090       1100       1110       1120       1130       1140
LILEKQKPDG VFQEDAPVIH QEMIGGLRNN NEKDMALTAF VLISLQEAKD ICEEQVNSLP

      1150       1160       1170       1180       1190       1200
GSITKAGDFL EANYMNLQRS YTVAIAGYAL AQMGRLKGPL LNKFLTTAKD KNRWEDPGKQ

      1210       1220       1230       1240       1250       1260
LYNVEATSYA LLALLQLKDF DFVPPVVRWL NEQRYYGGGY GSTQATFMVF QALAQYQKDA

      1270       1280       1290       1300       1310       1320
PDHQELNLDV SLQLPSRSSK ITHRIHWESA SLLRSEETKE NEGFTVTAEG KGQGTLSVVT

      1330       1340       1350       1360       1370       1380
MYHAKAKDQL TCNKFDLKVT IKPAPETEKR PQDAKNTMIL EICTRYRGDQ DATMSILDIS

      1390       1400       1410       1420       1430       1440
MMTGFAPDTD DLKQLANGVD RYISKYELDK AFSDRNTLII YLDKVSHSED DCLAFKVHQY

      1450       1460       1470       1480       1490       1500
FNVELIQPGA VKVYAYYNLE ESCTRFYHPE KEDGKLNKLC RDELCRCAEE NCFIQKSDDK

      1510       1520       1530       1540       1550       1560
```

```
VTLEERLDKA CEPGVDYVYK TRLVKVQLSN DFDEYIMAIE QTIKSGSDEV QVGQQRTFIS

      1570       1580       1590       1600       1610       1620
PIKCREALKL EEKKHYLMWG LSSDFWGEKP NLSYIIGKDT WVEHWPEEDE CQDEENQKQC

      1630
QDLGAFTESM VVFGCPN
```

**Molecular weight:** 184,326.5
**Theoretical pI:** 5.85
**Extinction Coefficient ($A_{280}$):** 180,550

## B5: Complement C3b

```
10         20         30         40         50         60
SPMYSIITPN ILRLESEETM VLEAHDAQGD VPVTVTVHDF PGKKLVLSSE KTVLTPATNH

        70         80         90        100        110        120
MGNVTFTIPA NREFKSEKGR NKFVTVQATF GTQVVEKVVL VSLQSGYLFI QTDKTIYTPG

       130        140        150        160        170        180
STVLYRIFTV NHKLLPVGRT VMVNIENPEG IPVKQDSLSS QNQLGVLPLS WDIPELVNMG

       190        200        210        220        230        240
QWKIRAYYEN SPQQVFSTEF EVKEYVLPSF EVIVEPTEKF YYIYNEKGLE VTITARFLYG

       250        260        270        280        290        300
KKVEGTAFVI FGIQDGEQRI SLPESLKRIP IEDGSGEVVL SRKVLLDGVQ NPRAEDLVGK

       310        320        330        340        350        360
SLYVSATVIL HSGSDMVQAE RSGIPIVTSP YQIHFTKTPK YFKPGMPFDL MVFVTNPDGS

       370        380        390        400        410        420
PAYRVPVAVQ GEDTVQSLTQ GDGVAKLSIN THPSQKPLSI TVRTKKQELS EAEQATRTMQ

       430        440        450        460        470        480
ALPYSTVGNS NNYLHLSVLR TELRPGETLN VNFLLRMDRA HEAKIRYYTY LIMNKGRLLK

       490        500        510        520        530        540
AGRQVREPGQ DLVVLPLSIT TDFIPSFRLV AYYTLIGASG QREVVADSVW VDVKDSCVGS

       550        560        570        580        590        600
LVVKSGQSED RQPVPGQQMT LKIEGDHGAR VVLVAVDKGV FVLNKKNKLT QSKIWDVVEK

       610        620        630        640        650        660
ADIGCTPGSG KDYAGVFSDA GLTFTSSSGQ QTAQRAELQC PQPAARRRRS NLDEDIIAEE

       670        680        690        700        710        720
NIVSRSEFPE SWLWNVEDLK EPPKNGISTK LMNIFLKDSI TTWEILAVSM SDKKGICVAD

       730        740        750        760        770        780
PFEVTVMQDF FIDLRLPYSV VRNEQVEIRA VLYNYRQNQE LKVRVELLHN PAFCSLATTK

       790        800        810        820        830        840
RRHQQTVTIP PKSSLSVPYV IVPLKTGLQE VEVKAAVYHH FISDGVRKSL KVVPEGIRMN
```

```
       850        860        870        880        890        900
KTVAVRTLDP ERLGREGVQK EDIPPADLSD QVPDTESETR ILLQGTPVAQ MTEDAVDAER

       910        920        930        940        950        960
LKHLIVTPSG CGEQNMIGMT PTVIAVHYLD ETEQWEKFGL EKRQGALELI KKGYTQQLAF

       970        980        990       1000       1010       1020
RQPSSAFAAF VKRAPSTWLT AYVVKVFSLA VNLIAIDSQV LCGAVKWLIL EKQKPDGVFQ

      1030       1040       1050       1060       1070       1080
EDAPVIHQEM IGGLRNNNEK DMALTAFVLI SLQEAKDICE EQVNSLPGSI TKAGDFLEAN

      1090       1100       1110       1120       1130       1140
YMNLQRSYTV AIAGYALAQM GRLKGPLLNK FLTTAKDKNR WEDPGKQLYN VEATSYALLA

      1150       1160       1170       1180       1190       1200
LLQLKDFDFV PPVVRWLNEQ RYYGGGYGST QATFMVFQAL AQYQKDAPDH QELNLDVSLQ

      1210       1220       1230       1240       1250       1260
LPSRSSKITH RIHWESASLL RSEETKENEG FTVTAEGKGQ GTLSVVTMYH AKAKDQLTCN

      1270       1280       1290       1300       1310       1320
KFDLKVTIKP APETEKRPQD AKNTMILEIC TRYRGDQDAT MSILDISMMT GFAPDTDDLK

      1330       1340       1350       1360       1370       1380
QLANGVDRYI SKYELDKAFS DRNTLIIYLD KVSHSEDDCL AFKVHQYFNV ELIQPGAVKV

      1390       1400       1410       1420       1430       1440
YAYYNLEESC TRFYHPEKED GKLNKLCRDE LCRCAEENCF IQKSDDKVTL EERLDKACEP

      1450       1460       1470       1480       1490       1500
GVDYVYKTRL VKVQLSNDFD EYIMAIEQTI KSGSDEVQVG QQRTFISPIK CREALKLEEK

      1510       1520       1530       1540       1550       1560
KHYLMWGLSS DFWGEKPNLS YIIGKDTWVE HWPEEDECQD EENQKQCQDL GAFTESMVVF


GCPN
```

**Molecular weight:** 176,774.7
**Theoretical pI:** 5.64
**Extinction Coefficient ($A_{280}$):** 176,700

**B6: 14-3-3 ς**

```
        10         20         30         40         50         60
MGSSHHHHHH SSGLVPRGSH MDKNELVQKA KLAEQAERYD DMAAAMKSVT EQGAELSNEE
        70         80         90        100        110        120
RNLLSVAYKN VVGARRSSWR VVSSIEQKTE GAEKKQQMAR EYREKIETEL RDIANDVLSL
       130        140        150        160        170        180
LEKFLIPNAS QAESKVFYLK MKGDYYRYLA EVAAGDDKKG IVDQSQQAYC CAFEISKKEM
       190        200        210        220        230        240
QPTHPIRLGL ALNFSVFYYE ILNSPEKAAS LAKTAFDEAI AELDTLSEES YKDSTLIMQL
       250
LRDNLTLWTS
```

**Molecular weight:** 28,300.9
**Theoretical pI:** 5.81
**Extinction Coefficient ($A_{280}$):** 27,390

**B7: Vps75**

```
        10         20         30         40         50         60
MSYYHHHHHH DYDIPTTENL YFQGAMADIG SMMSDQENEN EHAKAFLGLA KAEEEVDAIE
        70         80         90        100        110        120
REVELYRLNK MKPVYEKRDA YIDEIAEFWK IVLSQHVSFA NYIRASDFKY IDTIDKIKVE
       130        140        150        160        170        180
WLALESEMYD TRDFSITFHF HGIEGDFKEQ QVTKVFQIKK GKDDQEDGIL TSEPVPIEWP
       190        200        210        220        230        240
QSYDSINPDL IKDKRSPEGK KKYRQGMKTI FGWFRWTGLK PGKEFPHGDS LASLFSEEIY
       250                        260                        270
PFAVKYYAEA QRDLEDEEGE SGLSADGDSE G
```

**Molecular weight:** 31581.0
**Theoretical pI:** 4.82
**Extinction Coefficient ($A_{280}$):** 49,850

**B8: Myoglobin**

```
       10         20         30         40         50         60
MVLSEGEWQL VLHVWAKVEA DVAGHGQDIL IRLFKSHPET LEKFDRFKHL KTEAEMKASE
        70         80         90        100        110        120
DLKKHGVTVL TALGAILKKK GHHEAELKPL AQSHATKHKI PIKYLEFISE AIIHVLHSRH
       130                140                150                160
PGDFGADAQG AMNKALELFR KDIAAKYKEL GYQGGSGHHH HHH
```

**Molecular weight:** 18355.1
**Theoretical pI:** 8.70
**Extinction Coefficient ($A_{280}$):** 15,470 **($A_{409}$):** 125,000

**B9: Myosin Regulatory Light Chain**

```
        10         20         30         40         50         60
MAPKKAKKRA GGANSNVFSM FEQTQIQEFK EAFTIMDQNR DGFIDKNDLR DTFAALGRVN
        70         80         90        100        110        120
VKNEEIDEMI KEAPGPINFT VFLTMFGEKL KGADPEETIL NAFKVFDPEG KGVLKADYVR
       130                140                150                160
EMLTTQAERF SKEEVDQMFA AFPPDVTGNL DYKNLVHIIT GEEKD
```

**Molecular weight:** 18,652.1
**Theoretical pI:** 4.82
**Extinction Coefficient ($A_{280}$):** This protein does not contain any Trp residues.

As the RLC makes a complex with MiniHMM this is taken into consideration when calculating the Extinction Coefficient.

**B10: MiniHMM**

```
        10         20         30         40         50         60
MHHHHHHSTE NLYFQGSSER RDSLLVIQWN IRAFMGVKNW PWMKLYFKIK PLLKSAEREK
        70         80         90        100        110        120
EMASMKEEFT RLKEALEKSE ARRELEEKMV SLLQEKNDLQ LQVQAEQDNL ADAEERCDQL
       130        140        150        160        170
IKNKIQLEAK VKEMNERLED EEEMNAELTA KKRKLEDECS ELKRDIDDLE LTLAK
```

**Molecular weight:** 20,939.8
**Theoretical pI: 5.41**
**Extinction Coefficient ($A_{280}$):** 19,480

For the RLC/MiniHMM complex, which are co-expressed, the Extinction Coefficient ($A_{280}$) used is **20,970.** This is for half of the complex (one heavy chain + one rMLC). Molecular weight without of Histag is 37785.1.

**B11: cMyBP-C Domain C0**

```
        10         20         30         40         50         60
MHHHHHHSSG VDLGTENLYF QSMPEPGKKP VSAFSKKPRS VEVAAGSPAV FEAETERAGV
        70         80         90        100        110        120
KVRWQRGGSD ISASNKYGLA TEGTRHTLTV REVGPADQGS YAVIAGSSKV KFDLKVIEGS

TREEF
```

**Molecular weight:** 13,520.0
**Theoretical pI:** 8.07
**Extinction Coefficient ($A_{280}$):** 9,970

# Appendix C: Making Chemically Competent Cells

## For the preparation of Top 10, BL21, BL21 (*) and DH10β *E. coli* cells

**Table B.1 Media and Buffers for Cell Preparation**

| Media/Buffer | Composition |
|---|---|
| SOB | 1.25g yeast extract |
| | 5g tryptone |
| | 150mg NaCl |
| | 50mg KCl |
| | 60mg $MgSO_4$ |
| | ad 250mL $ddH_2O$ |
| | Adjust pH 7.5 with 1 M NaOH |
| | For agar add 15 g agar per 1000mL |
| SOC | As above supplemented with 20mM glucose |
| CCMB80 Buffer | 10 ml 1M KOAc pH 7.0 |
| | 80 mM $CaCl_2.2H_2O$ |
| | 20 mM $MnCl_2.4H_2O$ |
| | 10 mM MgCl2.6$H_2O$ |
| | 100 mL sterile glycerol |
| | adjusted to 1000mL with $ddH_2O$ |
| | and adjust pH 6.4 with 0.1 M HCl |
| | Sterile filter – store 4$^oC$ |

An aliquot of each cell line (purchased from Invitrogen Technologies) was used to streak some working material onto SOB agar and cells were grown at r.t to obtain single colonies. A single colony was used to inoculate 2ml of SOB-medium and the culture shaken overnight at r.t. To 250 mL of fresh SOB-medium, 1mL of the overnight culture was added and cells grown until an $OD_{600}$ of 0.3 was reached.

Cells were spun at 3000g at 4$^{o}$C for 10 min to pellet cells. Cells were resuspended in 80 mL of ice cold CCMB80 Buffer and incubated on ice for 20 min. Cells were spun again and resuspended in 10 ml of ice-cold CCMB80 Buffer. The $OD_{600}$ of 50$\mu$L of the resuspended cells mixed with 200$\mu$L of SOC was taken, and all cells diluted to yield a final $OD_{600}$ with this test of between 1.0 and 1.5. Cells were incubated on ice for 20 min and aliquoted into 50$\mu$L volumes and stored at -80$^{o}$C.

Efficiency of cells was tested by transforming cells with 1$\mu$L of standard pUC19 control plasmid (Life Technologies).

# Appendix D: Protein Sequence Analysis by Mass Spectrometry

### Sample Preparation for Trypsin Digest

Using standard methods, the protein sample being tested was run on SDS-PAGE in high enough quantity to produce a clean sharp band and the gel stained with GelCode Blue (Pierce ) for 1hr and de-stained in $ddH_2O$ for a further hour. Any tryptic mapping was done for the C3b protein sample which is shown as 2 separate bands on SDS PAGE.

Each band was excised separately and tightly avoiding excess acrylamide. Bands were incubated separately in 300$\mu$L of 200mM NH4HCO3 (ABC) in 50% Acetonitrile (ACN) at r.t for 30 min. This step was repeated twice, changing solutions in between to remove excess SDS. Gel pieces were the incubated in 300$\mu$L of 20mM DTT, 200mM ABC, 50% ACN at 32$^{o}$C for 1hr to reduce disulphides in the protein.

Three washes each with 300$\mu$L of 200mM ABC.50% CAN were carried out and 100$\mu$L of 50mM iodoacetamide , 200mM ABC.50%ACN added and left at r.t in the dark for 20 min. The gel pieces were then washed in 500$\mu$L of 20mM ABC.50%ACN three times and bands cut into pieces of approximately 2mm x 1mm in size. Gel pieces were spun at 13,000rpm for 2min and gel pieces covered in ACN, turning gel pieces white. ACN was decanted gel pieces allowed to dry.

To a vial of Promega mass spectrometry grade trypsin, 5*0*$\mu$L of 50mM ABC was added, and gel pieces swollen in a mix of 58$\mu$L 50mM ABC and 2$\mu$L Trypsin at 4$^{o}$C. After visible swelling of the gel pieces, the tubes were sealed with Nesco-film$^{TM}$ and incubated at 32$^{o}$C for 24hr. Samples were then sonicated using a sonicating water bath and given to the SIRCAMS facility to run. (University of Edinburgh).

## Appendix E: Forms in Support of Ethics Application PA10837

### E1: Participant Information Sheet

University of St Andrews

# Participant Information Sheet

**Project Title**

*Isolation and purification of complement protein C3b from human blood/plasma*

**What is the study about?**

*We invite you to participate in a research project where we extract human blood for the purpose of purifying proteins of interest (namely complement protein C3b) for use in structural analysis.*

*This study is being conducted as part of my PhD Thesis in the School of Physics & Astronomy.*

**Do I have to take Part?**

*This information sheet has been written to help you decide if you would like to take part. It is up to you and you alone whether or not to take part. If you do decide to take part you will be free to withdraw at any time without providing a reason.*

**What would I be required to do?**

*You will be invited to the medical school at the North Haugh where a trained individual will extract ~20ml of blood from you. This will be done in sterile conditions, similar to those used when extracting blood for medical purposes or for donation.*

## Will my participation be Anonymous and Confidential?

*Your participation will be entirely anonymous, as it will not be possible to identify any participants from their samples as all samples will be pooled, before purifying the desired proteins from the pool. Any data generated as a result will be used in PhD thesis and potential future publications but again I stress that your sample will not be identifiable in publication, or to the researcher working on the project.*

**Storage and Destruction of Data Collected**

*No 'data' as such is generated from the project. There will be basic assays carried out in order to assess the quality of protein purified and to determine whether we need to adjust our purification protocols. Any methods used in doing so will be published in future publication as well as included in the final PhD thesis. Again though, no information on any specific participant will be stored, as the desired end product is purified from a pooled blood sample from multiple participants.*

**What will happen to the results of the research study?**

*The results will be finalised by 2015 and written up as part of my PhD Thesis.*

**Are there any potential risks to taking part?**

*There are no potential risks in taking part, though the subject may experience brief discomfort during the blood extraction due to the use of needles. There are no risks involved with the procedure as all participants have blood taken from a trained individual. The technique is done in sterile conditions with new sharps used for each participant, and used sharps disposed of immediately and safely. Used sharps are then incinerated departmentally.*

## Questions

You will have the opportunity to ask any questions in relation to this project before completing a Consent Form.

**Consent and Approval**

This research proposal has been scrutinised and been granted Ethical Approval through the University ethical approval process.

**What should I do if I have concerns about this study?**

A full outline of the procedures governed by the University Teaching and Research Ethical Committee is available at http://www.st-andrews.ac.uk/utrec/Guidelines/complaints/

## Contact Details

*Researcher:* *Stacey Bell*

*Contact Details:* *sb258@st-andrews.ac.uk*

*Supervisor:* *Graham Smith*

*Contact Details:* *GMS@st-andrews.ac.uk* 01334 462669

**E2: Supervisor Statement in Support of Ethics Application PA10837**

(gms@st-andrews.ac.uk – Graham Smith)

Project : Isolation and Purification of Complement Protein C3b from human blood/plasma.

Researcher : Stacey Bell (sb258@st-andrews.ac.uk)
School of Physics and Astronomy

Having reviewed all aspects of this project and the associated risks involved, I can approve this application. Potential risks have been identified, and how these will be addressed is covered extensively throughout the submitted application. This particular protein cannot be made using recombinant methods, and so must be purified directly from source, which in this instance is human blood/plasma. This coupled with our interest in studying the human protein, means that there is no alternative other than purification from blood extracted from human participants.

Submitted with this application is the relevant documentation outlining the nature of the procedure to the participants, alongside relevant consent form. The participants are under no obligation to proceed, and are assured they can withdraw at any time. Blood is being taken by a trained individual from willing and informed donors, with volumes of blood not exceeding levels which could detrimentally affect the wellbeing of the participant. For these reasons I see no reason why this application should not be granted ethical approval.

**E3: Participant Consent Form**

# Participant Consent Form

# Anonymous Data

University of St Andrews

**Project Title**
*Isolation and Purification of Complement Protein C3b from human blood/plasma.*

**Researcher(s) Name(s)**
*Stacey Bell (sb258@st-andrews.ac.uk)*

**Supervisors Names**
*Janet E Lovett (janet.lovett@ed.ac.uk)*

*Graham Smith (gms@st-andrews.ac.uk)*

The University of St Andrews attaches high priority to the ethical conduct of research. We therefore ask you to consider the following points before signing this form. Your signature confirms that you are happy to participate in the study.

**What is Anonymous Data?**

The term 'Anonymous Data' refers to data collected by a researcher that has no identifier markers so that even the researcher cannot identify any participant. Consent is still required by the researcher, however no link between the participant's signed consent and the data collected can be made.

**Consent**

The purpose of this form is to ensure that you are willing to take part in this study and to let you understand what it entails. Signing this form does not commit you to anything you do not wish to do.

Material gathered during this research will be anonymous, so it is impossible to trace back to you. Blood will be taken from multiple participants and pooled together, and so it is impossible to identify any one participant from the pooled samples. After the desired protein has been isolated by multiple purification methods, it is necessary to assess the success of these purification protocols using basic lab techniques. Any data gathered from these

techniques will definitely be used outlining our methods in PhD theses, as well as potential usage in future publication. It is important to state however, that at no time is any one individual's sample identifiable from any other.

I have read and understood the information sheet.

I have been given the opportunity to ask questions about the study.

I have had my questions answered satisfactorily.

I understand that I can withdraw from the study without having to give an explanation.

I understand that my data once processed will be anonymous and that only the researcher(s) (and supervisors) will have access to the raw data which will be kept confidentially.

*I agree to my data (in line with conditions outlined above) being kept by the researcher and being archived and used for further research projects / by other bona fide researchers.*

I have been made fully aware of the potential risks associated with this research and am satisfied with the information provided.

I agree to take part in the study

Participation in this research is completely voluntary and your consent is required before you can participate in this research.

**Name in Block Capitals** _____

**Signature** _____

**Date** _____

# Appendix F: Record of Personal Development

**CONFERENCES ATTENDED**

- Annual Meeting of the St Andrews Centre of Magnetic Resonance (2012, 2013, 2014, 2015).
-Annual International Meeting of the ESR Spectroscopy Group of the Royal Society of Chemistry (Warwick 2012, Dundee 2014, Southampton 2015).
- IX$_{th}$ EF-EPR 2014 conference (Marseille 2014).

**PRESENTATIONS**

- Annual Meeting of the St Andrews Centre of Magnetic Resonance (2014, 2015 – poster presentations).
- Annual International Meeting of the ESR Spectroscopy Group of the Royal Society of Chemistry (Warwick 2012, Dundee 2014, Southampton 2015 – poster presentations).
- IX$_{th}$ EF-EPR 2014 conference (Marseille 2014 – poster presentation).
- University of Edinburgh, Departmental Talk – Using EPR Spectroscopy as a tool to study the protein – protein interactions in large biomacromolecular complexes (2011, 2012).
- University of St Andrews, Departmental Talk – Using EPR Spectroscopy as a tool to study the protein - protein interaction in large biomacromolecular complexes (2015).
- University of Edinburgh, School of Chemistry (organic section) conference, Firbush Outdoor Retreat Centre, (2013).

**PRIZES AWARDED**

- Annual Meeting of the St Andrews Centre of Magnetic Resonance, **2014 poster prize.**

*'Site-Directed Spin-Labelling : Cysteine Substitution Mutagenesis V.S. Genetic Code Expansion'*

**PUBLIC ENGAGEMENT & OUTREACH**

- Volunteer at the Edinburgh International Science Festival for two consecutive years, representing the University of Edinburgh, The Royal Society of Chemistry, and the BBC.
- Volunteer for the Solar Sparks, which involved visiting schools in Fife, teaching children from 4-11 about sustainability and renewable energy.

**SCIENCE WRITING & COMMUNICATION**

- Contributions to the University of St Andrews **'Sci@StAnd'** science publication.

**TEACHING & MENTORING**

- During my PhD, I have been fortunate to be employed by the University of Edinburgh as a part-time demonstrator to undergraduate students in both the schools of Biology and Chemistry.
- I have been responsible for the supervision and mentoring of three final year undergraduate students, each for a period of 4 months, in a research project on which their dissertations are based, and which contributes significantly to their final degree.
- I have also been the mentor to a visiting scholar at the University of St Andrews, from ETH in Zurich for a period of 6 months, introducing him to the field of molecular biology.

# References


Ababou, A., Gautel, M., & Pfuhl, M. (2007). Dissecting the N-terminal myosin binding site of human cardiac myosin-binding protein C: Structure and myosin binding of domain C2. *J. Biol. Chem.*, 23;282(12):9204-15.

Ababou, A., Rostkova, E., Mistry, S., Masurier, C. Le, Gautel, M., & Pfuhl, M. (2008). Myosin Binding Protein C Positioned to Play a Key Role in Regulation of Muscle Contraction: Structure and Interactions of Domain C1. *J. Mol. Biol.*, 384(3):615-30

Abdullin, D., Hagelueken, G., Hunter, R. I., Smith, G. M., & Schiemann, O. (2014). Geometric model-based fitting algorithm for orientation-selective PELDOR data. *Molecular Physics*, 113(6), 544–560.

Al-khayat, H. A. (2013). Review article Three-dimensional structure of the human myosin thick filament : clinical implications. *Glob. Cardiol. Sci. Pract*., 3: 280–302.

Amsterdam, I. M. C. Van, Ubbink, M., Canters, G. W., & Huber, M. (2003). Measurement of a Cu À Cu Distance of 26 ä by a Pulsed EPR Method. *Angew. Chem. Int. Ed.*, *42*(1), 62–64.

Aslam, M., & Perkins, S. J. (2001). Folded-back solution structure of monomeric factor H of human complement by synchrotron X-ray and neutron scattering, analytical ultracentrifugation and constrained molecular modelling. *J. Mol. Biol.*, 309(5), 1117–1138.

Bai, X., Yan, C., Yang, G., Lu, P., Ma, D., Sun, L., Shi, Y. (2015). An atomic structure of human γ-secretase. *Nature*. 10;525(7568):212-7

Baker, J. L., & Voth, G. a. (2013). Effects of ATP and Actin-filament binding on the dynamics of the myosin II S1 domain. *Biophys. J.*, 105(7), 1624–1634.

Banham, J. E., Baker, C. M., Ceola, S., Day, I. J., Grant, G. H., Groenen, E. J. J., Timmel, C. R. (2008). Distance measurements in the borderline region of applicability of CW EPR and DEER: A model study on a homologous series of spin-labelled peptides. *J. Magn. Reson.*, *191* (2), 202–218.

Banham, J. E., Timmel, C. R., Abbott, R. J. M., Lea, S. M., & Jeschke, G. (2006). The characterization of weak protein-protein interactions: Evidence from DEER for the trimerization of a von Willebrand Factor A domain in solution. *Angew. Chem. Int. Ed.*, 45, 1058–1061.

Barlow, P. N., Baron, M., Norman, D. G., Day, a J., Willis, a C., Sim, R. B., & Campbell, I. D. (1991). Secondary structure of a complement control protein module by two-dimensional 1H NMR. *Biochemistry*, 30(4), 997–1004.

Barlow, P. N., Steinkasserer, A., Norman, D. G., Kieffer, B., Wiles, A. P., Sim, R. B., & Campbell, I. D. (1993). Solution Structure of a Pair of Complement Modules by Nuclear Magnetic Resonance. *J. Mol. Bio.*, 323(1) 268–284.

Becker, C. F. W., Lausecker, K., Balog, M., Kálai, T., Hideg, K., Steinhoff, H. J., & Engelhard, M. (2005). Incorporation of spin-labelled amino acids into proteins. *Magn. Reson. Chem.*, *43* (SPEC. ISS.), 34–39.

Beier, C., & Steinhoff, H.-J. (2006). A structure-based simulation approach for electron paramagnetic resonance spectra using molecular and stochastic dynamics simulations. *Biophys. J.*, 91(7), 2647–2664.

Bhattacharjee, A., Lehtinen, M. J., Kajander, T., Goldman, A., & Jokiranta, T. S. (2010). Both domain 19 and domain 20 of factor H are involved in binding to complement C3b and C3d. *Mol. Immunol.*, 47(9), 1686–1691.

Blankenfeldt, W., Thomä, N. H., Wray, J. S., Gautel, M., & Schlichting, I. (2006). Crystal structures of human cardiac beta-myosin II S2-Delta provide insight into the functional role of the S2 subfragment. *Proc. Natl. Acad. Sci. U.S.A.*, 103(47), 17713–17717.

Blaum, B. S., Hannan, J. P., Herbert, A. P., Kavanagh, D., Uhrín, D., & Stehle, T. (2014). Structural basis for sialic acid–mediated self-recognition by complement factor H. *Nat. Chem. Biol.*, 11(1), 77–82.

Borbat, P. P., & Freed, J. H. (2013). Pulse Dipolar Electron Spin Resonance: Distance Measurements. *Struct. Bond*, 10(1007), 1–82.

Borbat, P. P., Surendhran, K., Bortolus, M., Zou, P., Freed, J. H., & Mchaourab, H. S. (2007). Conformational motion of the ABC transporter MsbA induced by ATP hydrolysis. *PLoS Biol.*, *5*(10), 2211–2219.

Bowman, A., Hammond, C. M., Stirling, A., Ward, R., Shang, W., El-Mkami, H.,Owen-Hughes, T. (2014). The histone chaperones Vps75 and Nap1 form ring-like, tetrameric structures in solution. *Nucleic Acids Res.*, 42(9), 6038–6051.

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.*, 72, 248–254.

Brown, L. J., Sale, K. L., Hills, R., Rouviere, C., Song, L., Zhang, X., & Fajer, P. G. (2002). Structure of the inhibitory region of troponin by site directed spin labeling electron paramagnetic resonance. *Proc. Natl. Acad. Sci. U.S.A.*, 99(20), 12765–12770.

Burns, J. A, Butler, J. C., Moran, J., & Whitesides, G. M. (1991). Selective reduction of disulfides by tris(2-carboxyethyl)phosphine. *J. Org. Chem.*, 56(8), 2648–2650.

Campbell, M. G., Smith, B. C., Potter, C. S., Carragher, B., & Marletta, M. a. (2014). Molecular architecture of mammalian nitric oxide synthases. *Proc. Natl. Acad. Sci. U.S.A.*, 111(35), E3614–E3623.

Carroll, M. C. (2004). The complement system in regulation of adaptive immunity. *Nat. Immunol.*, 5(10), 981–986.

Carroll, M. V., & Sim, R. B. (2011). Complement in health and disease. *Adv. Drug Deliv. Rev.*, 63(12), 965–975.

Chen, L. F., Winkler, H., Reedy, M. K., Reedy, M. C., & Taylor, K. a. (2002). Molecular modeling of averaged rigor crossbridges from tomograms of insect flight muscle. *J. Struct. Biol.*, 138(1-2), 92–104.

Chin, J. W. (2011). Reprogramming the genetic code. *EMBO J.*, 30(12), 2312–2324.

Chin, J. W., Martin, A. B., King, D. S., Wang, L., & Schultz, P. G. (2002). Addition of a photocrosslinking amino acid to the genetic code of Escherichia coli. *Proc. Natl. Acad. Sci. U.S.A.*, 99(17), 11020–11024.

Chin, J. W., Santoro, S. W., Martin, A. B., King, D. S., Wang, L., & Schultz, P. G. (2002). Addition of p -Azido- l -phenylalanine to the Genetic Code of Escherichia coli. *J. Am. Chem. Soc.*, 124(31), 9026–9027.

Clark, S. J., Ridge, L. a, Herbert, A. P., Hakobyan, S., Mulloy, B., Lennon, R., Day, A. J. (2013). Tissue-specific host recognition by complement factor H is mediated by differential activities of its glycosaminoglycan-binding regions. *J. Immunol.*, 190(5), 2049–57.

Cline, D. J., Redding, S. E., Brohawn, S. G., Psathas, J. N., Schneider, J. P., & Thorpe, C. (2004). New water-soluble phosphines as reductants of peptide and protein disulfide bonds: Reactivity and membrane permeability. *Biochemistry*, 43(48), 15195–15203.

Cole, M. A., Tully, S. E., Dodds, A. W., Arnold, J. N., Boldt, G. E., Sim, R. B., Wentworth, P. (2009). A chemical approach to immunoprotein engineering: Chemoselective functionalization of thioester proteins in their native state. *ChemBioChem.*, 10(8), 1340–1343.

Cornish, V. W., Benson, D. R., Altenbach, C. A, Hideg, K., Hubbell, W. L., & Schultz, P. G. (1994). Site-specific incorporation of biophysical probes into proteins. *Proc. Natl. Acad. Sci. U.S.A.*, 91(8), 2910–2914.

Craig, R., Lee, K.H., Mun,J.Y., Luther, P.K. (2014). Structure, sarcomeric organization and thin filament binding of cardiac myosin binding protein-C. *Pflugers Arch.*, 466(3), 425–431.

Cruickshank, P. A S., Bolton, D. R., Robertson, D. A., Hunter, R. I., Wylde, R. J., & Smith, G. M. (2009). A kilowatt pulsed 94 GHz electron paramagnetic resonance spectrometer with high concentration sensitivity, high instantaneous bandwidth, and low dead time. *Rev. Sci. Instrum.*, 80(10). 103102

Daff, S. (2010). NO synthase: Structures and mechanisms. *Nitric Oxide*, 23(1), 1–11.

De Córdoba, S. R., & De Jorge, E. G. (2008). Translational Mini-Review Series on Complement Factor H: Genetics and disease associations of human complement factor H. *Clin. Exp. Immunol*, 151(1), 1–13.

Deiters, A., & Schultz, P. G. (2005). In vivo incorporation of an alkyne into proteins in Escherichia coli. *Bioorgan. Med. Chem. Lett.*, 15(5), 1521–1524.

de Vera, I.M.S, Blackburn,M.E., Galiano, L., Fanucci, G.E. (2013). Pulsed EPR Distance Measurements in Soluble Proteins by Site-directed Spin-labeling (SDSL). *Curr Protoc Protein Sci.* 74,1-42

DiScipio, R. G. (1992). Ultrastructures and interactions of complement factors H and I. *J. Immunol.* 149(8), 2592–2599.

Dodds, A. W. (1993). Small-scale preparation of complement components C3 and C4. *Methods Enzymol.* 223, 46–61.

Dunkel, S., Pulagam, L. P., Steinhoff, H.-J., & Klare, J. P. (2015). In vivo EPR on spin labeled colicin A reveals an oligomeric assembly of the pore-forming domain in E. coli membranes. *Phys. Chem. Chem. Phys.* 17(7), 4875–4878.

El Mkami, H., Ward, R., Bowman, A., Owen-Hughes, T., & Norman, D. G. (2014). The spatial effect of protein deuteration on nitroxide spin-label relaxation: Implications for EPR distance measurement. *J. Magn. Reson.*, 248, 36–41.

Endeward, B., Butterwick, J. a., MacKinnon, R., & Prisner, T. F. (2009). Pulsed electron-electron double-resonance determination of spin-label distances and orientations on the tetrameric potassium ion channel KcsA. *J. Am. Chem. Soc.*, 131(42), 15246–15250.

Ezhevskaya, M., Bordignon, E., Polyhach, Y., Moens, L., Dewilde, S., Jeschke, G., & Van Doorslaer, S. (2013). Distance determination between low-spin ferric haem and nitroxide spin label using DEER: the neuroglobin case. *Molecular Physics, 111(18-19), 2855–2864.*

Ferreira, V. P., Herbert, A. P., Hocking, H. G., Barlow, P. N., & Pangburn, M. K. (2013). Critical role of the C-terminal domains of factor H in regulating complement activation at cell surfaces. *J. Immunol.*, 1;177(9):6308-16.

Fielding, A., Concilio, M., Heaven, G., & Hollas, M. (2014). New Developments in Spin Labels for Pulsed Dipolar EPR. *Molecules*, 19(10), 16998–17025.

Finkelstein, J. (2009). Metalloproteins. *Nature*, 460(7257), 813.

Fleissner, M. R., Bridges, M. D., Brooks, E. K., Cascio, D., Kalai, T., Hideg, K., & Hubbell, W. L. (2011). Structure and dynamics of a conformationally constrained nitroxide side chain and applications in EPR spectroscopy. *Proc. Natl. Acad. Sci. U.S.A.*, 108(39), 16241–16246.

Fleissner, M. R., Brustad, E. M., Kálai, T., Altenbach, C., Cascio, D., Peters, F. B., Hubbell, W. L. (2009). Site-directed spin labeling of a genetically encoded unnatural amino acid. *Proc. Natl. Acad. Sci. U.S.A.*, 106(51), 21637–21642.

Ford, J. E. (2000). Muscle Physiology and Cardiac Function. *Journal of Muscle Research and Cell Motility*. 22 (561)

Fritzinger, D. C., Hew, B. E., Thorne, M., Pangburn, M. K., Janssen, B. J. C., Gros, P., & Vogel, C. W. (2009). Functional characterization of human C3/cobra venom factor hybrid proteins for therapeutic complement depletion. *Dev. Comp. Immunol.*, 33(1), 105–116.

Garrigos, M., Mallam, S., Vachette, P., & Bordas, J. (1992). Structure of the myosin head in solution and the effect of light chain 2 removal. *Biophys. J.*, 63(6), 1462–1470.

Gautier, A., Nguyen, D. P., Lusic, H., An, W., Deiters, A., & Chin, J. W. (2010). Genetically encoded photocontrol of protein localization in mammalian cells. . *J. Am. Chem. Soc.*, 2010, 132 (12), pp 4086–4088

Getz, E. B., Xiao, M., Chakrabarty, T., Cooke, R., & Selvin, P. R. (1999). A comparison between the sulfhydryl reductants tris(2-carboxyethyl)phosphine and dithiothreitol for use in protein biochemistry. *Anal. Biochem.*, 273(1), 73–80.

Gigli, I., Fujita, T., & Nussenzweig, V. (1979). Modulation of the classical pathway C3 convertase by plasma proteins C4 binding protein and C3b inactivator. *Proc. Natl. Acad. Sci. U.S.A.*, 76(12), 6596–6600.

Gingras, A. R., Vogel, K., Steinhoff, H., Ziegler, W. H., Patel, B., Emsley, J., Barsukov, I. L. (2006). Structural and Dynamic Characterization of a Vinculin Binding Site in the Talin Rod Structural and Dynamic Characterization of a Vinculin Binding Site in the Talin, *Biochemistry*, 2006, 45 (6), pp 1805–1817.

Gophane, D. B., Endeward, B., Prisner, T. F., & Sigurdsson, S. T. (2014). Conformationally Restricted Isoindoline-Derived Spin Labels in Duplex DNA: Distances and Rotational Flexibility by Pulsed Electron-Electron Double Resonance Spectroscopy. *Chemistry*, 20(48), 15913–15919.

Gordon, D. L., Kaufman, R. M., Blackmore, T. K., Kwong, J., & Lublin, D. M. (1995). Identification of complement regulatory domains in human factor H. *J. Immunol.155*(1), 348–356.

Granzier, H. L., & Campbell, K. B. (2006). New insights in the role of cardiac myosin binding protein C as a regulator of cardiac contractility. *Circ. Res.*, 99(8), 795–797.

Graslund. (2008). Protein production and purification. *Nat. Methods*, 5(2), 135–146.

Greiss, S., & Chin, J. W. (2011). Expanding the Genetic Code of an Animal. *J. Am. Chem. Soc.* 133 (36), pp 14196–14199.

Gros, P. (2011). In self-defense. *Nat. Struct. Mol. Biol.*, 18(4), 401–402.

Gros, P., Milder, F. J., & Janssen, B. J. C. (2008). Complement driven by conformational changes. *Nat. Rev. Immunol.*, *8*(December 2007), 48–58.

Gruen, M., & Gautel, M. (1999). Mutations in beta-myosin S2 that cause familial hypertrophic cardiomyopathy (FHC) abolish the interaction with the regulatory domain of myosin-binding protein-C. *J. Mol. Biol.*, 286(3), 933–949.

Gurd FR, Falk KE, Malmström BG, Vänngård T.Tore, V. (1967). A Magnetic Resonance Study of Sperm and Its Complex with 1 Cupric Ion * Whale Ferrimyoglobin, *J. Biol. Chem.*, 242(24), 5724–5731.

Hagelueken, G., Ward, R., Naismith, J. H., & Schiemann, O. (2012). MtsslWizard: In Silico Spin-Labeling and Generation of Distance Distributions in PyMOL. *Appl. Magn. Reson.*, 42(3), 377–391.

Hahn, A., Reschke, S., Leimkühler, S., & Risse, T. (2014). Ketoxime coupling of p -acetylphenylalanine at neutral ph for site-directed spin labeling of human sulfite oxidase. *J. Phys. Chem. B.*, 118(25), 7077–7084.

Hakobyan, S., Harris, C. L., Van Den Berg, C. W., Fernandez-Alonso, M. C., De Jorge, E. G., De Cordoba, S. R., Morgan, B. P. (2008). Complement factor H binds to denatured rather than to native pentameric C-reactive protein. *J. Biol. Chem.*, 283(45), 30451–30460.

Hancock, S. M., Uprety, R., Deiters, A., & Chin, J. W. (2010). Expanding the genetic code of yeast for incorporation of diverse unnatural amino acids via a pyrrolysyl-tRNA synthetase/tRNA pair. *J. Am. Chem. Soc.*, 132(42), 14819–14824.

Hanson, J., & Huxley, H. E. (1953). Structural basis of the cross-striations in muscle. *Nature*, 172(4377), 530–532.

Harris, S. P., Bartley, C. R., Hacker, T. A., McDonald, K. S., Douglas, P. S., Greaser, M. L., Moss, R. L. (2002). Hypertrophic cardiomyopathy in cardiac myosin binding protein-C knockout mice. *Circ. Res.*, 90(5), 594–601.

Harris, S. P., Rostkova, E., Gautel, M., & Moss, R. L. (2004). Binding of myosin binding protein-C to myosin subfragment S2 affects contractility independent of a tether mechanism. *Circ. Res.*, *95*(9), 930–936.

Hemminga, M.A., Berliner, L.J. (2007). *ESR Spectroscopy in Membrane Biophysics*. London: Springer. 129-164.

Hilger, D., Polyhach, Y., Jung, H., & Jeschke, G. (2009). Backbone structure of transmembrane domain IX of the Na+ /proline transporter PutP of Escherichia coli. *Biophys. J.*, 96(1), 217–225.

Ho, G., & Chisholm, R. L. (1997). Substitution mutations in the myosin essential light chain lead to reduced actin-activated ATPase activity despite stoichiometric binding to the heavy chain. *J. Biol. Chem.*, *272(7), 4522–4527.*

Holm, L., Ackland, G. L., Edwards, M. R., Breckenridge, R. A., Sim, R. B., & Offer, J. (2012). Chemical labelling of active serum thioester proteins for quantification. *Immunobiology*, 217(2), 256–264.

Holmes, M. A., Geeves, K. C. (2000). Structural Mechanism of Muscle Contraction. *Annual Review of Biochemistry*, 68, 687–728.

Hopkins, S. C., Sabido-David, C., Corrie, J. E., Irving, M., & Goldman, Y. E. (1998). Fluorescence polarization transients from rhodamine isomers on the myosin regulatory light chain in skeletal muscle fibers. *Biophys. J.*, 74(6), 3093–3110.

Howarth, J. W., Ramisetti, S., Nolan, K., Sadayappan, S., & Rosevear, P. R. (2012). Structural insight into unique cardiac myosin-binding protein-C motif: A partially folded domain. *J. Biol. Chem.*, 287(11), 8254–8262.

Huang, C., Yin, Q., Zhu, W., Yang, Y., Wang, X., Qian, X., & Xu, Y. (2011). Highly selective fluorescent probe for vicinal-dithiol-containing proteins and in situ imaging in living cells. *Angew Chem. Int. Ed. Engl.* 50(33), 7551–7556.

Hubbell WL, Mchaourab HS, Altenbach C, Lietzow MA. (1996). Watching proteins move using site-directed spin labeling. *Nature*, 4(7):779–783.

Huber, M., Lindgren, M., Hammarström, P., Mårtensson, L. G., Carlsson, U., Eaton, G. R., & Eaton, S. S. (2001). Phase memory relaxation times of spin labels in human carbonic anhydrase II: Pulsed EPR to determine spin label location. *Biophys. Chem.*, 94(3), 245–256.

Hustedt, E. J., Smirnov, a I., Laub, C. F., Cobb, C. E., & Beth, A H. (1997). Molecular distances from dipolar coupled spin-labels: the global analysis of multifrequency continuous wave electron paramagnetic resonance data. *Biophys. J.*, 72(4), 1861–1877.

Huxley, H. E. (1957). The double array of filaments in cross-striated muscle. *J. Biophys. Biochem. Cytol.*, 3(5), 631–648.

Idowu, S. M., Gautel, M., Perkins, S. J., & Pfuhl, M. (2003). Structure, stability and dynamics of the central domain of cardiac myosin binding protein C (MyBP-C): Implications for multidomain assembly and causes for cardiomyopathy. *J. Mol. Biol.*, 329(4), 745–761.

Isenman, D. E., Kells, D. I., Cooper, N. R., Müller-Eberhard, H. J., & Pangburn, M. K. (1981). Nucleophilic modification of human complement protein C3: correlation of conformational changes with acquisition of C3b-like functional properties. *Biochemistry*, 20(1980), 4458–4467.

Jagtap, a. P., Krstic, I., Kunjir, N. C., Hänsel, R., Prisner, T. F., & Sigurdsson, S. T. (2015). Sterically shielded spin labels for in-cell EPR spectroscopy: Analysis of stability in reducing environment. *Free Radic. Res.*, 49(1), 78–85.

Janssen, B. J. C., Christodoulidou, A., McCarthy, A., Lambris, J. D., & Gros, P. (2006). Structure of C3b reveals conformational changes that underlie complement activity. *Nature*, 444(November), 213–216.

Janssen, B. J. C., & Gros, P. (2007). Structural insights into the central complement component C3. *Mol. Immunol.*, 44, 3–10.

Janssen, B. J. C., Huizinga, E. G., Raaijmakers, H. C. A, Roos, A., Daha, M. R., Nilsson-Ekdahl, K., Gros, P. (2005). Structures of complement component C3 provide insights into the function and evolution of immunity. *Nature*, 437(7058), 505–511.

Jeschke, G. (2006). DeerAnalysis 2006 User ' s Manual, *Springer*, Heidelberg, 1–30.

Jeschke, G. (2012). DEER Distance Measurements on Proteins. *Annu. Rev. Phys. Chem.*, 63, 419–446.

Johnson, D. B. F., Xu, J., Shen, Z., Takimoto, J. K., Schultz, M. D., Schmitz, R. J., Wang, L. (2011). RF1 knockout allows ribosomal incorporation of unnatural amino acids at multiple sites. *Nat. Chem. Biol.*, 7(september), 779–786.

Joseph, B., Jeschke, G., Goetz, B. a., Locher, K. P., & Bordignon, E. (2011). Transmembrane gate movements in the type II ATP-binding cassette (ABC) importer BtuCD-F during nucleotide cycle. *J. Biol. Chem.*, 286(47), 41008–41017.

Kajander, T., Lehtinen, M. J., Hyvärinen, S., Bhattacharjee, A., Leung, E., Isenman, D. E., Jokiranta, T. S. (2011). Dual interaction of factor H with C3d and glycosaminoglycans in host-nonhost discrimination by complement. *Proc Natl Acad Sci U S A.*, 108(7), 2897–2902.

Kanda, T., Takai, K., Hohsaka, T., Sisido, M., & Takaku, H. (2000). Sense codon-dependent introduction of unnatural amino acids into multiple sites of a protein. *Biochem. Biophys. Res. Commun.*, 270(3), 1136–1139.

Kendrew, J. C., Bodo, G., Dintzis, H. M., Parrish, R. G., Wyckoff, H., & Phillips, D. C. (1958). A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature*, 181(4610), 662–666.

Kinoshita, Y., Yamada, K. I., Yamasaki, T., Mito, F., Yamato, M., Kosem, N., Utsumi, H. (2010). In vivo evaluation of novel nitroxyl radicals with reduction stability. *Free Radic. Biol. Med.*, 49(11), 1703–1709.

Kolb, H. C., Finn, M. G., & Sharpless, K. B. (2001). Click Chemistry：Diverse Chemical Function from a Few Good Reactions. *Angew Chem Int Ed Engl.* 1;40 (11):2004-2021.

Kunjir, N. C., Reginsson, G. W., Schiemann, O., & Sigurdsson, S. T. (2013). Measurements of short distances between trityl spin labels with CW EPR, DQC and PELDOR. *Phys. Chem. Chem. Phys.*, 15(45), 19673–85.

Kunst, G., Kress, K. R., Gruen, M., Uttenweiler, D., Gautel, M., & Fink, R. H. (2000). Myosin binding protein C, a phosphorylation-dependent force regulator in muscle that controls the attachment of myosin heads by its interaction with myosin S2. *Circ. Res.*, 86(1), 51–58.

Lang, K., & Chin, J. W. (2014). Cellular incorporation of unnatural amino acids and bioorthogonal labeling of proteins. *Chem. Rev.*, 114(9), 4764–806.

Langford-Smith, A., Day, A. J., Bishop, P. N., & Clark, S. J. (2015). Complementing the Sugar Code: Role of GAGs and Sialic Acid in Complement Regulation. *Front. Immunol.*, 6(February), 1–7.

Law, S. K., & Dodds, a W. (1997b). The internal thioester and the covalent binding properties of the complement proteins C3 and C4. *Protein Sci,*. 6, 263–274.

Levitt, M. H. (2011). Short perspective on “nMR population inversion using a composite pulse” by M.H. Levitt and R. Freeman [J. Magn. Reson. 33 (1979) 473-476]. *J. Magn. Reson.*, 213(2), 274–275.

Liszewski K. & Atkinson, J. P. (2015). Complement regulators in human disease: lessons from modern genetics. *J. Intern. Med.*, 277(3), 294–305.

Longhi, S., Belle, V., Fournel, A., Guigliarelli, B., & Carrière, F. (2011). Probing structural transitions in both structured and disordered proteins using site-directed spin-labeling EPR spectroscopy. *J. Pept. Sci.*, 315–328.

Loscha, K. V., Herlt, A. J., Qi, R., Huber, T., Ozawa, K., & Otting, G. (2012). Multiple-site labeling of proteins with unnatural amino acids. *Angew Chem Int Ed Engl.* 51(9), 2243–2246.

Lovett, J. E., Abbott, R. J. M., Roversi, P., Caesar, J. J. E., Doria, M., Jeschke, G., Lea, S. M. (2013). Investigating the structure of the factor B vWF-A domain/CD55 protein–protein complex using DEER spectroscopy: successes and pitfalls. *Molecular Physics*, 111(18-19), 2865–2872.

Lovett, J. E., Lovett, B. W., & Harmer, J. (2012). DEER-Stitch: Combining three- and four-pulse DEER measurements for high sensitivity, deadtime free data. *J. Magn. Reson.*, 223, 98–106.

Lu, Y., Yeung, N., Sieracki, N., & Marshall, N. M. (2009). Design of functional Metalloproteins, *Nature*, 460(7257), 855–862.

Luther, P., & Squire, J. (2014). The Intriguing Dual Lattices of the Myosin Filaments in Vertebrate Striated Muscles: Evolution and Advantage. *Biology*, 3(4), 846–865.

Macauley-Patrick, S., Fazenda, M. L., McNeil, B., & Harvey, L. M. (2005). Heterologous protein production using the Pichia pastoris expression system. *Yeast*, 22(4), 249–270.

Makou, E. Herbert, A. P., & Barlow, P. N. (2013). Functional Anatomy of Complement Factor H. *Biochemistry*. 11;52(23):3949-62.

Makou, E., Matis, I., Schmidt, C. Q., Mertens, H. D. T., Svergun, D. I., & Barlow, P. N. (2010). High-resolution structures of module pairs 10-11, 11-12, and 12-13 combined with SAXS data for 10-15 and 10-18 constructs to define architecture of factor H. *Mol. Immunol.*, 47, 295–312.

Makou, E., Mertens, H. D. T., MacIejewski, M., Soares, D. C., Matis, I., Schmidt, C. Q., Barlow, P. N. (2012). Solution structure of CCP modules 10-12 illuminates functional architecture of the complement regulator, factor H. *J. Mol. Biol.*, 424(5), 295–312.

Malito, E., Carfi, A., & Bottomley, M. (2015). Protein Crystallography in Vaccine Research and Development. *Int. J. Mol. Sci.*, 16(6), 13106–13140.

Margossian, S. S. (1985). Reversible dissociation of dog cardiac myosin regulatory light chain 2 and its influence on ATP hydrolysis. *J. Biol. Chem.*, 260(25), 13747–13754.

Markiewski, M. M., & Lambris, J. D. (2007). The role of complement in inflammatory diseases from behind the scenes into the spotlight. *Am. J. Pathol*, 171(3), 715–727.

Martorana, A., Bellapadrona, G., Feintuch, A., Gregorio, E. Di, Aime, S., & Goldfarb, D. (2014). Probing Protein Conformation in Cells by EPR Distance Measurements using Gd 3+ Spin Labeling. *J. Am. Chem. Soc*. 136(38):13458-65.

Maslennikov, I., & Choe, S. (2013). Advances in NMR structures of integral membrane proteins. *Curr. Opin. Struct. Biol.*, 23(4), 555–562.

Merle, N. S., Church, S. E., Fremeaux-Bacchi, V., & Roumenina, L. T. (2015). Complement System Part I – Molecular Mechanisms of Activation and Regulation. *Front. Immunol*, 6, 1–30.

Merle, N. S., Noe, R., Halbwachs-Mecarelli, L., Fremeaux-Bacchi, V., & Roumenina, L. T. (2015). Complement System Part II: Role in Immunity. *Front. Immunol.*, *6(May), 1–26.*

Meyers, R. A. (1995). *Molecular Biology and Biotechnology. A Comprehensive Desktop Reference*. Wiley-VCH; 1 edition (June 29, 1995).

Mineev, K. S., Goncharuk, S. A., Kuzmichev, P. K., Vilar, M., & Arseniev, A. S. (2015). NMR Dynamics of Transmembrane and Intracellular Domains of p75NTR in Lipid-Protein Nanodiscs. *Biophys. J.*, 109(4), 772–782.

Morgan, H. P., Mertens, H. D. T., Guariento, M., Schmidt, C. Q., Soares, D. C., Svergun, D. I., Hannan, J. P. (2012). Structural analysis of the C-terminal region (modules 18-20) of complement regulator factor H (FH). *PLoS ONE*, 7(2), 19–24.

Morgan, H. P., Schmidt, C. Q., Guariento, M., Blaum, B. S., Gillespie, D., Herbert, A. P., Hannan, J. P. (2011). Structural basis for engagement by complement factor H of C3b on a self surface. *Nat. Struct. Mol. Biol.*, 18(4), 463–470.

Morita, F., Kondo, S., Tomari, K., Minowa, O., Ikura, M., & Hikichi, K. (1985). Calcium binding and conformation of regulatory light chains of smooth muscle myosin of scallop. *J. Biochem.*, 97(2), 553–561.

Mottagui-Tabar, S., & Isaksson, L. A. (1997). Only the last amino acids in the nascent peptide influence translation termination in Escherichia coli genes. *FEBS Lett.*, 414(1), 165–170.

Mullick, J., Bernet, J., Panse, Y., Hallihosur, S., Singh, A. K., & Sahu, A. (2005). Identification of complement regulatory domains in vaccinia virus complement control protein. *J. Virol.*, 79(19), 12382–12393.

Narr, E., Godt, A., & Jeschke, G. (2002). Selective measurements of a nitroxide-nitroxide separation of 5 nm and a nitroxide-copper separation of 2.5 nm in a terpyridine-based copper(II) complex by pulse EPR spectroscopy. *Angew Chem Int Ed Engl.*, 41(20), 3907–3910.

Netto, L. E., & Stadtman, E. R. (1996). The iron-catalyzed oxidation of dithiothreitol is a biphasic process: hydrogen peroxide is involved in the initiation of a free radical chain of reactions. *Arch. Biochem. Biophys.*, 333(1), 233–242.

Neumann, H., Wang, K., Davis, L., Garcia-Alai, M., & Chin, J. W. (2010). Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. *Nature*, 464(7287), 441–444.

Nguyen, D. P., Lusic, H., Neumann, H., Kapadnis, P. B., Deiters, A., & Chin, J. W. (2009). Genetic encoding and labeling of aliphatic azides and alkynes in recombinant proteins via a pyrrolysyl-tRNA synthetase/tRNACUA pair and click chemistry. *J. Am. Chem. Soc.*, 131(25), 8720–8721.

Nirenberg, M. W., & Matthaei, J. H. (1961). The dependence of cell-free protein synthesis in E. coli upon naturally occurring or synthetic polyribonucleotides. *Proc Natl Acad Sci U S A.*, 47(10), 1588–1602.

Oakley, C. E., Hambly, B. D., Curmi, P. M. G., & Brown, L. J. (2004). Myosin binding protein C: structural abnormalities in familial hypertrophic cardiomyopathy. *Cell Res.*, 14(2), 95–110.

Okemefuna, A. I., Nan, R., Gor, J., & Perkins, S. J. (2009). Electrostatic Interactions Contribute to the Folded-back Conformation of Wild Type Human Factor H. *J. Mol. Biol.*, 391(1), 98–118.

Padmavathi, P. V. L., & Steinhoff, H. J. (2008). Conformation of the Closed Channel State of Colicin A in Proteoliposomes: An Umbrella Model. *J. Mol. Biol.*, 378(1), 204–214.

Pangburn, M. K. (2000). Host recognition and target differentiation by factor H, a regulator of the alternative pathway of complement. *Immunopharmacology, 49(1-2), 149–157.*

Pangburn, M. K., Pangburn, K. L., Koistinen, V., Meri, S., & Sharma, A K. (2000). Molecular mechanisms of target recognition in an innate immune system: interactions among factor H, C3b, and target in the alternative pathway of human complement. *J. Immunol.*, 164(9), 4742–4751.

Pant, K., Watt, J., Greenberg, M., Jones, M., Szczesna-Cordary, D., & Moore, J. R. (2009). Removal of the cardiac myosin regulatory light chain increases isometric force production. *FASEB J.*, 23(10), 3571–3580.

Pechtl, I. C. (2010). Study of complement regulatory factor H based on Forster resonance energy transfer and investigation of disease-linked genetic variants. *Prot. Sci*.20(12): 2102–2112.

Pechtl, I. C., Kavanagh, D., Mcintosh, N., Harris, C. L., & Barlow, P. N. (2011). Disease-associated N-terminal complement factor H mutations perturb cofactor and decay-accelerating activities. *J. Biol. Chem.*, 286(13), 11082–11090.

Perkins, S. J., Nealis, A., & Sim, R. (1991). Oligomeric domain structure of human complement factor H by X-ray and neutron scattering. *Biochemistry*, 30(11), 2847–2857.

Pliotas, C., Ward, R., Branigan, E., Rasmussen, A., Hagelueken, G., Huang, H., Naismith, J. H. (2012). PNAS Plus: Conformational state of the MscS mechanosensitive channel in solution revealed by pulsed electron-electron double resonance (PELDOR) spectroscopy. *Proc Natl Acad Sci U S A.*, 109(40), E2675–E2682.

Polyhach, Y., Bordignon, E., & Jeschke, G. (2011). Rotamer libraries of spin labelled cysteines for protein studies. *Phys. Chem. Chem. Phys.*, 13(6), 2356–2366.

Pouw, R. B., Vredevoogd, D. W., Kuijpers, T. W., & Wouters, D. (2015). Of mice and men: The factor H protein family and complement regulation. *Mol. Immunol.*, 67(1), 12–20.

Prisner, T. F., Marko, a., & Sigurdsson, S. T. (2015). Conformational dynamics of nucleic acid molecules studied by PELDOR spectroscopy with rigid spin labels. *J. Magn. Reson.*, 252, 187–198.

Prosser, B. E., Johnson, S., Roversi, P., Herbert, A. P., Blaum, B. S., Tyrrell, J., Lea, S. M. (2007). Structural basis for complement factor H linked age-related macular degeneration. *J. Exp. Med.*, 204(10), 2277–2283.

Rabenstein, M. D., & Shin, Y. K. (1995). Determination of the distance between two spin labels attached to a macromolecule. *Proc Natl Acad Sci U S A.*, 92(18), 8239–8243.

Rakowsky, M. H., Zecevic, A, Eaton, G. R., & Eaton, S. S. (1998). Determination of high-spin iron(III)-nitroxyl distances in spin-labeled porphyrins by time-domain EPR. *J. Magn. Reson.* 131(1), 97–110.

Ratti, J., Rostkova, E., Gautel, M., & Pfuhl, M. (2011). Structure and interactions of myosin-binding protein C domain C0: Cardiac-specific regulation of myosin at its neck? *J. Biol. Chem.*, 286, 12650–12658.

Reginsson, G. W., Kunjir, N. C., Sigurdsson, S. T., & Schiemann, O. (2012). Trityl radicals: Spin labels for nanometer-distance measurements. *Chemistry*, 18(43), 13580–13584.

Reginsson, G. W., & Schiemann, O. (2011). Pulsed electron-electron double resonance: beyond nanometre distance measurements on biomacromolecules. *Biochem. J.*, 434, 353–363.

Ricklin, D., Hajishengallis, G., Yang, K., & Lambris, J. D. (2010). Complement: a key system for immune surveillance and homeostasis. *Nat. Immunol.* 11(9), 785–797.

Rodriguez, E., Rallapalli, P. M., Osborne, A. J., & Perkins, S. J. (2014). New functional and structural insights from updated mutational databases for complement factor H, Factor I, membrane cofactor protein and C3. *Biosci. Rep.*, 34(5), 635–649.

Root, D. D. (2002). The dance of actin and myosin: a structural and spectroscopic perspective. *Cell Biochem. Biophys.*, 37(2), 111–139.

Sadayappan, S., Gulick, J., Klevitsky, R., Lorenz, J. N., Sargent, M., Molkentin, J. D., & Robbins, J. (2009). Cardiac myosin binding protein-C phosphorylation in a myosin heavy chain background. *Circulation*, 119(9), 1253–1262.

Sahu, A, & Lambris, J. D. (2001). Structure and biology of complement protein C3, a connecting link between innate and acquired immunity. *Immunol. Rev.*, 180(4), 35–48.

Samygina, V. R., Sokolov, A. V., Bourenkov, G., Petoukhov, M. V., Pulina, M. O., Zakharova, E. T., Svergun, D. I. (2013). Ceruloplasmin: Macromolecular Assemblies with Iron-Containing Acute Phase Proteins. *PLoS ONE*, 8(7).

Schmidt, C. Q., Herbert, A. P., Kavanagh, D., Gandy, C., Fenton, C. J., Blaum, B. S., Barlow, P. N. (2008). A new map of glycosaminoglycan and C3b binding sites on factor H. *J. Immunol.*, 181(4), 2610–2619.

Schmidt, C. Q., Herbert, A. P., Mertens, H. D. T., Guariento, M., Soares, D. C., Uhrin, D., Barlow, P. N. (2010). The Central Portion of Factor H (Modules 10-15) Is Compact and Contains a Structurally Deviant CCP Module. *J. Mol. Biol.*, 395(1), 105–122.

Schmidt, C. Q., Slingsby, F. C., Richards, A., & Barlow, P. N. (2011). Production of biologically active complement factor H in therapeutically useful quantities. *Protein Expr.Purif.*, 76(2), 254–263.

Schmidt-Krey, I., & Rubinstein, J. L. (2011). Electron cryomicroscopy of membrane proteins: Specimen preparation for two-dimensional crystals and single particles. *Micron*, 42(2), 107–116.

Schreier, S., Bozelli, J. C., Marín, N., Vieira, R. F. F., & Nakaie, C. R. (2012). The spin label amino acid TOAC and its uses in studies of peptides: Chemical, physicochemical, spectroscopic, and conformational aspects. *Biophys. Rev.*, 4(1), 45–66.

Sedlak, J., & Lindsay, R. H. (1968). Estimation of total, protein-bound, and nonprotein sulfhydryl groups in tissue with Ellman's reagent. *Anal. Biochem.*, 25(1), 192–205.

Sharma, A K., & Pangburn, M. K. (1996). Identification of three physically and functionally distinct binding sites for C3b in human complement factor H by deletion mutagenesis. *Proc Natl Acad Sci U S A.*, 93 (October 1996), 10996–11001.

Shevelev, G. Y., Krumkacheva, O. A., Lomzov, A. A., Kuzhelev, A. A., Rogozhnikova, O. Y., Trukhin, D. V, Bagryanskaya, E. G. (2014) Physiological-temperature distance measurement in nucleic acid using triarylmethyl-based spin labels and pulsed dipolar EPR spectroscopy. *J. Am. Chem. Soc.*136(28): 9874-7

Short, G. F., Golovine, S. Y., & Hecht, S. M. (1999). Effects of release factor 1 on in vitro protein translation and the elaboration of proteins containing unnatural amino acids. *Biochemistry*, 38(27), 8808–8819.

Sim, R. B., & DiScipio, R. G. (1982). Purification and structural studies on the complement-system control protein beta 1H (Factor H). *Biochem. J.*, 205(2), 285–293.

Smith, M. E. B., Schumacher, F. F., Ryan, C. P., Tedaldi, L. M., Papaioannou, D., Waksman, G., Baker, J. R. (2010). Protein modification, bioconjugation, and disulfide bridging using bromomaleimides. *J. Am. Chem. Soc.*, 132(6), 1960–1965.

Sottrup-Jensen, L., Stepanik, T. M., Kristensen, T., Lønblad, P. B., Jones, C. M., Wierzbicki, D. M., Lundwall, a. (1985). Common evolutionary origin of alpha 2-macroglobulin and complement components C3 and C4. *Proc Natl Acad Sci U S A.*, 82(1), 9–13.

Steinhoff, H. J., Mollaaghababa, R., Altenbach, C., Hideg, K., Krebs, M., Khorana, H. G., & Hubbell, W. L. (1994). Time-resolved detection of structural changes during the photocycle of spin-labeled bacteriorhodopsin. *Science* 266(5182), 105–107.

Stone, T. J., Buckman, T., Nordiot, P. L., & Mcconnell, H. M. (1965). Spin-labeled biomolecules. *Proc. Natl. Acad Sci. U.S.A.*, 154(4):1010-7.

Szent-Györgyi, A. G. (2004). The early history of the biochemistry of muscle contraction. *J. Gen. Physiol.*, 123(6), 631–641.

Timmel C.R., Harmer, J.R. (2014). Structural Information from Spin-Labels and Intrinsic Paramagnetic Centres in the Biosciences. *London: Springer*. 1-329.

Tsvetkov, Y. D., & Grishin, Y. a. (2009). Techniques for EPR spectroscopy of pulsed electron double resonance (PELDOR): A review. *Instruments and Experimental Techniques*, 52(5), 615–636.

Van Doorslaer, S., Trandafir, F., Harmer, J. R., Moens, L., & Dewilde, S. (2014). {EPR} analysis of cyanide complexes of wild-type human neuroglobin and mutants in comparison to horse heart myoglobin. *Biophys. Chem.*, *190-{191C}*, 8–16.

Walden, M., Edwards, J. M., Dziewulska, A. M., Bergmann, R., Saalbach, G., Kan, S.-Y., Schwarz-Linek, U. (2015). An internal thioester in a pathogen surface protein mediates covalent host binding. *eLife*, 4, 1–24.

Walport, M. J. (2001). Complement: First of Two Parts. *N. Engl. J. Med.*, 344(14), 1058–1066.

Ward, R., Bowman, A., Sozudogru, E., El-Mkami, H., Owen-Hughes, T., & Norman, D. G. (2010). EPR distance measurements in deuterated proteins. *J. Magn. Reson.*, 207(1), 164–167.

Winegrad, S. (1999). Cardiac myosin binding protein C. *Circ. Res.*, 84, 1117–1126.

Witt, C. C., Gerull, B., Davies, M. J., Centner, T., Linke, W. A., & Thierfelder, L. (2001). Hypercontractile Properties of Cardiac Muscle Fibers in a Knock-in Mouse Model of Cardiac Myosin-binding Protein-C. *J. Biol. Chem.*, 276(7), 5353–5359.

Wu, J., Wu, Y., Ricklin, D., Janssen, B. J. C., & Lambris, J. D. (2010). Complement Regulators. *Structure*, 10(7), 728–733.

Wu, J., Wu, Y.-Q., Ricklin, D., Janssen, B. J. C., Lambris, J. D., & Gros, P. (2009). Structure of complement fragment C3b-factor H and implications for host protection by complement regulators. *Nat. Immunol.*, 10(7), 728–733.

Yang, Z., Becker, J., & Saxena, S. (2007). On Cu(II)-Cu(II) distance measurements using pulsed electron electron double resonance. *J. Magn. Reson.*, 188(2), 337–343.

Yang, X., Lee, W. H., Sobott, F., Papagrigoriou, E., Robinson, C. V, Gu, J., Elkins, J. M. (2006). Structural basis for protein – protein interactions in the 14-3-3 protein family, *Proc. Natl. Acad. Sci. U.S.A.* 103(46):17237-42.

Yang, Z., Liu, Y., Borbat, P., Zweier, J. L., Freed, J. H., & Hubbell, W. L. (2012). Pulsed ESR dipolar spectroscopy for distance measurements in immobilized spin labeled proteins in liquid solution. *J. Am. Chem. Soc.*, 134(24), 9950–9952.

Young, T. S., Ahmad, I., Brock, A., & Schultz, P. G. (2009). Expanding the genetic repertoire of the methylotrophic yeast Pichia pastoris. *Biochemistry*, 48(12), 2643–2653.

Young, T. S., Ahmad, I., Yin, J. A., & Schultz, P. G. (2010). An Enhanced System for Unnatural Amino Acid Mutagenesis in E. coli. *J. Mol. Biol.*, 395(2), 361–374.

Young, T. S., & Schultz, P. G. (2010). Beyond the canonical 20 amino acids: Expanding the genetic lexicon. *J. Biol.Chem.*, 285(15), 11039–11044.

Zhang, W., Hywood Potter, K. J., Plantz, B. A, Schlegel, V. L., Smith, L. A, & Meagher, M. M. (2003). Pichia pastoris fermentation with mixed-feeds of glycerol and methanol: growth kinetics and production improvement. *J. Ind. Microbiol. Biotechnol.*, 30, 210–215.

Zheng, H., Hou, J., Zimmerman, M. D., Wlodawer, A., & Minor, W. (2014). The future of crystallography in drug discovery. *Expert Opin. Drug Discov.*, 9(2), 125–37.