Structural and Functional Studies of Metabolic Enzymes and their Complexes

Henry J Bailey

University College



Supervisors: Assoc Prof Wyatt Yue, Dr Sean Froese, Dr Jolanta Kopec

A thesis submitted for the degree of *Doctor of Philosophy* in Clinical Medicine at the Structural Genomics Consortium

Nuffield Department of Medicine, University of Oxford

Declaration

I declare that no parts of this research or thesis have been submitted for another award, degree or diploma at any other university or learning institution. This thesis contains no other person's work, except where stated in text.

Acknowledgements

I would like to thank my supervisors Wyatt Yue, Sean Froese and Jolanta Kopec for guiding me through the last four years and providing a fantastic environment for learning.

Thank you to our collaborators in Zurich, Tanja Plessl, Sean Froese and Matthias R. Baumgartner for expertise in propionate catabolism pathways and our collaborators in the US, Jason R. Marcero and Harry A. Dailey for expertise in haem biosynthesis and protein kinetic assays. Many thanks to Gustavo Arruda Bezerra and William Foster in the lab for fragment screening experiments.

I would also like to say a massive thank you to the entire SGC particularly the MOB group past and present. The camaraderie in the group is unparalleled and I could not have wished for a more supportive and wonderful beer loving bunch of friends.

Thank you to Laura Diaz Saez for our combined cryo-EM screening and crochet sessions. Thank you to my lovely family. Finally, thank you to Fiona Sorrell both for teaching me how to use an X-ray beamline and for all the tap dance lessons and for eventually falling in love with me. Thanks for all the gin, wine, beer and other thesis writing essentials that you have provided and for the best 4 years of my life.

Abbreviations

A-SCS - ATP specific SCS isoform
AdoCbl - 5′-deoxyadenosylcobalamin
ALA - 5′-aminolevulinic acid
ALAS - 5′-aminolevulinate synthase
BC - biotin carboxylase
BCCP - biotin carboxylase carrier protein
Cbl - cobalamin
Cryo-EM - electron cryo microscopy
Ct - C-terminal
CT - carboxyl transferase
Dmax - longest scattering vector
DSG - di(N-succinimidyl) glutarate
ec - E. coli
FECH - ferrochelatase
FSC - Fourier shell correlation
G-SCS - GTP specific SCS isoform
GOF - gain of function disorder
hs - H. sapiens
I(0) - extrapolated intensities at zero scattering angle
LOF - loss of function
MCEE - methylmalonyl-CoA epimerase
MMA - methylmalonic aciduria
MMAA - methylmalonic aciduria type A protein
MMUT - methylmalonyl-CoA mutase
NDP - nucleotide di-phosphate
NDPK4 - nucleoside diphosphate kinase 4
Nt - N-terminal
PA - propionic aciduria
PCC - propionyl-CoA carboxylase
PLP - pyridoxal 5′-phosphate
PPIX - protoporphyrin IX
RBS - ribosomal binding site
rc - R. capsulatus
Rg - radios of gyration
S-CoA - succinyl-CoA
SAXS - small angle x-ray scattering
sc - S. cerevisiae

SCS - succinyl-CoA synthetase complex
SEC - size exclusion
ss - S. scrofa
SUCLA2 -
SUCLG1 -
SUCLG2 -
TCA - tricarboxylic acid
XLP - X-linked protoporphyria
XLSA - X-linked sideroblastic anaemia

Table of Contents

Declaration ***2***

Acknowledgements ***2***

Abbreviations ***3***

1 Abstract ***10***

2 Introduction ***11***

2.1 Thesis overview **11**
2.1.1 Enzyme pathways investigated 11
2.1.2 Protein-protein interactions in intermediate metabolism pathways 15
2.1.3 Overall Goal 17

2.2 Role of ALAS2 in haem biosynthesis **20**
2.2.1 ALAS catalytic mechanism 20
2.2.2 ALAS2 mutations in disease 22
2.2.3 Erythroid specific porphyrias 24
2.2.4 Wider interactome of ALAS2 25

2.3 The many roles of the succinyl-CoA synthetase complex **26**
2.3.1 Reaction mechanism and function 26
2.3.2 Two complex isoforms exist 28
2.3.3 The role of SCS in disease 30

2.4 Enzyme complexes in propionate metabolism **31**
2.4.1 The propionyl-CoA carboxylase complex 31
2.4.2 The MMUT-MMAA complex 33
2.4.3 The propionate catabolism metabolon 35
2.4.4 Propionate catabolism and disease 35

3 Methods ***36***

3.1 Cloning, mutagenesis **36**
3.1.1 hsALAS2 36
3.1.2 hsA-SCS 37
3.1.3 hsPCC 37
3.1.4 hsMMAA and hsMMUT 38

3.2 Protein expression and production procedure 38
3.2.1 Expression 38
3.2.2 Purification 38

3.3 hsALAS2 specific methods 39
3.3.1 Crystallisation 39
3.3.2 Data collection, Processing and refinement 39
3.3.3 SAXS solution studies 40
3.3.4 Differential scanning fluorimetry 41
3.3.5 Fluorescence intensity 41
3.3.6 Size exclusion co-elution 41
3.3.7 Cross-linking 42

3.4 SCS specific methods 42
3.4.1 Crystallisation 42
3.4.2 Data collection, Processing and refinement 42
3.4.3 SAXS solution studies 43
3.4.4 Cryo-EM grid making 43
3.4.5 EM data processing 44
3.4.6 EM model building and refinement 45
3.4.7 Activity assay 45

3.5 MMUT and MMAA specific methods 45
3.5.1 SAXS solution studies 45
3.5.2 GraFix 46
3.5.3 Cryo-EM grid making and screening 46

3.6 PCC specific methods 47
3.6.1 SAXS solution studies 47
3.6.2 Crystallisation optimisation 47
3.6.3 Cryo-EM grid making and data collection 48
3.6.4 EM data processing 48

4 Structure function study of hsALAS2 49

4.1 ALAS chapter introduction 49
4.1.1 Known ALAS structures 49
4.1.2 ALAS2 results chapter synopsis 53

4.2 Results 53
4.2.1 ALAS2 protein production and characterisation 53
4.2.2 Crystallisation and structure determination 57
4.2.3 hsALAS2 is a homodimer in crystal and in solution 60
4.2.4 ALAS2 structure and conserved features 63
4.2.5 hsALAS2 C-terminus is a self-inhibitory loop 66
4.2.6 Role of mobile active loop 71
4.2.7 hsALAS2 and scALAS have different Ct-extension 74
4.2.8 Revised understanding of disease-causing mutations of hsALAS2 76

4.3 ALAS2 fragment screening campaign reveals starting point for C-terminal stabilisation 78
4.3.1 Rational for the fragment screening approach 78
4.3.2 Analysis of ALAS2 fragment hits 80
4.3.3 Enzymatic studies 85

4.4 Conclusions 89

5 Structural and biophysical characterisation succinyl-CoA ligase Complex 90

5.1 Background to this study 90
5.1.1 Known structure function relationships of SCS 90
2.1.1 Unanswered questions and chapter synopsis 94

5.2 Complex Isolation and validation 95
5.2.1 Co-expression design of SUCLA2 and SUCLG1 95
5.2.2 Co-purification of SUCLA2 and SUCLG1 reveals stable complex formation 96

5.3 hsA-SCS structural studies 99
5.3.1 Crystal structure of hsA-SCS 99
5.3.2 Disease associated missense mutations disrupt the dimerisation interface 104

5.4 Oligomerisation of hsA-SCS 110
5.4.1 In solution oligomeric states as revealed SAXS studies 110
5.4.2 Ligand dependent oligomerisation determined by SAXS 117
5.4.3 Cryo-EM determination of the hsA-SCS hetero-octamer induced by succinyl-CoA 125
5.4.4 Intermediate assembly of the hetero-octamer 137

5.5 Interaction studies of succinyl-CoA ligase 142
5.5.1 Interaction studies of A-SCS with ALAS2 142

5.6 A-SCS results chapter summary 145

6 Structural studies of Propionate catabolism complexes 149

6.1 Introduction to MMUT-MMAA complex results 149
6.1.1 Current understanding of the MMUT-MMAA complex 149
6.1.2 Propionate catabolism results chapter synopsis: MMUT-MMAA 154

6.2 Structural studies of the MMUT-MMAA complex 154
6.2.1 Cloning and expression of MMAA-MMUT 154
6.2.2 SAXS analysis of MMAA-MMUT complex size and shape 154
6.2.3 Cryo-Electron microscopy studies of the MMAA and MMUT complex. 167
6.2.4 Modelling Filament formation 172

6.3 MMUT-MMAA results chapter summary 175

6.4 Introduction to PCC complex results 176
6.4.1 PCC structure function relationships 176
6.4.2 PCC results chapter synopsis 180

6.5 Structural studies of PCC complex 180
6.5.1 Cloning and expression of PCC complex 180
6.5.2 Purification and complex validation 181
6.5.3 PCC SAXS determination of complex size and shape 184
6.5.4 Probing conformational change of the PCC complex 187
6.5.5 Crystallography trials of PCC 190
6.5.6 Cryo-EM structure of PCC complex reveals conformational changes 192

6.6 PCC results chapter summary 200

7 Conclusion and Future Perspective 201

7.1 ALAS2 and haem biosynthesis 201
7.1.1 Significance of study 201
7.1.2 Limitations of work 201

7.1.3 Future experiments ... 203
7.2 Role of A-SCS in TCA cycle ... 205
7.2.1 Significance of study ... 205
7.2.2 Limitations of work ... 205
7.2.3 Future experiments ... 207
7.3 Complexes in propionate Catabolism ... 209
7.3.1 Significance of study ... 209
7.3.2 Limitations of data ... 209
7.3.3 Future experiments ... 210
7.4 Future perspective ... 212
Appendix 1: My publications ... 215
Appendix 2: PDB validation reports ... 216
References ... 216

List of Figures

Figure 2-1 Links between propionate catabolism, TCA cycle and haem biosynthesis pathways. ... 11
Figure 2-2: Haem biosynthesis pathway ... 14
Figure 2-3: Conserved reaction mechanism of ALAS enzymes. ... 22
Figure 2-4: Erythroid specific porphyria in haem biosynthetic pathway. ... 24
Figure 2-5: SCS reaction mechanism. ... 26
Figure 2-6 PCC reaction mechanism. ... 33
Figure 2-7: MMUT reaction mechanism. ... 34
Figure 4-1: Domain architecture and Structure-based sequence alignment of ALAS orthologues. ... 52
Figure 4-2: Expression of hsALAS2. ... 55
Figure 4-3: PLP loading of recombinant hsALAS2. ... 56
Figure 4-4: Crystal packing hsALAS2$_{\Delta N78}$ ... 59
Figure 4-5: Structure of hsALAS2. ... 61
Figure 4-6: Small angle X-ray scattering studies of hsALAS2$_{\Delta N142}$ and hsALAS2$_{\Delta N142-\Delta C545}$ ... 63
Figure 4-7: hsALAS2 ternary fold. ... 64
Figure 4-8: PLP binding site of hsALAS2. ... 66
Figure 4-9: Interactions of the Ct-extension with catalytic core. ... 68
Figure 4-10: C-terminal extension contacts mobile loop. ... 71
Figure 4-11: Role of C-terminal extension in structural and functional integrity: ... 73
Figure 4-12: comparison of hsALAS2 ct-extension with orthologue structures ... 75
Figure 4-13: Mapping of disease-causing missense mutations onto hsALAS2 structure. ... 77
Figure 4-14: hsALAS2 fragment binding pockets. ... 81
Figure 4-15: Cluster 1 of fragments around the C-terminal extension. ... 82
Figure 4-16: Activity of Progression curves for recombinant hsALAS2$_{\Delta N142}$ WT and R511E variant. ... 86

Figure 4-17: XLP causing variants of hsALAS2. 87
Figure 4-18 Enzyme activity of hsALAS2$_{\Delta N78}$ with fragments. 88
Figure 5-1: Enzyme pathways involving the succinyl-CoA synthetase complex. 91
Figure 5-2: Known quaternary states of SCS structures. 92
Figure 5-3: SCS 3 step reaction mechanism. 92
Figure 5-4: Active site annotation of ssG-SCS structure. 94
Figure 5-5: Bi-cistronic hsA-SCS expression vectors. 96
Figure 5-6: Purification of hsA-SCS. 97
Figure 5-7: Activity assay of hsA-SCS. 98
Figure 5-8: Crystal structure of hsA-SCS complex. 101
Figure 5-9: hsA-SCS Secondary structure elements. 102
Figure 5-10: Comparison of ATP grasp lid conformation with orthologue structures 103
Figure 5-11: Mutation analysis of hsA-SCS. 107
Figure 5-12: SUCLA2 interaction interface mutations. 109
Figure 5-13: Size exclusion chromatography coupled SAXS (SEC-SAXS) of as purified hsA-SCS. 111
Figure 5-14: SAXS Log10 intensity plots for hsA-SCS 112
Figure 5-15 Linear plots in the SAXS guinier region for hsA-SCS concentration series. 114
Figure 5-16: SAXS real-space P(r) distributions for the hsA-SCS concentration series. 116
Figure 5-17: Batch mode SAXS Log10 intensity plots of hsA-SCS with substrates. 119
Figure 5-18: SAXS P(r) distribution and MW calculation for A-SCS ligand bound states. 122
Figure 5-19: Ab initio SAXS models of A-SCS in presence of succinyl-CoA. 123
Figure 5-20: 2d classes hsA-SCS with succinyl-CoA: 126
Figure 5-21: Cryo-electron microscopy of hsA-SCS with succinyl-CoA: 127
Figure 5-22: Modelling into the EM and SAXS maps. 129
Figure 5-23: Two interaction interfaces of the A-SCS hetero-octamer 130
Figure 5-24 conformational change of the ATP grasp lid. 131
Figure 5-25: Conformational change of the C-terminal domain. 133
Figure 5-26: Sequence analysis at interaction interface 1 and 2. 135
Figure 5-27: Intermediate complex hsA-SCS assembly. 139
Figure 5-28: Nucleotide inhibition of tetramer one (interface one in the hetero-octamer) 141
Figure 5-29: investigating wider interactions of hsA-SCS. 145
Figure 5-30: Hypothesised model for A-SCS reaction. 147
Figure 6-1 Propionate catabolism pathway. 149
Figure 6-2: Structures of MMAA and MeaB GTPases. 150
Figure 6-3 Modelling Mutase MMAA interaction. 152
Figure 6-4: SEC-SAXS elution profiles 155
Figure 6-5: SAXS modelling of MMUT and MMAA homodimers. 157
Figure 6-6: SEC-SAXS elution profiles. 159
Figure 6-7: MW calculation for the SEC-SAXS isolated MMUT-MMAA complex. 161
Figure 6-8: Table of SAXS parameters. 162
Figure 6-9: SAXS modelling of peak 1 data with MMUT homodimer. 163
Figure 6-10: SAXS P(r) distance distribution. 166
Figure 6-11: Cryo-EM optimisation of MMAA-MMUT complex. 168

Figure 6-12: GraFix optimisation of MMAA-MMUT cryo-EM sample. 171
Figure 6-13: Model of the MMAA-MMUT filament based on extension of the IcmF structure. 173
Figure 6-14: Filament model fit to SAXS scattering data. 174
Figure 6-15: Filament formation as an alternative sequestering mechanism in the MMUT reactivation pathway. 176
Figure 6-16: Propionate catabolism pathway. 177
Figure 6-17: Structure of bacterial chimeric PCC holoenzyme: 179
Figure 6-18 Bi-cistronic hsPCC expression vectors. 180
Figure 6-19: Isolation of PCC complex. 183
Figure 6-20: Log10 intensity plot for the PCC complex. 185
Figure 6-21: Ab initio modelling of PCC SAXS envelope. 187
Figure 6-22: Probing conformational changes in PCC using SAXS. 189
Figure 6-23: PCC crystal degradation. 190
Figure 6-24: Cryo-EM 2D and 3D classifications of PCC complex. 194
Figure 6-25: Final cryo-EM map of PCC complex at 3.3 Å resolution. 196
Figure 6-26: Local resolution estimates for the PCC cryo-EM map. 197
Figure 6-27: symmetry expansion to deconvolute PCCA conformations. 200

List of Tables

Table 2-1: Thesis summary table. 19
Table 4-1: hsALAS2 Crystallographic table of parameters 60
Table 5-1: Crystallographic table of parameters for the hsA-SCS dataset. 100
Table 5-2 SAXS scattering table of parameters for hsA-SCS concentration range. 113
Table 5-3: SAXS table of parameters for A-SCS ligand bound states 124

1 Abstract

Enzymes in Intermediate metabolism participate in highly complex protein-protein interactions to carry out their cellular functions. To gain insight into these complexes, an integrative structural biology approach is adopted. Here the novel human structures of aminolevulinate synthase 2 (hsALAS2), involved in haem biosynthesis, and the ATP specific succinyl-CoA synthetase complex (hsA-SCS), involved in the tricarboxylic acid cycle are solved. In both cases the structures unveil novel enzymatic regulatory features important in the catalytic mechanism for the enzymes and provide insights into enzyme dysfunction in disease. The drugability of hsALAS2, for the treatment of haem related porphyria, is assessed with X-ray fragment screening campaigns and attempts are made to substantiate literature evidence of a quaternary complex between ALAS2 and A-SCS. In addition, the structures of protein complexes in the propionate catabolism pathway are pursued to understand conformational dynamics that underpin their respective reaction mechanisms. The data together builds our understanding of the complexity of protein-protein interactions and their proposed involvement in regulating metabolic flux and pathway cross talk.

2 Introduction

2.1 Thesis overview

2.1.1 Enzyme pathways investigated

This thesis is centred around the study of enzymes and their associated protein complexes involved in central metabolic processes within the mitochondria. These processes, termed 'Intermediary metabolism', involve reactions concerned with the storage and production of metabolic energy required for the biosynthesis of small molecules and energy storage compounds[1]. The study focuses on proteins found within three important pathways of intermediary metabolism, including haem biosynthesis, the tricarboxylic acid (TCA) cycle and propionate catabolism (Figure 2-1).

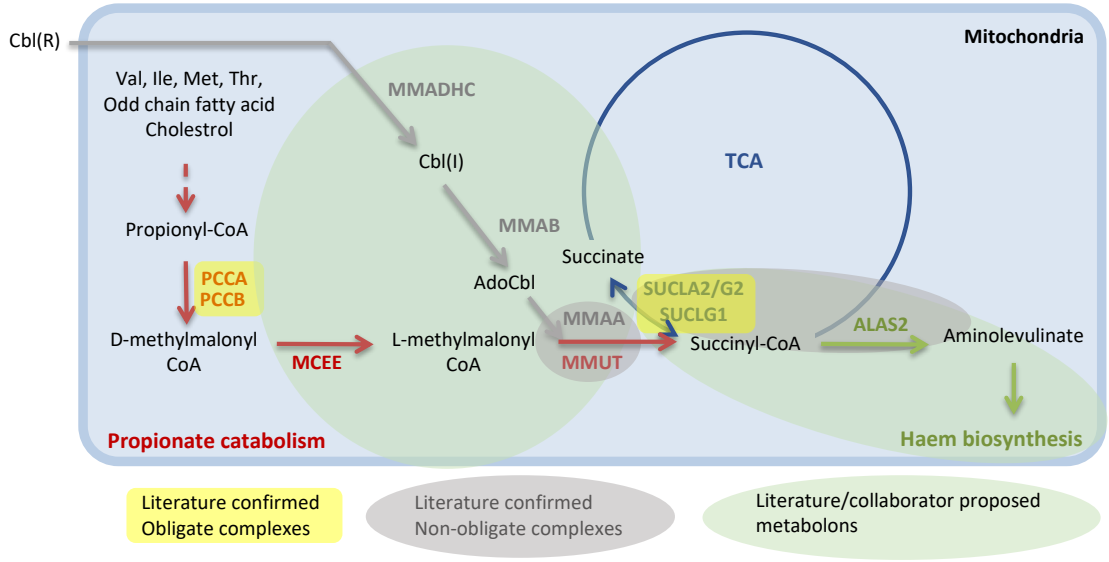


Figure 2-1 Links between propionate catabolism, TCA cycle and haem biosynthesis pathways. Propionate catabolism is displayed in red arrows, TCA cycle in blue arrows, haem biosynthesis in green arrows and the methylmalonyl-CoA mutase (MMUT) cofactor trafficking pathway in grey arrows. Yellow highlighted enzymes display obligate complexes, grey highlighted enzymes display non-obligate complexes and green highlighted enzymes reveal potential metabolons.

2.1.1.1 Propionate catabolism

Propionate catabolism occurs within the mitochondrial matrix. Its core function is to funnel a CoA derivatised propionate metabolite (propionyl-CoA), directly into the TCA cycle. Propionyl-CoA is the common product of the degradation of methionine, β-branched chain amino acids (valine, isoleucine and threonine), odd-chain fatty acids and the side chain of cholesterol. To allow entry into the TCA cycle, a three-step pathway has evolved converting propionyl-CoA into succinyl-CoA, which can be subsequently catabolised by succinyl-CoA synthetase for direct energy production (red pathway in Figure 2-1). A biotin dependent (vitamin B7) carboxylation reaction, carried out by propionyl-CoA carboxylase (PCC), catalyses the first step in the pathway converting the 3-carbon chain CoA derivative, propionyl-CoA into a 4-carbon chain derivative D-methylmalonyl-CoA. An epimerase reaction, catalysed by methylmalonyl-CoA epimerase (MCEE), subsequently converts D-methylmalonyl-CoA to its L-enantiomer, L-methylmalonyl-CoA. Finally, a second isomerisation catalysed by methylmalonyl-CoA mutase (MMUT) converts L-methylmalonyl-CoA to succinyl-CoA. The final isomerisation is dependent the cofactor vitamin B12 (cobalamin, Cbl). MMUT is one of two human proteins that require B12 for its catalytic reaction and a complex mitochondrial B12 trafficking (grey pathway in Figure 2-1) co-exists with the propionate catabolism pathway to provide the MMUT reaction with activated forms of Cbl.

2.1.1.2 The TCA cycle

The TCA cycle is core to central metabolism of nearly all life forms and is located within the mitochondria of eukaryotes. The primary function of the TCA cycle is to produce reducing equivalents for energy generation via oxidative phosphorylation. It

is formed by a series of enzyme reactions in a closed loop cycle which allows entry for multiple substrates. In this multi-step process, acetyl-CoA is oxidised to two molecules of CO_2 directly producing energy in the form of 1 ATP and additionally 3 NADH and 1 $FADH_2$ reducing molecules that can enter oxidative phosphorylation and further energy production. In addition to catabolism, the TCA cycle functions in anabolic processes, producing intermediate metabolites that are transported from the mitochondria to the cytosol where they can be synthesised into macromolecular structures. As reagents are exported out of the mitochondria, the cycle requires replenishment of substrates in order to continue energy production. Multiple anaplerotic pathways have evolved to carry out this function[2]. One such input pathway is the provision of succinyl-CoA by the propionate catabolism pathway to the succinyl-CoA synthetase complex (SCS). Within the TCA cycle, SCS subsequently converts succinyl-CoA to succinate, a reaction directly coupled to the production of ATP.

2.1.1.3 Role of Haem and its biosynthesis

Haem is a tetrapyrrole cofactor that functionalises a diverse range of haemoproteins, making it essential for many cellular processes across lifeforms. It is most well-known for its role as a co-factor for haemoglobin and myoglobin during oxygen transport and storage. However, haem also plays many other roles such as in sensing diatomic gases, detoxification of harmful molecules, regulation of translation and transcription, signalling pathways, cellular respiration, erythroblast maturation, processing microRNA and protein import into the mitochondria[3].

In metazoans, a conserved biosynthetic pathway consisting of 8 reactions, located within the mitochondria and cell cytoplasm, exists to convert iron, glycine, and succinyl-CoA into haem (Figure 2-2).

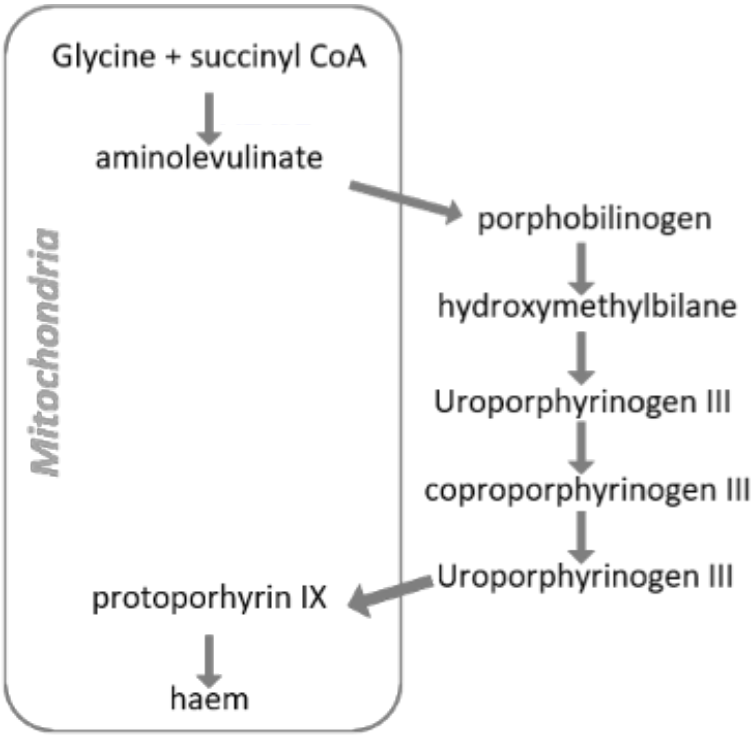


Figure 2-2: Haem biosynthesis pathway

Depiction of the 8 haem biosynthetic steps in the mitochondria and cytosol.

The enzyme 5′-aminolevulinate synthase (ALAS) catalyses the first and rate-limiting step of haem synthesis in the mitochondria[4,5]. ALA is subsequently exported from the mitochondria and the intermediate haem biosynthetic reactions are carried out in the cytoplasm, generating intermediate protoporphyrin forms (Figure 2-2). The final reaction takes place back in the mitochondria and requires ferrochelatase (FECH) to load iron into the newly synthesised tetrapyrrole scaffold known as protoporphyrin IX (PPIX) producing haem[6,].

2.1.2 Protein-protein interactions in intermediate metabolism pathways

Within the intermediate metabolic pathways described here proteins interact to form both obligate and non-obligate protein complexes. Obligate complexes are defined as permanent protein complexes where, for reasons of stability or evolution of functional co-dependence, the individual protein substituents of the complex do not exist in isolation. In contrast, substituents of non-obligate complexes can exist both independently and as hetero-oligomeric complexes. Their interactions tend to be lower in affinity to allow for their assembly and disassembly under specific spatial and temporal cellular conditions[7,8].

The PCC complex in propionate catabolism is an example of an obligate complex, where two separate proteins interact to catalyse separate parts of the reaction cycle. One polypeptide is required to couple ATP hydrolysis with the activation of the biotin cofactor via a carboxylation reaction whilst its obligate interaction partner is required to transfer the carboxyl group to the substrate propionyl-CoA (section 2.4.1) (highlighted yellow in Figure 2-1)[9,10]. SCS within the TCA cycle is another example of an obligate complex where one polypeptide is required to convert succinyl-CoA to succinate for the next step of the TCA cycle whilst its obligate interaction partner couples this to the production of ATP through substrate level phosphorylation (section 2.3.1) (highlighted yellow in Figure 2-1)[11,12].

The MMUT enzyme that catalyses the final step of the propionate catabolism pathway participates in non-obligate interactions with methylmalonic aciduria type A protein (MMAA)[13,14]. In this interaction, MMUT requires its co-chaperone MMAA to assist the reactivation of inactive MMUT (purple dashed circle Figure 2-1) (section

2.4.2) allowing for MMUT to dissociate from the complex and continue catalysis[15]. Literature reports[16,17] have also implicated ALAS enzymes to participate in non-obligate interactions with the TCA cycle enzyme SCS (purple dashed circle Figure 2-1) (section 2.2.4). However, to date strong biophysical evidence that supports the existence of this interaction is missing. The transient nature and requirement for specific cellular conditions of many non-obligate interactions increases the difficulty of their study. This has led to a long list of reported transient cellular complexes that have yet to be validated through stringent biophysical experiments, which has impeded the determination of their three-dimensional structure.

In addition to specific obligate and non-obligate complexes that exist in these pathways (yellow dashed circles Figure 2-1), there are further literature proposals of the existence of metabolons present within the pathways described here. In contrast to previously described complexes, metabolons are much larger enzyme-enzyme assemblies between sequential enzymes of a metabolic pathway[7,8,18]. They can be temporary in nature, requiring specific cellular conditions to form, but all act to physically channel metabolites between active sites of sequential enzymes in a pathway preventing diffusion with the bulk solvent[7,8]. One such example is the TCA cycle metabolon incorporating protein-protein interaction networks with sequential and non-sequential enzymes in the cycle shown by AP-MS[19]. Evidence for substrate channelling within this interaction network was also provided by isotope-based dilution[19].

A metabolon involving proteins of the propionate catabolism pathway (red pathway in Figure 2-1) with the Cbl cofactor trafficking pathway (grey pathway in Figure 2-1) (section 2.4.3)[20] and a separate metabolon involving enzymes in the haem biosynthetic pathway, encompassing the ALAS SCS interaction, (section 2.2.4) have

recently been proposed to exist in cells[21]. Due to their size, requirement of specific cellular environments and often transient nature they are very difficult to study, and to date there are only a few examples that successfully describe the true nature of these assemblies beyond their proposed existence in proteomics study.

2.1.3 Overall Goal

This thesis is focused on the structural composition of proteins at the individual or tandem level to gain mechanistic insights into how this relates to enzyme function, to understand the role of enzyme variants in disease and to provide resources for the future study or larger protein assemblies. As a starting point, the human isoforms of individual proteins within these pathways and their well-established obligate and non-obligate enzyme complexes were selected for *in vitro* structural and functional characterisation using recombinant techniques. These include erythroid specific 5′-aminolevulinic synthase 2 (ALAS2) within the haem biosynthesis pathway, the ATP specific succinyl-CoA synthetase complex (A-SCS) in the TCA cycle, and the propionyl-CoA carboxylase (PCC) and methylmalonyl-CoA mutase complexes (MMUT and MMAA complex) within propionate catabolism. Each protein or complex has already been previously characterised from a structure-function perspective, either through study of orthologues or paralogues. However, assumptions made from these homology studies have not yet fully explained the mechanistic details of these proteins. In each case, questions remain over how the enzymes work in concert within their complexes to adopt multiple conformations that allow them to catalyse individual reactions, the structural features that allow diversification of functional cellular roles between paralogues, and the structural features evolved to enable pathway regulation (Table 2-1). A complete understanding of these hitherto unknown mechanistic factors will serve to rationalise the role of mutations and enzyme

dysregulation in human disease. In addition to their individual study, the isolation of enzymes and their obligate and non-obligate complexes will produce reagents for the future study of larger metabolons.

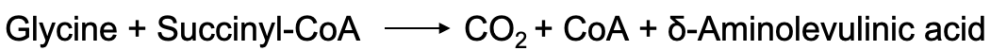
Protein	Pathway	Complex type	Evidence for interaction	Previous structures	Aims	Outcomes of this thesis
ALAS2	Haem-biosynthesis	NA	NA	*E.coli* and *S.cerevisiae homologue* structures solved by X-ray crystallography	To understand the structure and mechanism of the human (hs) ALAS2 C-terminal extension, evolutionarily distinct from lower eukaryotes and bacteria, in enzyme self regulation and porphyria disease.	1: Novel crystal structure of hsALAS2 reveals features of the Ct-extension that explain its role in both enzyme regulation and porphyria disorders. 2: Crystallography-based fragment screening campaign points towards future development of ALAS2 activity inhibitors as a therapy for porphyria disorders.
A-SCS	Tricarboxylic acid cycle	Obligate	High resolution orthologue and paralogue complex structures	*E.coli* and *S.scrofa* homologue complexes solved by X-ray crystallography	To determine the evolutionary features, evolved at protein level, that partition A-SCS and its related paralogue complex, G-SCS, in opposing cellular roles.	1: Novel crystal structure of hsA–SCS hetero-dimer. 2: Identification of in solution substrate induced oligomerisation 3: Novel hetero-octameric conformation of hsA-SCS solved by single particle cryo-electron microscopy in presence of succinyl-CoA 4: Coupling structures to intermediate reaction states of the enzyme complex to propose a ping-pong reaction mechanism.
A-SCS-ALAS2	Haem-biosynthesis?	Non-obligate	Affinity purification and affinity purification mass spectrometry	No complex structures	To isolate and characterise the proposed protein interaction between hsA-SCS and hsALAS2.	1: In-vitro isolation of the complex proved unsuccessful in this study
MMUT-MMAA	Propionate catabolism	Non-obligate	Size exclusion chromatography, Isothermal titration calorimetry, negative stain electron microscopy in human and orthologue complexes. Structure of an orthologue fusion complex (IcmF).	Individual human structures solved via X-ray crystallography. Negative stain electron microscopy 2D classes of the complex in various ligand states	To understand the structure of the human MMAA-MMUT complex in order to gain mechanistic details unpinning the exquisite metallocofactor trafficking and activation mechanism.	1: Novel identification of MMAA-MMUT filamentous oligomers revealed using small angle x-ray scattering and single particle cryo-electron microscopy.
PCC	Propionate catabolism	Obligate	Orthologue complex structures	Chimeric bacterial structure solved by X-ray crystallography	To gain structural insights into alternative catalytically relevant conformations previously unsolved in model systems	1: single particle cryo-electron microscopy reveals an alternative conformation of the BCCP domain engaging the PCCA active site.

Table 2-1: Thesis summary table.

2.2 Role of ALAS2 in haem biosynthesis

2.2.1 ALAS catalytic mechanism

ALAS enzymes use pyridoxal 5′-phosphate (PLP) as a cofactor to catalyse the condensation of succinyl-CoA and glycine to produce 5′-aminolevulinic acid (ALA) in the first and rate determining step of the haem biosynthetic pathway[4,5]. During this condensation reaction CoA and CO_2 are released as by-products:



Within vertebrates, two different nuclear encoded ALAS isozymes exist with ~60% amino acid sequence identity[22,23]. ALAS1 is the housekeeping enzyme providing a basal level of haem in non-erythroid cell types for use by cytochromes and other haemoproteins. ALAS1 expression is of particular importance in hepatic cells where haem production functionalises cytochrome P450 for drug detoxification[24,25,26]. Haem production in hepatocytes is highly regulated at multiple levels at this rate limiting step, as excess haem leads to significant damage of the liver[25]. The second isozyme, ALAS2, is predominantly expressed in erythroid progenitor cells. It is responsible for the production of around 85–90% of total body haem, where it acts as the cofactor for haemoglobin[25]. The accumulation of haem in erythroid cells results in significantly increased synthesis of haemoglobin and stimulates the maturation of erythroblast cells[25]. Underlining the importance of haem synthesis, pathogenic mutations exist in both the erythro- and hepato-specific biosynthetic genes that lead to the toxic accumulation porphyrin intermediates. This biochemical phenotype is typical of the group of inherited disorders known as porphyrias[27,6].

The ALAS catalytic mechanism is conserved among both isozymes[28,29,30,31]. It requires the formation of a Schiff-base covalent bond between the aldehyde carbonyl of the PLP cofactor and ε-amino group of a conserved lysine residue. The reaction mechanism proceeds in an ordered sequential manner. First, in the absence of any substrates, PLP exists as an internal aldimine adduct which involves the Schiff base interaction to the active site lysine (Lys391 in human ALAS2 (hsALAS2)) (Figure 2-3, state I). Glycine initiates catalysis via a transaldimation reaction, which replaces the lysine ε-amino group in the Schiff-base linkage with the α-amino group from glycine to form a new external aldimine adduct with the PLP cofactor (Figure 2-3, state II). Glycine subsequently undergoes deprotonation of the pro-R proton (on the Cα) that results in the formation of a transient quinonoid intermediate (Figure 2-3, state III). This intermediate is primed for nucleophilic attack by the second substrate, succinyl-CoA (Figure 2-3, state IV), which results in the release of CoA (Figure 2-3, state V). Next, decarboxylation, assisted by the active site histidine (His285 in hsALAS2), of the Cα–carboxylate bond in the remaining aldimine to α-amino-β-ketoadipate group releases CO_2 and leads to formation of a stabilised quinonoid intermediate (Figure 2-3, state VI) . Finally, re-protonation of the quinonoid intermediate regenerates the internal aldimine of PLP with the active site lysine (Figure 2-3, state VII and I) which in turn allows the release of the product, ALA. There have been a number of kinetic studies that suggest this final phase of the reaction, that leads to product release and regeneration of the internal aldimine, is the rate-determining step[28,29].

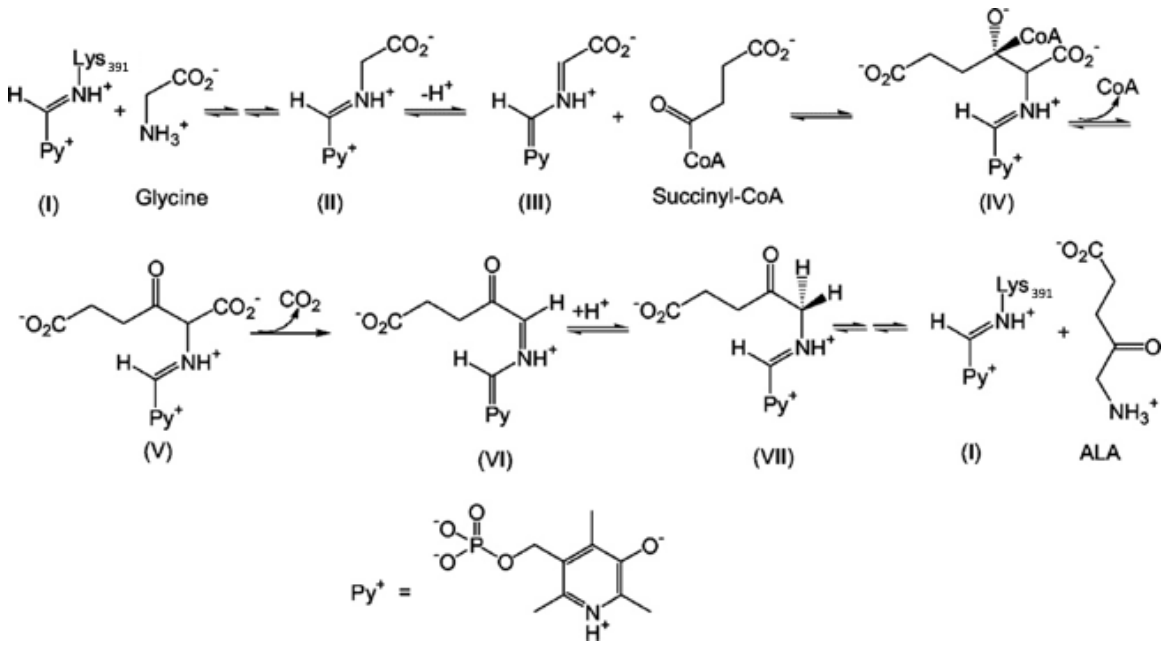


Figure 2-3: Conserved reaction mechanism of ALAS enzymes.

Py refers to the pyridine ring and phosphate portion of the cofactor PLP (shown at bottom). Displayed in the reaction mechanism is the internal or external aldimine linkage to the protein or substrate intermediates. Figure adapted[32].

2.2.2 ALAS2 mutations in disease

ALAS2 variants are reported and associated with two human diseases. Frameshift and indel mutations are reported in exon 11 of the human *ALAS2* gene[33,34]. Exon 11 encodes for a C-terminal extension that has evolved in eukaryotic forms of the enzyme and is not present in prokaryotes. Mutations in exon 11 result in deletion, replacement, or elongation of the Ct-extensions and are the molecular cause of the rare, erythroid specific, inherited disorder known as X-linked protoporphyria (XLP, OMIM #300752) where inheritance of a single copy of the mutated gene on the X chromosome leads to the disease in males. Females with one mutated copy of the gene also display symptoms, however, severity is broad and ranges from asymptomatic to severe. XLP is a gain of function disorder (GOF) resulting in

increased enzyme activity. The increased activity, after deletion of this sequence, implicates a possible function of the Ct-extension to provide a regulatory role, limiting catalytic activity of the enzyme[35,36,37]. The mechanism of how the Ct-extension regulates enzyme activity is currently unknown. The disease presents with painful phototoxicity with increased risk for liver dysfunction and failure[34,33]. At the enzymatic level the gene mutations causing increased ALAS2 activity result in the accumulation of downstream toxic haem intermediates such as PPIX. These intermediates accumulate in the bone marrow, blood plasma, and red blood cells leading to hypersensitivity of the skin to sunlight[34,33]. XLP is categorised as 'ultra-rare' and the true incidence of X-linked protoporphyria is unknown.

In addition to XLP, inherited loss-of-function (LOF) mutations in *ALAS2* have been reported to cause a blood disorder known as X-linked sideroblastic anaemia (XLSA, OMIM #300751). In this case, mutations that result in ALAS2 enzyme deficiency lead to haem-deficient and iron-overloaded red blood cells (ringed sideroblasts)[38,39,40,41]. XLSA is also classified as ultra-rare and the true incidence of this disease is unknown. XLSA is generally caused by missense mutations within the catalytic domain (exons 5–10) with some additional mutations reported in the Ct-extension (exon 11)[38,39,40,41]. Hence, there is a clear relationship between the location and type of mutation in the *ALAS2* gene with two opposing kinetic effects on the ALAS2 enzyme that lead to very different disease phenotypes. While this dichotomous nature of the genotype-phenotype correlation between XLSA and XLP is now widely accepted, the molecular mechanisms that result in the GOF phenotypes seen in XLP are largely unknown, beyond the proposal that the ALAS2 Ct-extension has evolved a self-inhibitory role, deletion of which leads to higher ALAS2 enzyme activity[35,36,37].

2.2.3 Erythroid specific porphyrias

XLP belongs to a group of inherited blood porphyria disorders that are also caused by deficiency of haem biosynthesis at the intermediate and terminal steps, including ALAS2 over-activity. Of this group of disorders, a further three (in addition to ALAS2-associated XLP) are classified as erythroid-specific porphyrias, whereby loss of function of a haem biosynthetic enzyme downstream of ALAS2 in the erythroid pathway results in the accumulation of toxic haem intermediates (Figure 2-4). These disorders have a similar clinical presentation to XLP[42,43]. The most common of these erythroid-specific porphyrias affects the terminal enzyme, FECH, with an estimate between 1 in 75,000 and 1 in 200,000 people affected by the disease in Europe[42,43].

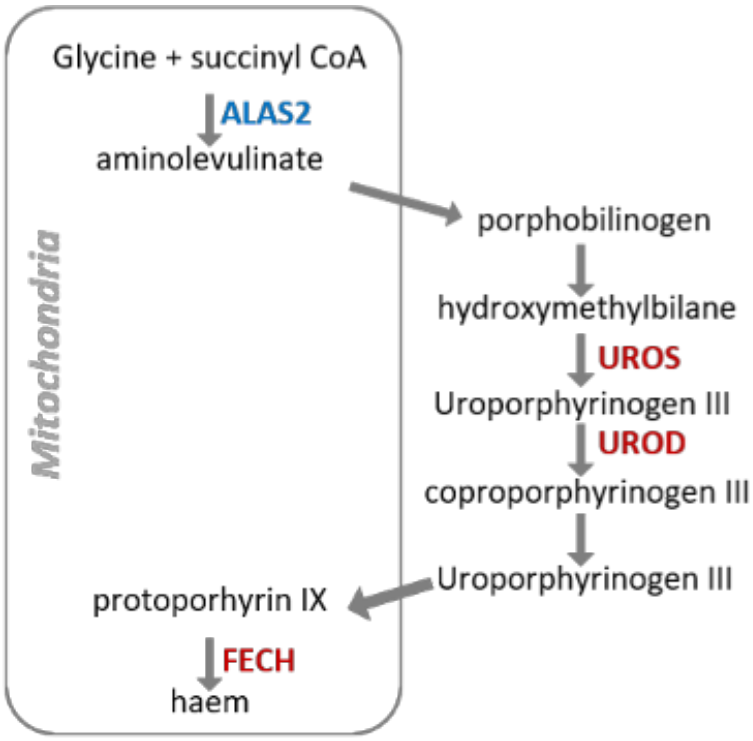


Figure 2-4: Erythroid specific porphyria in haem biosynthetic pathway.

Deficiency in UROS, UROD and FECH (red) lead to erythroid specific porphyria. Overactivity of ALAS2 (blue) leads to XLP.

To rescue or reactivate deficient enzymes in the treatment of any inherited disorder is extremely difficult. However, inhibition of ALAS2 as potential therapeutic target, via a substrate reduction approach[44], could provide a more realistic treatment for downstream porphyria and potentially XLP itself (Figure 2-2). Substrate reduction works on the premise that reducing the activity of upstream enzymes in a biosynthetic pathway reduces flux into that pathway, subsequently reducing the accumulation of toxic substrates that arise due to deficiency in enzymes downstream of the pathway[44]. In this case, reduced ALAS2 activity could decrease metabolic flux into the haem biosynthetic pathway, potentially reducing build-up of toxic haem intermediates, particularly PPIX. Such an approach is already FDA-approved and in practice, in the hepatic pathway, for the treatment of hepatic forms of porphyria. In this instance, siRNA knockdown of the ALAS1 gene shows significant rescue for acute intermittent porphyria, a form of hepatic specific porphyria[45,46]. This provides a proof of principle for the potential of a similar therapeutic benefit of ALAS2 inhibition towards the erythropoietic forms of porphyria.

2.2.4 Wider interactome of ALAS2

ALAS2 has been implicated in protein-protein interactions. ALAS2 is believed to take part in a haem biosynthesis metabolon involving FECH, PPOX and mitochondrial metabolite transporters (e.g. ABCB7, ABCB10, Tmem14c and Tfrc), as well as the TCA cycle enzyme SCS[21]. In fact, direct interactions have previously been shown between the ALAS2 and SUCLA2 subunit of the SCS complex though pull-down experiments *in vitro*[17]. The authors indicate that the interaction site could be mapped to the C-terminal extension of ALAS2 as deletion and point mutation studies appear to abrogate the interaction. However, the proposed substrate channelling explanation as a physiological function of this interaction is controversial, since SCS

has been shown to be unable to catalyse the quantity of succinyl-CoA required for haem synthesis in erythroid cells[47]. By contrast, a substrate channelling mechanism between ALAS2 and downstream haem biosynthetic enzymes is more plausible, as such a mechanism would protect the cell from exposure to intermediate haem metabolites, which are highly toxic.

2.3 The many roles of the succinyl-CoA synthetase complex

2.3.1 Reaction mechanism and function

Succinyl-CoA synthetase (SCS) is an enzyme complex found within the tricarboxylic acid (TCA) cycle. It is composed of an α and β subunit, encoded by different genes[12,11]. The two subunits together are responsible for catalysing the only reaction that is reversible under physiological conditions in the TCA cycle (Figure 2-5)[48].

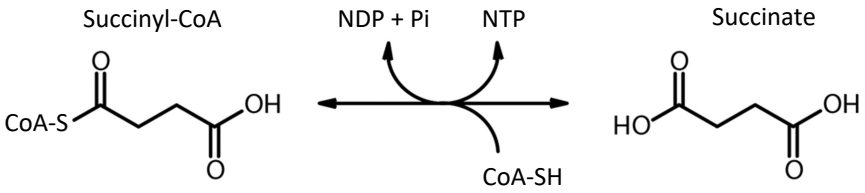


Figure 2-5: SCS reaction mechanism.

In the forward direction, the reaction is catabolic and drives energy production in the cycle. The thio-ester linkage of succinyl-CoA is hydrolysed to produce succinate and CoA and the free energy released from this energy rich bond is harnessed by the β subunit which can bind and phosphorylate a nucleotide triphosphate (NDP), specifically GDP or ADP, to produce GTP or ATP respectively[11,12,49]. This is the only example of substrate level phosphorylation in the TCA cycle where production of energy, in the form of ATP, is directly coupled to the catabolism of SCS substrates. In contrast, other members of the TCA cycle catabolise their substrates to produce

reducing agents to power energy production in the form of ATP via oxidative phosphorylation[2]. This makes the SCS reaction particularly important when oxidative phosphorylation is inhibited, e.g. during hypoxia[50]. To carry out this forward reaction, SCS receives the succinyl-CoA substrate either from upstream in the TCA cycle, produced by α-ketoglutarate dehydrogenase, or from methylmalonyl-CoA mutase via the propionate catabolism pathway (Figure 2-1). Therefore, SCS not only is a core TCA cycle enzyme, involved in canonical TCA function to catabolise molecules derived originally from sugars, but importantly, also provides an entry point for the conversion of carbon sources derived from fats and protein into energy by the cycle[2].

In the reverse reaction, energy in the form of ATP or GTP is utilised to catalyse the production of succinyl-CoA from succinate and CoA. The reverse reaction is of particular importance in anabolic tissues such as liver where succinyl-CoA is required for the regulation of ketone synthesis[51,52] and utilised in haem biosynthesis[53]. In ketone synthesis, SCS is proposed to play an important role in producing succinyl-CoA as a substrate for succinylation, a regulatory post translational modification of proteins. Succinylation is key to inhibiting the enzyme β-hydroxybutyrate dehydrogenase, involved in ketone synthesis. Its regulation is essential to prevent ketoacidosis, a disorder directly linked to the overproduction of ketone bodies in the liver[54,55,52]. Given the increasing number of proteins identified to be regulated by succinylation, the reversible nature of the SCS reaction may have even more broad reaching roles in mitochondrial metabolic regulation[56,57]. Succinyl-CoA is also a core substrate for the first step of haem biosynthesis, where it is condensed with glycine to form ALA by ALAS enzymes. Previous reports of the reverse reaction of SCS providing the majority of succinyl-CoA for erythrocyte maturation, have been refuted as SCS alone is unable to catalyse the required

amounts[47]. Instead, remodelling of cellular metabolism during erythropoiesis allows α ketoglutarate dehydrogenase to supply succinyl-CoA that is originally sourced from glutamine[47]. However, in liver cells, where equivalent levels of haem synthesis are not required, the reverse reaction could be sufficient to supply ALAS1 with the required level of substrates for the production of haem[47].

2.3.2 Two complex isoforms exist

To reflect the differing biological roles that the enzyme is responsible for, dependent on the reaction direction and tissue specificity, two isoforms of the SCS complex have evolved in eukaryotes[52,11]. Both complexes share the same α-subunit (SUCLG1) but differ in their β subunits, allowing SUCLG1 to heterodimerise with either SUCLG2 or SUCLA2. The SUCLG1-SUCLG2 complex forms the GTP specific isoform (G-SCS) and the SUCLG1-SUCLA2 complex forms the ATP specific isoform (A-SCS). Both forms of the complex are expressed to some degree in nearly all tissues, however, the relative proportions of each complex vary significantly dependent on cell type. A-SCS is more highly expressed in tissues that require increased oxidative phosphorylation, such as the heart, brain, testes and skeletal muscle, whereas the G-SCS complex is expressed more highly in anabolic tissues such as the liver[49,58]. The different nucleotide specificity provided by the two complexes has been speculated to give rise to a favoured directionality of the reaction that could explain the tissue dependent expression[52]. This is because the reaction equilibrium that each SCS complex reaches is expected to be sensitive to the relative concentrations of its specific nucleotide substrate and product in the particular tissue.

2.3.2.1 The roles of G-SCS

Within the mitochondria, the GTP to GDP ratio is estimated to be one hundred[51,59,52]. This excess of GTP allows G-SCS to participate more readily in the reverse reaction (the hydrolysis of GTP to catalyse formation of succinyl-CoA from succinate and CoA) and is therefore most likely tasked to regulate succinyl-CoA production in ketone body metabolism[51]. Tissue specific temporal changes in G-SCS complex formation from expression regulation of SUCLG2 support its participation in two seemingly opposing functional roles of ketone metabolism. G-SCS has been shown to be upregulated during ketone body catabolism in the brain during times of starvation[52]. In this case, the reverse reaction is specifically required to recycle succinate generated from the conversion of acetoacetate and succinyl-CoA to acetoacetyl-CoA during the extraction of energy step from ketone bodies delivered to the brain. Therefore, G-SCS plays a very important role in energy production in the brain specifically in times of glucose starvation. In contrast, within the liver, where induction of ketogenesis is important to provide ketone bodies as a source of energy to other tissues, there is a significant down regulation of G-SCS[52]. This is likely because liver cells already lack core enzymes that are required to utilise ketone bodies for energy production in order to favour their synthesis and delivery to other tissues. As G-SCS does not have a recycling role in ketone catabolism in the liver, it is therefore presumed that downregulation of SUCLG2 serves to reduce its role in the production of succinyl-CoA, which can otherwise serve as an inhibitor of β-hydroxybutyrate dehydrogenase through succinylation[52,54].

2.3.2.2 The roles of A-SCS

In the mitochondria, the ATP to ADP ratio of one[51,59,52], suggests that the A-SCS complex is better suited to drive the reaction in the forward direction of the TCA cycle. This function, as a result of continued TCA cycling, increases overall TCA output of reducing agents for cellular energy production in the form of oxidative phosphorylation. In addition, the forward reaction in A-SCS allows for the direct production of ATP though substrate level phosphorylation. It also implicates the A-SCS complex as the specific TCA entry point for succinyl-CoA that is derived from propionate catabolism.

2.3.3 The role of SCS in disease

Considering the aforementioned importance of SCS in many different metabolic processes, it is not surprising that defective SCS is linked to human disease. SCS deficiency represents a group of inherited metabolic disorders that can be caused by autosomal recessive mutations of the genes coding for SCS, resulting in loss-of-function phenotypes. SCS deficiency is extremely rare, with approximately 71 known cases[60]. Prognosis is poor and usually results in death in the neonatal or early childhood period[60]. To date, loss of function mutations have been reported for genes encoding SUCLG1 and SUCLA2, but not SUCLG2.

Deficiency of SUCLA2 results in biochemical and clinical phenotypes specific to A-SCS biological function. These include impaired oxidative metabolism, encephalomyopathy and mitochondrial DNA depletion[60–62]. The metabolic block arising from SUCLA2 deficiency results in increased levels of toxic metabolites associated with succinyl-CoA, such as succinylcarnitine, as well as toxic metabolites

associated with intermediate substrates of propionic metabolism, such as methylmalonylcarnitine and propionylcarnitine[63]. The accumulation of these propionic intermediates leads to phenotypic overlap with a group of metabolic disorders known as organic acidemias, associated with defective propionic metabolism (Section 2.4.4).

The pathological mechanism that results in mtDNA depletion is currently unknown, despite this being the most common phenotype of SUCLA2 deficiency. The connection with enzymes involved in mtDNA maintenance has previously been explored[64]. A reported interaction of SUCLA2 with the nucleoside diphosphate kinase 4 (NDPK4) is speculated to be important in the regulation of the nucleotide salvage pathway, a process that recycles nucleosides formed in the catabolism of DNA and RNA[65–67]. NDPKs are required to convert ATP into other DNA and RNA precursors; however, no evidence exists that definitively describes the functional nature of this interaction, or association with disease.

Deficiency in SUCLG1 is likely to affect both A-SCS and G-SCS complexes, in addition to the symptoms found in SUCLA2 deficiency, it is associated with more severe phenotypes such as hepatomyopathy. It is likely that these additional SUCLG1 specific defects are related to the deficiency of G-SCS complex in liver cell metabolism[60].

2.4 Enzyme complexes in propionate metabolism

2.4.1 The propionyl-CoA carboxylase complex

Propionyl-CoA carboxylase (PCC) exists in the matrix of the mitochondria with evidence of loose association to the inner mitochondrial membrane[68]. The enzyme forms a hetero-oligomeric protein complex comprised of two proteins, PCCA and

PCCB[69,10,9]. The overall reaction mechanism occurs via two distinct half reactions where PCCA is required to catalyse the first half of the reaction and PCCB is subsequently required for the second half (Figure 2-6). PCC itself is a member of the larger family of biotin-dependent carboxylases and is found within the subfamily of carboxylases that process acyl-CoA esters.

The acyl-CoA carboxylase subfamily family have all evolved 3 core domains that allow the conserved reaction mechanism to take place (Figure 2-6)[70,71]. PCCA houses 2 of the 3 domains named the biotin carboxylase (BC) domain and the biotin carboxylase carrier protein (BCCP). The BCCP domain sits at the C-terminus of PCCA residues (653-728). The function of BCCP is to covalently bind the biotin cofactor at Lys699 via amide bond formation, a reaction catalysed in the mitochondria by the holocarboxylase synthase enzyme. The BCCP domain in its biotinylated form is required by both halves of the reaction. In the first half, the N-terminal BC component uses ammonium bicarbonate as a source of CO_2 for the carboxylation of the biotin cofactor at the N1′ atom (Figure 2-6)[70,71]. The reaction is dependent on energy derived from ATP hydrolysis and encoded within the BC domain is also an ATP-grasp fold responsible for binding Mg-ATP and coupling cleavage of the gamma phosphate to power carboxylation of the biotin cofactor. The second half of the reaction takes place within the PCCB subunit. PCCB is formed by a single carboxyltransferase (CT) domain that is responsible for the binding of propionyl-CoA and catalysing the CO_2 exchange between the carboxybiotin group bound to BCCP of PCCA and propionyl-CoA, producing D-methylmalonyl-CoA[70,71].

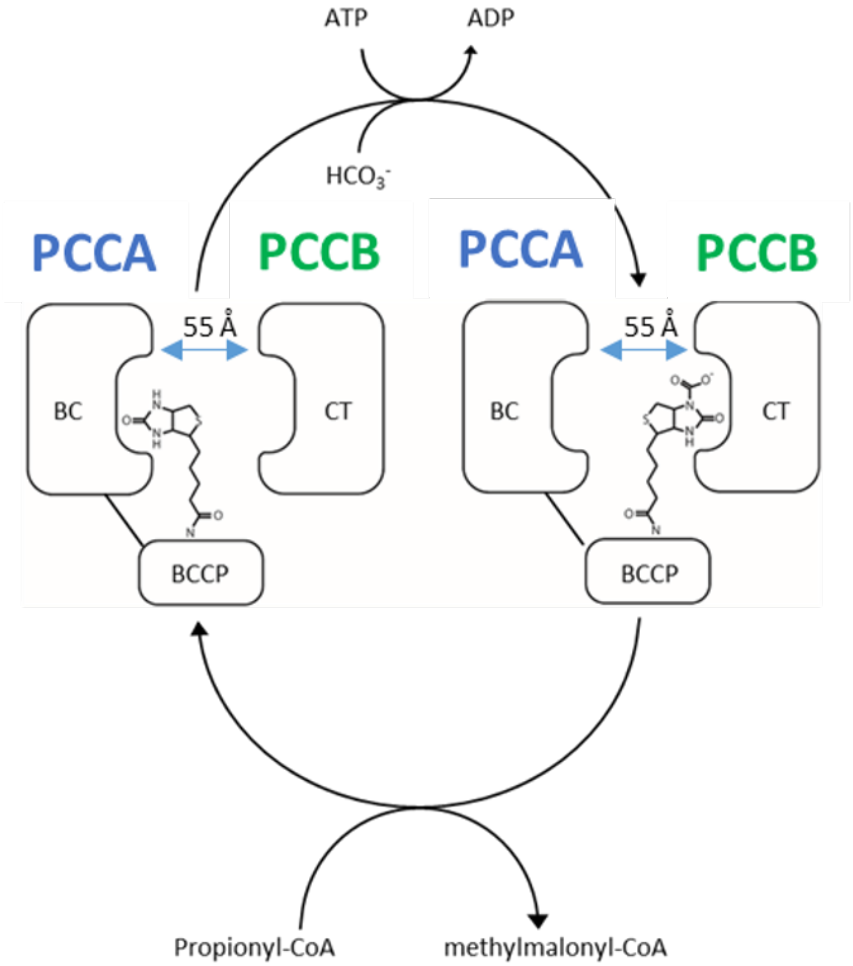


Figure 2-6 PCC reaction mechanism.

PCCA biotin carboxylase subunit labelled BC and biotin carboxylase carrier protein labelled BCCP. PCCb carboxyltransferase domain labelled CT. The chemical structure of biotin Is displayed in the caboxylated (right) and decarboxylated (left) forms.

2.4.2 The MMUT-MMAA complex

MMUT is also found in the mitochondrial matrix and is one of two human enzymes that requires the essential cofactor vitamin B12 (cobalamin, Cbl). The crystal structure of hsMMUT was solved revealing a homo-dimeric arrangement of subunits that each can bind one reactive Cbl cofactor[14]. MMUT uses an activated form of Cbl, 5′-deoxyadenosylcobalamin (AdoCbl), primed to carry out the isomerisation reaction.

The reaction mechanism utilises radical chemistry to enable the 1,2 rearrangement of methylmalonyl-CoA producing succinyl-CoA[72]. The first step requires homolytic cleavage of the upper axial ligand of AdoCbl(III) reducing the cofactor to Cbl(II) and forming a 5′-deoxyadenosyl radical. This radical can abstract a hydrogen atom from the methylmalonyl-CoA substrate. The newly formed substrate radical subsequently undergoes 1,2 rearrangement of the carboxyl group on the vicinal carbon to produce a product radical that can reaccept a hydrogen atom, producing succinyl-CoA and in turn regenerating the 5′-deoxyadenosyl radical ready for the next cycle (Figure 2-7).

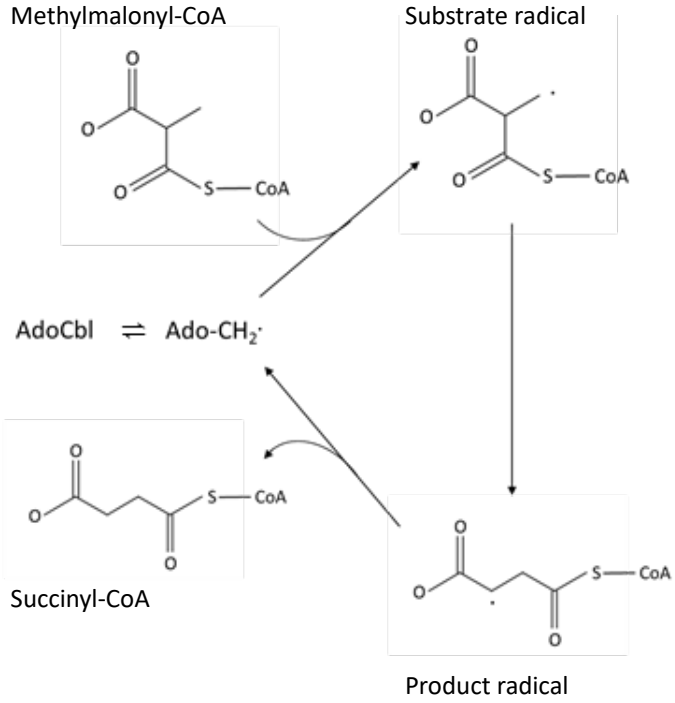


Figure 2-7: MMUT reaction mechanism.

Continued catalysis leads to reduced MMUT activity via a gradual loss of the 5′-deoxyadenosyl radical and the accumulation of the inactive Cbl(II) form of the co-factor bound to MMUT[15,73]. To combat this, a highly complex trafficking and cofactor activation pathway has evolved that serves to strip MMUT from inactive cofactor and deliver new AdoCbl[15,73,74,75]. Free cobalamin in its inactive form is bound by the

activation enzyme MMAB. MMAB is a homo-trimeric enzyme that catalyses a reaction to convert the reduced form of Cbl (Cbl(I)) to the oxidised form AdoCbl through adenosylation with ATP, producing the active cofactor and PPPi[76,77]. To bind AdoCbl *in vivo*, MMUT requires the assistance of the homodimeric co-chaperone MMAA. It is proposed that MMAA uses its GTPase activity (one GTPase per monomer) to act as a gating mechanism, likely tasked with transferring the correct form of Cbl, in this case AdoCbl originally derived from MMAB[15,73]. The same GTPase activity has also been implicated in allowing the unloading of Cbl(II) in inactivated forms of MMUT[15,73].

2.4.3 The propionate catabolism metabolon

In addition to essential MMAA-MMUT interaction, our collaborators[20] have been investigating the wider interactome of MMUT using immunoprecipitation experiments. To date they have gathered strong evidence to suggest that within cells the MMUT-MMAA complex could form part of a larger protein metabolon involving both upstream cofactor loading factors including MMAB and Cbl mitochondrial import proteins (MMADHC) (grey pathway Figure 2-1) as well as the intermediate propionate catabolism enzyme MCEE that catalyses conversion between D and L forms of methylmalonyl-CoA[20].

2.4.4 Propionate catabolism and disease

Defects in propionate catabolism result in two severe disorders, called propionic aciduria (PA, OMIM 606054) and methylmalonic aciduria (MMA, OMIM 251000), which form part of a larger group of disorders known as the branched chain organic acidurias. They are both examples of rare inborn errors of metabolism with autosomal recessive inheritance that share highly similar clinical and biochemical

presentation. Disease incidence within Europe is not accurately recorded, however, it is estimated to fall between 1:48,000 and 1:61,000 births for MMA and between 1:50,000 and 1:500,000 births for PA in Western populations[78,79].

PA is the consequence of mutations in either the *PCCA* or *PCCB* gene[79], whereas MMA is caused by genetic defects in either *MMUT*, *MCEE* or Cbl activating or trafficking enzymes (*MMAA*, *MMAB*, *MMADHC*)[76,80,81]. Clinical presentation for both diseases in the most severe cases may include metabolic acidosis and hyperammonaemia, and even be followed by coma and death[78,79]. Mutation annotations within PCCA, PCCB, MMUT and the associated Cbl processing enzymes are vast, with examples affecting multiple domains within each enzyme likely resulting in diverse modes of deficiency[78,79]. Mechanistic understanding, from a structure to function perspective, of these enzymes that have evolved complex tertiary and quaternary conformations with interaction dependent reaction mechanisms is required to fully understand pathogenesis.

3 Methods

3.1 Cloning, mutagenesis

3.1.1 hsALAS2

To clone ALAS2 expression plasmids, DNA encoding human ALAS2 (hsALAS2) with various N- and C-terminal boundaries were inserted via ligation independent cloning into pFB-LIC-Bse or pNIC28-Bsa4 vectors in-frame with a tobacco etch virus protease cleavable N-terminal His_6-tag (Figure 4-2). Site directed mutagenesis was carried out on $hsALAS2_{\Delta N142}$ and $hsALAS2_{\Delta N142\Delta C545}$ cloned into pNIC28-Bsa4 vector using *QuikChange* method and confirmed by sequencing. $hsALAS2_{\Delta N78}$ containing

NheI and BamHI restriction sites was cloned into the pET28a vector in-frame with the N-terminal His_6-tag by collaborators (Jason R. Marcero and Harry A. Dailey, Jr from the Department of Biochemistry and Molecular Biology, University of Georgia, Athens). Variants of the $hsALAS2_{\Delta N78}$ pET28a fusion were prepared via *Quickchange* mutagenesis XLP mutants delAT, delAGTG, delG, and Q548X.

3.1.2 hsA-SCS

To clone has-SCS complex co-expression plasmids, DNA fragments encoding $hsSUCLG1_{\Delta N51}$, aa 52-346; and $hsSUCLA2_{\Delta N41}$, aa 42-463 were both amplified to contain a 5' lic adaptor with a 3' ribosomal binding site (RBS) and a 5' ribosomal binding site with a 3' lic adaptor sequence. Fragments of each gene with opposing lic and RBS sites were subcloned together into the pNIC28-Bsa4 vector (GenBank ID: EF198106) producing a bicistronic co-expression vector with His_6-tag-$hsSUCLA2_{\Delta N41}$ and untagged $hsSUCLG1_{\Delta N51}$ and vice versa. Both genes are subsequently separated by a ribosomal binding site.

3.1.3 hsPCC

To clone hsPCC complex co-expression plasmids, a DNA fragment encoding hsPCCA (aa 53-728) was amplified to contain a 5' lic adaptor with a 3' RBS and a DNA fragment encoding hsPCCB (aa 31-539) was amplified to contain a 5' RBS with a 3' lic adaptor sequence. Fragments of each gene with the opposing lic and RBS sites were subcloned together into the pNIC28-Bsa4 vector (GenBank ID: EF198106) producing a bicistronic co-expression vector with His_6-tag-hsPCCA and untagged hsPCCB. Both genes are subsequently separated by a ribosomal binding site.

3.1.4 hsMMAA and hsMMUT

Constructs were cloned in previous studies and used here[14].

3.2 Protein expression and production procedure

The following protocol describes the expression and purification protocol for all constructs used in this study.

3.2.1 Expression

To express proteins, constructs cloned into pNIC28-Bsa4 and pNIC-CTHF expression vectors were transformed into BL21(DE3) competent cells (PCC bi-cistronic vectors were co-transformed with BirA-expressing plasmid). Single-colony inoculations were induced at an OD_{600} of 0.5-0.7 with 0.1 mM IPTG for 16-18 h at 18 °C (For the PCC expression, 25 mg/L biotin powder was supplemented in the expression media). pFB-LIC-Bse vectors containing hsALAS2 were expressed in baculovirus-infected insect cells in Sf9 media at 27 °C at a cell density of 2.0 x 10^6 per ml in roller bottles for 72 hr. All cultures were centrifuged and cell pellets stored at -80°C.

3.2.2 Purification

To purify proteins, all cells containing expressed proteins were resuspended in lysis buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 10 mM imidazole, 5% Glycerol, 0.5 mM TCEP) (total volume of 30mL per 1 L of original cell expression) and lysed by sonication (for PCC a further 5 mg of biotin was supplemented per 180 mL of cell lysate) and centrifuged at 35,000 x*g*. The clarified cell extract was incubated with 2.5 mL of Ni-NTA resin per 200 mL of cell lysate. The column was washed with 100 mL

Binding Buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 5% glycerol, 10 mM imidazole, 0.5 mM TCEP), 50 mL Wash Buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 5% glycerol, 40 mM imidazole, 0.5 mM TCEP) and eluted with 15 mL of Elution Buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 5% glycerol, 250 mM imidazole, 0.5 mM TCEP) per 2.5 mL of Ni-NTA resin. The eluent fractions were concentrated to 5 mL and applied to a Superdex 200 (ALAS2, A-SCS, MMAA and MMUT) or Superose 6 (PCC) 16/60 column pre-equilibrated in GF Buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 0.5 mM TCEP, 5% glycerol). Eluted protein fractions were pooled and concentrated to 5-20 mg/mL stocks.

3.3 hsALAS2 specific methods

3.3.1 Crystallisation

To crystalise hsALAS2 for structural studies using x-ray crystallography, crystals were grown by the vapour diffusion method. hsALAS2$_{\Delta N142}$, protein was pre-incubated with 5 mM hydroxylamine for 30 minutes on ice to convert PLP into the non-covalent form. Sitting drops containing 75 nL protein (17 mg/mL) and 75 nL well solution containing 25% (w/v) PEG 3350, 0.1 M Bis-Tris pH 6.7 and 0.3 M $MgCl_2$ were equilibrated at 20 ^{o}C. hsALAS2$_{\Delta N78}$, sitting drops containing 200 nL protein (11 mg/mL), 133 nL well solution containing 20% (w/v) PEG3350 and 0.2 M NH_4Cl, and 67 nL seed solution derived from previously obtained micro crystals. Crystals were cryo-protected using 25% (v/v) ethylene glycol and flash-cooled in liquid nitrogen.

3.3.2 Data collection, Processing and refinement

To solve the hsALAS2 structures, diffraction data were collected at the DLS beamline i04 and i24. hsALAS2$_{\Delta N142}$ was crystallized in the monoclinic space group C2 with two molecules in the asymmetric unit. The data was processed using the

Xia2[82] auto processing dials pipeline. hsALAS2$_{\Delta N78}$ was crystallized in the monoclinic space group $P2_1$ with four molecules in the asymmetric unit. The data were processed using non-ellipsoidal anisotropy truncations by the *STARANISO*[83] server, based on a lower limit of 1.2 for the local I/sigI producing resolution cut off of 2.65 Å in the best direction and 3.55 Å in worst. The human structures were solved by molecular replacement using the program *PHASER*[84] and the rcALAS structure (PDB code 2bwn) as the search model. The final model was produced by iterative cycles of restrained refinement and model building using *COOT*[85], *REFMAC5*[86] and *Phenix.refine*[87]. Coordinates and structure factors for hsALAS2$_{\Delta N142}$ were deposited in the PDB with the accession code 6HRH.

3.3.3 SAXS solution studies

To analyse the solution structures of hsALAS2, SAXS experiments for the hsALAS2$_{\Delta N142}$ and hsALAS2$_{\Delta N142\Delta C545}$ were performed at 0.99 Å wavelength on the Diamond Light Source beamline B21 coupled to the Shodex KW403-4F size exclusion column (Harwell, UK) and equipped with Pilatus 2M two-dimensional detector at 4.014 m distance from the sample, $0.005 < q < 0.4$ Å^{-1} ($q = 4\pi \sin\theta/\lambda$, 2θ is the scattering angle). Individual SAXS frames were recorded every 3 seconds of SEC elution. The samples were prepared at 10 mg/ml in a buffer containing 300 mM NaCl, 25 mM HEPES pH 7.5, 1 mM TCEP, 2 % glycerol, 1% sucrose and the measurements were performed at 20℃. The data were processed and analysed with *Scatter*[88] and the *ATSAS* program package[89]. *Scatter* was used to calculate the radius of gyration Rg and forward scattering I(0) via Guinier approximation, and to derive the maximum particle dimension Dmax and P(r) function. The *ab initio* model was derived using *DAMMIF*[89]. Ten individual models were created, then overlaid and averaged using *DAMAVER*[89].

3.3.4 Differential scanning fluorimetry

To determine the thermostability, hsALAS2$_{\Delta N142}$ was assayed for shifts in melting temperature caused by titration of PLP and succinyl-CoA in a 96-well PCR plate using an Mx3005p RT-PCR machine (Stratagene). Each reaction well (20 µl) consisted of protein (20 µM in a buffer containing 50 mM HEPES, pH 7.5, 500 mM NaCl), SYPRO-Orange (Invitrogen) diluted 1,000×. Fluorescence intensities were measured from 25 °C to 96 °C with a ramp rate of 1 °C/min, as described. The temperature shifts, $\Delta T_{m,\,obs}$, for each ligand were measured using real-time quantitative polymerase chain reaction (RT-QPCR). The experiment was repeated three times and an error was calculated based on the standard deviation.

3.3.5 Fluorescence intensity

To analyse the PLP binding status, hsALAS2 $_{\Delta N142}$ was assayed for shifts in fluorescence intensity caused by titration of PLP. Each well (10 µl) consisted of protein (20 µM in a buffer containing 50 mM Hepes, pH 7.5, 500 mM NaCl) and varying concentrations of PLP. Schiff base formation was monitored via excitation at 440 nm and emission at 520 nm using an Omega plate reader. The experiment was repeated three times and an error was calculated based on the standard deviation.

3.3.6 Size exclusion co-elution

To study interactions between hsA-SCS and hsALAS2, co-elution studies were performed with hsALAS2$_{\Delta N78}$ and the hsA-SCS complex isolated by His$_6$-tag-hsSUCLA2$_{\Delta N41}$ and untagged hsSUCLG1$_{\Delta N51.}$ Protein samples were pre incubated at 1:1 molar ratio and 5 mg/ml total protein concentration for 30 mins and passed over and analytical superdex 200 increase column. Eluent samples were analysed via SDS-PAGE.

3.3.7 Cross-linking

To study interactions between hsA-SCS and hsALAS2 covalent crosslinking studies were performed. 0.5 mg/mL of $sALAS2_{\Delta N78}$ and the hsA-SCS complex isolated by His_6-tag-$hsSUCLA2_{\Delta N41}$ and untagged $hsSUCLG1_{\Delta N51}$ were incubated with DSG lysine Crosslinker added to 0.2 mM in 10uL reaction volumes at room temperature (500 mM NaCl, 50 mM Hepes pH 7.5, 0.5 mM TCEP, 5% glycerol) for 30 min. The reaction was quenched with 150 mM amonium bicarbonate for 10 mins and samples were analysed by SDS-PAGE gel and triptic digest MS/MS.

3.4 SCS specific methods

3.4.1 Crystallisation

To crystalise A-SCS for structural studies using x-ray crystallography, crystals of as purified hsA-SCS isolated by His_6-tag-$hsSUCLG1_{\Delta N51}$ and untagged $hsSUCLA2_{\Delta N41}$ were grown by vapour diffusion method. Sitting drops containing 100 nL protein (6.2 mg/mL) and 50 nL well solution containing 2.1M DL- malic acid were equilibrated at 20°C. Crystals were cryo-protected using 25% (v/v) ethylene glycol and flash-cooled in liquid nitrogen.

3.4.2 Data collection, Processing and refinement

To solve the hsA-SCS crystal structure, diffraction data were collected at the DLS beamline i04 and processed using Xia2 auto processing pipeline. The protein crystallised in the trigonal $P3_121$ space group and contained 1 copy of each subunit in the asymmetric unit. The structure was solved via molecular replacement with 1EUC. The final model was produced by iterative cycles of restrained refinement and model building using *Coot*[85] and *REFMAC5*[86].

3.4.3 SAXS solution studies

To analyse the solution structures of hsALAS2, SAXS experiments for as purified hsA-SCS complex isolated by His_6-tag-hsSUCLA2$_{\Delta N41}$ and untagged hsSUCLG1$_{\Delta N51}$ were performed at 0.99 Å wavelength on the Diamond Light Source beamline B21 coupled to the Shodex KW403-4F size exclusion column (Harwell, UK) and equipped with Pilatus 2M two-dimensional detector at 4.014 m distance from the sample, $0.005 < q < 0.4$ Å-1 ($q = 4\pi \sin\theta/\lambda$, 2θ is the scattering angle). Individual SAXS frames were recorded every 3 seconds of SEC elution. The samples were prepared at 30, 15, 7.5, and 3.75 mg/mL in a buffer containing 300 mM NaCl, 25 mM Hepes 7.5, 1 mM TCEP, 2 % glycerol, 1% sucrose and the measurements were performed at 20℃. For ligand studies in batch mode protein samples at 6, 3 and 1.5 mg/mL samples (300 mM NaCl, 25 mM Hepes 7.5, 1 mM TCEP, 2 % glycerol, 1% sucrose, 1 mM $MgCl_2$) in the presence of 5 mM appropriate ligand, were directly measured through a capillary using the same procedure as above without pre filtration over a SEC column. For each batch dataset a total of 30 frames were recorded. The data were processed and analysed with *Scatter*[88] and the *ATSAS* program package[89]. *Scatter* was used to calculate the radius of gyration Rg and forward scattering I(0) via Guinier approximation, and to derive the maximum particle dimension Dmax and P(r) function. The *ab initio* model was derived using *DAMMIF*[89]. 20 individual models were created, then overlaid and averaged using *DAMAVER*[89]. The *Oligomer*[89] programme was used to calculate fits to known high resolution structures.

3.4.4 Cryo-EM grid making

To determine the structure of A-SCS in the presence of succinyl-CoA using single particle cryo-electron microscopy, 3 µl of 1 mg/mL purified complex were applied to

the glow-discharged Quantifoil Au R1.2/1.3 grid (Structure Probe). Blotting and vitrification in liquid ethane was carried out using a Vitrobot Mark IV (FEI Company) at 4°C and 95% humidity with a 10 s wait and a 3 s blot at zero blotting force from both sides. Grids were loaded into a Glacios transmission electron microscope (ThermoFisher Scientific) operating at 200 keV with a Falcon3 camera. Movies were recorded in linear mode with a pixel size of 0.96 Å and a defocus set to mode with a pixel size of 0.96 Å and a defocus range of −1.5 to −3.0 µm (steps of 0.3 µm). Data were collected with a total dose of 32.52 e $Å^{-2}$ and images were recorded with a 1 s exposure over 19 frames.

3.4.5 EM data processing

To calculate 3D electron density maps of hsA-SCS, A total of 200 dose-fractioned movies were corrected for drift using *RELION's MotionCor2*[90] with the dose-weighting option. CTF parameters were determined by *CTFFIND*4.1[91]. A subset of particles was manually picked and extracted. 4 classes were selected from one round of 2D classification to use in reference-based autopicking using *RELION* 3.0[90]. Downstream particle classification, refinement and post-processing steps were performed in *RELION* 3.0[90]*.* Junk particles were removed using 2D classification resulting in 6175 particles. These particles were used to create an Initial model 3D model. This initial model was used as a reference for 3D classification into 4 classes. The 4 classes were homogenous indicating little structural variation in the data at this resolution. A masked 3D refinement, with dihedral D2 symmetry imposed, was therefore performed using the original 6175 particles classified by 2D classification. After postprocessing and particle polishing the global resolution of the map was calculated based on the Fourier shell correlation (FSC) 0.143 threshold criterion.

3.4.6 EM model building and refinement

To model protein coordinates into the hsA-SCS EM map 4 hsA-SCS heterodimers derived from the crystal structure were rigid body fit to the EM density map. Automatic molecular dynamics flexible fitting was carried out using Namdimator to improve structure fit to density.

3.4.7 Activity assay

To determine whether the hsA-SCS complex is active, hsA-SCS complex at 2 mg/mL was incubated with substrates at 5 mM (500 mM NaCl, 50 mM Hepes pH 7.5, 0.5 mM TCEP, 5% glycerol and 1 mM $MgCl_2$) in a total reaction volume of 15 uL for 20 min at room temperature. Samples were quenched with 0.1 % formic acid and analysed on denature MS

3.5 MMUT and MMAA specific methods

3.5.1 SAXS solution studies

To determine the in solution structures of hsMMUT and hsMMAA, SAXS experiments for hsMMUT and hsMMAA and a 1:1 molar mix of both were performed at 0.99 Å wavelength on the Diamond Light Source beamline B21 coupled to the Shodex KW403-4F size exclusion column (Harwell, UK) and equipped with Pilatus 2M two-dimensional detector at 4.014 m distance from the sample, $0.005 < q < 0.4$ Å-1 ($q = 4\pi \sin\theta/\lambda$, 2θ is the scattering angle). Individual SAXS frames were recorded every 3 seconds of SEC elution. The individual samples were run at 10 mg/mL in a buffer containing 300 mM NaCl, 25 mM HEPES 7.5, 1 mM TCEP, 2 % glycerol, 1% sucrose, 1 mM $MgCl_2$ and 500 µM GTPPCP and the measurements

were performed at 20℃. The data were processed and analysed with *Scatter*[88] and the *ATSAS* program package[89]. *Scatter* was used to calculate the radius of gyration Rg and forward scattering I(0) via Guinier approximation, and to derive the maximum particle dimension Dmax and P(r) function. The *ab initio* model was derived using *DAMMIF*[89]. 20 individual models were created, then overlaid and averaged using *DAMAVER*[89]. The *Oligomer*[89] programme was used to calculate fits to known high resolution structures.

3.5.2 GraFix

To isolate stabilised hsMMAA and hsMMUT complex samples for analysis by cryo-electron microscopy, GraFix experiments were performed. 6 mg/mL of hsMMUT was incubated with 3 mg/mL of hsMMAA in 150 mM KCl, 1 mM $MgCl_2$, 20 mM HEPES pH 7.5, 0.5 mM TCEP and 100 uM GMPPCP in a total volume of 200 uL for 10 mins. The sample was applied to a 12 ml 10-30 % glycerol, 0-0.15% glutaraldehyde (GRADE I) gradient and ultra-centrifuged using SW41 rotor at 29300rpm for 16 hr at 4ºC. 500 uL fractions were collected and analysed on SDS-PAGE.

3.5.3 Cryo-EM grid making and screening

To attempt to determine the structure of the hsMMAA and hsMMUT, the GraFix samples were applied to the glow-discharged Quantifoil Au R1.2/1.3 grid (Structure Probe). Blotting and vitrification in liquid ethane was carried out using a Vitrobot Mark IV (FEI Company) at 4°C and 95% humidity with a 9 s wait and a 3 s blot at zero blotting force from both sides. Grids were loaded into a Glacios transmission electron microscope (ThermoFisher Scientific) operating at 200 keV with a Falcon3 camera. Images were recorded in linear mode with a pixel size of 0.96 Å and a defocus set to mode with a pixel size of 0.96 Å at a defocus of -2.5 µm. Data were

collected with a total dose of 32.52 e $Å^{-2}$ and images were recorded with a 1 s exposure.

3.6 PCC specific methods

3.6.1 SAXS solution studies

To determine the solution structure of the PCC complex, SAXS experiments for the hsPCC complex and PCCA protein were performed at 0.99 Å wavelength on the Diamond Light Source beamline B21 coupled to the Shodex KW403-4F size exclusion column (Harwell, UK) and equipped with Pilatus 2M two-dimensional detector at 4.014 m distance from the sample, 0.005 < q < 0.4 Å-1 ($q = 4\pi \sin\theta/\lambda$, 2θ is the scattering angle). Individual SAXS frames were recorded every 3 seconds of SEC elution. The samples were prepared at 8 mg/mL in a buffer containing 300 mM NaCl, 25 mM HEPES 7.5, 1 mM TCEP, 2 % glycerol, 1% sucrose, 1 mM $MgCl_2$ with and without 5 mM ATPγS and the measurements were performed at 20℃. The data were processed and analysed with *Scatter*[88] and the *ATSAS* program package[89]. *Scatter* was used to calculate the radius of gyration Rg and forward scattering I(0) via Guinier approximation, and to derive the maximum particle dimension Dmax and P(r) function. The *ab initio* model was derived using *DAMMIF*[89]. 20 individual models were created, then overlaid and averaged using *DAMAVER*[89]. The *Oligomer*[89] programme was used to calculate fits to known high resolution structures.

3.6.2 Crystallisation optimisation

To attempt to optimise crystals for determination of the hsPCC complex structure using x-ray crystallography, the complex was crystallised in course screens (HCS3, JCSG7, MCSG1, HIN3) and selected follow up fine screens. Plates were subsequently sent to VMXi for in situ room temperature screening. Multi-crystal

datasets with same unit cells were scaled and merged together using *Xia2*[82] auto processing pipeline.

3.6.3 Cryo-EM grid making and data collection

To determine the structure of hsPCC complex in the presence of ATPγS, 3 µl of 0.3 mg/mL purified complex in the presence of 1 mM $MgCl_2$ and 5 mM ATPγS was applied to the glow-discharged Quantifoil Au R1.2/1.3 grid (Structure Probe). Blotting and vitrification in liquid ethane was carried out using a Vitrobot Mark IV (FEI Company) at 4°C and 95% humidity with a 9 s wait and a 3 s blot at zero blotting force from both sides. Grids were loaded into a Glacios transmission electron microscope (ThermoFisher Scientific) operating at 200 keV with a Falcon3 camera. Movies were recorded in linear mode with a pixel size of 0.96 Å and a defocus set to mode with a pixel size of 0.96 Å and a defocus range of −1.5 to −3.0 µm (steps of 0.3 µm). Data were collected with a total dose of 32.52 e $Å^{-2}$ and images were recorded with a 1 s exposure over 19 frames.

3.6.4 EM data processing

To calculate 3D electron density maps of hsPCC, a total of 813 dose-fractioned movies were corrected for drift using *RELION's MotionCor2*[90] with the dose-weighting option. CTF parameters were determined by *CTFFIND*4.1[91]. A subset of particles was manually picked and extracted. 4 classes were selected from one round of 2D classification to use in reference-based autopicking using *RELION* 3.0[90]. Downstream particle classification, refinement and post-processing steps were performed in *RELION* 3.0[90]. Junk particles were removed using 2D classification resulting in 67946 particles. Using this subset of particles, an initial 3D model was created. This initial model was used as a reference for 3D classification into 8

classes. One 3D class showed high resolution density for the intact complex and consisted of 26722 particles. A masked 3D refinement, with dihedral D3 symmetry imposed, was performed using the 26722 particles classified by 3D classification. After postprocessing and particle polishing the global resolution of the map was calculated based on the Fourier shell correlation (FSC) 0.143 threshold criterion. Symmetry expansion was applied to the original 67946. Masked 3D classification of a single PCCA subunit was performed on symmetry expanded particles to produce 8 classes.

4 Structure function study of hsALAS2

4.1 ALAS chapter introduction

4.1.1 Known ALAS structures

The first crystal structure of ALAS was solved from the bacterium *R. capsulatus* (rcALAS)[92]. ALAS enzymes are members of the fold type I of the PLP-dependent enzyme superfamily, existing as obligate homodimers and binding one PLP molecule per monomer at a large dimer interface[92]. rcALAS structures reveal an induced-fit reaction mechanism upon substrate binding, mediated by open-to-close transitions of a mobile active site loop (504-515 in hsALAS2). In the open state this loop moves away from the active site and the enzyme is poised for substrate binding, subsequent closure of the loop in the active site allows for catalysis to take place. Reopening of the loop allows for product release after catalysis[92]. This conformational flexibility of the active site loop is proposed to be the kinetic barrier for the rate determining step (product release) as previously shown by elegant stopped flow kinetic assays to sample intermediate reaction steps of the enzyme[31].

In addition to the highly conserved catalytic core, solved in the rcALAS structures, eukaryotic ALAS enzymes have evolved extensions at the N and C-termini (Figure 4-1A). The first 49 amino acids of the 142 amino acid long N-terminal extension (human ALAS2 numbering) contain the mitochondrial targeting sequence[93,94]. This region is cleaved upon mitochondrial import, and shows low sequence identity between lower eukaryotic ALAS enzymes and metazoan ALAS1 (Figure 4-1A,B). In addition to this targeting sequence, at the N-terminus, metazoans encode three Cys-Pro motifs conserved in both ALAS paralogues. The first of these two motifs appear within the targeting sequence and are presumably cleaved in the mature forms of the enzyme (Figure 4-1A,B). They are responsible for a haem sensing feedback regulatory mechanism that modulates mitochondrial import based on cellular haem concentration[95,96]. The final Cys-Pro motif, together with an invariant Gln-Glu-Asp-Val motif, are found within the remaining N-terminal extension (Nt-extension) (residues 50 to 143 in human) of unknown function. The remaining Nt-extension displays low sequence identity across ALAS isozymes and lower eukaryotic orthologues, and varies significantly in length (Figure 4-1A, B). At the C-terminus, eukaryotes have evolved an extension that varies between 35 and 60 aa in length within different organisms. Metazoan ALAS1 and ALAS2 share ~53% sequence identity in their Ct-extensions with the major differences occurring in the terminal 9 amino acids. Lower eukaryotic forms of ALAS are more varied and encode an entirely different Ct-extension sequence from metazoans (Figure 4-1B). In humans, mutations that cause deletion or frameshifts of the Ct-terminus extension result in the gain-of-function (GOF) disease XLP (section 2.2.2).

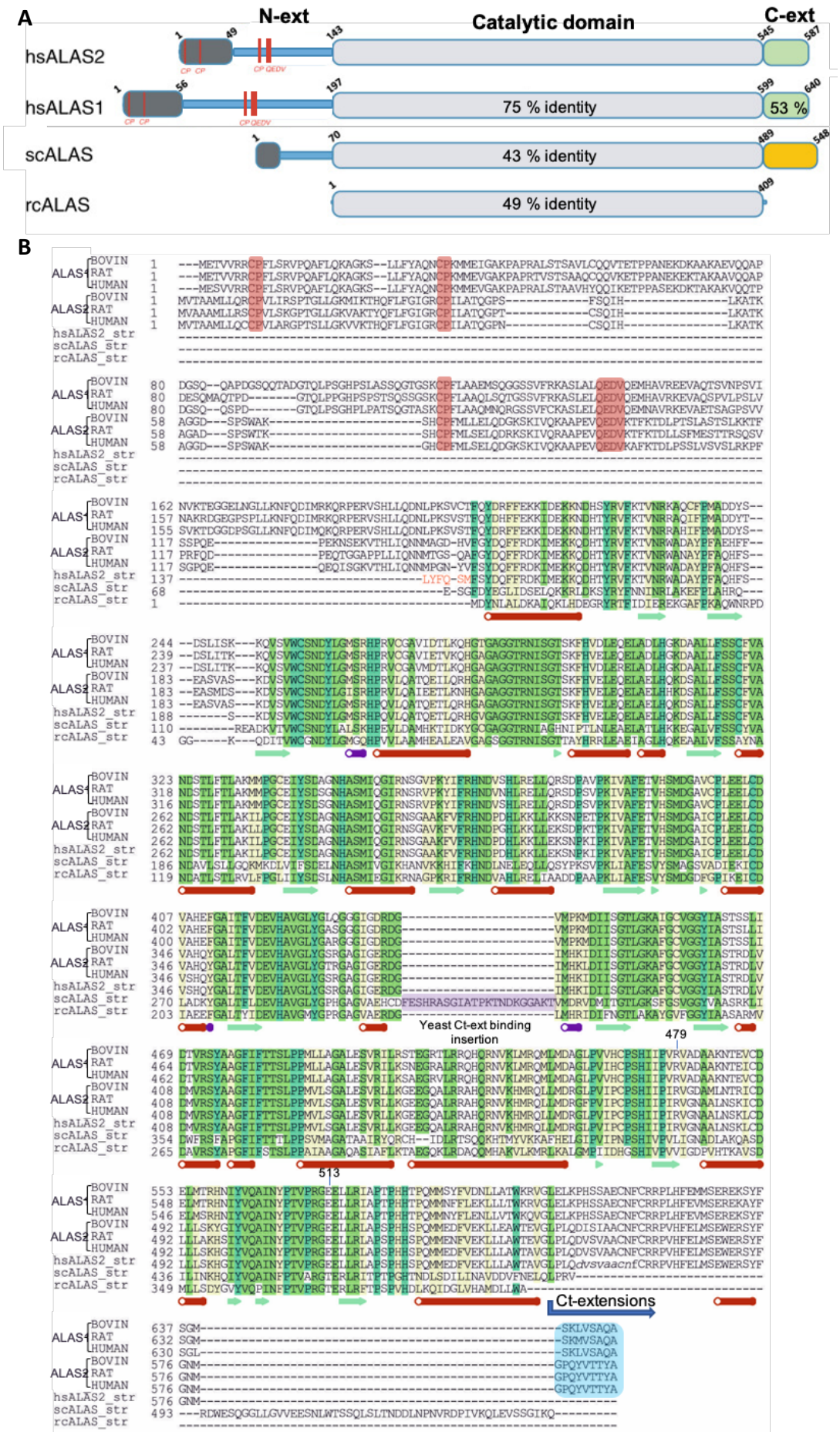

A
N-ext
Catalytic domain
C-ext
hsALAS2
hsALAS1
scALAS
rcALAS
75 % identity
53 %
43 % identity
49 % identity
B
Yeast Ct-ext binding insertion
Ct-extensions

Figure 4-1: Domain architecture and Structure-based sequence alignment of ALAS orthologues.

A: Domain architecture of hsALAS1, hsALAS2, *S. cerevisiae* ALAS (scALAS), and *R. capsulatus* ALAS *(*rcALAS), highlighting the catalytic core (gray box) flanked by N-terminal (black and blue bar) and C-terminal extension (green in higher eukaryotes, orange in lower eukaryotes). The conserved Cys-Pro motifs and invariant Gln-Glu-Asp-Val motif are marked red line and labelled CP and QEDV **B:** Sequences of the crystallized constructs used in the structure determination of hsALAS2 (hsALAS2_str), *S. cerevisiae* ALAS (scALAS_str), and *R. capsulatus* ALAS (rcALAS_str) are aligned, on the basis of Cα-superposition of their structures. Additionally, full-length sequences of ALAS1 and ALAS2 from bovine, rat and human are appended to the alignment. For the hsALAS2_str sequence, vector-encoded residues at the beginning of the construct are denoted in orange letters, and secondary structure assignment is shown at the bottom of the alignment (α-helix, red; 3_{10} helix, purple; β-strand, green). Nt-extension motifs conserved among higher eukaryotes are highlighted in red. The yeast catalytic core insertion that binds yeast Ct-extension is highlighted purple. Blue arrow denotes the start of eukaryotic Ct-extensions. Last un-conserved 9 residues are highlighted blue in metazoan ALAS sequences.

The recently published ALAS2 structure from the budding yeast *S. cerevisiae* (scALAS) provides the first insights into the structure of eukaryotic ALAS2 C-terminal extensions[97]. The structure showed each Ct-extension wrapping around the surface of the ALAS homodimer and is used to speculate the involvement of this region in XLP. However, the sequence of scALAS2 Ct-extension has very poor similarity with other eukaryotic counterparts (Figure 4-1B) and is likely to be evolutionarily distinct from the mammalian ALAS2 Ct-extension. Furthermore, no direct interaction or conformational modulation is detected between the Ct-extension and the active site of the scALAS structure, making it difficult to infer the self-inhibitory mechanism of the C-terminus. Finally, in direct contrast with human ALAS2 XLP-causing mutations, deletion of the scALAS C-terminal extension does not lead to increased enzyme activity[97].

4.1.2 ALAS2 results chapter synopsis

This chapter reports the crystal structure determination of hsALAS2, which, in combination with enzyme kinetics, reveals features of the Ct-extension that explain its role in both enzyme regulation and the XLP disease. The structure is used for a crystallography-based fragment screening campaign revealing a binding hotspot around the Ct-extension, where fragments form interactions that would likely interfere with the Ct-conformational dynamics and point towards future development of ALAS2 activity inhibitors. These fragments represent a starting point to develop ALAS2 inhibitors as substrate reduction therapy for porphyria disorders that accumulate toxic haem intermediates.

4.2 Results

4.2.1 ALAS2 protein production and characterisation

To generate recombinant protein for structural studies specifically aimed at identifying a regulatory role of the human Ct-extension, multiple hsALAS2 constructs were cloned into pNIC28-Bsa4 vectors containing Nt-His_6-tag for *E.coli* expression system. Constructs tested for expression contained either full length ALAS2 lacking the mitochondrial import sequences or have either removed/shortened the N-terminal extension, whilst retaining the catalytic core and Ct-extension intact (Figure 4-2 Blue and black bars). Constructs containing truncation of the N-terminus were generated as this region is predicted to be largely unstructured and may consequently interfere with enzyme solubility or structural studies using X-ray crystallisation. Vectors were initially expressed in a 3 ml culture scale to test for soluble isolation of ALAS2 using Ni^{2+} affinity chromatography as the first purification step (Figure 4-2B). Of the constructs tested, $hsALAS2_{\Delta N142}$ (containing the catalytic

core domain and Ct-extension), was successfully isolated in the eluent. However, this construct showed an additional truncated band, potentially due to proteolysis (Figure 4-2B, lane 1), an outcome observed previously by other studies[37]. Using intact mass spectrometry, the truncated product was attributed to degradation of the 40-aa Ct-extension, producing a mix of 3 truncated subspecies with varying lengths (Figure 4-2C). The predominant species in the sample appeared to be the intact hsALAS2$_{\Delta N142}$ construct (aa 143-587), and crystallisation trials were attempted for this protein without any success. Degradation was not tested over a significant length of time and it is not clear how a relatively small portion of truncated monomer would affect untruncated monomers within the homodimer.

With this in mind, the expression system was changed to baculovirus-infected insect Sf9 cells, with the hope that a different proteolytic background from the expression host would make hsALAS2 less susceptible to cleavage at the C-terminus. Several constructs of hsALAS2 (Figure 4-2A, green bars and blue bars) were isolated as soluble proteins during expression tests, this time without any sign of proteolytic degradation at the Ct-extension (Figure 4-2B, Lanes 2 and 3 and Figure 4-2D). These constructs include the same hsALAS2$_{\Delta N142}$ truncation (aa 143-587) that was isolated from *E.coli* cells and a longer form of the enzyme hsALAS2$_{\Delta N78}$ (aa 79-587). In addition to the N-terminal truncations, hsALAS2 constructs with deleted Ct-extensions were produced (Figure 4-2A, hsALAS2$_{\Delta N78\Delta C545}$ and hsALAS2$_{\Delta N142\Delta C545}$) to serve as controls in mechanistic studies of the proposed C-terminal regulatory domain (Figure 4-2A).

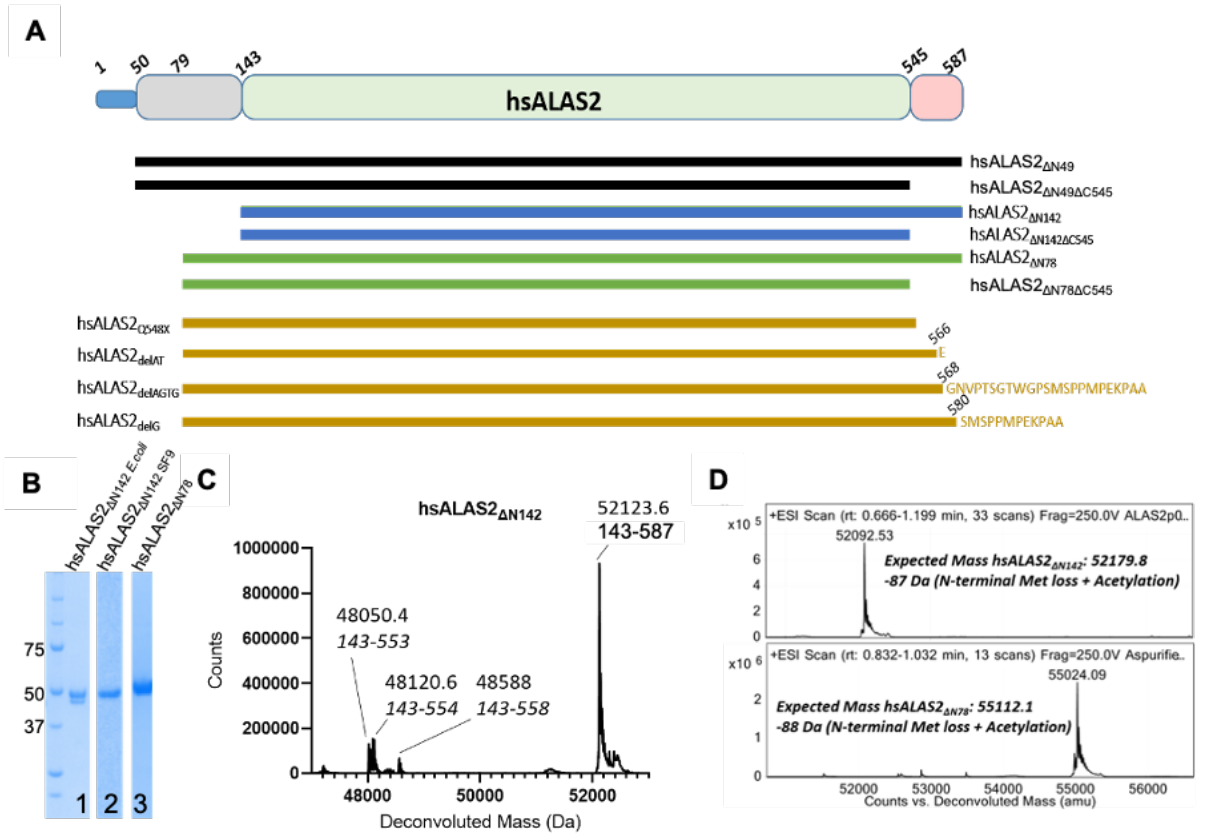


Figure 4-2: Expression of hsALAS2.

A: hsALAS2 is a 587-aa polypeptide (top) containing a mitochondrial targeting sequence (blue), an unstructured domain of unknown function (grey), a conserved catalytic domain (light green) and a C-terminal extension unique to eukaryotes (pink). Black bars denote attempted constructs that were insoluble during expression tests, green and blue bars represent soluble wild-type constructs in this study, and orange bars indicate constructs of disease-associated variants in this study. **B:** SDS-PAGE of hsALAS2 constructs purified from Ni^{2+} affinity chromatography, including *E.coli* hsALAS2$_{\Delta N142}$ (lane 1), SF9-expressed hsALAS2$_{\Delta N142}$ (lane 2), and SF9-expressed hsALAS2$_{\Delta N78}$ (lane 3)**. C:** Deconvoluted mass spectrum of *E.coli* expressed hsALAS2$_{\Delta N142}$ 143-587 shows predominantly intact protein (expected MW 52122.7 Da, measured MW 52123.6 Da) with three additional degradation peaks at low relative abundance to intact protein. Peak at 48588 Da corresponds to hsALAS2 143-558, peak at 48120.6 Da corresponds to hsALAS2 143-554 and peak at 48050.4 corresponds to hsALAS2 143-553. **D:** Deconvoluted mass spectrum of both purified hsALAS2$_{\Delta N142}$ and hsALAS2$_{\Delta N78}$ proteins expressed in SF9 cells and used in structural studies are shown to have a fully intact N-terminus and C-terminus, after expression and purification.

UV-visible light absorbance spectrums of the hsALAS2 purified constructs were next taken to assess the loading and tautomer state of the PLP cofactor. An absorbance peak at 335 nm indicates the presence of PLP in the active site. A second absorbance peak at 420 nm occurs if there is a protonated aldimine linkage between the cofactor and active site lysine[98]. *E. coli* expressed constructs reveal isolation of an enzyme cofactor complex in the protonated aldimine, shown by the absorbance peak at 420 nm (Figure 4-3A). Insect cell expressed constructs reveal isolation of an enzyme cofactor complex in an inactive form, shown by the lack of 420 nm peak and presence of the 335 nm peak (Figure 4-3A) [98]. The equilibrium between PLP states can be pushed in favour of the protonated aldimine bond formation by titration of PLP, as shown by increased fluorescence of the internal aldimine bond (Figure 4-3B). Internal aldimine bond formation through titration of fresh PLP also leads to increased thermal stability of the enzyme with similar binding affinities of PLP between the hsALAS2 constructs with and without the Ct-extension (Figure 4-3C). Interestingly, the construct lacking the Ct-extension (hsALAS2 79-545) was overall 2-3 °C less thermally stable than the construct containing the predicted regulatory domain (hsALAS2$_{\Delta N78}$, 79-587) (Figure 4-3C). The activity of the recombinant protein has been assayed by our collaborator (Section 4.3.3).

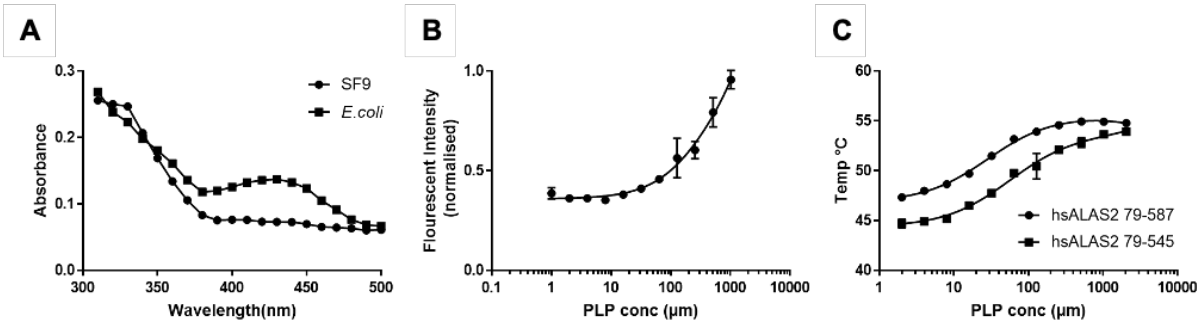


Figure 4-3: PLP loading of recombinant hsALAS2.

A: UV-Vis spectrum of hsALAS2$_{\Delta N142}$ comparing different ratios of the PLP tautomeric states of holoenzyme when isolated from insect cell (SF9) or *E. coli* cell expression systems. *E. coli*-expressed hsALAS2 produces holoenzyme in the protonated aldimine (shown by absorption maximum at 420 nm), whereas insect cell-expressed hsALAS2 likely yields holoenzyme in the deprotonated inactive state (shown by absorption maximum at 335 nm and lack of absorbance peak at 420 nm). Raw data points plotted from one experiment. **B**: A plot of fluorescence intensity increases with excitation at 440 nm and emission at 520 nm. Titration of excess PLP to insect cell expressed hsALAS2$_{\Delta N78}$ shifts the equilibrium towards the protonated form, shown by increased fluorescence of the internal aldimine. Data are presented as mean values +/- SEM from three technical replicates (n=3). **C**: Midpoint melting temperatures, determined by differential scanning fluorimetry, of insect cell-expressed hsALAS2$_{\Delta N78}$ and hsALAS2$_{\Delta N78\text{-}\Delta C545}$. Internal aldimine formation resulting from titration of PLP to both proteins is thermally more stable. Data are presented as mean values +/- SEM, from n=3 technical replicates.

4.2.2 Crystallisation and structure determination

X-ray crystallography was chosen to attempt to understand from a structural perspective how the regulatory C-t extension might impose an autoinhibitory effect on the enzyme. For crystallisation trials, the longest soluble construct, hsALAS2$_{\Delta N78}$ (expressed in insect cells) was chosen first as it retains the intact Ct-extension and retains a large portion of the previously uncharacterised N-terminal domain. In the as purified state initial crystallisation trials were unsuccessful. To optimise the as purified protein for crystallisation, prior to crystallisation set up, protein was treated with 5 mM hydroxylamine as described in crystallisation trials of scALAS[97]. Treatment with hydroxylamine removes the covalent linkage with the PLP cofactor[97] and allowed for the nucleation of poorly diffracting hsALAS2 micro crystals. These crystals were obtained through random matrix coarse screens, set up with a concentration of 5 and 10 mg/ml protein at 20°C and 4°C. Fine screening of protein

precipitant concentrations versus pH led to a high yield of microcrystals but no improvement in diffraction. These micro crystals were subsequently harvested for random matrix micro seeding experiments[99], with the aim to bypass the requirement to find conditions that allow for crystal nucleation and immediately probe conditions supportive of stable crystal growth by providing pre nucleated seeds. By this method, new conditions were discovered that led to larger crystals. A fine screen of this condition, varying pH and precipitant as well as seed concentration, consistently produced diffracting crystals in multiple drops. In addition, a control experiment, in parallel, without the use of seeds showed no nucleation and crystal growth.

From this batch of crystals, an anisotropic dataset diffracting to 2.7 Å resolution in the best direction and 3.6 Å in the worst was collected at the Diamond Light Source synchrotron. The data were truncated using the *staraniso* server[83]. Instead of a classic isotropic truncation based on a global I/σ(I) average in spherical shells, *staraniso* calculates locally-averaged value of I/σ(I), to define an anisotropic diffraction cut-off. This simultaneously prevents the loss of high-resolution data and incorporation of poor quality diffraction into the dataset, and is an improvement on a blanket ellipsoidal fitting in cases anisotropy deviates from an basic ellipsoid description[83].

The structure of hsALAS2$_{\Delta N78}$ from this dataset was determined by molecular replacement, using the rcALAS structure (PDB: 2BWN, sequence identity 49 %) as a search model. The Nt-extension, encompassing residues 79-142, displayed no electron density, while the catalytic core and the C-terminus could be accounted for. Intact mass spectrometry of purified protein revealed that the N-terminal residues were not degraded at the point of crystallisation (Figure 4-2D). Analysis of the crystal packing protomers also suggested there would be space for the missing N-terminal

density within the crystal lattice (Figure 4-4). Both factors therefore suggest that the N-terminal sequence is still present in the crystal but does not form in a single discrete conformation like the rest of the structure.

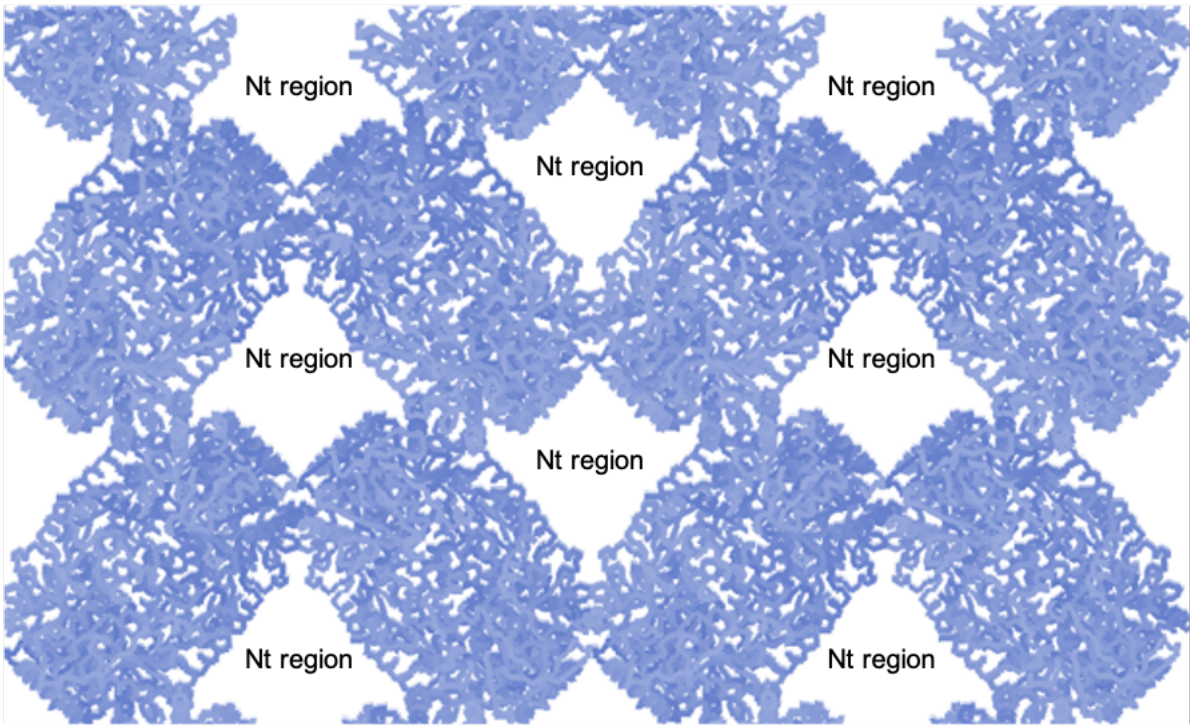


Figure 4-4: Crystal packing hsALAS2$_{\Delta N78}$

Blue backbone trace shows ALAS2 dimers arranged in crystal lattice. Empty regions labelled Nt region shows the area in which the Nt-extension could extend.

Due to this large disordered region, the shorter hsALAS2$_{\Delta N142}$ protein was next pursued for crystallisation. Using the same sample preparation methods, conditions that enabled growth of large crystals without requirement for seeding were identified in random course matrix screens. With this batch of crystals, the structure of the shorter construct was solved to a higher resolution of 2.3 Å and the diffraction was more isotropic than the previous lattice. The following structural description is based on this dataset. Data collection, model building and refinement statistics are shown in Table 4-1.

ataset	hsALAS2$^{\Delta N142}$	hsALAS2$^{\Delta N78}$
Beamline	Diamond beamline I04	Diamond beamline I24
Wavelength (Å)	0.9795	0.9795
Unit cell parameters (a, b, c)	125.77 Å, 107.7 Å, 75.71 Å	78.21 Å, 155.62 Å, 86.24 Å
(α, β, γ)	90.0°, 109.0°, 90.0°	90.0°, 98.45°, 90.0°
Space group	*C* 1 2 1	*P* 1 21 1
Resolution range (Å)	71.53–2.30	85.32–2.65[a]
Unique reflections	40,141	58,818
Rsym (%)	0.09 (0.71)	0.21 (0.76)
I/sig(I)	5.8 (1.0)	4.9 (1.5)
Completeness (%)	94.8 (95.0)	72.6 (19.3)[a]
Multiplicity	2.3 (2.3)	3.4 (3.0)
R_{cryst} (%)	0.21	0.21
R_{free} (%)	0.23	0.26
Wilson *B* factor (Å^2)	43.16	26.8
Average total *B* factor (Å^2)	55.2	25.8
R.m.s.d. bond length (Å^2)	0.003	0.004
R.m.s.d. bond angle (°)	0.69	0.68
Clashscore	6.7	5.78
Ramachadran favored (%)	95.6	95.6
Ramachandran disallowed (%)	0	0
Rotamer outliers (%)	0	0.15

1.Data from highest resolution shell shown in parenthesis.
2.Anisotropic data truncated in STARANISO using local I/sigI cut off at 1.2 results in the inclusion of data to 2.65 with low isotropic completeness.

Table 4-1: hsALAS2 Crystallographic table of parameters

4.2.3 hsALAS2 is a homodimer in crystal and in solution

The asymmetric unit of the crystal lattice contains one hsALAS2 homodimer (Figure 4-5). In order to confirm this assembly in solution, small angle x-ray scattering (SAXS) was used. SAXS measurements can be used to calculate information on the average shape, conformation and MW, giving details on the assembly state of proteins in a solution environment. Size exclusion chromatography coupled to small angle X-ray scattering (SEC-SAXS) was used to filter potential aggregates in sample preparation that may affect overall scattering of the ensemble of potential solution states.

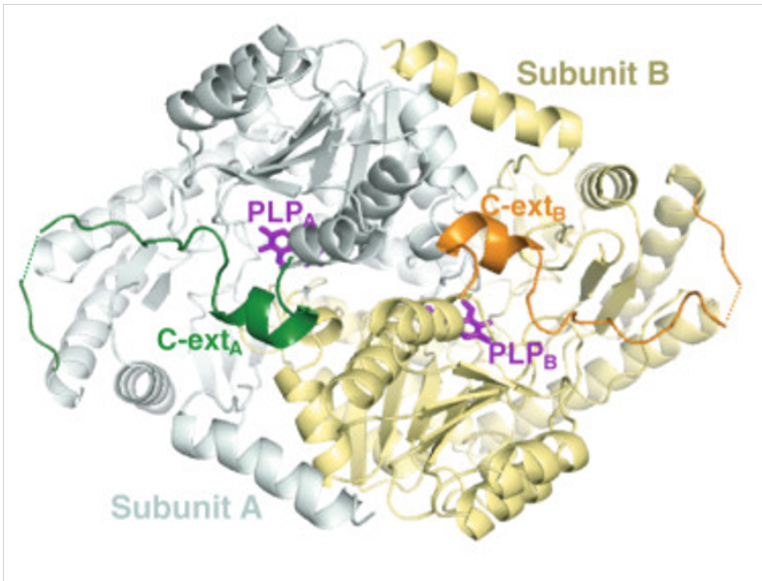


Figure 4-5: Structure of hsALAS2.

Cartoon representation of hsALAS2 homodimer (this study) composed of protomer A (catalytic domain in gray, Ct-extension in dark green) and the opposite protomer B (catalytic domain in yellow, Ct-extension in orange. PLP cofactor is shown in purple sticks.

The size exclusion trace for the crystallisable construct (hsALAS2$_{\Delta N142}$) was a single sharp peak indicating a homogenously sized enzyme was isolated for scattering measurement (Figure 4-6A). The elution peak containing frames with similar auto calculated radios of gyration (Rg) values were buffer subtracted, merged and displayed in a Log*10* intensity plot (Figure 4-6C) using the software *Scatter*[88]. From the plot MW can be calculated using the Porod Debye volume estimate, provided the protein is globular, or the Vc/Rg^2 method[100] as an invariant proportional to mass. The two methods suggest a MW of 110 or 92 kDa respectively, sizes similar to a homo-dimeric state with an expected mass of 104 kDa. Furthermore, *ab initio* modelling using *Dammif*[89] results in a volume which is a good match to the homo-dimeric crystal structure (Figure 4-6D), as expected from previous orthologue structures solved in scALAS and rcALAS.

The two monomers within the homodimer overlay with a Cα rmsd of 0.167 Å. The dimer interface is mostly formed by highly conserved residues within the catalytic domain, previously shown in the rcALAS structure[92]. In addition, the Ct-extension, not present in the rcALAS sequence or conserved in the scALAS structure, contributes three aromatic residues and hydrogen bonding interactions to the dimeric interface (section 4.2.5). To test whether in the human ALAS2 isoform the additional contacts provided by the Ct-extension is required for dimerization, SEC-SAXS experiments of the catalytic core domain construct hsALAS2$_{\Delta N142\text{-}\Delta N545}$ were performed. The size exclusion trace for this construct was also a single sharp peak indicating a homogenously sized enzyme is isolated for scattering measurement (Figure 4-6B). The elution peak containing frames with similar auto calculated ratios of gyration (Rg) values were buffer subtracted, merged and displayed in a Log*10* intensity plot overlaid with the ALAS2$_{\Delta N142}$ construct (Figure 4-6C)[89]. The two plots overlay well suggesting a similar overall size and shape. The Porod Debye volume estimate and the Vc/Rg^2 method[100] suggested a MW of 100 or 84 kDa respectively, consistent with a homo-dimeric state of an expected mass of 94 kDa. *Ab initio* modelling using *Dammif*[89] also resulted in a volume which is a good match to the homo-dimeric crystal structure of hsALAS$_{\Delta N142}$. (Figure 4-6E).

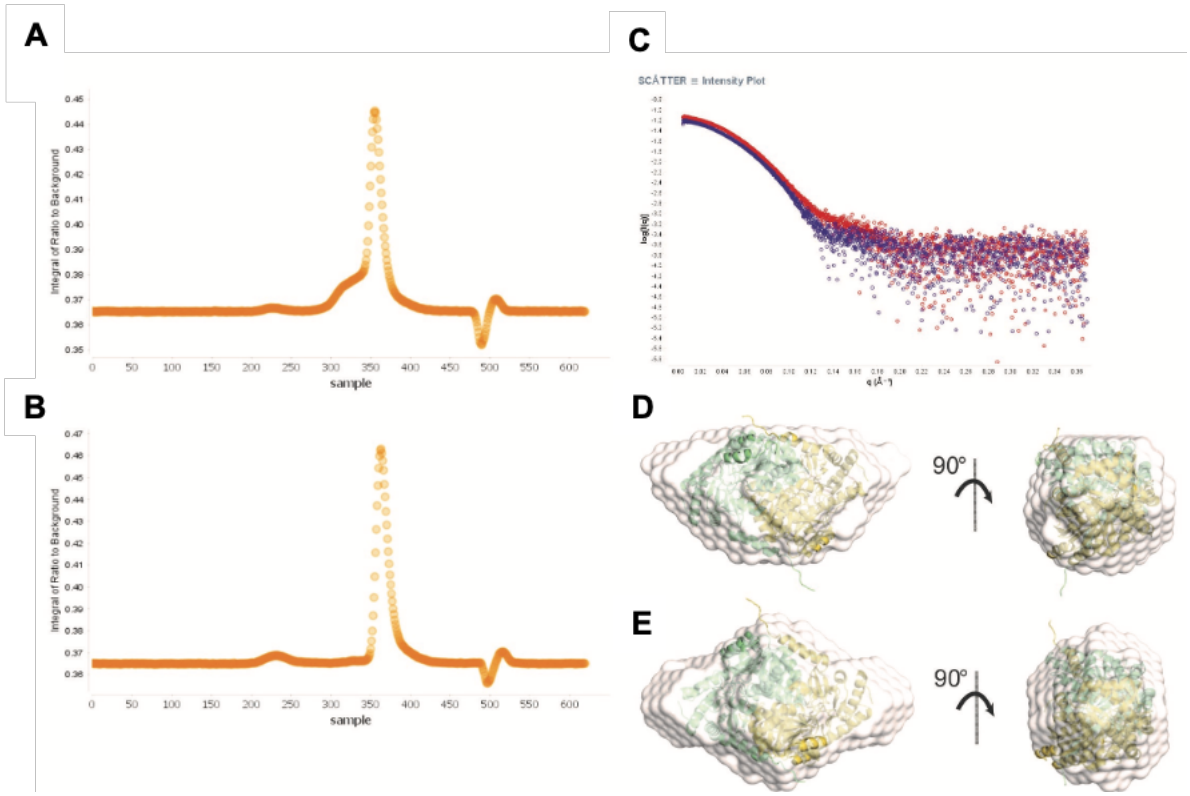


Figure 4-6: Small angle X-ray scattering studies of hsALAS2$_{\Delta N142}$ and hsALAS2$_{\Delta N142\text{-}\Delta C545}$

A: SEC-SAXS signal plot of hsALAS2$_{\Delta N142}$, with each data point (orange circle) representing the integrated area of the ratio of the sample SAXS curve to the estimated background (Y-axis) against elution SAXS frame (X-axis). **B:** SEC-SAXS signal plot of hsALAS2$_{\Delta N142\text{-}\Delta C545}$, construct that contains only the catalytic domain without the C-terminal extension. **C:** Log10 intensity plot of subtracted and merged SAXS frames. Red circles represent scattering data points from hsALAS2$_{\Delta N142\text{-}\Delta C545}$, and blue circles represent scattering data points from hsALAS2$_{\Delta N142}$. **D:** Orthogonal views of the *ab initio* envelopes calculated from SAXS data for hsALAS2$_{\Delta N142}$ **E:** Orthogonal views of the *ab initio* envelopes calculated from SAXS data for hsALAS2$_{\Delta N142\text{-}\Delta C545}$.

4.2.4 ALAS2 structure and conserved features

hsALAS2 has a highly similar catalytic domain to rcALAS. The two domains overlay with a Cα rmsd 1.1 Å, and 49 % overall sequence identity with each other. The core is also very similar in structure and sequence to scALAS, albeit to a slightly lesser degree (overall sequence identity 43% and Cα RMSD of 1.5 Å) (Figure 4-7A). The

three structures all belong to the same superfamily fold of type I PLP-dependent enzymes. The ternary structure of this conserved catalytic domain is formed by three discrete sub-domains. In hsALAS2 (Figure 4-7B), subdomain 1 is formed by aa 143-195 which fold into a single α-helix and a three-stranded, antiparallel β-sheet. Subdomain 2, formed by aa 196-439, is the largest domain folding into a seven-stranded β-sheet that is enclosed by nine α-helices in repeated β/α motifs. Finally, subdomain 3 is formed by residues 440-544 and includes a three-stranded antiparallel β-sheet, three α-helices and a fourth β-strand that also contributes to the β-sheet of sub-domain 1.

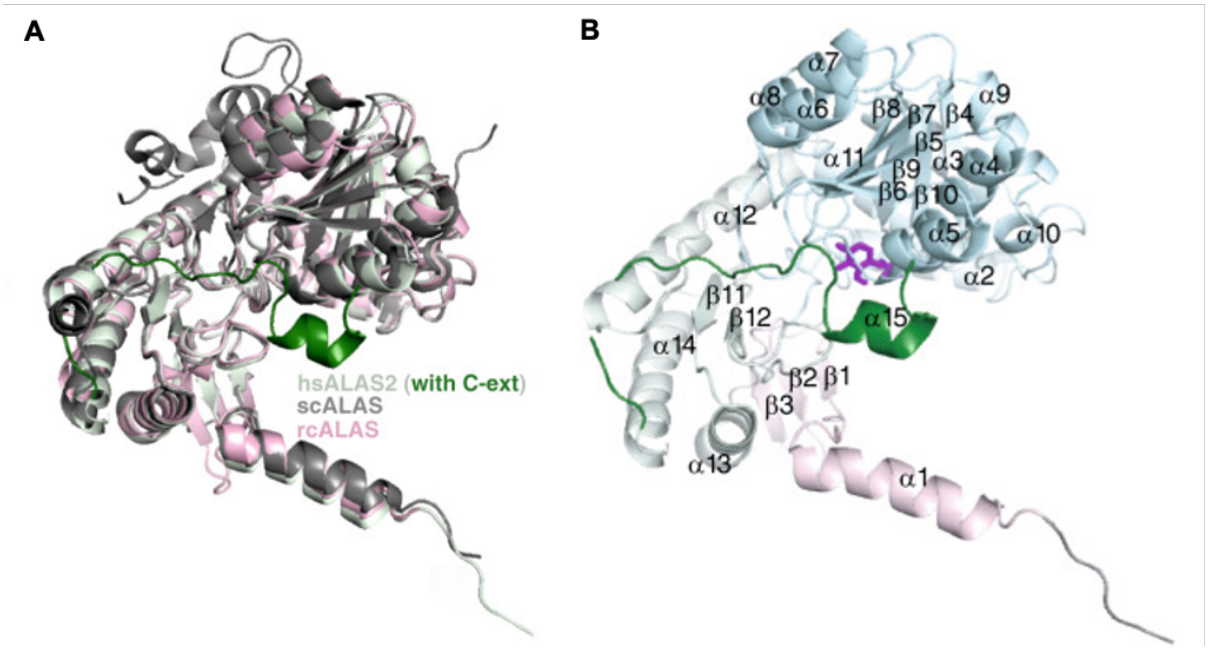


Figure 4-7: hsALAS2 ternary fold.

A: Structure superimposition of protomer from hsALAS2 (this study), scALAS (PDB 5TXR) and rcALAS (PDB 2BWN). **B:** Domain organization and secondary structure assignment for hsALAS2. Subdomain 1 is shown in pink, subdomain 2 in gray, subdomain 3 in cyan, and the Ct-extension in dark green.

The hsALAS2 obligate homodimer forms two active sites with residues contributed by both monomeric subunits at the symmetrical dimerisation interface. Within this

interface a single molecule of the PLP cofactor is bound to each active site (Figure 4-8). To achieve reliable crystal diffraction at high resolution, it was necessary to treat the enzyme with hydroxylamine. The effect of this is to convert any covalently bound PLP cofactors into a single PLP-oxime species that is non-covalently bound to the enzyme[97] (Figure 4-8, inset). In the structure, hsALAS retains the extensive non-covalent interactions with PLP that are highly conserved within ALAS orthologues as seen in sequence alignments (Figure 4-8). These interactions are contributed mainly from sub-domain 2 of the catalytic core where the PLP pyridinium is tightly bound between the hydrophobic residues His285 and Val359, above and below the plane of the ring respectively. The PLP ring makes additional hydrogen bond contacts to His360 and Asp357. The PLP phosphate forms hydrogen bonds to both Ser257 and Thr388 side-chains and Cys258 and Phe259 main-chains within its own subunit and, additionally, hydrogen bonds to both Thr420 and Thr421 side-chains and Thr421 main-chain of the opposing monomer, across the dimer interface (Figure 4-8).

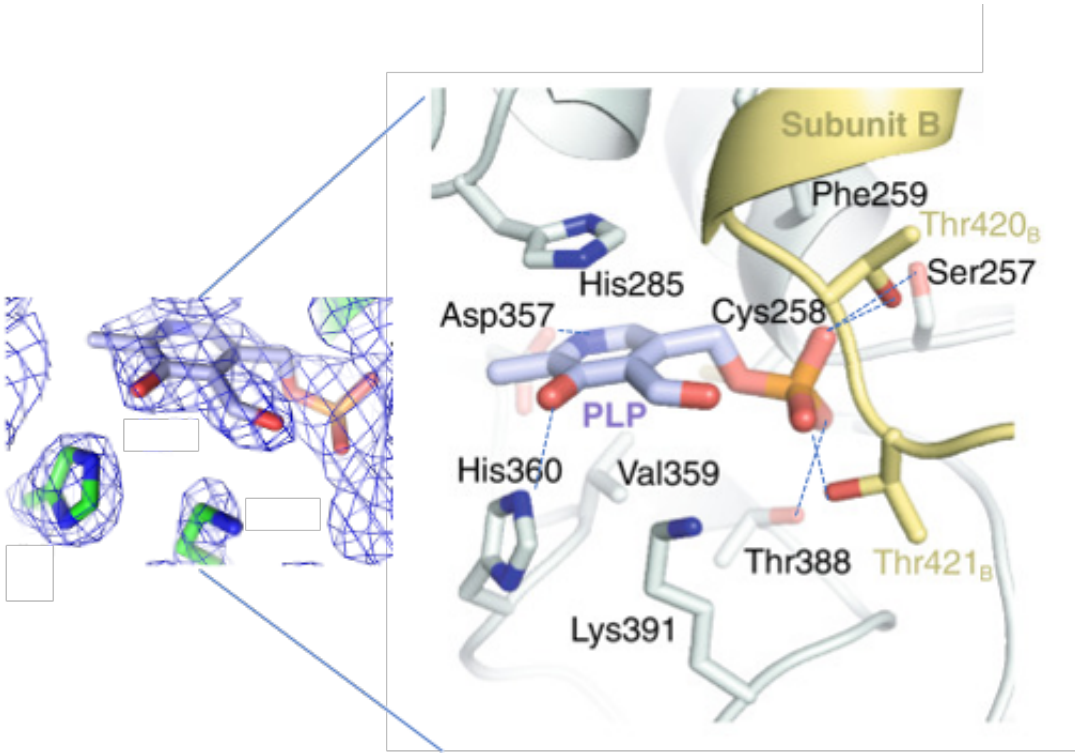


Figure 4-8: PLP binding site of hsALAS2.

PLP-interacting residues from monomer A are shown in gray and from opposing monomer B in yellow. Hydrogen bonding displayed with dashed lines. PLP is shown in mauve (carbon color). Inset shows the 2Fo − Fc electron density map of the active site region, contoured at 1σ, revealing residues of hsALAS2 that interact with PLP. As expected, no covalent linkage was observed between PLP and its Schiff base Lys391 in the crystal.

4.2.5 hsALAS2 C-terminus is a self-inhibitory loop

The most significant structural differences in the hsALAS2 enzyme compared to the previous orthologue structures can be found in the Ct-extension directly after the catalytic core domain. The 40-residue hsALAS2 Ct-extension (Figure 4-1) begins with a largely unstructured section formed by residues 544-557 which extends outward from α helix 14 at the protein periphery and forms hydrophobic interactions with the loop between α helix 14 and β strand 11 (Figure 4-9A). For a large part of this region, no clear electron density is resolved (residues 549-555) (Figure 4-9D), suggesting this unstructured loop is capable of adopting multiple conformations across the crystal lattice. Residues 560-568 form the next distinguishable portion of the Ct-extension (Figure 4-9A). Here the C-terminus forms a linear polypeptide that sits in a binding groove on the surface of the enzyme along the interface between sub-domains 2 and 3, directing the Ct-extension toward the dimer interface and active site. Main-chain amides and carbonyls from this sequence directly interact through hydrogen bonding with catalytic core residues. The main catalytic core residues that take part in these interactions, Arg479 and Glu513, are not conserved in either rcALAS or scALAS (Figure 4-1B), indicating that this particular binding mode has evolved in higher order eukaryotes only. The third segment of the Ct-extension is formed by residues 568-575 which folds into a two-turn helix labelled α15 (Figure 4-9B). The helix creates a lid, sitting directly on top of the PLP-bound

active site. In-depth examination of the interactions of helix α15 within the active site (Figure 4-9B) shows two glutamate residues, Glu569 and Glu571, are able to form charged interactions with Arg511, situated in catalytic subdomain 3, and with Lys299/Arg293 in catalytic subdomain 2. Three aromatic residues within α15, Tyr570, Tyr574 and Phe575, form π-π stacking interactions with themselves, and also act as an anchor for the Ct-extension to the dimerisation interface through further hydrophobic interactions with Phe267, Thr268 and Ala415 of the opposite dimer subunit. Hydrogen bonds are formed from the carbonyl group of residue Phe575 to the side chain of residue Lys271 and also from the carbonyl group of residue G576 to the side chain of residue Arg293; both interactions are again at the dimer interface between opposing subunits. Lastly, residues 575-578 direct the Ct-extension back to the protein exterior, where the last 9 residues, 579-587, are surface-exposed without clear electron density to enable modelling. These terminal 9 residues also create one of the largest sequence divergences between ALAS1 and ALAS2 (Figure 4-1B, Figure 4-9D). An omit map of the entire Ct-extension confirmed the correct interpretation of electron density in this region (Figure 4-9D).

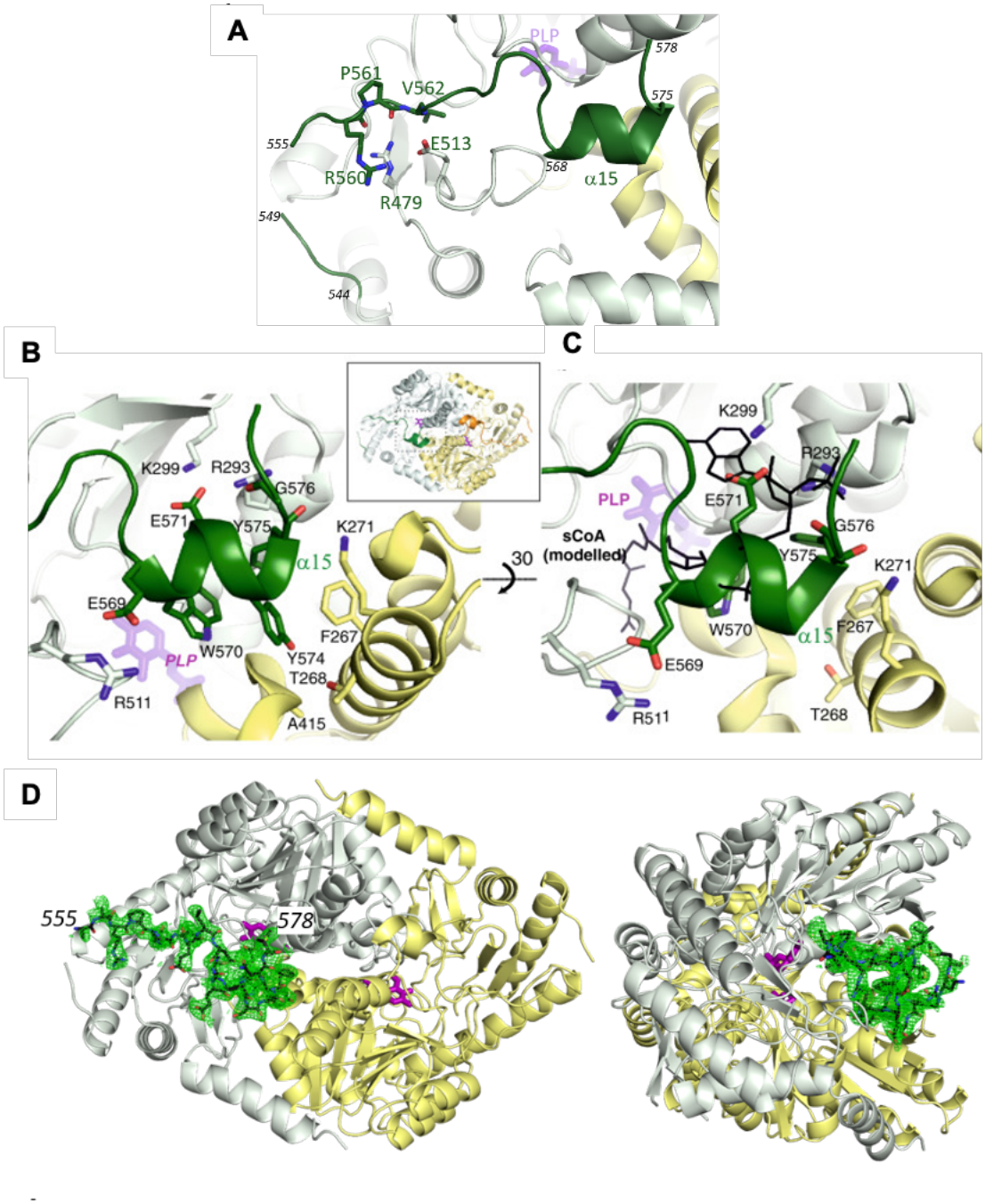


Figure 4-9: Interactions of the Ct-extension with catalytic core.

A: Topology of an hsALAS2 Ct-extension (dark green), packing against the catalytic core of its own protomer (light green) and of the opposing protomer in the homodimer (yellow). PLP is shown as purple sticks. Interacting residues of the catalytic core and the unstructured portion of the Ct-extension are displayed. Residues 550-554 are disordered in the structure and hence not modelled. **B:** Interactions of helix α15 from the Ct-extension (dark green) with the catalytic core of monomer A shown in light green and the opposing monomer B shown in yellow. **C:** Similar view as panel A,

related by 30° rotation along the horizontal axis. The succinyl-CoA (S-CoA) ligand, modeled into hsALAS2 from rcALAS (PDB 2BWO), is shown in black lines to highlight steric clashes with helix α15. **D:** Composite 2mFo-DFc omit map (by Phenix software) of the hsALAS2 Ct-extension, contoured at 1σ, shown for one subunit of the homodimer. Black lines represent Ct-extension residues 555-578 built onto the density. Two different views rotated by 90 along the Y axis are presented.

The rcALAS structure was also solved in the presence of the substrate succinyl-CoA (PDB 2BWO)[92]. This structure revealed key residues that bind and orient the succinyl-CoA into the active site pocket, presenting its thioester bond to the bound PLP cofactor. Previous studies have shown that the interactions of ALAS with succinyl-CoA are mostly invariant across phyla, suggesting that hsALAS2 has evolved a similar binding mode to that of rcALAS[92]. However, after superimposing the hsALAS2 and succinyl-CoA-bound rcALAS structures, the two-turn helix (α15) of the Ct-extension occupies the succinyl-CoA binding pocket, completely blocking the access of succinyl-CoA to the hsALAS2 active site (Figure 4-9C). In this instance, α15 would create steric clashes with the ADP-diphosphate moiety of a modelled succinyl-CoA molecule. It can therefore be assumed that the hsALAS2 structure is representative of a conformation of the Ct-extension that inhibits binding of succinyl-CoA to the active site. This inhibition is not possible in the previously solved isoforms of ALAS as they either lack the regulatory Ct-extension (rcALAS) or have evolved a completely different extension and binding mode to the catalytic core (scALAS) (section 4.2.7).

In the hsALAS2 structure, multiple interactions of helix α15 with the catalytic core mediate this steric blockade (Figure 4-9B, C). This includes the formation of salt bridges between Glu571 with both Lys299 and Arg293, which overlap with the expected position of the adenine group of the succinyl-CoA ADP moiety. In addition,

the 3'-phosphoadenosine moiety is blocked by a hydrogen bond between the Glu576 carbonyl and Arg293. Finally, the salt bridge between Glu569 and Arg511 helps bind α15 so that the two-turn helix directly blocks the diphosphate and pantothenic acid groups of CoA.

Of particular interest, Arg511 is positioned in a previously described mobile active site loop[101,31,92] that is comprised of residues 505-514, between β strands 11 and 12. In addition to the salt bridge with Glu569, Arg511 forms a salt bridge with Asp159 from helix α1, where the equivalent interaction in the rcALAS structure is heavily implicated to play a role in succinyl-CoA-induced conformational change during catalysis[101,31,92] (Figure 4-10). During catalysis, loss of interaction between Arg511 of the mobile loop and Asp159 allows transition of ALAS to an open state compatible with succinyl-CoA binding. Therefore, this additional interaction of the two-turn helix, present only in metazoans, could increase affinity for the mobile loop to remain in a closed state for hsALAS2.

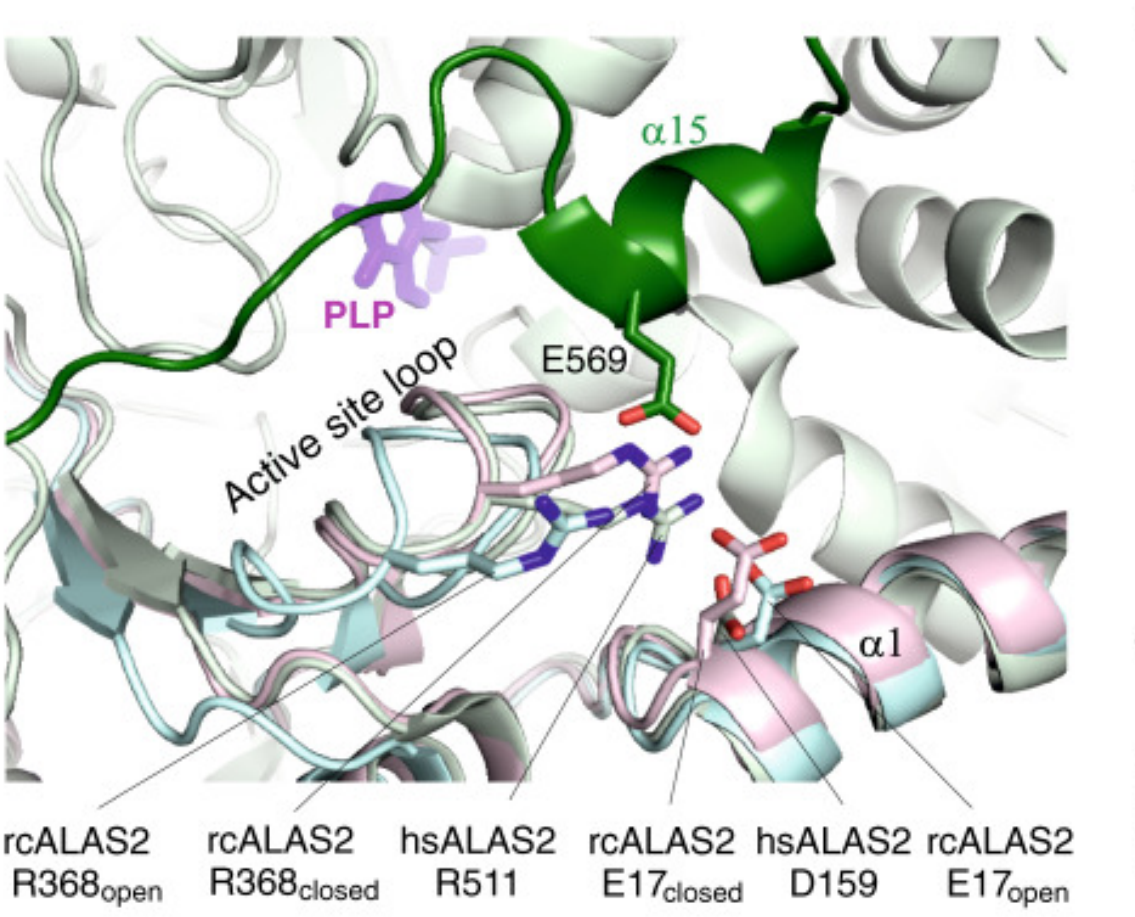


Figure 4-10: C-terminal extension contacts mobile loop.

View of the hsALAS2 active site loop conformation within one monomer (catalytic domain light green, Ct-extension dark green). The active site loop from the open (cyan) and closed (pink) forms of rcALAS (PDB 2BWN) are overlayed onto hsALAS2.

4.2.6 Role of mobile active loop

The conformational changes within the mobile loop are also required at the point of product release. Current understanding[101,31,92] of the mechanism that mediates these changes during the catalytic cycle suggest that the mobile loop first adopts the open conformation during succinyl-CoA binding and subsequently transitions to a closed form upon CoA release. This form, induced by the salt bridge formation of Arg511 with Asp159 (human numbering) (as seen in the rcALAS open and closed state

structure, PDB 2BWN) allows for the intermediate catalytic steps, which include decarboxylation followed by re-protonation to regenerate the internal aldimine, to take place. Finally, the enzyme is required to transition back to the open form to allow for product release. It is this transition that is described to be the rate determining step during catalysis[101,31,92]. The interaction between the Ct-extension and the active site loop, via Arg511, suggests that the mobility of the active site loop during both succinyl-CoA binding and product release must require significant conformational change of the Ct-extension. Therefore, these two functional regions must act in concert to drive the ALAS2 catalytic cycle.

To understand the role of the Arg511 interaction with the Ct-extension, site-directed mutagenesis was used on the crystallised construct, hsALAS2$_{\Delta N142}$. An Arg511Glu (R511E) variant protein was generated, designed specifically to abolish the charged interactions of Arg511 with both Glu569 in the Ct-extension and Asp159 in the active site loop. After affinity purification from *E.coli* expression system, the R511E variant protein appeared to be more prone to proteolysis at the Ct-extension in comparison with the wild type protein, as shown by SDS-PAGE (Figure 4-11A). Here lane one (labelled with circle) shows that hsALAS2$_{\Delta N142}$ purified from *E. coli* is represented by two bands, the upper band is expected to be intact protein while the lower represents a C-terminally degraded form, as confirmed previously in this report by mass spectrometry (Figure 4-2C). Conversely, lane three (labelled with a triangle), containing the hsALAS2$_{\Delta N142}$ R511E mutation is predominantly isolated in the degraded form. Intact mass spectrometry (Figure 4-11B) identifies this increased degradation to also occur at the Ct-extension with no intact protein detected. This would suggest that the interaction of the Ct-extension with the catalytic domain active site loop has been significantly disrupted, leaving the C-terminus in a more

exposed position allowing for increased degradation. Consistent with this, the hsALAS2$_{\Delta N142}$ R511E variant is >2 ^{o}C more thermolabile than hsALAS2$_{\Delta N142}$ WT protein (Figure 4-11C). hsALAS2$_{\Delta N142}$ R511E instead has the same melting temperature as hsALAS2$_{\Delta N142\Delta C545}$ (incorporating aa 143-545), whereby the Ct-extension has been fully truncated from the enzyme (Figure 4-11C). The increased thermolability of WT hsALAS2 $_{\Delta N142\Delta C545}$ remains unchanged when the R511E mutation is incorporated into this truncated construct as a control experiment (Figure 4-11C). In addition, the hsALAS2 $_{\Delta N142\Delta C545}$ WT and R511E variant behave similarly on SDS-PAGE with no signs of degradation (Figure 4-11A, Lane 2 and 4). Therefore, the reduced stability can be attributed specifically to the loss of interactions with the Ct-extension to the catalytic core only and not due to alternative destabilisation effects of R511E within the catalytic core (Figure 4-11C). The impact of R511E variant on hsALAS2 enzyme activity was also determined by our collaborator (see Section 4.3.3).

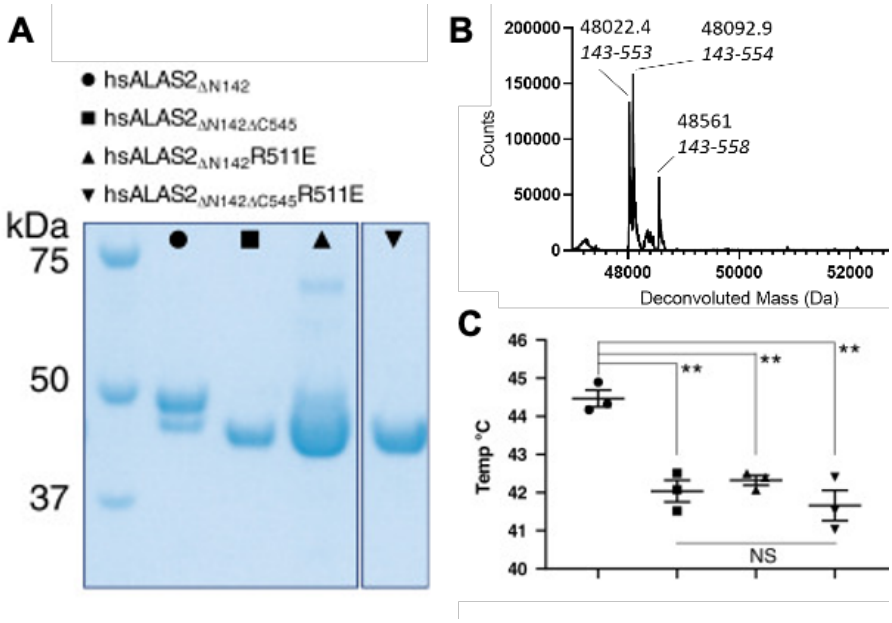


Figure 4-11: Role of C-terminal extension in structural and functional integrity:

A: SDS-PAGE of the purified samples from hsALAS2 143-587 WT (square) and R511E variant (up triangle), as well as hsALAS2 $_{\Delta N142-\Delta 545}$ WT (circle) and R511E variant (down triangle). **B:** Deconvoluted mass spectrum of *E.coli* expressed R511E variant of hsALAS2$_{\Delta N142}$ shows no signal for intact protein (expected MW 52095.49 Da) and high abundance of the three degradation peaks seen previously in WT (Figure 4-2C). Peak at 48561 Da corresponds to R511E hsALAS2 143-558, peak at 48092.9 Da corresponds to R511E hsALAS2 143-554 and peak at 48022.4 corresponds to R511E hsALAS2 143-553. **C:** Midpoint melting temperatures, determined by differential scanning fluorimetry, of the four samples in panel b. Error bars represent mean values ±SEM from n = 3 technical replicates. **p-values determined by two sample t-test; p = 0.0026, 0.0011, and 0.0002 from left to right.

4.2.7 hsALAS2 and scALAS have different Ct-extension

A comparison of the hsALAS2 structure with the previously described scALAS structure[97] shows that their equivalent Ct-extensions not only differ in sequence (Figure 4-1B) but also in the location and conformation within the quaternary structure of the ALAS enzyme. The point of structural divergence occurs immediately after the last helix (α14) of the catalytic domain, at position Leu545 in hsALAS2 or Leu489 in scALAS. The hsALAS2 extension threads across the catalytic core between interface of subdomain 2 and 3, and subsequently shielding the substrate binding site from the exterior solution (Figure 4-12). This differs significantly in scALAS where a two-turn helix, that is not conserved in metazoan ALAS constructs, instead directs its Ct-extension back along the periphery of sub-domain 3 (Figure 4-12A). The scALAS Ct-extension then packs against a 22 amino acid insertion within sub-domain 2 of the catalytic core domain that is unique to the scALAS sequence (302-323, scALAS numbering) and not present in hsALAS2 (Figure 4-12B). Its Ct-extension then continues to wrap around the homodimer surface reaching as far as the N-terminal sub-domain 1 of the opposing monomer. Unlike hsALAS2, the Ct-extension in scALAS does not directly interact with residues that

form the substrate binding site, nor does its presence indirectly alter the conformation of any of the residues within the active site (Figure 4-12C), leaving it exposed as seen in the rcALAS structure that lacks any Ct-extension (Figure 4-12D). Due to the different lengths, lack of sequence conservation between the two Ct-extensions, and the evolution of completely different binding residues within the catalytic core domains, it is highly unlikely that the two structures represent interchangeable alternate conformations of the Ct-extension.

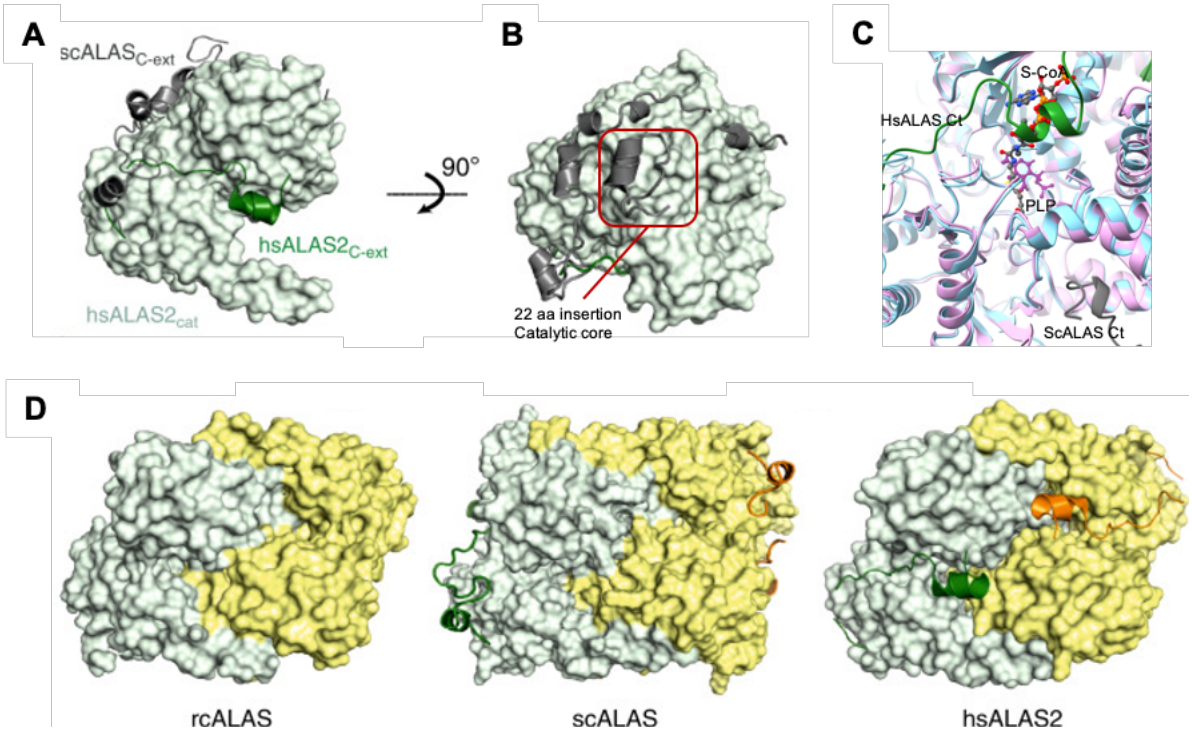


Figure 4-12: comparison of hsALAS2 ct-extension with orthologue structures

A: hsALAS2 monomer with catalytic domain in surface representation (light green) and Ct-extension in cartoon representation (dark green). The scALAS Ct-extension is also superimposed in cartoon representation (gray). **B:** Same as A, rotated 90 degrees along the X axis. **C:** scALAS (pink) superimposed to rcALAS (cyan) in cartoon representation. scALAS Ct-extension shown in grey does not bind nor alter the active site conformation of ALAS. hsALAS2 Ct-extension shown in dark green for reference. **D:** rcALAS (PDB 2BWN), scALAS (PDB 5TXR) and hsALAS2 homodimers. Catalytic domains are shown in surface representation (monomer A light green, monomer B yellow). Ct-

extensions are shown in cartoon representation for scALAS and rcALAS (dark green for monomer A, orange for monomer B).

4.2.8 Revised understanding of disease-causing mutations of hsALAS2

The hsALAS2 crystal structure provides a new template to understand the molecular basis of disease-causing mutations for XLSA and XLP, the two inherited disorders associated with *ALAS2* gene mutations. To date, over 82 missense changes affecting 69 different residues of hsALAS2 have been identified as causing XLSA and mapped onto the hsALAS2 (Figure 4-13). These substitutions are largely concentrated in the catalytic core, and many of these have previously been interpreted in the context of the rcALAS structure[92], revealing putative molecular defects such as disruption of the dimerization interface, direct disruption of catalytic regions, or a global destabilization of the enzymatic fold.

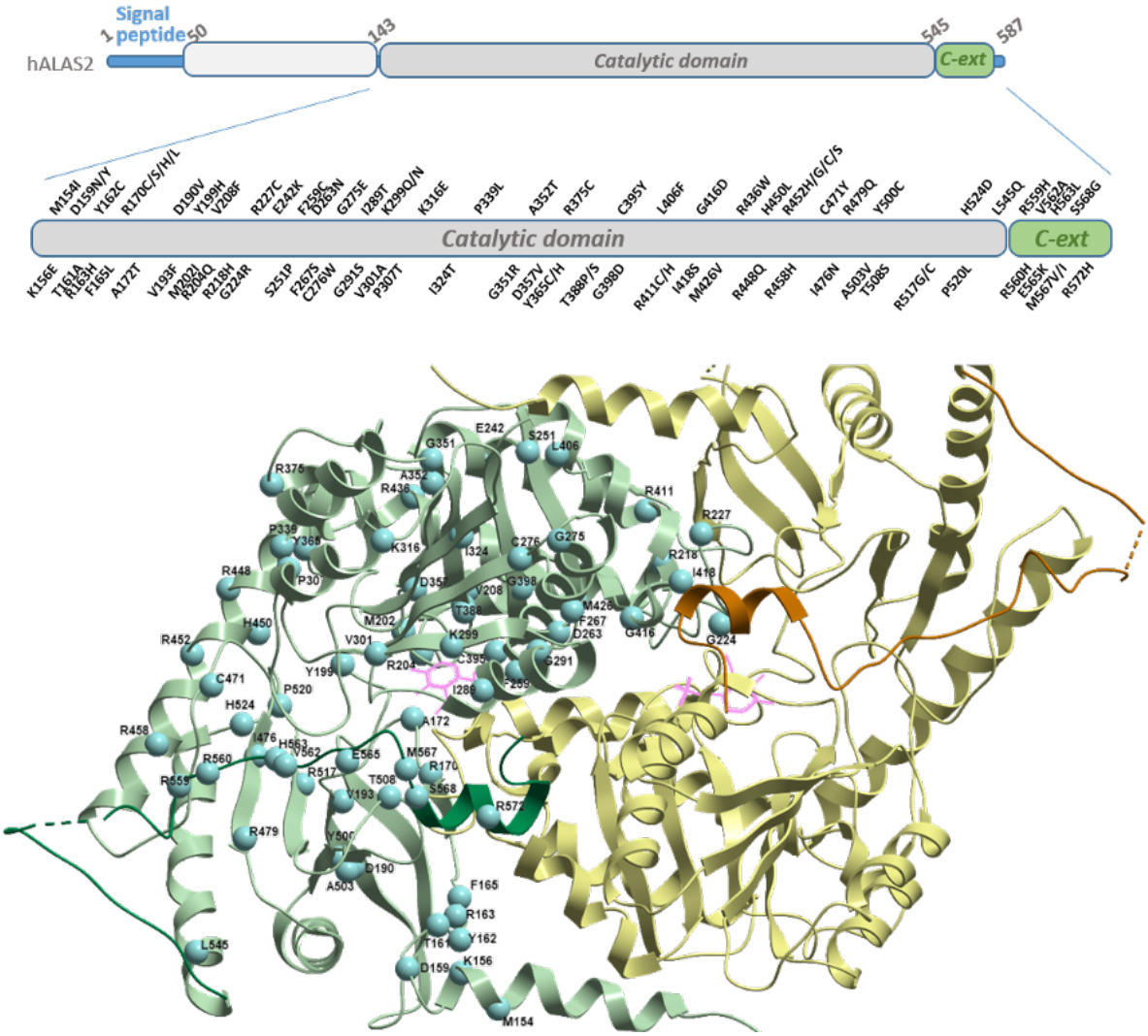


Figure 4-13: Mapping of disease-causing missense mutations onto hsALAS2 structure.

Sites of missense mutations that lead to the XLSA disorder (spheres) are mapped onto a domain organisation schematic (top) and the hsALAS2 structure shown as a homodimer (green and yellow subunits) (bottom).

The hsALAS2 Ct-extension, structurally characterised in this study, provides new insights on some previously characterised mutations in the catalytic core, namely Asp159Asn and Asp159Tyr. These substitutions would clearly disrupt the aforementioned salt bridge network between Asp159, Arg511 and Glu569 in the hsALAS2 structure, leading to a decoupling of the concerted interaction of the Ct-extension and the catalytic core during the transitioning between open and closed

states (Figure 4-10). The importance of this salt bridge network has been already demonstrated by the earlier reported results from DSF, SDS-PAGE and mass spectrometry of the R511E variant (Figure 4-11). With the Asp159Asn/Tyr variants, deficiency of the enzyme, which is the opposing effect on enzyme activity compared to the R511E variant, is likely observed because the Arg511 to Glu569 steric block is still retained and perhaps even more optimally positioned for interaction.

The gain of function (GOF) XLP mutations reported to date are all located within the final exon that encodes the hsALAS2 Ct-extension. Three of the reported indel mutations (p.Gln548*, c.1699_1670delAT and c.1706_1709delAGTG)[37] result in the deletion of the entire two-turn α15 helical region that blocks the ALAS2 active site. A fourth known indel mutation (c.1734delG)[37] leaves the Ct-extension helix α15 intact, though it results in a frameshift and an additional 12-aa sequence downstream of that region (Figure 4-2A yellow bars). Disruption of the α15 helix through these mutations is expected to alleviate the restriction of conformational change imposed by the interaction of the Ct-extension with the active site loop, providing an explanation for the increased rates of product release as shown previously in stopped-flow experiments[29]. To verify this hypothesis, variant proteins harbouring the Ct-extension deletion or frameshift changes were generated from recombinant expression and steady-state kinetics measured by our collaborators (see Section 4.3.3).

4.3 ALAS2 fragment screening campaign reveals starting point for C-terminal stabilisation

4.3.1 Rational for the fragment screening approach

As described (section 2.2.3), ALAS2 could be a target for substate reduction therapy to alleviate flux into the haem biosynthetic pathway for treatment of downstream

erythroid specific porphyria's. Additionally, inhibitors developed for this purpose also have the potential to treat XLP itself, because the main biochemical hallmark is hyperactivity of ALAS2 enzyme. To date there are no reported inhibitors or chemical starting points specific for hsALAS2. To this end, crystallography-based fragment screening was used to identify chemical matters that are capable of binding functional regions of ALAS2. This type of experimental set up, accessing high throughput facilities at the Diamond Light Source XCHEM facility at Beamline I04-1[102,103], can routinely be used to screen thousands of low molecular weight (<300 Da) small molecule fragments. One successful example of this was the development of nano molar inhibitor of the E3 ligase Keap1 that blocks protein-protein interaction with the substrate NRF2[104].

The screening approach uses X-ray crystallography to find fragment hits and visualise their interaction sites to identify druggable protein hotspots. The premise of this screening approach is to survey a very broad chemical space initially with low affinity binders ('fragments'), to identify as many potential ligand-protein interactions as possible[102]. Fragment hits can subsequently be converted to chemical leads by growing a single fragment or by combining multiple fragments[105]. For example, the Keap1 campaign combined multiple fragment hits which clustered around substrate protein protein interaction hotspots[104]. The availability of fragment-bound crystal structures, an integral part of the readout of this screening approach, will help to facilitate this bottom up approach to small molecule drug design.

Here, the Diamond-SGC poised library of fragments was screened against ALAS2 crystals. The library consists of 776 fragments, and importantly each fragment is associated with a range of commercially available or easily synthesised small molecules for potential follow up. The construct used for crystallisation was the same

as the original ALAS2 dataset (hsALAS2 $_{\Delta N142}$) and the crystallisation conditions applied were nearly identical, apart from the use of micro ALAS2 crystals to seed nucleation in each drop. This additional step was necessary to reliably obtain the required number of well diffracting crystals. The following crystal handling and data collection were carried out by Gustavo Arruda Bezerra and William Foster in our lab. Downstream data processing was carried out by Gustavo and me. My contributions presented here also include Interpretation and annotation of data. Each crystal was soaked with a different compound from the diverse fragment library and mounted for data collection. A total of 295 X-ray data sets were collected at the Diamond Light Source beamline I04-1, with resolution ranging between 1.44-1.93 Å. Diffraction data were auto processed in the *xia2*[103,82] pipeline and molecular replacement (using the hsALAS2 $_{\Delta N142}$ structure) and initial refinement on all suitable datasets run in parallel using *DIMPLE*[103,106]. After phasing, *PanDDA*[103,107]. was used to identify fragment densities in difference maps. *PanDDA* performs local alignment of electron density maps from all of the phased structures generating spatial Z-score maps which quantify relative variability at each position in every dataset. This processing flattens random solvent noise allowing for the identification of low occupancy fragment hits based on density modification relative to fragment free datasets. After fragment placement, structures were refined using the *REFMAC* programme[86].

4.3.2 Analysis of ALAS2 fragment hits

19 datasets had fragments bound to the ALAS2 homodimer, (fragments binding at packing interfaces excluded) The 19 fragments clustered into 3 different hotspot regions on the hsALAS2 enzyme (Figure 4-14).

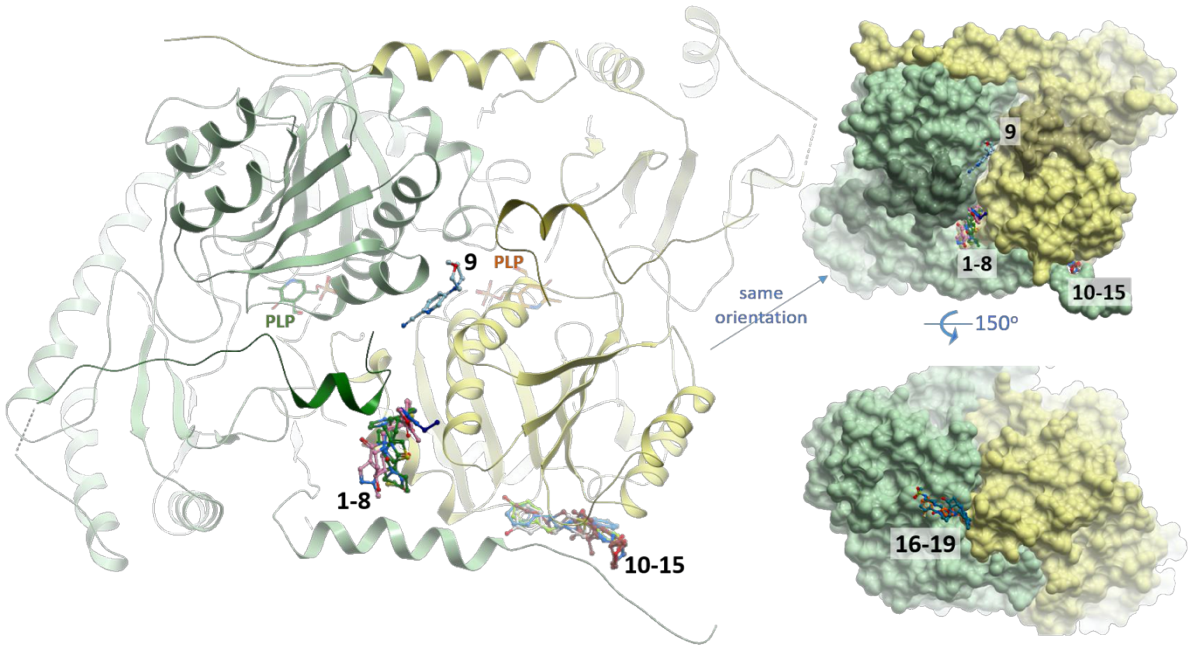


Figure 4-14: hsALAS2 fragment binding pockets.

Fragments are clustered into three distinct regions of the protein. The two subunits of an ALAS2 homodimer are coloured green and yellow. Inset shows two rotated views of ALAS2 in surface representation.

The first and most biologically significant of these hotspots bound fragments 1-8 (Figure 4-14, left and top inset). Each of these fragments were positioned in a pocket formed between Ct-extension helix α15 and the catalytic core helix α1, with additional interactions contributed by the catalytic core helices α4, α9 and α10 from the opposing subunit (Figure 4-15). This pocket sits amongst the stabilising interactions of the Ct-extension with the catalytic core at the active site. For example, directly adjacent to the fragments is the salt bridge network that forms between Glu569 of α15, Arg511 of the mobile active site loop, and Asp159 of the catalytic core helix α1 (Figure 4-15). Also close to the fragment pocket are the inter-subunit contacts which include hydrophobic interactions between Tyr570, Tyr574 and Phe575 of one subunit to residues Phe267 and Thr268 of α4 in the opposing subunit, and additionally hydrogen bond between carbonyl group of Phe575 and

Lys271 of α4 in the opposing subunit (Figure 4-9B). Given that this region is already important for stabilising the Ct-extension to the catalytic core and across the dimerization interface, the additional interactions of a small molecule in this pocket could be hypothesised to enhance this stabilising effect. This increased affinity of the Ct-extension to the catalytic core and across the dimerization interface may help to lock the enzyme into the auto-inhibitory state.

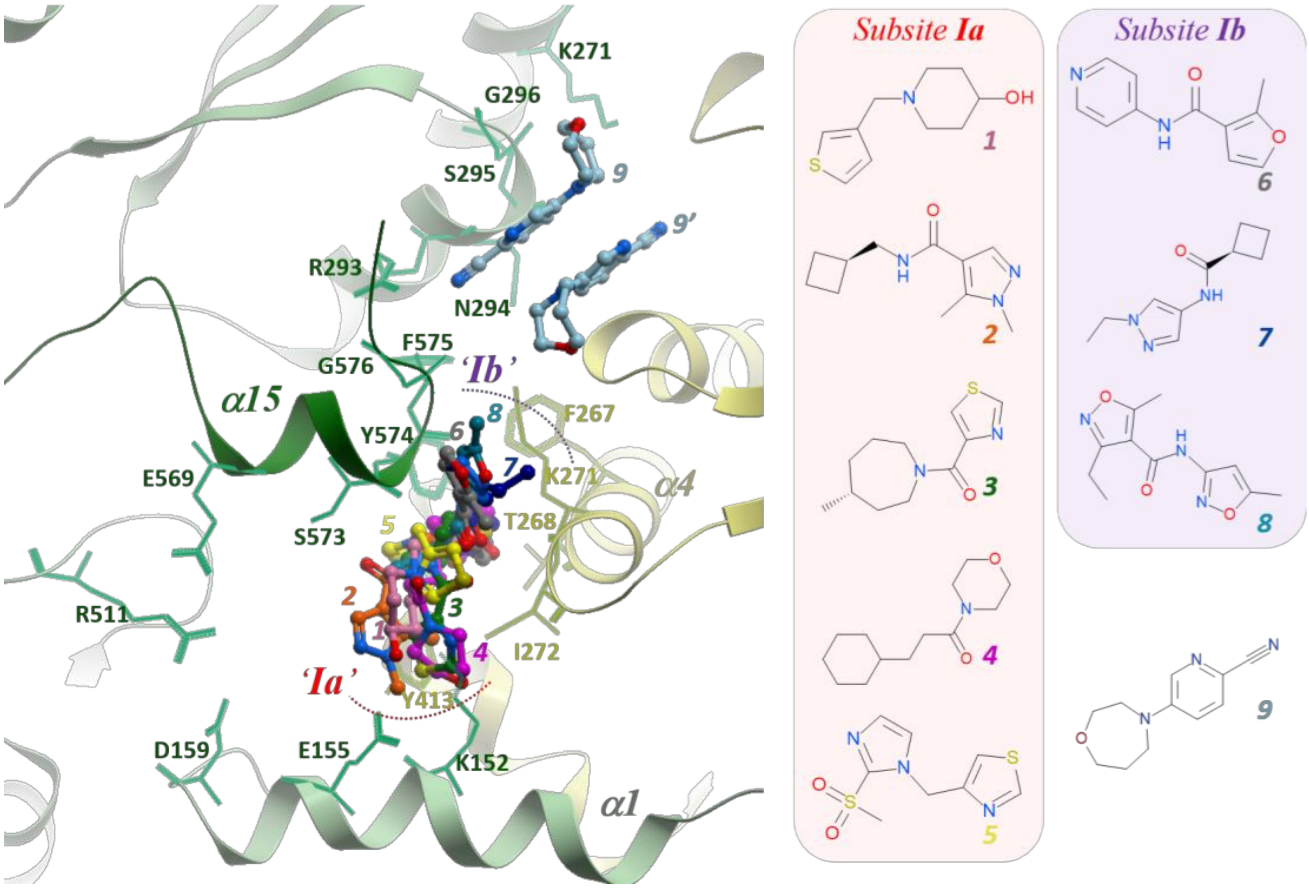


Figure 4-15: Cluster 1 of fragments around the C-terminal extension.

Close-up view of the C-terminal extension region of one subunit (green) and its neighbouring subunit (yellow) is shown. For fragment 9, the second molecule (9') bound to the yellow dimeric subunit is also shown. The labels of the 2D chemical structures of fragments 1-9 are individually coloured-coded according to the stick representation in the left image.

The interactions of the fragments within the binding pocket itself are largely hydrophobic. Each of the fragments are all sandwiched directly between Tyr574 from

helix α15 and Ile272 α4. The aromatic side-chain of Tyr574 (helix α15) adopts an alternate conformation when bound to fragments 2-4, relative to the other fragments and fragment free datasets. Additional hydrophobic interactions with all 8 of the fragments are contributed by residues Phe267, Thr268 and Lys271 again from α4 and Tyr413 from α9 of the opposing subunit.

In addition to this stabilising affect across the dimer interface, fragments 1 to 5 also extend into a further subsite (referred to as '*Ia*' in Figure 4-15), interacting directly with Glu155 and Lys152 from helix α1 of the catalytic core. Again, it is conceivable that these fragment interactions may enhance the stabilising effect of the salt bridge network Glu569, Arg511 and Asp159 that also anchors the Ct-extension into the active site. (Figure 4-15). Fragments 6 to 8 also occupy a further subsite ('*Ib*' in Figure 4-15) in direct proximity to the last turn of helix α15, at residues Phe575 and Gly576, although no specific interactions are discernible between the protein and portion of the fragment that is directed into this subsite.

A single fragment (9, Figure 4-15) was also modelled at the C-terminal helix α15, this time positioned on the opposite side with respect to the first 8 fragments, directly adjacent to the hydrogen bond interactions between Gly576 of the Ct-extension and Arg293 of the catalytic core. This site sits directly on the 2-fold symmetry axis of the hsALAS2 homodimer and therefore, two fragment 9 molecules are displayed next to each other. One side of the fragment makes extensive hydrophobic contacts formed by residues Asn294, Ser295, Gly296 and Lys271. The fragment nitrile group also hydrogen bonds with the main chain carbonyl of Phe575 (in C-terminal helix α15) and guanidinium side-chain from Arg293. It is therefore likely that this fragment may again act to stabilise interactions of the Ct-extension with the catalytic core of the protein further anchoring it over the active site. The same nitrile group also extends

towards the fragment pocket subsite Ib. The opposite face of the fragment forms van der Waals contact with the symmetry related fragment bound to the opposing subunit in the dimer. Due to the close proximity of fragment 9 to the cluster of 8 fragments in site Ia and Ib it may be possible to design follow up compounds that could join common groups.

The second cluster of fragments 10 to 15 is situated again at the homodimeric interface, but at a site that is distant from the enzyme's active site and Ct-extension (Figure 4-14). These fragments are accommodated between the stretch before helix α1 (partially incorporating the His tag linker and TEV cleavage site) of one subunit and the central αβ core of the other subunit. This extensive surface, structurally conserved among ALAS orthologues, is important for maintaining the dimer assembly through several hydrophobic interactions (e.g. Asp145, Asp383 backbone, Ile384, Ser402, Thr403 and Leu 406). A future angle for development of these fragments, to this end, might be to evaluate their potential in disrupting hsALAS2 dimerisation, an essential catalytic feature of the enzyme. However, a small molecule campaign to achieve disruption of a very extensive dimer interface would be difficult due to the high affinity of the obligate ALAS homodimer interaction.

The final four fragments 16-19, form a third cluster bound within a pocket of the enzyme situated on the opposite face of the active site (Figure 4-14). This surface interaction is formed by Arg218 from one ALAS2 subunit and residues Trp171-Phe177 from the opposite subunit. Since this is a shallow pocket, distant from the active site, the functional relevance is not clear and further exploration has not been considered.

4.3.3 Enzymatic studies

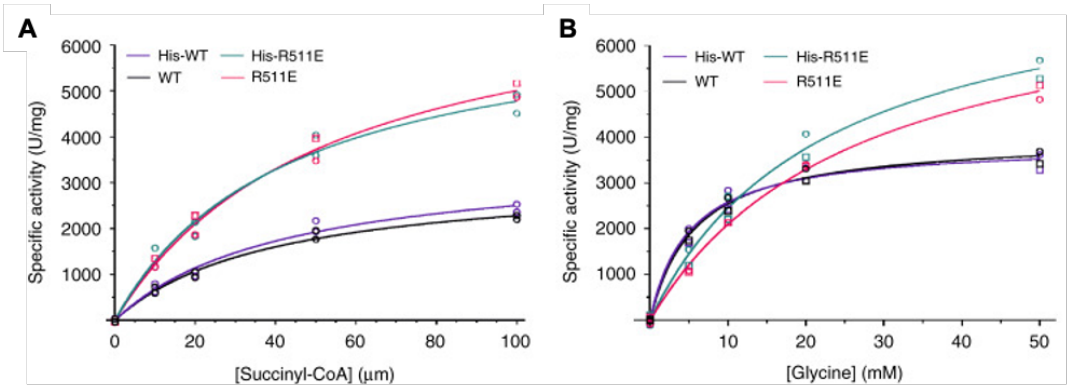
The following enzymatic assay data has been carried out by collaborators (Jason R. Marcero and Harry A. Dailey, Jr from the Department of Biochemistry and Molecular Biology, University of Georgia, Athens). My contributions include all communications with the collaborators, sharing protein reagents for assay measurement, and data interpretation and annotation. Characterising ALAS2 activity is important to validate structural interpretations of the hsALAS2 structure, understand the role of XLP disease associated variants on enzyme activity and to identify any inhibitory effects of the 8 fragments found to bind hsALAS2 in the XChem campaign. To do so, a discontinuous colorimetric activity assay was used to monitor ALA production from the condensation of glycine and succinyl-CoA (Figure 2-3) through modification with Ehrlichs reagent, as previously described[108].

4.3.3.1 Wild type enzyme activity

First, the activity of the crystallized construct ($ALAS2_{\Delta N142}$) was tested in the assay, to verify whether activity of the enzyme was similar to previously reported studies. The insect cell-expressed hsALAS2 is active with K_m values (31±4 μM for succinyl-CoA, 11.8±1.0 μM for glycine; (Figure 4-16), close to those reported by others using a similar assay.

To confirm the structural observations that Arg511 is important for maintaining the Ct-extension in an auto-inhibitory state, the proteolysis-prone R511E variant has a 2-fold higher enzyme activity than WT, as indicated by Vmax and kcat values (Figure 4-16). As expected, the R511E variant demonstrates a higher specificity (kcat/Km) for succinyl-CoA than WT protein, indicating the mutated enzyme is able to adopt an

open state compatible with succinyl-CoA binding. Although interestingly, the opposite effect on specific activity was observed for glycine (Figure 4-16). Altogether, the data suggests that the salt bridge network between Glu569, Arg511 and Asp159 plays an important role in regulating the efficiency of the hsALAS2 reaction.



Enzyme (His6-tagged or tag-free)	k_{cat} (min^{-1})	V_{max}, SCoA (U mg^{-1})	K_m, SCoA (μM)	k_{cat}/K_m, SCoA (min^{-1} μM^{-1})	V_{max}, Gly (U mg^{-1})	K_m, Gly (mM)	k_{cat}/K_m, Gly (min^{-1} mM^{-1})
WT	3.0 ± 0.2	3200 ± 200	39 ± 6	0.075 ± 0.012	4000 ± 130	5.7 ± 0.7	0.52 ± 0.07
His-WT	3.2 ± 0.3	3600 ± 300	42 ± 8	0.076 ± 0.015	3900 ± 180	5.0 ± 0.9	0.64 ± 0.12
R511E	6.3 ± 0.5	7700 ± 500	53 ± 7	0.120 ± 0.019	7700 ± 300	26 ± 2	0.24 ± 0.03
His-R511E	6.4 ± 0.7	6800 ± 500	43 ± 8	0.15 ± 0.03	8000 ± 500	22 ± 3	0.29 ± 0.05

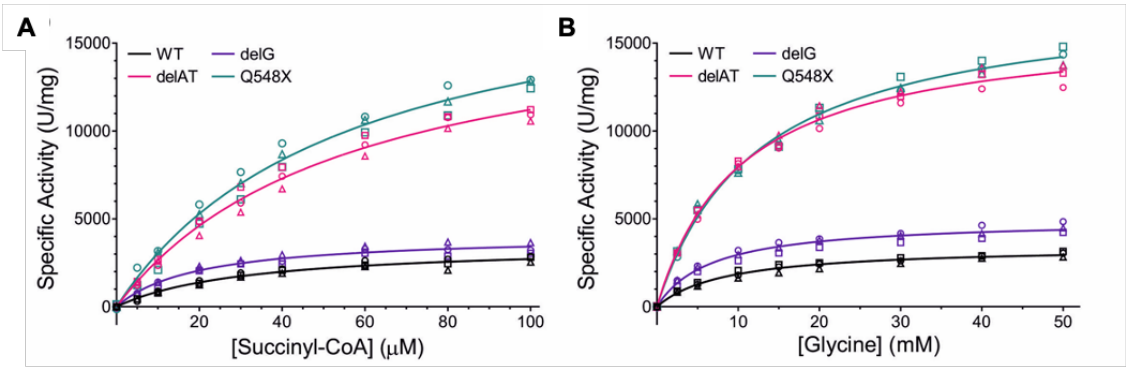
Figure 4-16: Activity of Progression curves for recombinant hsALAS2$_{\Delta N142}$ WT and R511E variant.

Specific activity is equivalent to the initial velocity normalized to protein in each assay. Succinyl-CoA and glycine titrations contained 50 mM glycine and 100 μM succinyl-CoA, respectively. Each nonlinear regression line was fitted to data from two biological replicates (n = 2), with the data for each replicate (average of two technical replicates) plotted as distinct symbols (circle and square). Table reveals kinetic parameters for recombinant N-terminal His6-tagged and tag-free hsALAS2$_{\Delta N142}$ WT and R511E proteins. Reported parameters are best-fit values ±SE for progression curves. Data generated by collaborators (see acknowledgments)

4.3.3.2 Characterisation of XLP-causing deletion variants

Next, hsALAS2 wild-type and XLP variants were measured to understand the effect of XLP variants on ALAS2 activity (Figure 4-17). k_{cat} and V_{max} values for delAT and Gln548* were found to be 3- to 6-fold greater than wild-type and delG proteins, in

agreement with those previously reported in the literature (delAGTG was not assayed due to C-terminal proteolysis). The association of the delG mutation with XLP has been hypothesised to be the result of increased *in vivo* enzyme stability relative to that of the wild-type[37]. It was found that delG has the same 2- to 3-fold wild-type specificity (k_{cat}/K_m) with respect to succinyl-CoA as delAT and Q548X Figure 4-17). Interestingly, this appears to be due in part to the lower K_m of delG for succinyl-CoA compared to delAT, Q548X, and wild-type forms, suggesting that the greater affinity for succinyl-CoA may in part contribute to the hyperactivity of delG.



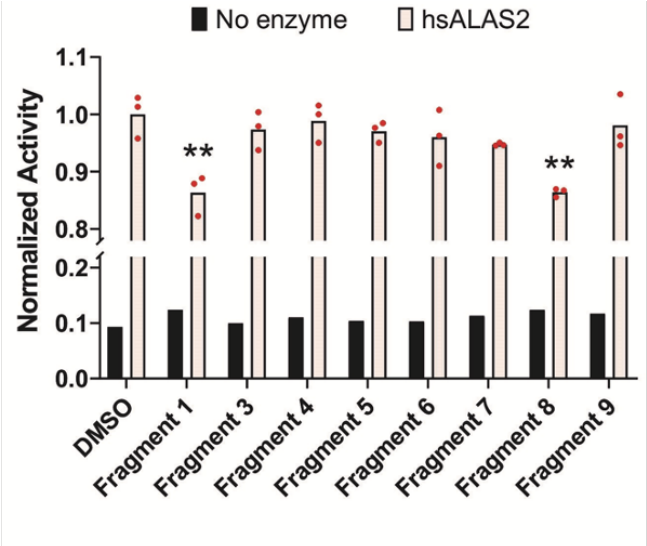
Enzyme (His6-tagged)	k_{cat}(min^{-1})	V_{max}, SCoA (U mg^{-1})	K_m, SCoA (μM)	k_{cat}/K_m, SCoA (min^{-1} μM^{-1})	V_{max}, Gly (U mg^{-1})	K_m, Gly (mM)	k_{cat}/K_m, Gly (min^{-1} mM^{-1})
WT	3.5 ± 0.2	3600 ± 200	32 ± 4	0.110 ± 0.014	3440 ± 90	8.5 ± 0.8	0.41 ± 0.04
delAT	15.4 ± 0.8	17,500 ± 800	56 ± 5	0.28 ± 0.03	16100 ± 310	10.3 ± 0.6	1.50 ± 0.12
Q548X	17.3 ± 1.1	19,900 ± 1000	55 ± 5	0.31 ± 0.04	17700 ± 330	12.4 ± 0.6	1.40 ± 0.11
delG	4.5 ± 0.2	4000 ± 180	20 ± 3	0.23 ± 0.03	4960 ± 140	6.7 ± 0.7	0.66 ± 0.08

Figure 4-17: XLP causing variants of hsALAS2.

Michaelis-Menten curves for purified, E. coli-expressed hsALAS2$_{\Delta N78}$ WT and XLP variants with titration of succinyl-CoA (b) and glycine (c). Regression lines represent the mean of three biological replicates (n=3), with the data for each replicate plotted as distinct symbols (circles, squares and triangles). Specific activity represents initial velocity normalized to protein in each assay (1-4 μg mL^{-1}), where succinyl-CoA and glycine titrations were conducted in the presence of 50 mM glycine and 100 μM succinyl-CoA, respectively. Table shows kinetic parameters for recombinant His6-tagged hsALAS2$_{\Delta N78}$ WT and XLP variant proteins expressed in *E. coli*. Reported parameters are best-fit values ±SE for progression curves.

4.3.3.3 *Characterisation of fragments bound to ALAS2 Ct-extension*

The enzymatic activity assay was finally used to test for inhibitory effects of the identified fragment hits from the crystallography screening campaign (section 4.3.2). It was found that fragments 1 and 8 inhibited hsALAS2 by 15% in the presence of 1 mM fragments and showed concentration dependency, reaching 28 and 22% inhibition with 5 mM fragments of 1 and 8, respectively (Figure 4-18). Such inhibition would suggest the fragments restrict conformational dynamics of the Ct-extension preventing binding of succinyl-CoA and ALAS2 open-closed transitions.



Additive	Concentration	WT Activity	% WT Activity
DMSO	–	1.00 ± 0.06	100 ± 6
Fragment **1**	1 mM	0.84 ± 0.03	84 ± 3
	5 mM	0.72 ± 0.02	72 ± 2
Fragment **8**	1 mM	0.85 ± 0.04	85 ± 4
	5 mM	0.78 ± 0.03	78 ± 3

Figure 4-18 Enzyme activity of hsALAS2$_{\Delta N78}$ with fragments.

Activity values represent the mean of three biological replicates ± SD normalised to WT enzyme treated with DMSO (diluent).

4.4 Conclusions

It was previously determined that a conserved mobile loop in the rcALAS active site, undergoing conformational change to bind succinyl-CoA, was proposed as a regulatory feature that committed the enzyme to catalysis through an open-to-close active site transition[92]. Here, the hsALAS2 Ct-extension was found to impose additional conformational restriction on this active site loop (through the salt bridge network Glu569:Arg511:Asp159), preventing transition to the open conformation. To overcome these restraints, displacement of the Ct-extension from the conformation depicted in our structure, will be essential for both substrate binding and product release. This data therefore adds to the described complexity of regulatory mechanisms known to exist to modulate ALAS2 activity[109] underlining the importance of its tight regulation to maintain metabolic flux during erythropoiesis.

The structure also provides insight into the role of previously engineered mutations of the active site loop from earlier studies[101]. Specifically, engineered mutations directed to the active site loop of murine ALAS2 (including the equivalent Arg511 residue in human) also led to a similar hyperactivity as the clinical mutations causing XLP. This observation can be easily reconciled with the hsALAS2 structure, as they would be likely to abolish interactions with the Ct-extension helix α15 and relieve the autoinhibition.

It can be speculated that the C-terminal regulatory feature, observed here for ALAS2, will also be present with a similar role in the ubiquitously expressed ALAS1 isozyme, considering their sequence conservation within this region (e.g. helix α15).

However, beyond helix α15 at the far C-terminus, ALAS1 and ALAS2 have evolved very diverse amino acid sequences (Figure 4-1B). Being surface-exposed, these residues may endow the two isoforms with different protein interaction functions. The hsALAS2 Ct-extension is implicated from in vitro pull-down studies as the binding site for the β subunit SUCLA2 of the A-SCS complex[17] although follow up studies here (section 5.5.1) are yet to describe this interaction further.

Finally, the recent siRNA knockdown of the *ALAS1* mRNA showing effective prevention of adverse effects from acute hepatic porphyrias[45] points to a similar therapeutic benefit of ALAS2 inhibition for porphyrias in the erythropoietic pathway. Fragments 1-9 identified from the crystallography-based screening campaign have the potential for chemistry optimisation into lead inhibitors in the future. These inhibitors would be aimed at increasing residence time at the Ct-extension region to prevent substrate access and product release during catalysis.

5 Structural and biophysical characterisation succinyl-CoA ligase Complex

5.1 Background to this study

5.1.1 Known structure function relationships of SCS

Succinyl-CoA synthetase complexes catalyse the reversible conversion of succinyl-CoA to succinate (Figure 5-1), a reaction directly coupled to the production or hydrolysis, depending on reaction directionality, of ATP or GTP.

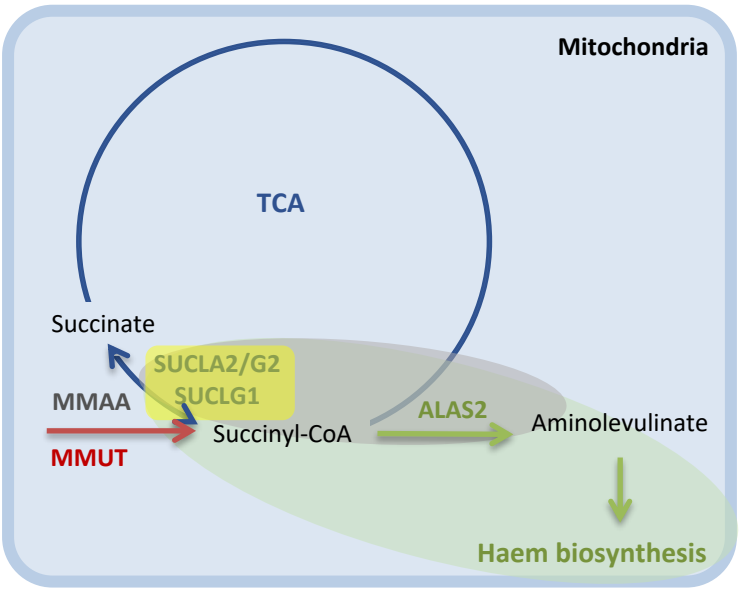


Figure 5-1: Enzyme pathways involving the succinyl-CoA synthetase complex.

The SCS complex highlighted in yellow (G-SCS formed by SUCLG1 and SUCLG2 and A-SCS formed by SUCLG1 and SUCLA2) is shown to participate in the TCA cycle. The anaplerotic input reaction involving MMUT and MMAA of the propionic catabolism pathway is shown in red and the proposed non-obligate and metabolon complexes between A-SCS and haem biosynthesis enzymes are shown in grey and green highlighting respectively.

It has been well established that the GTP and ATP specific succinyl-CoA synthetase complexes (G-SCS and A-SCS) participate in differing cellular roles[52]. Therefore, it is important to understand the structure function relationships of the SCS isoforms to ascertain how they are separately implicated in multiple cellular processes. Although all SCS complexes are composed of α and β subunits, quaternary structures affecting the composition of α and β appear to vary within the known paralogues. The *E. coli* (ec) enzyme (ecSCS) forms an $(\alpha\beta)_2$-hetero-tetramer, the structure of which was solved in the presence CoA[110,111] (PDB: 1SCU) (Figure 5-2A). Unlike eukaryotes, ecSCS only encodes a single β subunit which is promiscuous in nucleotide specificity, using both ATP and GTP for catalysis. By contrast, isolation of G-SCS purified from pig heart[58] revealed the evolution of a eukaryotic complex that

is active as an αβ-hetero-dimer (Figure 5-2B). The structures of *Sus scrofa* G-SCS (ssG-SCS) have been solved in multiple states including in the APO state (1EUC), in the presence of GTP (2FP4), GDP (2FPG), succinate (5CAE), CoA (4XX0), and in the intermediate state of the complex involving phosphorylation of a catalytic histidine (see below) (1EUD)[58,112–115].

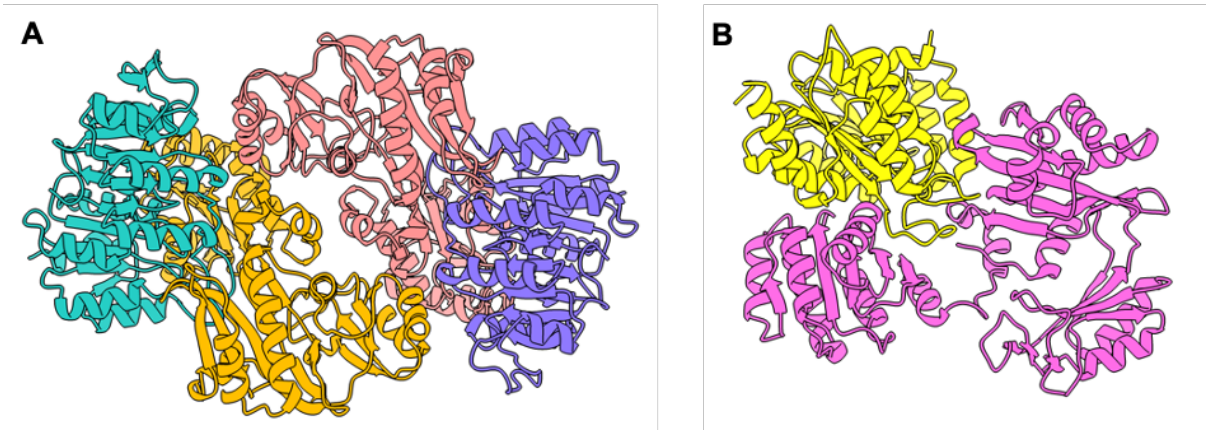


Figure 5-2: Known quaternary states of SCS structures.

A: ecSCS $(\alpha\beta)_2$-hetero-tetramer (1SCU). α subunits are coloured purple and green and β subunits coloured orange and salmon. **B:** ssG-SCS hetero-dimer (1EUC). α subunit is coloured yellow and β subunits coloured pink.

Structural and enzymatic studies, primarily using purified recombinant ssG-SCS and ecSCS, have helped to delineate in detail the SCS reaction mechanisms. Current predictions suggest that the reaction takes place within two separate active sites (1 and 2) of the complex, via a mechanism that progresses through 3 steps (Figure 5-3):

$$1: \text{SCS} + \text{succinyl-CoA} + P_i \leftrightharpoons \text{SCS}\cdot\text{succinyl-PO}_4 + \text{CoASH}$$
$$2: \quad \text{SCS}\cdot\text{succinyl-PO}_4 \leftrightharpoons \text{SCS-PO}_3 + \text{succinate}$$
$$3: \quad \text{SCS-PO}_3 + \text{NDP} \leftrightharpoons \text{SCS} + \text{NTP}$$

Figure 5-3: SCS 3 step reaction mechanism.

Reaction scheme adapted[112]

The ecSCS structure first identified that the α-subunit is responsible for binding CoA-SH[111]. Although no structures exist in the presence of succinyl-CoA (substrate of the first step), the CoA moiety of succinyl-CoA is likely to bind in the same pocket as CoA alone. Such an orientation would present the thioester bond into active site 1 (Figure 5-4) positioned at the hetero-dimer interface. An organic phosphate that binds near a catalytic histidine carries out a nucleophilic attack of this bond, leading to hydrolysis of succinyl-CoA to form succinyl-PO_4 and CoA in the first step.

The succinyl-PO_4 product from the first step is believed to stay bound to the enzyme[116], in a pocket very close to active site 1, contributed by residues from both SUCLG1 (α subunit) and the C-terminal domain of the β subunit. The second step requires transfer of the phosphoryl group from succinyl-PO_4 to the catalytic histidine residue of SUCLG1[117]. Two α-helices, termed the `power helices', one contributed by SUCLG1 and the other contributed by β-subunit, are believed to stabilise the enzyme in the conformation that retains the phosphorylated histidine in active site 1[111].

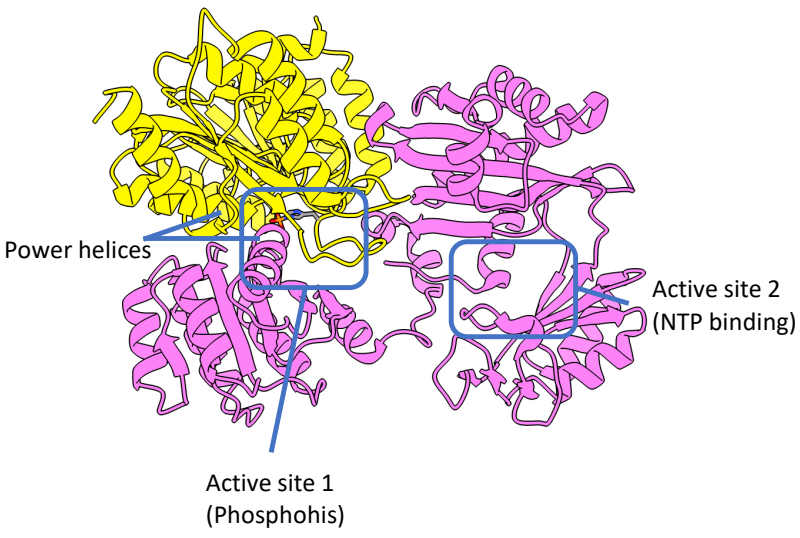


Figure 5-4: Active site annotation of ssG-SCS structure.

ssG-SCS hetero-dimer (1EUD). The α subunit is coloured yellow and β subunits are coloured pink.

The third step of the reaction involves substrate level phosphorylation of the incoming nucleotide. The nucleotide-binding site was identified in the N-terminal ATP grasp domain of the β-subunit. This forms active site 2, and is approximately 35 Å from the position of the catalytic histidine in active site 1[115,118] (Figure 5-4). The catalytic histidine is situated in the middle of a long loop consisting of 24 amino acids. It is postulated that this phosphohistidine loop can transition to active site 2, juxtaposed in the nucleotide-binding pocket, via long-range conformational change, to deliver a phosphoryl group to the bound nucleotide di-phosphate (NDP). Evidence for such a conformational change is found in the related enzyme acetyl-CoA synthetase from *Candidatus Korachaeum cryptofilum*, the structure of which showed extension of the entire catalytic loop positioning the histidine within active site 2[119].

2.1.1 Unanswered questions and chapter synopsis

Until now, there has been preferential enzymatic characterisation of the G-SCS complex compared to A-SCS, in part because G-SCS was the first eukaryotic SCS complex successfully identified and characterised, and also because structural characterisation of any higher order eukaryotic A-SCS complex has to date remained difficult. It also remains unclear whether tissue expression and substrate/product ratios are the only significant regulatory factors that modulate the apparent involvement of the two complexes in very different biological pathways. Thirdly, the mechanisms underlying the involvement of A-SCS specifically in mitochondrial DNA depletion syndromes and haem biosynthesis through proposed protein-protein interactions remain unresolved (.

Here, the A-SCS complex is investigated using X-ray crystallography, cryo-electron microscopy and small angle X-ray scattering, to better understand the structure/function features that have evolved in the polypeptide of SUCLA2 for its participation in energy production within the TCA cycle, and to provide the appropriate molecular template to understand disease mutations affecting specifically the A-SCS complex, caused by *SUCLA2* gene mutations. Interactions of A-SCS and ALAS2 in relation to its involvement in erythroid specific haem biosynthesis pathway, are also investigated as validation of results from existing literature (Figure 5-1).

5.2 Complex Isolation and validation

5.2.1 Co-expression design of SUCLA2 and SUCLG1

In order to obtain, for structural and biophysical studies, the soluble complex of human A-SCS (hsA-SCS) with mature forms of hsSUCLG1 and hsSUCLA2, i.e. lacking their N-terminal mitochondrial targeting sequences (hsSUCLG1$_{\Delta N51}$, aa 52-346; and hsSUCLA2$_{\Delta N41}$, aa 42-463), a bicistronic co-expression method was used. DNA fragments encoding hsSUCLG1$_{\Delta N51}$ and hsSUCLA2$_{\Delta N41}$ were PCR-amplified, incorporating as overhangs either a ligation independent cloning (lic) site at the N-terminus (Nt) combined with a C-terminal (Ct) ribosomal binding site (RBS), or an Nt-RBS with a Ct-lic site. Amplified coding sequences containing the above overhangs were subsequently mixed for the second PCR reaction, in the following combinations: Nt-lic-hsSUCLG1-CtRBS with Nt-RBS-hsSUCLA2-Ct-lic (to generate His-tagged SUCLG1 and untagged SUCLA2), or Nt-lic-hsSUCLA2-Ct-RBS with Nt-RBS-hsSUCLG1-Ct-lic (to generate untagged SUCLG1 and His-tagged SUCLA2). These fragments were then sub-cloned into pNIC28-Bsa4 vectors to form a bi-cistronic reading frame. This results in two coding sequences within one open

reading frame that are interspersed by an RBS, with the first coding sequence containing a vector-encoded tobacco etch virus protease cleavable Nt-His_6-tag and the second coding sequence untagged (Figure 5-5).

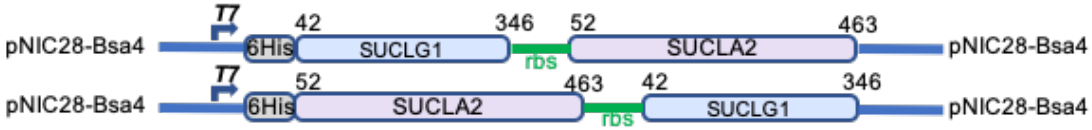


Figure 5-5: Bi-cistronic hsA-SCS expression vectors.

DNA fragments encoding Nt-His_6-tagged SUCLG1 (42-346) or SUCLA2 (52-463) and non-tagged SUCLA2 (52-463) or SUCLG1 (42-346), interspersed by an in-frame ribosomal binding site (rbs) (green bar) are sub-cloned into the pNIC28-Bsa4 vector (blue bar).

DNA sequencing confirmed presence of both polypeptides. The complex with a His_6-tag on SUCLG1 contained a point mutation in the stop codon leading to a two-residue extension at the Ct-terminus (after residue 463) before the next stop codon. This construct was used in determination of the crystal structure. All SAXS and cryo-EM experiments were performed on the WT complex using a His_6-tag on SUCLA2 (see later sections)

5.2.2 Co-purification of SUCLA2 and SUCLG1 reveals stable complex formation

In order to test for yield, solubility and complex formation, the two bi-cistronic hsA-SCS vectors were initially expressed in three millilitre culture scale and purified using Ni^{2+} affinity chromatography. Stoichiometric proportions of hsSUCLG1$_{\Delta N51}$ and hsSUCLA2 $_{\Delta N41}$ were observed in the eluant fractions to high purity, as shown in SDS-PAGE with similar yields observed when the complex was pulled-down with a Nt-His_6-tag on either of the two subunits (Figure 5-6, labelled Ni-eluent). The complex expression was next scaled up to six litre cultures and an additional purification step using gel filtration chromatography was performed. The complex

resolved as a single broad peak using gel filtration chromatography spanning nearly 20% of the column volume (Figure 5-6, chromatogram). A superdex 200 column with a resolving range for globular proteins between 10 to 600 kDa was chosen based on previous expected quaternary arrangements of the α and β subunits. Subsequent SDS-PAGE analysis reveals the stoichiometric co-elution of both proteins across the entire broad peak (Figure 5-6, labelled Gel filtration), indicating both proteins form a stable complex over the time course of the chromatography. However, the broad nature of the peak may reflect a degree of unresolved conformational or size heterogeneity within the sample.

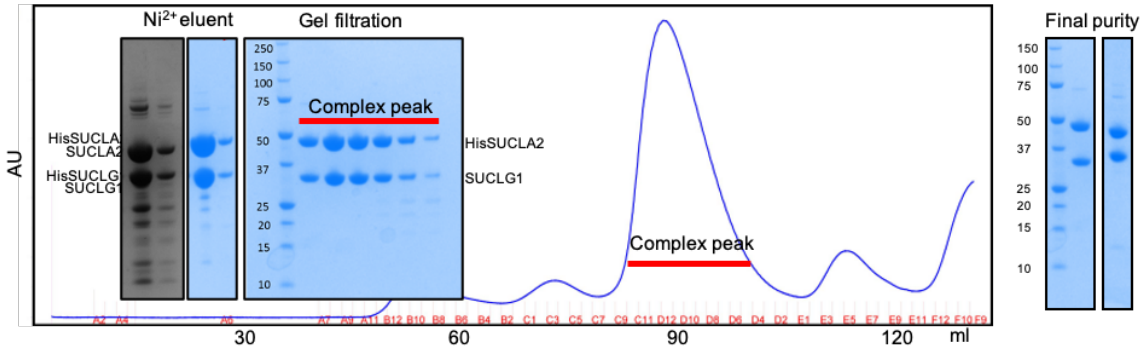


Figure 5-6: Purification of hsA-SCS.

SDS page analysis (left gels) of hsA-SCS purity using either His6-taghsSUCLG1$_{\Delta N51}$ or His6-taghsSUCLA2 $_{\Delta N41}$ to pull out the complex. Size exclusion chromatogram shows broad elution profile of the complex that representative of both purification methods. SDS page analysis (middle gel) shows stoichiometric co-elution of hsA-SCS subunits across the gel filtration peak. SDS page analysis (right gel) shows final purity for the two purification methods. Lane one is the complex pulled out by His6-taghsSUCLA2 $_{\Delta N41}$ and lane two is the complex pulled out by His6-taghsSUCLG1$_{\Delta N51}$.

Intact mass spectrometry analysis of the complex reveals a sub-population of SUCLG1 to have an observed mass which is 80 da higher than the expected theoretical mass. The 80 Da difference in mass suggests that the complex was isolated in a mixed phospho-histidine. Given prior knowledge of the rection mechanism (Figure 5-3), this could be attributed to phosphorylation of the catalytic

His299 residue of SUCLG1. The population in the phospho-histidine state would be poised in the third step of the reaction, ready to proceed in the forward direction if the catalytic histidine is found in active site 2, or the reverse reaction if in active site 1 (Figure 5-7, blue). To test whether the phosphorylated residue observed is indeed His299, the enzymatic activity of hsA-SCS was tested by monitoring the conversion between the unmodified and phosphorylated states using intact mass spectrometry. As expected from the proposed reaction mechanism (Figure 5-3), incubation with succinate and CoA or with ADP resulted in dephosphorylation of the intermediate phospho-His299 species i.e. loss of 80 Da (Figure 5-7, green and yellow), whereas incubation with ATP lead to phosphorylation (Figure 5-7, pink). Unexpectedly, incubation with succinyl-CoA also lead to dephosphorylation of His299 (Figure 5-7, black), It is plausible that in the absence of free organic phosphate, the phosphate group on the catalytic His is sufficient to cleave succinyl-CoA. Without ADP, the enzyme is likely to be trapped in a state ready to carry out the second reaction step, where the complex is bound to the succinyl-phosphate as opposed to a phospho-histidine.

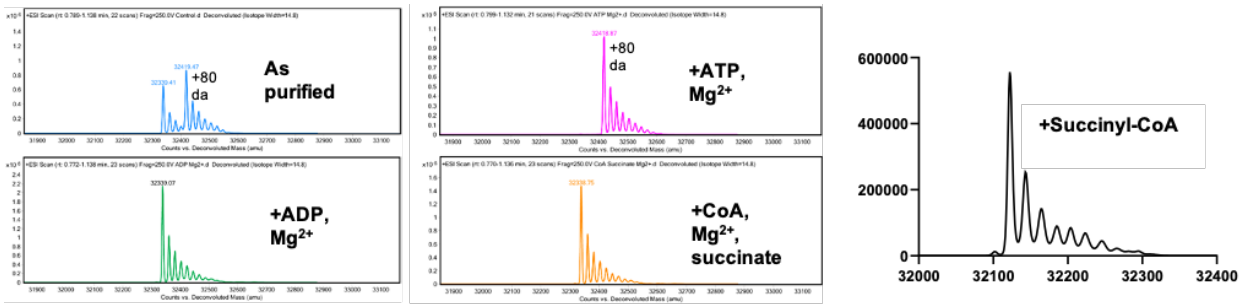


Figure 5-7: Activity assay of hsA-SCS.

Intact mass spectrometry analysis of SUCLG1 showing phosphorylation state of His299 in the presence of different substrates.

5.3 hsA-SCS structural studies

5.3.1 Crystal structure of hsA-SCS

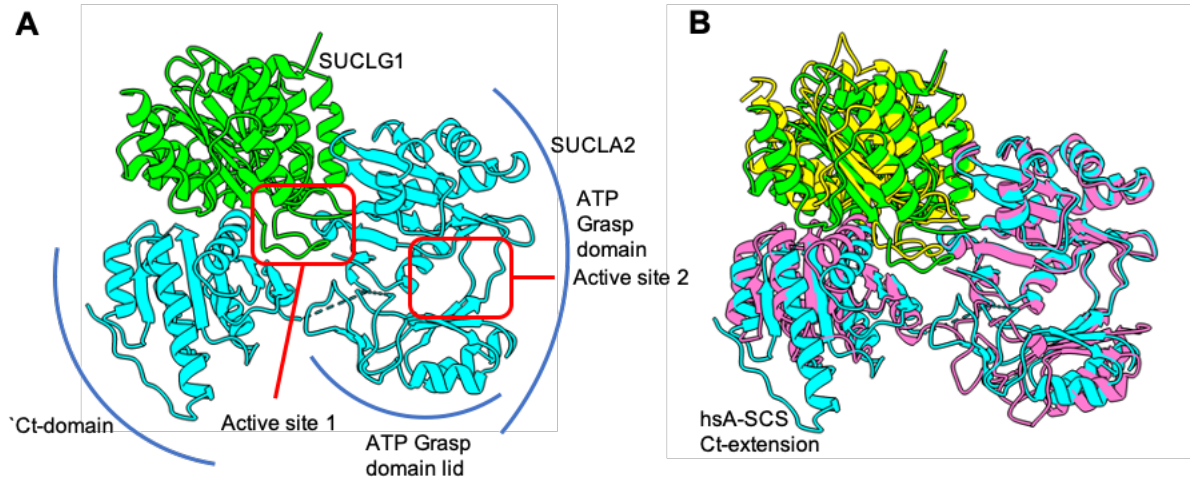
To understand from a structural perspective the features evolved in the SUCLA2 polypeptide that distinguish it from SUCLG2 and therefore enable it to participate in alternative metabolic pathways, X-ray crystallography was employed. hsA-SCS, comprised of His_6hsSUCLG1$_{\Delta N51}$ and hsSUCLA2$_{\Delta N41}$, in the as purified state crystallised promiscuously in multiple coarse screen conditions. Diffraction datasets were recorded and auto processed using Xia2[82] processing pipelines at the Diamond Light Source synchrotron. The structure was solved to 2.6 Å resolution using molecular replacement with the ssG-SCS structure (1EUC)[58] as search model. Crystallographic table of parameters are displayed in Table 5-1.

Dataset	hsA-SCS
Beamline	Diamond Beamline I04
Wavelength (Å)	0.9795
Unit cell parameters (a,b,c)	130.20 Å, 130.20 Å, 119.84 Å
(α, β, γ)	90.0°, 90°, 120°
Space group	P 31 2 1
Resolution range (Å)	65.10 – 2.59
Observed/Unique reflections	36946
Rsym (%)	0.09 (0.94)
I/sig(I)	14.4 (2.1)
Completeness (%)	100 (100)
Multiplicity	9.7 (9.9)
CC(1/2)	1(0.55)
R_{cryst} (%)	0.21
R_{free} (%)	0.25
Wilson *B* factor (Å^2)	63
R.m.s.d. bond length (Å^2)	0.002
R.m.s.d. bond angle (°)	0.47
Clashscore	5.42
Ramachadran favoured (%)	97.37
Ramachandran disallowed (%)	0
Rotamer outliers (%)	0

Table 5-1: Crystallographic table of parameters for the hsA-SCS dataset.

Values in parenthesis represent highest resolution shell.

The asymmetric unit of the hsA-SCS structure contained a single hetero-dimer (Figure 5-8A), with a total buried surface area of 5310 Å^2 at the hetero-dimer interface. The hsA-SCS structure shares the same ternary fold as the homologous ssG-SCS and ecSCS structures. The structure overlays most closely with phosphorylated ssG-SCS (1EUD)[58], (Figure 4-7B), with an Cα-rmsd of 0.39 Å. Overlay with other SCS structures results in an Cα-rmsd of 0.41 Å to un-phosphorylated ssG-SCS (1EUC)[58], 0.50 Å to nucleotide bound ssG-SCS structures (2FP4, 2FPG)[115] and 0.79 Å to the ecSCS structure (1SCU)[110]. Sequence identity for the α subunit of hsA-SCS with ssG-SCS is 94.5% and with ecSCS is 65.3%. By contrast, the β subunits differ more substantially compared to the previously solved structures, with a sequence identity of 50.8% ssG-SCS. This greater sequence diversion for β subunits reflects that they are encoded at different gene loci, providing the differing nucleotide specificity



and evolutionarily distinct roles.

Figure 5-8: Crystal structure of hsA-SCS complex.

A: structure of hsA-SCS with labelled active sites and domains. **B:** Overlay of hsA-SCS structure with the phosphorylated ssG-SCS structure (1EUD). hsA-SCS: SUCLA2 is coloured cyan and SUCLG1 is green. ssG-SCS: SUCLG2 is coloured pink and SUCLG1 coloured yellow.

In the hsA-SCS structure, the α subunit, SUCLG1, is formed of 10 helices and 14 β-strands (Figure 5-9A,C). The 14 strands are arranged into two separate β-sheets containing six parallel strands and one antiparallel strand each (Figure 5-9C). The first sheet is formed by strands β1 to β7 and the second sheet is formed by strands β8 to β14. The two β-sheets form the core of the structure and are sandwiched by the 11 helices (nine α-helical, α1- α9; and two 3_{10}-helical, η1-η2). Like the ssG-SCS structure (1EUD), hsSUCLG1 is captured in the phosphorylated state at His299 (Figure 5-9D), and in line with all previous structures of ssG-SCS and bacterial homologues the active site loop (aa Ala287-Gly310) containing the phospho-histidine is bound within active site 1 (Figure 5-9E). This position of the loop in the phosphorylated state has previously been described as most likely primed for the forward reaction that produces succinyl-CoA in the presence of succinate and CoA.

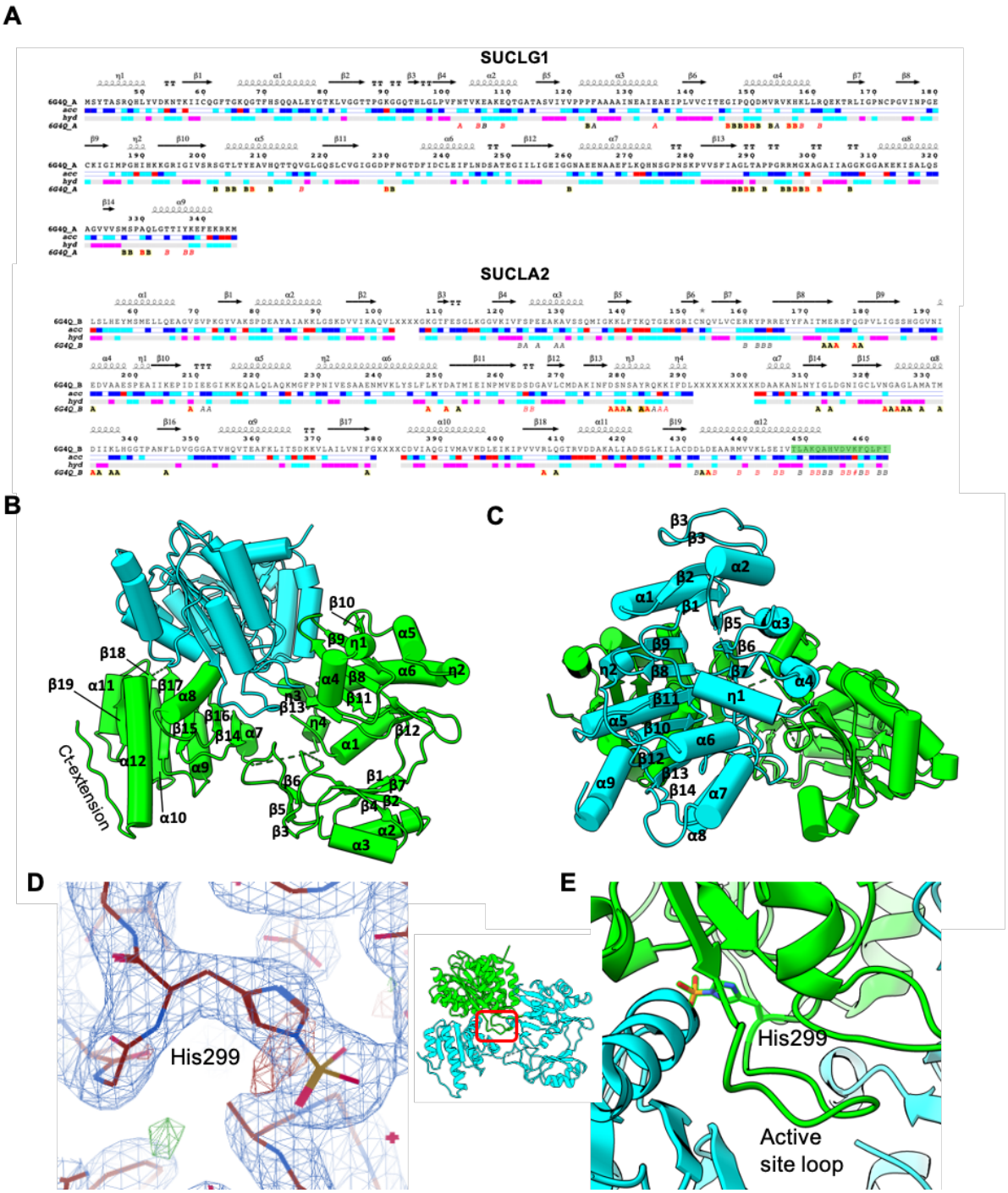


Figure 5-9: hsA-SCS Secondary structure elements.

A: Assigned secondary structure elements to the hsA-SCS complex. The Ct-extension sequence of SUCLA2 is highlighted green in panel A. **B-C:** Cartoon representation of the hsA-SCS structure. SUCLG1 coloured blue and SUCLA2 coloured green. Views in B and C are rotated 90° along the X axis. **D:** 2Fo-FC map (contoured at 1 σ) showing active site 1 with density for the phospho-histidine. **E:** Zoom of active site 1 showing position of active site loop.

The β subunit SUCLA2 is separated into two distinct domains (Figure 5-8A). The N-terminus, residues 66–288, display the canonical ATP grasp domain fold formed by 9 α-helices and 13 β-strands arranged into 3 separate β-sheets. Within this domain, the ATP grasp lid, (aa Gly74-Glu159), composed of β-strands 1, 2, 7 and 4 and α helices 2 and 3 (Figure 5-9A,B), adopt a significantly less open conformation when compared to the equivalent nucleotide free state of ssG-SCS (1EUD and 1EUC) (Figure 5-10A). Instead, the position of the ATP grasp lid is more similar to that of the GTP bound state of ssG-SCS (2FP4) (Figure 5-10B), although no density for the nucleotide is present in our structure. The lid itself is a catalytically essential region of all active ATP grasp folds, providing a flexible lid on top of the active site that is separate to the main body of the domain[120]. Its established role is to provide a phosphate binding loop (named the P-loop) to interact with the nucleotide. Previous structures of related domains frequently reveal the lid to adopt open and closed conformations dependent on nucleotide occupancy[120]. In order for nucleotide phosphorylation or dephosphorylation to proceed, like ssG-SCS, the active site loop would have to travel approximately 35 Å to position His299 into active site 2 (Figure 5-8A), which is located within the ATP grasp domain of SUCLA2.

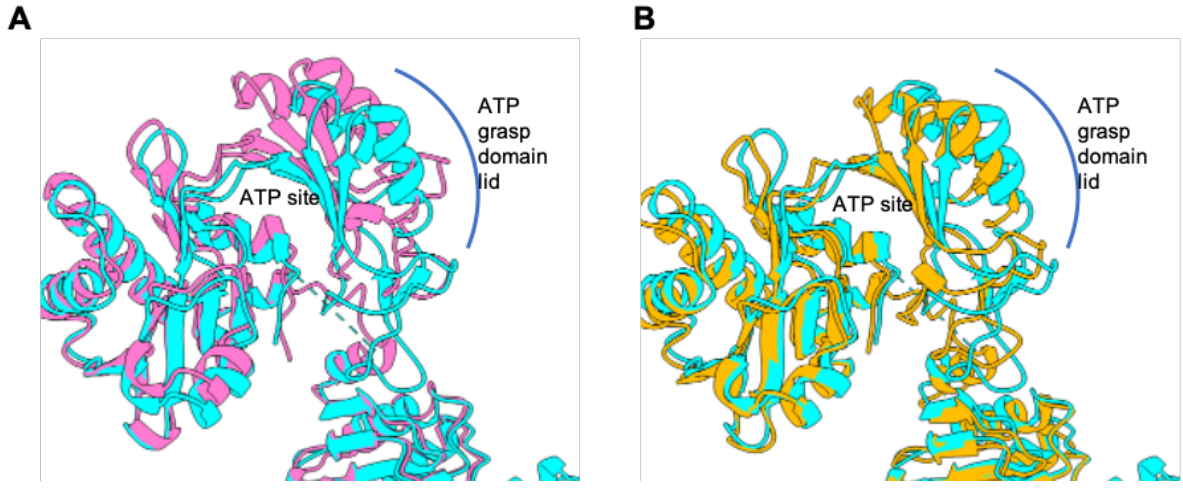


Figure 5-10: Comparison of ATP grasp lid conformation with orthologue structures

A: Overlay of the ATP-grasp domain lid of hsSUCLA2 (cyan) with apo ssSUCLG2 (1EUD) (pink). **B:** Overlay of the ATP-grasp domain lid of hsSUCLA2 (cyan) with GTP bound ssSUCLG2 (2FP4) (gold).

The C-terminal domain of hsSUCLA2 is connected to the N-terminal domain by a loop, formed by residues 290 to 303, that has been shown to interact with the bottom of the ATP-grasp domain lid in both bacterial and ssG-SCS structures. However, this loop region has no density in the hsA-structure and is presumably disordered. The C-terminal domain, consisting of residues 303 to 463, has a similar fold to the C-terminus of hsSUCLG1, but lacks its first β-strand. The remaining β-strands (β14 – β19) form a predominantly parallel β-sheet sandwiched by α-helices 7 to 12. (Figure 5-9A,B).

Overall, the two domains of SUCLA2 form a significant part of the hetero-dimer interface at active site 1, as illustrated in Figure 4-7A. Both subunits have previously been shown to contribute residues to succinate binding in this region[112]. Of particular interest, higher order eukaryotic A-SCS sequences have evolved a C-terminal extension of 16 amino acids relative to eukaryotic SUCLG2 polypeptides and bacterial homologues (Figure 5-9A). In our structure, this results in an extra 2 turns of α-helix 12 and an unstructured loop of the terminal 11 amino acids, aa Ala153-Ile463 (Figure 5-8B). The relevance of this extension with respect to hsA-SCS structure and function remains unclear. The additional two amino acids at the C-terminus of this construct, due to a mutated stop codon are disordered and not modelled into the data.

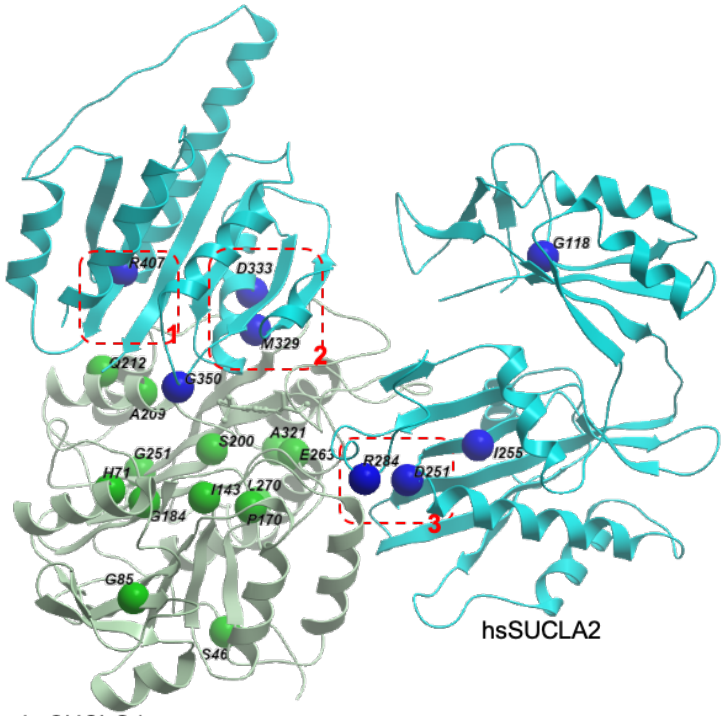
5.3.2 Disease associated missense mutations disrupt the dimerisation interface

The hsA-SCS structure provides a template to map and rationalise the effect of disease relevant variants on the SUCLG1 and SUCLA2 subunits, as well as on the

complex (Figure 5-11A). To date, there are 21 recorded cases of SUCLG1 deficiency[60]. The gene mutations are a mixture of missense, aberrant splicing, one nonsense and a 2 bp deletion. From the 21 cases there are 16 reported missense mutations (Figure 5-11B). The cases with bi-allelic missense mutations tend to survive longer (reported median life expectancy of 1.5 years) than complete loss of function mutations (0 median live expectancy after birth)[60]. When mapped onto the structure of hsA-SCS, disease causing missense variants of SUCLG1 are mainly buried within the core fold of the enzyme (Figure 5-11A) and are predicted by the *FoldX* protein stability web servers[121] to disrupt protein stability (Figure 5-11B).

50 patients are recorded to have *SUCLA2* gene mutations. The distribution of mutations include, 10 recorded missense mutations, whole gene deletions, exon skipping and several frameshift mutations[60]. Patients with mutations in the *SUCLA2* gene tend to have a milder disease course than those with *SUCLG1* mutations, as they have a median expected survival of 20 years for bi-allelic missense mutations and 15 years for complete loss of function mutations[60]. So far there have been no recorded disease causing mutations in the GTP specific β subunit of SUCLG2. The most common site of missense mutation in SUCLA2 is at the conserved interaction interface with SUCLG1, harbouring 7 of the 10 mutations and predicted by the *FoldX* server[121] to have severe effects on protein stability. (Figure 5-11).

A



B

SUCLA2 variants	1. Structural Location	2. FoldX Stability Prediction	3. Sequence Conservation
Ala103Asp	Not mapped		
Gly118Arg	Surface	-0.57	Invariant
Asp251Asn	Interface	1.10	Predominantly Asp
Arg284Cys	Interface	2.90	Predominantly Arg
Ala307Val	Surface	-0.68	Variable; Val tolerated
Met329Val	Interface	4.08	Invariant
Asp333Gly	Interface	3.00	Invariant
Gly350Ser	Surface	-0.01	Predominantly Gly
Gly350Val	Surface	0.28	Predominantly Gly
Arg407Trp	Interface	7.27	Predominantly Arg
SUCLG1 variants			
Met14Leu	Not mapped		
Met14Thr	Not mapped		
Ser46Phe	Buried	46.09	Conserved
His71Arg	Surface	1.35	Majority His
Gly85Ala	Buried	7.05	Predominantly Gly
Ile143Ser	Surface	1.39	Majority Ile
Pro170Arg	Buried	14.58	Predominantly Pro
Gly184Ser	Surface	11.49	Invariant
Ser200Phe	Buried	18.68	Majority Ser
Ala209Glu	Buried	4.19	Majority Ala
Gln212Arg	Surface	-0.08	Majority Gln; Arg tolerated
Gly251Ser	Buried	3.05	Variable; Ser tolerated
Glu263Gly	Buried	5.88	Invariant
Glu263Lys	Buried	8.48	Invariant
Leu270trp	Buried	8.08	Variable
Ala321Pro	Surface		Majority Ala

Figure 5-11: Mutation analysis of hsA-SCS.

A: Reported SUCLG1 and SUCLA2 missense variants mapped onto the hsA-SCS crystal structure. Green and blue spheres represent Cα atoms of residues that harbour missense variants. **B:** Table of all missense mutations annotated in the structure shown in panel A with predicted changes in stability (ΔΔG, kcal/mol) (calculated by *FoldX*)[121], and amino acid variations at that position among orthologues (calculated via sequence alignments by *ConSurf*)[122].

The 7 disease associated missense changes of SUCLA2 are located at the subunit interface in 3 distinct interaction sites (Figure 5-11A). The first site (labelled 1 in Figure 5-11A and Figure 5-12A) involves interaction between Arg284 and Asp251 of the SUCLA2 ATP grasp domain with Asp151 and Gln150 in helix α4 of SUCLG1. Mutation of Arg284 to Cys would prevent charged interactions forming with Asp151. The second site (labelled 2 in Figure 5-11A and Figure 5-12A) involves interaction of Asp333 and Met329 of helix α8 in SUCLA2 Ct-domain with the active sit loop of SUCLG1 that contains the catalytic His299 residue. The Asp333Gly variant would likely remove hydrogen bonding to the SUCLG1 backbone of Gly288, while Met329Val may reduce hydrophobic interaction with the His299 residue. Close to this interaction site is Gly350, which has been identified to be mutated to valine or serine (Figure 5-11A). Although Gly350 does not participate in interactions at the dimer interface, it is juxtaposed next to the phosphate binding site for the reactive His299 residue in SUCLG1. Therefore, mutation to a bulkier residue may also disrupt hetero-dimeric interactions that form part of this active site. The final interaction site (Labelled 3 in Figure 5-11A and Figure 5-12A) involves residue Arg407 of the terminal β strand in SUCLA2 and Glu208 of SUCLG1 α5, the same helix which is responsible for binding the phosphorylated His299 residue. Mutation of Arg407 to tryptophan would be expected to replace charged interactions with a bulky hydrophobic residue that may disrupt interactions at the dimeric interface. The final 3

mutations are either not able to be modelled in the current structure or are surface exposed and therefore difficult to determine their overall effect.

To test experimentally whether the mutated interaction sites are crucial to complex formation, three of the missense changes (one from each of the three sites) were separately incorporated into the hsA-SCS bi-cistronic expression system by site-directed mutagenesis. These include the enzyme variants Asp333Gly, Arg284Cys and Arg407Trp, and the amino acid substitution was confirmed by sequencing. Complex formation between the SUCLA2 variants and SUCLG1 were assessed through Ni^{2+} affinity chromatography pull-down experiments. As shown in in Figure 5-12B, while wild-type His_6-SUCLG1 can pull down wild-type SUCLA2 (lane 1), the Asp333Gly (D333G) variant of SUCLA2 (lane 3) completely abrogates this interaction, as shown by a lack of co-elution of SUCLA2. A similar abrogation of interaction is seen with the Arg284Cys (R284C) variant of SUCLA2 in lane 4, although to a lesser degree as some complex is successfully pulled down. Finally, the Arg407Trp (R407W) mutation of SUCLA2 interestingly results in a loss of both SUCLA2 and SUCLG1 from the eluants (lane 2). It is possible that this particular mutation leads to a destabilised conformation of the complex that resulted in insoluble expression.

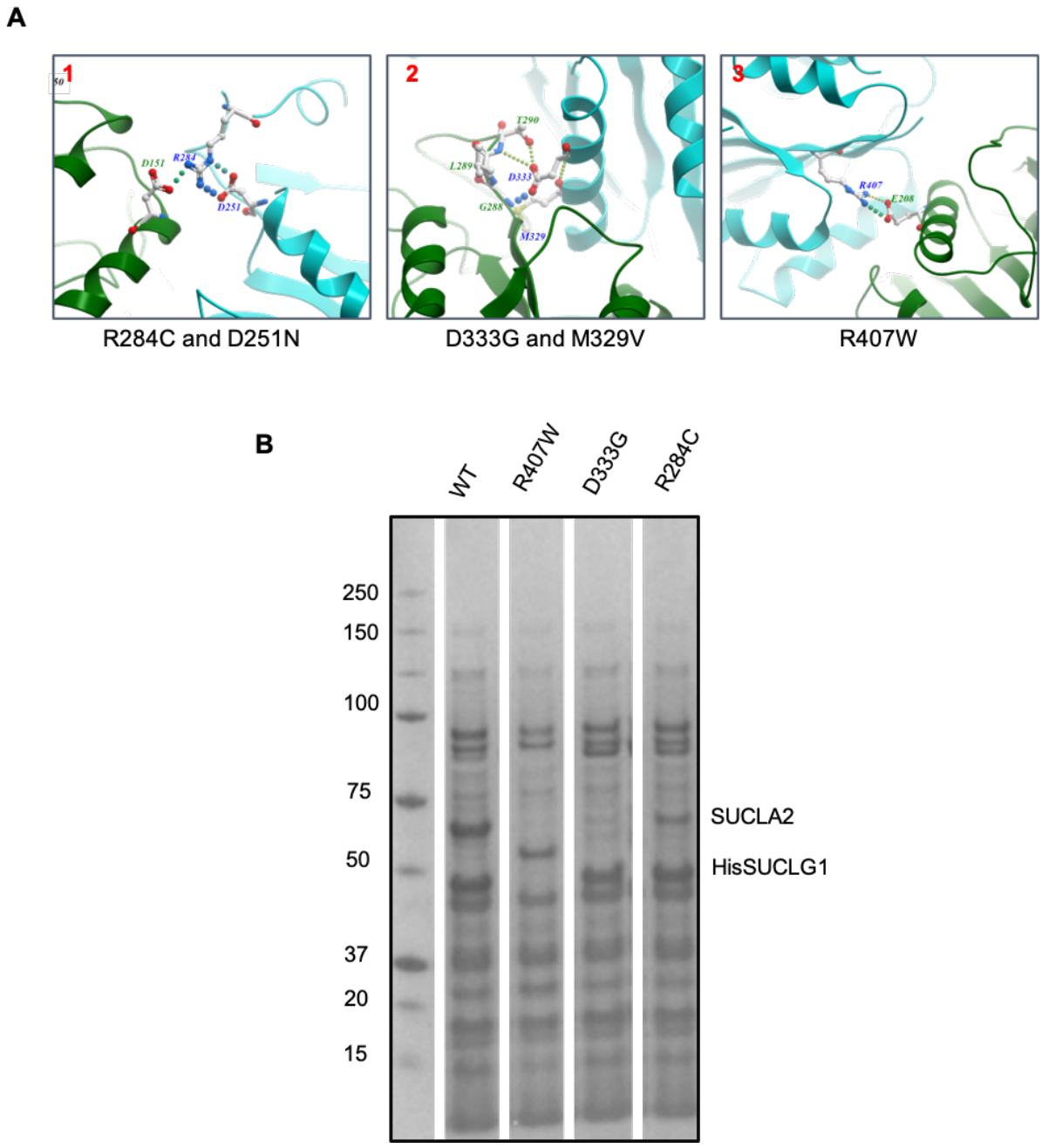


Figure 5-12: SUCLA2 interaction interface mutations.

A: Residues mutated at the hetero-dimer interface and predicted to disrupt dimer formation. What is blue chain what is green chain what is blue dashed line what is green dashed line **B:** SDS PAGE of IMAC pull down showing SUCLA2 mutations that disrupt complex isolation by breaking the hetero-dimer interface.

At the sequence level, all 6 SUCLA2 residues at the interface harbouring the 7 missense changes are highly conserved or invariant among β subunit homologues

(Figure 5-11B) and participate in identical interactions[58]. The lack of tolerated reported mutations at the conserved interaction sites in the G-SCS β-subunit (SUCLG2), or any interface missense mutations found in the shared SUCLG1 subunit that may uncouple the GTP specific complex, suggests that the ATP specific form is not able to compensate the differing roles of the GTP specific complex. However, the existence of hsA-SCS disease mutations could suggest the cellular roles of the A-SCS complex can be compensated to a degree by the GTP specific complex. This interpretation is supported by the difference in median life expectancy where complete loss of SUCLG1 function is not tolerated after birth in contrast to the milder disease associated with SUCLA2 loss of function.

5.4 Oligomerisation of hsA-SCS

5.4.1 In solution oligomeric states as revealed SAXS studies

The hsA-SCS biological assembly is believed to be hetero-dimeric, as interpreted from previous ssG-SCS complex studies[58] and the crystal structure of hsA-SCS. To determine whether this matches the in-solution state of the complex, small angle x-ray scattering (SAXS) was used. SAXS measurements can be used to calculate information on the average shape, conformation and MW, giving details on the assembly state of proteins in a solution environment. This can be compared against the higher resolution hsA-SCS structure, recorded in crystalline environment, to ensure that the structural interpretation accurately describes the ensemble of potential solution conformations.

To calculate the solution state of hsA-SCS (consisting of hsSUCLG1$_{\Delta N51}$ and His$_6$hsSUCLA2$_{\Delta N41}$) using SAXS, four protein concentrations (30 mg/ml, 15 mg/ml, 7.5 mg/ml and 3.75 mg/ml) of hsA-SCS in the as purified form were tested in size exclusion chromatography coupled SAXS (SEC-SAXS) experiments. SEC coupling

was used as SAXS measurements represent the average state of particles in solution. Therefore, any aggregates or size heterogeneity that could affect overall scattering of the particles in solution would be separated out by a SEC column for independent analysis, provided the resolving power of the column is capable of separating different states. Initial analysis of the SEC elution profiles reveal an inverse correlation between the concentration of the sample tested and the particle elution volume (Figure 5-13), indicating an increase in the hydrodynamic radius of the particle at high concentration. Plotted onto the size exclusion traces for each concentration is the auto estimated radius of gyration (Rg) of each frame measured in the SAXS experiment, which also increases with protein concentration (Figure 5-13 y-axis right hand side). From this behaviour it can be inferred that at higher concentration, the particles representing hsA-SCS complex have either increased in size or have significantly altered conformation.

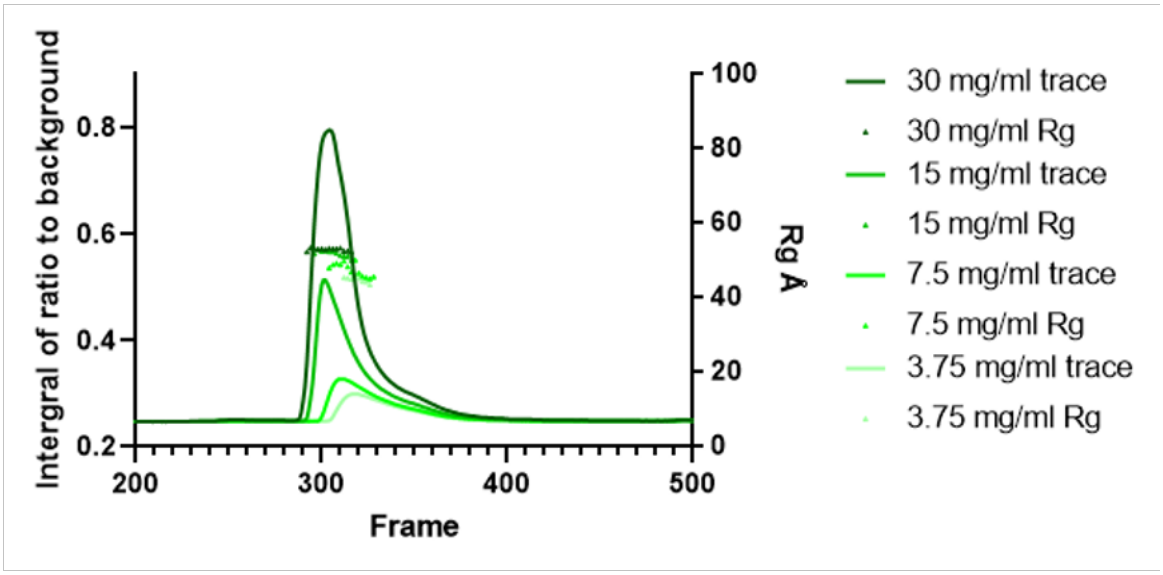


Figure 5-13: Size exclusion chromatography coupled SAXS (SEC-SAXS) of as purified hsA-SCS.

Size exclusion profile overlay of the as purified hsA-SCS concentration range with estimated Rg values plotted on right Y axis for each frame. Trace represents the integrated area of the ratio, of the

SAXS curve representing each sample frame (X-axis, SAXS measurement taken every 3 seconds of the SEC elution time course), to the estimated background (left Y-axis).

To determine the SAXS scattering profile of the hsA-SCS complex, each concentration, elution peaks containing frames with similar Rg values were buffer subtracted, merged and displayed in a Log*10* intensity plot for the concentration series (Figure 5-14). The overlaid scattering curves reveal significant changes in the shapes of each scattering curve as concentration is increased, indicating concentration induced changes in the conformation and/or size of hsA-SCS.

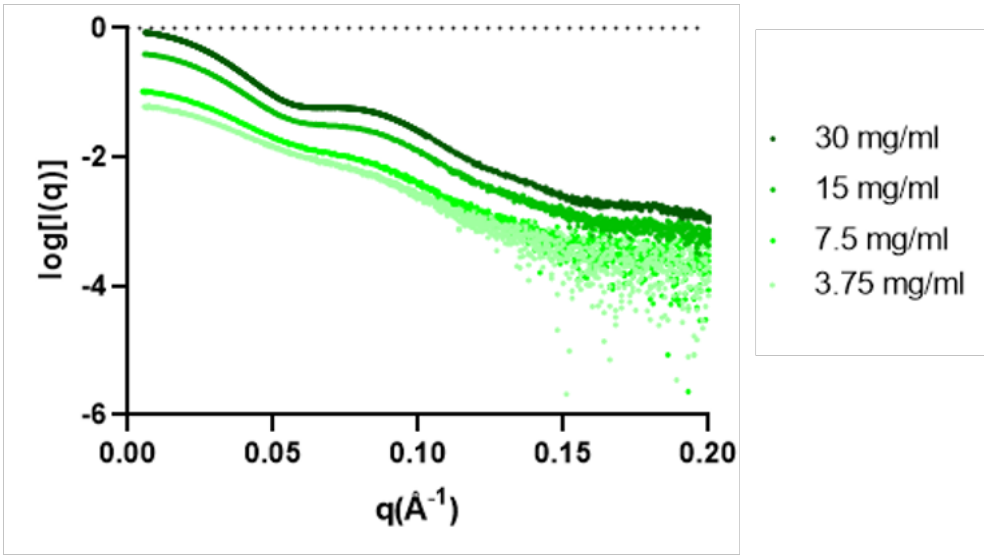


Figure 5-14: SAXS Log10 intensity plots for hsA-SCS

Plots show SAXS scattering profiles from the averaged and buffer frames under the elution peak at each concentration range. Intensity of scatter (y axis) is plotted against angle of scatter (X axis). If shape and size was independent of concentration plots one would see a parallel increase in intensity (y axis) between curves as concentration increases. Deviation from this, as seen here, indicates concentration dependency on size and shape of the average particle.

The more accurately calculated Rg (Table 5-2) using both Guinier approximation (slope of the low q region (Figure 5-15)) and the P(r) distribution (a real space depiction of the scattering data showing distribution of vectors within the average particle) (Figure 5-16) and extrapolated intensities at zero scattering angle (I(0))

(normalised for concentration), also increase with concentration (Table 5-2). Both traits are indicative of particles that undergo a concentration dependent increase in size.

Protein conc mg/ml	3.75	7.5	15	30
Guinier				
I(0)	0.06	0.11	0.41	0.88
I(0)/protein conc)	0.016	0.014	0.027	0.029
Rg (Å)	45.86	48.46	52.38	53.27
P(r)				
I(0)	0.06	0.11	0.41	0.88
I(0)/protein conc)	0.016	0.014	0.027	0.029
Rg (Å)	46.63	48.75	52.05	52.75
Dmax	138	144	147	140
Particle size				
MW (kDa)	110	120	160	180
Model fitting				
Crystal dimer fit (X^2)	91.2	427.4	4626.7	16351.4

Table 5-2 SAXS scattering table of parameters for hsA-SCS concentration range.

Values for radios of gyration (Rg), zero scattering angle (I(0)), and normalised zero scattering angle (I(0)) for protein concentration (I(0)/protein conc) are calculated by Guinier fitting and real space distribution plots are. Dmax values calculated by real space distribution plots. MW values calculated by the Vc/Rg^2 method[100]. Crystal dimer fit using the hsA-SCS structure is calculated using *Oligomer*.[89]

Linear fits in the Guinier region show that the sample is free from aggregation effects across the concentration range (Figure 5-15), indicating that during the size transition the particles remain monodisperse.

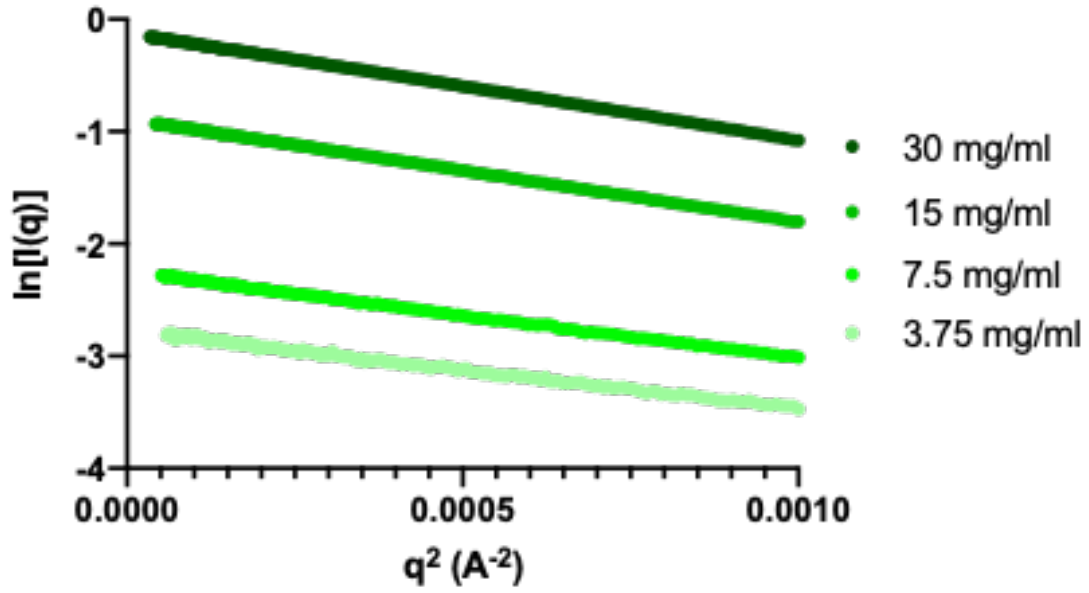


Figure 5-15 Linear plots in the SAXS guinier region for hsA-SCS concentration series.

Transform of the log10 intensity plot in the low q region for linear guinier fitting. Intensity of scatter (Y-axis) vs by angle of scatter squared (X axis).

Since it is apparent that the hsA-SCS complex undergoes a concentration dependent increase in size, it will be useful to know the MW of these potential states. There are multiple ways to estimate MW using SAXS data. The traditional approach using Porod Debye volume estimation is reliable for globular proteins only, and often results in overestimation of size for particles that deviate from the globular form, such as protein complexes. For this reason, MW can be calculated alternatively using the Vc/Rg^2 method[100] as an invariant proportional to molecular mass. The expected smallest particle size of the hsA-SCS complex would be 80 kDa (30 kDa corresponding to one SUCLG1 protomer and 50 kDa corresponding to one SUCLA2 protomer), corresponding to the hetero-dimer revealed by X-ray crystallography. At the lowest concentration, the calculated MW is already above the mass of the hetero-dimer at 110 kDa, and increases to 180 kDa at the highest concentration, revealing clear evidence of concentration dependent oligomerisation (Table 5-2).

The maximum size of the oligomer is not possible to calculate, as no plateau is reached across the concentration range, indicating the size transition has not gone to completion. In addition, as SAXS data indicate an average of all possible states, it is most likely that each data point represents a change in the equilibrium between dimer and an oligomer of unknown size, as opposed to the observation of multiple discrete single states. For these reasons it is difficult to interpret a single quaternary assembly for the datasets. SAXS data is usually used to calculate *ab initio* models providing a low-resolution model of the average particle in solution. Again, as it is expected that each dataset represents a mixture of unknown particle sizes it would be difficult to calculate individual models of these states from this data alone.

P(r) distribution functions are a plot of the real space distribution (Y-axis) of inter particle distances (vectors) (X-axis) and reveal the probability of finding a point within the particle at a distance r (Å) from another given point. For example a homogeneous globular protein would be expected to show a bell shaped distribution of vectors in P(r) plot[123]. In this study, the P(r) function derived from the hsA-SCS as purified sample reveals a double peak (Figure 5-16), which is often indicative of multidomain proteins or homo-oligomerisation[123]. The first peak would represent the proportion of vectors contributed by the asymmetric unit of the oligomer, and the second peak would then represent the proportion of vectors in the larger oligomeric form. A switch from a positive to negative ratio between the heights of first and second peaks, as concentration increases, provides further evidence of concentration dependent oligomerisation in the sample (Figure 5-16). Across the concentration range the reported Dmax (longest scattering vector within sample on X axis) remains the same at 140 Å. No decrease in Dmax indicates that even at the lowest concentration tested, a proportion of higher order oligomer is present (Figure

5-16). This result is in line with the MW calculation of the particle at this concentration which is already above the expected dimer MW.

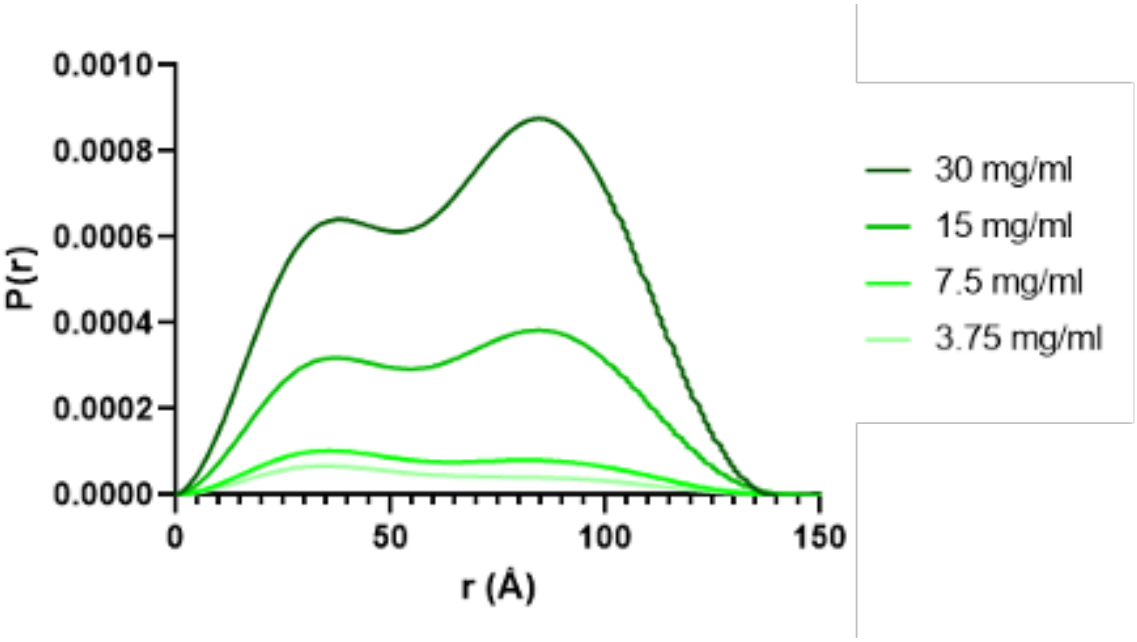


Figure 5-16: SAXS real-space P(r) distributions for the hsA-SCS concentration series.

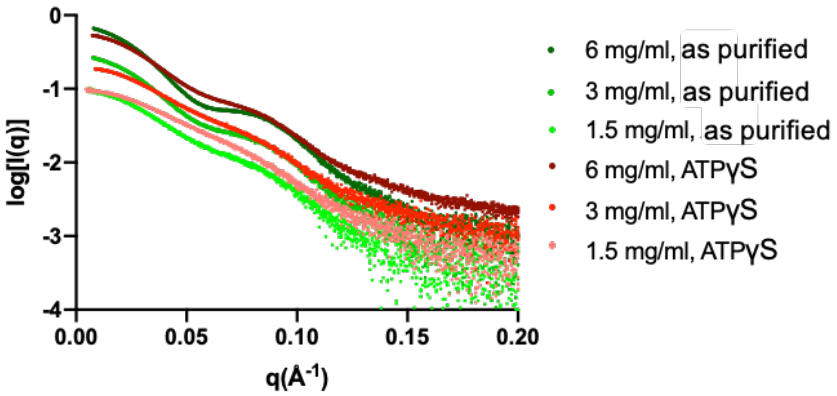
Finally, the theoretical scattering curve for the crystallographic hetero-dimer was calculated and fitted to the experimental curves using the programme *Oligomer*[89]. The purpose of this was to estimate how well the high-resolution model describes the ensemble of potential solution states. The χ^2 (a test for goodness of fit) results clearly reveal a very poor fit for all concentrations, indicating the dimer alone, as seen in the crystal structure, does not sufficiently describe all the states that are present in solution (Table 5-2). Additionally, the quality of fit negatively correlates with protein concentration, again indicating size increases relative to the hetero-dimer (Table 5-2).

5.4.2 Ligand dependent oligomerisation determined by SAXS

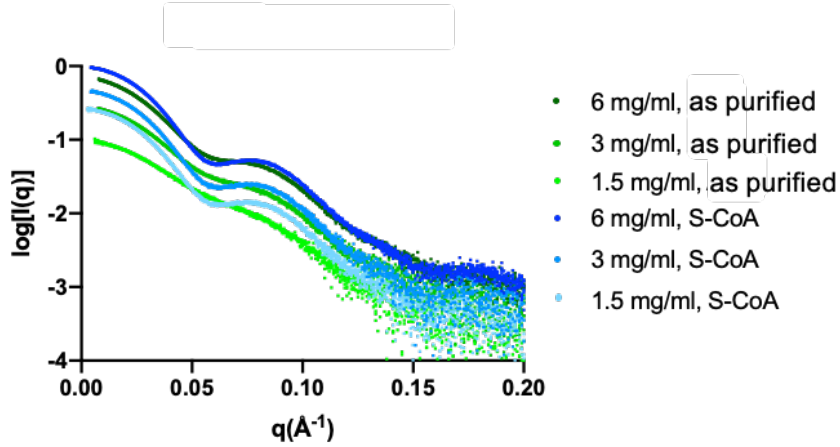
One consideration as to whether the oligomerisation observed in the SAXS studies of the as purified datasets is physiologically relevant is that it requires high protein concentration, above 7.5 mg/ml, to become a prominent feature in solution. These concentrations would not be expected in the cellular environment. In order to establish whether the oligomer might exist at more physiological concentrations, SAXS was again used to screen for substrates of A-SCS that may help to stabilise the different states of the complex at lower protein concentrations.

Three concentrations of hsA-SCS at 6 mg/ml, 3 mg/ml and 1.5 mg/ml were tested using batch mode SAXS (not coupled with SEC) with 5 mM of either ATPγS (a non-cleavable analogue of ATP), ADP, succinyl-CoA (S-CoA), succinate or CoA (all at ≥ 66 fold molar excess). SEC was omitted from the experiment as the column used did not have the resolving power to separate the different oligomers and no aggregates were visible in the elution profiles. Additionally, experiments without coupling to size exclusion allow for more rapid testing of multiple concentrations and ligand states, and enable determination of the effect of ligands on the equilibrium of hsA-SCS particle size. One drawback of the batch mode is that the introduction of aggregates induced after incubation with ligands are not filtered during the experiment. Samples were inspected visually and centrifuged to overcome this. After incubation with ADP, the protein immediately precipitated and consequently this sample was not tested. For the other ligands, frames with similar Rg and I(0) values were buffer subtracted, merged and displayed in a Log*10* intensity plot for each protein concentration series overlaid with the as purified dataset (Figure 5-17).

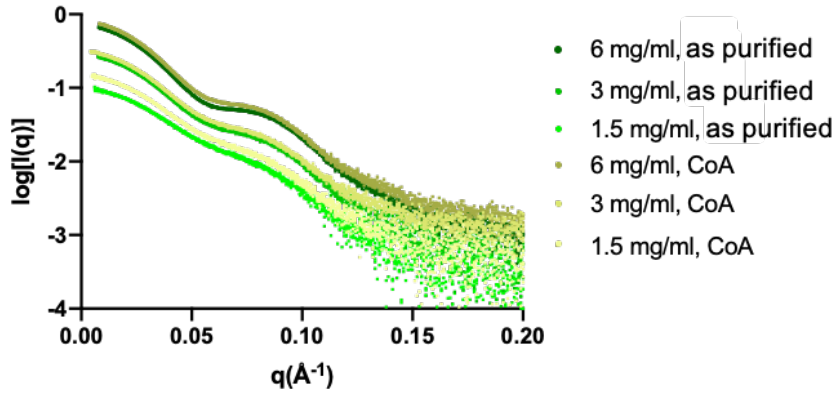
A


log[I(q)]
0
-1
-2
-3
-4
0.00
0.05
0.10
0.15
0.20
q(Å-1)
6 mg/ml, as purified
3 mg/ml, as purified
1.5 mg/ml, as purified
6 mg/ml, ATPγS
3 mg/ml, ATPγS
1.5 mg/ml, ATPγS

B


log[I(q)]
0
-1
-2
-3
-4
0.00
0.05
0.10
0.15
0.20
q(Å-1)
6 mg/ml, as purified
3 mg/ml, as purified
1.5 mg/ml, as purified
6 mg/ml, S-CoA
3 mg/ml, S-CoA
1.5 mg/ml, S-CoA

C


log[I(q)]
0
-1
-2
-3
-4
0.00
0.05
0.10
0.15
0.20
q(Å-1)
6 mg/ml, as purified
3 mg/ml, as purified
1.5 mg/ml, as purified
6 mg/ml, CoA
3 mg/ml, CoA
1.5 mg/ml, CoA

D

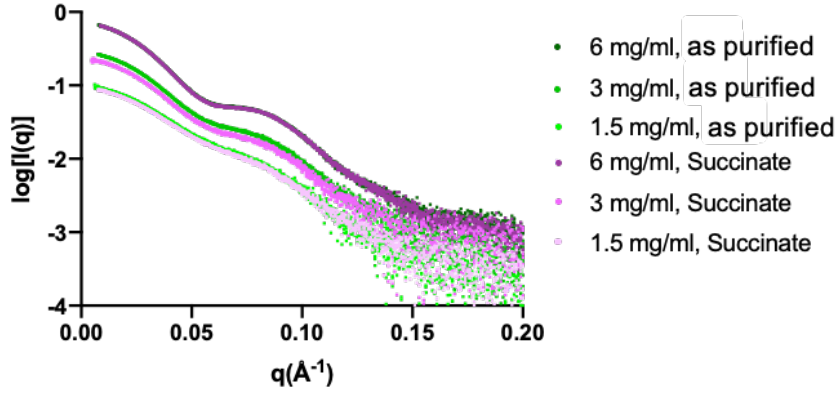

log[I(q)]
0
-1
-2
-3
-4
0.00
0.05
0.10
0.15
0.20
q(Å-1)
6 mg/ml, as purified
3 mg/ml, as purified
1.5 mg/ml, as purified
6 mg/ml, Succinate
3 mg/ml, Succinate
1.5 mg/ml, Succinate

Figure 5-17: Batch mode SAXS Log10 intensity plots of hsA-SCS with substrates.

A-D: Log10 intensity plots revealing SAXS scattering profiles from the averaged and buffer subtracted frames. Each substrate concentration series if overlaid with the as purified batch concentration series. Intensity of scatter (y axis) is plotted against angle of scatter (X axis).

Calculated Rg, I(0) (concentration normalised), MW and dmax are shown in Table 5-3 and reveal that the as purified sample in the batch mode behaves very similarly to the original SEC-SAXS experiment with a similar concentration dependent oligomerisation reported. Neither succinate nor CoA, which are products of the first and second reaction in the forward direction (Figure 5-3), appear to alter the equilibrium of dimer to oligomer as shown by an overlay of log intensity plots (Figure 5-17C,D) and reported MW (Figure 5-18D).

Incubation with ATPγS, an analogue of the final product in the forward direction and also a substrate of reverse reaction, however, causes a significant change across the scattering profile of the complex relative to the as purified dataset (Figure 5-17A). In this case, ATPγS shifts the equilibrium in favour of hetero-dimer formation, shown by both an increase in the ratio of peak 1 to peak 2 in the P(r) distribution (Figure 5-18B) and a reported MW of 86 kDa at the lowest concentration, which is very close to the expected hetero-dimer of 80 kDa (Figure 5-18D)(Table 5-3). The calculated Rg is much lower than seen in as purified, and the determination of Rg, normalised I(0) and MW shows less concentration dependent increases relative to the as purified dataset. However, a double peak in the P(r) distribution and some increase in MW at the highest concentration (Figure 5-18B, Figure 5-18D)(Table 5-3) would still indicate that a small proportion of oligomer is present across the 3 concentrations with some concentration dependent increase.

Addition of succinyl-CoA, the first substrate of the forward reaction or final product of the reverse reaction, also significantly changes the Log10 intensity profile compared to the as purified dataset (Figure 5-17B). The equilibrium appears to shift in favour of oligomerisation, shown by a decrease in the ratio of peak 1 to peak 2 in the P(r) distribution (Figure 5-18C), and increase in the reported MW relative to the as purified dataset (Figure 5-18D). Unlike previous conditions, in the presence of S-CoA, the calculated values for Rg, MW, normalised I(0) (Table 5-3) and the ratio of the two peaks in the P(r) distribution (Figure 5-18C) do not change with concentration, indicating that there is very little variation in particle behaviour across the three concentrations. One can therefore assume that the S-CoA bound sample better represents a homogenous population of the oligomeric state. The calculated MW (aprox 310 kDa) for the higher order oligomeric species might indicate a hetero-octamer comprised of 4 hetero-dimers with a theoretical MW of 320 kDa.

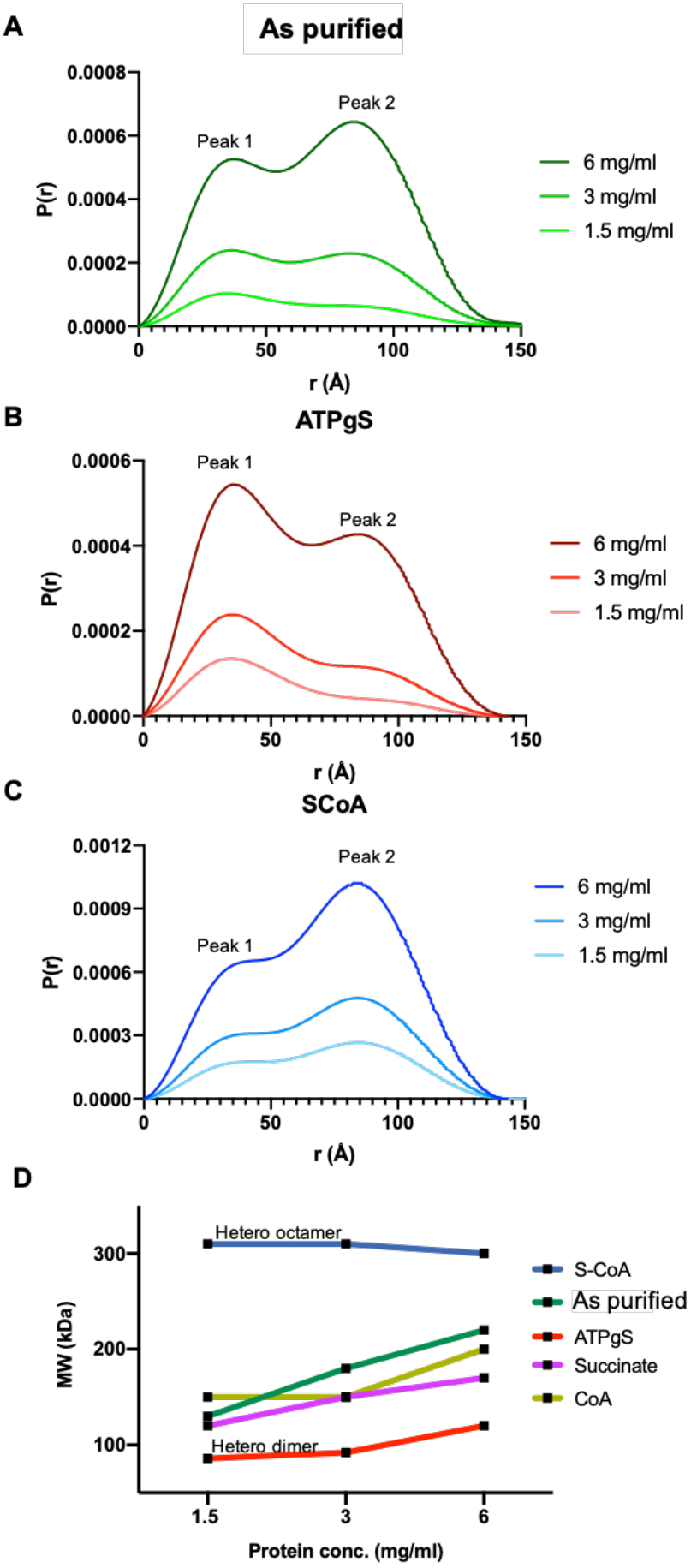

A
As purified
0.0008
0.0006
0.0004
0.0002
0.0000
P(r)
Peak 1
Peak 2
0
50
100
150
r (Å)
6 mg/ml
3 mg/ml
1.5 mg/ml
B
ATPgS
0.0006
0.0004
0.0002
0.0000
P(r)
Peak 1
Peak 2
0
50
100
150
r (Å)
6 mg/ml
3 mg/ml
1.5 mg/ml
C
SCoA
0.0012
0.0009
0.0006
0.0003
0.0000
P(r)
Peak 2
Peak 1
0
50
100
150
r (Å)
6 mg/ml
3 mg/ml
1.5 mg/ml
D
300
200
100
MW (kDa)
Hetero octamer
Hetero dimer
1.5
3
6
Protein conc. (mg/ml)
S-CoA
As purified
ATPgS
Succinate
CoA

Figure 5-18: SAXS P(r) distribution and MW calculation for A-SCS ligand bound states.

A-C: Real-space p(r) distributions for the concentration series of each ligand state **D:** MW trends of each concentration A-SCS concentration series in the presence of the different substrates.

To rationalise how a hetero-octameric complex can be assembled, three plausible arrangements of the four hetero-dimers can be envisaged. These arrangements are mediated by different potential symmetries and include an asymmetric (C1), cyclic (C4 symmetry) or dihedral (D2 symmetry) arrangement. To view the overall shape of these arrangements, *ab initio* models imposing the three possible symmetries were created using twenty rounds of *Dammif*[89]. These models were aligned, averaged and filtered with the *Damaver* program[89] to provide final low-resolution envelopes of the complex with C1, C4 or D2 symmetry imposed. Processing in C1 reveals an asymmetric envelope with a clear helical turn and enough volume to fit four hetero-dimers (Figure 5-19A). This type of helical shape has previously been shown to be a common artefactual reconstruction of a ring-shaped particle with a central cavity without imposing symmetry[124]. Processing in C4 symmetry results in a less plausible solution with loose and disconnected density (Figure 5-19B). Processing in D2 symmetry results in a ring-shaped envelope containing a central cavity with enough volume to fit four hetero-dimers (Figure 5-19C). Although anecdotal evidence points towards a D2 symmetry arrangement, it is impossible to prove the exact arrangement of A-SCS using SAXS data alone.

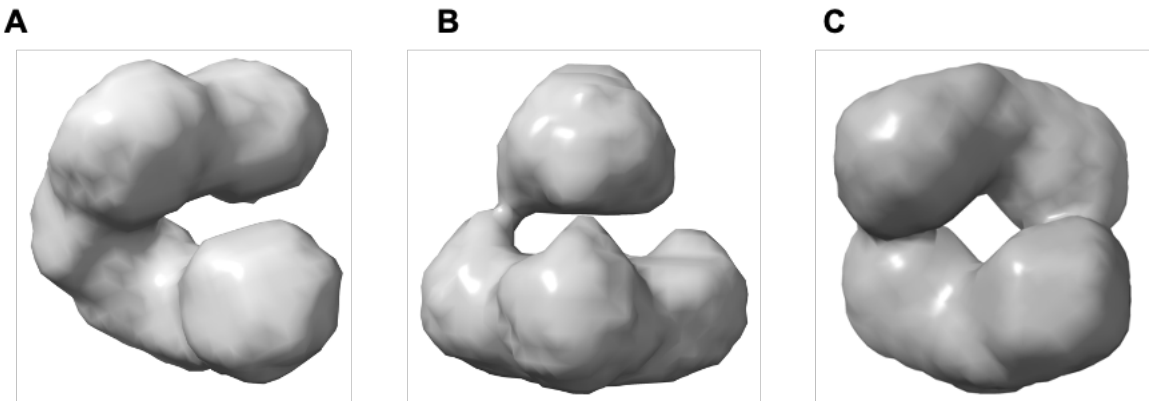


Figure 5-19: Ab initio SAXS models of A-SCS in presence of succinyl-CoA.

A: *Dammif* model calculated with C1 symmetry imposed. **B:** *Dammif* model calculated with C4 symmetry imposed. **C:** *Dammif* model calculated with D2 symmetry imposed.

Finally, the *Oligomer* programme was used to determine the fit of the crystallographic hsA-SCS hetero-dimer theoretical scattering curve to the experimental datasets in presence of each ligand (shown by crystal dimer fit χ^2 row in Table 5-3). Tor the batch as purified dataset, like the SEC-SAXS experiments, the χ^2 fit is again poor and negatively correlates with protein concentration. In the presence of ATPγS, while still poor, the fit improved relative to as purified across the concentration range, indicating the ATPγS dataset is closer to the dimer state but cannot be fully described by the dimer alone. The fit still showed negative correlation with protein concentration. As expected from the MW calculations in the presence of succinyl-CoA, the dimer fit was significantly worse across the entire concentration range (Table 5-3).

	As purified			ATPγS			S-CoA			Succinate			CoA		
Protein conc mg/ml	1.5	3	6	1.5	3	6	1.5	3	6	1.5	3	6	1.5	3	6
Guinier															
I(0)	0.01	0.27	0.69	0.01	0.20	0.56	0.26	0.46	0.97	0.09	0.22	0.69	0.14	0.31	0.79
I(0)/protein conc)	0.06	0.09	0.12	0.07	0.07	0.09	0.17	0.15	0.16	0.06	0.07	0.12	0.09	0.10	0.13
Rg (Å)	46.5	50.7	53.4	38.9	42.8	47.8	54.9	54.7	54.1	46.9	51.1	53.6	49.1	51.0	53.6
P(r)															
I(0)	0.1	0.27	0.69	0.1	0.2	0.56	0.26	0.46	0.97	0.09	0.22	0.69	0.14	0.30	0.78
I(0)/protein conc)	0.067	0.09	0.12	0.067	0.066	0.093	0.17	0.15	0.16	0.06	0.07	0.12	0.09	0.10	0.13
Rg (Å)	47.0	50.9	52.5	40.3	44.3	48.5	53.9	53.88	53.4	47.2	50.6	52.5	48.4	50.6	52.5
Dmax	136.8	154.9	156.5	122.6	139.3	143.4	148.1	150.0	147.0	135.0	137.8	143.5	128.8	140.3	0.13
Particle size															
MW (kDa)	130	180	220	86	92	120	310	310	300	120	150	170	150	150	200
Model fitting															
Crystal dimer fit (X^2)	88.8	347.5	1595	31.7	80.7	959.8	1898	1218	4471	169.1	133.4	1806	127.7	387.1	1079

Table 5-3: SAXS table of parameters for A-SCS ligand bound states

Values for radios of gyration (Rg), zero scattering angle (I(0)), and normalised zero scattering angle (I(0)) for protein concentration (I(0)/protein conc) are calculated by Guinier fitting and real space distribution plots are. Dmax values calculated by real space distribution plots. MW values calculated by the Vc/Rg^2 method[100]. Crystal dimer fit using the hsA-SCS structure is calculated using *Oligomer*.[89]

The SAXS experiments have successfully confirmed the presence of a predominant hetero-dimeric species in the presence of ATPγS and conversion to a potential hetero-octameric species in the presence of succinyl-CoA. This latter larger species is of sufficient size to be studied by cryo-EM. Direct visualisation of single particles by cryo-EM would serve to confirm that the hetero-octameric quaternary assembly exists, identify the symmetrical arrangement of A-SCS subunits within the oligomer, and hopefully provide higher resolution information about its composition.

5.4.3 Cryo-EM determination of the hsA-SCS hetero-octamer induced by succinyl-CoA

In the previous section, the SAXS data provided proof of principle that the A-SCS complex, in the presence of succinyl-CoA, could be studied further with single particle cryo-electron microscopy, because it represents a homogenous population at a MW that can be easily visible using electron microscopy.

The hsA-SCS samples pre-incubated with succinyl-CoA were vitrified directly onto cryo-EM grids. Grids were screened using a Glacios 200 keV microscope (at the Oxford Particle Imaging Centre (OPIC)) with a magnification of 150000x and directly imaged with a Falcon III direct electron detector. A preliminary dataset containing 200 micrographs was collected. Particle auto picking using *Relion*[90] and filtering away junk particles using rounds of 2D classification led to the selection of 6175 particles, representing a range of orientations of the complex (Figure 5-20).

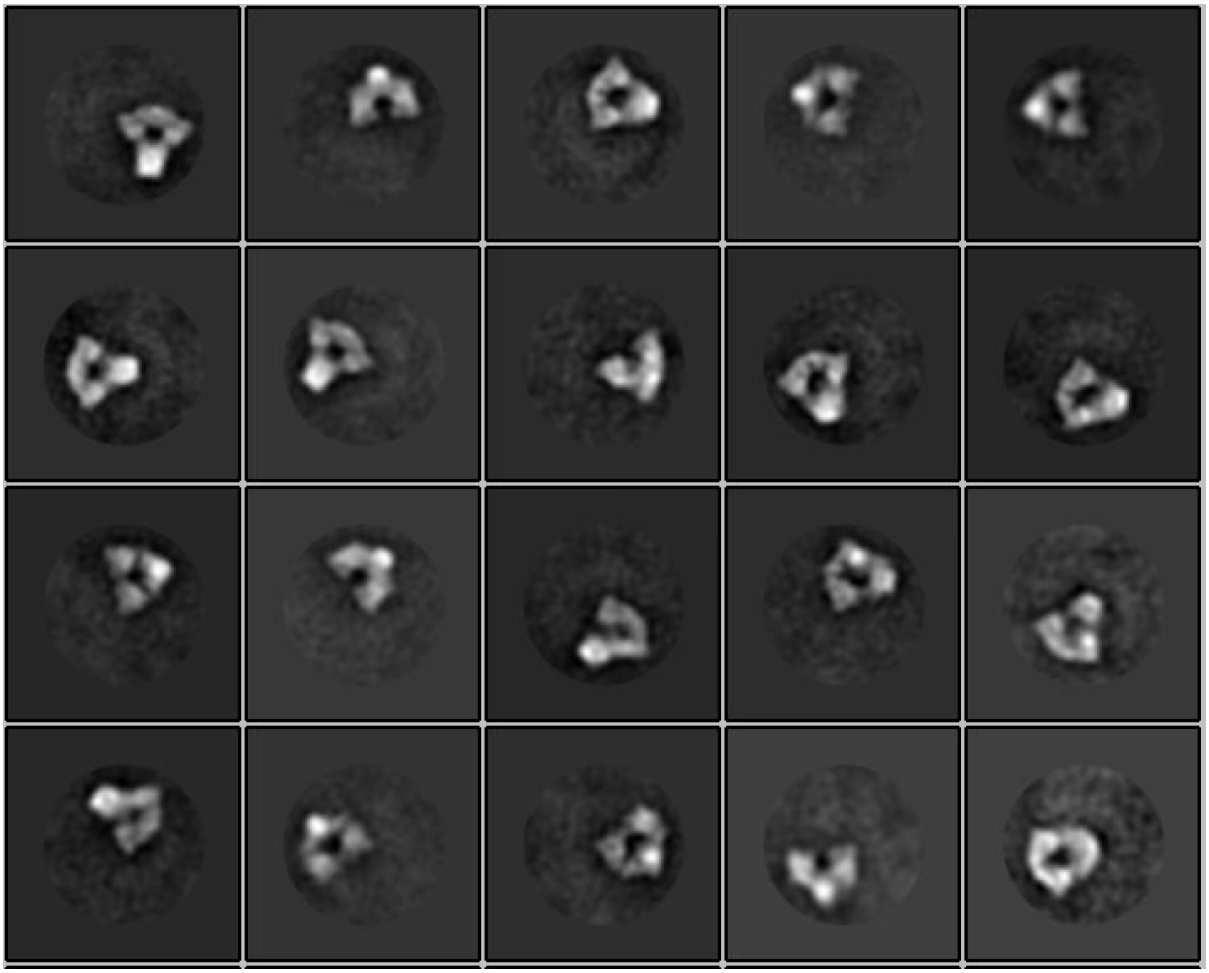


Figure 5-20: 2d classes hsA-SCS with succinyl-CoA:

Each panel represents a group of 2D aligned and averaged particles occupying the same orientation in the micrographs. Particles are displayed as white density.

These particles were used to generate a 3D reconstruction with initial model building in *Relion*[90]. The initial model was used as a reference for a round of 3D classification to identify and group discrete conformations into 4 separate classes, however the classes did not reveal visable heterogeneity within the pool of 6175 particles. The initial model was therefore used as a reference for 3D refinement of the 6175 particles resulting in a final map at 8 Å global resolution (Figure 5-21A). The global resolution was determined by Fourier shell correlation (FSC) to be 8 Å using a standard cut-off of 0.143, proposed to provide a resolution to density scale that is

comparable with X-ray crystallography. Like crystallography, the resolution can vary substantially across the map. For this dataset, the local resolution estimates[90] predict the resolution to range from 6.8 Å to 8.9 Å (Figure 5-21B).

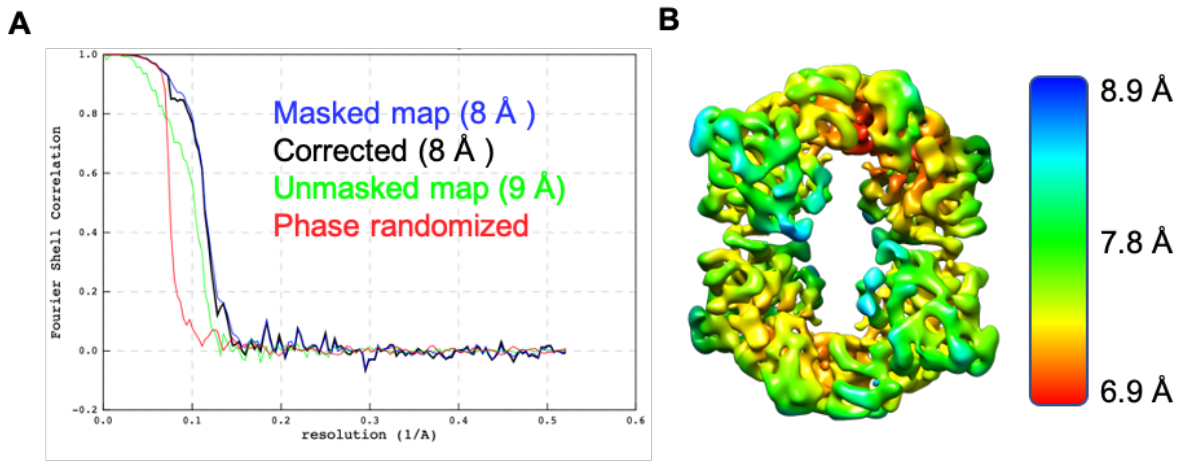


Figure 5-21: Cryo-electron microscopy of hsA-SCS with succinyl-CoA:

A: FSC curves of phase randomised map showing global resolution of 8 Å for the masked map and 9 Å without masking. **B:** Local resolution of the maps scaled between 6.9 and 8.9 Å.

This reconstruction allowed the rigid body placement of 4 αβ hetero-dimers, derived from the crystallographic model in this study, into the density (Figure 5-22A). Like the *E.coli* hetero-tetrameric forms of SCS[110], the higher order oligomerisation of the A-SCS αβ-hetero-dimer, within the hetero-octamer, is mediated entirely by the β-subunits SUCLA2. However, the interaction interfaces in the hetero-octameric form are markedly different from the *E.coli* hetero-tetramer, and occur at completely different sites of SUCLA2.

Instead of a single large interaction interface (buried surface area 1363 Å^2) with cyclic (C2) symmetry between the two αβ-hetero-dimers in ecSCS[110], human A-SCS displays dihedral (D2) symmetry between the four αβ-hetero-dimers (Figure 5-22CA). In this arrangement two smaller interaction interfaces (1 and 2, Figure 5-22CA), previously unobserved, occur in two of the three two-fold symmetry axis.

The third two-fold symmetry axis does not contain any interaction interface, resulting in a ring shaped structure with a central cavity size of approx. 80 Å in the longest direction and 30 Å in the shortest (Figure 5-22CA). Overlay of the D2 symmetry imposed dammif model previously calculated from the SAXS data (Figure 5-19C) shows that the in-solution derived model is a very close fit to the cryo-EM model (Figure 5-22B).

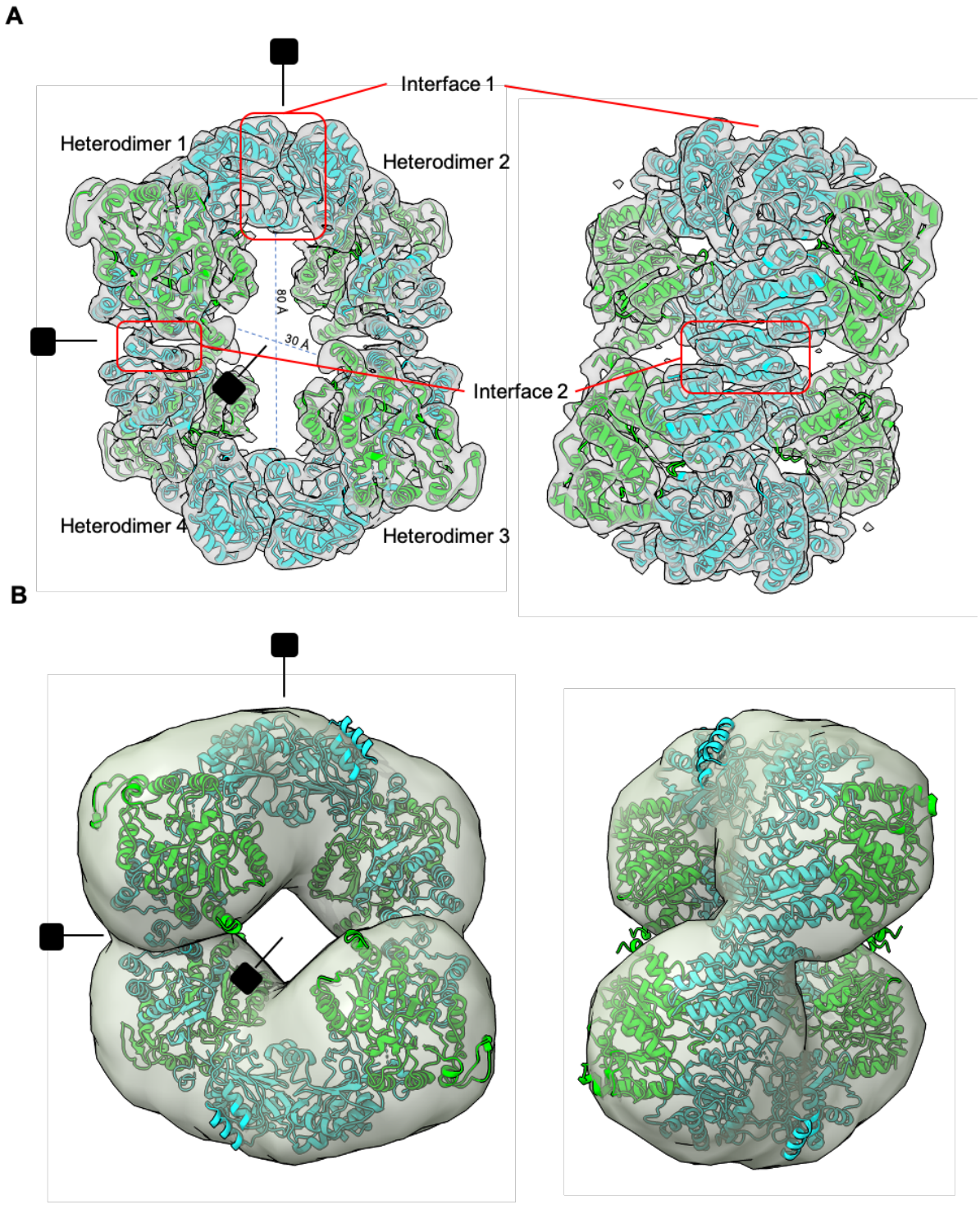


Figure 5-22: Modelling into the EM and SAXS maps.

A: Electron microscopy map showing D2 symmetry with hetero-octameric model of hsA-SCS overlaid after rigid body fitting and subsequent modelling using *Namdinator*[125]. The 4 SUCLA2 subunits are coloured cyan and the 4 SUCLG1 subunits are coloured green. **B:** SAXS *ab initio Dammif*[89] model showing D2 symmetry with hetero-octameric cryo-EM model of hsA-SCS overlaid using *Supcomb*[89].

In both panels, two-fold (black box) symmetry axes (stick) are indicated. Right structure rotated 90° along Y axis relative to left in both panels.

Within the dihedral arrangement, the four α-subunits (one from each hetero-dimer) are positioned at each pole of the complex, separated by a minimum 30 Å in the shortest direction, preventing direct contacts with α or β subunits from opposing hetero-dimers. Interface 1 in the hetero-octameric state is formed by loop Glu159-Arg194 and the ATP grasp domains of two SUCLA2 subunits (Figure 5-23A). Interface 2 is formed at the C-terminus of two SUCLA2 subunits, residues Arg438 to Ile463 (Figure 5-23B) which contains the 16 residue C-terminal extension unique to SUCLA2.

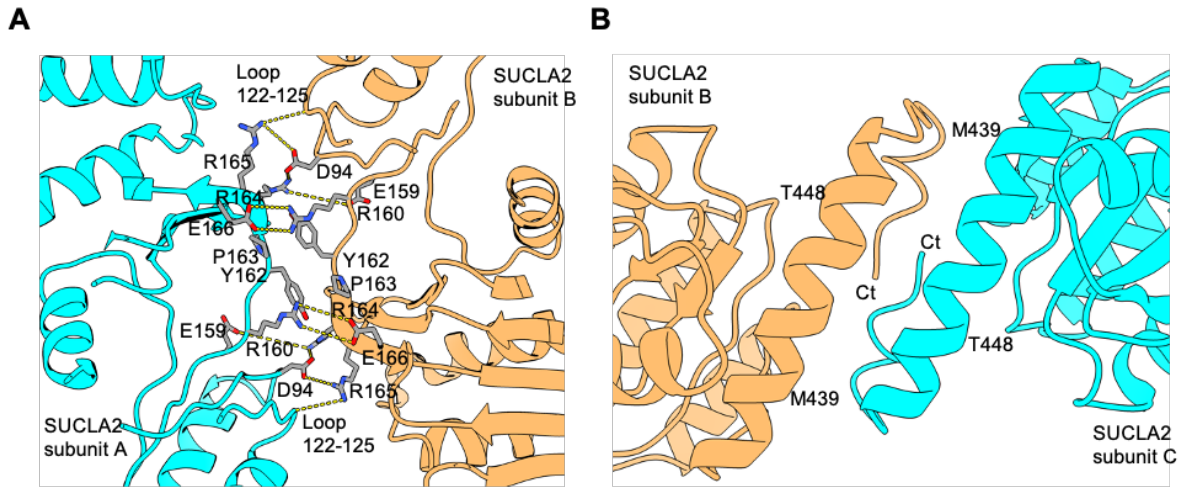


Figure 5-23: Two interaction interfaces of the A-SCS hetero-octamer

A: Dimerisation interface 1 of the cryo-EM model. The two SUCLA2 subunits from different hetero-dimers are coloured orange and cyan. Potential interacting residues are displayed as sticks. **B:** Dimer interface 2 of the cryo-EM model. The two interfacing SUCLA2 subunits are coloured orange and cyan.

The program *Namdinator*[125] was used to model conformational changes of the rigid-body fitted hetero-dimers, to better match the low-resolution EM density. It is clear that SUCLA2 subunits undergo significant rotational movement, in order to form both

interfaces. In the case of interface 1, superimposition of the two structures shows that part of the ATP grasp lid domain, residues 74 to 159, rotates at a hinge point formed by the loop connecting Glu159-Arg164, which also directly takes part in the oligomerisation interaction, and the loop connecting Glu66-Gly74. This movement leads to a 4.5 Å shift of the ATP grasp domain lid, resulting in a more open state of the ATP binding sites that is presumably ready for nucleotide binding (Figure 5-24A).

The conformational change of the ATP grasp lid appears to be driven by ordering or movement of loop 290 to 303 leading to the rotation at the identified hinge points. The majority of loop 290 to 303 is disordered in the crystal structure of hsA-SCS, however, the same region directly interacts with the nucleotide grasp domain in ssG-SCS through hydrogen bonding of Glu244 (Gln296 in hsA-SCS) to the backbone of Leu88 (Leu102 in hsA-SCS). The low-resolution density present in the EM data appears to follow a similar trajectory as the G-SCS structure in this region (Figure 5-24B), indicating a similar interaction may occur in hsA-SCS.

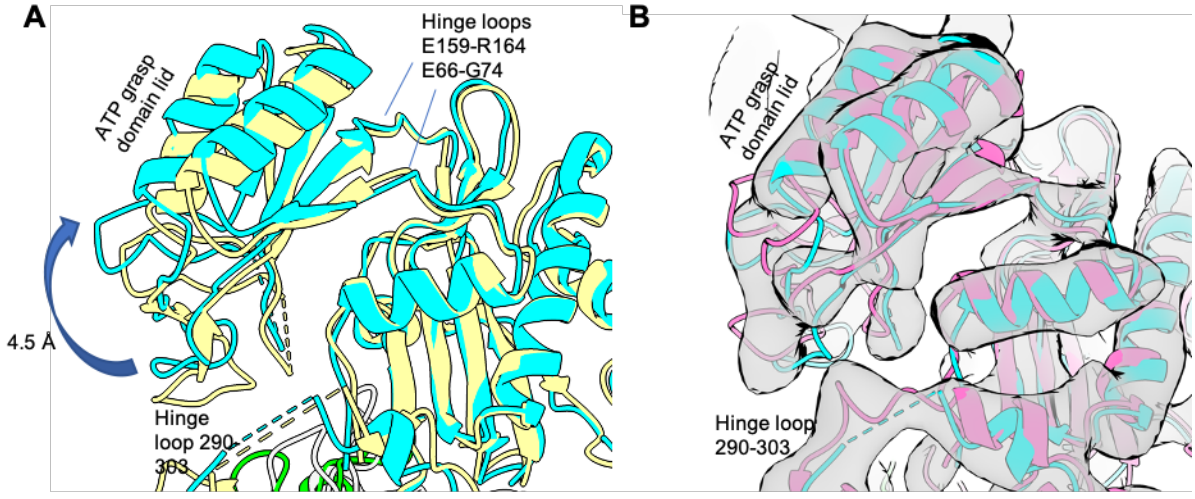


Figure 5-24 conformational change of the ATP grasp lid.

A: Zoomed cartoon depiction of the cryo-EM model showing ATP grasp domain lid of SUCLA2 (cyan) overlaid with the crystal structure of hsA-SCS in this study (yellow). **B:** Zoomed cartoon depiction of

cryo-EM model showing ATP grasp domain lid (cyan) overlaid with the crystal structure of ssG-SCS (pink) and EM density (grey), revealing the ordering of proposed hinge loop 290-303.

In addition to altering the ATP grasp domain lid conformation, the 290 to 303 loop also appears to allow conformational change at the C-terminus by forming a hinge region for rotation of the entire C-terminal domain, residues 303 to 463, leading to a 9.4 Å shift at the C-terminus relative to the crystal structure of hsA-SCS (Figure 5-25A). The C-terminal extension residues 457 to 463, only present in the SUCLA2 subunit in A-SCS and not in the equivalent SUCLG2 of G-SCS, also undergo conformational change from a unstructured extension seen in the crystal structure to a two-turn helix directly at interaction interface 2 (Figure 5-25A). Since residues 380-382, and 350 of the SUCLA2 C-terminal domain form part of the shared substrate binding site for succinate with SUCLG1[112], the conformational change of the entire C-terminus could be driven through interaction of succinyl-CoA or succinyl-phosphate with this region of SUCLA2. This could represent a state of the enzyme when it has turned over reaction step 1 (Figure 5-3) and is poised ready for reaction step two to take place. The rotation of the entire C-terminal domain is comparable to the rotation of this domain seen within the ssG-SCS succinate bound structure (5CAE) (Figure 5-25B), although in the ssG-SCS structure the movement is more subtle and does not result in rotation of loop 290 to 303. In the hsA-SCS structure, reordering of this hinge loop, which shows low sequence conservation with SUCLG2, propagates a C-terminal conformational change to the more distant ATP grasp domain at the first interaction interfaces, providing a mechanism whereby substrate binding at the C-terminus could allow for both, seemingly distant, domains to be favourably positioned for oligomerisation. The central ATP grasp domain formed by residues 53 to 71 and 165 to 290 does not change significantly in conformation

(Figure 5-24A) relative to the crystal structure. *HingeProt*[126] was next used to independently validate identified hinge regions using the hetero-dimer crystal structure alone. Three hinge sites were predicted at residues 71, 161 and 291 in SUCLA2 that are well matched to the identified hinge loops by superimposition of the two hsA-SCS conformations. (Figure 5-24A).

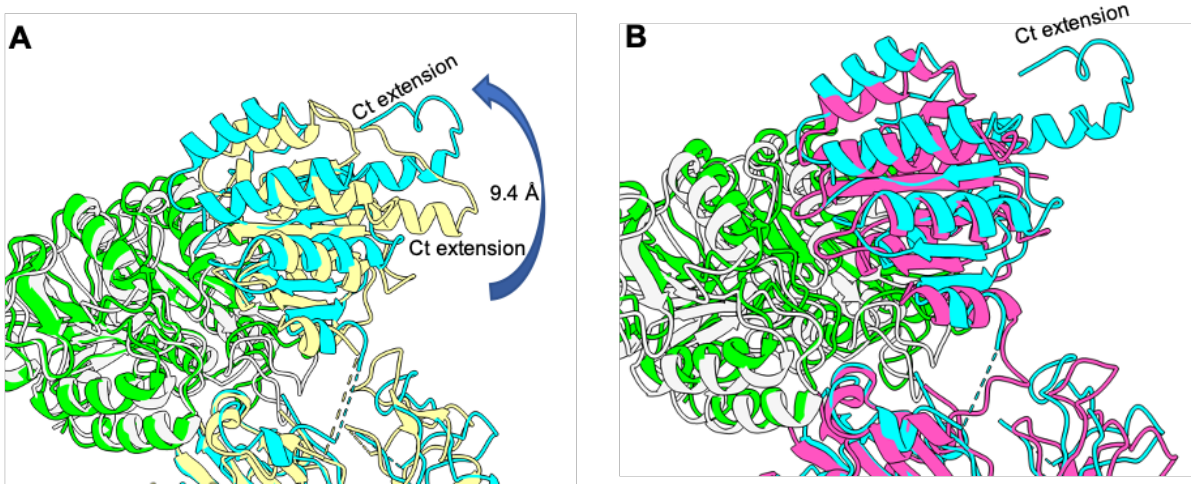


Figure 5-25: Conformational change of the C-terminal domain.

A: Zoomed image of cryo-EM model showing SUCLA2 Ct-domain (cyan) overlaid with the crystal structure of hsA-SCS (yellow) SUCLG1 cryo-EM structure is coloured Green and SUCLG1 from the crystal structure is coloured white. **B:** Zoomed image of cryo-EM model showing SUCLA2 Ct-domain (cyan) overlaid with the crystal structure of succinate bound ssG-SCS (5CAE) (pink). SUCLG1 cryo-EM structure is coloured Green and SUCLG1 from ssG-SCS structure is coloured white.

Although the resolution of the data does not permit modelling at the amino acid sidechain level, it is possible to predict potential interactions at the hetero-octamer interfaces (labelled in Figure 5-22A) based on proximity in the structure (Figure 5-23A). At interface 1 it is likely that Arg164 is in close enough proximity to form salt bridge interactions with Glu159 and or Asp94 of the opposing SUCLA2 subunit. Arg165 is also in close proximity to interact Asp94 and or hydrogen bond with backbone residues of the ATP grasp domain loop formed by residues 122 to 125.

Tyr162 could form π/ π stacking interactions with Tyr162 or CH/π interactions with Pro163 of the opposing SUCLA2 subunit. Finally, Arg160 is also in close enough proximity to interact with Glu166 of the opposing SUCLA2 subunit.

Analysis of the interaction interface 1 at the sequence level also shows a clear diversion of hsA-SCS from both ssG-SCS and bacterial isoforms of the SCS complex (Figure 5-26A). The residue properties in the loop (residues 160-164, green in Figure 5-26A) involved in the interaction at the ATP grasp domain are not conserved between A-SCS and G-SCS. Of the potential interacting residues, positively charged Arg160 and Arg164 have been substituted by small hydrophobic or polar residues in G-SCS and bacterial SCS, such that they would no longer be able to form salt bridges or hydrogen bonds with opposing G-SCS β-subunits (Figure 5-26A, 1st and 5th residue in the sequence highlighted green). The equivalent residue of Tyr162 is aspartic acid in G-SCS that would not form π/ π or π/CH stacking interactions (Figure 5-26A, 3rd residue in the sequence highlighted green).

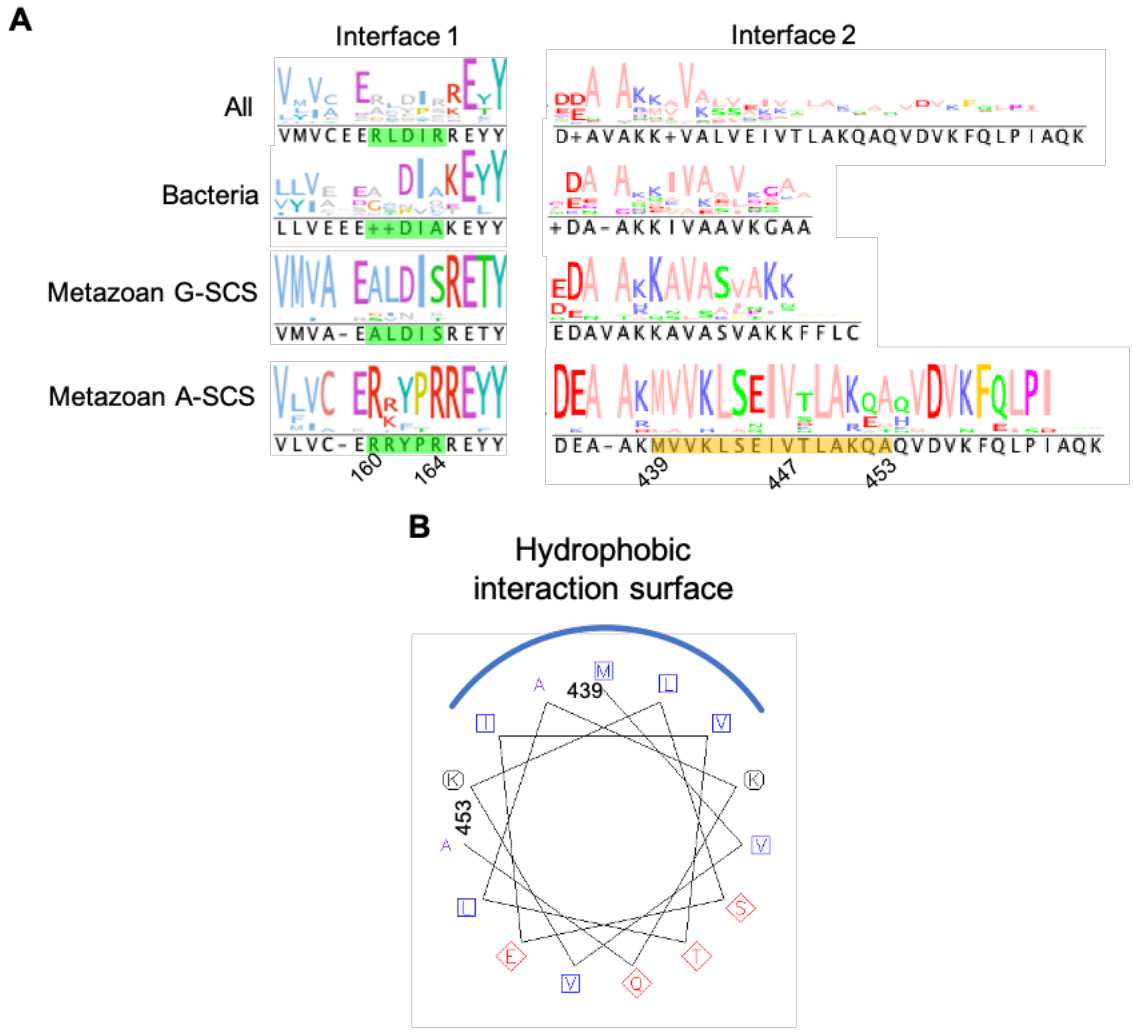


Figure 5-26: Sequence analysis at interaction interface 1 and 2.

A: Sequence consensus comparing sequences at interface 1 and interface 2 between ssG-SCS, hsA-SCS and bacterial SCS. **B:** Helical wheel diagram of SUCLA2 sequence Met439-Ala453 showing hydrophobic interface participating in dimeric, anti-parallel coiled-coil interactions.

At the C-terminal interaction interface (interface 2 in Figure 5-22A), residues within the sequence Met439 to Ile463 appear to be involved in interactions with the equivalent sequence of the opposing SUCLA2 subunit (Figure 5-23B). The sequence from Met439 to Val447 shows very low conservation compared to G-SCS and bacterial homologues (Figure 5-26A, first 9 residues in highlighted orange sequence). Additionally, the sequence from Thr448 to the C-terminal residue Ile463

is absent in G-SCS and bacterial homologues (Figure 5-26A, 10th residue in orange section to C-terminus). At the interaction site the helical region from Met439 to Ala453 (Figure 5-26A, entire orange sequence) displays a hydrophobic surface across the interface in a dimeric, anti-parallel manner flanked by charged or polar amino acids, indicative of a tight coiled-coil interaction (Figure 5-26B). In line with these observations, the coiled-coil prediction server *waggawagga*[127] indeed identifies Leu433 to Ala453 as a possible coiled-coil region based on sequence alone. The *logicoil* server[128] also predicts that the most likely orientation for the shorter interacting sequence Met439 to Ala453, as identified in the EM structure, is an antiparallel coiled-coil dimer. The loop and helical residues from residue 454 to the C-terminus are in close proximity to form further interactions with the tight dimeric coiled coil residues and residues within the loop 424 to 432 found within the opposing SUCLA2 subunit. Altogether, the sequence diversions at both the C-terminal and ATP grasp domain interfaces, and the lack of reported homo-oligomerisation of G-SCS, would indicate that the hetero-octamer formation has evolved in A-SCS only.

Inspection of active site 1 in the EM map reveals that the density for the catalytic loop containing His299 is missing in the EM data set, indicating conformational change has occurred relative to the crystal structure due to the presence of succinyl-CoA. The new loop conformation cannot be modelled as no density is visible in the dataset even with local sharpening. This loop could adopt either a flexible range of continuous conformations or multiple discrete conformations that did not uphold the D2 symmetry restraints applied during data processing. Analysis of the lower resolution data processed without symmetry also does not provide further information about the exact loop positioning. A larger dataset at higher resolution in

future may therefore provide better coverage of the discrete conformations present within these particles.

5.4.4 Intermediate assembly of the hetero-octamer

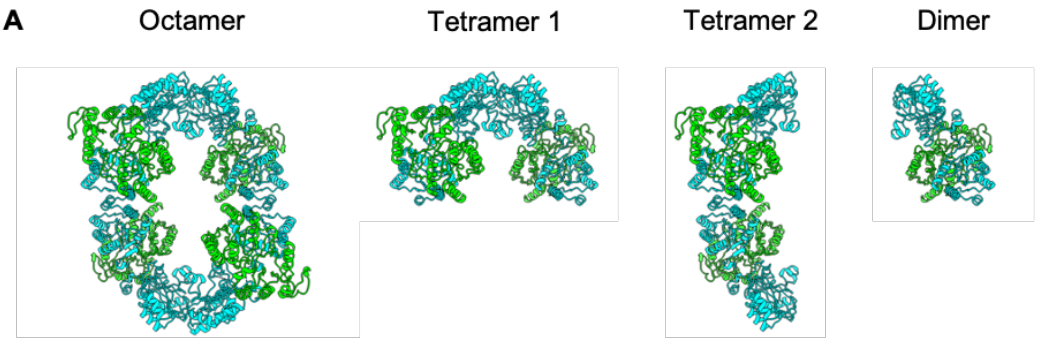
Now that the hetero-octameric structure has been successfully determined using cryo-EM, it can be used alongside the hetero-dimeric model determined by X-ray crystallography with the ligand dependent SAXS data to gain a better insight into the complex assembly mechanism. It will be interesting to understand the order of interactions that must occur during hetero-octamer formation. To test this, the *Oligomer*[89] programme was used to fit different proportions of the hetero-octamer, hetero-dimer and two intermediate hetero-tetramer states that could arise during formation of the dihedral complex to the SAXS data in the as purified, ATPγS and succinyl-CoA configurations of hsA-SCS (Figure 4.25). The two new states are named tetramer 1 (representing a hetero-tetramer formed by the interface 1 interaction in the context of the hetero-octamer) and tetramer 2 (representing the alternative hetero-tetramer formed by interface 2 in the context of the hetero-octamer) (Figure 5-27A). The result of such a fitting is a X^2 value corresponding to the proportions of each state that best describe the theoretical scattering curves. Using the 4-model fit, X^2 dramatically improved across all conditions tested relative to fitting the hetero-dimer alone (Figure 5-27B). Figure 5-27C-E plots the proportions of the 4 models to each ligand state across the concentration series that result in the X^2 reported in Figure 5-27B. In the as purified state, as protein concentration increases, the proportion of the hetero-dimer decreases from 69 % to 34 %, whilst the proportion of the hetero-octamer increases from 28 % to 59 %. Additionally, a small amount of tetramer 1 (representing formation of interface 1) increases with concentration, whereas, no proportion of tetramer 2 (representing formation of

interface 2) can be fitted to the data (Figure 5-27C). This indicates that the concentration dependent assembly of the hetero-octamer must first occur through formation of tetramer 1 and subsequent dimerisation of tetramer 1 at interface 2 to form the hetero-octamer, therefore, bypassing a population of tetramer 2.

In the presence of succinyl-CoA (Figure 5-27D), at the lowest concentration, the dimer population seen in the as purified form has already fully converted to 76% of hetero-octamer and 24 % of tetramer 1, with little variation across the concentration range tested. Again, no proportion of tetramer 2 can be fitted to the data, also indicating that it does not appear as an intermediate species during complex formation (Figure 5-27D).

In the presence of ATPγS (Figure 5-27E), 77% of hetero-dimer is present at the lowest concentration. With increasing concentration, the hetero-dimer population decreases. However, in contrast to the as purified state, no tetramer 1 formation is observed and only negligible amounts of hetero-octamer is observed. Instead, tetramer 2 increases with concentration from 15% to 79% (Figure 5-27E). This would suggest that ATPγS binding prevents hetero-octamer formation through inhibition of the interaction at interface 1. It is unlikely that tetramer 2 represents a physiologically relevant state in the presence of ATPγS, as it is only present at high concentrations.

A



B

	APO			ATPγS			S-CoA		
Protein conc mg/ml	1.5	3	6	1.5	3	6	1.5	3	6
Crystal dimer fit (X^2)	88.8	347.5	1595	31.7	80.7	959.8	1898	1218	4471
4 state fit (X^2)	1.24	1.20	2.86	1.09	1.46	4.89	3.1	2.05	7.94

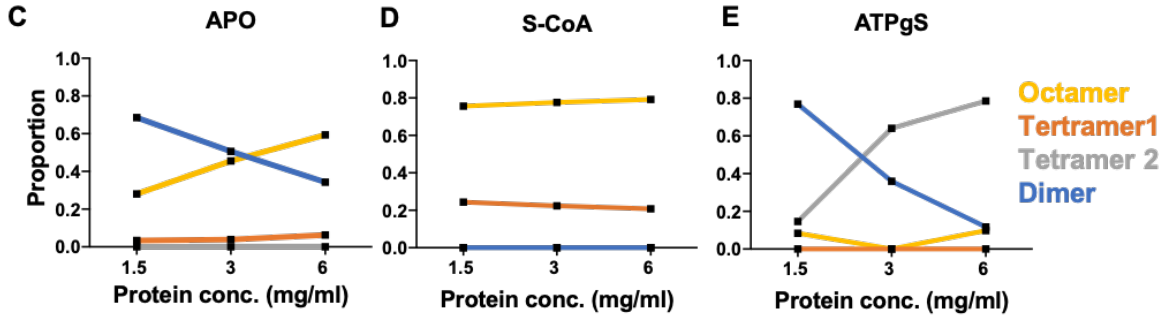


Figure 5-27: Intermediate complex hsA-SCS assembly.

A: Models of the hetero-dimer, hetero-octamer and two potential intermediate hetero-tetramers (hsSUCLG1 in green and hsSUCLA2 in cyan). **B:** Table comparing X^2 of the dimer only fit vs the 4-model fit of the SAXS data in as purified, ATPγS and succinyl-CoA bound SAXS datasets. **C-E:** Proportion trends of each oligomeric state fitted to the as purified (C) succinyl-CoA (D) and ATPγS (E) bound SAXS datasets across the protein concentration series.

To visualise where the nucleotides would bind within the hsA-SCS tetramer 1 state and therefore understand the ATPγS-induced changes in the context of interface 1 inhibition in the hetero-octamer, the GTP bound ssG-SCS hetero-dimer (PDB 2FP4) was duplicated and the two hetero-dimers were overlaid with two of the hsA-SCS hetero-dimers that interact at interface 1 to form the tetramer 1 (Figure 5-28). This reveals that the nucleotide binding site occurs directly at the SUCLA2 ATP grasp

domain interaction at interface 1. At this interaction site, the adenosine groups of nucleotides would be juxtaposed across the interface directly below loop Glu159 to Pro163 (shown to be key for interaction formation; Figure 5-23A and Figure 5-26A) (Figure 5-28A). Although no direct clashing of atoms is observed between both modelled nucleotides, it is possible that the conformational change arising from ATP binding would be a plausible explanation of the disrupted interface.

The resolution of the EM data does not allow accurate modelling at the residue level, however, there are candidate residues that could be hypothesised to take part in both ATP binding and dimerisation at interface one in the absence of ATP. Glu166 is one candidate residue, highly conserved across all isoforms of SCS (Figure 5-26A) and previously shown to be important in nucleotide binding[115]. In the absence of nucleotide, this residue may also be involved in formation of interface 1 in the ATP specific isoform only. The equivalent residue in the GTP specific ssG-SCS structure (Glu114)(PDB 2FP4) has been shown to form a hydrogen bond with the 3' hydroxyl group in the ribose moiety of GTP. In the hsA-SCS hetero-octameric state, Glu166 of one SUCLA2 subunit is also in close proximity (<6 Å) to Arg160 of the opposing SUCLA2 subunit across interaction interface 1 (Figure 5-28B). It could therefore be speculated that in the absence of ATP, Glu166 is free to form a salt bridge with Arg160 during formation of interface 1. In the presence of ATP, this interaction could be broken in favour of ATP binding, leading to disassembly of the hetero-octamer. Arg160 is also only conserved in metazoan ATP specific isoforms of the SCS complex (corresponds to Ala108 of ssG-SCS) (Figure 5-26A, 1st residue in the sequence highlighted green) which could provide further evidence for why hetero-octamer formation is only observed in A-SCS. However, higher resolution data, with

corrected backbone and rotamer placement, would be required to confirm these interactions.

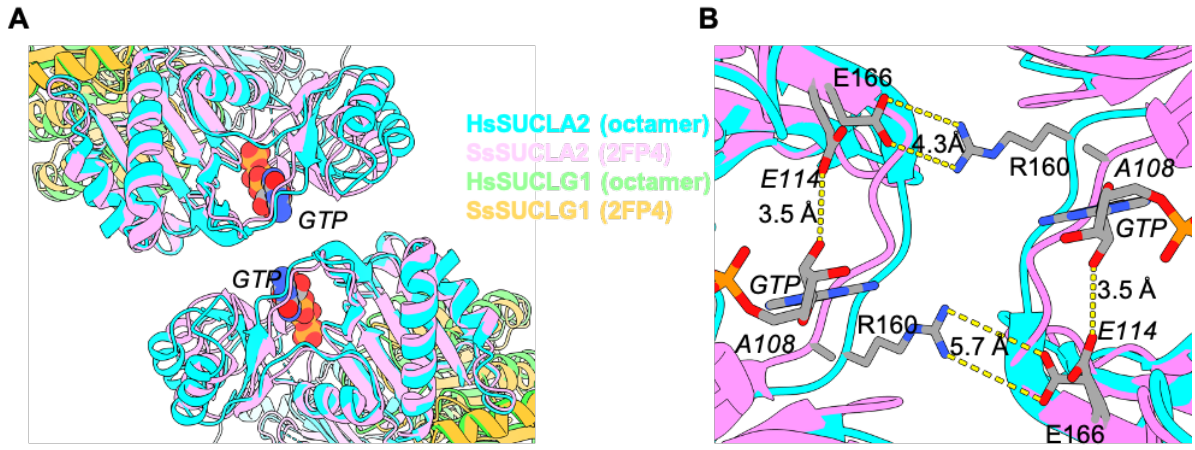


Figure 5-28: Nucleotide inhibition of tetramer one (interface one in the hetero-octamer)

A: Tetramer one interface of hsA-SCS **(**hsSUCLA2 in cyan and hsSUCLG1 in green) overlaid with two copies of the GTP bound ssG-SCS hetero-dimer (2FP4) (ssSUCLG2 in pink and ssSUCLG1 in yellow) (italic labels are of molecules from the G-SCS structure). **B**: Zoom of the tetramer one interface overlay showing residues in hsA-SCS that may be involved in both ATP binding and oligomerisation (italic labels are of molecules/sidechains from the G-SCS structure).

In summary, the novel hetero-octameric state identified in SAXS and EM is stabilised in the presence of succinyl-CoA via, first, interaction of the ATP grasp domain lids at interface 1 and subsequent interaction of the Ct-extension at interface 2. ATP binding shifts the complex in favour of hetero-dimer formation through inhibition of interface 1. In the ATP bound state interface 2 is able to still interact to form a hetero-tetramer; however, this only occurs in very high protein concentrations. The interactions are shown to be only conserved with metazoan SUCLA2 sequences and are not shared by the paralogue SUCLG2 in the GTP forming complex.

5.5 Interaction studies of succinyl-CoA ligase

In addition to the sequence and structure diversion between SCS complexes reported in above sections, the SCS complexes have also diverted in both biological roles and their interactomes. Unlike G-SCS, A-SCS is implicated in the forward reaction participating in the TCA cycle converting succinyl-CoA to succinate and producing ATP, however, it has also been shown to interact with ALAS2 from the erythroid specific haem biosynthesis pathway[16,17]. ALAS2 also requires the substrate of A-SCS, succinyl-CoA to initiate haem biosynthesis. To attempt to link sequence and structure divergence with a biological context, interactions studies with ALAS2 were pursued using the recombinantly isolated proteins *in vitro*.

5.5.1 Interaction studies of A-SCS with ALAS2

To investigate the proposed interaction between ALAS2 and A-SCS in the haem biosynthetic pathway, co-elution studies using size exclusion chromatography were performed with the purified forms of hsALAS2 and hsA-SCS proteins. The insect cell expressed ALAS2 construct incorporating amino acids 79-587 was used as it represented the longest soluble construct. The proteins were mixed at a 1:1 molar ratio in the absence of substrates and in the presence of succinyl-CoA to test for interactions of A-SCS in the different oligomeric states. Protein mixtures were passed over a size exclusion column (superdex 200) with a resolving range of 10-600 kDa. SDS-PAGE of the elution fractions was performed to ascertain whether the two proteins co-eluted in a stable complex. Under the conditions tested, SDS-PAGE analysis showed that the two proteins did not co-elute, indicating no physical interaction has been detected (Figure 5-29A, representative gel of all conditions tested).

It is possible that the A-SCS-ALAS2 interaction is transient under the current conditions tested and therefore cannot be detected by SEC. Additionally, Other protein components or molecules may be required to stabilise the complex, or perhaps the missing region in the ALAS2 N-terminus is also required. To test for the possibility of transient interaction, cross-linking studies between hsA-SCS and hsALAS2 were performed. The crosslinker disuccinimidyl glutarate (DSG) was chosen due to its ability to link lysine residues that are within close proximity (30 Å from Ca to Ca). When purified A-SCS alone was treated with DSG (lane 3), bands resolved at 80 kDa, 130 kDa and 160 kDa, which were absent in the equivalent sample without DSG (Figure 5-29C, lane 2), were visualized. Based on size, these bands correspond to hetero-dimer, and perhaps hetero-trimer and hetero-tetramer species, respectively. It may not be possible to capture the entire hetero-octamer using crosslinking as it requires surface exposed lysines to be within proximity at all interaction interfaces. Therefore, identified crosslinked assemblies only represent potential subcomplexes where the interface is compatible with crosslinking. Similarly, treatment of ALAS2 alone with DSG (lane 5) yielded bands that correspond to the expected cross-linked ALAS2 homodimers (100 kDa), but also potentially homo-trimers (150 kDa) and homo-pentamers (250 kDa) again not present in control sample without DSG (lane 4). The existence of higher order oligomers in ALAS2 enzymes have not previously been observed and should be reconciled by future studies.

When the A-SCS and ALAS2 proteins were mixed at a 1:1 molar ratio (lane 6), there was an upshift of bands with different pattern compared to the proteins incubated alone with cross linker. These bands marked with red asterisk could represent cross-linked complexes between A-SCS and ALAS2, however the exact stoichiometry of

interaction is difficult to deconvolute. The protein identities for these bands were confirmed by tryptic digest MS/MS which mapped peptides corresponding to each of the three component proteins (ALAS2, hsSUCLG1, hsSUCLA2) in the experiment (Figure 5-29C), and therefore would suggest an interaction between SCS-and ALAS2.

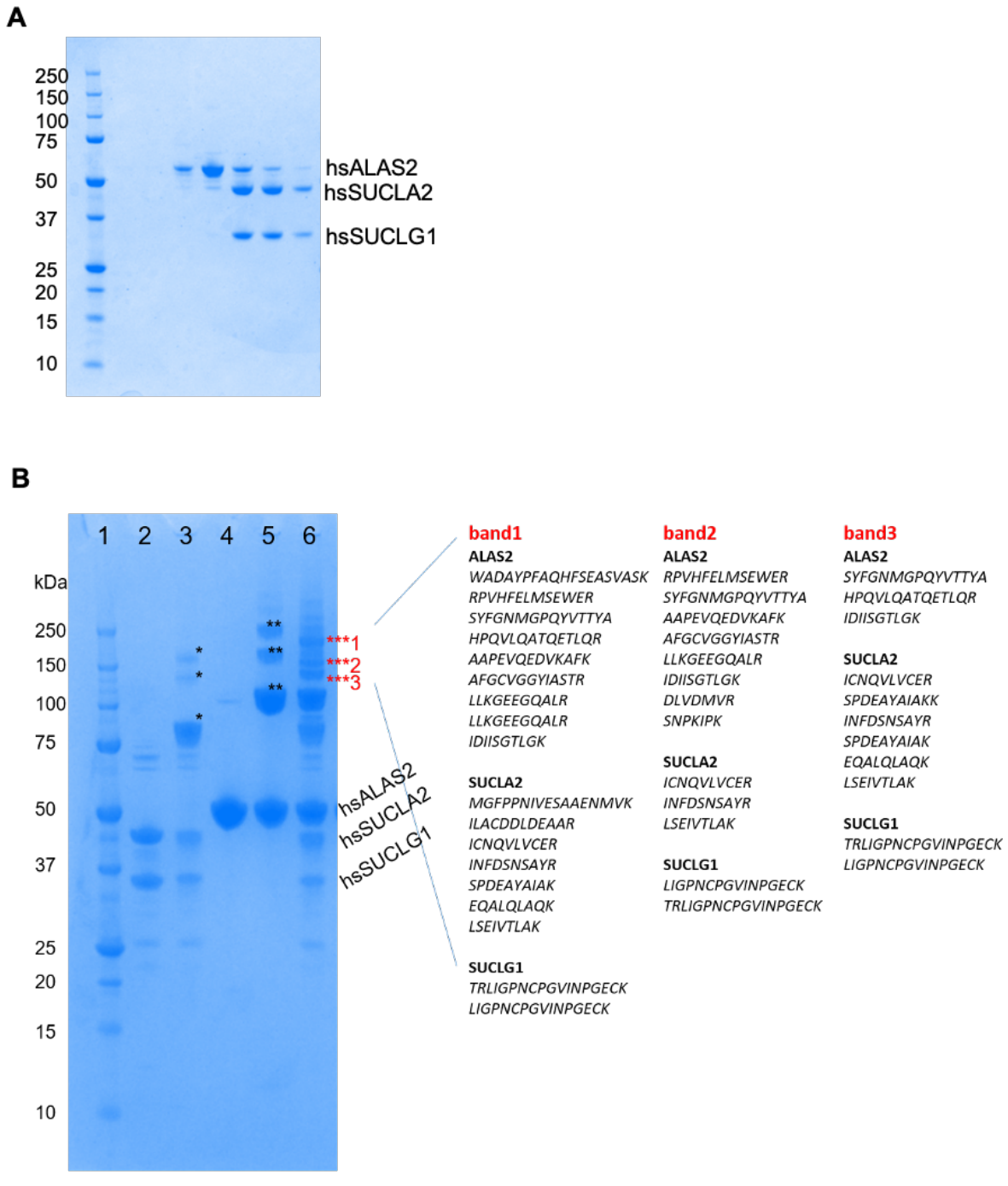


Figure 5-29: investigating wider interactions of hsA-SCS.

A: SDS-PAGE gel of ALAS2 and hsA-SCS SEC interaction assay showing lack of co-elution of the two protein samples. **C:** Crosslinking SDS-PAGE of interactions between hsALAS2 79-578 and succinyl-CoA synthetase. Results shown are representative of three independent experiments. Lane 1, molecular weight markers in kDa; Lane 2, purified succinyl-CoA ligase (SUCLG1-SUCLA2); Lane 3, SUCLG1-SUCLA2 after incubation with di(N-succinimidyl) glutarate crosslinker (DSG), showing crosslinked oligomeric species of SUCLG1-SUCLA2 (*); Lane 4, purified hsALAS2ΔN78; Lane 5, hsALAS2 79-578 incubated with DSG, showing crosslinked oligomeric species of hsALAS2 (**); Lane 6, mixture of SUCLG1-SUCLA2 and hsALAS2 79-578 incubated with DSG, showing crosslinked species representing the complex between SUCLG1-SUCLA2 and hsALAS2 79-578 (***). Protein identity of all asterisked bands (1-3) were confirmed by tryptic digest MS/MS, in which mapped peptides corresponding to each of the three component proteins are shown in the inset figure.

5.6 A-SCS results chapter summary

It expected that A-SCS and its related paralogous complex G-SCS participate in opposing cellular roles[52], through their ability to catalyse opposite directions of a reversible reaction[52]. These cellular roles are strongly supported by the differing tissue specific expression levels of the two complexes[49]. The evolution of differential nucleotide specificity was previously believed to be the only factor that controlled directionality of the SCS reaction, and hence its partitioning into different cellular roles. It was therefore assumed that the hetero-dimeric structural features revealed from studies of the ssG-SCS complex would be applicable to the A-SCS complex, and subsequently the overall reaction mechanisms would also be equivalent. Here, isolation of the hetero-octameric state of the A-SCS complex in the presence of succinyl-CoA shows significant divergence in the assembly of A-SCS compared with G-SCS. This divergence is attributed to the evolution of sequences for homo-oligomerisation in the SUCLA2 subunit at the Ct-extension and within the ATP grasp lid domain linkers, that are not present in the equivalent β subunit of G-SCS.

Extensive SAXS study revealing substrate- and product-induced oligomeric transitions lead to the expectation that in order for the reaction to proceed in the forward direction, A-SCS must first bind succinyl-CoA and adopt the hetero-octameric conformation. Based on mass spectrometry experiments that measured half reactions of A-SCS, succinyl-CoA binding is expected to induce reaction steps 1 (Figure 5-3) leading to the release of CoA. After CoA release, the enzyme would be bound to succinyl-phosphate and poised for nucleotide binding in the hetero-octameric state. It is not clear yet whether binding of succinyl-CoA or completion of reaction step 1 is the cause of hetero-octamerisation. After binding of ADP, the complex would transition to the hetero-dimeric form allowing for substrate level phosphorylation of ADP to produce and release the product ATP (Figure 5-30A).

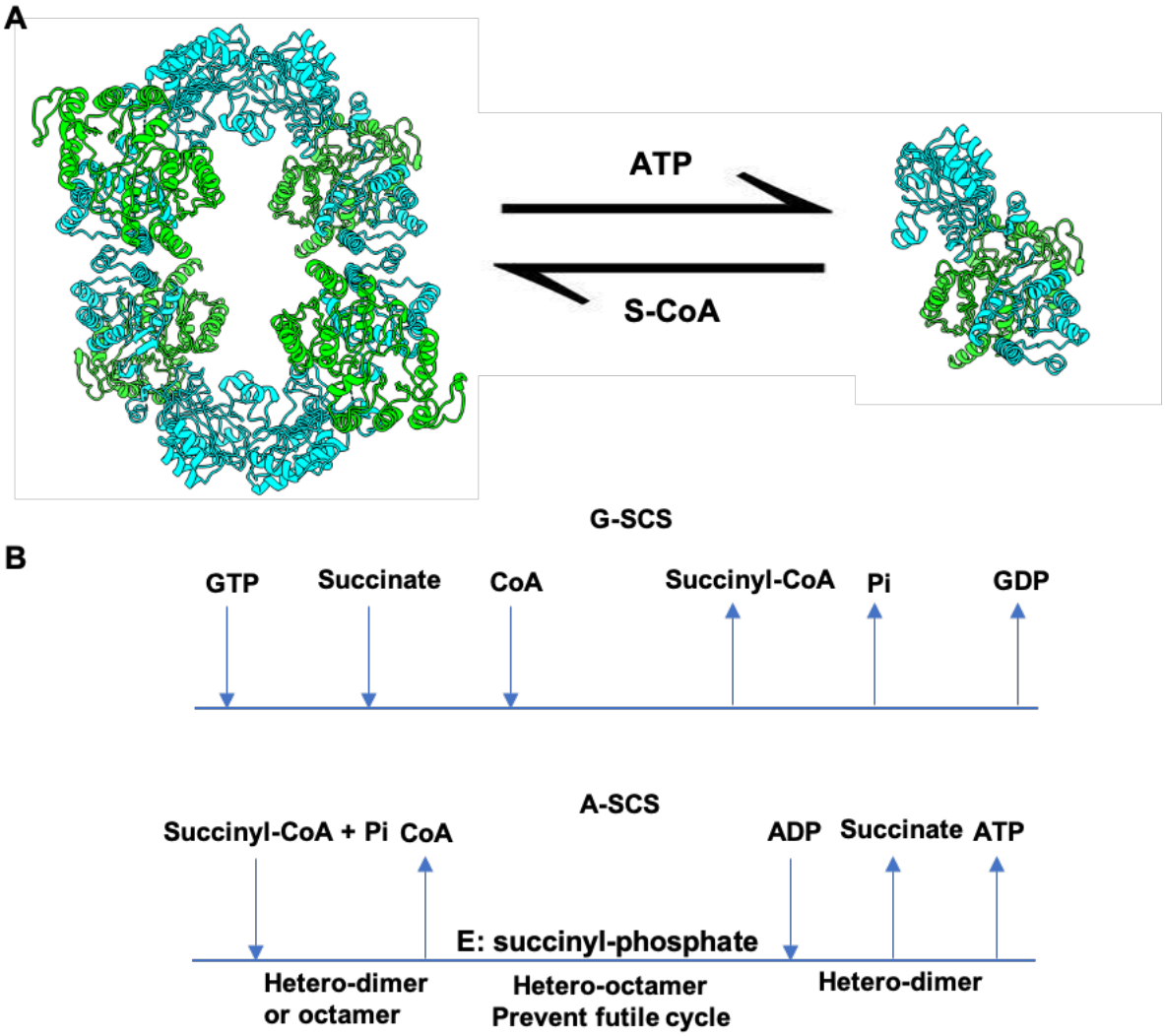


Figure 5-30: Hypothesised model for A-SCS reaction.

A: Substrate induced transitions in A-SCS oligomeric state. **B:** Proposed reaction mechanism for G-SCS based from previous study and A-SCS based on structural data presented here.

Previously, there have been many proposed models for the reaction mechanism of SCS. The most recent and widely accepted model for G-SCS is an ordered *ter ter* reaction mechanism[129] (Figure 5-30B), where all three substrates in the reverse reaction (succinate, CoA and GTP) bind before product release. This mechanism however cannot account for the complex intermediate transitions in oligomeric states described above in the A-SCS complex. Instead, a *ping pong* mechanism may be a more appropriate description of the reaction mechanism (Figure 5-30B). In this case,

in the hsA-SCS catalysed forward reaction, succinyl-CoA binding and CoA product release occur to produce the hetero-octameric conformation followed by ADP binding and ATP product release in the dimeric form.

The existence of multiple oligomeric states that alter the proposed reaction mechanism relative to the dimeric G-SCS complex may enforce the forward directionality to the reaction. Keeping the enzyme poised in the hetero-octameric state when bound to succinyl-phosphate could prevent futile cycling between the enzyme reaction steps 1 and 2 (Figure 5-3). From a structure perspective this would align well with the observation that the active site histidine loop becomes disordered in the hetero-octameric state preventing transfer of the phosphate group in the absence of nucleotide binding. This model would rely on the speculation that binding and cleavage of succinyl-CoA occurs in the hetero-dimer state and that transition from the hetero-dimer to a hetero-octamer occurs only after cleavage of succinyl-CoA in reaction step 1 (Figure 5-3) in the succinyl-phosphate bound state (currently untested here). Only when nucleotide is present would the transition to the hetero-dimer occur again allowing for reaction step 2 and 3 to take place. In this hypothesis, when succinyl-CoA is abundant in the milieu, the hetero-octamer could drive the reaction towards the direct production of ATP and increased oxidative phosphorylation through enhanced TCA cycling. When succinyl-CoA levels are low and ATP is abundant, the enzyme may adopt the dimeric form that, like G-SCS, is compatible with the reverse reaction (Figure 5-30) The G-SCS reverse reaction may not require such exquisite regulation as the high ratio of GTP to GDP in the mitochondria will serve to prevent futile cycling.

6 Structural studies of Propionate catabolism complexes

6.1 Introduction to MMUT-MMAA complex results

6.1.1 Current understanding of the MMUT-MMAA complex

MMUT and MMAA participates in the final step of propionate catabolism producing succinyl-CoA for entry into the TCA cycle. The MMUT reaction requires the supply of active AdoCbl cofactor in order for catalytic turnover. To overcome cofactor inactivation, a highly complex MMUT reactivation pathway has evolved. This pathway relies on the gating action of the p-loop GTPase co-chaperone MMAA[13,15,73]. Through its GTPase activity, MMAA can both assist loading of MMUT with active Cbl cofactor and unloading of the inactivated forms of Cbl. This process also requires a 3rd component, MMAB, to convert inactive Cbl to the active AdoCbl form ready for reloading to MMUT.

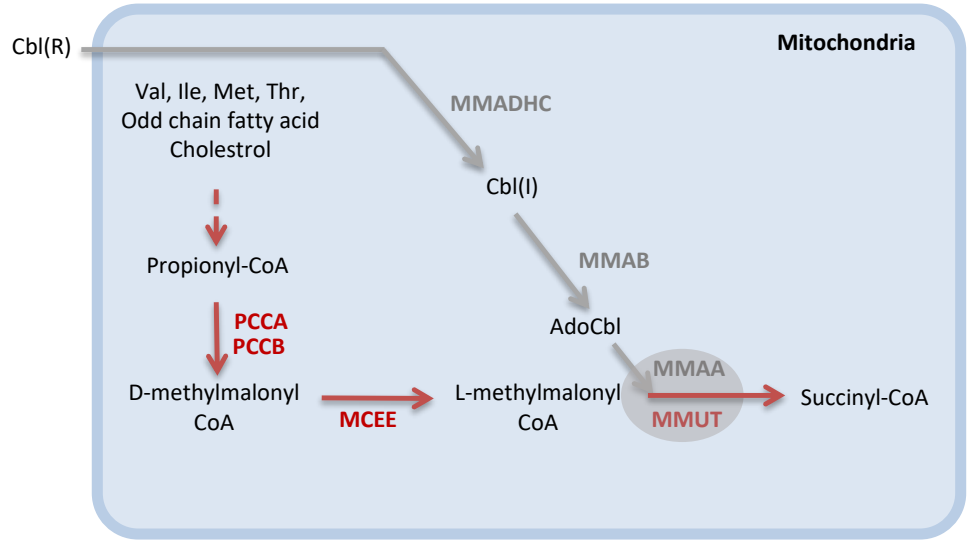


Figure 6-1 Propionate catabolism pathway.

MMAA-MMUT non obligate complex is highlighted grey in the third step of the propionate catabolism pathway.

Different orthologues and paralogues have evolved quite varied mechanisms to tackle this reactivation pathway. Studies of the direct MMUT othologue in *M. extorquens* were the first to show an interaction *in vitro* with the separately encoded GTPase, MeaB, using native-PAGE[13] and later by isothermal titration calorimetry[130]. The structures MeaB and hsMMAA have both been solved[14,131] revealing a homodimeric arrangement of both proteins (Figure 6-2, left and middle) with a single GTPase domain per subunit in the homodimer. Overlay of one MMAA and one MeaB subunit (Figure 6-2, right) and sequence analysis of the GTPase components (Cα rmsd 1 Å, sequence similarity 45 %) show both proteins are highly conserved. However, significant divergence of the homo-dimeric interface has evolved in hsMMAA that results in the juxtaposition of the two GTP binding sites (Figure 6-2). This divergence may alter the GTPase interaction mechanism with human forms of MMUT.

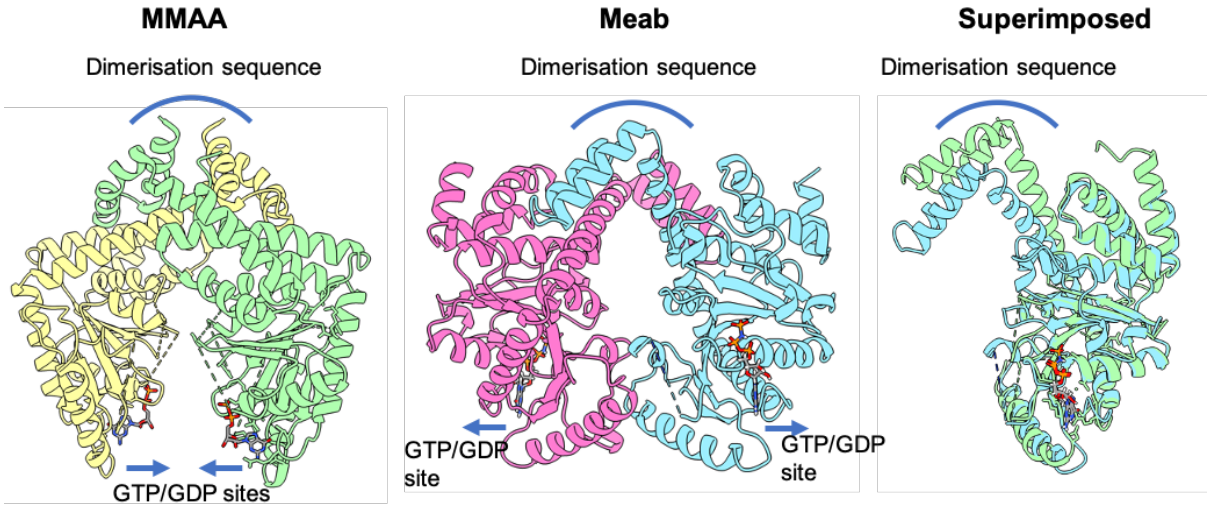


Figure 6-2: Structures of MMAA and MeaB GTPases.

Left: MMAA homodimer (PDB 2WWW). GDP shown in stick, individual subunits coloured green and yellow and arrows indicate juxtaposition of nucleotides. **Middle:** MeaB homodimer (PDB 4JYB). GTP

shown in stick, individual subunits coloured pink and cyan and arrows indicate inverted nucleotide binding sites relative to MMAA. **Right:** Overlay of one subunit of MMAA (colour green) with one subunit of MeaB (colour cyan) shows structural divergence at the dimerisation interface that causes subunit inversion.

The structure of hsMMUT has been solved[14] and also reveals a homodimeric arrangement, with a Cbl cofactor bound active site per subunit. Direct interactions of hsMMUT and hsMMAA have been reported using size exclusion chromatography[14]. These studies reveal complex formation has ligand dependency requiring apo MMUT (no Cbl present) and GTP bound MMAA (using a GTP non hydrolysable analogue) to interact in a wide range of potential higher order oligomeric complexes.

The crystal structure of a related Cbl dependent radical enzyme, isobutyryl-CoA mutase (Icm) (PDB 4XC6 (MMUT paralogue) in bacteria, reveals its p-loop GTPase is encoded as a fusion protein with Icm (IcmF)[132,133]. This fusion protein forms a homodimer in solution with a single GTPase domain per mutase active site (Figure 6-3A). The sequence and structural similarity between hsMMUT, hsMMAA and the fusion structure of IcmF allows speculation of the interaction interface of the human enzymes by super imposing the 3 structures (Figure 6-3B). The GTPase domains overlay with an Cα rmsd 1 Å and the MMUT domains with an Cα rmsd 1.1 Å. Due to the distance between the two MMUT active sites within one homodimer, two GTPase homodimers are required to interact with one MMUT homodimer to ensure both MMUT active sites can engage a GTPase. This results in a heterohexamer with stoichiometry of $2MMAA_{dimer}$:$1MMUT_{dimer}$, however, in this arrangement, two GTPase domains are left unoccupied (Figure 6-3B). Interaction of additional MMUT homodimers with these unsaturated binding sites may indicate a mechanism for the formation of higher order oligomers[14].

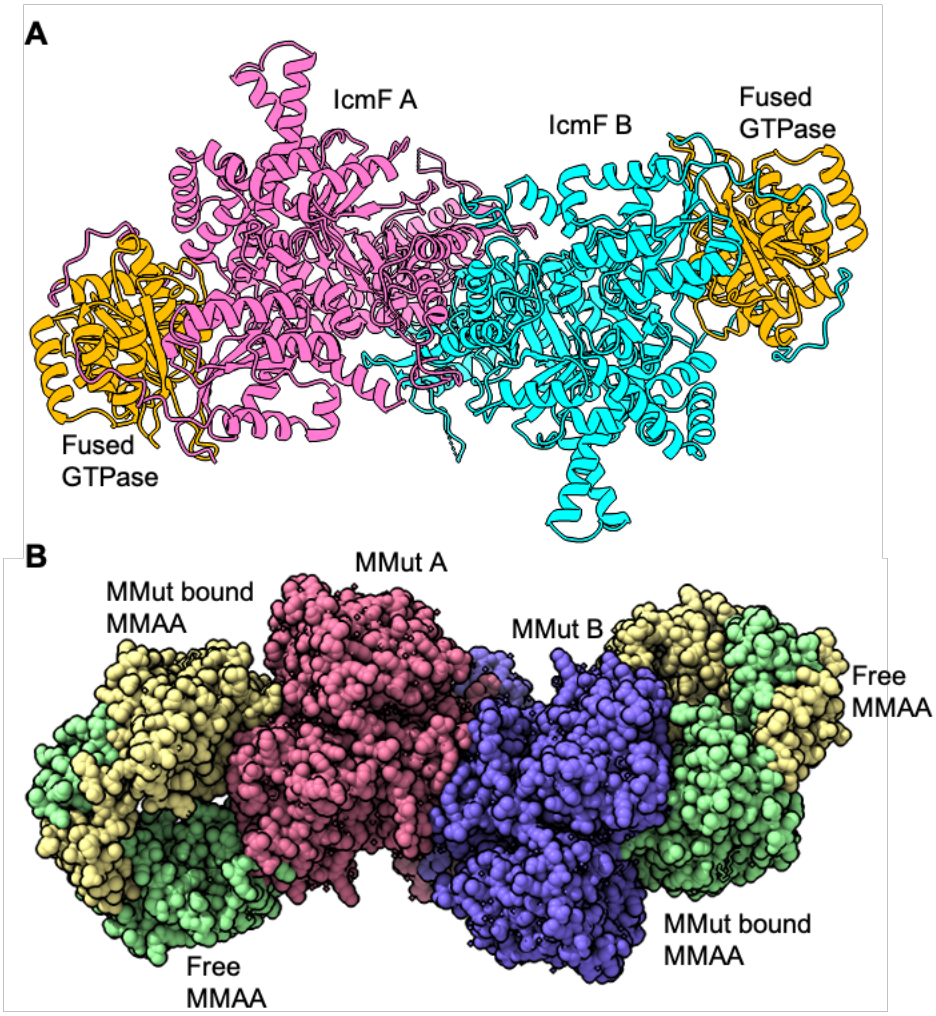


Figure 6-3 Modelling Mutase MMAA interaction.

A: Homodimeric structure of IcmF containing two isobutyryl-CoA mutase domains (pink and cyan) each fused to a single GTPase domain (orange). **B:** Model for the interaction of 2 hsMMAA homodimers (yellow and green) and one hsMMUT homodimer (red and purple) generated through alignment of one hsMMUT homodimer to the isobutyryl-CoA mutase domain of IcmF and a single hsMMAA homodimer per fused IcmF GTPase domain giving a final stoichiometry for the modelled heterohexamerix complex $2MMAA_{dimer}$:$1MMUT_{dimer}$.

Negative stain electron microscopy experiments of the human MMAA-MMUT complex have started to identify and delineate the role of potential functional heterooligomers in the presence of different substrates and cofactor conditions[15]. In

the current model, GTP bound MMAA binds inactive MMUT, subsequent GTP hydrolysis offloads inactive cofactor (Cbl(II)) for cofactor reactivation by MMAB. MMAA in the GDP state remains bound to MMUT in a hetero-hexameric assembly with a proposed stoichiometry of 1 MMAA homodimer to 2 MMUT homodimers. In this state, the complex is primed to receive the active form of the cofactor AdoCbl from MMAB. However, in this arrangement two MMUT sites would remain unoccupied in contrast to the related IcmF structure. Once regenerated to the active form, MMUT dissociates from its chaperone to catalyse multiple cycles of the terminal reaction in propionate catabolism producing succinyl-CoA for the TCA cycle[15].

In addition to the above potential oligomeric assemblies, the existence of hetero-octamers and hetero-dodecamers that form ring like structures with 1:1 stoichiometry of hsMMUT and hsMMAA have been reported in the same study[15]. These higher order oligomers are speculated to act as a sequestering mechanism where in the absence of avialble Cbl cofactor, MMUT homodimers could be held in a temporarily inactive state[15]. To date, no high-resolution information is available to substantiate or provide further mechanistic detail on this potentially exquisite metallocofactor trafficking and activation mechanism. In this chapter, different structural biology approaches are employed to find potential ways to confirm the existence of the different MMAA-MMUT oligomers and to attempt to analyse their exact composition with single particle cryo-EM.

6.1.2 Propionate catabolism results chapter synopsis: MMUT-MMAA

In this chapter, the interaction of hsMMAA with hsMMUT is investigated using SAXS experiments. Attempts are made to optimise identified complexes for high resolution structure solution using cryo-EM.

6.2 Structural studies of the MMUT-MMAA complex

6.2.1 Cloning and expression of MMAA-MMUT

Previous work has shown that formation of a MMAA-MMUT complex is highly dependent upon nucleotide and Cbl bound states[14,15]. As a result, co-expression was presumed not to be a plausible method for complex production. Instead, mature forms of hsMMAA and hsMMUT, lacking their N-terminal mitochondrial targeting sequences (hsMMAA aa 76-418 and hsMMUT aa 12-750), were expressed individually, purified and subsequently reconstituted into the two-way complex. MMAA and MMUT constructs were already previously cloned, purified and crystallised in the lab[14], ready for scale-up. MMUT contained a vector-encoded tobacco etch virus protease cleavable Ct-His_6-tag (pNIC-CTHF) whereas MMAA contained a vector-encoded tobacco etch virus protease cleavable Nt-His_6-tag (pNIC28-Bsa4) enabling them to be purified using the same methods as described here[14].

6.2.2 SAXS analysis of MMAA-MMUT complex size and shape

As a starting point, SEC-SAXS was used to investigate the average size and shape of the individual MMUT and MMAA proteins. Both MMAA and MMUT eluted from the size exclusion column as single peaks (Figure 6-4 top and bottom).

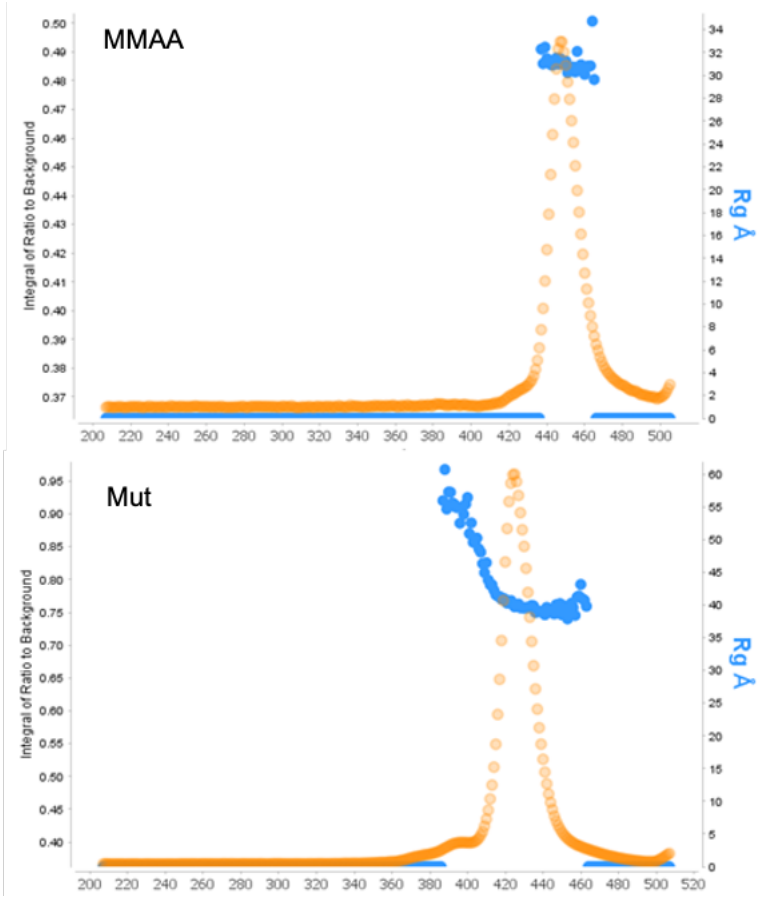


Figure 6-4: SEC-SAXS elution profiles

The elution profiles for MMAA alone (top), MMUT alone (bottom). Each point (yellow) represents the integrated area of the ratio, of the SAXS curve representing each sample frame (X-axis, SAXS measurement taken every 3 seconds of the SEC elution time course), to the estimated background (left Y-axis). Displayed in blue is the auto estimated Rg (right, Y-axis) from the Guinier region for each sample frame (X-axis).

Frames with similar estimated Rg values were buffer-subtracted and averaged. The MW calculation, using the Vc/Rg^2 method[100] as an invariant proportional to molecular mass, is predicted to be 150 kDa for MMUT and 76 kDa for MMAA. The theoretical monomeric MW for MMUT and MMAA is 83 and 38 kDa respectively, indicating that in both cases the homo-dimeric form of the individual proteins have been isolated, consistent with published structures (PDB 2WWW, 3BIC). Back calculation, using *Oligomer*[89], of the SAXS curve for the MMUT dimer crystal structure (PDB 3BIC) and

fitting to the SAXS data gives a χ^2 value of 36 for MMUT, indicating some deviation between the solution structure and the crystal structure (Figure 6-5A). Solution modelling using 20 rounds of *dammif*[89] with C2 symmetry enforced and averaging and filtering with the *Damaver*[89] program provided a final low-resolution envelope of the dimeric state. Overlay of the crystal structure with the SAXS model also shows a fairly poor fit (Figure 6-5B). The poor fitting could be due to the presence of non-specific interparticle interactions resulting from high protein concentrations affecting the average conformation in solution[123]. Later modelling at lower protein concentrations show improved fits for MMUT (see Figure 6-9). Back calculation of the SAXS curve for the MMAA dimer crystal structure (2WWW) and fitting to the SAXS data gives a χ^2 value of 13, indicating a reasonable correlation to the solution structure (Figure 6-5C). Solution modelling using 20 rounds of *dammif*[89] with C2 symmetry enforced and averaging and filtering with *Damaver*[89] provided a final low-resolution envelope of the dimeric state. Overlay of the crystal structure with the SAXS model shows a very good fit (Figure 6-5D).

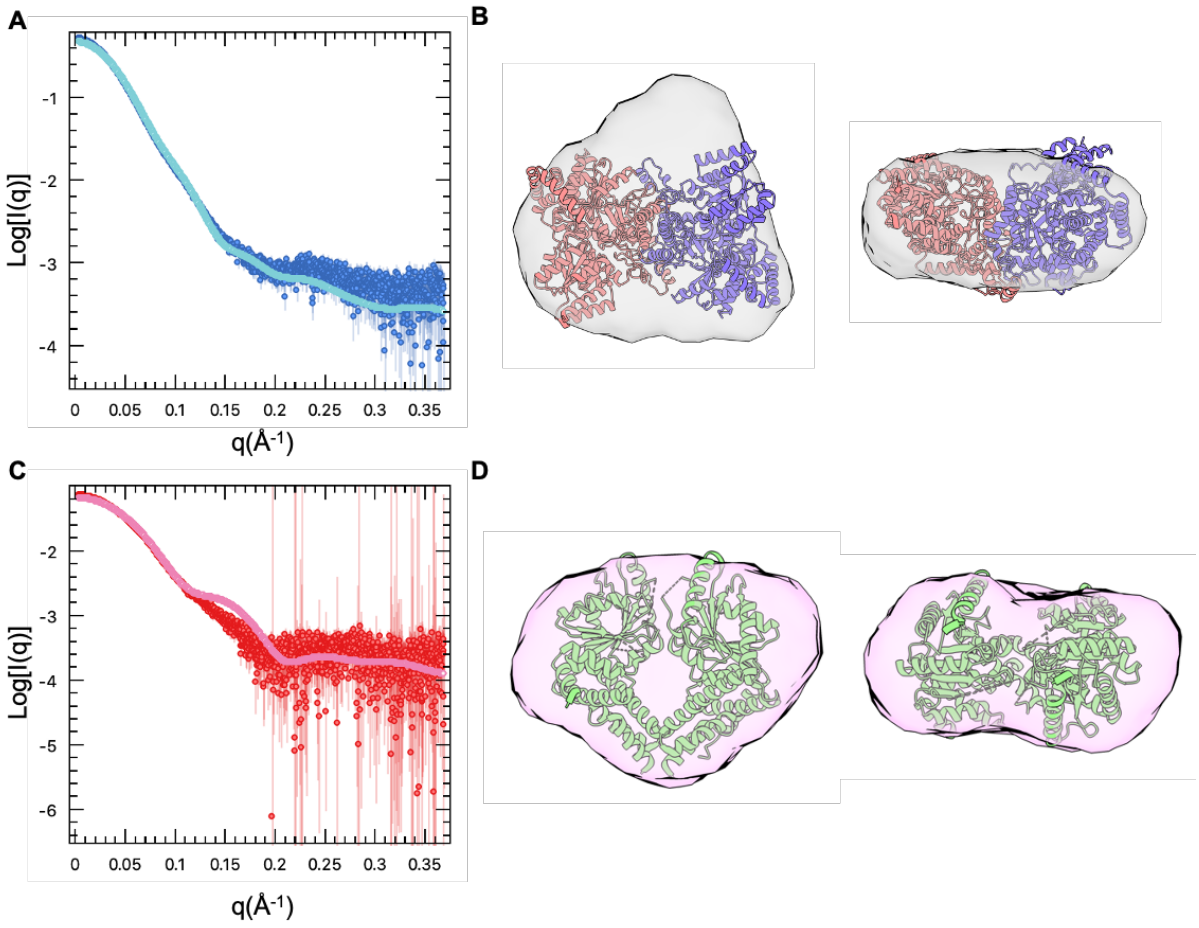


Figure 6-5: SAXS modelling of MMUT and MMAA homodimers.

A: Log10 intensity plot showing the MMUT experimental scattering data (blue dots) overlaid with the theoretical scattering curve (light blue line) of the MMUT crystal structure dimer (PDB 3BIC). **B:** *Ab initio* modelling of the MMUT SAXS envelope overlaid with the dimeric model of the MMUT crystal structure. **C:** Log10 intensity plot showing the MMAA experimental scattering data (red dots) overlaid with the theoretical scattering curve (pink line) of the MMAA crystal structure dimer (PDB 2WWW). **D:** *ab initio* modelling of the MMAA SAXS envelope overlaid with the dimeric model of the MMAA crystal structure.

To study the binary MMUT-MMAA complex, the two proteins were mixed at 1:1 molar ratio in the presence of the GTP analogue (GTPPCP) and the SEC-SAXS dataset was recorded. Analysis of the SEC chromatogram (Figure 6-6, bottom) shows multiple peaks shifted to a lower elution volume when compared to the individual runs of MMUT and MMAA alone (Figure 6-6, top and middle), indicating

successful formation of a two-way complex. These shifts are in line with previously published gel filtration results for the complex[14] showing MMUT-MMAA co-elute via SDS-PAGE (see section 6.2.3 for a similar SEC experiment and SDS-PAGE analysis of complex formation). The SEC complex elution spans the entirety of the elution volume with several peak spikes (Figure 6-6, bottom). Toward the end of the elution volume individual peaks begin to be resolved. The presence of multiple distinct peaks, albeit not well resolved throughout elution, is an initial indication of the existence of differently sized oligomeric states as opposed to random aggregations that would more likely be manifested as a continuum. The last peak (1) elutes in the same elution volume as the MMUT only sample (Figure 6-6, middle). This may indicate excess MMUT that either did not participate in the higher order interactions or had since dissociated from a complex species. No elution peak is seen for excess MMAA in the complex sample, indicating all soluble MMAA is converted into complex.

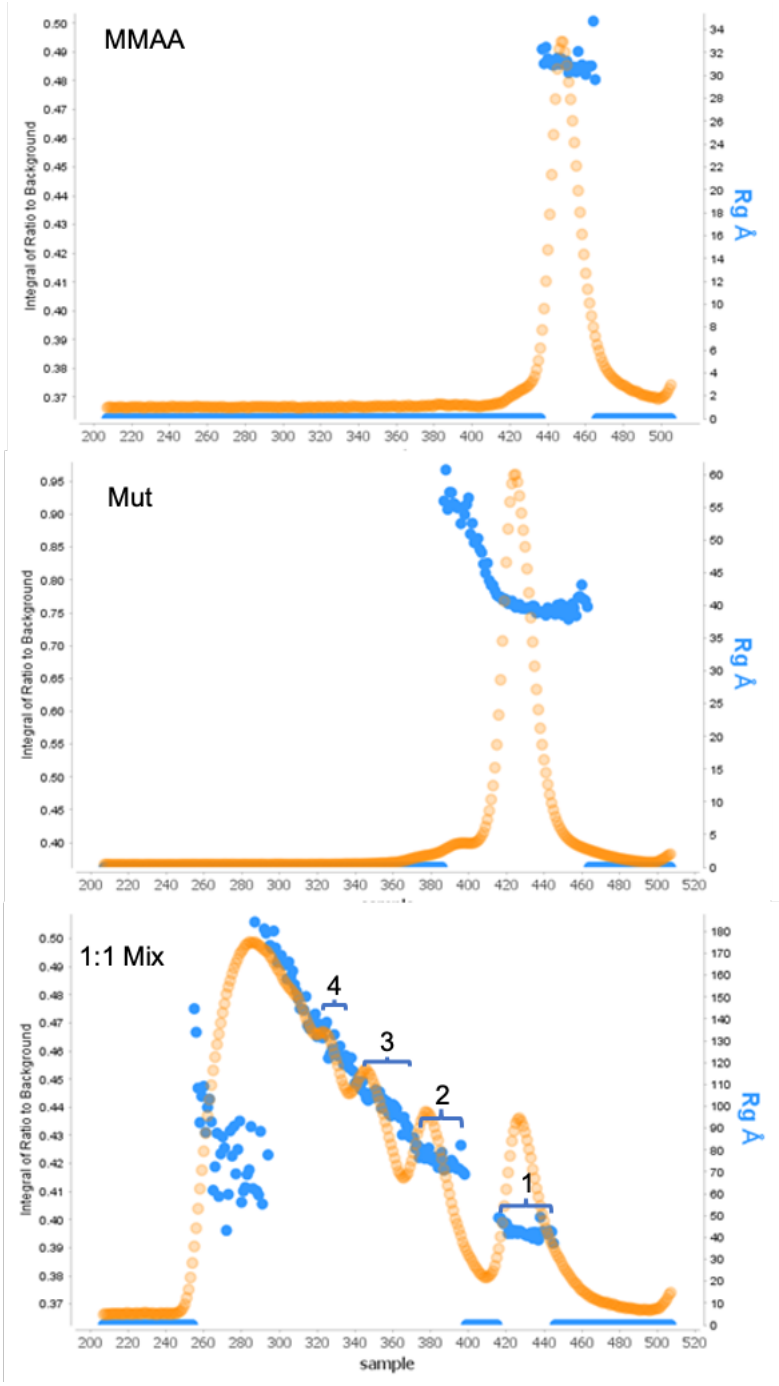


Figure 6-6: SEC-SAXS elution profiles.

The elution profiles for MMAA alone (top), MMUT alone (middle) and the 1:1 molar mix of the two (bottom). Each point (yellow) represents the integrated area of the ratio, of the SAXS curve representing each sample frame (X-axis SAXS measurement taken every 3 seconds of the SEC elution time course), to the estimated background (left Y-axis). Displayed in blue is the auto estimated Rg (right, Y-axis) from the Guinier region for each sample frame (X-axis). Brackets 1-4 represent

elution peaks of varied resolution that are buffer subtracted and averaged for further SAXS characterisation.

The lack of fully resolvable peaks poses a problem for exact characterisation of each possible state in the MMUT-MMAA complex, as a single frame is the average of all possible states in the eluent at a particular measurement timepoint. Nevertheless, attempts to analyse the scattering data were pursued further. As a starting point, 3 frame rolling average was calculated and buffer scattering was subtracted for every datapoint under the elution. The datasets were used to estimate average MW changes across the various complex peaks. The calculated MWs range from 150 kDa at the largest elution volume to 1-2 MDa as elution volume decreases (Figure 6-7). As peaks are not fully resolved, individual MW estimates are less reliable and are contributed by an average of multiple sized particles. However, a plateau over adjacent elution fractions is indicative of a major single-size population, either contributed by one specific state or an average of multiple stable states at that particular concentration. Each inflection point between plateaus might indicate transitions to higher order oligomers. From this analysis, 3 plateaus are clearly distinguished and labelled 1, 2 and 3 in inverse order of elution. In addition, a 4th plateau (labelled 4) is somewhat discernible before individual species are no longer resolved in the elution (Figure 6-7). The locations of the plateaus align nicely with the gel filtration peaks in (Figure 6-6, bottom).

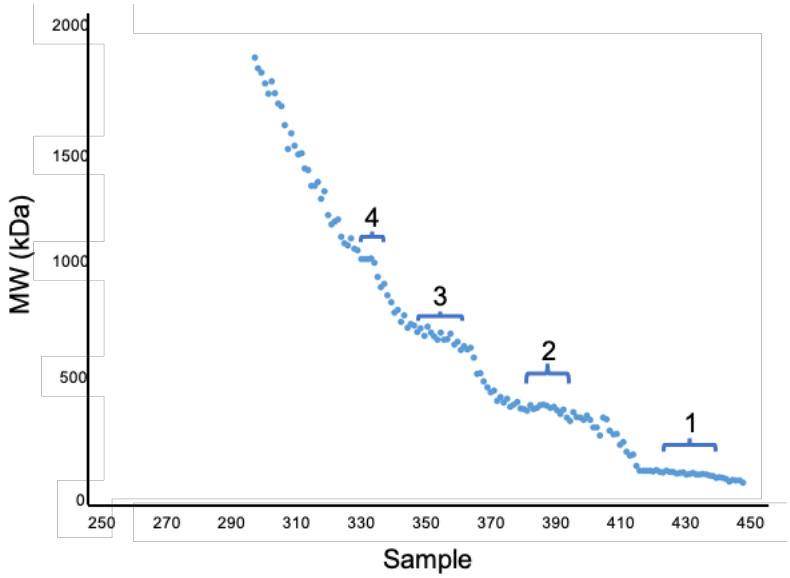


Figure 6-7: MW calculation for the SEC-SAXS isolated MMUT-MMAA complex.

MW (y-axis) against SAXS sample frame (X-axis, taken every 3 seconds across the SEC elution time course) MW plateaus are labelled 1 to 4 and correspond to labelled peaks in Figure 6-6 bottom.

Frames under each plateau were next averaged and buffer subtracted for more accurate MW calculations. For plateau 1 the MW estimation of 150 kDa is in agreement with the size of MMUT-only homodimer (Figure 6-8). This alongside similar Rg and Dmax estimates (Figure 6-8) confirms peak 1 to be excess MMUT.

Protein sample	Mut	MMAA	Complex peak 1	Complex peak 2	Complex peak 3	Complex peak 4
Guinier						
I(0)	0.51	0.07	0.06	0.13	0.24	0.38
Rg (Å)	40.52	31.02	41.34	78.28	107.77	137.91
P(r)						
I(0)	0.50	0.07	0.06	0.14	0.24	0.38
Rg (Å)	40.27	30.92	42.47	81.57	109.72	140.99
Dmax	157	119	160	320	443	550
Particle size						
MW (kDa)	150	76	150	440	760	>1000
Model fitting						
Crystal structure fit (X^2)	36 (3BIC)	13 (2www)	5 (3BIC)			

Figure 6-8: Table of SAXS parameters.

comparing complex peaks 1-4 with individual MMUT and MMAA data. Values for radios of gyration (Rg), zero scattering angle (I(0)), and normalised zero scattering angle (I(0)) for protein concentration (I(0)/protein conc) are calculated by Guinier fitting and real space distribution plots are. Dmax values calculated by real space distribution plots. MW values calculated by the Vc/Rg^2 method[100]. Crystal fit using the MMAA structure (2WWW) or MMUT structure (3BIC) to the experimental data is calculated using *Oligomer*.[89]

Interestingly, in comparison with the MMUT individual SAXS experiment (Figure 6-5A), fitting the SAXS scattering curve to the back calculated curve for the MMUT crystal structure (PDB 3BIC), significantly improved the X^2 value from 40 (previous individual MMUT SAXS experiment shown in Figure 6-5A) to 5 (Figure 6-9A). This suggests the non-complexed MMUT protein in this sample adopts a conformation that better reflects the conformation in the crystal structure. Furthermore, *ab initio* model calculation produced a model of the SAXS data that overlays significantly better to the crystallographic model (Figure 6-9B) than calculated previously (Figure 6-5B). Since the concentration of MMUT within peak 1 of the complex sample is significantly lower (shown by lower a much lower intensity of scatter across the curve Y-axis and lower I(0)), this improved fit from the MMUT individual sample (Figure 6-5A) may be due to less non-specific interparticle interactions affecting the overall scattering curve.

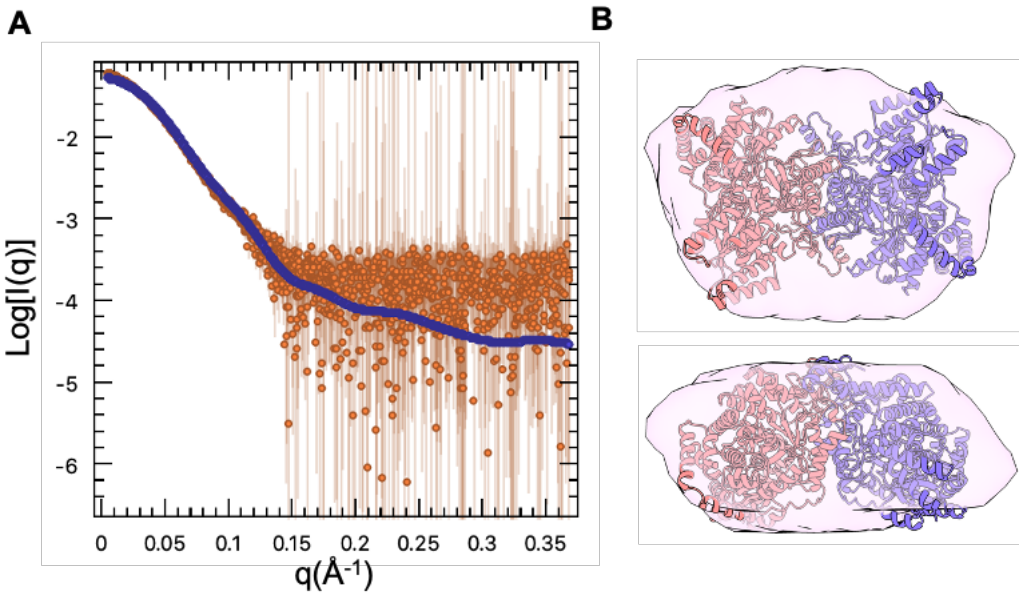


Figure 6-9: SAXS modelling of peak 1 data with MMUT homodimer.

A: Log10 intensity plot showing the peak 1 experimental scattering data (brown dots) overlaid with the theoretical scattering curve (dark blue line) of the MMUT crystal structure dimer (PDB 3BIC). **B:** *Ab initio* modelling of the peak 1 SAXS envelope overlaid with the dimeric model of the MMUT crystal structure.

For Plateau 2, MW calculations suggest an average particle size of 440 kDa, which could correspond to an oligomer of 2 MMAA dimers and 2 MMUT dimers (theoretical MW of 480 kDa) (Figure 6-8)). The third plateau, is estimated to be 760 kDa, approximately corresponding to 3 MMAA dimers and 3 MMUT dimers (theoretical MW of 720 kDa) (Figure 6-8)). Both arrangements have previously been reported from negative stain electron microscopy in the form of discrete ring-like structures[15]. However, results here show particles continue to significantly increase in total size, indicating the existence of many higher order oligomeric states. It is difficult to rationalise how ring structures could continue to grow in size to accommodate higher-order oligomers such as those shown in plateau 4 and the further unresolved peaks (Figure 6-7, Figure 6-6, bottom).

Attempts to calculate an *ab initio* model of the oligomeric states in peaks 2-4 are prohibited as this procedure requires isolation of single discrete conformations that have not been resolved here. However, some basic qualitative shape descriptions can be obtained from analysis of the p(r) distance distribution vectors calculated from each peak dataset. P(r) distribution function is a plot of the real space distribution (Y-axis) of inter particle distances (vectors) (X-axis) and reveal the probability of finding a point within the particle at a distance r (Å) from another given point. Figure 6-10A shows theoretical p(r) distribution plots that correspond to different discrete particle shapes such as discs, rods and spheres[123].

An overlay of the P(r) distributions derived from all SAXS datasets analysed in this section, cross-referenced with theoretical p(r) plots, shows complex formation in peaks 2 (Figure 6-10B, dark red), 3 (Figure 6-10B, red) and 4 (Figure 6-10B, light red) deviates from the bell-shaped distribution of vectors corresponding to a globular particle, as seen in the MMAA (Figure 6-10B, purple) and MMUT (Figure 6-10B, green) individual datasets and the MMUT population in peak 1 (Figure 6-10B, maroon).

The complex datasets in peaks 2 (Figure 6-10B, dark red), 3 (Figure 6-10B, red) and 4 (Figure 6-10B, light red) transition to a shape that becomes increasingly elongated. This is noted in the graph by an initial large proportion of small inter particle vectors (with a P(r) peak maximum of 75 Å for the peak 2 dataset) and a gradual gradient decrease to a large Dmax (Dmax is the longest possible interparticle distance or largest X value when Y = 0) (280 Å in the peak 2 dataset). This distribution is similar to the green ‘rod’ like particle plotted in Figure 6-10A. As the particles increase in size from peak 2 to peak 4, the P(r) distribution peak maxima remains the same at 75 Å, however, the gradient of decrease becomes more gradual resulting in an ever

increasing Dmax (longest particle vector) (Figure 6-10B). This reflects the fact that the particle size is increasing in one dimension only (i.e. increasing in length but not width for example). This one-dimensional elongation effect is characteristically observed in filament formation. Among the complex peaks 2-4 the average Dmax appears to regularly increase by a discrete unit of 140 Å with every SAXS peak, perhaps indicating the increment of a repeating unit as the filament grows. For peaks 3 and 4, a second maxima in the P(r) distribution at approximately 150 Å begins to appear alongside the elongation, which may be reflective of the helical repeat as the filament grows longer.

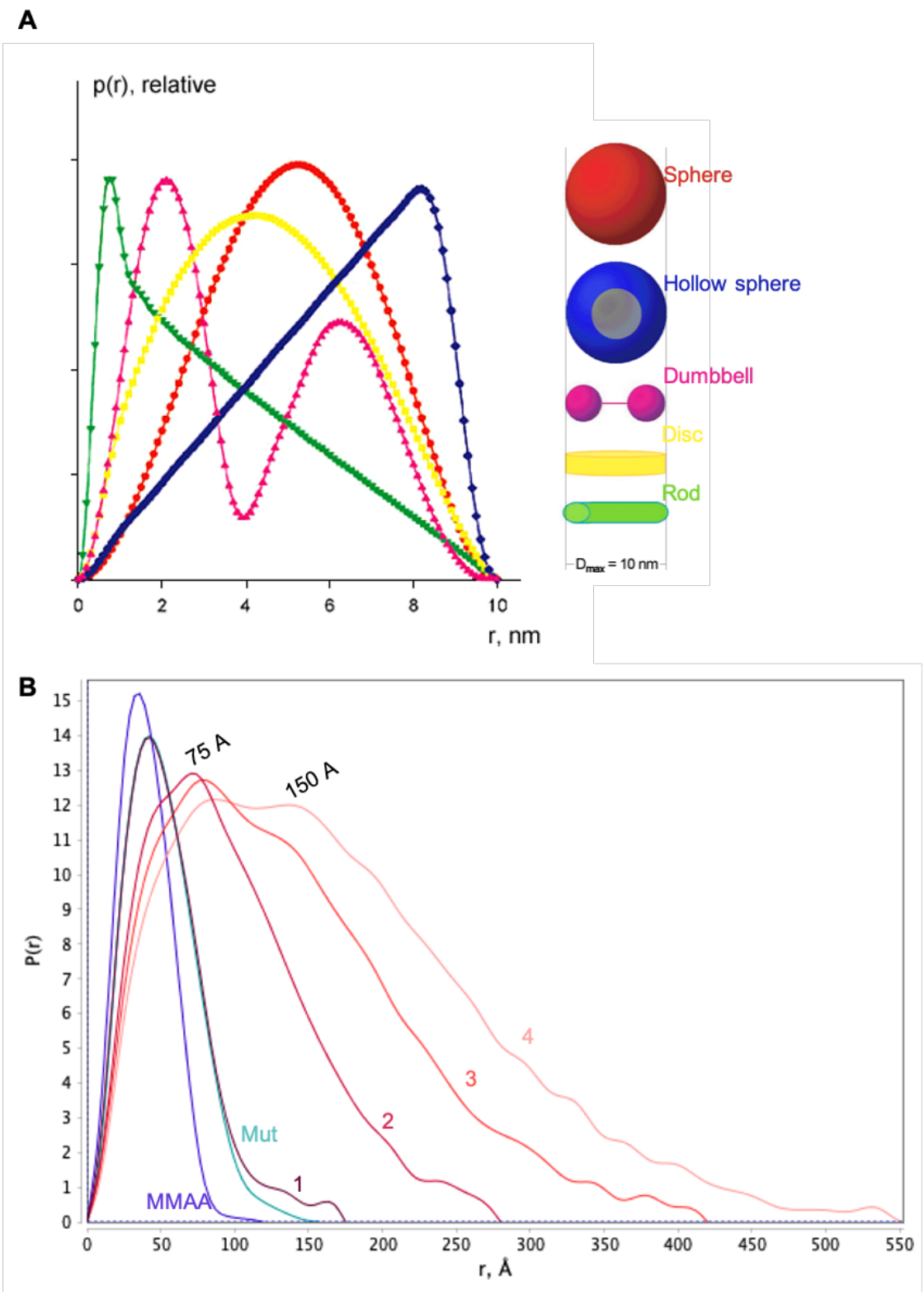


Figure 6-10: SAXS P(r) distance distribution.

P(r) distribution function is a plot of the real space distribution (Y-axis) of inter particle distances (vectors) (X-axis) and reveal the probability of finding a point within the particle at a distance r (Å) from another given point. **A:** Theoretical distance distributions for different shape categories adapted

from[123]. **B:** P(r) distance distribution comparing complex datasets labelled 1 to 4 (corresponding to peaks labelled in Figure 6-6 and Figure 6-7) with individual MMUT and MMAA data.

With the above P(r) analysis suggesting complex formation between MMUT and MMAA results in production of elongated particles, it would be interesting to attempt to directly visualise these individual particles using cryo-EM to assess whether they can be described as filamentous complexes.

6.2.3 Cryo-Electron microscopy studies of the MMAA and MMUT complex.

Cryo-EM was used in an attempt to directly visualise MMAA-MMUT complexes, interpretated to be filamentous forms of MMUT-MMAA of varying length through SAXS data analysis. The MMUT-MMAA complex was reconstituted over size exclusion chromatography using the same conditions as the SAXS studies, but this time with MMUT and MMAA mixed at 1:2 molar concentration to account for the initial excess of MMUT seen in SAXS (Figure 6-11A). The size exclusion results were in line with the previous SEC-SAXS results, giving a large broad complex peak representing the different oligomeric states with co-elution of MMAA and MMUT as shown by SDS-PAGE (Figure 6-11A). In this instance MMAA was in excess as opposed to MMUT, likely due to the different initial molar ratio chosen for this reconstitution experiment.

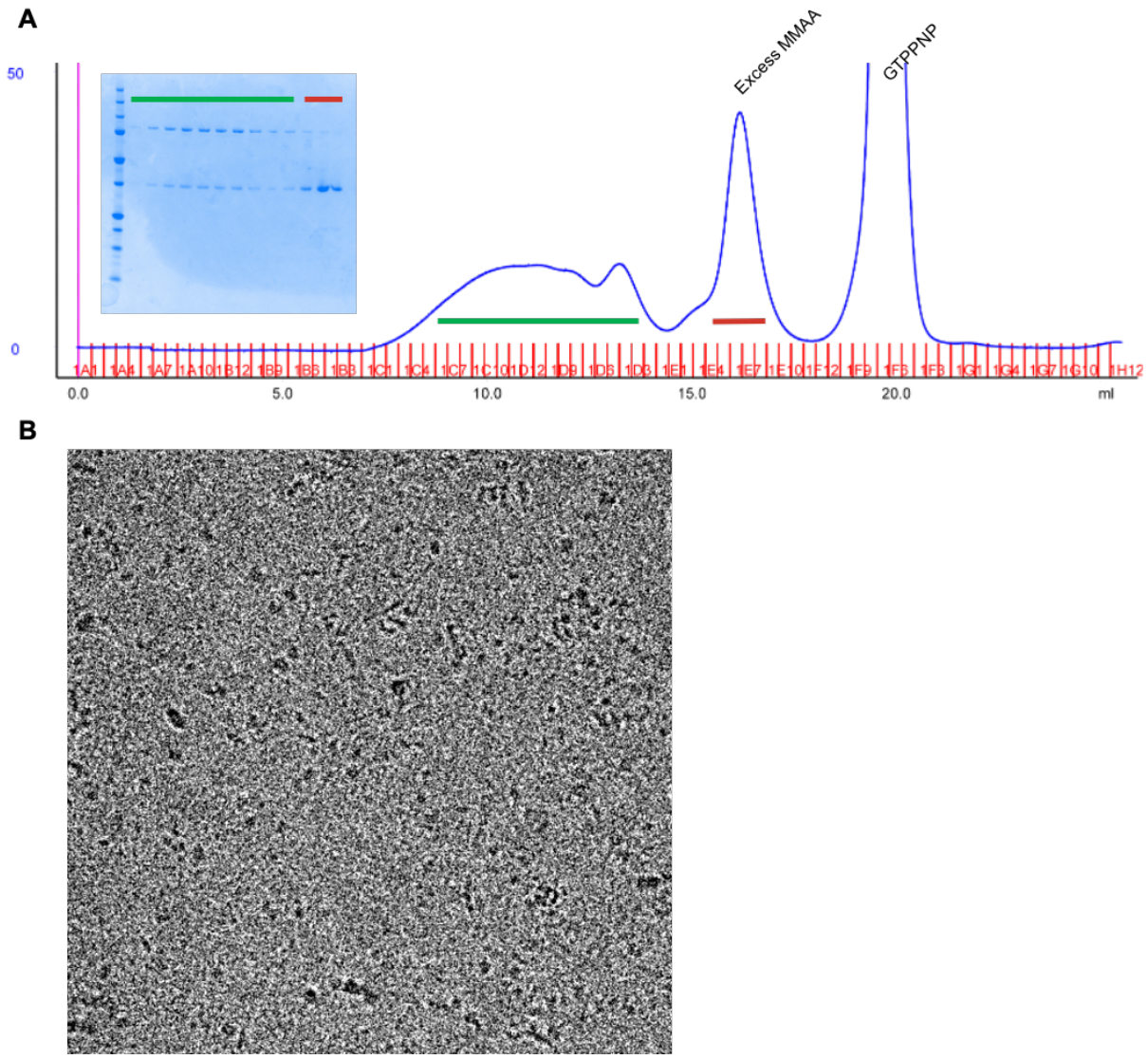


Figure 6-11: Cryo-EM optimisation of MMAA-MMUT complex.

A: Size exclusion gel filtration chromatogram and SDS-PAGE analysis of elutions. Coloured bar in SDS-PAGE corresponds to the elution peaks with the equivalent coloured bar on the chromatogram.

B: Initial micrograph screen of MMAA-MMUT complex isolated from gel filtration.

Samples from the elution peaks were vitrified directly onto cryo-electron microscopy grids. Grids were screened using Glacios 200 keV microscope at magnification of 150000x and directly imaged with a Falcon III direct electron detector, at the Oxford Particle Imaging Centre. Initial images showed particle over-crowding and low signal to noise making direct visualisation of single particles difficult (Figure 6-11B). To combat this, a gradient fixation approach (GraFix)[134] was next used in an attempt to

better isolate the different oligomeric states and obtain stable solution samples that can be better visualised on a micrograph. This technique uses ultracentrifugation to purify macromolecules by size through a glycerol density gradient, whilst simultaneously allowing for weak intramolecular chemical cross-linking using the lysine crosslinker glutaraldehyde within the density gradient. The effect of this is stabilisation and separation of heterogeneously sized macromolecules. The gradient is first analysed on SDS-PAGE and isolated fractions are subsequently vitrified onto grids and screened using cryo-EM.

SDS-PAGE analysis of the MMUT-MMAA complex after GraFix (Figure 6-12A) reveals a similar profile to SEC-SAXS size exclusion (Figure 6-6, bottom). In this case, as the density of the gradient increases (from left to right of the gel) an increase in MW of the bands is observed due to chemical crosslinking preventing disassociation under denaturing conditions of SDS-PAGE. Just like the SEC-SAXS experiments, the lower molecular weight species are better resolved as individual bands and likely correspond to the individual peaks in the isolated in SEC-SAXS elution (seen in Figure 6-6, bottom). The higher MW species lose the discrete band separation and merge into a smear. Tryptic digest followed by MS/MS and MW weight of approximately 160 kDa (based on the gel marker Figure 6-12A) confirms that the dominant band (*) from an early gradient fraction (labelled fraction 1 on the gel) contains a dimer species of excess MMUT only. Bands (** and ***) from two later gradient fractions, labelled fraction 2 and fraction 3 on the gel (Figure 6-12A), were confirmed by tryptic digest MS/MS as containing both MMUT and MMAA peptides, indicating they are isolated in complex with each other. These fractions containing complex (2 and 3) were next vitrified on EM grids and screened with the same Glacios microscope settings previously described. Initial interpretation of the

grids reveals individual particles with some filamentous properties observed in the sample corresponding to fraction 2 on the GraFix SDS-PAGE gel (Figure 6-12A) however, the predominant species are individual low MW macromolecules and some short filaments (Figure 6-12B, left). Fraction 3 of the GraFix SDS-PAGE gel (Figure 6-12A), containing higher molecular weight species, shows a clear transition to a higher abundance of short filaments (Figure 6-12B, right). Filament formation predicted in SAXS experiments was directly visualised using cryo-EM at the micrograph level. In the higher MW fractions (Figure 6-12B, right) filaments appear monodisperse with high signal to noise. This sample is ready for follow up data collection to attempt to gain high resolution detail of the filament helical repeat.

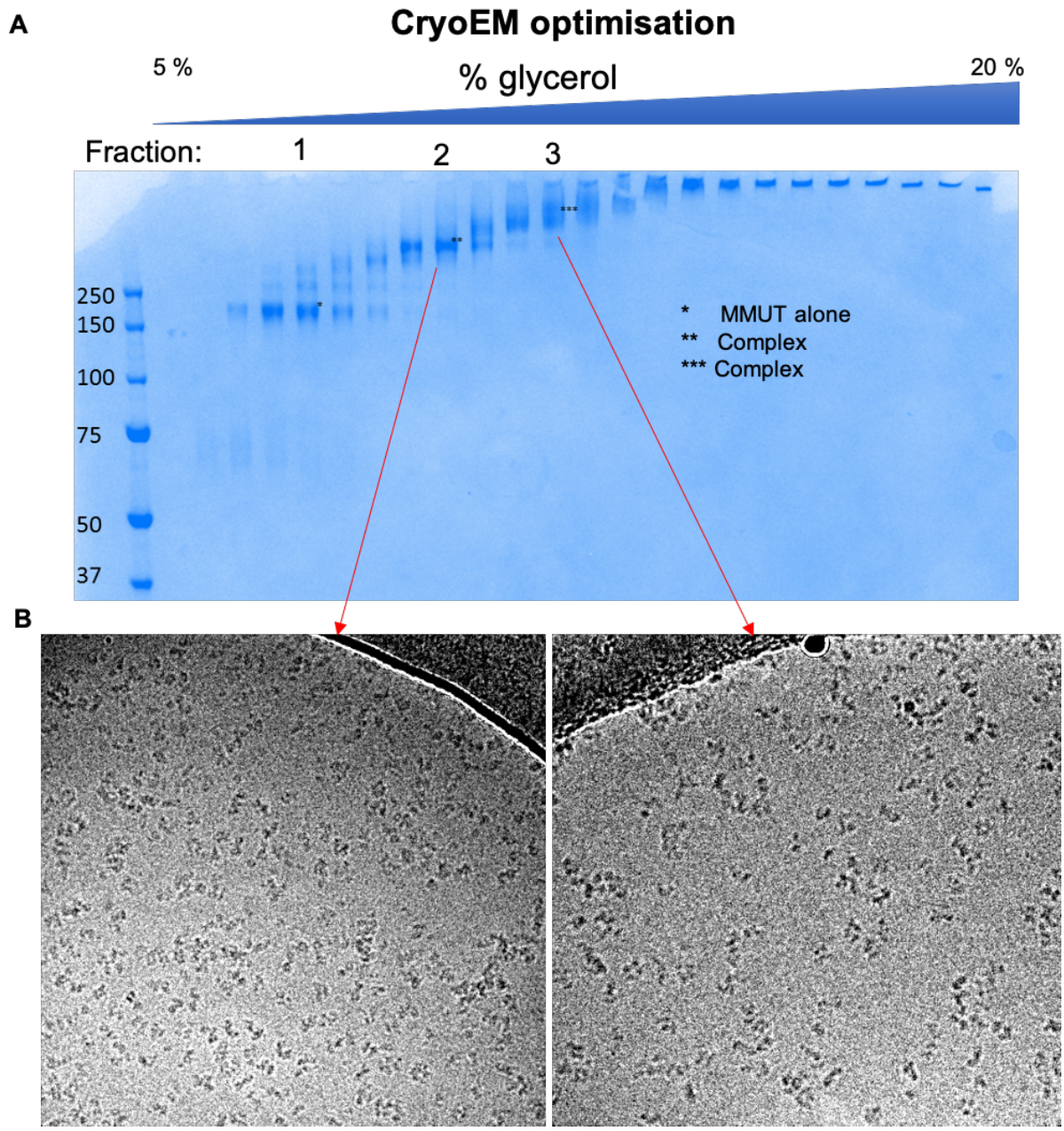


Figure 6-12: GraFix optimisation of MMAA-MMUT cryo-EM sample.

A: SDS-PAGE showing separation of different MMAA-MMUT oligomers using GraFix. Bands labelled *, ** and *** were cut out tryptic digest MS/MS analysis and confirmed to contain MMUT alone (*) or the MMAA-MMUT complex (** and ***). Change mutase to MMUT, Blue triangle shows glycerol gradient increase across the gel. **B:** Cryo-EM micrographs of different MMUT-MMAA oligomers separated by GraFix. Left micrograph corresponds to fraction two on the gel. Right micrograph corresponds to fraction 3 on the gel.

6.2.4 Modelling Filament formation

In the absence of high-resolution information, it is interesting to speculate how a filament consisting of repeating units of MMUT-MMAA might grow. Modelling the of the direct MMUT-MMAA interaction interface using the paralogue fusion enzyme IcmF has been shown earlier (Figure 6-3). Due to the homodimeric nature of MMAA, the initial $2MMAA_{dimer}$:$1MMUT_{dimer}$ stoichiometry left two potential interaction sites on MMAA un-occupied (Figure 6-3B). Given the evidence for filament formation here, attempts to grow this model into a higher order oligomer were made. To do this, the $2MMAA_{dimer}$:$1MMUT_{dimer}$ model was duplicated and one MMAA homodimer from one model was aligned to the opposing MMAA homodimer of the second $2MMAA_{dimer}$:$1MMUT_{dimer}$. By overlapping opposing MMAA dimers from duplicated models, a single MMUT and MMAA homodimer was added to the end of the complex creating an oligomeric stoichiometry of $3MMAA_{dimer}$:$2MMUT_{dimer.}$ In this arrangement, 1 MMAA homodimer is fully saturated with interactions to two MMUT homodimers. The remaining two MMAA homodimers have unsaturated interaction sites. Continuing this process by adding MMAA-MMUT homodimer repeats results in the formation of a filamentous structure containing 3 MMAA dimers and 3 MMUT dimers per helical repeat. In the model, MMAA homodimers sit at the repeated apices of the helical filament with MMUT homodimers forming the vertices that bridge two MMAA apices (Figure 6-13).

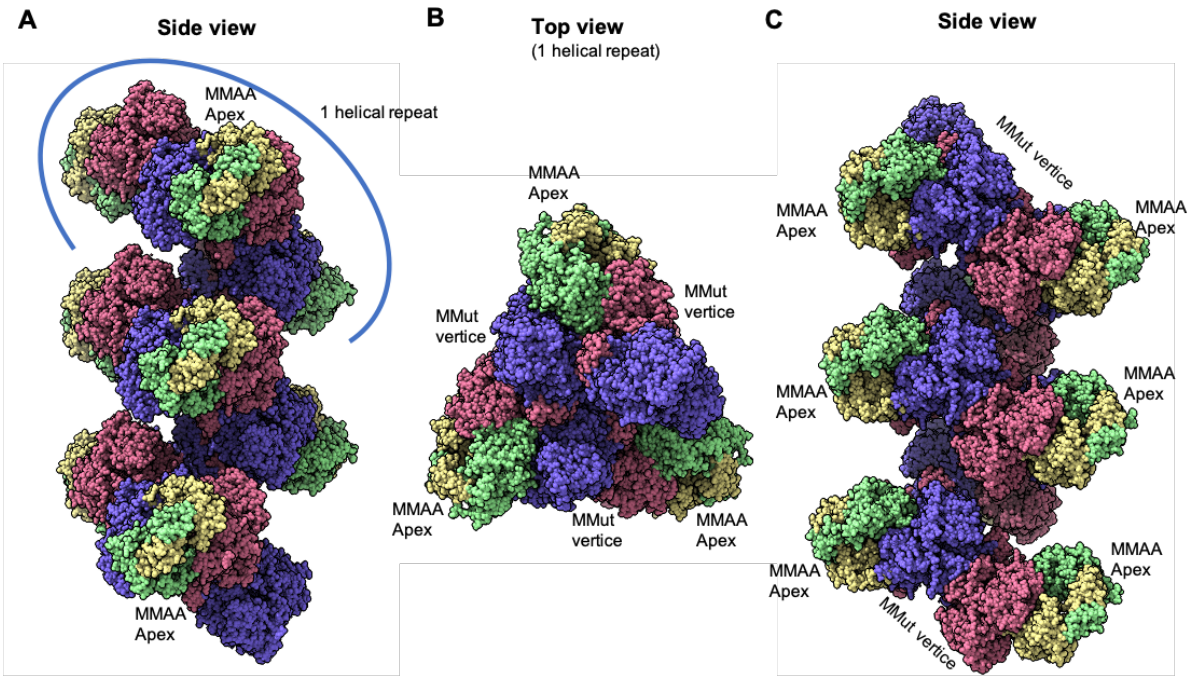


Figure 6-13: Model of the MMAA-MMUT filament based on extension of the IcmF structure. **A:** Shows MMAA homodimers (green and yellow) at apex of the filament. **B:** Provides a top view showing 3 MMAA homodimers and 3 MMUT homodimers per helical turn. **C:** shows the MMUT homodimers (red and purple) along the vertices of the filament.

8 filaments of different lengths were built, based on as few as 1 homodimer of each protein to as large as 8 homodimers of each (each with a sequential addition of 1 MMAA homodimer and 1 MMUT homodimer) and termed here 1:1, 2:2, etc to 8:8. To determine whether a filament model can describe the SAXS data, the *Oligomer* program[89] was used to fit proportions of these filaments to the scattering curve of plateau 2 resulting in a very good X^2 fit of 1.88 using 60% 1:1, 20% 2:2, 1% 6:6 and 9% 8:8 (Figure 6-14). There are of course many more intermediate states that could be modelled and used in the fitting. However, the results already show that the filamentous shape category is correctly attributed to the MMUT-MMAA SAXS data, and that the modelled structure itself is a very good fit to the experimental data.

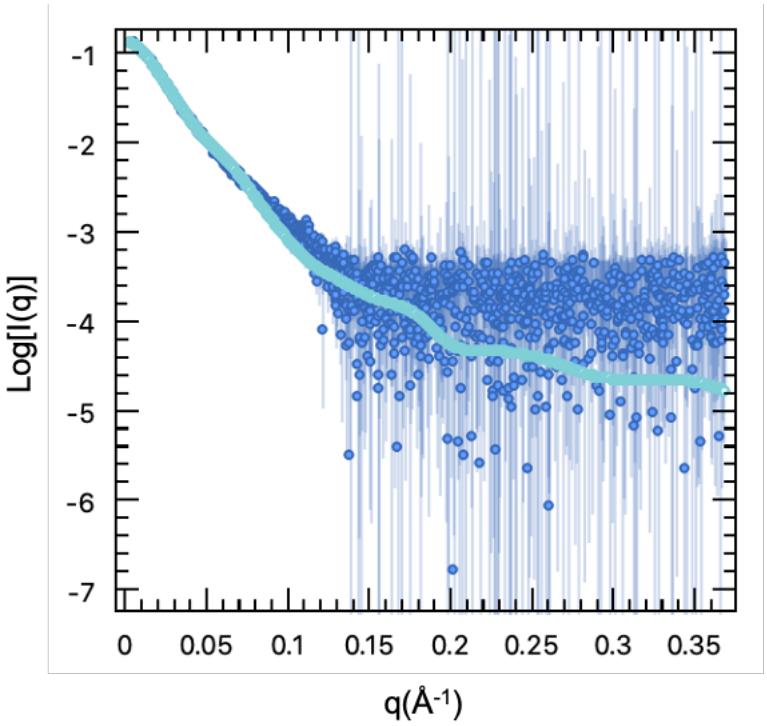


Figure 6-14: Filament model fit to SAXS scattering data.

Overlay of the theoretical filament scattering using proportions of different filaments lengths to the experimental scattering data to give X^2 1.88.

The current model creates some clashing between two MMUT homodimers and one MMAA subunit with the MMUT homodimer within a single helical repeat. This is not unexpected as the MMAA structure used in filament formation is the GDP bound state (PDB 2www). Previous studies have shown that only the GTP bound state is capable of heterogenous higher order oligomerisation, whereas, the GDP form of MMAA is only capable of homogenous hetero-hexamer formation only[15]. Therefore, GTP specific conformational change of MMAA may be required to alleviate steric clashing of the current model. No clashing is observed between two helical repeats.

The structure of MMAA bound to GDP suggests a possible hinge point at the homodimerision interface. Evolution of an inverted homodimer interaction relative to the orthologue MeaB that juxtaposes the nucleotide binding sites may induce

discrete conformational states of the homodimer at this hinge region that are sensitive to the guanosine nucleotide loading state (Figure 6-2). This could explain nucleotide dependent transitions between MMUT hetero-oligomers and active homo-dimers previously described in literature[15].

The continuous nature of the filaments creates a more plausible description of the MMAA-MMUT interaction data then previously reported ring structures that occur under the equivalent ligand bound states[15]. A multiple ring explanation representing different oligomeric arrangements requiring different subunit numbers would enforce a continuum of conformational change in the individual protein building blocks. In contrast, filament repeats can increase particle length without the need for continual conformational change. A filament can therefore be better rationalised from the observed data. In this model a single conformational change would be required to convert higher order oligomers into the predicted hetero-hexamer state induced by MMAA in the GDP bound form.

6.3 MMUT-MMAA results chapter summary

The previously reported negative stain cryo-EM studies, of ligand dependent MMUT MMAA complexes, have helped identify transitions between active homo dimeric MMUT enzymes and the reactivation pathway that requires formation of hetero-hexamers with the GTPase MMAA[15]. The reported higher order oligomeric ring complexes of the MMAA-MMUT interaction are described to be a MMUT sequestering mechanism in both the absence of Cbl and the GTP bound conformation of MMAA[15]. Through in solution study of equivalent ligand states, data in this work offers an alternative filamentous conformation to describe the potential MMUT MMAA sequestering mechanism. These structural models provide a more

plausible explanation of the multiple oligomeric states that are isolated through *in vitro* reconstitution of the complex. Here, it is hypothesised that the evolution of separate MMAA homodimers, in comparison to the MMUT-GTPase fusion enzyme IcmF, provides unsaturated interaction sites in the human complex that could be utilised in filament formation. A mechanistic model describing transitions between the filamentous structure (shown in this study), the hetero-hexameric form and homodimeric form (shown previously in literature[15]), with dependency on ligand bound state of the two enzymes, is shown in Figure 6-15.

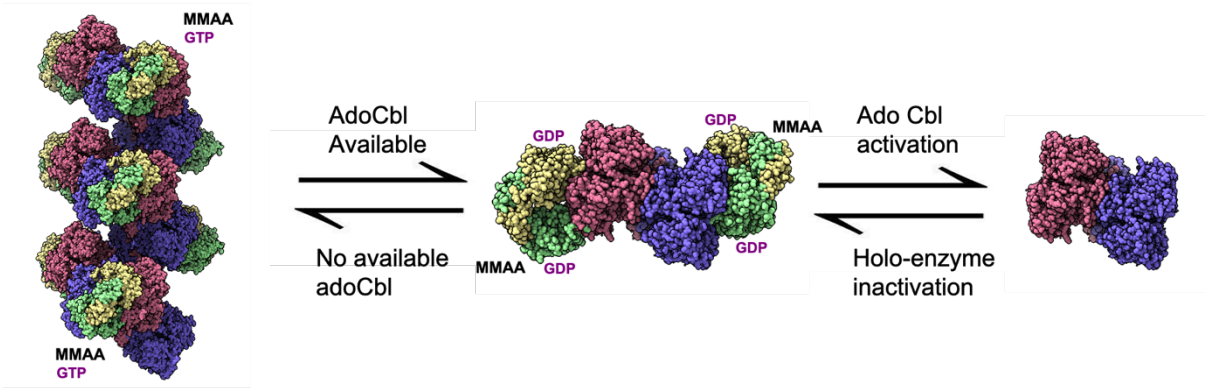


Figure 6-15: Filament formation as an alternative sequestering mechanism in the MMUT reactivation pathway.

The hetero-hexameric to homodimer transitions (right equilibrium symbol) are proposed in previous studies[15]. Here the existence of MMAA-MMUT filaments (left structure) suggests a MMUT sequestering structure in the absence of Cbl. MMAA homodimers are shown in green and yellow. MMUT homodimers are shown in red and blue.

6.4 Introduction to PCC complex results

6.4.1 PCC structure function relationships

PCC catalyses a biotin dependent (vitamin B7) carboxylation reaction in the first step of the propionate catabolism pathway converting the 3-carbon chain CoA derivative, propionyl-CoA into a 4-carbon chain derivative D-methylmalonyl-CoA (Figure 6-16).

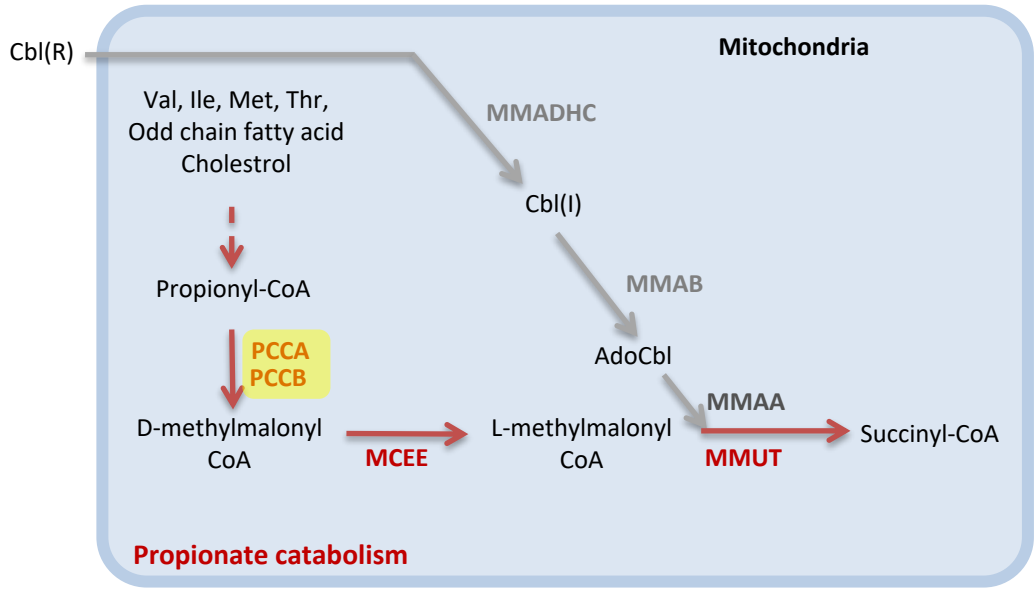


Figure 6-16: Propionate catabolism pathway.

PCC obligate complex consisting of PCCA and PCCB subunits is highlighted yellow in the first step of the propionate catabolism pathway.

PCC falls within a larger family of acyl-CoA carboxylases that share conserved sequence components, namely the BC, BCCP (encoded by PCCA) that catalyse the first half reaction (Figure 2-6) and the CoA esters binding portion of the CT domain (encoded by PCCB)[70,79]. The reaction mechanisms that occur at the BC and CT active sites, among the family members, are also conserved. However, interestingly the family has evolved strikingly different quaternary structures to enable presentation of the BCCP bound biotin group to the BC domain for carboxylation and subsequently divert carboxyl biotin group to the CT domain for carboxyl-transfer to the substrate[79]. The crystal structure of a bacterial chimeric version of the PCC holoenzyme (where PCCA and PCCB constructs where taken from separate bacterial organisms for reasons of protein stability) was solved revealing a hetero-dodecameric complex comprised of 6 chains of PCCA and 6 chains of PCCB arranged with dihedral (D3) symmetry[71]. This arrangement consists of a homo-

hexameric core consisting of two rings of PCCB trimers with cyclic (C3) symmetry that can bind 6 PCCA subunits. The PCCA subunits associate to the top and bottom of the complex resulting in a four-ring configuration α3-β3-β3-α3 (Figure 6-17A)[71].

In the structure the biotinylated BCCP domain interacts directly with the active site of one PCCB CT domain (Figure 6-17). Whilst the major interaction that stabilises PCCA with the PCCB core is between a hook domain that sits between the BC and BCCP domains of one PCCA subunit and the CT domain an opposing PCCB subunit (Figure 6-17). This domain forms a large central helix encase by a β-barrel formed by eight-stranded anti-parallel β-strands (Fig. 7C). The helix extends into the core of the CT domain and a long loop region connects the C-terminal BCCP domain positioned in the active site of the next PCCB subunit in the ring. The two interaction sites formed by the hook domain and the BCCP domain within one PCCA subunit are therefore contacting two separate PCCB subunits bridging the cyclic interaction interfaces within the homo-trimeric rings of PCCB. PCCA subunits do not form interactions with one another creating upper and lower PCCA rings that are offset from the central PCCB rings (Figure 6-17)[71]. To date high resolution structural information only records the conformation where the BCCP is poised in the active site of PCCB ready to carry out the second half of the reaction. No information exists for how BCCP engages its own active site in the BC domain of PCCA. Approximately ~55 Å separate the two repeated active sites within the hetero-complex (Figure 6-17B). The covalent biotin cofactor linkage itself is highly flexible, however the range of conformations it could adopt only allow the N1' reactive atom to extend to a total distance of ~16 Å from the BCCP backbone. Subsequently, significant conformational changes are required within the PCCA subunit in order to relocate and orient the biotin cofactor in a favourable position to catalyse the first half reaction

in the BC domain. Given the length and unstructured nature of the linker connecting the hook domain to the BCCP domain (Figure 6-17B) it is expected that most of this conformational change occurs through translocation of the entire BCCP domain.

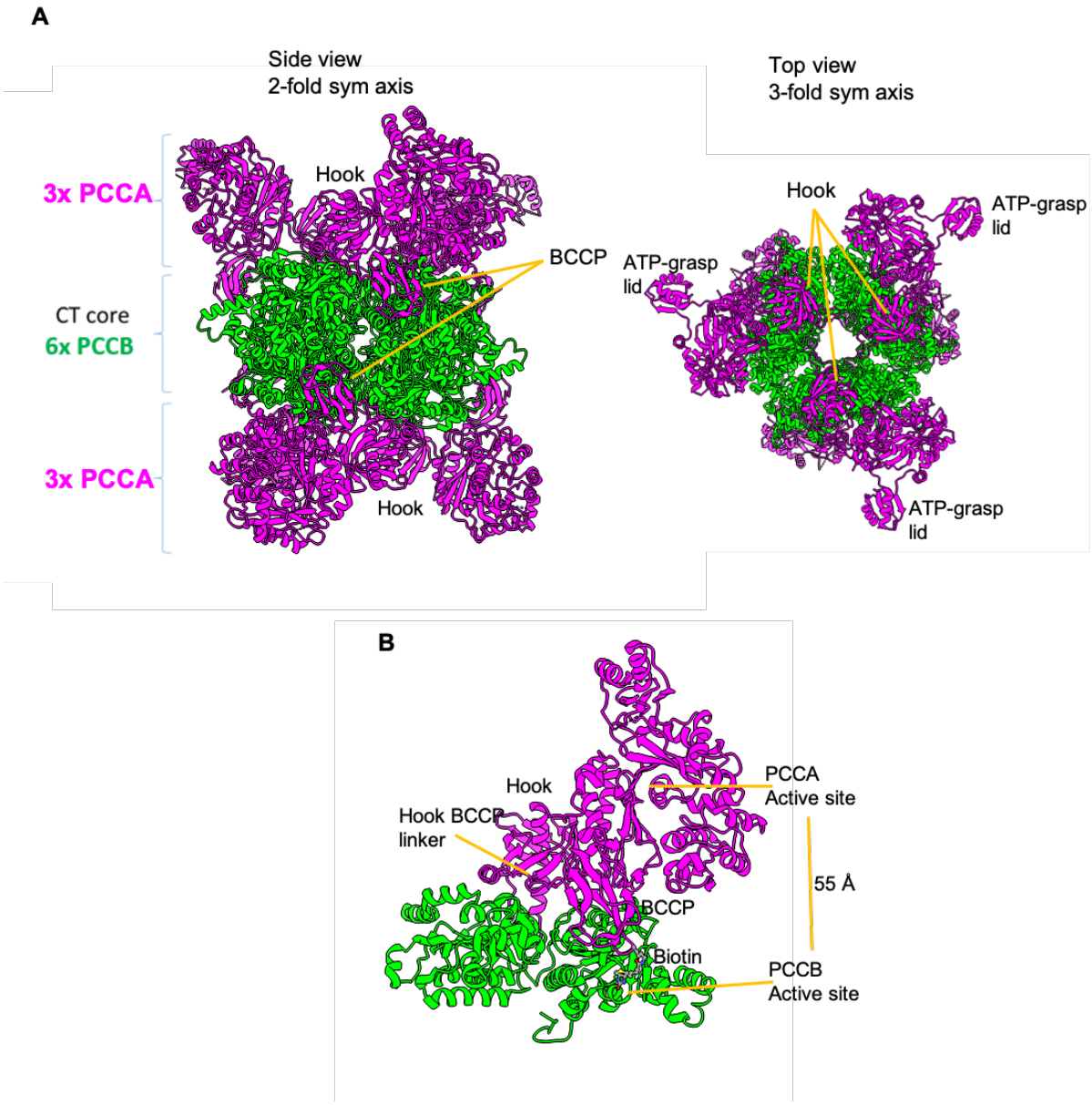


Figure 6-17: Structure of bacterial chimeric PCC holoenzyme:

A: Cartoon depictions of the dodecameric PCC holoenzyme showing side view, along 2-fold symmetry axis (left) and top view, along the 3-fold symmetry axis (right). PCCA chains are coloured purple and PCCB chains coloured green. **B:** Cartoon depiction of a single PCCA chain (purple) engaging the PCCB (green) active site via the BCCP domain.

6.4.2 PCC results chapter synopsis

In this chapter, Attempts are made to isolate novel conformations of the PCC complex using SAXS, and cryo-EM to add further detail to our understanding of the conformational plasticity required in the PCC catalytic mechanism.

6.5 Structural studies of PCC complex

6.5.1 Cloning and expression of PCC complex

In order to obtain, for structural and biophysical studies, the soluble complex of human PCC with mature forms of hsPCCA and hsPCCB, lacking their N-terminal (Nt) mitochondrial targeting sequences (hsPCCA aa 53-728 and hsPCCB aa 31-539), the same bicistronic *E. coli* co-expression method applied to the isolation of hsA-SCS (Section 4.2.1) was used. This resulted in a pNIC28-Bsa4 plasmid with a bi-cistronic open reading frame, where the first coding sequence contained the PCCA subunit with a vector-encoded tobacco etch virus protease cleavable Nt-His_6-tag and the second coding sequence containing untagged PCCB (Figure 6-18).

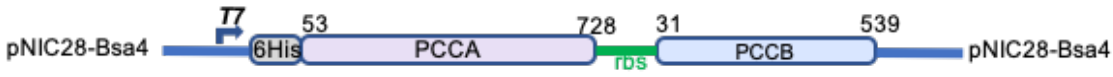


Figure 6-18 Bi-cistronic hsPCC expression vectors.

DNA fragments encoding Nt-His_6-tagged PCCA (53-728) and non-tagged PCCB (31-539), interspersed by an in-frame ribosomal binding site (rbs) (green bar) are sub-cloned into the pNIC28-Bsa4 vector (blue bar).

6.5.2 Purification and complex validation

To obtain adequate amounts of protein material, to test for complex production and structural studies, expression of the PCC complex carried out in 6 L volumes and purified from *E.coli* cell lysates using a 3 step strategy consisting of Ni^{2+} affinity chromatography, size exclusion gel filtration chromatography and reverse Ni^{2+} affinity chromatography, following His tag cleavage. Both components of the PCC complex were present in large portions after capture of His-tagged PCCA via Ni^{2+} affinity chromatography (Figure 6-19A, gel inset 'Ni^{2+} elution'). The presence of co-eluting PCCB provides initial indication of successful complex formation. For size exclusion chromatography, a gel filtration column (Superose 6) with a resolving range for globular proteins between 5 to 5000 kDa was chosen, based on the expected 750 kDa hetero-dodecameric arrangement of the PCCA and PCCB subunits[71]. The elution profile revealed two absorbance peaks measured at 280 nm across the elution volume (Ve) of the column. Subsequent SDS-PAGE analysis reveals the first peak at Ve 70 ml represented a stoichiometric co-elution of both proteins (Figure 6-19A, gel inset 'peak 1'), indicating both proteins form a stable complex over the time course of the chromatography experiment. The second peak (Ve 90 ml) contained a large amount of excess PCCA only (Figure 6-19A, gel inset 'peak 2'). Intact mass spectrometry analysis of the complex peak reveals the presence of both proteins with a mixture of PCCA in the biotinylated and apo forms (Figure 6-19B). The PCCA His-tag was cleaved by TEV protease and the complex was further purified by reverse affinity chromatography. Pure complex was isolated in the flow-through, while the contaminant GroEL, which has a similar mass to the PCC complex and is a common contaminant from *E.coli* expression, was retained in the affinity resin. To overcome the lack of fully biotinylated complex, PCC was co-

expressed in *E coli* with a BirA-expressing plasmid and excess biotin supplemented in the media. BirA is the *E coli* enzyme responsible for recognising the conserved, sequence specific biotinylation site and subsequently catalyses its covalent linkage to Lys699. After purification, mass spectrometry analysis revealed 100 % incorporation of biotin in PCCA (Figure 6-19B). This form of the complex was used in all subsequent experiments.

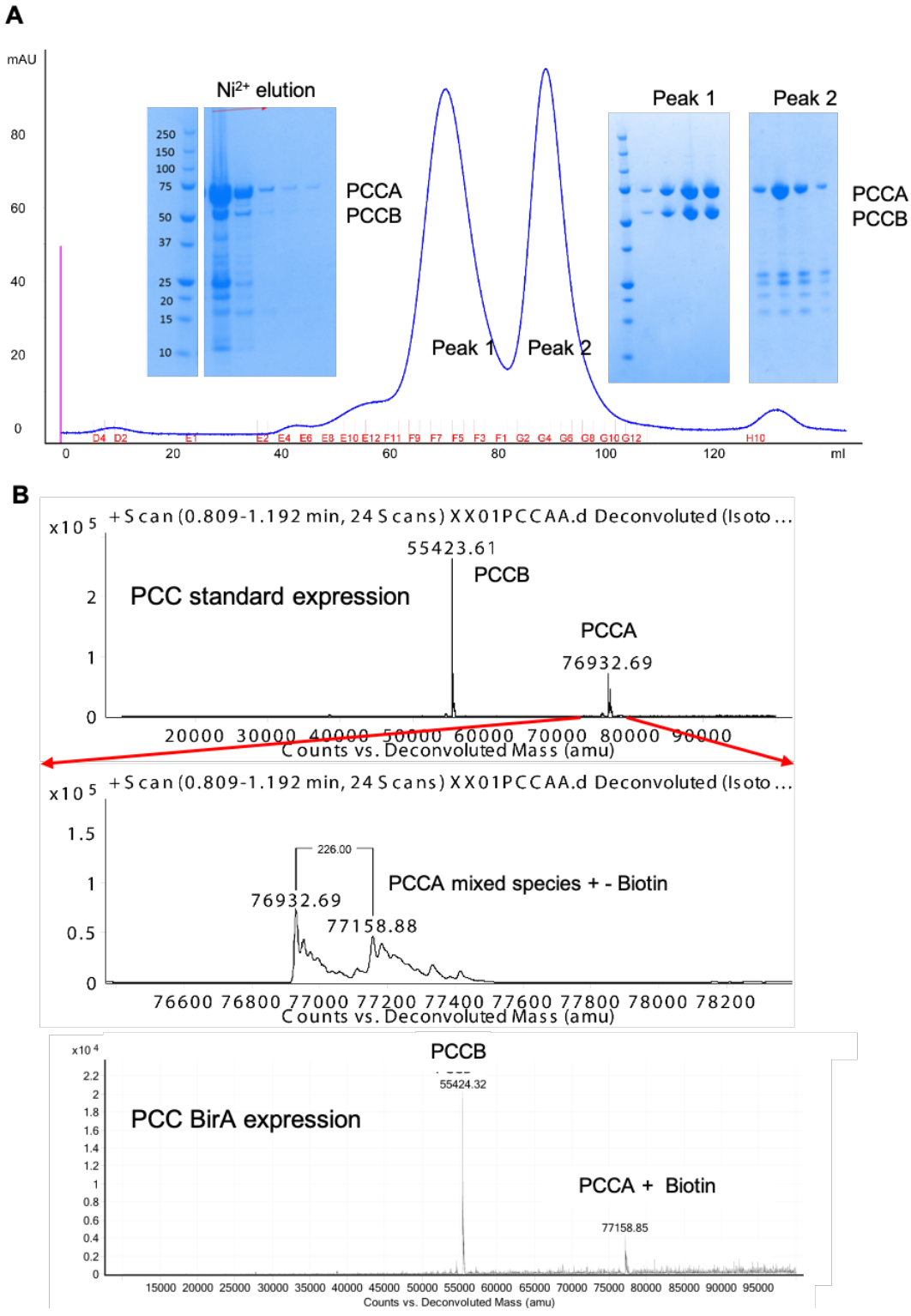


Figure 6-19: Isolation of PCC complex.

A: SDS PAGE of Ni^{2+} affinity purification and size exclusion purification with chromatogram showing stoichiometric complex isolation (peak 1) and separation of excess PCCA (peak 2)**. B:** Intact mass spectrometry analysis showing mass of complex pool from peak 1 of size exclusion chromatography.

Using standard expression, PCCA is isolated in mixed biotinylated state (top and middle). Using BirA co-expression, PCCA is fully biotinylated (bottom).

6.5.3 PCC SAXS determination of complex size and shape

Small angle X-ray scattering (SAXS) was used following purification of PCC to confirm successful isolation of a stable intact complex and to determine the average composition, size and shape of each complex in solution for comparison with previously reported literature. The SEC elution profile in the SEC-SAXS experiment reveals a single peak for the PCC complex (Figure 6-20A). Unlike the purification profile, no excess PCCA is observed, indicating that the complex as isolated and purified remained a single stable species, making it ideal for size and shape determination to compare against the high-resolution orthologue models of the PCC. As with previously described SEC-SAXS datasets, frames with similar estimated Rg values were buffer-subtracted, averaged and displayed in a Log*10* intensity plot (Figure 6-20B).

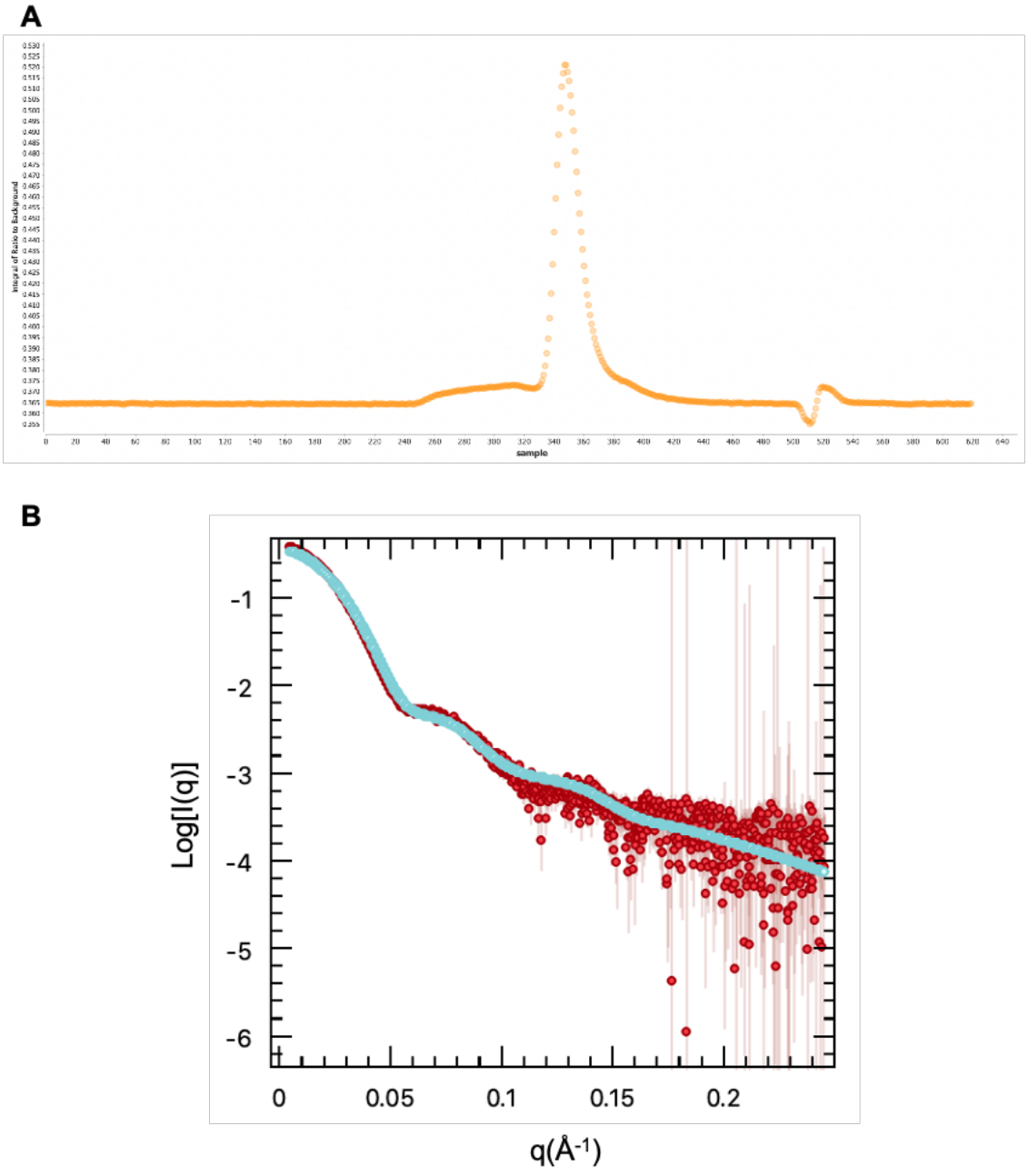


Figure 6-20: Log10 intensity plot for the PCC complex.

A: SEC-SAXS elution profile. Each point (yellow) represents the integrated area of the ratio, of the SAXS curve representing each sample frame (X-axis SAXS measurement taken every 3 seconds of the SEC elution time course), to the estimated background (left Y-axis) **B:** Log10 intensity plot for the PCC complex**.** Red circles correspond to the experimental scattering curve and the blue line represents the theoretical scattering curve for the orthologue bacterial crystal structure**.**

The MW calculation, using the Vc/Rg^2 method[100], as an invariant proportional to molecular mass, is predicted to be 830 kDa. This is close to the expected hetero-dodecameric size of 773 kDa, calculated with 6 PCCA subunits (each 74 kDa) and 6 PCCB subunits (each 55 kDa). In addition to confirming particle size, the agreement of particle shape to the previously known structure is shown by back calculating the theoretical scattering curve for the bacterial crystallographic hetero-dodecamer (PDB 3N6R[71]) and fitting this to the experimental curves determined here using the programme *Oligomer*[89](Figure 6-20B). A χ^2 value of 20 reveals a reasonable correlation between the solution structure and the back-calculated bacterial structure, indicating that the average conformation of this sample at low resolution is similar although not identical. As a final check, the *ab initio* model is calculated using 20 rounds of *Dammif*[89] runs. These models were aligned, averaged and filtered with the *Damaver*[89] program to provide a final low-resolution envelope of the complex with D3 symmetry imposed (Figure 6-21). D3 symmetry is assumed based on the calculated MW size indicating isolation of a dodecamer and the experimental scattering correlation with the theoretical scattering of the known bacterial structure. Given the pre-existing knowledge of the low-resolution shape of the bacterial structure[71], the prolate anisometry was forced in reconstruction ensuring correct alignment of the symmetry axis. The result, when overlaid with the crystal structure, gives a very good fit (Figure 6-21).

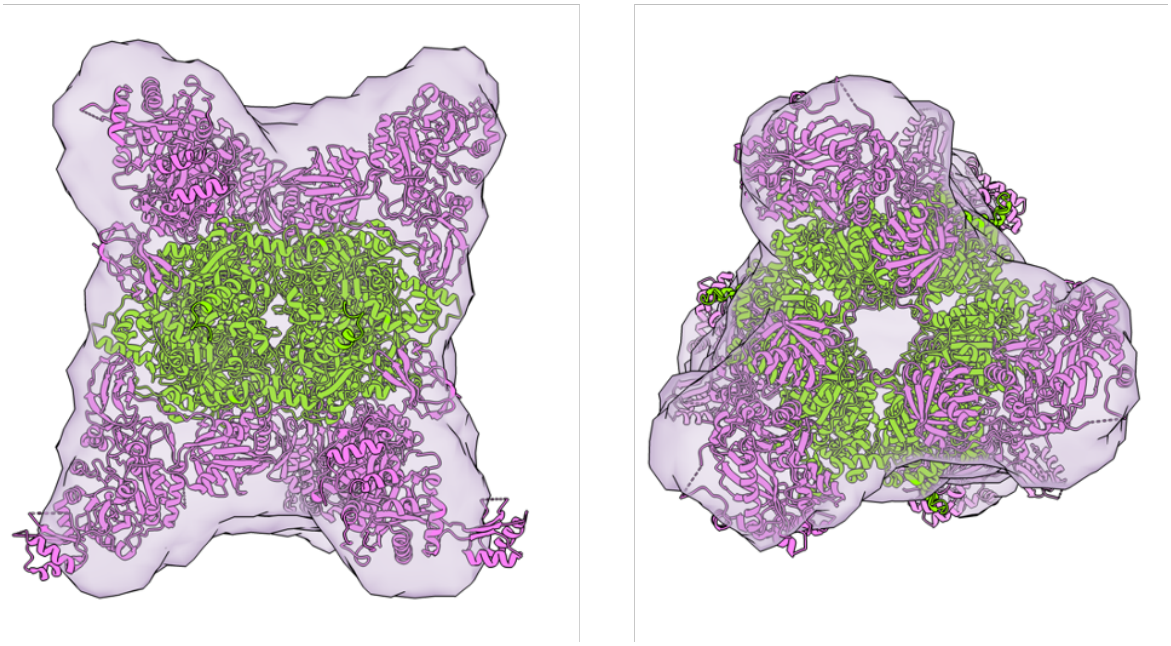


Figure 6-21: Ab initio modelling of PCC SAXS envelope.

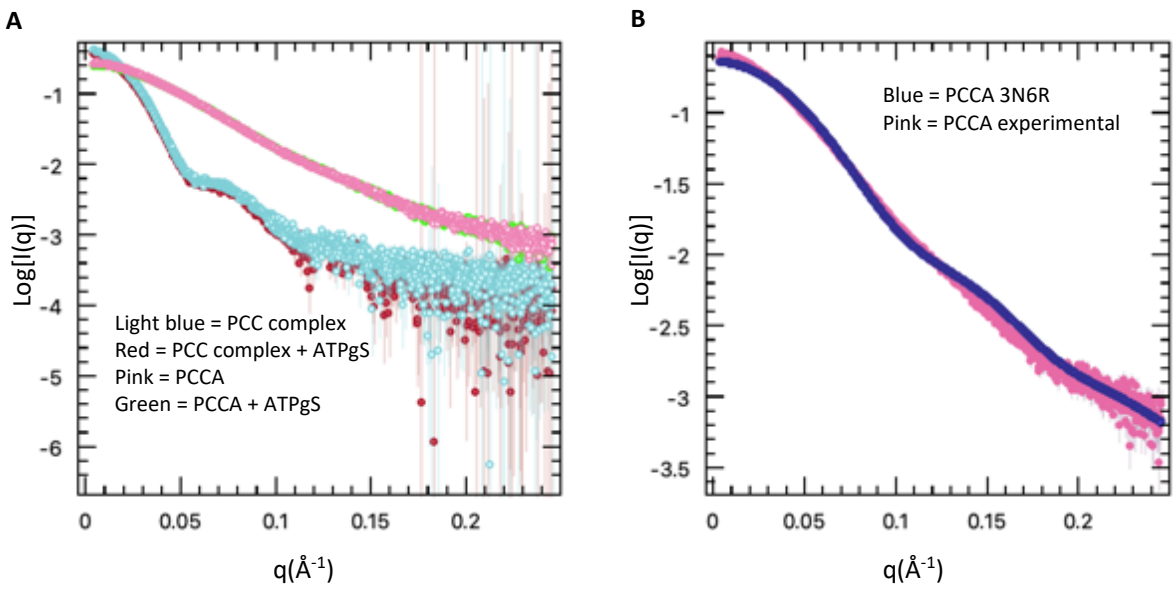
PCC SAXS model produced by *dammif* and *damaver*[89] overlaid with the bacterial orthologue crystal structure (PDB 3N6R)[71]. The six PCCA subunits are coloured purple, the six PCCB subunits coloured green and SAXS density mauve. Left is a side view aligned on the 2-fold symmetry axis whilst right is a top down view aligned along the 3-fold symmetry axis.

6.5.4 Probing conformational change of the PCC complex

To study conformational changes that might be observed in the SAXS scattering curves, the PCC complex was pre-incubated with an ATP analogue (ATPγS). Trapping the complex in an ATP bound form may help to isolate a conformation where the BCCP domain engages the BC domain ready for the first half reaction (Figure 2-6). The SEC-SAXS profile was collected for comparison with the presumed ligand free dataset. The same experiment, in the presence and absence of ATPγS, was performed on the PCCA-only sample, separated from the complex via gel filtration to again test for conformational change upon ligand binding (Figure 6-19A). MW calculations of the PCCA-only SAXS dataset revealed an average size of 73 kDa, indicating that in isolation PCCA exists as a monomer (expected MW 74 kDa)

(Figure 6-22C). The existence of PCCA as monomers is expected due to the lack of homo-oligomerisation interfaces contributed by PCCA within the bacterial complex structure[71]. The SAXS scattering curve for one PCCA subunit taken from the bacterial complex structure was back calculated and fit to the experimental SAXS scattering data again using *Oligomer*[89]. The χ^2 value of 22 reveals a reasonable correlation between PCCA monomers from the solution scattering data and previously calculated bacterial structure (Figure 6-22B).

In the presence of ATPγS, visual inspection of an overlay of the log10 scattering curves did not reveal any detectable differences in particle shape or size when compared against the equivalent as purified datasets for both the PCC complex (Figure 6-22A, light blue circles ligand free, red circles ATPγS bound) and PCCA only samples (Figure 6-22A, pink circles ligand free, Green circles ATPγS bound). In addition, Rg, I(0), Dmax and MW calculations remain unchanged providing quantitative evidence for a lack of detectable conformational changes (Figure 6-22C). This perhaps is only an indication that the resolution limit of the experiment is not high enough to detect the potentially small domain movement in relation to the rest of complex structure. It is therefore clear that structure determination using x-ray crystallography or cryo-EM would be required to provide higher resolution information that can deconvolute different conformations.



C

Protein sample	PCC	PCC + ATPgS	PCCA	PCCA + ATPgS
Guinier				
I(0)	0.37	0.41	0.26	0.27
Rg (Å)	67.76	67.80	35.61	36.70
P(r)				
I(0)	0.38	0.42	0.25	0.26
Rg (Å)	67.13	67.20	36.70	37.65
Dmax	220	220	145	145
Particle size				
MW (kDa)	830	820	73	76
Model fitting				
Crystal structure fit (X^2)	20 (3N6R)	16 (3N6R)	22 (3N6R PCCA)	25 (3N6R PCCA)

Figure 6-22: Probing conformational changes in PCC using SAXS.

A: Log10 scattering intensity plots for the PCC complex and PCCA subunit in the presence and absence of ATPγS. **B:** Log10 scattering intensity plot of the PCCA subunit. Pink circles correspond to the experimental scattering curve and the dark blue line represents the theoretical scattering curve for the orthologue PCCA bacterial crystal structure (3N6R)**. C:** SAXS scattering table parameters for the PCC and PCCA datasets. Values for radios of gyration (Rg), zero scattering angle (I(0)), and normalised zero scattering angle (I(0)) for protein concentration (I(0)/protein conc) are calculated by Guinier fitting and real space distribution plots are. Dmax values calculated by real space distribution plots. MW values calculated by the Vc/Rg^2 method[100]. Crystal fit using the PCC structure (3N6R) or PCCA monomeric structure (extracted from 3N6R)[71] to the experimental data is calculated using *Oligomer*.[89]

6.5.5 Crystallography trials of PCC

In order to gain higher resolution data of the PCC complex, that may serve to capture different conformational states of the BCCP domain engaging either the BC domain of PCCA or the CT domain of PCCB, crystallisation trials were set up. As per standard procedures, commercial random matrix crystallisation screens were first implemented to test broadly for conditions that yield crystalline material, under several different protein concentrations and temperatures. Crystals of the as purified and ATPγS incubated PCC complex were readily obtained from a number of crystallisation conditions, which generally appeared after 12 hours of setup. However, in many conditions, crystal growth was rapidly followed by deterioration within days of first nucleation, particularly in conditions buffered to high pH (Figure 6-23).

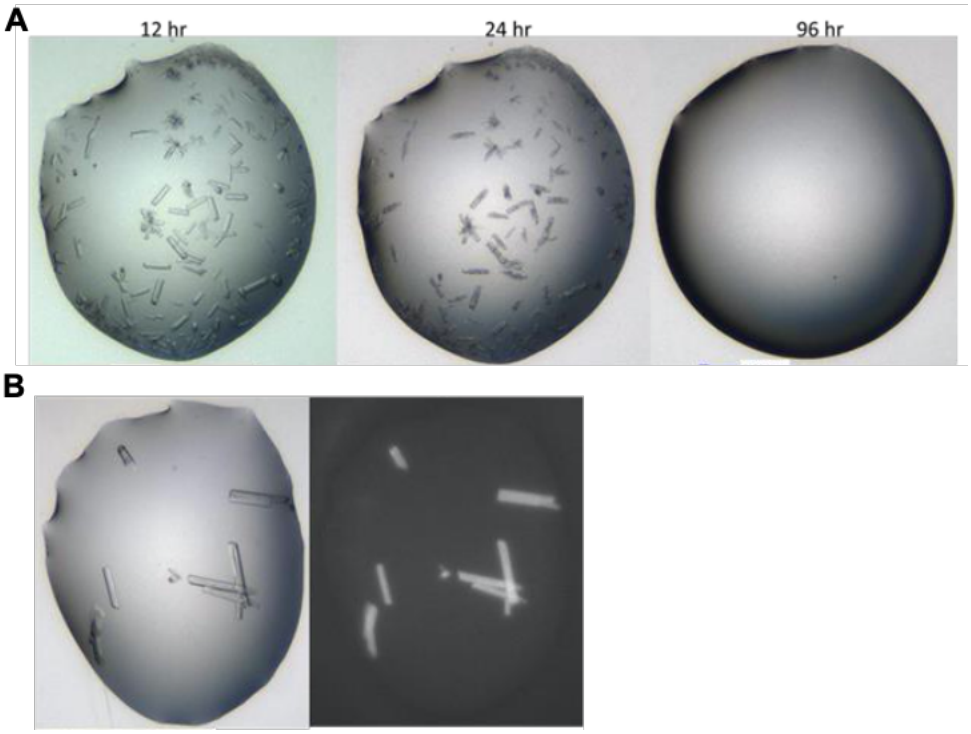


Figure 6-23: PCC crystal degradation.

A: Crystal nucleation and growth is rapid however conditions, particularly at high pH, led to rapid crystal re-solubilisation. **B:** UV imaging right confirmed absorbance of crystal at 280 nm indicating presence of protein.

Some crystals where successfully harvested prior to visible deterioration allowing x-ray measurement at the diamond light source synchrotron with diffraction recorded to around 11 Å resolution, providing further evidence that these are protein crystals. The conditions that yielded the most stable crystals were further optimised in follow-up screens using fine grids of pH and precipitant gradients. Crystals harvested form this round diffracted to around 5 Å resolution. Random matrix micro seeding (described in section 4.2.2), was attempted but with no improvement in diffraction quality. To test whether cryo-protection and crystal handling was a limiting factor in resolution, in situ diffraction was tested at room temperature using the VMXi beamline at the Diamond Light Source. The benefit of this diffraction experiment is that crystals are not subject to any unnecessary change in conditions from the point of growth to the point of data collection such as exposure to air, dilution in cryo protectant or “rough handling” from myself. An additional rationale is that those crystals that rapidly degrade could have datasets potentially recorded *in situ*. Several crystals were shot using this method with the highest diffraction unfortunately reaching just 5.7 Å resolution. After merging the highest resolution frames from two datasets using Xia2 auto processing pipelines only 30 degrees of data was collected despite the potential of 60-degree wedge collections per crystal shot, hence resulting in just 38 % overall data completeness. The combination of no resolution improvement *in situ* and poor crystal alignment quality (the latter a consequence of early phase of VMXi beamline development) resulted in the abandonment of attempts to solve the crystal structure of the PCC complex.

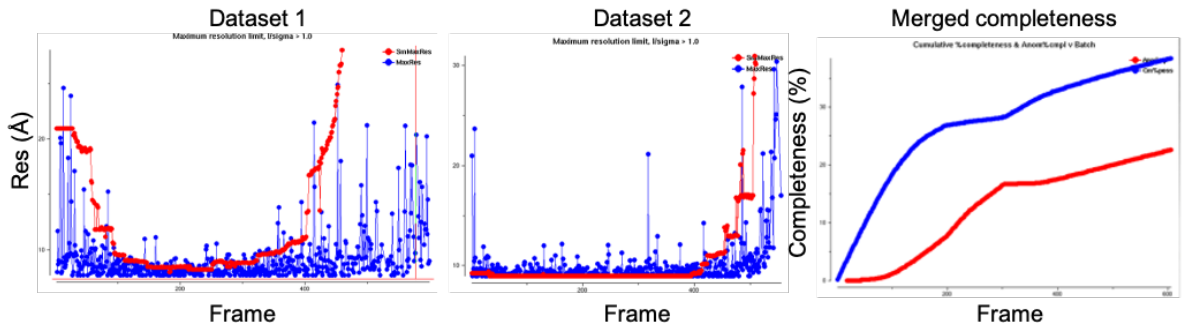


Figure 14: VMXi collection of PCC complex. Left and middle show frame by frame resolution estimates based on signal cut off of I/sigma >1 for two datasets with equivalent space groups. Right shows completeness after merging.

6.5.6 Cryo-EM structure of PCC complex reveals conformational changes

Due to its size and internal symmetry, the PCC complex, theoretically, is very amenable to cryo-EM experiments. cryo-EM is also capable of resolving discrete protein conformations at high resolution, making it a viable option to capture the conformational dynamics of the BCCP domain that allow for biotin carboxylation and carboxyl transfer to occur in separate active sites. Previous bacterial structures of the PCC complex[71] have captured the enzyme in a conformation that is competent for the second half-reaction (Figure 2-6), where the BCCP domain of PCCA is engaged in the PCCB active site for carboxy transfer to the substrate. In this study, to trap the complex in a conformation poised for the first half–reaction (Figure 2-6), the ATP analogue ATPγS was preincubated with purified PCC, and the sample was subsequently vitrified on cryo-EM grids. The addition of an ATP analogue was used in an attempt to flip the BCCP domain into the PCCA active site ready for the biotin carboxylation half-reaction to take place.

The complex sample was screened on the Glacios microscope at the Oxford Particle Imaging Centre, using the same set us as previously described (section 5.4.3). Particles were monodispersed with high contrast and multiple orientations where

visible including some top down 3-fold symmetry views several different side views of the complex. The data quality warranted a full dataset collection using the same microscope settings. A total of 813 micrograph movies were recorded through a defocus range of -1.5 uM to -3 uM (with 0.3 um steps). The movies were motion corrected and ctf corrected[90]. 88781 Particles were picked using the *relion* auto picking programme[90]. Rounds of 2D classification were used to filter out junk particles revealing high resolution 2D projections of multiple orientations of the complex from a total of 67946 particles (Figure 6-24A)[90]. Using this subset of particles, an initial 3D model was created[90]. This was used as a reference for 3D classification into 8 different classes (Figure 6-24B)[90]. The 8 classes all resembled the expected PCC complex, but appeared to vary in resolution quality particularly within the PCCA subunits, indicating their flexibility. One 3D class showed high resolution density for the intact complex and consisted of 26722 particles (Figure 6-24B, red box).

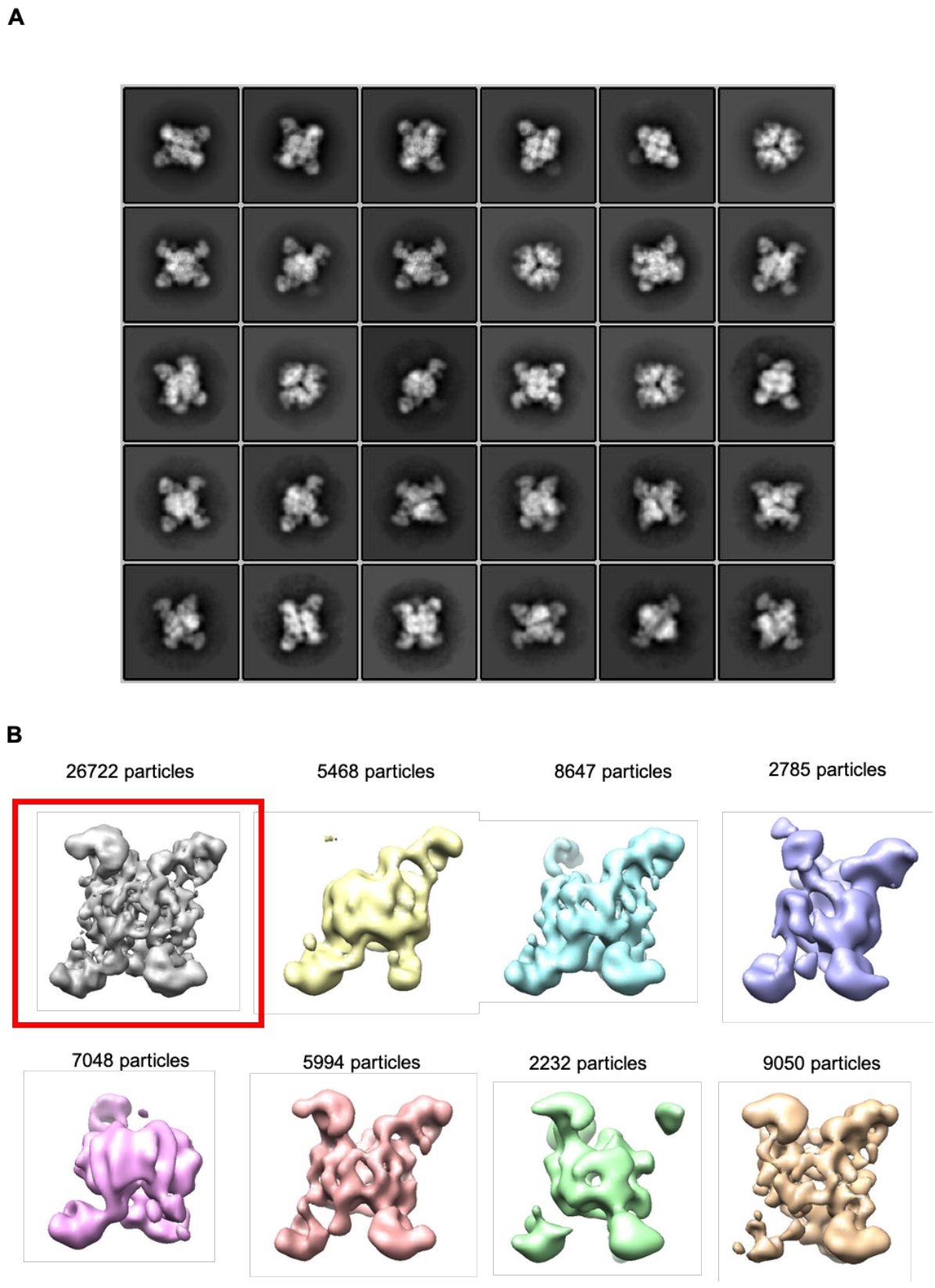


Figure 6-24: Cryo-EM 2D and 3D classifications of PCC complex.

A: 2D classes after filtering junk particles. Final number of particles 67946. **B:** 8 3D classes after 3D classification of the 67946 particles taken from the 2D classes. Red box indicates class taken forward

for refinement. The 67946 particles from all 8 3D classes were used in symmetry expand strategy (Figure 6-27).

These particles from the largest class were subsequently 3D refined to a global resolution of 3.3 Å, with D3 symmetry imposed on the map (Figure 6-25A). Overlay of the bacterial PCC complex, via rigid body fitting, shows that it is a very good fit to the human EM structure, in line with the SAXS model. Based on this map overlay, human PCC adopts the same arrangement of subunits as the bacterial complex forming a heterododecamer, with a core of 2 trimer rings of PCCB subunits flanked on both sides of the core by 3 PCCA subunits each (Figure 6-25B). Significant local divergence in conformation relative to the bacterial structure is detected in the PCCA subunit. This subunit houses the catalytic BCCP and BC domains as well as the hook domain responsible for binding PCCB. A substantial portion of the BC domain is formed by an ATP grasp fold which utilises ATP hydrolysis to power carboxylation of the biotinylated BCCP domain. (Figure 6-25C). In contrast to the bacterial structure, the lid of the ATP grasp domain has closed over the ATP binding pocket, likely a result of ATP binding. However, ATP is not yet modelled and confirmed in the map (due to time constraints at point of writing). Initial analysis of the BCCP domain shows it engaging the PCCB core, with density for the biotin cofactor reaching into the active site of PCCB. This orientation of the BCCP domain with respect to PCCB is predominantly observed in all previous structures of biotin dependent carboxylases including PCC and other members.

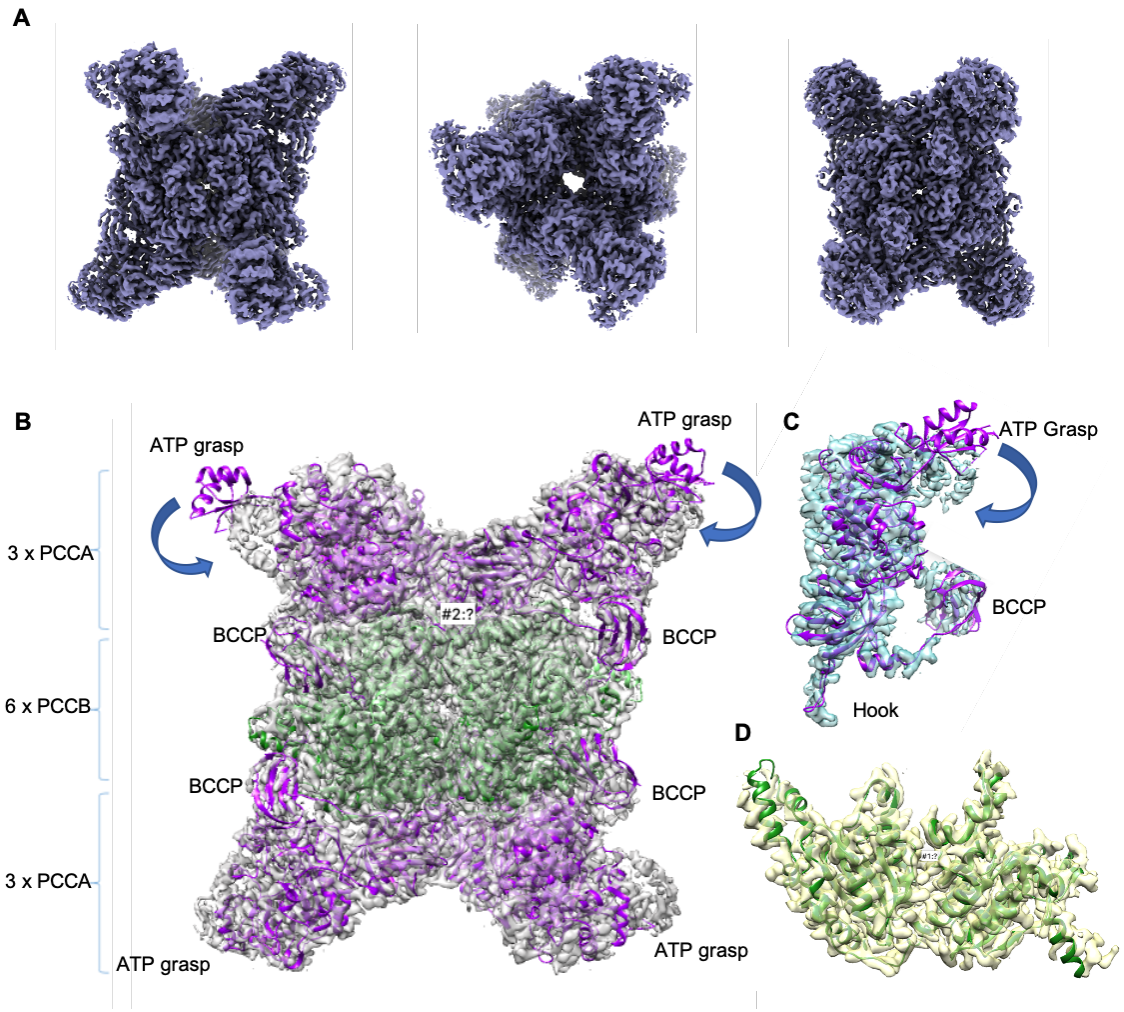


Figure 6-25: Final cryo-EM map of PCC complex at 3.3 Å resolution.

A: D3 symmetry refined map (dark blue) of the PCC complex displayed along C2 symmetry axis (left and right) and C3 symmetry axis (middle). **B:** rigid body fitting of the bacterial crystal structure 3N6R within the 3.3 Å cryo-EM density map**.** Bacterial PCCA subunits coloured purple and PCCA subunits coloured green and density map for whole complex grey. **C:** Rigid body fit of single bacterial PCCA subunit (purple) (taken from PDB 3N6R) within extracted density map for a single PCCA subunit (blue) **D:** Rigid body fit of single bacterial PCCB subunit (green) (taken from PDB 3N6R) within extracted density map for a single PCCB subunit (yellow)

Analysis of the local resolution in this cryo-EM reconstruction shows the majority of the PCCB core reaching beyond the 3.3 Å global resolution estimate. The hook domain of the PCCA subunit, that makes substantial interactions with the PCCB core, is also high resolution of around 3.3 Å. However, the PCCA BCCP domain and

ATP binding domains are predicted to be at lower resolution of 5 Å (Figure 6-26). Lower resolution areas of a cryo-EM map often reflect the increased conformational heterogeneity present in this area within the particles relative to more homogenous parts of the structure. These could be regions that adopt multiple discrete conformations or continuous movement that locally disobey the D3 symmetry rules applied to the processing. In the case of PCC, this is consistent with the notion that these elements must undergo conformational change in order for catalysis to occur.

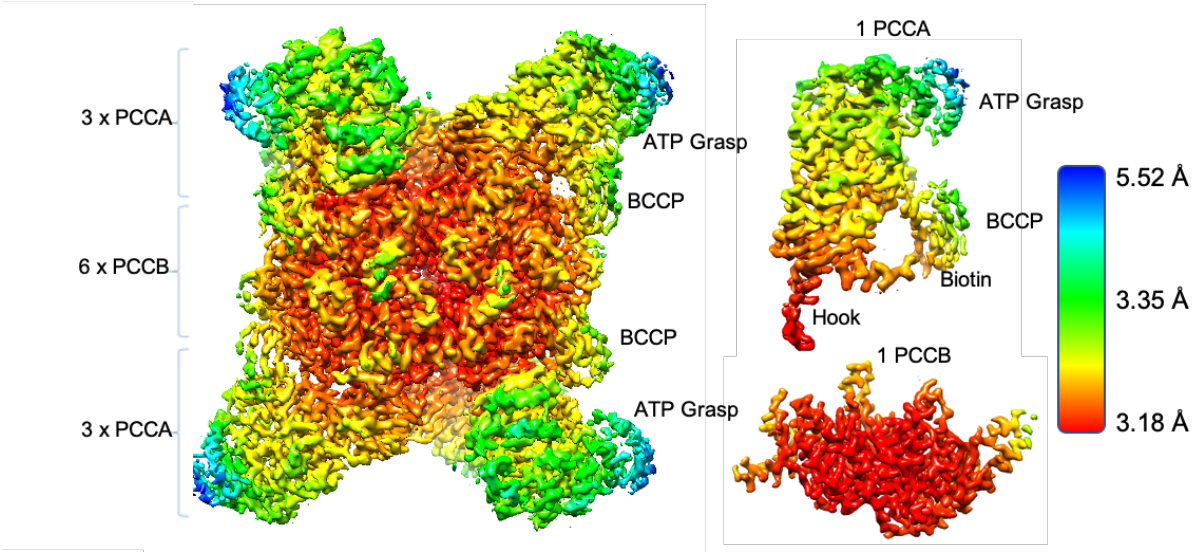


Figure 6-26: Local resolution estimates for the PCC cryo-EM map.

Local resolution estimate for the PCC cryo-EM map. Left shows holocomplex, right top shows density for 1 PCCA subunit and right bottom shows density for 1 PCCB subunit. Maps are coloured by resolution using scale bar right.

The effect of applying D3 symmetry only serves to increase the resolution of the most stable areas of the map and could have averaged away differences in local conformational states between particles. It is therefore plausible that within this subset of particles there may be a further subset of instances where the BCCP domains adopt alternative conformations such as engaging the BC domain PCCA subunit, instead of the more universally observed orientation towards the CT domain

of PCCB. To attempt to isolate different conformations within this subset of particles, a second round of 3D classification was carried out without imposing symmetry. In this classification, the 3D refined model was used as a reference and the 3D classification 'T value' was increased to improve stringency of classification. This strategy was still not sufficient to separate different conformations, indicating that there are not enough particles in the dataset that adopt identical conformations to deconvolute the heterogeneity.

To observe different conformations to high resolution, a larger dataset would therefore be required. In the absence of additional data, however, it is possible to use the internal symmetry within the PCC complex to our advantage. So far, the complex has been analysed as an intact dodecamer only. Within the complex the heterodimeric asymmetric unit formed by one PCCA and one PCCB subunit is repeated 6 times within each particle. This means the total number of asymmetric units that have the potential to adopt different conformations is six-fold higher than the total number of complex particles. By classifying the entire complex, as per standard procedures, the data is restricted to requiring all 6 asymmetric units to be in equivalent conformation for high-resolution structure determination. To overcome this limitation, symmetry expansion was carried out to align, average and classify the asymmetric units individually. In this instance individual particle complexes are duplicated to 6 identical particles each containing the same 6 asymmetric units that may adopt slightly different local conformations (Figure 6-27A). The duplicated particles are rotated along the symmetry D3 axis 6 times so that each asymmetric unit within the duplicates of one complex will be aligned with all other asymmetric units. A mask is then applied around one of these asymmetric units which allows 3D classification to ignore the rest of the complex and only classify different

conformations within this region (Figure 6-27B). This effectively increases the number of particles 6-fold and allows the classification of all potential conformations against each other, circumventing the issue of averaging out local conformation when the entire complexes are classified against each other. This approach is well documented method to deal with local deviation from global particle symmetry[135] and was originally applied to refine the vertices of the Kelp fly virus capsids[136].

This strategy has been applied here to particles from all 8 3D classes that were classified in the first round of 3D classification, (Figure 6-24B) in order to increase the quantity of potential PCCA-PCCB conformations captured. After symmetry expansion, 8 new classes where created using masked 3D classification of a single PCCA subunit, 3 of these classes isolated PCCA in the original conformation observed in the first round of EM processing using the full complex. Importantly, 2 of the classes revealed that the PCCA subunit was able to adopt an alternative conformation in the dataset. The alternative conformation, formed by 118963 particles, reveals both the BCCP domain swinging in the direction of the PCCA active site coupled with small closure of the PCCA ATP grasp domain lid over the activesite (Figure 6-27C). To achieve the new conformation the BCCP domain appears to have rotated around the linker region causing an approximate 35 Å translocation of the covalently bound biotin into the active site of the PCCA subunit. The data in the current state is low resolution however, both this subset of particles and the classes representing the conformation where the BCCP interacts with PCCB now represents more homogenous pools of PCCA subunits. Different strategies for refinement, using various sized masks, will be used next in order to attempt to refine the two alternative conformations to higher resolution. These refinements are computationally time consuming are currently ongoing.

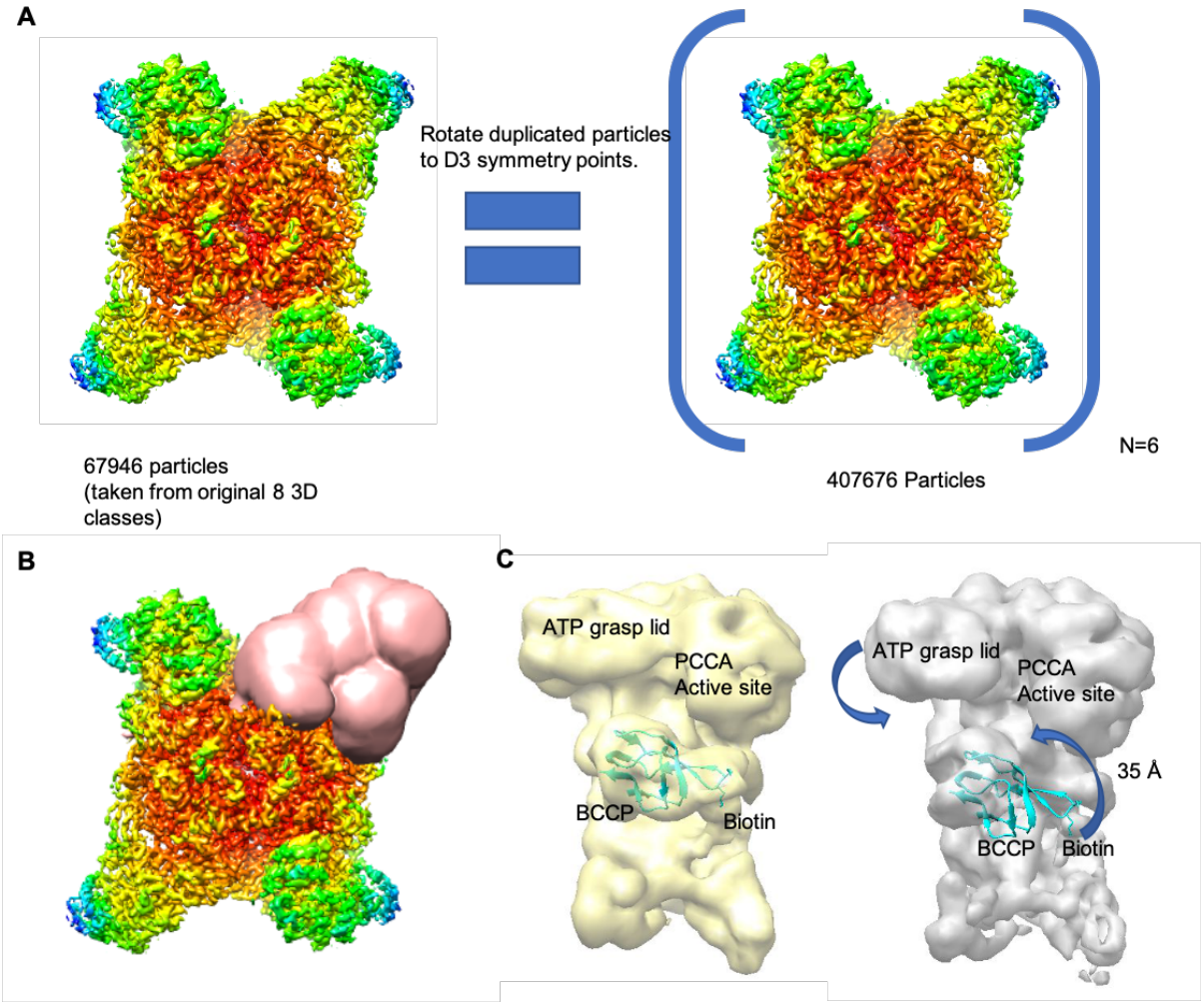


Figure 6-27: symmetry expansion to deconvolute PCCA conformations.

A: Step 1 duplication and re-orientation of particles using D3 symmetry. **B:** Step 2 masked classification of an individual PCCA subunit, using the duplicated particle set, into discrete conformational groups. **C:** 2 PCCA 3D classes revealing alternative conformations where BCCP domain swings toward the PCCA active site/ATP grasp domain. Coloured grey is the new conformation compared to previously isolated class coloured yellow.

6.6 PCC results chapter summary

Previous structures of the PCC complex and other acyl-CoA carboxylases have all been isolated in the conformation where the BCCP domain engages the PCCB subunit poised for the second half of the reaction. Here, a mechanism of cryo-EM

processing taking advantage of the global internal symmetry within the complex has successfully isolated a subset of PCC particles that adopt an alternative conformation of the BCCP domain engaging the PCCA active site. Although processing is still ongoing, the refinement to higher resolution is expected, allowing for modelling of the intra subunit interaction at the residue level that will hopefully provide new insights into reaction mechanism and disease associated mutations.

7 Conclusion and Future Perspective

7.1 ALAS2 and haem biosynthesis

7.1.1 Significance of study

This study provides exciting new structural insights into the regulatory mechanism that has evolved in higher eukaryotes which allows the autoinhibition of ALAS2 via its Ct-extension domain. The data helps explain how the frameshift and indel mutations in the ALAS2 C-terminus lead to the gain-of-function disorder XLP. Finally, the XChem fragment screening campaign provides a starting point for small molecule drug discovery aimed at providing inhibitors that will be clinically effective against the three erythroid specific porphyria disorders downstream of ALAS2 in the haem biosynthetic pathway.

7.1.2 Limitations of work

Whilst the structure of the Nt-truncated hsALAS2 enzyme provides previously unknown insights into the function of the Ct-extension, it is unfortunately only a single snapshot of the conformational dynamics that have likely evolved in higher order eukaryotes to regulate ALAS activity. At least one other conformational state

has to exist, where the ct-extension adopts an alternative conformation, in order to alleviate the inhibitory regulation of the ALAS2 reaction. Significant questions remain as to how, when appropriate in the cellular environment, ALAS2 overcomes this autoinhibitory state. Some suggestions have described the A-SCS interaction with ALAS2 as a potential mechanism to overcome this, because A-SCS has previously been linked to an interaction with the Ct-extension. However, the lack of strong biophysical evidence for this, coupled with a lack of success in the isolation of an ALAS2-A-SCS complex here, makes this hypothesis still quite speculative.

The XChem fragment campaign reveals exciting prospects for the development of small molecule inhibitors for substrate reduction therapies that target the most common downstream porphyrias. However, due to the experimental design the prospect of producing inhibitors for the less common disease, XLP, is much less promising. XLP is caused by ALAS2 Ct-extension indel mutations leading to hyperactivity of the enzyme. The current template to screen for fragment binding is the WT ALAS2 enzyme, where the Ct-extension is blocking the active site. WT ALAS2 was chosen due to an initial academic interest in understanding Ct-autoinhibitory mechanisms but may not necessarily provide the most effective construct for development of clinically useful inhibitors. It is argued here that the observed mode of fragment binding is through interactions with the catalytic core and the Ct-extension. Therefore, an intact Ct-extension could be useful to lock the enzyme in the auto-inhibited state. However, it will only be effective in cases in which the ALAS2 Ct-extension is not truncated. For XLP, however, this is not the case and it would be more practical to repeat the fragment screen on a construct that is not blocked by the presence of the Ct-extension. This way, active site competitors could

be developed to both be useful as substrate reduction approach for the treatment of downstream porphyrias and also for the treatment of XLP.

Detection of weak fragment binding through specific interaction to the protein using X-ray crystallography is unambiguous. However, the use of fragments in ALAS2 activity assays to determine whether the current set of fragments offer some inhibition, as well as the weak binding detected by X-ray crystallography, is limiting. The high concentrations of fragments required to overcome their low affinity could induce non-specific inhibitory effects on the enzyme that would elevate the false positive rate in this experiment. For meaningful results, assays should be performed on larger XCHEM follow up compounds once optimised for affinity and binding specificity.

7.1.3 Future experiments

In the immediate term, the ALAS2 work can be progressed by two pathways. The first is to continue the pursuit of regulatory mechanisms that control haem biosynthesis through regulation of ALAS enzymes. It will be of significant interest to explore and isolate more functional states of hsALAS2 that reveal further mechanistic details of the Ct-extension. That could involve a more comprehensive study of interaction partners that may bind to relieve autoinhibition of the Ct-extension, or simply attempting to optimise conditions to solve the full length hsALAS2 structure in the presence of different substrates. Of the latter point, it would be interesting to us X-ray crystallography to study the conformational changes expected to occur through succinyl-CoA binding to attempt to understand the proposed displacement of the Ct-extension and whether the C-terminus continues to play further roles during different phases of catalysis such as product release[31].

Another way to explore Ct-extension conformational dynamics further is to compare from a structural perspective the similarities and differences of ALAS1 enzymes.

The major inhibitory residues of the ALAS Ct-extension are conserved between isoforms and likely adopt the same binding mode, however, the terminal 9 residues show strong divergence between ALAS1 and ALAS2. To investigate this further, as a starting point isolation of ALAS1 is currently being optimised for future structural study. Protein with different Nt-boundaries have been expressed but are prone to aggregation in solution. Currents attempts are to optimise buffer conditions for protein stability. Successful production of ALAS1 will also be important for testing cross reactivity of potential follow up ALAS2 compounds with ALAS1.

The second way to progress this project further is in the pursuit of small molecule inhibitors to continue the process of drug discovery. Collaborations with chemists will be required to provide the first round of development of the initial fragment hits. Fragment merging through linking fragments within the cluster that bind the catalytic core, ct-extension and dimer interface, represents one possible direction to take. In addition, growing individual fragments into larger molecules that share the same functional groups shown to bind in the fragment molecules will be explored. Both methods will serve to find larger molecules that have stronger binding affinities and subsequently represent more realistic molecules for assessing enzyme inhibition and cross reactivity with ALAS1. Constructs are also in development for follow up fragment screens that lack the C-terminal two-turn helix blocking the active site of the enzyme, whilst still retaining some regions of crystal packing contacts in the early part of the Ct-extension sequence. These constructs will hopefully be used for crystallisation trials, and if reliably crystallisable to eventually find active site binders

by fragment screening campaign that may be better placed for the treatment of XLP as well as the downstream porphyrias.

7.2 Role of A-SCS in TCA cycle

7.2.1 Significance of study

This study provides the first insights into a novel hetero-octameric conformational state of the A-SCS complex through the combined use of X-ray crystallography, SAXS and cryo-EM. The SAXS experiments have delineated transitions between functional oligomeric states induced by different substrate bound states. Such conformational re-landscaping of the A-SCS complex is hypothesised here to direct opposing biological roles of the two SCS isoforms. This functional partitioning through evolution of alternative structural conformations have new implications in understanding of A-SCS enzymatic mechanism and its dysregulation in human disease.

7.2.2 Limitations of work

Although the structural data here indicate that within the sequence the existence of octameric and dimeric oligomer transitions is possible, one limitation of this data is the lack of supporting evidence for existence of these states in an *in vivo* environment. So far, the assumption until this work has been that A-SCS adopts the same dimeric state and reaction mechanism as the G-SCS complex. Consequently, experiments have not yet been designed in literature or here to prove the existence of oligomers in cells. Future *in vivo* experiments will significantly improve the impact of this work.

In addition, the current low resolution of the cryo-EM data, preventing the modelling of side chain residues, makes the interpretation of the novel oligomeric interfaces mostly theoretical. It will be important to validate these interpreted interfaces with higher resolution structures, or through engineered site-specific mutagenesis studies. High resolution structures will also help confirm the mechanism for substrate induced conformational change at the amino acid level.

So far, a complete range of SCS substrates have also not yet been tested for their effect on A-SCS oligomerisation. In this study, it is possible to attribute the expected oligomeric states to the start and end steps of the reaction. It will be useful to test more ligand combinations in order to ascertain the exact transition point between oligomeric states. It is also not yet known whether initial succinyl-CoA binding or the partial enzyme turnover from step 1 to step 2 in the reaction mechanism is the cause of oligomerisation. Kinetic studies of oligomerisation combined with reaction kinetics will be important to establish exactly where in the reaction mechanism oligomerisation takes place.

Assumptions are also made that these oligomeric states are only in existence in A-SCS based on sequence divergence at the predicted oligomerisation interfaces. Although the complete lack of oligomerisation sequences in G-SCS is compelling evidence, no equivalent succinyl-CoA bound structures of the G-SCS protein exist to confirm that it cannot form higher order oligomers. Considering that the alternative oligomeric states are expected to direct each form of the complex to alternative reported functions *in vivo*, it will be important to confirm this experimentally.

Unfortunately, attempts to isolate the proposed ALAS2 A-SCS complex, based on report of *in vitro* pull-down using recombinant proteins[17], were unsuccessful. The

physiological explanation for the ALAS2-SUCLA2 interaction is suggested to be a possible substrate channelling mechanism in combination with potential ALAS2 activation through direct interaction with the ALAS2 ct-extension[17]. However, questions remain as to how and why such a channelling mechanism would be required. Firstly, the A-SCS complex is implicated in catalysing the forward reaction, converting succinyl-CoA to succinate in the TCA cycle. Haem biosynthesis would require the production of succinyl-CoA via the reverse reaction, catalysed by G-SCS. Additionally, previous work has determined that SCS enzymes are in fact not the major source of succinyl-CoA during erythroid differentiation[47] bringing into question how substrate channelling may explain the physiological role of the interaction. There are many ways to speculate the role of the proposed interaction. One alternative idea could be that the ALAS2 interaction with A-SCS, if it does indeed exist, could serve to inhibit SUCLA2. During erythropoiesis, inhibition of A-SCS would prevent TCA consumption of succinyl-CoA, increasing flux into the haem biosynthesis pathway. It will be important to find ways to investigate these macromolecular interactions further to understand the roles A-SCS plays in the proposed haem biosynthetic metabolon.

7.2.3 Future experiments

To tackle the limitations of the data, follow-up data collections using cryo-EM on optimised sample will be important. The current preliminary data set has only a small number of micrographs and already leads to relatively reasonable resolution. Due to the internal symmetry and stability of the complex and the broad range of orientations captured in this dataset, the path to high resolution in future should only require a larger data collection using the same protein conditions without significant need to re-optimise the sample. This higher resolution data will help validate, at the

residue level, conformational changes that occur within the complex to allow for oligomerisation and identification of the specific residues involved in oligomeric interactions. This data will guide the appropriate selection of residues for site-directed mutagenesis to confirm regions of importance for formation of the different oligomeric states in solution, using the same SAXS methods reported here. In the meantime, study of a SUCLA2 Ct-extension deleted complex offers a good starting point in the immediate future for validation of the octameric structure. In the absence of higher resolution, some of the predicted residues to interact at the ATP grasp lid domain could also be selected for mutation studies to validate this interaction region.

Fluorescence fluctuation spectroscopy would be a useful method to detect the different oligomeric states of A-SCS in human cell lines and to identify cellular conditions that impact on these functional states. To achieve this, transient expression vectors containing fluorescently tagged A-SCS coding sequences in inducible expression systems could be cloned. Simple confocal microscopy experiments that follow published imaging protocols[137] would then be performed to investigate quaternary structure in live cells.

Finally, it will be interesting to perform follow up SAXS experiments testing the effect of oligomerisation on different combinations of A-SCS substrates. So far optimising a sample that allows binding to ADP without protein precipitation will be important. Also, intermediate ligand combinations such as succinate and CoA combined are yet to be tested. These states should be investigated using the same SAXS methods reported here and in addition it would be exciting to use a novel photo spectroscopy method called interferometric scattering mass spectrometry (Iscam)[138]. This allows very accurate measurement of quaternary structure *in vitro* at physiological relevant

protein concentrations, allowing rapid screening of conditions that affect oligomerisation.

7.3 Complexes in propionate Catabolism

7.3.1 Significance of study

Using a combination of SAXS and cryo-EM, this study has identified the existence of heterogenous filamentous forms of the MMAA-MMUT complex. Such structures provide an insight into the exquisite cofactor reactivation process that has evolved to deliver functional Cbl cofactor to the MMUT enzyme.

More subtle conformational changes observed in the PCC complex have also been isolated using cryo-EM. Future high-resolution determination of these states will provide unprecedented details of underrepresented, ATP specific, conformations in the acyl-CoA carboxylase family.

7.3.2 Limitations of data

Although the data here points towards the potential for the existence of filamentous structures as one plausible state in several possible functional conformations of the MMAA-MMUT complex, the data is significantly limited. The direct visualisation of filaments using cryo-EM provides compelling evidence for the interpretations of filaments in the SAXS data however, so far filaments are only visualised in low resolution at the micrograph level. High resolution cryo-EM collections are required to provide stronger evidence for the existence of this modelled structure. The current model for induced filament formation is also heavily reliant on comparison with homology models. The mode of interaction has clearly evolved to be more complicated than the IcmF fusion protein and the nature of the interaction interface between the MMAA-MMUT may no longer reflect the IcmF intra-action.

Experimental data that allows more accurate modelling of the complex at residue level is a must in order to substantiate predictions made in this study. Like the A-SCS study, another limitation of this data is the lack of supporting evidence for the existence of these oligomeric states in an *in vivo* setting. It will be important to help understand the proposed MMUT sequestering mechanism *in vivo.*

For the PCC complex no modelling has yet been performed into the high-resolution structure, making interpretations at this point preliminary. 3D refinement of the symmetry expanded data is still required to extract high resolution information that will allow the appropriate characterisation of the alternative conformational states, adopted by the PCCA BCCP domain, that are present within the dataset. This step will provide insights into the intra subunit interactions that allow for carboxylation of the biotin group. Such information will be very useful in the interpretation of disease specific mutations that occur in the PCCA subunit which likely affect catalysis of the first half reaction and provide insights into the catalytic mechanism of biotin carboxylation by PCCA.

7.3.3 Future experiments

The next step in this project will be to optimise the oligomeric forms of the MMAA-MMUT complex for high resolution structure determination using cryo-EM. GraFix will continue to be used and higher molecular weight species identified in the GraFix SDS-PAGE gel (Figure 6-12) could be tested for data collections. In addition, buffer screens and protein concentration optimisation could be performed on the MMAA-MMUT complex samples without crosslinking, to improve monodispersity and signal to noise ratios in the non-crosslinked samples.

Follow up SAXS studies using a range of protein concentrations and proposed ligand states will also provide very useful insights and validation into the oligomeric transitions that are proposed in previous studies[15]. It will be interesting to study different nucleotide combinations and varying cofactor states at the upper axial ligand to ascertain whether these affect the oligomeric state of the complex. Alongside these SAXS studies (Iscam) (Gavin Young, et al. 2018) could also be used here to gain accurate measurement of quaternary structure *in vitro* at physiological relevant protein concentrations, allowing the screening of many ligand conditions that affect oligomerisation.

Fluorescence fluctuation spectroscopy would be a useful method to first detect different oligomeric states of MMUT-MMAA complex in human cell lines. It would also be interesting to identify cellular conditions that impact on these functional states using the same method. To achieve this, it would useful to perform these experiments in cells where Cbl cofactor levels and GTP/GDP ratios can be manipulated. A crude manipulation approach could be to simply add excess Cbl cofactor or specific nucleotide to the cell media or exclude Cbl cofactor from the media and probe effects of filament formation/abrogation. Alternatively, these experiments could be performed in cell lines that block Cbl cofactor entry into the mitochondria or cell through knock down or knock out of the membrane transporters responsible for trafficking Cbl.

The cryo-EM data for the PCC complex requires rounds of model building, refinement and model validation to enhance the interpretation of data. Alongside modelling the human amino acid sequences into the density, different strategies[135] can be adopted for refining the alternative conformations seen in the PCCA subunit through symmetry expansion of the data. In the first instance, focused refinement

using a mask for the PCCA subunit that ignores data from the remaining complex will be performed on the homogenous pool of PCCA conformations that have been classified, after symmetry expansion, into a single group of particles using 3D classification. If this does not allow high resolution reconstruction of density, signal subtraction can be performed that removes completely the extra density from the rest of the complex to reduce the background noise in the refinement.

A drawback to refining the PCCA subunit alone in the above strategies is that PCCA is just under the recommended size conditions for successful focused refinement[135]. To resolve this, the mask could therefore be extended to include a PCCB subunit in the refinement. This may unfortunately reintroduce some conformational heterogeneity back into the data if there is an element of continuous conformational states between the PCCA and PCCB subunits within this subset of particles. It may be a requirement to undergo further 3D classifications with this larger set of particles, to once again isolate a homogenous pool of particles that is large enough for focused refinement. Such tasks are computationally demanding and will likely take weeks to months of data processing to find the best solution.

7.4 Future perspective

I have discussed in the short term the need to confirm many of the specific mechanistic oligomeric states, witnessed using structural biology techniques on *in vitro* isolated proteins and their binary complexes, in the cellular environment to ascertain their physiological relevance. Conversely, over the last few years there has been a wealth of cell biology and proteomics data promising the existence of larger cellular metabolons involving many different proteins from multiple pathways. Validation and more detailed structure function analysis of such metabolons *in vitro*

has not yet seen the same amount of prolific success. The challenge in the field will be to convert these promised complexes, such as the haem metabolon and mitochondrial B12 metabolon discussed here, into real structural data to establish their composition, and exact molecular functions for the interactions.

Difficulties in reconstituting these metabolons are perhaps explained by their transient low affinity interactions, unknown components of complex, the sheer number of implicated genes posing a problem for *in vitr*o isolation, and the requirement for specific spatial temporal cellular environments for such complexes to form. Advances in cloning systems that enable multi gene co-expressions in recombinant hosts such as MultiBac[139] and InteBac[140] will help to isolate many large stable subcomplexes. However, to overcome the dynamic nature of these metabolons when attempting to study their composition, it will be interesting to follow advancements in technologies which will allow the fields of cell biology and structural biology to cross. Future endeavours into *in situ* structural biology such as cryo-electron tomography will enable direct visualisation of the proposed complexes in their native cellular or organelle environments as opposed to reconstructed stable complexes that are studied by single particle cryo-EM. This may provide insights into dynamic changes in complex composition under different cellular conditions. Advances in endogenous capture of complexes, for example introducing affinity capture fusion tags by knock-in using Crispr Cas9 gene editing, will allow isolation from a range of cellular locations and cell types for their study via single particle cryo-EM. Such targeted isolation would be very useful for erythroid specific protein-protein interaction such as A-SCS and ALAS2 that may only occur in significant quantities during erythropoiesis and could even be used to pull out more components of the haem metabolon. Finally, improvements in cellular crosslinking

studies coupled with LC-MS/MS technologies that can map directly peptide interfaces involved in specific interactions will provide stronger evidence for the endless proteomics studies claiming the many protein-protein interactions that exist in the cell.

It is my hope that the study of individuals proteins and their stable subcomplexes using now a full range of modern structural biology techniques, as shown in this thesis, will continue to uncover surprisingly complex mechanisms of the binary protein-protein interactions that were previously believed to be well characterised. These tractable studies will help isolate and characterise many more unknown conformational states of sub-components predicted to participate in speculative cellular metabolons. This information alongside the better determinisation of the role of protein-protein interactions in the cellular environment will form the building blocks for successful study of the more elusive large macromolecular protein complexes. Such studies will provide unprecedented detail on the complex catalytic pathway regulatory mechanisms, provide insights into the multi functionality of enzymes and allow for better understanding of the links between gene mutation and pathophysiology of metabolic disease.

Appendix 1: My publications

1. **Henry J. Bailey***, Gustavo A. Bezerra*, Jason R. Marcero*, Siladitya Padhi*, William R. Foster, Elzbieta Rembeza, Arijit Roy, David F. Bishop, Robert J. Desnick, Gopalakrishnan Bulusu, Harry A. Dailey, Wyatt W. Yue. Human aminolevulinate synthase structure reveals a eukaryotic-specific autoinhibitory loop regulating substrate binding and product release. **Nat Commun.** 2020 Jun 4;11(1):2813.

2. Gustavo A. Bezerra*, William R. Foster*, **Henry J. Bailey***, Kevin G. Hicks, Sven W. Sauer, Bianca Dimitrov, Thomas J. McCorvie, Jürgen G. Okun, Jared Rutter, Stefan Kölker, Wyatt W. Yue. Crystal structure and interaction studies of human DHTKD1 provide insight into a mitochondrial megacomplex in lysine catabolism. **IUCrJ.** 2020 Jun 10;7(Pt 4):693-706.

3. Hassler M, Shaltiel IA, Kschonsak M, Simon B, Merkel F, Thärichen L, **Bailey HJ**, Macošek J, Bravo S, Metz J, Hennig J, Haering CH**.** Structural Basis of an Asymmetric Condensin ATPase Cycle. **Mol Cell.** 2019 Jun 20;74(6):1175-1188.e9.

4. Fox NG, Yu X, Feng X, **Bailey HJ**, Martelli A, Nabhan JF, Strain-Damerell C, Bulawa C, Yue WW, Han S. Structure of the human frataxin-bound iron-sulfur cluster assembly complex provides insight into its activation mechanism**. Nat Commun.** 2019 May 17;10(1):2210.

5. Heuberger K*, **Bailey HJ***, Burda P, Chaikuad A, Krysztofinska E, Suormala T, Bürer C, Lutz S, Fowler B, Froese DS, Yue WW, Baumgartner MR. Genetic, structural, and functional analysis of pathogenic variations causing methylmalonyl-CoA epimerase deficiency. **Biochim Biophys Acta Mol Basis Dis.** 2019 Jun 1;1865(6):1265-1272.

6. Bilyard MK, **Bailey HJ**, Raich L, Gafitescu MA, Machida T, Iglésias-Fernández J, Lee SS, Spicer CD, Rovira C, Yue WW, Davis BG. Palladium-mediated enzyme activation suggests multiphase initiation of glycogenesis. **Nature.** 2018 Nov;563(7730):235-240.

7. Oerum S, Roovers M, Rambo RP, Kopec J, **Bailey HJ**, Fitzpatrick F, Newman JA, Newman WG, Amberger A, Zschocke J, Droogmans L, Oppermann U, Yue WW. Structural insight into the human mitochondrial tRNA purine N1-methyltransferase and ribonuclease P complexes**. J Biol Chem.** 2018 Aug 17;293(33):12862-12876.

8. Oerum S, Roovers M, Leichsenring M, Acquaviva-Bourdain C, Beermann F, Gemperle-Britschgi C, Fouilhoux A, Korwitz-Reichelt A, **Bailey HJ**, Droogmans L,Oppermann U, Sass JO, Yue WW. Novel patient missense mutations in the HSD17B10 gene affect dehydrogenase and mitochondrial tRNA modification functions of the encoded protein**. Biochim Biophys Acta Mol Basis Dis.** 2017 Dec;1863(12):3294-3302.

9. Watkins DW, Jenkins JMX, Grayson KJ, Wood N, Steventon JW, Le Vay KK, Goodwin MI, Mullen AS, **Bailey HJ**, Crump MP, MacMillan F, Mulholland AJ, Cameron G, Sessions RB, Mann S, Anderson JLR. Construction and in vivo assembly of a catalytically proficient and hyperthermostable de novo enzyme. **Nat Commun.** 2017 Aug 25;8(1):358.

*denotes equal contribution of authors

Appendix 2: PDB validation reports

6G4Q: Structure of human ADP-forming succinyl-CoA ligase complex SUCLG1-SUCLA2

https://www.ebi.ac.uk/pdbe/entry-files/download/6g4q_full_validation.pdf

6HRH: Structure of human erythroid-specific 5'-aminolevulinate synthase, ALAS2

https://files.rcsb.org/pub/pdb/validation_reports/hr/6hrh/6hrh_full_validation.pdf

References

1. Biochemistry (3rd edn). C.K. Mathews, K.E. Van Holde, K.G. Ahern (eds). *Int Microbiol - Off J Spanish Soc Microbiol*. 2001. doi:10.2436/im.v4i3.9330

2. Martínez-Reyes I, Chandel NS. Mitochondrial TCA cycle metabolites control physiology and disease. *Nat Commun*. 2020. doi:10.1038/s41467-019-13668-3

3. Kim HJ. Haem Structure and Function. *eLS*. April 2018:1-9. doi:doi:10.1002/9780470015902.a0000605.pub2

4. GIBSON KD, LAVER WG, NEUBERGER A. Initial stages in the biosynthesis of porphyrins. 2. The formation of delta-aminolaevulic acid from glycine and succinyl-coenzyme A by particles from chicken erythrocytes. *Biochem J*. 1958. doi:10.1042/bj0700071

5. Akhtar M, Abboud MM, Barnard G, Jordan P, Zaman Z. Mechanism and stereochemistry of enzymic reactions involved in porphyrin biosynthesis. *Philos Trans R Soc Lond B Biol Sci*. 1976. doi:10.1098/rstb.1976.0005

6. Dailey HA, Meissner PN. Erythroid heme biosynthesis and its disorders. *Cold*

Spring Harb Perspect Med. 2013. doi:10.1101/cshperspect.a011676

7. Van Schaftingen E, Veiga-da-Cunha M, Linster CL. Enzyme complexity in intermediary metabolism. *J Inherit Metab Dis*. 2015. doi:10.1007/s10545-015-9821-0

8. Zhang Y, Fernie AR. Stable and temporary enzyme complexes and metabolons involved in energy and redox metabolism. *Antioxid Redox Signal*. 2020. doi:10.1089/ars.2019.7981

9. Lamhonwah AM, Barankiewicz TJ, Willard HF, Mahuran DJ, Quan F, Gravel RA. Isolation of cDNA clones coding for the α and β chains of human propionyl-CoA carboxylase: Chromosomal assignments and DNA polymorphisms associated with PCCA and PCCB genes. *Proc Natl Acad Sci U S A*. 1986. doi:10.1073/pnas.83.13.4864

10. Lau EP, Cochran BC, Munson L, Fall RR. Bovine kidney 3-methylcrotonyl-CoA and propionyl-CoA carboxylases: each enzyme contains nonidentical subunits. *Proc Natl Acad Sci U S A*. 1979. doi:10.1073/pnas.76.1.214

11. Weitzman PDJ, Jenkins T, Else AJ, Holt RA. Occurrence of two distinct succinate thiokinases in animal tissues. *FEBS Lett*. 1986. doi:10.1016/0014-5793(86)81223-6

12. Johnson JD, Muhonen WW, Lambeth DO. Characterization of the ATP- and GTP-specific succinyl-CoA synthetases in pigeon: The enzymes incorporate the same α-subunit. *J Biol Chem*. 1998. doi:10.1074/jbc.273.42.27573

13. Korotkova N, Lidstrom ME. MeaB Is A Component of the Methylmalonyl-CoA

Mutase Complex Required for Protection of the Enzyme from Inactivation. *J Biol Chem*. 2004. doi:10.1074/jbc.M312852200

14. Froese DS, Kochan G, Muniz JRC, et al. Structures of the human GTPase MMAA and vitamin B12-dependent methylmalonyl-CoA mutase and insight into their complex formation. *J Biol Chem*. 2010. doi:10.1074/jbc.M110.177717

15. Ruetz M, Campanello GC, McDevitt L, et al. Allosteric Regulation of Oligomerization by a B12 Trafficking G-Protein Is Corrupted in Methylmalonic Aciduria. *Cell Chem Biol*. 2019. doi:10.1016/j.chembiol.2019.03.014

16. Furuyama K, Sassa S. Interaction between succinyl CoA synthetase and the heme-biosynthetic enzyme ALAS-E is disrupted in sideroblastic anemia. *J Clin Invest*. 2000. doi:10.1172/JCI6816

17. Bishop DF, Tchaikovskii V, Hoffbrand AV, Fraser ME, Margolis S. X-linked sideroblastic anemia due to carboxyl-terminal ALAS2 mutations that cause loss of binding to the β-subunit of succinyl-CoA synthetase (SUCLA2). *J Biol Chem*. 2012. doi:10.1074/jbc.M111.306423

18. Srere PA. The metabolon. *Trends Biochem Sci*. 1985. doi:10.1016/0968-0004(85)90266-X

19. Zhang Y, Beard KFM, Swart C, et al. Protein-protein interactions and metabolite channelling in the plant tricarboxylic acid cycle. *Nat Commun*. 2017. doi:10.1038/ncomms15212

20. Plessl T, Bürer C, Lutz S, Yue WW, Baumgartner MR, Froese DS. Protein destabilization and loss of protein-protein interaction are fundamental

mechanisms in cblA-type methylmalonic aciduria. *Hum Mutat*. 2017. doi:10.1002/humu.23251

21. Piel RB, Dailey HA, Medlock AE. The mitochondrial heme metabolon: Insights into the complex(ity) of heme synthesis and distribution. *Mol Genet Metab*. 2019. doi:10.1016/j.ymgme.2019.01.006

22. Bishop DF, Henderson AS, Astrin KH. Human δ-aminolevulinate synthase: Assignment of the housekeeping gene to 3p21 and the erythroid-specific gene to the X chromosome. *Genomics*. 1990. doi:10.1016/0888-7543(90)90542-3

23. Riddle RD, Yamamoto M, Engel JD. Expression of δ-aminolevulinate synthase in avian cells: Separate genes encode erythroid-specific and nonspecific isozymes. *Proc Natl Acad Sci U S A*. 1989. doi:10.1073/pnas.86.3.792

24. Bonkowsky HL, Sinclair PR, Sinclair JF. Hepatic heme metabolism and its control. *Yale J Biol Med*. 1979.

25. Ogun AS, Valentine M. *Biochemistry, Heme Synthesis*.; 2018.

26. Ferreira GC. Heme Synthesis. In: *Encyclopedia of Biological Chemistry: Second Edition*. ; 2013. doi:10.1016/B978-0-12-378630-2.00145-6

27. Besur S, Hou W, Schmeltzer P, Bonkovsky HL. Clinically important features of porphyrin and heme metabolism and the porphyrias. *Metabolites*. 2014. doi:10.3390/metabo4040977

28. Hunter GA, Zhang J, Ferreira GC. Transient kinetic studies support refinements to the chemical and kinetic mechanisms of aminolevulinate synthase. *J Biol Chem*. 2007. doi:10.1074/jbc.M609330200

29. Fratz EJ, Clayton J, Hunter GA, et al. Human Erythroid 5-Aminolevulinate Synthase Mutations Associated with X-Linked Protoporphyria Disrupt the Conformational Equilibrium and Enhance Product Release. *Biochemistry*. 2015. doi:10.1021/acs.biochem.5b00407

30. Stojanovski BM, Hunter GA, Jahn M, Jahn D, Ferreira GC. Unstable reaction intermediates and hysteresis during the catalytic cycle of 5-aminolevulinate synthase: Implications from using pseudo and alternate substrates and a promiscuous enzyme variant. *J Biol Chem*. 2014. doi:10.1074/jbc.M114.574731

31. Hunter GA, Ferreira GC. Pre-steady-state reaction of 5-aminolevulinate synthase: Evidence for a rate-determining product release. *J Biol Chem*. 1999. doi:10.1074/jbc.274.18.12222

32. Zhang J, Ferreira GC. Transient state kinetic investigation of 5-aminolevulinate synthase reaction mechanism. *J Biol Chem*. 2002. doi:10.1074/jbc.M203584200

33. Whatley SD, Ducamp S, Gouya L, et al. C-Terminal Deletions in the ALAS2 Gene Lead to Gain of Function and Cause X-linked Dominant Protoporphyria without Anemia or Iron Overload. *Am J Hum Genet*. 2008. doi:10.1016/j.ajhg.2008.08.003

34. Ducamp S, Schneider-Yin X, De Rooij F, et al. Molecular and functional analysis of the C-terminal region of human erythroid-specific 5-aminolevulinic synthase associated with X-linked dominant protoporphyria (XLDPP). *Hum Mol Genet*. 2013. doi:10.1093/hmg/dds531

35. Balwani M, Doheny D, Bishop DF, et al. Loss-of-function ferrochelatase and gain-of-function erythroid-specific 5-aminolevulinate synthase mutations causing erythropoietic protoporphyria and X-linked protoporphyria in North American patients reveal novel mutations and a high prevalence of X-lin. *Mol Med*. 2013. doi:10.2119/molmed.2012.00340

36. Tchaikovskii V, Desnick RJ, Bishop DF. Molecular expression, characterization and mechanism of ALAS2 gain-of-function mutants. *Mol Med*. 2019. doi:10.1186/s10020-019-0070-9

37. Bishop DF, Tchaikovskii V, Nazarenko I, Desnick RJ. Molecular expression and characterization of erythroid-specific 5-aminolevulinate synthase gain-of-function mutations causing X-linked protoporphyria. *Mol Med*. 2013. doi:10.2119/molmed.2013.00003

38. Harigae H, Furuyama K. Hereditary sideroblastic anemia: Pathophysiology and gene mutations. *Int J Hematol*. 2010. doi:10.1007/s12185-010-0688-4

39. Cotter PD, Baumann M, Bishop DF. Enzymatic defect in “X-linked” sideroblastic anemia: Molecular evidence for erythroid δ-aminolevulinate synthase deficiency. *Proc Natl Acad Sci U S A*. 1992. doi:10.1073/pnas.89.9.4028

40. Ducamp S, Kannengiesser C, Touati M, et al. Sideroblastic anemia: Molecular analysis of the ALAS2 gene in a series of 29 probands and functional studies of 10 missense mutations. *Hum Mutat*. 2011. doi:10.1002/humu.21455

41. Bottomley SS, May BK, Cox TC, Cotter PD, Bishop DF. Molecular defects of erythroid 5-aminolevulinate synthase in X-linked sideroblastic anemia. *J*

Bioenerg Biomembr. 1995. doi:10.1007/BF02110031

42. Manceau H, Gouya L, Puy H. Acute hepatic and erythropoietic porphyrias: From ALA synthases 1 and 2 to new molecular bases and treatments. *Curr Opin Hematol*. 2017. doi:10.1097/MOH.0000000000000330

43. Balwani M, Desnick RJ. The porphyrias: Advances in diagnosis and treatment. *Blood*. 2012. doi:10.1182/blood-2012-05-423186

44. Yue WW, Mackinnon S, Bezerra GA. Substrate reduction therapy for inborn errors of metabolism. *Emerg Top Life Sci*. 2019. doi:10.1042/etls20180058

45. Yasuda M, Gan L, Chen B, et al. RNAi-mediated silencing of hepatic Alas1 effectively prevents and treats the induced acute attacks in acute intermittent porphyria mice. *Proc Natl Acad Sci U S A*. 2014. doi:10.1073/pnas.1406228111

46. Chan A, Liebow A, Yasuda M, et al. Preclinical Development of a Subcutaneous ALAS1 RNAi Therapeutic for Treatment of Hepatic Porphyrias Using Circulating RNA Quantification. *Mol Ther - Nucleic Acids*. 2015. doi:10.1038/mtna.2015.36

47. Burch JS, Marcero JR, Maschek JA, et al. Glutamine via a-ketoglutarate dehydrogenase provides succinyl-CoA for heme synthesis during erythropoiesis. *Blood*. 2018. doi:10.1182/blood-2018-01-829036

48. Bridger WA. 18. Succinyl-Coa Synthetase. *Enzymes*. 1974. doi:10.1016/S1874-6047(08)60150-X

49. Johnson JD, Mehus JG, Tews K, Milavetz BI, Lambeth DO. Genetic evidence

for the expression of ATP- and GTP-specific succinyl- CoA synthetases in multicellular eucaryotes. *J Biol Chem*. 1998. doi:10.1074/jbc.273.42.27580

50. Weinberg JM, Venkatachalam MA, Roeser NF, Nissim I. Mitochondrial dysfunction during hypoxia/reoxygenation and its correction by anaerobic metabolism of citric acid cycle intermediates. *Proc Natl Acad Sci U S A*. 2000. doi:10.1073/pnas.97.6.2826

51. Ottaway JH, McClellan JA, Saunderson CL. Succinic thiokinase and metabolic control. *Int J Biochem*. 1981. doi:10.1016/0020-711X(81)90111-7

52. Jenkins TM, Weitzman PDJ. Distinct physiological roles of animal succinate thiokinases Association of guanine nucleotide-linked succinate thiokinase with ketone body utilization. *FEBS Lett*. 1986. doi:10.1016/0014-5793(86)80900-0

53. Labbe RE, Kurumada T, Onisawa J. The role of succinyl-CoA synthetase in the control of heme biosynthesis. *BBA - Gen Subj*. 1965. doi:10.1016/0304-4165(65)90050-4

54. Lowe DM, Tubbs PK. Succinylation and inactivation of 3-hydroxy-3-methylglutaryl-CoA synthase by succinyl-CoA and its possible relevance to the control of ketogenesis. *Biochem J*. 1985. doi:10.1042/bj2320037

55. Siess EA, Fahimi FM, Wieland OH. Decrease by glucagon in hepatic succinyl-CoA. *Top Catal*. 1980. doi:10.1016/0006-291X(80)90725-1

56. Zhang Z, Tan M, Xie Z, Dai L, Chen Y, Zhao Y. Identification of lysine succinylation as a new post-translational modification. *Nat Chem Biol*. 2011. doi:10.1038/nchembio.495

57. Park J, Chen Y, Tishkoff DX, et al. SIRT5-Mediated Lysine Desuccinylation Impacts Diverse Metabolic Pathways. *Mol Cell*. 2013. doi:10.1016/j.molcel.2013.06.001

58. Fraser ME, James MNG, Bridger WA, Wolodko WT. Phosphorylated and dephosphorylated structures of pig heart, GTP-specific succinyl-CoA synthetase. *J Mol Biol*. 2000. doi:10.1006/jmbi.2000.3807

59. Smith CM, Bryla J, Williamson JR. Regulation of mitochondrial α ketoglutarate metabolism by product inhibition a α ketoglutarate dehydrogenase. *J Biol Chem*. 1974.

60. Carrozzo R, Verrigni D, Rasmussen M, et al. Succinate-CoA ligase deficiency due to mutations in SUCLA2 and SUCLG1: phenotype and genotype correlations in 71 patients. *J Inherit Metab Dis*. 2016. doi:10.1007/s10545-015-9894-9

61. Van Hove JLK, Saenz MS, Thomas JA, et al. Succinyl-coa ligase deficiency: A mitochondrial hepatoencephalomyopathy. *Pediatr Res*. 2010. doi:10.1203/PDR.0b013e3181e5c3a4

62. Huang X, Bedoyan JK, Demirbas D, et al. Succinyl-CoA synthetase (SUCLA2) deficiency in two siblings with impaired activity of other mitochondrial oxidative enzymes in skeletal muscle without mitochondrial DNA depletion. *Mol Genet Metab*. 2017;120(3):213-222. doi:10.1016/j.ymgme.2016.11.005

63. Carrozzo R, Dionisi-Vici C, Steuerwald U, et al. SUCLA2 mutations are associated with mild methylmalonic aciduria, Leigh-like encephalomyopathy, dystonia and deafness. *Brain*. 2007. doi:10.1093/brain/awl389

64. Zhao Y, Tian J, Sui S, et al. Loss of succinyl-CoA synthase ADP-forming β subunit disrupts mtDNA stability and mitochondrial dynamics in neurons. *Sci Rep*. 2017. doi:10.1038/s41598-017-05168-5

65. Kadrmas EF, Ray PD, Lambeth DO. Apparent ATP-linked succinate thiokinase activity and its relation to nucleoside diphosphate kinase in mitochondrial matrix preparations from rabbit. *BBA - Gen Subj*. 1991. doi:10.1016/0304-4165(91)90083-S

66. Kavanaugh-Black A, Connolly DM, Chugani SA, Chakrabarty AM. Characterization of nucleoside-diphosphate kinase from Pseudomonas aeruginosa: Complex formation with succinyl-CoA synthetase. *Proc Natl Acad Sci U S A*. 1994. doi:10.1073/pnas.91.13.5883

67. Kowluru A, Tannous M, Chen HQ. Localization and characterization of the mitochondrial isoform of the nucleoside diphosphate kinase in the pancreatic β cell: Evidence for its complexation with mitochondrial succinyl-CoA synthetase. *Arch Biochem Biophys*. 2002. doi:10.1006/abbi.2001.2710

68. Frenkel EP, Kitchens RL. Intracellular Localization of Hepatic Propionyl-CoA Carboxylase and Methylmalonyl-CoA Mutase in Humans and Normal and Vitamin B12 Deficient Rats. *Br J Haematol*. 1975. doi:10.1111/j.1365-2141.1975.tb00885.x

69. Wah AMLH, Lam KF, Tsui F. Assignment of the α and β chains of human propionyl-CoA carboxylase to genetic complementation groups. *Am J Hum Genet*. 1983.

70. Tong L. Structure and function of biotin-dependent carboxylases. *Cell Mol Life*

Sci. 2013. doi:10.1007/s00018-012-1096-0

71. Huang CS, Sadre-Bazzaz K, Shen Y, Deng B, Zhou ZH, Tong L. Crystal structure of the α6β6 holoenzyme of propionyl-coenzyme A carboxylase. *Nature*. 2010. doi:10.1038/nature09302

72. Banerjee R. Radical carbon skeleton rearrangements: Catalysis by coenzyme B12-dependent mutases. *Chem Rev*. 2003. doi:10.1021/cr0204395

73. Padovani D, Banerjee R. A G-protein editor gates coenzyme B12 loading and is corrupted in methylmalonic aciduria. *Proc Natl Acad Sci U S A*. 2009. doi:10.1073/pnas.0908106106

74. Banerjee R, Gherasim C, Padovani D. The tinker, tailor, soldier in intracellular B12 trafficking. *Curr Opin Chem Biol*. 2009. doi:10.1016/j.cbpa.2009.07.007

75. Gherasim C, Lofgren M, Banerjee R. Navigating the B12 road: Assimilation, delivery, and disorders of cobalamin. *J Biol Chem*. 2013. doi:10.1074/jbc.R113.458810

76. Dobson CM, Wai T, Leclerc D, et al. Identification of the gene responsible for the cblB complementation group of vitamin B12-dependent methylmalonic aciduria. *Hum Mol Genet*. 2002. doi:10.1093/hmg/11.26.3361

77. Leal NA, Park SD, Kima PE, Bobik TA. Identification of the human and bovine ATP:cob(I)alamin adenosyltransferase cDNAs based on complementation of a bacterial mutant. *J Biol Chem*. 2003. doi:10.1074/jbc.M212739200

78. Huemer M, Baumgartner MR. The clinical presentation of cobalamin-related disorders: From acquired deficiencies to inborn errors of absorption and

intracellular pathways. *J Inherit Metab Dis*. 2019. doi:10.1002/jimd.12012

79. Wongkittichote P, Ah Mew N, Chapman KA. Propionyl-CoA carboxylase – A review. *Mol Genet Metab*. 2017. doi:10.1016/j.ymgme.2017.10.002

80. Dobson CM, Wai T, Leclerc D, et al. Identification of the gene responsible for the cblA complementation group of vitamin B12-responsive methylmalonic acidemia based on analysis of prokaryotic gene arrangements. *Proc Natl Acad Sci U S A*. 2002. doi:10.1073/pnas.242614799

81. Ledley FD, Lumetta M, Nguyen PN, Kolhouse JF, Allen RH. Molecular cloning of L-methylmalonyl-CoA mutase: Gene transfer and analysis of mut cell lines. *Proc Natl Acad Sci U S A*. 1988. doi:10.1073/pnas.85.10.3518

82. Winter G. Xia2: An expert system for macromolecular crystallography data reduction. *J Appl Crystallogr*. 2010. doi:10.1107/S0021889809045701

83. Tickle IJ, Flensburg C, Keller P, et al. STARANISO. *Cambridge, United Kingdom Glob Phasing Ltd*. 2018.

84. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. *J Appl Crystallogr*. 2007. doi:10.1107/S0021889807021206

85. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. *Acta Crystallogr Sect D Biol Crystallogr*. 2010. doi:10.1107/S0907444910007493

86. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr Sect D Biol*

Crystallogr. 1997. doi:10.1107/S0907444996012255

87. Afonine P V., Grosse-Kunstleve RW, Echols N, et al. Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr Sect D Biol Crystallogr*. 2012. doi:10.1107/S0907444912001308

88. Förster S, Apostol L, Bras W. Scatter: Software for the analysis of nano-and mesoscale small-angle scattering. *J Appl Crystallogr*. 2010. doi:10.1107/S0021889810008289

89. Petoukhov M V., Franke D, Shkumatov A V., et al. New developments in the ATSAS program package for small-angle scattering data analysis. *J Appl Crystallogr*. 2012. doi:10.1107/S0021889812007662

90. Scheres SHW. RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *J Struct Biol*. 2012. doi:10.1016/j.jsb.2012.09.006

91. Rohou A, Grigorieff N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J Struct Biol*. 2015. doi:10.1016/j.jsb.2015.08.008

92. Astner I, Schulze JO, Van Den Heuvel J, Jahn D, Schubert WD, Heinz DW. Crystal structure of 5-aminolevulinate synthase, the first enzyme of heme biosynthesis, and its link to XLSA in humans. *EMBO J*. 2005. doi:10.1038/sj.emboj.7600792

93. Munakata H, Yamagami T, Nagai T, Yamamoto M, Hayashi N. Purification and structure of rat erythroid-specific δ-aminolevulinate synthase. *J Biochem*. 1993. doi:10.1093/oxfordjournals.jbchem.a124123

94. Schoenhaut DS, Curtis PJ. Nucleotide sequence of mouse 5-aminolevulinic

acid synthase cDNA and expression of its gene in hepatic and erythroid tissues. *Gene*. 1986. doi:10.1016/0378-1119(86)90351-3

95. Lathrop JT, Timko MP. Regulation by heme of mitochondrial protein transport through a conserved amino acid motif. *Science (80-)*. 1993. doi:10.1126/science.8424176

96. Munakata H, Sun JY, Yoshida K, et al. Role of the heme regulatory motif in the heme-mediated inhibition of mitochondrial import of 5-aminolevulinate synthase. *J Biochem*. 2004. doi:10.1093/jb/mvh112

97. Brown BL, Kardon JR, Sauer RT, Baker TA. Structure of the Mitochondrial Aminolevulinic Acid Synthase, a Key Heme Biosynthetic Enzyme. *Structure*. 2018. doi:10.1016/j.str.2018.02.012

98. Yadav PK, Xie P, Banerjee R. Allosteric communication between the pyridoxal 5'-phosphate (PLP) and heme sites in the H2S generator human cystathionine β-synthase. *J Biol Chem*. 2012. doi:10.1074/jbc.M112.414706

99. Till M, Robson A, Byrne MJ, et al. Improving the success rate of protein crystallization by random microseed matrix screening. *J Vis Exp*. 2013. doi:10.3791/50548

100. Rambo RP, Tainer JA. Accurate assessment of mass, models and resolution by small-angle scattering. *Nature*. 2013. doi:10.1038/nature12070

101. Lendrihas T, Hunter GA, Ferreira GC. Targeting the active site gate to yield hyperactive variants of 5-aminolevulinate synthase. *J Biol Chem*. 2010. doi:10.1074/jbc.M109.074237

102. Cox OB, Krojer T, Collins P, et al. A poised fragment library enables rapid synthetic expansion yielding the first reported inhibitors of PHIP(2), an atypical bromodomain. *Chem Sci*. 2016. doi:10.1039/c5sc03115j

103. Krojer T, Talon R, Pearce N, et al. The XChemExplorer graphical workflow tool for routine or large-scale protein-ligand structure determination. *Acta Crystallogr Sect D Struct Biol*. 2017. doi:10.1107/S2059798316020234

104. Davies TG, Wixted WE, Coyle JE, et al. Monoacidic Inhibitors of the Kelch-like ECH-Associated Protein 1: Nuclear Factor Erythroid 2-Related Factor 2 (KEAP1:NRF2) Protein-Protein Interaction with High Cell Potency Identified by Fragment-Based Discovery. *J Med Chem*. 2016. doi:10.1021/acs.jmedchem.6b00228

105. Wood DJ, Lopez-Fernandez JD, Knight LE, et al. FragLites - Minimal, Halogenated Fragments Displaying Pharmacophore Doublets. An Efficient Approach to Druggability Assessment and Hit Generation. *J Med Chem*. 2019. doi:10.1021/acs.jmedchem.9b00304

106. Wojdyr M, Keegan R, Winter G, Ashton A, Lebedev A, Krissinel E. A pipeline for the rapid generation of difference maps from protein crystals. *Acta Crystallogr Sect A Found Adv*. 2014. doi:10.1107/s2053273314085520

107. Pearce NM, Bradley AR, Krojer T, Marsden BD, Deane CM, Delft F Von. Partial-occupancy binders identified by the Pan-Dataset Density Analysis method offer new chemical opportunities and reveal cryptic binding sites. *Struct Dyn*. 2017. doi:10.1063/1.4974176

108. Shoolingin-Jordan PM, Al-Daihan S, Alexeev D, et al. 5-Aminolevulinic acid

synthase: Mechanism, mutations and medicine. In: *Biochimica et Biophysica Acta - Proteins and Proteomics*. ; 2003. doi:10.1016/S1570-9639(03)00095-5

109. Peoc'h K, Nicolas G, Schmitt C, et al. Regulation and tissue-specific expression of δ-aminolevulinic acid synthases in non-syndromic sideroblastic anemias and porphyrias. *Mol Genet Metab*. 2019. doi:10.1016/j.ymgme.2019.01.015

110. Fraser ME, James MNG, Bridger WA, Wolodko WT. A detailed structural description of Escherichia coli succinyl-CoA synthetase. *J Mol Biol*. 1999. doi:10.1006/jmbi.1998.2324

111. Wolodko WT, Fraser ME, James MNG, Bridger WA. The crystal structure of succinyl-CoA synthetase from Escherichia coli at 2.5-Å resolution. *J Biol Chem*. 1994. doi:10.2210/pdb1scu/pdb

112. Huang J, Fraser ME. Structural basis for the binding of succinate to succinyl-CoA synthetase: *Acta Crystallogr Sect D Struct Biol*. 2016. doi:10.1107/S2059798316010044

113. Joyce MA, Hayakawa K, Wolodko WT, Fraser ME. Biochemical and structural characterization of the GTP-preferring succinyl-CoA synthetase from Thermus aquaticus. *Acta Crystallogr Sect D Biol Crystallogr*. 2012. doi:10.1107/S0907444912010852

114. Huang J, Malhi M, Deneke J, Fraser ME. Structure of GTP-specific succinyl-CoA synthetase in complex with CoA. *Acta Crystallogr Sect Struct Biol Commun*. 2015. doi:10.1107/S2053230X15011188

115. Fraser ME, Hayakawa K, Hume MS, Ryan DG, Brownie ER. Interactions of GTP with the ATP-grasp domain of GTP-specific succinyl-CoA synthetase. *J Biol Chem*. 2006. doi:10.1074/jbc.M511785200

116. Nishimura JS, Meister A. Evidence for Succinyl Phosphate as an Enzyme-bound Intermediate in the Reaction Catalyzed by Succinyl Coenzyme A Synthetase. *Biochemistry*. 1965. doi:10.1021/bi00883a034

117. Bridger WA. Evidence for two types of subunits in succinyl coenzyme a synthetase. *Biochem Biophys Res Commun*. 1971. doi:10.1016/0006-291X(71)90522-5

118. Joyce MA, Fraser ME, James MNG, Bridger WA, Wolodko WT. ADP-binding site of Escherichia coli succinyl-CoA synthetase revealed by X-ray crystallography. *Biochemistry*. 2000. doi:10.1021/bi991696f

119. Weiße RHJ, Faust A, Schmidt M, Schönheit P, Scheidig AJ. Structure of NDP-forming Acetyl-CoA synthetase ACD1 reveals a large rearrangement for phosphoryl transfer. *Proc Natl Acad Sci U S A*. 2016. doi:10.1073/pnas.1518614113

120. Fawaz M V., Topper ME, Firestine SM. The ATP-grasp enzymes. *Bioorg Chem*. 2011. doi:10.1016/j.bioorg.2011.08.004

121. Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F, Serrano L. The FoldX web server: An online force field. *Nucleic Acids Res*. 2005. doi:10.1093/nar/gki387

122. Ashkenazy H, Abadi S, Martz E, et al. ConSurf 2016: an improved

methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res*. 2016. doi:10.1093/nar/gkw408

123. Svergun DI, Koch MHJ. Small-angle scattering studies of biological macromolecules in solution. *Reports Prog Phys*. 2003. doi:10.1088/0034-4885/66/10/R05

124. Volkov V V., Svergun DI. Uniqueness of ab initio shape determination in small-angle scattering. In: *Journal of Applied Crystallography*. ; 2003. doi:10.1107/S0021889803000268

125. Kidmose RT, Juhl J, Nissen P, Boesen T, Karlsen JL, Pedersen BP. Namdinator - Automatic molecular dynamics flexible fitting of structural models into cryo-EM and crystallography experimental maps. *IUCrJ*. 2019. doi:10.1107/S2052252519007619

126. Emekli U, Schneidman-Duhovny D, Wolfson HJ, Nussinov R, Haliloglu T. HingeProt: Automated prediction of hinges in protein structures. *Proteins Struct Funct Genet*. 2008. doi:10.1002/prot.21613

127. Simm D, Hatje K, Kollmar M. Waggawagga: Comparative visualization of coiled-coil predictions and detection of stable single α-helices (SAH domains). *Bioinformatics*. 2015. doi:10.1093/bioinformatics/btu700

128. Vincent TL, Green PJ, Woolfson DN. LOGICOIL - Multi-state prediction of coiled-coil oligomeric state. *Bioinformatics*. 2013. doi:10.1093/bioinformatics/bts648

129. Li X, Wu F, Beard DA. Identification of the kinetic mechanism of succinyl-CoA

synthetase. *Biosci Rep*. 2013. doi:10.1042/BSR20120069

130. Padovani D, Labunska T, Banerjee R. Energetics of interaction between the G-protein chaperone, MeaB, and B 12-dependent methylmalonyl-CoA mutase. *J Biol Chem*. 2006. doi:10.1074/jbc.M600047200

131. Hubbard PA, Padovani D, Labunska T, Mahlstedt SA, Banerjee R, Drennan CL. Crystal structure and mutagenesis of the metallochaperone MeaB: Insight into the causes of methylmalonic aciduria. *J Biol Chem*. 2007. doi:10.1074/jbc.M704850200

132. Jost M, Cracan V, Hubbard PA, Banerjee R, Drennan CL. Visualization of a radical B12 enzyme with its G-protein chaperone. *Proc Natl Acad Sci U S A*. 2015. doi:10.1073/pnas.1419582112

133. Jost M, Born DA, Cracan V, Banerjee R, Drennan CL. Structural basis for substrate specificity in adenosylcobalamin-dependent Isobutyryl-CoA Mutase and related Acyl-CoA Mutases. *J Biol Chem*. 2015. doi:10.1074/jbc.M115.676890

134. Stark H. GraFix: Stabilization of fragile macromolecular complexes for single particle Cryo-EM. In: *Methods in Enzymology*. ; 2010. doi:10.1016/S0076-6879(10)81005-5

135. Serna M. Hands on methods for high resolution cryo-electron microscopy structures of heterogeneous macromolecular complexes. *Front Mol Biosci*. 2019. doi:10.3389/fmolb.2019.00033

136. Briggs JAG, Huiskonen JT, Fernando KV, et al. Classification and three-

dimensional reconstruction of unevenly distributed or symmetry mismatched features of icosahedral particles. *J Struct Biol*. 2005. doi:10.1016/j.jsb.2005.03.009

137. Dunsing V, Luckner M, Zühlke B, Petazzi RA, Herrmann A, Chiantia S. Optimal fluorescent protein tags for quantifying protein oligomerization in living cells. *Sci Rep*. 2018. doi:10.1038/s41598-018-28858-0

138. Strack R. Weighing single proteins with light. *Nat Methods*. 2018. doi:10.1038/s41592-018-0055-6

139. Pelosse M, Crocker H, Gorda B, Lemaire P, Rauch J, Berger I. MultiBac: From protein complex structures to synthetic viral nanosystems. *BMC Biol*. 2017. doi:10.1186/s12915-017-0447-6

140. Altmannova V, Blaha A, Astrinidis S, Reichle H, Weir JR. InteBac - An integrated bacterial and baculovirus expression vector suite . *Protein Sci*. 2020. doi:10.1002/pro.3957