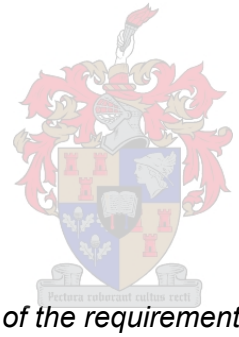
Investigation of protein complexes involved in coenzyme A biosynthesis

By

Warrick Lyle Sitzer



Thesis presented in partial fulfilment of the requirements for the degree of Master of Science in the Faculty of Science at Stellenbosch University

Supervisor: Prof Erick Strauss

Co-supervisor: Dr. Jeremy Woodward (University of Cape Town)

Department of Biochemistry

December 2022

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Warrick L. Sitzer
December 2022

Abstract

The biochemistry of the CoA biosynthetic pathway has been extensively studied and the differences in how the pathway functions between domains of life has led to its validation in recent years as target for drug development against many infectious disease-causing organisms. However, the structural organization of the CoA biosynthetic machinery remains to be investigated. Compelling evidence of a potential CoA-synthesizing protein complex (CoA-SPC) in yeast was recently demonstrated by pairwise interaction studies of the CoA biosynthetic enzymes from *Saccharomyces cerevisiae* which were found to interact with each other. In bacteria, the formation of a CoA biosynthetic enzyme complex has not been reported. However, the elucidation of the *Mycobacterium smegmatis* CoaBC (*Msm*CoaBC) structure and the recent findings showing CoaBC to be a vulnerable point in the CoA pathway suggested that it could form the basis for the formation of such complexes in bacteria.

The first aim of this study focused on the isolation of the proposed CoA-SPC from *S. cerevisiae* and its characterisation by means of affinity pull-down assays. In this investigation, we were able to successfully overexpress yeast-related CoA biosynthetic proteins but unsuccessful in isolating the CoA-SPC from yeast lysate using pull-downs. High performance liquid chromatography (HPLC) was performed as an alternative means to assay for CoA formation by partially purified yeast protein samples. However, no detectable amounts of CoA were observed in the samples after repeated attempts. As CoA is readily detected within yeast lysates, we speculate that errors in either the preparation methods or assays procedure is a result that hindered our progress. Overall, as this complex has not yet been isolated before, the conditions that would lead to its formation could impact its overall isolation and therefore needs to be further explored.

In the second aim, we explored the concept of a CoA biosynthetic complex in bacteria by investigating the structure of the bi-functional CoaBC of *Staphylococcus aureu*s utilising single particle electron microscopy techniques. We were able to produce a low-resolution model of the *S. aureus* CoaBC (*Sa*CoaBC) at ~20Å under negative stain conditions. Two distinct models were reconstructed with one accommodating the overall architecture of *Msm*CoaBC and the other significantly different along its vertices. 3D classifications were unable to clearly establish that the two models are a result of a true conformational change. A detailed single particle cryo-EM study is necessary to determine this, and it will also contribute greatly to the proposed idea of multiple conformations within the CoaBC structure.

Opsomming

Die biochemie van die CoA biosintetiese pad is omvattend bestudeer en die verskille in hoe die pad funksioneer tussen lewensdomeine het in onlangse jare gelei tot die validering daarvan as teiken vir geneesmiddelontwikkeling teen verskeie aansteeklike siekte-veroorsakende organismes. Die strukturele organisasie van die CoA biosintetiese masjinerie moet egter nog ondersoek word. Dwingende bewyse van 'n potensiële KoA-sintetiserende proteïenkompleks (CoA-SPC) in gis is onlangs gelewer deur paargewyse interaksiestudies van die KoA biosintetiese ensieme van *Saccharomyces cerevisiae* wat gevind is om met mekaar te reageer. By bakterieë is die vorming van 'n KoA biosintetiese ensiemkompleks nog nie aangemeld nie. Die oplossing van die *Mycobacterium smegmatis* CoaBC (MsmCoaBC) struktuur en die onlangse bevindinge waarin CoaBC as 'n kwesbare punt in die KoA-padweg uitgelig is, het egter aangewys dat dit die basis kan wees vir die vorming van sulke komplekse in bakterieë.

Die eerste doel van hierdie studie het gefokus op die isolasie van die voorgestelde CoA-SPC van *S. cerevisiae* en die karakterisering daarvan deur middel van affiniteits-uittrektoetse. In hierdie ondersoeke was ons in staat om gisverwante KoA biosintetiese proteïene suksesvol uit te druk, maar onsuksesvol in die isolering van die KoA-SPC vanuit gislisaat. Hoëprestasie vloeistofchromatografie (HPLC) is gebruik as 'n alternatiewe manier om vir KoA-vorming deur gedeeltelik gesuiwerde gisproteïenmonsters. Geen waarneembare hoeveelhede KoA is egter in die monsters na herhaalde pogings waargeneem nie. Aangesien CoA geredelik binne gislisate opgespoor kan word, spekuleer ons dat foute in óf die voorbereidingsmetodes óf toetsprosedure 'n resultaat is wat ons vordering belemmer het. Oor die algemeen, aangesien hierdie kompleks nog nie voorheen geïsoleer is nie, kan die toestande wat tot sy vorming sou lei, die algehele isolasie daarvan beïnvloed en moet dus verder ondersoek word.

In die tweede doel het ons die voorstel van 'n KoA biosintetiese kompleks in bakterieë ondersoek deur die struktuur van die bi-funksionele KoaBC van *Staphylococcus aureus* te ondersoek deur gebruik te maak van enkeldeeltjieelektronmikroskopie-tegnieke. Ons was in staat om 'n lae-resolusie model van die *S. aureus* CoaBC (SaCoaBC) te produseer by ~20Å onder negatiewe kleur toestande. Twee verskillende modelle is gerekonstrueer met een wat die algehele argitektuur van *Msm*CoaBC akkommodeer en die ander beduidend verskillend langs sy hoekpunte. 3D-klassifikasies kon nie duidelik vasstel of die twee modelle die gevolg is van 'n ware konformasieverandering nie. 'n Gedetailleerde enkeldeeltjie-krio-EM-studie is nodig om dit te bepaal, en dit sal ook grootliks bydra tot die voorgestelde idee van meervoudige konformasies binne die CoaBC-struktuur.

Acknowledgements

The pursuit of my MSc degree has been a challenging one and at the same time a very rewarding experience.

A great deal of thanks goes out to my supervisor Prof. Erick Strauss for his guidance and especially the patience throughout the course of my post-graduate journey. I couldn't have done this without you. Times were quite challenging during the final course of my studies and your continuous support and encouragement throughout the time not only made things easier but helped me learn a lot about myself and as a developing scientist. I appreciate the opportunity that you have given me to be part of such and amazing lab and I look forward to what the future has in store!

Many thanks also goes out to my co-supervisor Dr Jeremy Woodward and the microscopy unit team at the University of Cape Town. You have helped a ton and re-sparked my interest in microscopy. It was a deep learning curve but you and everyone else turned it into a pleasant experience.

A special thanks goes out to the entire METH lab group. Every single one of you both past and present made this a wonderful and unforgettable experience. I hope you all continue your paths to success. I wish you all nothing but the best! To the greatest lab manager Dr Helba Bredell, you are the most important part of the team and the genuine care and joy that you radiate on all of us is unmatched.

Finally, and most importantly - to my parents and brother. You all have supported me so much throughout my life and I am truly and eternally grateful. The pure unconditional love and unwavering support you all have given me during my academic journey made me achieve what I thought would be impossible. You all have sacrificed so much for me to achieve my goals and I hope to repay that debt as I progress throughout the next chapters of my life.

Table of Contents

List of Abbreviations

Å	Ångstrom
ADP	Adenosine 5’-diphosphate
ATP	Adenosine 5’-triphosphate
CoA	Coenzyme A
CoaBC	Phosphopantothenoylcysteine synthase/Phosphopantothenoylcysteine decarboxylase
CoA-SPC	Coenzyme A-synthesising protein complex
CoASy	Coenzyme A synthase
CPM	7-Diethylamino-3-(4-maleimidophenyl)-4-methylcoumarin
cryo-EM	Cryogenic electron microscopy
CTF	Contrast Transfer Function
DPCK	Dephospho-coenzyme A kinase
dPCoA	3′-Dephospho-coenzyme A
DTT	Dithiotheitol
Ec	*Escherichia coli*
EM	Electron microscopy
FAS	Fatty acid synthase
GST	Glutathione-*S*-transferase
His	Histidine
HPLC	High performance liquid chromatography
kDa	KiloDalton
LB	Luria-Bertani
MeCN	Acetonitrile
MeOH	Methanol
Msm	*Mycobacterium smegmatis*
Mtb	*Mycobacterium tuberculosis*
NMR	Nuclear magnetic resonance

NS-EM	Negative stain electron microscopy
Pan	Pantothenate
PanK	Pantothenate kinase
PBS	Phosphate buffered saline
PDB	Protein data bank
PDC	Pyruvate dehydrogenase complex
PPCDC	Phosphopantothenoylcysteine decarboxylase
PPCS	Phosphopantothenoylcysteine synthase
RELION	REgularised Likelihood OptimisatioN
Sa	*Staphylococcus aureus*
SAXS	Small angle X-ray scattering
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
TCA	Trichloroacetic acid
TCEP	Tris(2-carboxyethyl) phosphine
TEM	Transmission Electron Microscope/Microscopy
TEMED	Tetramethylethylenediamine
Tris-HCl	Tris(hydroxymethyl)aminomethane-HCl
Y2H	Yeast two-hybrid

List of Figures

List of Tables

Chapter 1: Introduction to protein complexes in CoA Biosynthesis

1.1 An introduction to biosynthetic complexes and protein multifunctionality

The dynamic environments within living cells are handled by a vast spectrum of biological processes governed by proteins and their complexes. Our understanding of the principles that dictate protein structural organization have improved significantly over the past decades. The variation and complexity of these proteins that has been elucidated over the years can be attributed to a multitude of mechanisms operating at the genome, transcriptome and proteome level.

A large proportion of proteins do not operate alone but function as a part of multiprotein complexes (Yanagida, 2002; Berggard *et al.*, 2007). Their interactions can be classified in several ways based on their contrasting functional and structural characteristics. Coordination of proteins on the grounds of their interactions may be defined as either homo-oligomeric or hetero-oligomeric. Furthermore, interactions as a measure of stability are classified as either obligate or non-obligate. Associations may also be transient or permanent determined by their persistence, with transient interactions being commonly associated within signalling pathways while permanent interactions form stable protein complexes.

In order to function, many proteins are assembled into complexes that are comprised of self-interacting copies of a single type of subunit or two or more distinct polypeptide chains. Multifunctional proteins contain two or more enzymatic activities that catalyse consecutive steps of a metabolic pathway and are located on the same polypeptide chain. This is fundamental to many biological processes as the functioning of proteins in this manner make it possible to coordinate and/or control more than one activity at the same time.

Cellular systems also possess multienzyme complexes in which different proteins are bound in highly ordered structures containing various enzymatic activities. The organization of these complexes is a result of an association of functionally related enzymes. Multienzyme complexes optimize cellular organization of biological process and can be considered a step forward in the evolution of catalytic efficiency as they provide advantages that individual enzymes, even those that have achieved near catalytic perfection, would not have. It is fair to assume that both new enzyme functions and multifunctionality arise from existing ones through evolution. This presents a mechanism that allows for the genes encoding the enzyme to be duplicated or fused to preserve original metabolic function and to conserve complex assembly pathways.

In metabolic pathways, enzymes that catalyze the sequence of reactions are linked such that the product of one enzyme catalyzed reaction becomes the substrate for the next. The overall rate of conversion of reactants to products will partly depend upon the extent of coordination between the enzymes. The idea of multifunctional complexes is based on a fundamental concept that reactions that are carried out in stable assemblies of enzymes involved in sequential catalytic transformations would enhance this rate of conversion (Huang, Holden and Raushel, 2001).

The channeling of substrates is a mechanism for the direct delivery of a metabolic intermediate from the active site of one enzyme to the active site of a second enzyme without prior dissociation into the bulk solvent. Thus, the reaction products from one active site are actively or passively translocated directly to another active site. Channeling of this kind can occur within multifunctional enzymes, tightly-associated multienzyme complexes, or transient enzyme complexes. Many essential metabolic enzyme complexes employ the process of substrate channeling to achieve efficient catalysis, by siphoning the substrates and intermediates between their active sites (Figure 1).

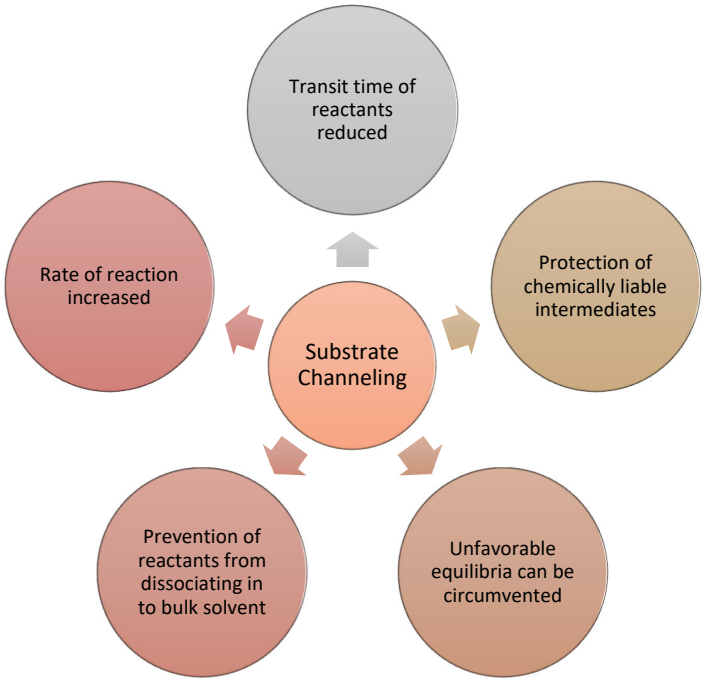


***Figure 1**: Flow chart representing the advantages of substrate channelling that occurs within multifunctional enzymes, tightly associated multienzyme complexes, or transient enzyme complexes.*

Substrate channeling has many advantages over the free diffusion of reaction products within the bulk solvent (Figure 1). The transit time for the movement of reaction products from one active site to the next is reduced (Westerhoff, 1992). Chemically labile intermediates can be protected from decomposition by the aqueous external environment. Unfavorable equilibria

can be circumvented, and reaction intermediates can be segregated from competing enzymatic transformations (Ovadi *et al*., 1994). Examples of substrate channeling have been cited for numerous biochemical pathways, including purine and pyrimidine biosynthesis, amino acid metabolism, lipid metabolism, glycolysis, the tricarboxylic acid cycle, DNA replication, RNA synthesis, and protein biosynthesis (Spivey *et al*., 1999).

The rate at which enzymatic reactions are carried out in the intracellular environment is partly determined by the frequency in which enzymes randomly collide, with the rate of reaction increasing with increased concentration of enzymes and substrates. However, the intracellular environment is a dense system that contains many biological processes carried out in nanomolar concentrations and an increase in rate is not feasible by this avenue. The assembly of multifunctional proteins or multienzyme complexes is an approach whereby this cellular compactness is bypassed by optimizing spatial organization of enzymes in metabolic pathways. The tryptophan synthase complex, fatty acid synthase complex, and pyruvate hydrogenase complex are three well-known examples of such multienzyme complexes.

1.1.1 Tryptophan synthase complex

Tryptophan synthase has served a key role in characterizing the enzymology and structural biology of multifunctional proteins as it was the first enzyme identified having two distinct catalytic activities that catalyses the final two steps in the biosynthesis of L-tryptophan. The α-subunits produce indole and D-glyceraldehyde 3-phosphate through the of cleavage indole 3-glycerol phosphate (IGP). Each β-subunit harbours a pyridoxal phosphate moiety that is used in the generation of L-tryptophan from the endogenously generated IGP and L-serine. There is a substantial body of biochemical and structural evidence to support the idea of substrate channelling occurring. It exists as a $(\alpha\beta)_2$ complex in bacteria and in higher plants but are single bifunctional proteins in yeasts and moulds, with the homologous α and β regions covalently connected by linker peptides (Hyde *et al.*, 1988; Crawford, 1989). The three-dimensional arrangement of the four subunits is assembled in a near linear complex with two α-subunits situated on opposite ends of a centralised ββ dimer, leading to an overall length of roughly 50 Å (Figure 2). The α-subunit folds into a single domain that contains eight parallel β-strands and 13 α-helices in a classical TIM barrel topology. The N- and C-terminal domains of the β-subunit are nearly equal in size and are characterized by four-stranded parallel β-sheet and six-stranded mixed β-sheet, respectively.

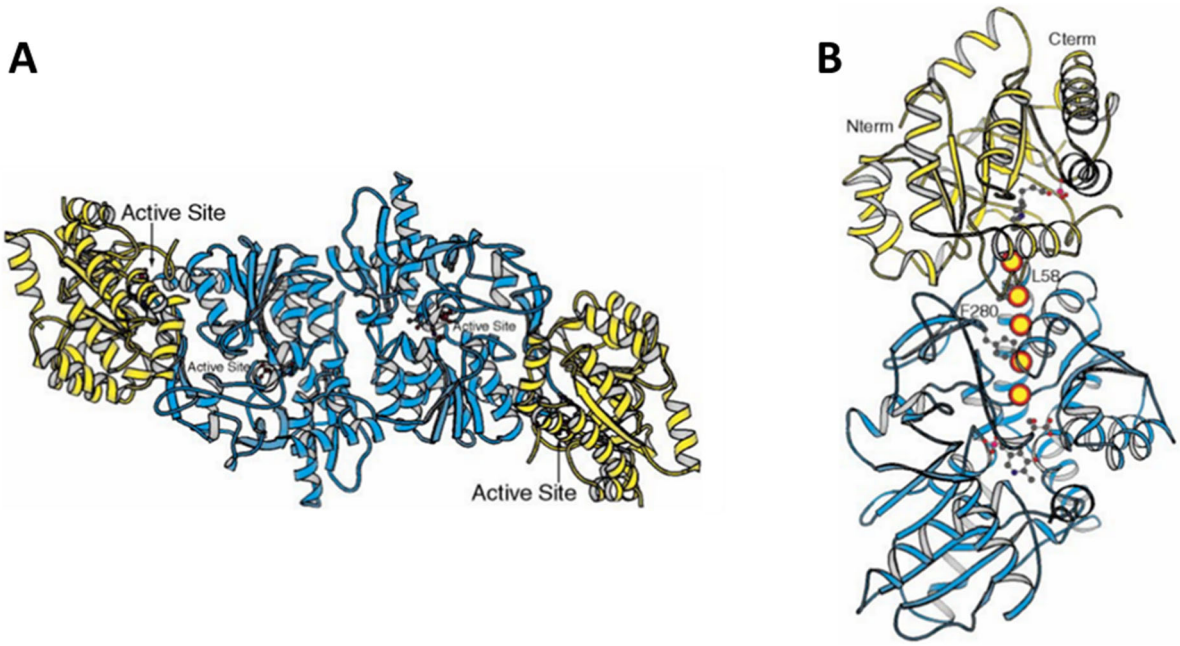


Figure 2: *Tryptophan synthase complex.* ***A)*** *Cartoon representation of the $(\alpha\beta)_2$ complex of tryptophan synthase. The α and β-subunits are color-coded in yellow and cyan, respectively.* ***B)*** *Stereoview of one (αβ)-heterodimer of tryptophan synthase. The orange spheres represent the course of the tunnel running from the α- to the β-subunits. This tunnel is ~25 Å long. Image was reproduced by Huang, X., Holden, H. M., and Raushel, F. M. (2001). Channeling of substrates and intermediates in enzyme-catalyzed reactions. Annual Reviews of Biochemistry. 70(1): 149-180. Reproduced with permission.*

The first direct evidence of tunnelling was captured for the *Salmonella typhimurium* enzyme with a largely hydrophobic tunnel that is contained within the α/β pairs and that spans a distance of 25 Å (Dunn *et al.*, 1990). The tunnel predominantly resides within the β-subunit but extends from the active site of the α-subunit to the centre of the binding site for pyridoxal phosphate. Each α-subunit produces indole and D-glyceraldehyde 3-phosphate through the cleavage of indole 3-glycerol phosphate (IGP). The pyridoxal phosphate containing β-subunits utilizes the endogenously generated indole intermediate with L-serine to form L-tryptophan. Evidence of efficient substrate tunnelling were supported by transient state kinetic investigations that revealed a rapid rate of channelling events by monitoring the overall conversion of IGP and L-serine to L-tryptophan (Creighton, 1970). Supporting studies later demonstrated the total coupling of the reaction at the α- and β-subunits as a result of the allosteric communication between the two sites that allow for open and closed conformational changes to trap the generated indole (Pan and Dunn, 1996). Structural analysis of the tryptophan synthesis complex highlights the importance of molecular tunnels that are thought to protect reactive intermediates from contacting the external medium. To date, all the biochemical data suggest that during the synthesis of L-tryptophan from IGP and L-serine,

indole is not released into solution, which is fully consistent with the tunnelling of this intermediate from the active site of the α-subunit to that of the β-subunit

1.1.2 Fatty acid synthase complex

Another noteworthy example of an multienzyme complex can be found in the fatty acid biosynthesis pathway where a series of discrete enzymatic reactions are directed sequentially in a highly organized system. This pathway entails the iterative elongation of fatty acid chains through a set of chemical reactions conserved in all kingdoms of life. Four basic reactions constitute a single round of elongation. In the first reaction, acetyl primer and malonyl elongation substrates are loaded from coenzyme A (CoA) to acyl carrier protein (ACP) by acetyltransferase (AT) and malonyl/palmitoyl transferase (MPT) and condensed to acetoacetyl-ACP in a decarboxylative reaction catalyzed by ketoacyl synthase (KS). In three subsequent reaction steps, the β-carbon groups are processed by ketoacyl reductase (KR), dehydratase (DH), and enoyl reductase (ER), which results in fully saturated acyl-ACP that can serve directly as a primer for the next condensation reaction. In each reaction cycle, the growing acyl-chain is elongated by two carbon units until C16 or C18 fatty acids are back transferred to CoA by MPT.

The organization of the catalytic units in fatty acid synthase complex (FAS) are remarkably diverse and are divided into two basic types. Type II FAS (FAS-II) are mainly found in bacteria, plants and parasites and, are comprised of small discrete monofunctional proteins each encoded by a separate gene that catalyzes a single step of the cyclic reaction (White *et al*., 2005). Type I FAS (FAS-I) systems, however, are highly integrated multienzyme complexes that have been extensively studied supported by the technical advances in both X-ray crystallography and cryogenic electron microscopy (cryo-EM). Type I FAS systems are structurally elaborate and functionally highly developed multifunctional protein complexes which evolved by gene fusion events from type II systems. Although FAS-II systems are capable of producing a diversity of products for cellular metabolism, FAS-I is usually considered to be a more efficient biosynthetic machine containing the entirety of catalytic activities required for the cyclic reaction as discrete domains on a single polypeptide chain (α) or on two different (α and β) polypeptides (Figure 3).

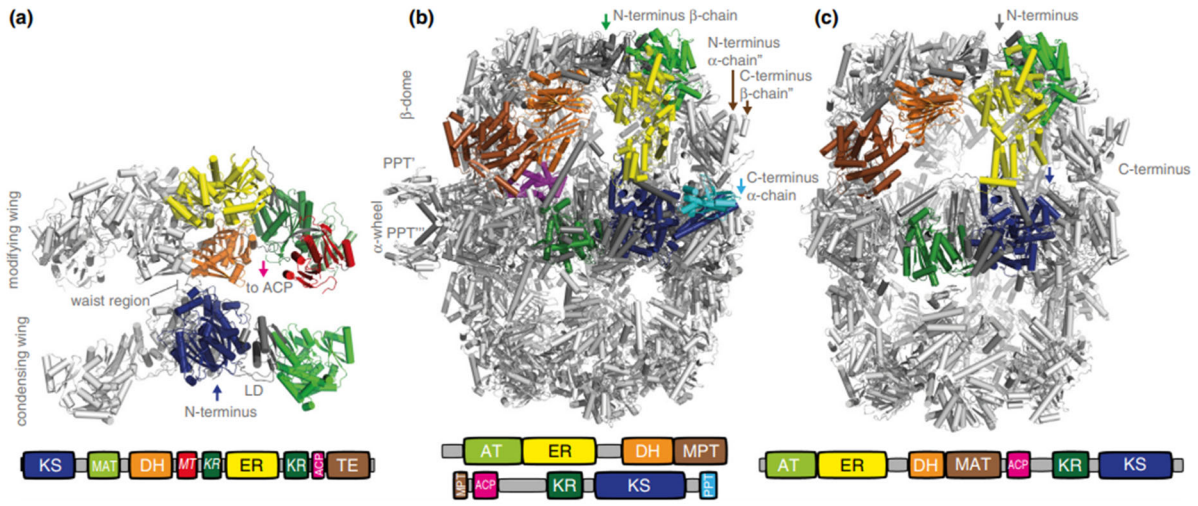


***Figure 3**: Organization and structure of fatty acid synthase complexes. Structure and domain organization of the mammalian **(a)**, the fungal **(b)** (full-length model presented), and the bacterial **(c)** FAS I system. Single chains of FAS type I systems are colored according to the attached domain overview. Images reproduced from Grininger, M. (2014). Perspectives on the evolution, assembly and conformational dynamics of fatty acid synthase type I (FAS I) systems. Current Opinion in Structural Biology. 25: 49-56. (https://doi.org/10.1016/j.sbi.2013.12.004). Reproduced with permission.*

In mammalian cells, FAS-I are recognized as active X-shaped α_2 homodimers 540 kDa in size whereas fungal FAS-I are expressed as a large 2.5 MDa highly symmetric, homohexameric or heterododecameric $\alpha_6\beta_6$ complex (Lomakin, Xiong and Steitz, 2007; & Maier, Leibundgut and Ban, 2008). In recent years the structural organization of the *Mycobacterium smegmatis* FAS was solved and revealed as a minimized version of fungal FAS-I from flexible fitting of the fungal FAS-I into the derived electron microscopy data. ACP-mediated substrate shuttling represents a key property of FAS-I systems guided by the higher order organizational and conformational concept studies from eukaryotic FAS. The movement of ACP within FAS-I systems is described to be stochastic, as the increase of this systems productivity is primarily carried out by retaining intermediates within tightly confined enzymatic compartments. Enzymatic efficiency is enhanced in this way when compared to FAS-II systems in which the individual enzymes are present in high copy numbers to make up as a result of its lower organizational level.

The architecture of FAS-I systems in fungi are particularly efficient in achieving high catalytic efficiency as the rigid cage-like machinery harbouring mobile ACP domains facilitates the transfer of intermediates. The mobile ACP is connected by two flexible linkers which confine the ACP path. Acyl carrier domains are isolated in the reaction chambers that shuttles the growing acyl chains to the adjacent active sites while end products and substrates are shuttled between the active sites (Jenni *et al*., 2007; Lomakin, Xiong and Steitz, 2007). The enzymatic

domains are architecturally arranged to minimize diffusion distances of the consecutive step of a fatty acid cycle, so that the probability for a productive interaction of ACP increases. Furthermore, active site coupling in this manner increases the reaction rate by the shuttling of ACP and likely prevents acyl-ACP intermediates from unwanted side reactions. Differences in the positioning of the domains and barrel hinges was illustrated in cryo-EM studies implied breathing within the reaction chambers. Similarly, bacterial FAS shares this concept as it is architecturally similar to the fungal FAS.

In a similar manner, the mammalian FAS-I fold might fulfil structural purposes for improving catalytic efficiency. Mammalian FAS can be conveniently characterized as comprising a body (ER and KS dimers and pseudo dimeric DH pairs) with two arms (KR monomers) and legs (monomeric MAT domains). This organization of mammalian FAS can be further grouped into two clefts that serve as entry sites to the active centres (Maier, Jenni and Ban, 2006; Maier, Leibundgut and Ban, 2008). Each cleft is oriented facing the active sites of the two sets of catalytic domains which in turn forms two chambers. The synthesis of fatty acids at these two active sites have been proposed to function asynchronously between open and closed states (one chamber involved in carbon-chain elongation and the other engaged in β-carbon processing). The flexible connection of the mammalian FAS halves may allow rotational motion around the dimer axis or a certain degree of tilting. Such motion would drag the ACP between the two faces of FAS and may contribute considerably to productive substrate shuttling.

1.1.3 Pyruvate dehydrogenase complex

Another well-known example of a multienzyme complex is the pyruvate dehydrogenase complex (PDC), which catalyses the key step in carbohydrate utilization: the linking of glycolysis to the citric acid cycle and to fatty acid biosynthesis through the conversion of pyruvate into acetyl-CoA and NADH. It is among the largest highly organised multienzyme systems that is found in both eukaryotes and prokaryotes. The biochemistry of this multicomponent enzyme assembly has been extensively studied and the individual enzymes and subdomains have been elucidated in great detail under nuclear magnetic resonance (NMR) and X-ray crystallographic methods, whereas the overall architecture of PDC complexes have been illuminated by means of cryo-EM.

The PDC enzyme assembly consists of multiple copies of three different catalytic components: pyruvate decarboxylase (E1), dihydrolipoamide acetyltransferase (E2), and dihydrolipoamide dehydrogenase (E3). E1 domains are known to be expressed as dimers (α_2) in *Escherichia coli* or heterotetramers ($\alpha_2\beta_2$) within yeast and mammalian systems. The E3 component is known to form dimers (α_2). The core structure of eukaryotic PDCs harbour an additional

noncatalytic putative subunit (E3-binding protein) that binds E3 specifically which is not present in prokaryotic counterparts.

The PDC is based on icosahedral symmetry surrounding E2 that forms the core of the macromolecular structure to which it associates tightly but non-covalently with multiple copies of E1 and E3 (Figure 4). The inner core is composed of either twenty four or sixty copies of the E2 catalytic domains arranged to form a cubic assembly with octahedral symmetry (gram negative bacteria) or pentagonal dodecahedron assembled into icosahedral symmetry respectively (eukaryotes and gram-positive bacteria). The E2 component polypeptide chain is composed of a 9 kDa lipoyl domain at the N-terminus; a small 4 kDa peripheral subunit-binding domain; and a large 24 kDa C-terminal catalytic domain responsible for directing the inner core assembly of E2 and expressing acyltransferase activity (Milne *et al.*, 2006; & Byron and Lindsay, 2017). The lipoyl, catalytic and subunit binding domains are all connected by an extended mobile linker regions.

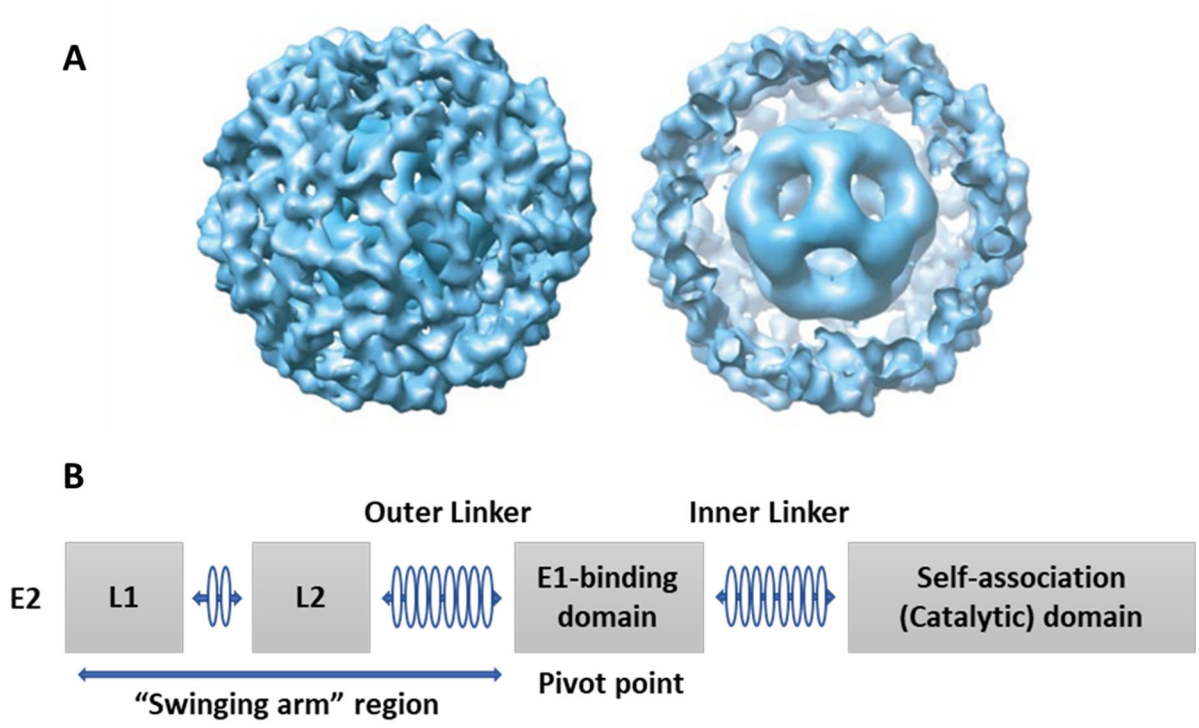


***Figure 4**: Architecture of the pyruvate dehydrogenase complex. **A** Surface representation of the refined three-dimensional model of E2E3 viewed along a 3-fold axis of symmetry with the same representation with a portion of the outer protein shell removed to aid visualization of the inner E2 core. **B** A simplified scheme representing the components responsible for substrate shuttling. Image **A** was reproduced from Milne, J. L., Wu, X., Borgnia, M. J., Lengyel, J. S., Brooks, B. R., Shi, D. and Subramaniam, S. (2006). Molecular structure of a 9-MDa icosahedral pyruvate dehydrogenase subcomplex containing the E2 and E3 enzymes using cryoelectron microscopy. Journal of Biological Chemistry, 281(7), 4364-4370. CC-BY 4.0.*

In the native PDC complex, the pyruvate decarboxylase and dihydrolipoamide acetyltransferase catalytic components coordinate to transfer an acetyl group derived from the oxidative decarboxylation of pyruvate to CoA. This is accomplished via a dihydrolipoyl group covalently attached to a lysine residue in the lipoyl domain of E2, whereas E3 regenerates an oxidised lipoyl domain to complete the cycle.

A key feature is that the E1 molecules are located on the periphery of the assembly in an orientation that allows each of the lipoyl domains tethered to the inner E2 core to access multiple E1 and E2 active sites from inside the complex. Furthermore, E1 and E3 components are located as well-defined radial shells exemplified by cryo-EM analysis of the fully assembled E1E2 and E2E3 subcomplexes. These components are separated by a clear ~75-90 Å gap from the inner icosahedral core in which the three sections of the E2 resides. The lipoyl movement is not entirely free and has been suggested to occur through constrained trajectories. The outer surface of the icosahedral E2 core was revealed to be predominantly positive charged extending toward each active site whereas the lipoyl domain in contrast is overall negatively charged. This suggested that that the movement of the lipoyl domain is guided by means of electrostatic interactions within its interior as the basis of substrate channelling.

1.2 Protein complex formation in CoA biosynthesis

1.2.1 Importance of CoA

The assembly of multienzyme complexes within biologically important pathways provides a mechanism to confine crucial reaction processes that are essential for growth and energy homeostasis within organisms. Of particular interest and a focus of this study is to investigate if such an assembly exists within the coenzyme A (CoA) biosynthetic pathway. CoA is composed of 3′-phosphoadenosine-5′-phosphate (PAP) moiety and a thiol-containing 4′-phosphopantetheine arm. Both moieties have been demonstrated to be essential for the intracellular functions of CoA. CoA and its other metabolically active derivatives are paramount to central cellular metabolic and signalling pathways with an estimation that it is involved in 9% of all known enzyme activities as defined by EC number (Strauss, 2010). CoA is also the source for the 4′-phosphopantetheine prosthetic group mediating the tricarboxylic acid cycle, degradation of fatty acids and its synthesis, amino acid synthesis, polyketide and non-ribosomal peptide biosynthesis and many other biological important processes as illustrated in Figure 5 (Leonardi *et al*., 2005; Spry, Kirk and Saliba, 2008).

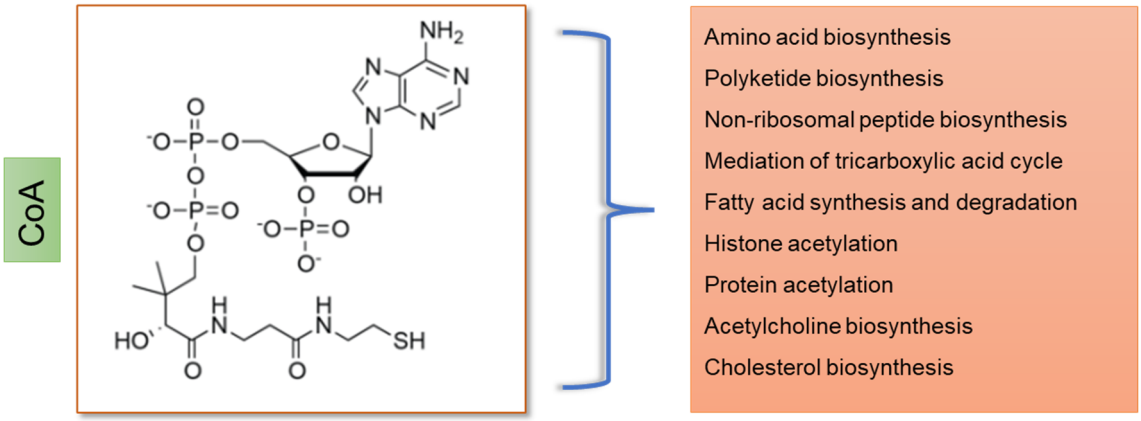


***Figure 5**: Image representing the CoA-dependent processes.*

In addition, recent studies illustrated the importance of CoA functioning as an important antioxidant in cellular response to oxidative and metabolic stress through protein "CoAlation" mediation (Gout, 2018; 2019). Furthermore, many determinants including hormones, nutrients, various cellular stresses, and metabolites strictly maintain the levels of CoA and its derivatives (Tubbs and Garland, 1964; Smith and Savage, 1980; Robishaw *et al.*, 1982). Although ubiquitous in nature, CoA cannot be taken up from the extracellular environment and must be synthesised *de novo* through a highly conserved five-step CoA biosynthetic pathway (Strauss, 2010).

1.2.2 Universal mechanism of CoA biosynthesis

The universal CoA biosynthetic pathway consists of five enzyme activities that convert pantothenate (vitamin B5) to CoA (Figure 6). The first catalytic step is executed by pantothenate kinase (PanK) to generate 4′-phosphopantothenate by the adenosine triphosphate (ATP) - dependent phosphorylation of pantothenate. PanK has been considered a potential target for antimicrobial drug development as it is having been proposed to be "rate-limiting" in most organisms through feedback inhibition by the pathway product CoA or its derivatives (Jackowski and Rock, 1981; Strauss, 2010). The second step involves the nucleoside triphosphate - dependent ligation of 4′-phosphopantothenate and L-cysteine by phosphopantothenoylcysteine synthetase (PPCS) to produce 4′-phosphopantothenoylcysteine. Two forms of PPCS have been identified that differ based on their nucleotide triphosphate requirements. Bacterial enzymes catalyse the formation of 4'-phosphopantothenoylcysteine using cytidine triphosphate (CTP), whereas ATP is used in eukaryotic organisms. Phosphopantothenoylcysteine decarboxylase (PPCDC) then utilizes the formed 4′-phosphopantothenoylcysteine in a decarboxylation reaction to generate 4′-phosphopantetheine. The penultimate catalytic step involves the transfer of an AMP moiety from ATP through the action of phosphopantetheine adenylyltransferase (PPAT) to form 3′-dephospho-CoA, which is then phosphorylated at its 3′-hydroxyl group by dephospho-CoA

kinase (DPCK), the final enzyme in the biosynthetic pathway, to produce CoA (Leonardi and Jackowski, 2007).

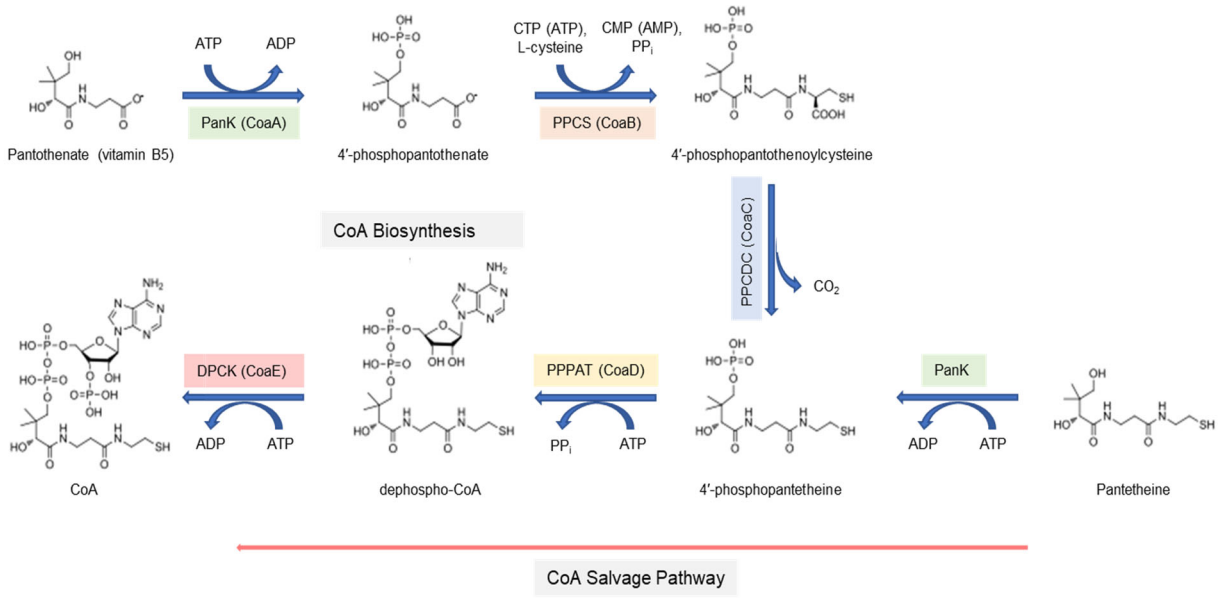


Figure 6: *The CoA biosynthetic pathway. The biosynthesis of CoA from pantothenic acid (vitamin B_5) is catalysed sequentially by pantothenate kinase (PanK), phosphopantothenoylcysteine synthetase (PPCS), phosphopantothenoylcysteine decarboxylase (PPCDC), phosphopantetheine adenylyltransferase (PPAT) and dephospho-CoA kinase (DPCK). The proteins representing the biosynthetic enzymes in bacteria are CoaA, CoaB, CoaC, CoaD and CoaE, respectively.*

1.2.3 Enzymes of the Coenzyme A biosynthetic pathway

Although the CoA pathway is conserved in all domains of life, a considerable amount of genetic, mechanistic, and structural diversity has been observed between the pathway enzymes of prokaryotic and eukaryotic organisms. A striking example is PanK, which have been shown to exist as three distinct types that are characterized based on their catalytic properties, sequence homology and also their overall structural fold (Brand and Strauss, 2005; & Yang *et al,* 2006). Type I and Type III PanKs are expressed in prokaryotic microorganisms whereas Type II PanKs are most often found in eukaryotic organisms. Type I PanKs are recognized as the prototypical bacterial PanK and has been extensively studied among the three types (Song and Jackowski, 1992). These proteins are homodimeric in structure and belong to the family of P-loop kinases (Vallari *et al.*, 1987; Song and Jackowski, 1992). Although type II PanKs are most commonly found in eukaryotic organisms, it is also present in *Staphylococcus* species, such as *Staphylococcus aureus*. The type II PanKs found in humans are produced by four distinct PanK genes that exhibit tissue-specific expression with some genes capable of expressing two distinct isoforms (Rock *et al*, 2000; Zhou *et al*, 2001; Hortnagel *et al*, 2003). Lastly, type III PanKs are unique in the sense that they predominantly

found in pathogenic bacteria, and unlike other PanKs do not experience feedback inhibition by CoA or acetyl-CoA. Additionally, these PanKs also do not accept alternative substrates such as pantothenate analogues.

The PPCS enzymes can be classified into two distinct types based on their utilization of either CTP or ATP, but these enzymes also feature differences with regards to their structural fold and arrangement of genes. Eukaryotic PPCS enzymes are mainly expressed as dimers with identical monomers. Some features of the monomer fold motifs are comparable to that of the ribokinase-fold found in a group of NAD-dependent enzymes. Conversely, bacterial CoaB and CoaC proteins are encoded by a single gene to produce a fused bifunctional enzyme (CoaBC) expressing both PPCS and PPCDC activities. In the same way PPAT activity in eukaryotes is fused to DPCK, the last enzyme in the pathway, forming a bifunctional PPAT/DPCK protein that is also referred to as CoA synthase (CoASy). Bacterial PPAT (CoaD) is expressed as a single monofunctional protein with that is homohexameric in structure that can be described as a dimer of distinct trimers. Lastly, the DPCKs in bacteria vary in their oligomeric structure depending on the organism of interest, with reports of monomeric, dimeric and also trimeric arrangements.

1.2.4 Evidence for complex formation in CoA biosynthesis

The biochemistry of the CoA biosynthetic pathway has been extensively studied and the differences in how the pathway operates in various domains of life allowed for the validation of this pathway in recent years as a target for drug development against many infectious disease-causing organisms (Spry *et al.*, 2008; Strauss, 2010; Moolman *et al.*, 2014). Except for 4′-phosphopantetheine, the CoA biosynthetic intermediates are not known to be used for other biological roles. As such, the existence of higher order assemblies makes sense in view of the molecular organization of CoA biosynthetic genes in other organisms, and also to promote flux through the CoA biosynthetic pathway by ensuring proximity of the pathway enzymes (Figure 7). While no structure of any CoA biosynthetic complex has been reported, the following provides a summary of the evidence that such complexes may form, even if only transiently.

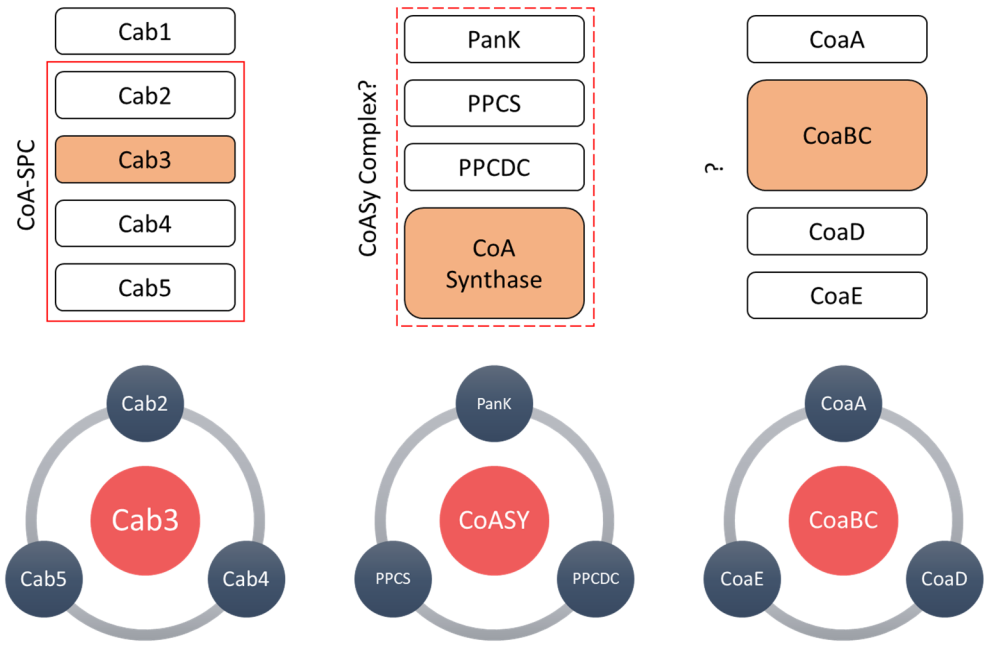


Figure 7: *Enzymes involved in the biosynthesis of CoA for yeast, mammalian, and bacterial cells respectively. Solid rectangle represents the enzymes involved in the formation of the CoA-SPC. Dotted lines represent a potential biosynthetic complex in mammalian cells.*

1.2.4.1 Bi-functional CoA Synthase (CoASy)

Reports on the localisation of CoASy and its interplay with other intracellular metabolites and signalling pathways suggests a biosynthetic complex exist in mammalian cells. An initial study described the possibility of CoASy interacting with other cellular proteins through its characterization by means of gel filtration chromatography. The study demonstrated that CoASy was eluted in two separate peaks, one in a low molecular-mass range of 60-150 kDa and another on a higher peak range of around 450 kDa. This suggested that CoASy may exhibit interactions with other proteins including those associated with CoA biosynthesis (Nemazanyy *et al.*, 2004).

Proteins that were found to interact with CoASy include those from signalling pathways coordinated by serum stimulation or starvation in response to stresses. These proteins include p85α, which is the regulatory subunit of PI3K (phosphoinositide 3-kinase), ribosomal protein S6K1, Shp2PTP (Src homology 2 domain-containing protein tyrosine phosphatase) and Src, which are tyrosine phosphatase and tyrosine kinases respectively, and EDC4 (enhancer of mRNA-decapping protein 4), a central scaffold constituent of processing bodies (Nemazanyy *et al*, 2004; Breus *et al.*, 2009; Breus *et al.*, 2010; Gudkova *et al.*, 2012).

Many studies also reported on the modulation of CoASy activity by modifications induced post-translationally. An increase in PPAT activity has been demonstrated as a response to tyrosine dephosphorylation of CoASy by Shp2PTP, while DPCK activity of CoASy was found to be

inhibited by EDC4, a protein that has also been shown to interact with CoASy in response to growth factors and cellular stresses (Breus *et al.*, 2010; Gudkova *et al.*, 2012). Furthermore, the interaction of CoASy with p85a is facilitated by tyrosine phosphorylation carried out by the Src tyrosine kinase family. Additionally, a recent study utilizing a proximity ligation assay has illustrated an increase in the interaction of PanK1β and CoASy in response to oxidative stress and induced serum starvation (Bakovic *et al.*, 2021). The significance of these interactions is not clearly understood but it provides insights into the regulation of CoA biosynthetic complex assembly.

The exact subcellular localization of CoASy remains unresolved with studies reporting on its confinement within the nucleus, outer mitochondrial membrane or within the mitochondrial matrix (Zhyvoloup *et al.*, 2003; Rhee *et al.*, 2013; Dusi *et al.*, 2014; Lin *et al.*, 2018). It has been hypothesized that CoASy can function as a scaffold for the formation of a CoA biosynthetic complex in mammalian cells as it was found it be anchored to the outer mitochondrial membrane via its N-terminal hydrophobic region (Zhyvoloup *et al.*, 2003). This was further supported by the discovery of CoASy being largely upregulated by phosphatidylcholine and phosphatidylethanolamine, which are the predominant components of the outer mitochondrial membrane. It was revealed that both enzymatic domains of CoASy are aligned to the cytosol suggesting that clustering of the CoA biosynthetic enzymes around the CoASy at the outer mitochondrial membrane may provide an efficient means of channelling intermediates between the cellular compartments in response to external stimuli.

Altogether, localization of individual CoA biosynthetic enzymes regulated by external factors such as nutrients, growth factors and stress signalling may create an interplay of these enzymes with the anchored CoASY and in this way can also allow for the produced CoA to conciliate processes in both the mitochondria and cytosol.

1.2.4.2 Coenzyme A synthesizing protein complex in Yeast

Several studies have presented compelling evidence for the existence of a multienzyme CoA synthesizing protein complex (CoA-SPC) in yeast (*Saccharomyces cerevisiae*) cells. The first of these was a study by Bucovaz *et al*. (1997) that demonstrated the existence of such a CoA-SPC in yeast with a molecular weight between 300 - 400 kDa. In their efforts, they proposed that the complex harbours four activities, with the first of these being the ATP-dependent transfer of ADP to pantothenate to form 5′-ADP-4′-pantothenate. However, later investigations revealed that the products of the genes encoding the five CoA biosynthetic enzymes in *S. cerevisiae* (Cab1–Cab5, where Cab is CoA biosynthesis) are similar to those of the universal CoA biosynthetic pathway (Olzhausen, Schübbe and Schüller, 2009). Yeast Cab null mutants Δ*cab2* and Δ*cab3* have also been demonstrated to individually complement the *coaBC* gene

from *E. coli*. Together, these investigations strongly suggested that in yeast CoA is synthesized according to the universal five step pathway, conclusively refuting the alternative pathway proposed by Bucovaz *et al*. (1997). Nonetheless, these findings did not negate the likelihood of a CoA-SPC existing in yeast, as it could still be formed by interactions between the various Cab proteins.

Evidence of a molecular assembly within the yeast CoA biosynthetic pathway was initially provided by the characterization of a heterotrimeric PPCDC in yeast that is comprised of Cab3 with nonessential yeast proteins Hal3 (also referred to as Sis2) and Vhs3 (Ruiz *et al*., 2009). Both Hal3 and Vhs3 act as regularity subunits of protein phosphatase Ppz1 that participates in regulating cellular integrity, halotolerance and monovalent cation homeostasis (Ruiz *et al.*, 2009). This study documented the first potential crosstalk between unrelated biochemical pathways, also referred to as moonlighting of proteins. These proteins individually act as inhibitors of Ppz1 but forms part of the unique heterotrimeric PPCDC enzyme when in combination with Cab3, catalyzing the third step of the CoA biosynthetic pathway with a molecular mass of 210-240 kDa (Abrie *et al.*, 2015). Conversely, other eukaryotic PPCDCs such as in mammalian cells are expressed as homotrimers with the active sites forming at the interacting interfaces of the oligomer. Compelling evidence for the existence of a CoA-SPC was demonstrated by pair-wise interaction studies among the Cab proteins in which Cab3 was suggested to form the scaffold for its assembly (Olzhausen *et al*., 2013). The results of this investigation are discussed in more detail in Chapter 3.

The levels of CoA and its derivatives varies significantly depending on the subcellular compartment that is examined. Cytosolic CoA pools are much lower compared to those found in the mitochondria and peroxisomes (Horie *et al.*, 1978; Idell-Wenger *et al.*, 1978; Williamson and Corkey *et al.*, 1979). This segregation may imply that the biosynthetic machinery of CoA assembles differently within each compartment, and/or alludes to the differential clustering of the enzymes in response to extracellular stimuli and stresses. In this way synthesised CoA can be mediated between intercellular compartments considering the transport of CoA into mitochondria has been reported to occur in yeast and also in mammalian cells (Tahiliani, 1989).

Taken together, these studies provide important insights of a molecular assembly of the CoA biosynthesis pathway enzymes that might be present within selected organisms. The clustering of enzymes to form a CoA-synthesizing complex provides many structural and functional benefits over individual enzymes acting separately. It may allow for the five consecutive reactions of CoA synthesis to proceed efficiently and without the intermediates

diffusing from the biosynthetic complex. However, further studies are required to characterize such a CoA-SPC, and to determine whether such proposed benefits are indeed real.

1.3 Identification and characterization of protein complexes

1.3.1 Methods to identify protein interactions

A range of methods currently exits to establish if two or more proteins are interacting either transiently or in a more stable manner, forming complexes. Of these, affinity purification and *in vivo* yeast two-hybrid assays (Y2H) are some of the most widely used for protein interaction analysis (Abu-Farha *et al*, 2009). For Y2H assays, the method is based on the use of transcription factors that require a DNA binding domain (DBD) that is fused with the protein of interest and an activation domain (AD) that enables gene expression which is merged with the prey protein. The interaction between the protein partners unifies the DBD and AD to activate the transcription of the reporter gene. Direct recognition of protein pairs is accomplished in this manner. However, it can only be applied to validate binary interactions and many false positives may be generated in the process when applied in high throughput-screens. Although robust in practice, it cannot be applied to all protein-protein interactions and to detect transient interactions.

In situ imaging methods such as the bimolecular fluorescence complementation assay (BiFC) and fluorescence resonance energy transfer (FRET) are also common techniques used to detect binary interactions. In short, FRET is based on the direct energy transfer between overlapping fluorophores that are linked to the proteins of interest. The interaction between the bait and prey results in a fluorescence emission when in close proximity, and is recorded in real time in live cells (Yu *et al.*, 2008). Similarly, BiFC entails the use of fluorescent proteins but that are fragmented. The bait and prey proteins contain a fragment of a preferred fluorescent protein, and its interaction will reconstitute it to produce a measurable readout (Zhang *et al.*, 2014). Both techniques have an advantage over Y2H in which the information obtained is in real time and provides information on the cellular location of interacting proteins. However, the success in each case is based on the quality of antibodies that are highly specific.

Unlike Y2H and fluorescence-based techniques that are used to validate pairwise interactions, recent advancements in affinity purification that is coupled to MS methods (AP-MS) are used for the purpose of identifying protein complexes (Berggard *et al.*, 2007). Briefly, affinity beads are pre-saturated with the tagged protein of choice is allowed to incubate with the biological extract containing proteins that are present in their native form. Proteins that have a high affinity for the immobilised proteins will bind while proteins that interact non-specifically is removed by washing. Multiprotein complexes are isolated directly from cell lysates in this

manner which is identified by MS and validated by blotting if antibodies of the candidate protein are available. A benefit of this technique is that it can be performed under near physiological conditions and also in the relevant organism and cell type.

When used in combination with MS and quantitative proteomics techniques, AP-MS can be used to explore the dynamics of protein complex compositions and altogether offers a lower risk of perturbing relevant post-translational modifications that are important for protein complex activity and organization (Berggard *et al.*, 2007; Choudhary and Mann, 2010). The limitations arise from the proteins that need to survive the wash steps which influences overall analysis, and the technique falls short when detecting fast exchanging complexes. Furthermore, the tag implements can carry the burden of interfering with protein function. Technological advancements between the techniques described are used to routinely probe dynamic changes in protein complex composition, large-scale analysis of protein interactions and their implication in biological processes between important pathways.

1.3.2 Methods to structurally characterize protein complexes

The structural characterization of stable protein complexes is often an invaluable tool in the elucidation of the mechanism whereby such complexes are formed and/or maintained, and the relation of this to their cellular function. This is especially the case when the characterization can be done for multiple functional states of the complex of interest.

The methods that are currently used to structurally characterize protein complexes are summarised in Table 1. Of these, the static three-dimensional (3D) structure determined by X-ray crystallography is still the most widely used. While such structural elucidations have been very successful, the need to obtain protein crystals is a major bottleneck - especially in the case of large and conformationally labile macromolecular assemblies. Additionally, the crystal lattice forces may unnaturally constrain the structure. However, macromolecular complexes are not static; structural fluctuations in proteins can occur simultaneously and at different protein structure levels. As such, the structure determined by X-ray crystallography often cannot provide information on the dynamics of multienzyme complexes.

Table 1: Methods used to structurally characterize protein complexes.

Biophysical Method	Sample Type	Advantages	Limitations
X-Ray Crystallography	Crystals	Very high resolution revealing fine detail of atomic structure	Crystals required Flexible portions are not seen. Structure may be influenced by crystal packing forces
Cryogenic Electron Microscopy	Frozen and dilute solutions (≤1 $mg.mL^{-1}$)	Low amount of material required. Direct visualization of particle shape and symmetry at near atomic details (0.3-0.4nm)	High computational power Low contrast presents challenges to determine structures below 100 kDa
Nuclear Magnetic Resonance	Dilute solutions (~5–10 $mg.mL^{-1}$)	High resolution (0.2-0.3 nm) in solution	Difficult to apply to proteins with MW exceeding 50 kDa
Small Angle X-Ray Scattering	Semi-dilute solutions (~1–100 $mg.mL^{-1}$)	Analysis of structure, kinetics and interactions in nearly native conditions. Study of mixtures and non-equilibrium systems. Wide MW range	Low resolution (~1–2 nm) Requires additional information to resolve ambiguity in model building

Solution methods such as NMR spectroscopy can reveal short-range dynamic behaviour in complex information. The method can yield an ensemble of models, potentially identifying flexible arrangements in solution, but it is limited by the size of the macromolecule that can be studied and typically requires that the complex molecular weight does not exceed 50 kDa for conventional applications. NMR spectroscopy was actively employed to aid in the elucidation of the lipoyl swinging arm domains of the PDC and functional interactions of the components that make up the FAS complexes.

Conformational dynamics is often an important feature of the function of protein complexes, and the existence of a dynamic compositional and conformational ensemble presents challenges for obtaining molecular-level snapshots of macromolecular complexes in defined states. The use of cryo-EM has greatly expanded to allow for the determination of the conformational and compositional heterogeneity of multiple structures of a single specimen preparation. The refinement procedures and focused classification techniques employed in

the process can be used to sort out localised structural heterogeneity and improve feature resolvability of flexible domains by means of local refinement.

Advances in cryo-EM in hardware and software have also made tremendous progress toward achieving atomic resolution for a broad range of large molecular complexes (Nogales, 2015; Cheng, 2018; Beplar *et al*, 2019). One of the key examples was the determination of the mammalian transient receptor potential channel TRPV1 at 3.4 Å resolution which represented the first example of a membrane protein to break the side chain resolution barrier without the use of crystallization. Its structural information aided in the elucidation on how this channel responds to physical and chemical stimuli. A robust structural strategy for resolving independent fluctuations, as well as the potential correlations among fluctuations within a single molecular complex, has recently been developed where focused classifications was used to decipher the protein dynamics of identical subunits in the bacterial chaperone GroEL. In this study, the 3.5 Å cryo-EM reconstruction was used to further analyse structural heterogeneity among chemically identical subunits in each GroEL oligomer (Roh *et al*., 2017).

Recent advances have also demonstrated that cryo-EM can also be combined with other methods to investigate higher order protein complexes. In a study by Diego *et al*. (2019), they introduced an integrated structure determination approach that simultaneously uses NMR and cryo-EM data to overcome the limits of each of these methods. The approach enabled determination of the structure of the 468 kDa dodecameric aminopeptidase TET2 at high precision and accuracy. This was accomplished by combining secondary-structure information obtained by NMR assignments of the 39 kDa-subunits and distance restraints from backbone amides with a 4.1Å resolution from the cryo-EM map. Their approach exceeded current standards of NMR and cryo-EM structure determination in terms of molecular weight and precision, and was demonstrated to be successful even in cases where only medium-resolution cryo-EM data are available.

Finally, the assembly of macromolecular complexes can also be studied through the docking of individual components into *ab initio* shapes obtained using small angle X-ray scattering (SAXS) (Wriggers and Chacón, 2001). The resolution of the shapes derived from SAXS analysis are generally low but is actively employed for the analysis of macromolecular complexes and provides information about conformational and compositional states of complex in solution. The architecture of the centralized portion of the human complement factor H (fH) was successfully demonstrated by means of the *ab initio* and rigid body modelling approach (Schmidt *et al*., 2010). The complement fH is an important component in the regulation in the humane innate immune system and comprises 20 modular complement control modules that are associated by short linkers. Structures of the 12 terminal complement

control modules have been previously determined in isolation or as two to four domain fragments at high resolution, but the structure of the full-length protein was not known until the SAXS study (Schmidt *et al*., 2008).

1.4 Study Aims

1.4.1 Problem statement and overall aim of the study

Despite much circumstantial evidence for the CoA biosynthesis enzymes forming at least transient complexes, no such complex has been characterized. The overall aim of this study was to obtain structural information that would provide direct evidence in support of the formation of a CoA-synthesizing protein complex in any organism.

1.4.2 Objective 1: Characterization of the putative CoA-SPC in *S. cerevisiae*

The first objective was to pursue the isolation of the proposed CoA-SPC from *S. cerevisae.* This was to be done by firstly purifying recombinant Cab proteins that are fused to glutathione S-transferase (GST, GST-tagged proteins) through large scale expression and purification via glutathione affinity chromatography. The isolated enzymes were used in a series of pull-down assays with yeast whole cell lysate to fish out the CoA-SPC complex and evaluate the interacting partners under denaturing (SDS-PAGE) and non-denaturing conditions (PAGE). As an alternative strategy, yeast whole cell protein extracts would be subjected to ammonium sulphate precipitation and assayed for CoA production using high performance liquid chromatography (HPLC) as a preliminary step for the isolation of the complex from such partially purified fractions.

1.4.3 Objective 2: Structural characterization of CoaBC bifunctional proteins

The known bifunctional proteins involved in CoA biosynthesis have been suggested to act as scaffolds for the formation of a CoA-synthesizing complex. Of particular interest is the bacterial CoaBC protein, which has both PPCS and PPCDC activities. For this objective, the CoaBC proteins from *S. aureus*, *E. coli* and *M. tuberculosis* were characterized using cryo-EM. This was accomplished by firstly purifying the protein via Ni^{2+}-based immobilised metal-affinity chromatography, followed by analysis using size exclusion chromatography to ensure that pure protein fractions are obtained, free from contaminating constituents that may interfere with structural determination. Single particle analysis were carried out by means of transmission electron microscopy under negative stain and cryogenic conditions for structural determination.

Chapter 2: Experimental Methods

2.1 Materials

All chemicals and solvents were of analytical grade and were obtained from Sigma-Aldrich, or Roche, unless stated otherwise. Expression vectors for overexpression and competent cell strains (*E. coli* BL21 (DE3)) were purchased from Novagen. A GeneJET Plasmid Miniprep Kit (Thermo-Fisher Scientific.) was used to recover plasmid DNA for transformations. An ÄKTA prime protein purification system (GE Healthcare.) or Glutathione Sepharose™ 4B slurry (GE Healthcare) was used to carry protein purification. HiTrap™ Chelating HP (1 mL) and Desalting (5 mL) columns (GE Healthcare.) used for protein purification were purchased through Sigma-Aldrich. The Quick Start Bradford Protein Assay Kit (Bio-Rad), containing bovine serum albumin standards and Bradford reagent was used for protein concentration determination. Large-scale (50–500 mL), medium-scale (5–50 mL) and small-scale (10–2000 µL) centrifugation was carried out using a Beckman Coulter Avanti J26 XPI, Heraeus Multifuge 3S/3S-R, and Heraeus Biofuge pico centrifuge respectively. SDS-PAGE analysis was done using either Hoefer, Inc. (USA.) or Bio-Rad PowerPac HC™ gel systems. An Agilent 1200 series system equipped with an in line FLD fluorescence detector was used for high performance liquid chromatography (HPLC) analyses.

2.2 Media and culture conditions

Maintenance of *S. cerevisiae* Y294 and *E. coli* BL21 (DE3) parental strains was carried out on YPD agar (10 $g.L^{-1}$ yeast extract, 20 $g.L^{-1}$ peptone, 20 $g.L^{-1}$ glucose, 12 $g.L^{-1}$ bacterial agar) and Luria Bertani (LB) agar plates (5 $g.L^{-1}$ yeast extract, 10 $g.L^{-1}$ tryptone, 10 $g.L^{-1}$ sodium chloride and 12 $g.L^{-1}$ bacterial agar) respectively. The plasmid vector containing the gene that encodes for *Sa*CoaBC (pPROEX-Hta base vector) were available in house. The plasmids were transformed into *E. coli* BL21 (DE3) for the heterologous expression of the protein with an N-terminal 6×histidine tag. Similarly, plasmid vectors containing the genes that encode for *S. cerevisiae* Cab1–Cab5 available from previous studies in the lab or generously provided by Hans-Joachim Schüller (all in pGEX-2T) and were transformed into *E. coli* BL21 (DE3) for expression of the proteins with GST-tags. Expression of recombinant cells used in this study were carried out aerobically in LB broth and maintained on LB agar supplemented with 100 µg/mL ampicillin.

2.3 Expression and purification

Expression and purification of *Sa*CoaBC was achieved according to previously published methods (van der Westhuyzen, 2012). For large scale overexpression of Cab proteins, overnight starter cultures were prepared in LB media (5 mL) grown aerobically in 500–1000 mL LB media supplemented with 100 µg/mL ampicillin at 37ºC until a OD_{600} value of between

0.6–0.8 was reached. Isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to induce the expression of the proteins according to the conditions listed in Table 2. After expression, the cultured cells were harvested by centrifugation at 8000 rpm for 20 min. The cell pellets were resuspended in Binding buffer (1× Phosphate-buffered saline, pH 7.3) supplemented with cOmplete™ Protease Inhibitor Cocktail (Roche) according to the manufacturer's instructions and sonicated for complete cell lysis. The cell lysate was centrifuged at 20 000 rpm for 20 min to pellet cellular debris. The supernatant containing the soluble protein fraction was retained and clarified using a 0.45 μm syringe filter (Corning®). Filtered protein fractions were purified in a batch purification utilizing the Glutathione Sepharose 4B resin according to the manufacturer's instructions. The clarified cell lysates harboring the GST-tagged Cab proteins were allowed to incubate with the resin for 1–2 h at 4 ºC using gentle end-over-end agitation. Non-specifically bound proteins were washed off the resin using Binding buffer for a total of 3 washes. To preserve protein functionality, proteins were eluted under mild non-denaturing conditions by adding an Elution buffer containing 50 mM Tris-HCl, 10 mM reduced glutathione, pH 8.0 to the resin and incubating for 5–10 min at 4 ºC. After each incubation, wash and elution step, the resin was sedimented at 500×g for 5 min and the supernatant was retained for analysis.

Table 2: Conditions used for expression of the yeast Cab proteins and their expected sizes

	Cab1	**Cab2**	**Cab3**	**Cab4**	**Cab5**
Induction Temperature	37°C	18°C	28°C	30°C	37°C
IPTG Concentration	0.5 mM	0.4 mM	0.1 mM	1 mM	1 mM
Incubation Time	Overnight	3 hours	Overnight	3 hours	Overnight
Expected Size	~ 66.9 kDa	67.8 kDa	91.2 kDa	60.3 kDa	53.3 kDa

2.4 Protein characterization

2.4.1 SDS-PAGE

Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) analysis of proteins was performed under standard methods as described by Smith, (1984). A modification of Lemberg's staining method was used for protein visualisation as outlined by Fairbanks (1925). Briefly, the initial stain (Solution A) used for gel submersion consists of solution containing 25% isopropanol, 10% acetic acid, and 0.05% Coomassie R-250. Gels submerged in this solution were heat-treated in a microwave and incubated for 30 min for

staining. Solution B (10% isopropanol, 10% acetic Acid, 0.005% Coomassie R-250), Solution C (10% acetic acid, 0.002% Coomassie R-250) and Solution D (10% acetic acid) each contain varying degrees of Coomassie brilliant blue stain and related components from Solution A to progressively de-stain the gel for visualisation. The de-staining process was achieved similarly by microwave heating and allowing it to incubate with the gel for 10 min before proceeding to the next de-staining solution.

2.4.2 Western Blot

Western blot analysis was carried out using a standard protocol. Briefly, an SDS-PAGE gel was loaded onto a nitrocellulose membrane at a constant 120 mA for 18 hours. The membrane was washed with Milli-Q® water after transferal of proteins and blocked with casein buffer (154 mM NaCl, 10 mM Tris-HCl, 0.5% (m/v) casein, 0.02% (m/v) thimerosal, pH 7.6) for 1 h at 37 ºC. Thereafter, the membrane was incubated with a primary antibody solution diluted 1:7500 in casein-Tween buffer (154 mM NaCl, 10 mM Tris-HCl, 0.5% (m/v) casein, 0.02% (m/v) thimerosal, 0.1% (v/v) Tween-20, pH 7.6) at 37 ºC for 1 h. Goat anti-rabbit horseradish peroxidase conjugate (Sigma) secondary antibody solution diluted 1:2000 in casein-Tween buffer (154 mM NaCl, 10 mM Tris-HCl, 0.5% (m/v) casein, 0.02% (m/v) thimerosal, 0.1% (v/v) Tween-20, pH 7.6) was used for further incubation of the membrane at 37 ºC for 1 h. After each incubation step, the solution was decanted, and the membrane was washed 3 times in PBS-Tween wash buffer for 5 min whilst agitating at 25 rpm on the GyroTwister (Labnet, USA). Membrane development was carried out in a solution containing 0.05% 4-chloro-1-naphtol, 16% MeOH, 1X PBS, and H_2O_2 at room temperature for 30 min through agitation. The developed membranes were rinsed in MilliQ® water for 5 min for evaluation.

2.4.3 Size exclusion chromatography

Purified *Sa*CoaBC sample was loaded onto a gel filtration column (PWXL5000 Tosoh Biosciences, Japan) equilibrated with 25 mM Tris–HCl, 5mM $MgCl_2$, 200 mM NaCl and 5% glycerol at pH 8.0. Protein samples were eluted at a flow rate of 0.5 mL/min for 1 column volume. Samples were immediately used for analysis and stored at -80 ºC. Sample fractionation was operated on a Gilson System Workcenter using a Gilson 203G fraction collector. Gilson Unipoint software was used to interpret and visualize the generated chromatograms.

2.5 Negative stain electron microscopy

2.5.1 Grid Preparation

An EMS100× Glow Discharge Unit (Electron Microscopy Sciences.) was used to glow-discharge carbon coated copper grids (Electron Microscopy Sciences.) to render the grids more hydrophilic and negatively charged at a plasma current of 25 mA for 30 seconds. The

proteins were negatively stained according to standard practices described by Booth *et al*., (2011). Three microlitres of *Sa*CoaBC protein preparation was allowed to adhere for 30–35 seconds on the glow discharged grid. For *Sa*CoaBC, grids were prepared at concentrations ranging from 0.1–5 mg/ml. Filter paper was used to blot excess sample off each grid followed by washing with 3 drops of distilled water and staining with 2 drops of 2% aqueous uranyl acetate (SPI supplies.). Grids were indirectly blotted with filter paper between each wash and staining step. Grids were allowed to air-dry before transfer to transmission electron microscopy.

2.5.2 Transmission electron microscopy and image processing

The prepared negative stain grids were loaded onto a single side-entry holder (Gatan) and imaged using a Tecnai T20 Transmission Electron Microscope (FEI) equipped with a CCD camera (2048 x 2048 pixel: Gatan) and a LaB6 electron gun operating at 200 kV. Micrographs were collected at SA 67 000× magnification at an electron dose of ~30 e/Å^2 with a defocus of approximately -1.5 µm. RELION was used for image processing and reconstruction (Scheres, 2016). Unbiased reference-free autopicking using Laplacian-of-Gaussian with a filter diameter range between 150 and 200 Å was used (Zivanov *et al.*, 2018). Picking threshold in all cases were adjusted to 0.05 with a minimum inter-particle distance of 10 nm. Particle co-ordinates were extracted from the micrographs by applying a box size of 128 pixels without downscaling. Poor particles were eliminated by multiple rounds of 2D classification (Punjani *et al*., 2017). This 2D classification process was optimized and performed using a regularization parameter T of 1.5 and applying a mask diameter of 24 nm for 25 iterations for a total of 50 classes. Image alignment was applied to sampling method with an in-plane angular sampling (rotational angle) set to 6 degrees and an offset search range and offset search step set to 5 and 1 pixels respectively. Three-dimensional reconstruction was conducted initially by generating an initial model from the selected 2D classes using RELION and the Stochastic Gradient Descent algorithm and applying tetrahedral (T) symmetry. Initial angular sampling of 15° was applied and an offset search range and offset search step adjusted for 6 pixels and 2 pixels respectively (Scheres, 2012). A mask diameter of 24 nm was applied. The resulting model was used as a reference model for 3D refinement in which the initial offset range of 5 pixels and an initial offset step of 1 pixel was used with an angular sampling of 7.5°. Data visualization and molecular dynamics using default parameters of reconstructed models were performed using UCSF Chimera (Pettersen, 2004).

2.6 HPLC analysis and quantification of CoA

HPLC analysis was carried out on an Agilent 1100 series system comprised of a Supelcosil LC-DP (250 x 4.60 mm, 5 µm particle size) reverse phase column protected by a Supelcosil LC-DP Supelguard cartridge guard column (Merck) and equipped with an in-line fluorescence

detector. Operation of the HPLC was carried out using a solvent system consisting of Solvent A (50 mM potassium phosphate, pH 8), Solvent B (60% aqueous acetonitrile) and Solvent C (100% acetonitrile) with a flow rate of 1 mL/min. HPLC workflow, sample extraction and derivatisation developed by Goosen and Strauss (2017) was implemented in this study. Yeast cell lysates prepared for HPLC analysis were carried out as outlined by Szymanski and Kerscher (2013). Prepared lysates were subjected to 4 rounds of ammonium sulphate precipitation (20%, 40%, 60% and 80%) and the recovered pellets were dissolved and filtered in PBS supplemented with 1 mM PMSF and cOmplete™ Protease Inhibitor Cocktail (Roche). Each fraction was assayed for the formation of CoA by the addition of 50 mM Tris-HCl pH 7.6,10 mM $MgCl_2$, 20 mM KCl, 500 µM Pan, 5 mM ATP, 1 mM L-cysteine and 2 mM TCEP to the protein sample (1–10 mg/mL) for a final reaction mixture of 50 µL. The reaction was allowed to proceed for 30-90 min at 37 ºC for conversion. Following incubation, the fractions were quenched with 10 µL of 90% TCA followed by neutralization with 40 µL 2.25 M NH_4OAc. The resulting particulates were centrifuged for 10 min at 13 000 rpm and 30 µL of the supernatant was recovered to use for overnight derivatization with CPM for HPLC analysis, as described by Goosen and Strauss (2017). Software package Chemstation was used for chromatogram analysis and interpretation.

Chapter 3: Attempts at the purification of the yeast CoA-SPC

3.1 Introduction

Olzhausen *et al.* (2013) explored the original proposal made by Bucovaz *et al.* (1997) that a CoA-SPC exists in yeast. This was done by systematically analysing the binary interactions of the yeast Cab1–Cab5 proteins as well moonlighting proteins Hal3 and Vhs3. Experimentally, this was accomplished by performing pairwise interactions of bacterial synthesised GST-Cab fusion proteins which were immobilised on glutathione sepharose *in vitro*, against yeast protein extracts harboring HA-tagged Cab proteins. Their results presented compelling evidence of a potential CoA-SPC in yeast as the Cab proteins were found to interact with each other. More specifically, Cab3 was shown to interact with itself and with Cab2, Cab4 and Cab5; Cab4 was unable to interact with itself but could bind with Cab2, Cab3 and Cab5. Cab2 and Cab5 were both capable of binding to Cab3 and Cab4 but could not interact with itself. Further analysis also confirmed the interaction of Cab3 with Hal3 and Vhs3; however, these moonlighting proteins were unable to interact with the remaining Cab proteins used in their study. They also demonstrated that Cab3 and Cab5 co-fractionated as a complex in a gel-filtration assay approximately 330 kDa in size.

Yeast two-hybrid assays were also conducted to further their investigations of the interactions of the Cab proteins *in vivo*. All tested combinations in this case were in agreement with what was observed in the pairwise studies. The authors also mapped the regions of these proteins that took part in these interactions. These results suggested that Cab3 acts as a scaffold for the complex, primarily via interactions with its N-domain. Altogether, this strongly suggested that the CoA-SPC complex indeed exists and that it contain at least some of the proteins interacting and varying degrees. However, as their studies only revealed binary interaction between the individual Cab proteins, we set out to investigate whether the intact complex can be purified from yeast cell extracts. This was done with the ultimate goal of structurally characterizing the complex.

3.2 Results and Discussion

3.2.1 Purification of Yeast Cab proteins

Our first strategy to purify the intact yeast CoA-SPC was to use purified and immobilized Cab proteins to fish out the CoA-SPC from yeast lysate by using pull-down assays. Plasmid vectors based on pGEX-2TK that encode Cab1–Cab5 for their expression as GST-tagged proteins we available in the laboratory or was obtained from Hans Joachim Schuller; these expression plasmids are under the transcriptional control of the tac promoter. Large scale purification of Cab protein was conducted using the conditions as shown on Table 2. These conditions were based on previous expression studies that were conducted to determine the conditions that would lead to soluble expression of these proteins. Harvested cells were purified using the method as outlined in Chapter 2.

Briefly, in all cases purification of the proteins of interest were carried out at 4°C and the clarified lysate harboring the Cab proteins were allowed to adhere to the affinity resin for at least of 30 min to allow for sufficient binding. Each purification procedure was subjected to 2–3 wash steps to remove unwanted GST-rich proteins that could be weakly bound to the resin. To preserve protein functionality, proteins were eluted under mild non-denaturing conditions using an elution buffer containing 50 mM Tris-HCl, 10 mM reduced glutathione, pH 8.0 for 5-10 min at 4 ºC. All wash steps and crude lysate were retained and included in the analysis on 12% SDS-PAGE gel to evaluate the overall purification efficiency for size, degradation, and purity of the Cab proteins (Figure 8).

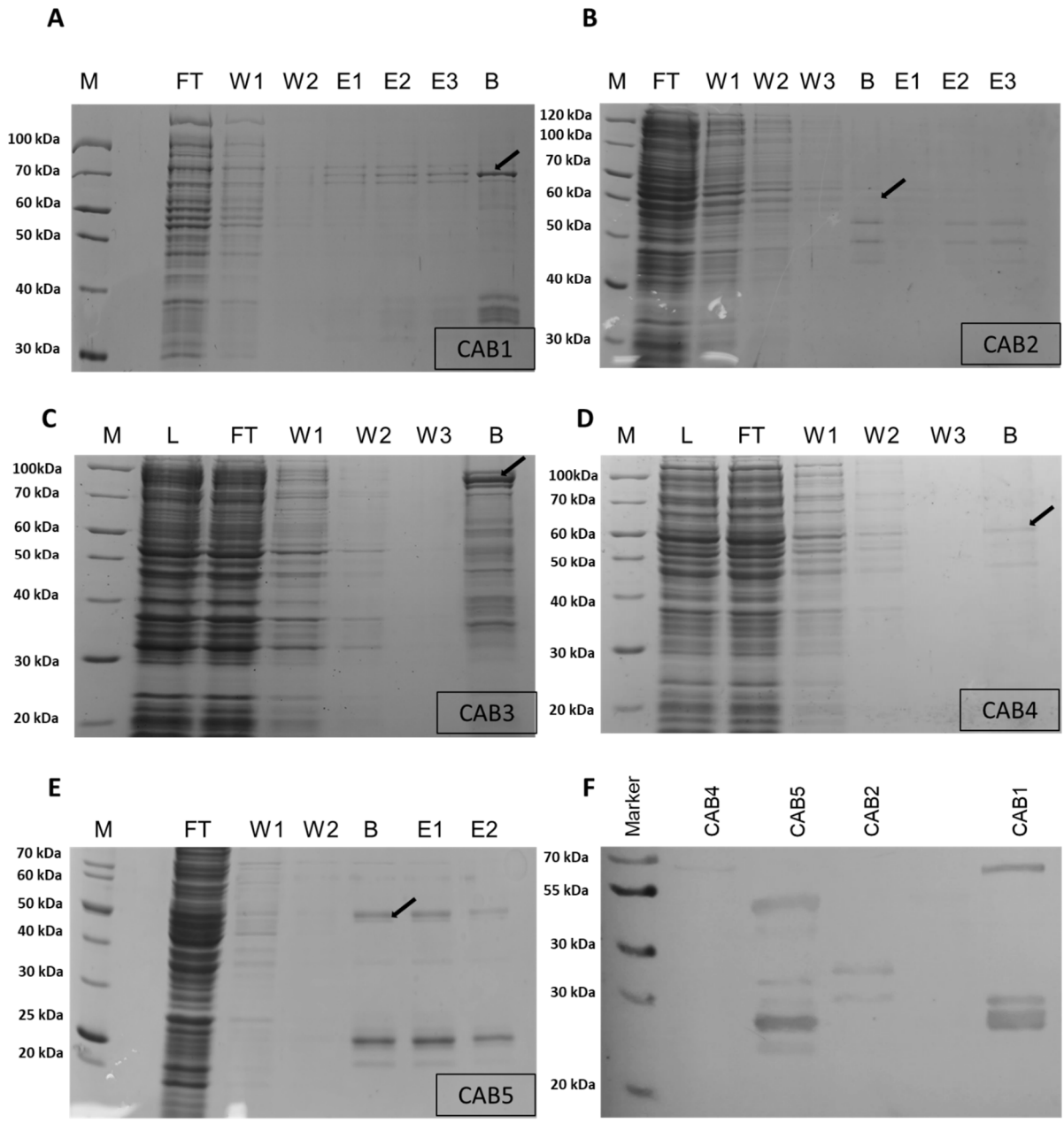


***Figure 8**: SDS-PAGE analysis of clarified bacterial lysates harbouring the GST-tagged Cab proteins. The Cab containing bacterial lysates immobilised on glutathione sepharose beads is represented as "L" while the unbound flow through indicated as "FT". In each case, the removal of proteins that are non-specifically bound was accomplished by multiple wash steps and is indicated as "W1–3". The Cab proteins of interest after wash steps are depicted as "B". Fractions E1-3 represent the eluted Cab proteins from the glutathione resin with "M" indicating the molecular weight marker. The expected molecular weight of the target protein in each case, as well as of other proteins/fragments of interest, are indicated for each gel using arrows. Western blot analysis (Anti-GST antibody) of purified Cab proteins is portrayed at the bottom right.*

Purification of Cab1, Cab3, Cab4 and Cab5 was successful as confirmed by SDS-PAGE analysis, with bands corresponding to the expected sizes accompanied in most cases by

degradation products (Figure 8A, C, D and E). The observed bands at 20–26 kDa represent GST that is bound to the resin. The bands appear lower due to the GST-fusion tag that often degrades upon denaturation and under protein gel electrophoresis. SDS-PAGE analysis of Cab2 revealed feint bands that do not correspond to the expected size of the protein. These bands might represent degradation of bound GST on the resin or other contaminating GST-rich proteins. Small scale expression trails were conducted to determine the conditions that would lead to soluble expression for Cab2. This was carried out by varying the IPTG concentrations, temperature, and incubation times but with no success as growth stagnation and low bacterial yield being the major contributing factors affecting the overall process. We also observed this from the initial repeated large-scale expression attempts which suggests that the expressed protein is exerting a toxic effect that is detrimental to host cellular functions. This has been reported in previous studies as slower growth rate, lower cellular density and death are visible results of this effect (Doherty *et al.*, 1993; Dong *et al.*, 1995). Western blot analysis was conducted as an additional confirmation step before proceeding with the pull-down assays (Figure 8F). As expected, no detectable proteins bands were observed for Cab2 other that the bands that were observed in the SDS PAGE gel (Figure 8B). The remaining Cab proteins were all detected after blotting and thus used for the pull-down assays.

3.3 CoA-SPC purification attempts by pull-down affinity chromatography

Pull-downs were carried out similarly to previously reported studies (Abrie *et al.*, 2015). In each case, Cab proteins (Cab1, Cab4, Cab3 and Cab5) were immobilized on glutathione agarose. For this, crude extracts from the cells overexpressing the protein of interest were used. This was done to reduce the amount of undesirable degradation that occurred when the Cab proteins were subjected to a full purification (Figure 8). The bound proteins were allowed to incubate with yeast lysate for 1 hour at 4°C using gentle end-over-end mixing to fish out any bound proteins, the CoA-SPC complex being the main target. After binding, the resin was sedimented and washed as previously described and evaluated under denaturing conditions.

The separation profile between the initial binding and after yeast lysate incubation was compared to observe any noticeable protein bands that could represent a potential interaction of the individual Cab proteins to the proposed CoA-SPC complex (Figure 9). The dotted box (in red) in the figures represents the profile of the initially bound protein (the bait), while the box on the right represents the protein bands after incubation with yeast lysates and subsequent wash steps. We expected to see bands corresponding to the pairwise combinations as demonstrated by Olzhausen *et al.* (2013) with Cab3 being of particular interest due to its proposed molecular scaffold nature (27–65 kDa range).

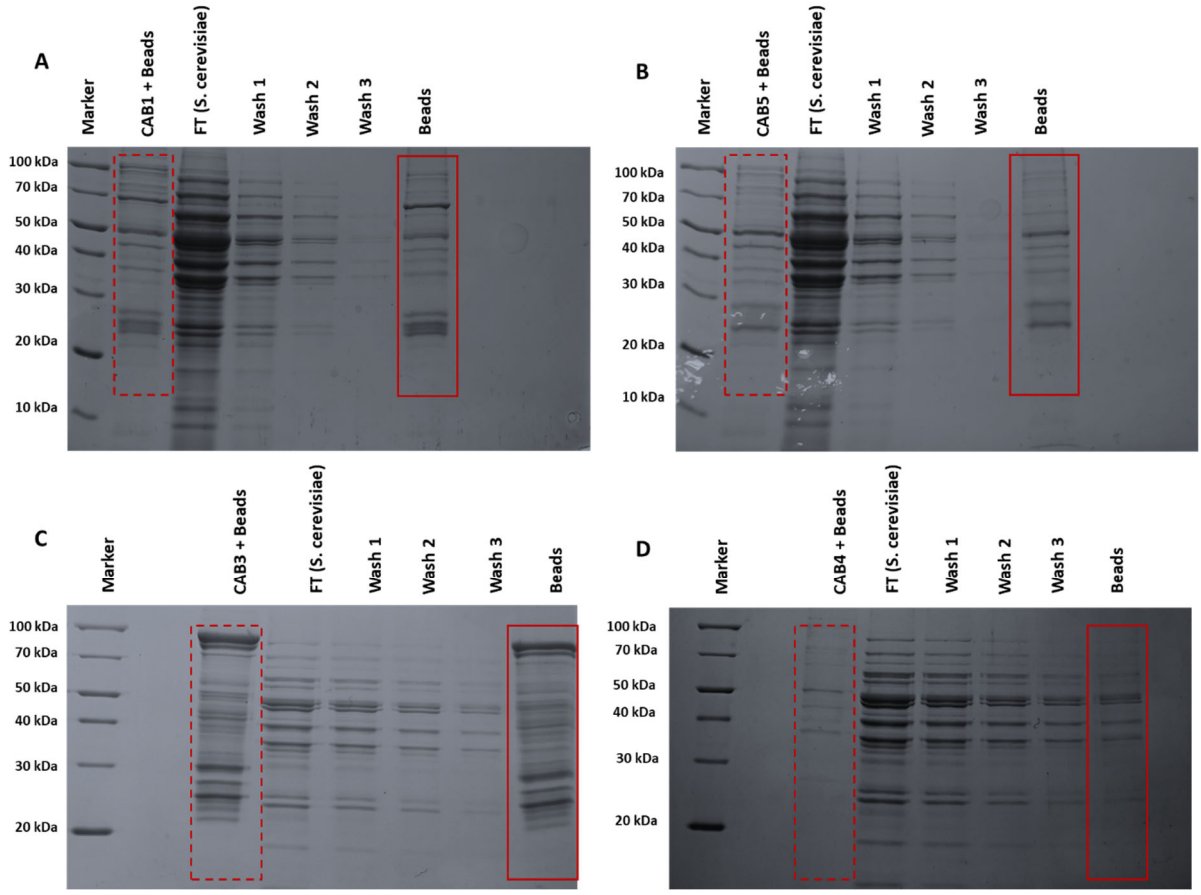


***Figure 9**: Pull-down assays of Cab proteins with clarified S. cerevisiae lysate. In each case, the GST-tagged Cab protein used as bait was immobilised on glutathione agarose beads (dotted box on left, lane marked "CABx + Beads"). Subsequently, clarified S. cerevisiae lysate was introduced to the resin; the lane marked "FT" represents the unbound flow-through. The beads were then sequentially washed three times (lanes marked "Wash 1–3"). The last lane ("Beads") represents the profile of the proteins adhered to the affinity resin after these wash steps (red box on right). Molecular weight markers are indicated as "M".*

Closer inspection of the bound proteins after incubation with the clarified yeast lysates revealed no noticeable bands indicative of potential interactions; instead, of the profiles of protein bands pre- and post-incubation are noticeably similar. As the incubation time applied for complex interactions in pull-down studies may influence the amount of bound protein, the duration that the yeast lysate was incubated with the bait protein was increased. Furthermore, adjustments were made to the washing parameters as too stringent washing may result in the loss of bound proteins if the interactions are transiently. We approached this in two ways: (1) using several quick wash steps with small volumes of washing buffer and (2) using prolonged incubation times with the prepared yeast lysate at 3 and 16 hours at 4 ºC. However, neither prolonged incubations times nor changes to the wash steps had any observable effect on the

outcome, as similar separation profiles were still observed in all these cases with no apparent bands that could portray potential interactions. We therefore attempted to establish the existence of a CoA-SPC by other means.

3.3 CoA-SPC purification attempts from yeast lysate

As we were unable to fish out the CoA-SPC by means of pull-down assays using the individual Cab proteins, efforts were made to purify the CoA-SPC from whole yeast lysate. We approached the whole cell isolation method by performing a series of ammonium sulphate precipitation reactions followed by the determination of CoA formation by each purified fraction to established whether the CoA-SPC was enriched. This approach was followed as the only report of the isolation and characterization of the CoA-SPC did not provide information on its enrichment during the followed purification steps (Bucovaz *et al*. 1997). Yeast lysates were prepared and fractionated by four rounds of increasing precipitation using ammonium sulphate at 20%, 40%, 60% and 80% concentrations. All operations were conducted at 4°C to prevent protein degradation.

Following fractionation, activity assays were carried out to determine the formation of CoA by the proteins in each fraction from an assay mixture in which all the components required for CoA biosynthesis were present. Approximately 10 mg/ml total protein were used from each preparation and the reaction was allowed to continue for between 60 min to allow for full conversion. The yeast lysate used for the fractionation was included as a reference. Following incubation, the fractions were quenched with 10 µL of 90% TCA, followed by neutralization with 2.25 M NH_4OAc. Samples were subsequently derivatized with the thiol-reactive fluorescent probe CPM for allow for CoA analysis according to an established method (Goosen and Strauss, 2017). As we were only interested in CoA formation as indicator of the presence of the CoA-SPC, standards were prepared for CoA and the expected retention time for this analyte was established to be at approximately 8 min. HPLC analysis of the yeast lysate and ammonium precipitated fractions revealed no detectable peaks of CoA in the resulting chromatograms (Figure 10).

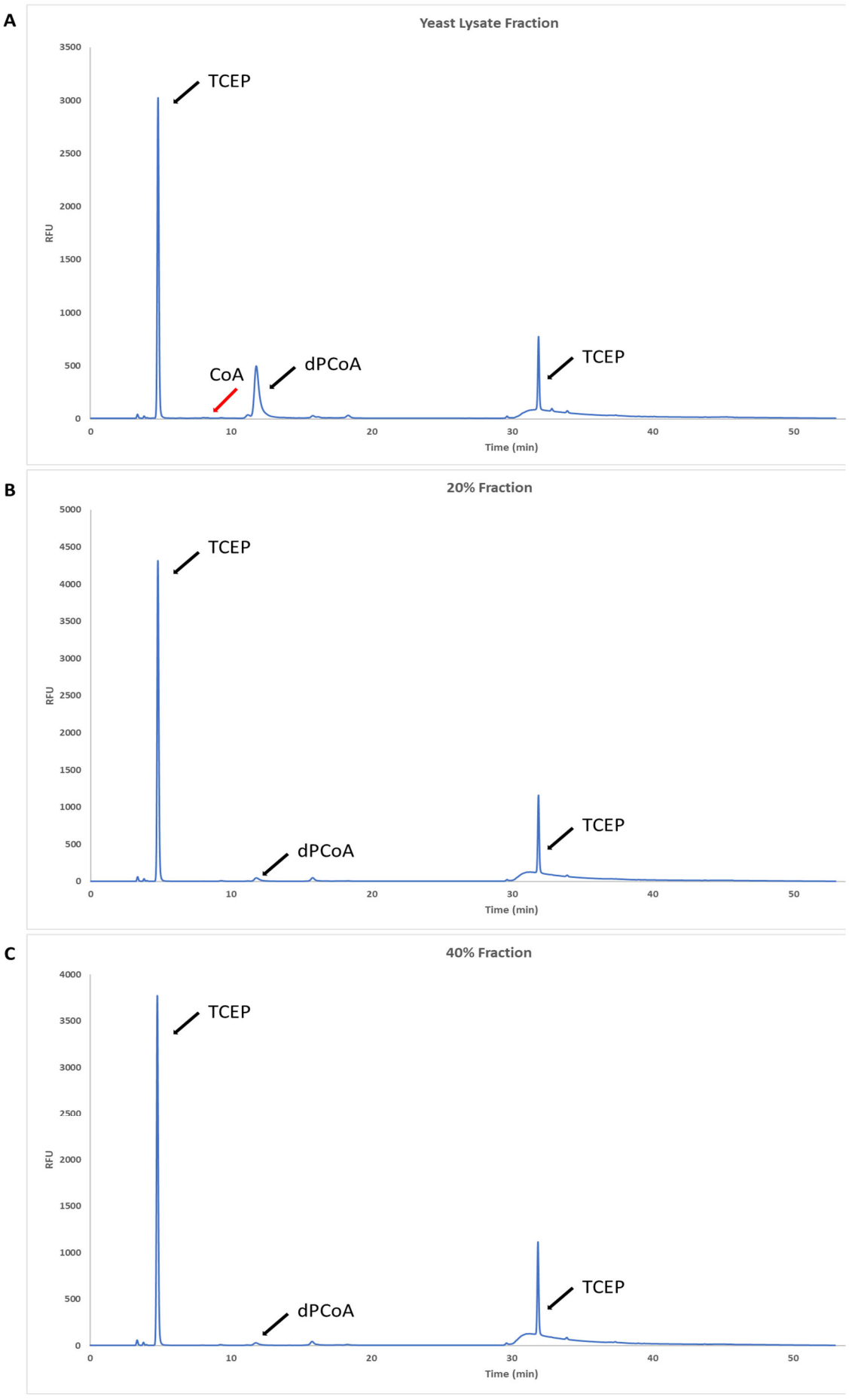


***Figure 10 (previous page)**: HPLC analysis of CoA formation to establish the enrichment of protein samples in the putative CoA-SPC. CoA formation was analysed by derivatization with a thiol-reactive fluorescent probe, followed by separation on an LC column. Emission of fluorescent derivatives was detected at 465 nm and excitation at 387 nm. Chromatograms represent the analysis of CoA formation by proteins present in prepared yeast lysate (A), the 20% $(NH_4)_2SO_4$ precipitated fraction (B) and the 40% $(NH_4)_2SO_4$ precipitated fraction (C). Following the incubation of each protein in an assay mixture containing 50 mM Tris-HCl pH 7.6, 10 mM $MgCl_2$, 20 mM KCl, 500 µM pantothenate, 5 mM ATP, 1 mM L-cysteine and 2 mM TCEP for activity determination. CoA is expected to elute at 8 min under these conditions. Other peaks representing known analytes are labelled.*

The first pronounced peak represents TCEP as it included within the reaction mix and frequently detected in the method. A noticeable peak around the 11 min mark was observed from the yeast lysate reaction which corresponds to the retention time for dephospho-CoA observed in previous analyses. However, as this metabolite is not as readily available as CoA, standards were not prepared for these analyses to confirm its identity. A small peak at the same retention time was also observed in the samples incubated with the ammonium sulphate precipitated fractions. Finally, a peak was observed in each fraction at 32 min. Its identity is unknown but is likely related to TCEP as its ratio to the TCEP peak remained constant in all the samples.

Following this result the activity assays were repeated as a time course assay over a period of three hours and six hours. However, HPLC analysis of these reactions gave a similar result to the first analysis. To account for the possibility that the CoA-SPC is only present at very low concentrations, the total protein concentration used in each reaction was increased (>100 mg/ml) and the reactions assayed for CoA formation after 30 min, three hours and six hours. In all these experiments we were unable to detect noticeable peaks for CoA formation, the putative dephospho-CoA peak being the only one being observed.

3.4 Conclusion

We were unsuccessful in our attempts to fish out the CoA-SPC by pull-down studies or to provide evidence of its existence in partially purified lysates. The absence of evidence of CoA formation even in the original lysates could indicate that the assays were not conducted correctly, or that issues with Cab protein stability arise in the preparation of lysates. The putative observation of dephospho-CoA being formed by yeast lysates suggests that at least Cab1, Cab2, Cab3 and Cab4 are present in active form in the lysates. However, our attempts at the purification of Cab5 showed that it is prone to significant degradation, and this might be the reason for the failed conversion of dephospho-CoA to CoA. In addition, the degradation of Cab5 might also affect the stability of a putative CoA-SPC.

The degradation of Cab5 and its impact on CoA-SPC formation could also explain why we were unable to purify the complex using the pull-down method. However, in the case of pull-down assays it is also likely that the presence of the GST-fusion tag interferes with complex formation due to its relatively large size compared to other expression tags. Efforts can be thus made by using tags such as a poly-histidine tag as it usually does not interfere protein interactions due to its small size, and simple and robust structure (Graeslund and Hammarstroem, 2013). Introduction of poly-histidine tag may in turn improve purification for both Cab2 and Cab4 to allow them to be included in future investigations (Zhao *et al.*, 2013).

As the isolation and characterization of this complex were not successfully performed, investigations into conditions that would lead to the formation of the putative CoA-SPC can be carried out as future work. External factors such as nutrients, growth factors and stress signalling could have a direct impact on its formation and abundance within the yeast cells. For example, interactions of the human CoASy protein with proteins from signalling pathways are coordinated by serum stimulation or starvation in response to stresses (Zhyvoloup *et al.*, 2003; Nemazanyy *et al.*, 2004; Breus *et al.*, 2009; Breus *et al.*, 2010; Gudkova *et al.*, 2012). Once these issues are resolved, native MS techniques coupled to other biophysical methods such as affinity pull-downs, single particle electron microscopy analysis, X-ray crystallography and other solution-based methods will be important to further investigate the structure and dynamics of the putative CoA-SPC.

Chapter 4: Biophysical characterization of CoaBC proteins

4.1 Introduction

In bacteria the CoaB and CoaC proteins of the CoA biosynthetic pathway are encoded by a single gene to produce a fused bifunctional enzyme CoaBC (*coaBC* gene product, previously known as *dfp*) harbouring both PPCS and PPCDC activities. It has been previously proposed that the structural organization of CoaBC proteins is dodecameric based on gel-filtration experiments and by comparison to enzymes belonging to the homo-oligomeric flavin-containing decarboxylase (HFCD) protein family (Kupke *et al.*, 2000). These proteins include the lantibiotic-synthesizing enzyme EpiD and the *Arabidopsis thaliana* flavoprotein AtHAL3a, a protein related to plant growth and salt and osmotic tolerance (Albert *et al.*, 2000; Kupke *et al.*, 2000). Together they share a flavin-binding motif, have active sites with conserved histidine and cysteine residues implicated in catalysis, and are known to form homo-oligomers.

The organisation of these proteins is homo-trimeric (AtHAL3a) or feature a homo-dodecameric arrangement with trimers disposed on the vertices of a tetrahedron (EpiD) with each of the monomers consisting of a single domain with a Rossmann-type fold. The molecular characterization of EpiD, which catalyses the oxidative decarboxylation of peptidylcysteines to peptidyl-aminoenethiols, formed the foundation for studying the enzymatic activity of the CoaBC protein from *E. coli*. It was demonstrated that the CoaC domain of the *E. coli* CoaBC protein shows sequence homology to EpiD and also with AtHAL3a. Although CoaBC shares similarities to HFCD protein family in terms of its organization, it is unique in that it contains more than one domain (Blaesse *et al.*, 2000).

It has been shown that the CoA biosynthetic pathway is regulated through feedback inhibition of PanK, PPAT and the PPAT-domain of CoASy (Vallari *et al.*, 1987; Yun *et al*., 2000; Miller *et al*., 2007). Drug vulnerability studies of CoaBC from *M. tuberculosis* recently validated this protein as a bactericidal target in the CoA biosynthetic pathway. It was demonstrated that knockdown of the expression of the CoaBC from *M. tuberculosis* had bactericidal growth effects both in culture as well as in *in vivo* mouse infection models (Evans *et al*, 2016). Recent reports indicated that the CoaBC from *M. smegmatis* (*Msm*CoaBC) is also regulated by several CoA thioesters that inhibit CoaB activity and act competitively towards the binding of CTP and 4′-phosphopantothenate and non-competitively towards the binding of L-cysteine (Mendes *et al,* 2021).

Structures of the individual CoaB and CoaC enzymes from several organisms have been reported but recently the overall full-length structure of *Msm*CoaBC was solved at 2.5 Å, showing it to exclusively exhibit a dodecameric assembly (Figure 11a). Furthermore, the fused arrangement of the CoaB and CoaC domains as depicted in their work demonstrates that the

CoaC core is important for CoaB dimerization. A flexible loop that can be seen covering and extending away from the 4′-phosphopantothenate binding site to the CoaC active site might also help to direct the CoaB product to the CoaC active site more effectively (Figure 11b).

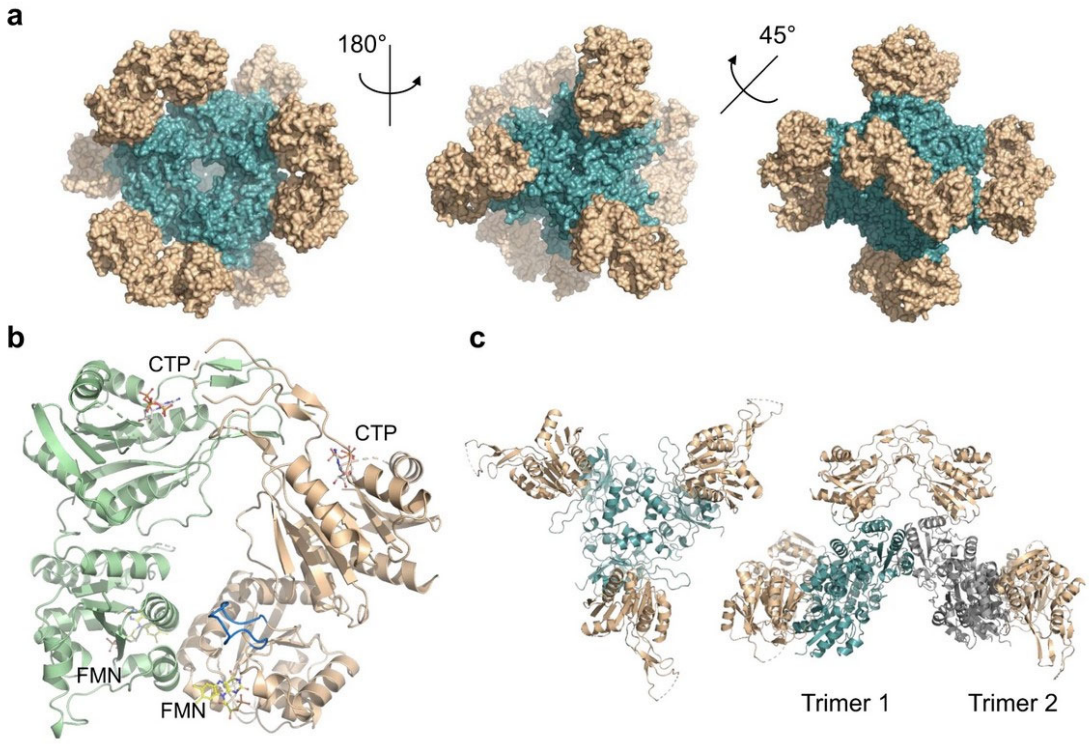


Figure 11*: X-ray crystal structure of FMN- and CTP-bound M. smegmatis CoaBC (MsmCoaBC).* ***Panel a:*** *Dodecameric MsmCoaBC with CoaC and CoaB domains illustrated in teal and in gold respectively.* ***Panel b:*** *View of a CoaBC dimer bound with FMN and CTP. Active site flexible flap of CoaC is emphasized in blue and with each protomer being coloured differently.* ***Panel c*** *Left: representation of the CoaBC trimer in which CoaC is coloured in teal and CoaB in gold. Right: Dimerisation of two CoaBC trimers is shown with CoaC coloured in teal or grey for different trimers. Each CoaB forms a dimer with protomers from different trimers. Image reproduced from Mendes et al., (2021). CC-BY 4.0.*

To the best of our knowledge, no evidence has been reported that suggests the formation of a multienzyme CoA biosynthetic complex in bacteria. However, the report of the structural organization of this enzyme in *M. smegmatis* and the recent findings detailing CoaBC to be a key vulnerability point in the CoA pathway indicates that it may act as a crucial scaffold for the complexation of the CoA biosynthetic enzymes (Evans *et al.*, 2016). Thus, to further explore the idea of a CoA biosynthetic complex operating in bacteria, this study aimed to characterize the structural diversity between related CoaBC proteins as reference for future studies in this regard.

4.2 Results and Discussion

4.2.1 Purification and negative stain transmission electron evaluation of *Sa*CoaBC

In this study we focused on obtaining the structures for related CoaBC proteins from *S. aureus* (*Sa*CoaBC), *M. tuberculosis* (*Mt*CoaBC) and *E. coli* (*Ec*CoaBC). Previous crystallography-based efforts were unsuccessful in determining the structures of these proteins. We thus explored the use of single particle analysis for structural determination. Our investigations firstly employed negative stain electron microscopy (NS-EM) which offers an observation of particles by the introduction of contrast enhancing stains such low-concentration solutions of heavy metals (typically uranium salts). Although limited in resolution to a maximum of ~18 Å NS-EM is quite useful for the study of a variety of biological problems as it allows for the evaluation of conditions such as homogeneity, dispersity, size, and shape of the proteins of interest in our samples, to establish their suitability for further study by cryo-EM.

Sample purity greatly influences 2D and 3D image classifications for negative stain evaluation, as both the stain depth and sample concentration are critical to generating reliable data. Consequently, we first evaluated the purity of the proteins under denaturing conditions. Using the standard practices as reported in Chapter 2, we successfully expressed and purified our proteins of interest. SDS-PAGE analysis revealed protein bands at the expected sizes for *Ec*CoaBC (43 kDa), *Sa*CoaBC (44 kDa) and *Mt*CoaBC (43 kDa). However, all three samples contained noticeable contaminating bands which would hinder the structural determination process (Figure 12). As a result, size exclusion chromatography (SEC) was used to further clean up the samples, which were subsequently evaluated under NS-EM. Due to time constraints, and because some structural information on the CoaB domain of the *Ec*CoaBC and *Mt*CoabC proteins have been reported, the results detailed in this section only includes the work done for *Sa*CoaBC, for which no structural data is currently known.

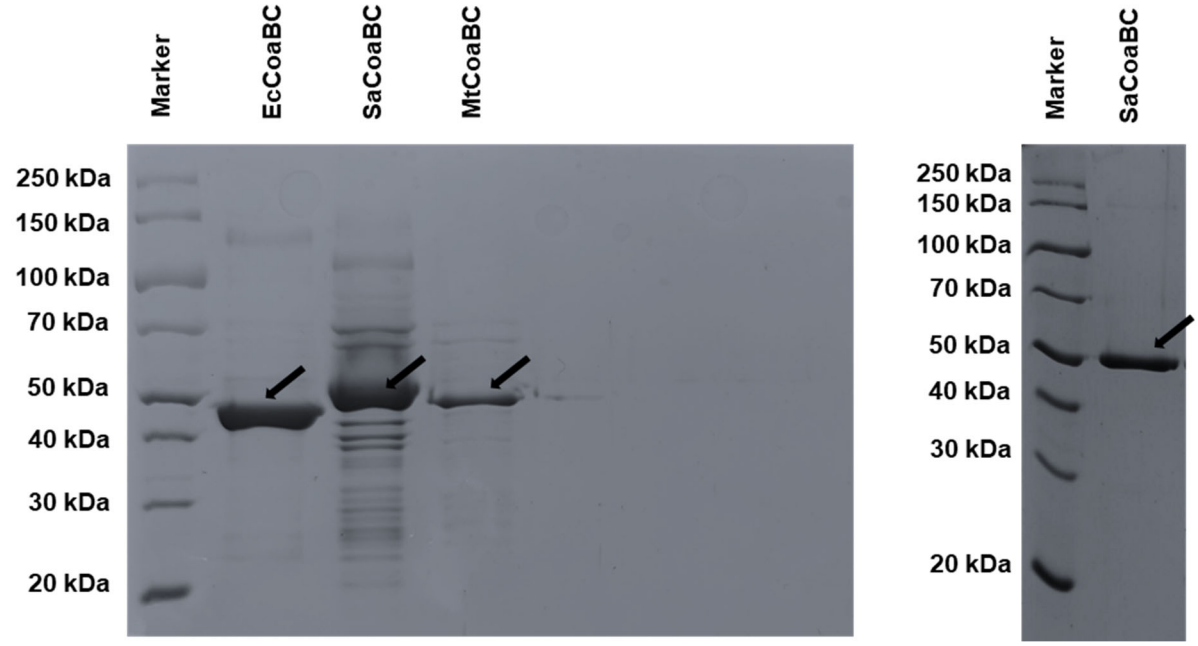


***Figure 12**: SDS-PAGE analysis of purified bifunctional CoaBC proteins. Left) SDS-PAGE analysis of His-tagged versions of EcCoaBC, SaCoaBC and MtCoaBC following purification by immobilized metal affinity chromatography (IMAC). Right) SDS-PAGE analysis of SaCoaBC after further purification by size exclusion chromatography. Proteins of interest are indicated with arrows.*

The grids harbouring the *Sa*CoaBC were prepared and stained using the conventional negative staining procedure and assessed in all cases at a magnification of 67 000× with an applied EFTEM filter. Alterations were made to the general staining protocol to achieve the best results for 2D and 3D classifications. This includes both decreasing the wash steps to prevent possible protein complex degradation and increasing uranyl acetate incubation times for improved contrast for imaging. The proteins prepared were selected from the leading edge from the size exclusion chromatogram and evaluated at concentrations ranging from 0.1–1 mg/mL.

A concentration of 0.1 mg/mL was found to be the most suitable concentration to use in this case as the evaluation of the samples after SEC revealed evenly distributed particles approximately 15 nm in length that are void of aggregated species when compared to the protein sample prior to SEC (Figure 13). Furthermore, sample preparation immediately following purification greatly facilitated this result as cryo-preserving samples lead to unwanted degradation/aggregation. Clear distinctions between the distributed particles could be made by visual inspection from the micrographs as some show clear 3-fold symmetries while others appear round and featureless. These distinctions were not previously seen from our previous

evaluations and allowed us to further our investigations. Multiple micrographs were captured and imported into REgularised Likelihood OptimisatioN (RELION 3.1) for processing.

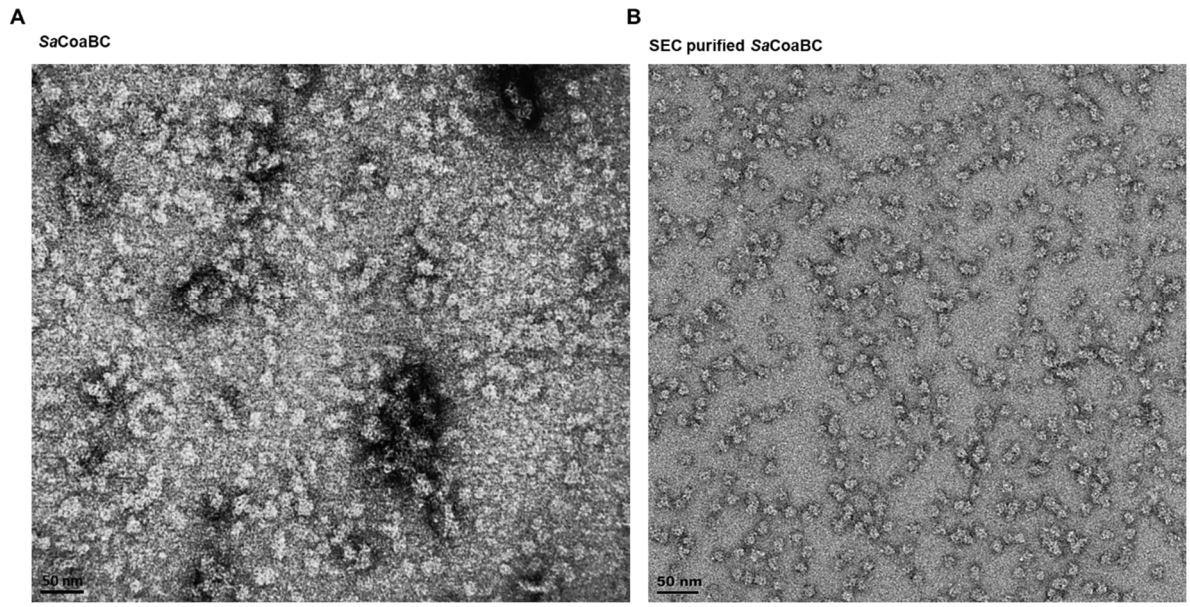


Figure 13: *NS-EM evaluation of CoaBC proteins.* ***A****) NS-EM evaluation of SaCoaBC following purification by IMAC (0.1 mg/mL),* ***B****): NS-EM evaluation of IMAC-purified SaCoaBC after further purification by size exclusion chromatography (0.1 mg/mL). A 50 nm scale bar is represented at the bottom left.*

4.2.2 Two-dimensional classification of *Sa*CoaBC

As stipulated in Table 3, the samples were imaged and collected at a dosage of 36.4 $e^-/Å^2$ to minimize radiation damage with a defocus range between -1.0 to -2.0 µm and an exposure time of 1.4 seconds. Micrographs were imported into RELION 3.1 without CTF correction. Raw micrographs had a pixel size of 3.2 Å and images displaying noticeable microscope drift, poor staining and astigmatism were excluded from the computational analysis. Particles from the micrographs were selected in an unbiased manner by reference-free auto-picking and applying Laplacian of Gaussian filtering with a diameter range of 15-20 nm. This eliminates biases that are imposed when using template-based picking or when manually picking particles. Manual picking biases the data towards particles that have recognizable symmetries (Maco *et al.*, 2011; Kastritis *et al.*, 2017; Ho *et al.*, 2020) while template-based picking introduces the risk of “Einstein from noise” (Henderson, 2013; Verbeke *et al.*, 2018; Verbeke *et al.*, 2020). This model bias phenomenon leads to a situation where the averages of the picked particles reproduce the reference image that is used for particle picking even if it only contains background noise.

Table 3: NS-EM data collection parameters and refinement statistics for low resolution dataset of *Sa*CoaBC with and without the addition of an inhibitor

	***Sa*CoaBC**	***Sa*CoaBC + Inhibitor**
Data collection parameters		
Microscope	T20	T20
Detector	CCD camera (Gatan)	CCD camera (Gatan)
Voltage (kV)	200	200
Pixel Size (Å)	3.2	3.2
Defocus Range (μm)	-1.0 to -2.0	-1.0 to -2.0
Exposure time (s)	1.4	1.4
Dose ($e^-/Å^2$)	36.4	36.4
Number of images	131	110
Magnification	67 000×	67 000×
Map Reconstruction		
Initial particle number	105 736	95 857
Final Particle Number	30 427	24 605
Box size (pixels)	128	128
Applied symmetry	T	T
Resolution (Å)	20	22

Picking threshold in all cases were adjusted to 0.05 with a minimum inter-particle distance of 10 nm. Particle co-ordinates were extracted from the micrographs by applying a box size of 128 pixels without downscaling. On average, over 800 particles were picked per micrograph and the resulting particle co-ordinates were subjected to 2D classification. The 2D classification procedure was used to classify the extracted particles into 2D class averages based on projection image matching (Scheres, 2012). In essence, the classification procedure is based on the comparison of pixel intensities between images. The 2D classification method curates the dataset, organizing particles that share similar views to enhance the signal-to-noise ratio. It is important to note that class averages derived from the 2D classification process may be composed of particles that have different conformations, but that are organised in a similar manner when compared to other particles that are included in the process. Therefore, the end image for a class average may produce a physically impossible

conformation. To account for this, multiple rounds of 2D classification are important as each round discards artifacts that do not reflect the particle's structure (Gallagher *et al.*, 2019). In this work, four successive rounds of reference-free 2D classification were enforced to eliminate false-positive picks, neighbouring and partially resolved particles after each round of classification which could influence image alignment.

Protein symmetry was determined by assessing the symmetry of sorted 2D class averages by evaluating the in-plane rotational symmetry. We observed noticeable 3-fold projections from the 2D class dataset (Figure 14A). From these classes we observed many harboring distinct "protrusions" along the vertices surrounding the central projection density. Many of the projections appear to average between 1-3 protrusions with some being more recognizable than the rest. We also observed projections from the 2D classes that are rounded and have no recognizable features/protrusions surrounding the central core. Furthermore, a noticeable area of reduced density is centered at the core of all the generated 2D classes.

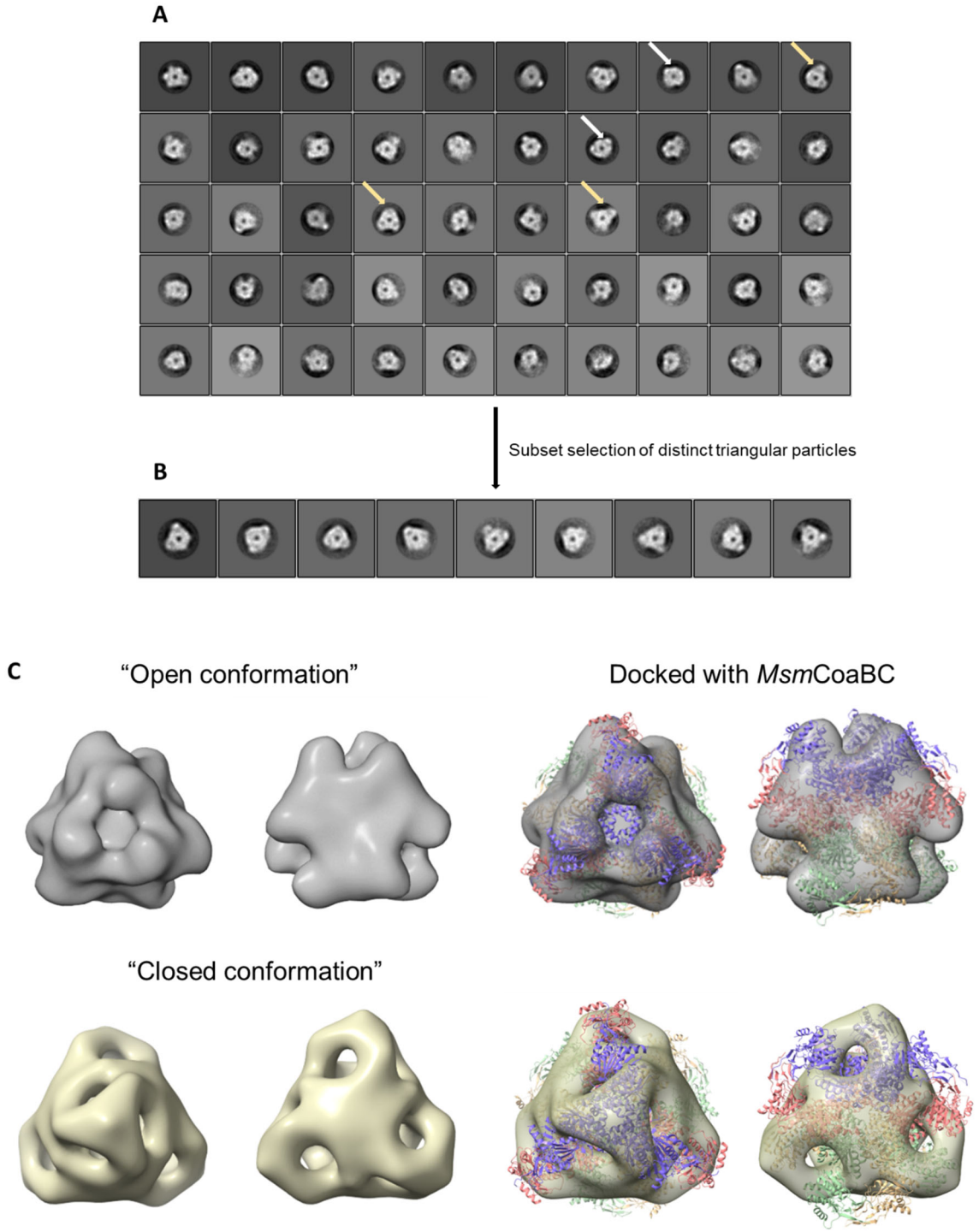


***Figure 14**: 2D Classification and 3D reconstruction of SaCoaBC. **A)** 2D class averages of SaCoaBC distributed over 50 classes after four successive rounds of in silico purification. White arrows illustrate rounded classes whereas yellow arrows represent distinct classes that have "protrusions" along the vertices. **B)** Subset selection of 2D class averages representing 3-fold projections. **C)** Left: Tetrahedral reconstruction produced after 3D refinement. Open and closed conformation illustrated along the 3-fold axis in grey and yellow respectively. Right: Docking of MsmCoaBC illustrated along the 3-fold axis.*

4.2.3 3D reconstruction of *Sa*CoaBC in negative stain and its comparison to *Msm*CoaBC

We assessed the in-plane rotational symmetry of 2D classes and applied particle symmetries in our reconstructions that were consistent with all views. To obtain a full 3D reconstruction from the data set, we first divided the 2D classes into two discrete groups based on the presence or absence of the protrusions. We selected and grouped the first set of 2D classes that represent an overall 3-fold symmetry containing at least two of the three protrusions (Figure 14B). The secondary group selections were made from the average of all the states containing both the 3-fold classes and those that appear round and featureless.

Three-dimensional reconstruction was conducted by first generating an initial model from the selected 2D classes using RELION and the Stochastic Gradient Descent algorithm and applying tetrahedral (T) symmetry (Punjani *et al.*, 2017; Zivanov *et al.*, 2018). The chosen symmetry was applied based on the information from the *Msm*CoaBC structure that exclusively exhibits tetrahedral symmetry. The reconstructions from the two datasets produced two distinct conformations at 20 Å as illustrated in Figure 14C. The models refined from the classes representing the stubbed 3-fold selection produced what we termed the "*closed conformation*", with the protrusions equating to the four points on the tetrahedron. These are absent from the models that were produced when refining classes representing the average of the projections, and these we termed the "*open conformation*". We retrieved the structure of *Msm*CoaBC from the Protein Data Bank (PDB ID: 6TGV) for structural comparison against the two generated reconstructions.

The overall structure of *Msm*CoaBC resembles a tetrahedron with CoaB positioned at the six edges of the tetrahedron where it dimerizes with another CoaB belonging to an adjacent trimer. The CoaC domain of *Msm*CoaBC exhibits a dodecameric arrangement similar to that seen in EpiD. Docking studies of the *open conformation* of *Sa*CoaBC closely resembles the crystal structure of *Msm*CoaBC, as illustrated along the 3-fold axis (Figure 14C). The CoaB dimer interfaces are at the vertices with the CoaC centered at the core of the density map. Conversely, the *closed conformation* is surprisingly different and the *Msm*CoaBC crystal structure cannot be incorporated fully into the map. The vertices where the CoaB dimer resides appear stretched with areas that are void of density as depicted in the figure above, with distinct protrusions at the four points of the reconstruction. These points correlate to the observed protrusions from the 2D classes that appear to be partially disjointed along the vertices from the central 3-fold density. We reprojected the *Msm*CoaBC structure and were unable to reproduce an orientation aligning to the *closed conformation*. Comparisons made by merging of the two conformations revealed a possible tightening and loosening effect of *Sa*CoaBC which may impact the overall reconstructions that was generated from this dataset. We evaluated this effect by applying a morph command between the two reconstructions. The

transition between the two states shows conformational changes that could occur along the density surrounding the CoaB interface that shifts towards each point of the reconstructed tetrahedron.

Previous studies have demonstrated that several CoaB dimer interfaces have an undefined loop. The diversity of this region between CoaB proteins may impact the overall interaction between these regions allowing for the transition to take effect. The structural fold of the CoaB domain of *Msm*CoaBC is consistent with the structures of several eukaryotic and bacterial CoaB proteins which displays a Rossmann-fold. Unlike the CoaB domain of *Ec*CoaBC and the CoaB proteins from many eukaryotic organisms that are capable of dimerization in the absence of CoaC, interactions between CoaC and CoaB in *M. smegmatis* is fundamental for CoaB dimerisation and activity. If this is also the case for *Sa*CoaBC, the easiest way for the protein to rearrange into the *closed conformation* would be if the monomers from the CoaB domain dimers hinge upon the loop and form a trimer (Figure 15B). In this way it will equate to the observed four-point interactions while conserving the overall tetrahedral symmetry and without affecting the CoaC core. Upon closer inspection of the *Msm*CoaBC dimer interaction, we have uncovered that it is held together by only a few hydrogen bonds (Figure 15C). Disruption in this area may allow for the transition of the *open conformation* into the *closed conformation* forming the interactions at each point.

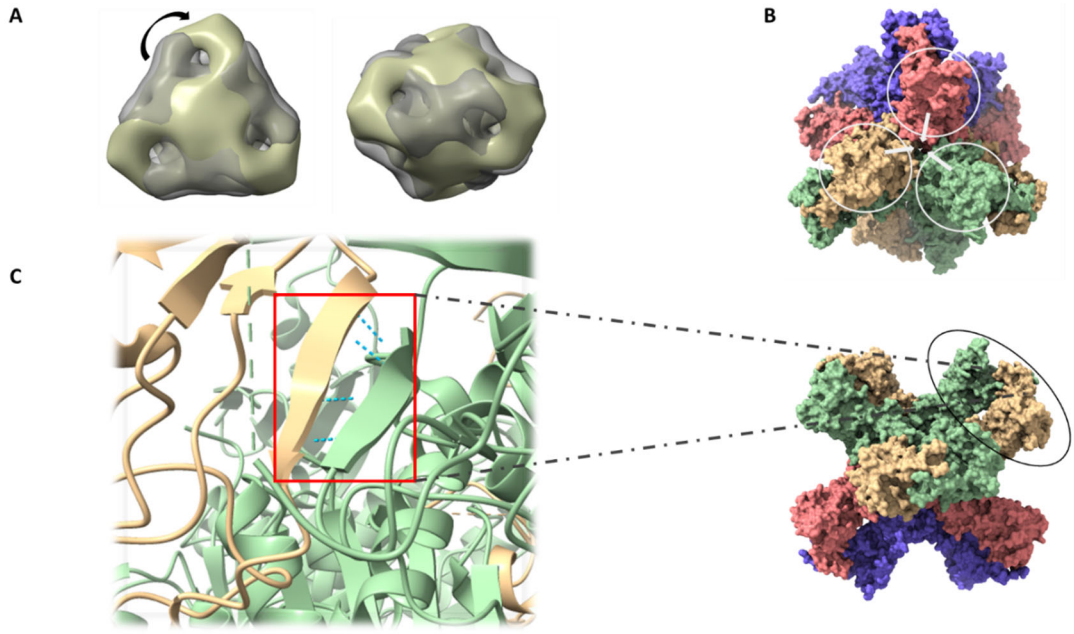


***Figure 15**: Comparison of open and closed conformation from SaCoaBC reconstructions. **A** Merging of closed (gold) and open conformation (grey) maps of SaCoaBC illustrated along the 3-fold and 2-fold axis. **B** Surface view of MsmCoaBC structure with monomers of the CoaB dimers circled in white. **C**: Emphasized image of CoaB dimer interface of MsmCoaBC. Red box illustrates H-bonding (blue dashed lines) where the CoaB dimer interface resides.*

4.2.4 Conformational validation by 3D classification

If the proposed conformational changes are derived from the disruption at the dimer interfaces, this would also mean that the CoaB domains would likely impact one another in the conformational change. Specifically, if one of the CoaB domains change from the *open conformation* to the *closed conformation*, it would lower energy barrier to those adjacent to it to form trimer interactions at each point/apex of the tetrahedron trimers. Consequently, there will be a higher likelihood for the remaining CoaB domains to interact and assume the *closed conformation*.

We therefore investigated the possible outcomes of the CoaB domains forming trimer interactions at each point/apex by performing a 3D classification of the dataset to computationally purify the different conformations. Briefly, the 3D classification process aims to iteratively categorize particles according to their similarities and is essentially used to differentiate between different conformations of the particle of interest. Unlike 2D classification that marginalizes over in-plane orientations, 3D classification contributes to both the orientational and class assignments of the particle images to treat sample heterogeneity.

Provided that CoaBC forms a tetrahedral complex that equates to four points where the proposed CoaB monomers interact to form the trimers as seen in the *closed conformation* (which has not yet been seen before in any CoaB structure), and another that corresponds to the open conformation (similar to that seen in the *Msm*CoaBC crystal structure), our classifications were limited to a small set of possible conformations assuming that the 3-fold axes are symmetric:

- CoaB interaction at **1** point: 4 possibilities related by tetrahedral symmetry
- CoaB interactions at **2** points: 6 possibilities related by tetrahedral symmetry
- CoaB interactions at **3** points: 4 possibilities related by tetrahedral symmetry
- CoaB interactions at **4** points: 1 possibility
- No CoaB interactions at any points: No point 3-folds

We approached this classification in a supervised manner by applying C3 symmetry using the best "*open conformation*" starting model and the best "*closed conformation*" starting model to computationally isolate the possible five structures. This was implemented to the entire dataset and to the subset selections of the 2D classes. In this way it also allowed us to test the overall consistency of the dataset between the two groups and whether there is a trigger whereby one assumes the open/closed point interactions in which the others follow. C3 symmetry was applied to account for the tetrahedral structure imposed in our refinements.

The outcome of this procedure is purely a geometric result with the aim to simply separate the differences at the points of the tetrahedron. Refinements and postprocessing were beyond the scope of this investigation and as a result, not implemented after the 3D classification process. The classes generated from the 3D classification procedure show a more defined density for one of the points while lacking density for the rest (Figure 16). This was observed from all classes displaying no apparent architectural changes that could indicate the proposed conformational change of the CoaB domains. Overall, while this classification procedure is feasible, the data generated could not provide enough evidence to determine if the conformational change exists. The particles of the dataset may have adopted a preferred orientation or it is has manifested as a result of a shortage of different 2D class averages.

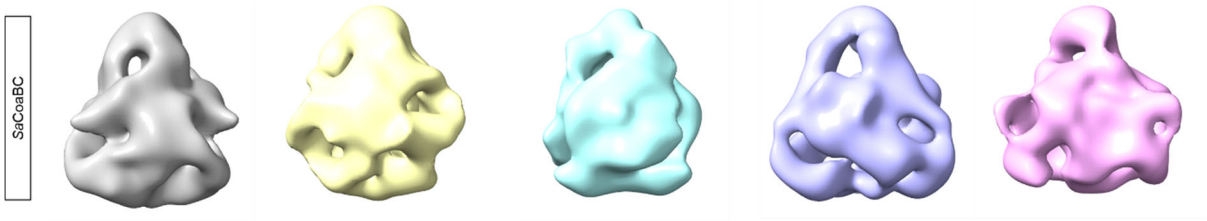


***Figure 16**: Supervised 3D Classification (without refinements) of SaCoaBC distributed over 5 classes. The classes were generated from entire 2D classification dataset in which both the "closed" and "open" reconstructions were used as reference to address sample heterogeneity.*

4.2.5 Data processing and assessment of SaCoaBC coupled with P-CJ

Following the previous results, we wished to further explore whether the introduction of 4'-phospho-CJ-15,801 (P-CJ), the active form of the known PPCS inhibitor CJ-15,801, will have impact on the overall particle distribution and end reconstruction. P-CJ was previously shown to be a potent, tight-binding inhibitor of the CoaB domain of several CoaBC proteins upon reaction with CTP. In the case of *Sa*CoaBC, treatment with P-CJ and CTP significantly stabilized the protein in thermal denaturation assays compare to the protein treated with its natural substrates only. We hypothesized that the reconstructions should favor the *open conformation* due to the tight binding interactions of P-CJ that act upon the CoaB domain, preventing it from forming the trimeric interactions required to form the *closed conformation*.

The parameters used for data collection and processing of the *Sa*CoaBC treated with P-CJ and CTP is identical to that used for the apo-structure. Five rounds of 2D classifications were carried out and then distributed over 50 classes for evaluation. From this we could clearly see many classes harboring distinct protrusions that are more noticeable and in some cases appear to be more disjointed from the central density (Figure 17A). Furthermore, a fraction of the 2D classes appear to be fuzzy/blurred accompanied by stain clouds. Additionally, the central density from some of the 2D classes appear to be less defined when compared to the

2D classes from the untreated *Sa*CoaBC sample. Similarly, as detailed in the 2D classes from the untreated *Sa*CoaBC, to obtain the 3D reconstructions from this dataset we selected the average of all the classes for use in our refinements. We also selected the classes that clearly show projections that have 3-fold axes accompanied by the protrusions present alongside the central density (Figure 17B).

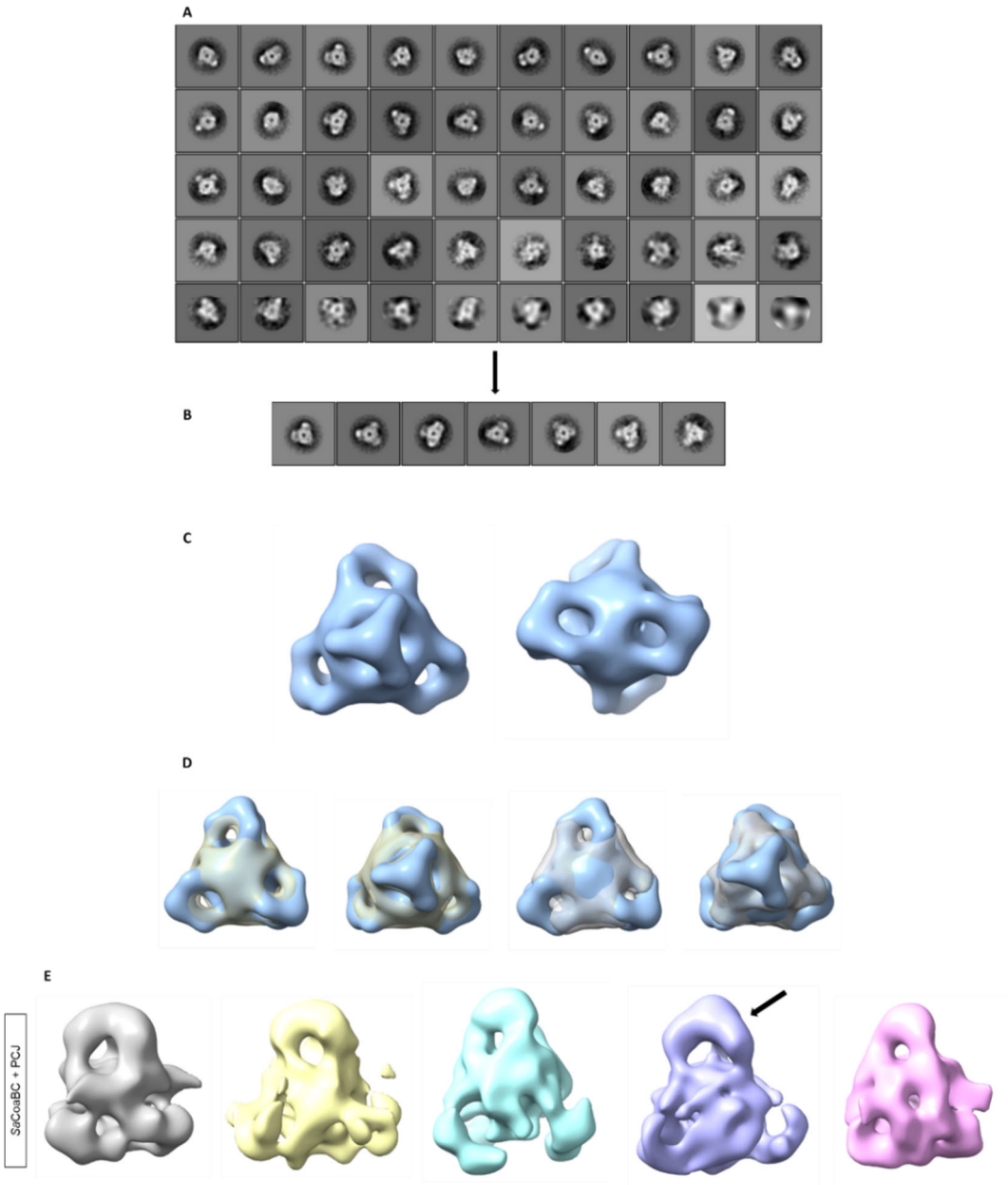


***Figure 17**: 2D and 3D classifications of SaCoaBC bound to P-CJ and CTP. **A** 2D class averages of SaCoaBC bound to P-CJ and CTP. **B** Selection of 2D classes harboring distinct 3-fold projections with protrusions at vertices. **C** 3D reconstruction generated from 2D subset selections. **D** Comparison of the treated SaCoaBC (P-CJ and CTP) reconstruction (blue) against the "closed conformation" (gold) and "open conformation" (grey) of the uninhibited SaCoaBC. **E** 3D classification of SaCoaBC (with P-CJ and CTP) distributed over 5 classes.*

Final 3D reconstructions generated for both the groups produced a model similar to the *closed conformation* reconstruction of the apo-protein dataset, with four distinct points on the tetrahedron at 22 Å (Figure 17C). Merging of the inhibitor-treated *Sa*CoaBC reconstruction with the apo-protein structure show only minor variations along the vertices and is overall similar. The areas of density around the points of the tetrahedron that is proposed to form the trimer interactions are more pronounced (Figure 17D). Merging of the inhibitor-treated *Sa*CoaBC reconstruction with the *open conformation* of the apo-protein reconstruction shares similarities as described previously from the comparisons in Figure 15A. As the models are very similar overall, these variations are likely caused by sample flattening or incomplete stain embedding. This is commonly observed in NS-EM as sample dehydration, deformations due to adsorption interactions and staining patterns all depend on the orientation of the protein on the carbon grid (Gallagher *et al*, 2019). This is also noticeable from the 2D classes as some of the backgrounds appear much darker in some regions from the projections as a result of inconsistent staining that we observed from the processed micrographs.

Overall, the entire dataset favors the structure that resembles the *closed conformation.* This was not expected as we hypothesized P-CJ to tightly act upon the CoaB domain in favour of the *open conformation*. In contrast, the selected 2D classes that were used to reconstruct the *closed conformation* in the apo-protein sample only accounts for approximately 20% of the particles used from the dataset as the overall distribution of those 2D classes favors the *open conformation* which closely resembles the structure from *Msm*CoaBC. We addressed this by carrying out 2D subset selections of the circular classes from the treated *Sa*CoaBC, but were unable to produce a valid model as it manifested as streaky reconstructions. As a result, the 3D classifications were carried out as described but limited to only one reference model. We applied it once again to both the selected 2D classes and to the entire data set. We were once again unable to observe regions from generated models that may indicate conformational variation from the points (Figure 17E). The computed dataset fell into one dominant conformational class that is more pronounced at only one of the points. Although our 3D classifications from both the inhibitor-treated and apo-protein datasets could not demonstrate that conformational changes exist for *Sa*CoaBC, it is possible that the proposed conformations could be identified using a larger image dataset. Larger datasets greatly facilitate structural information gained from EM studies. In NS-EM, image acquisition using low dose mode can overall improve resolution to gain better understanding of this effect close to 15 Å.

4.3 Conclusion

Our investigations of the structure of the *Sa*CoaBC protein complex were limited to the use of NS-EM due to both time constraints and the nature of the protein. Specifically, it proved to be quite challenging to work with *Sa*CoaBC due to inconsistent sample quality, and therefore

poor reproducibility. We have found that immediate sample preparation after purification and size exclusion chromatography greatly facilitated suitable particle acquisition and structural determination. Although limited, we were able to obtain valuable data into the overall architecture of the protein which was not possible with crystallography studies. Our findings suggests that *Sa*CoaBC exists in two forms in an equilibrium: an open and a closed forms that interchange. While our results are still inconclusive, it does provide evidence that warrants further investigation. For example, a detailed single particle cryo-EM study of *Sa*CoaBC in its apo state and in the presence of the P-CJ/CTP inhibitor mix may provide important insights in our understanding of the structure and the dynamics of CoaBC proteins. As the methods used to successfully prepare and process *Sa*CoaBC has been established, both NS-EM screening and cryo-EM acquisition can also be carried out for the other CoaBC proteins considered in this study (*Ec*CoaBC and *Mt*CoaBC) to provide further information on the structural and conformational diversity of CoaBC proteins.

Chapter 5: General conclusions and future work

The biochemistry of the CoA biosynthetic pathway has been broadly explored with many discrepancies of the pathways function observed between domains leading to its validation in recent years as target for drug development studies. Although extensively studied, the organization of the CoA biosynthetic machinery remains to be investigated. This study focused on investigating protein complex formation within the CoA biosynthetic pathway in two discrete ways.

First, we wished to investigate the potential existence of CoA-SPC in yeast by means of pull-down assays and whole cell protein extraction methods for its isolation. In this work we were able to successfully purify the Cab proteins (exception of Cab2) but unsuccessful in our attempts to isolate the CoA-SPC from *S. cerevisiae* under affinity pull-down assays. In the case of pull-down assays it is important to note that presence of the GST-fusion tag may compete with complex formation due to its relatively large size compared to other expression tags and thus introducing a poly-histidine tag can be pursued as an additional study for CoA-SPC isolation. This might in turn aid in the overall expression studies of Cab4 as its expression remained low and the purification of Cab2. Furthermore, crosslinking of the Cab protein can be explored to aid in solidifying the interactions if it's too weak to notice. The HPLC assays performed on the ammonium sulphate precipitated fractions were unable to detect CoA metabolites from both precipitated fractions and from the whole yeast lysate. We suspect that lysate preparatory factors may be the cause as CoA metabolites are readily detected in yeast lysates using HPLC or errors in assay procedure. Future work on the isolation and characterization of CoA-SPC will focus on repeating current experiments to address both the preparatory factors and the assay procedures that have limited the continuation of this work.

Many protein complexes confound individual structural biology techniques often as a result of having low abundance or displaying intrinsic dynamics. If this is the case for CoA-SPC, varying solution conditions and investigations with regards to external stimuli/stresses will be important for the pursual of the CoA-SPC isolation. Additional investigations such as single particle EM analysis under cryogenic temperatures coupled with NMR and crystallography of the individual Cab proteins can be pursued to determine the conformational dynamics to further report on the structure-function relationships between the proteins, which will contribute greatly to the context of CoA-SPC assembly.

Hybrid strategies that incorporate MS with soft ionization processes coupled to biophysical techniques can be used to further probe the properties of the CoA-SPC. This will in turn generate information of the assembly of the complex by varying solution conditions, help in

identifying protein-protein contacts, determining the relative binding strength and subunit exchange, and the determination of stoichiometry of subunits within the complex.

Our second aim pursued the characterization of bacterial bifunctional CoaBC proteins in the context of multienzyme complexes as it plays a key fragility point in CoA regulation in bacteria of which only one of the structures have been solved recently in *M. smegmatis* (*Msm*CoaBC). We were successful in determining the structure of *Sa*CoaBC at low resolution by means of negative staining single particle EM analysis. The data generated in this work demonstrated two key conformations after *in silico* 2D classification purification and 3D refinements. One of the reconstructions that we coined "*open conformation*" accommodated the overall architecture of *Msm*CoaBC refined to 20Å. The "*closed conformation*" is significantly different in which the vertices of the tetrahedron appear distinctly stretched forming for points at each apex.

Merging of the two maps hinted at the possibility of multiple conformations as a result of disruption within the CoaB dimer interface; however, 3D classifications were unable to computationally separate these states. We further probed the idea of multiple conformations by means of the incorporation of a tight binding inhibitor (PCJ +CTP) to assess the overall particle distribution. This investigation yielded unexpected results as the overall particle distribution within the dataset refined a model resembling the *closed conformation*. Three-dimensional classifications were unable to validate the effect by means of computationally purifying the proposed conformational states.

As future work, a detailed single particle cryo-EM analysis of the protein will be important to help decipher the structure of this protein at near atomic resolution and also for *Ec*CoaBC and *Mt*CoaBC. Unlike NS-EM, for cryo-EM samples are frozen rapidly in liquid ethane, producing a thin film of non-crystalline ice in which the proteins are orientated at random. The rapid freezing process aids in the preservation of native structure of the protein and the structural determination follows a well-defined path, imaged and maintained at cryogenic temperatures (Figure 18). Focused classification techniques in this manner will aid in probing the structural dynamics between the CoaB and CoaC domains of these bi-functional proteins.

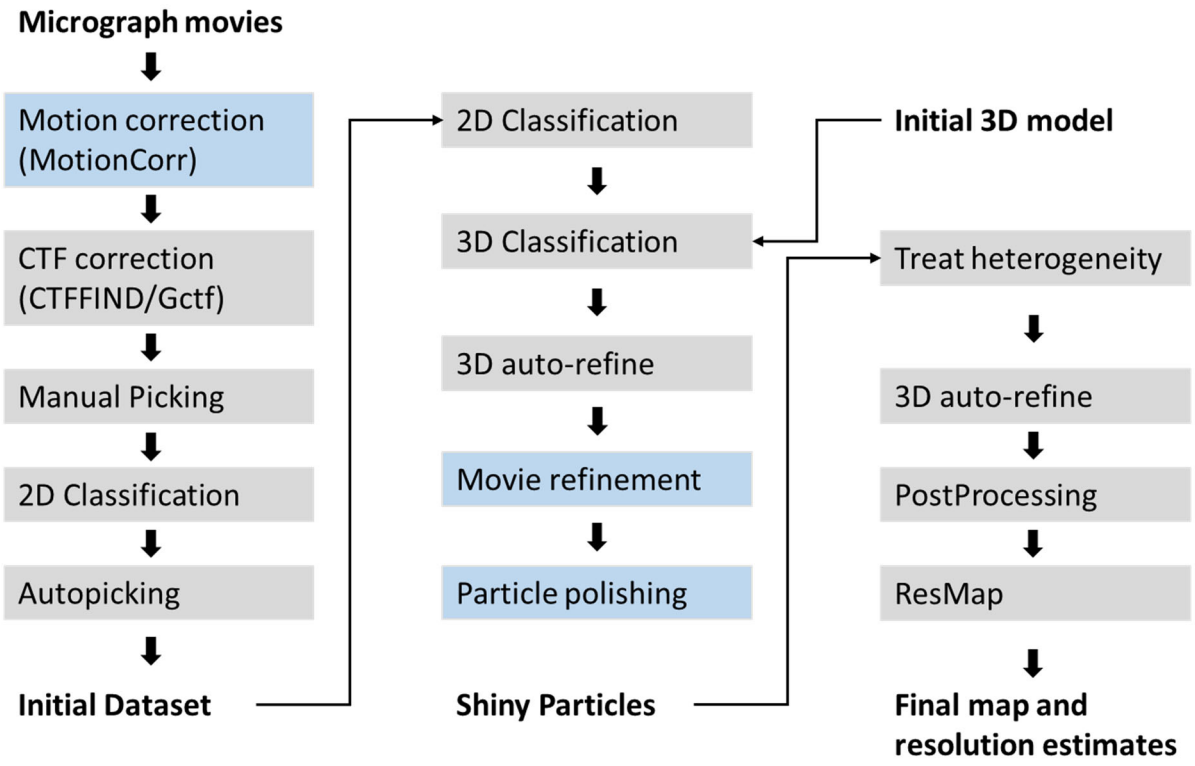


***Figure 18**: Workflow of a typical structure determination project in RELION. Image adapted from Scheres, 2016 with permission.*

References

Abrie, J. A., Molero, C., Ariño, J., & Strauss, E. (2015). Complex stability and dynamic subunit interchange modulates the disparate activities of the yeast moonlighting proteins Hal3 and Vhs3. *Scientific Reports*, *5*(1), 1-17.

Abu-Farha, M., Elisma, F., Zhou, H., Tian, R., Zhou, H., Asmer, M. S., & Figeys, D. (2009). Proteomics: from technology developments to biological applications. *Analytical Chemistry*, *81*(12), 4585-4599.

Albert, A., Martínez-Ripoll, M., Espinosa-Ruiz, A., Yenush, L., Culiáñez-Macià, F. A., & Serrano, R. (2000). The X-ray structure of the FMN-binding protein AtHal3 provides the structural basis for the activity of a regulatory subunit involved in signal transduction. *Structure*, *8*(9), 961-969.

Baković, J., López Martínez, D., Nikolaou, S., Yu, B. Y. K., Tossounian, M. A., Tsuchiya, Y., & Gout, I. (2021). Regulation of the CoA biosynthetic complex assembly in mammalian cells. *International Journal of Molecular Sciences*, *22*(3), 1131.

Baković, J., Yu, B. Y. K., Silva, D., Chew, S. P., Kim, S., Ahn, S. H., & Gout, I. (2019). A key metabolic integrator, coenzyme A, modulates the activity of peroxiredoxin 5 via covalent modification. *Molecular and Cellular Biochemistry*, *461*(1), 91-102.

Bepler, T., Morin, A., Rapp, M., Brasch, J., Shapiro, L., Noble, A. J., & Berger, B. (2019). Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nature Methods*, *16*(11), 1153-1160.

Berggård, T., Linse, S., & James, P. (2007). Methods for the detection and analysis of protein–protein interactions. *Proteomics*, *7*(16), 2833-2842.

Blaesse, M., Kupke, T., Huber, R., & Steinbacher, S. (2000). Crystal structure of the peptidyl-cysteine decarboxylase EpiD complexed with a pentapeptide substrate. *The EMBO journal*, *19*(23), 6299-6310.

Booth, D. S., Avila-Sakar, A., & Cheng, Y. (2011). Visualizing proteins and macromolecular complexes by negative stain EM: from grid preparation to image acquisition. *JoVE (Journal of Visualized Experiments)*, (58), e3227.

Brand, L. A., & Strauss, E. (2005). Characterization of a new pantothenate kinase isoform from Helicobacter pylori. *Journal of Biological Chemistry*, *280*(21), 20185-20188.

Breus, O., Panasyuk, G., Gout, I. T., Filonenko, V., & Nemazanyy, I. (2009). CoA synthase is in complex with p85αPI3K and affects PI3K signaling pathway. *Biochemical and Biophysical Research Communications*, *385*(4), 581-585.

Breus, O., Panasyuk, G., Gout, I. T., Filonenko, V., & Nemazanyy, I. (2010). CoA Synthase is phosphorylated on tyrosines in mammalian cells, interacts with and is dephosphorylated by Shp2PTP. *Molecular and Cellular Biochemistry*, *335*(1), 195-202.

Bucovaz, E. T., Macleod, R. M., Morrison, J. C., & Whybrew, W. D. (1997). The coenzyme A-synthesizing protein complex and its proposed role in CoA biosynthesis in baker's yeast. *Biochimie*, *79*(12), 787-798.

Byron, O., & Lindsay, J. G. (2017). The pyruvate dehydrogenase complex and related assemblies in health and disease. *Macromolecular Protein Complexes*, 523-550.

Cheng, Y. (2018). Single-particle cryo-EM—How did it get here and where will it go. *Science*, *361*(6405), 876-880.

Choudhary, C., & Mann, M. (2010). Decoding signalling networks by mass spectrometry-based proteomics. *Nature reviews Molecular cell biology*, *11*(6), 427-439.

Crawford, I. P. (1989). Evolution of a biosynthetic pathway: the tryptophan paradigm. *Annual Review of Microbiology*, *43*(1), 567-600.

Creighton, T. E. (1970). A steady-state kinetic investigation of the reaction mechanism of the tryptophan synthetase of *Escherichia coli*. *European Journal of Biochemistry*, *13*(1), 1-10.

Doherty, A. J., Connolly, B. A., & Worrall, A. F. (1993). Overproduction of the toxic protein, bovine pancreatic DNaseI, in Escherichia coli using a tightly controlled T7-promoter-based vector. *Gene*, *136*(1-2), 337-340.

Dong, H., Nilsson, L., & Kurland, C. G. (1995). Gratuitous overexpression of genes in *Escherichia coli* leads to growth inhibition and ribosome destruction. *Journal of Bacteriology*, *177*(6), 1497-1504.

Dunn, M. F., Aguilar, V., Brzovic, P., Drewe Jr, W. F., Houben, K. F., Leja, C. A., & Roy, M. (1990). The tryptophan synthase bi-enzyme complex transfers indole between the α- and β-sites via a 25–30 Å long tunnel. *Biochemistry,* 29(37), 8598-8607.

Dusi, S., Valletta, L., Haack, T. B., Tsuchiya, Y., Venco, P., Pasqualato, S., & Tiranti, V. (2014). Exome sequence reveals mutations in CoA synthase as a cause of neurodegeneration with brain iron accumulation. *The American Journal of Human Genetics*, *94*(1), 11-22.

Evans, J. C., Trujillo, C., Wang, Z., Eoh, H., Ehrt, S., Schnappinger, D., & Mizrahi, V. (2016). Validation of CoaBC as a bactericidal target in the coenzyme A pathway of *Mycobacterium tuberculosis. ACS Infectious Diseases*, *2*(12), 958-968.

Fairbanks, E. E. (1925). A modification of Lemberg's staining method. *American Mineralogist: Journal of Earth and Planetary Materials*, *10*(5), 126-127.

Gallagher, J. R., Kim, A. J., Gulati, N. M., & Harris, A. K. (2019). Negative-Stain Transmission Electron Microscopy of Molecular Complexes for Image Analysis by 2D Class Averaging. *Current Protocols in Microbiology*, *54*(1), e90.

Goosen, R., & Strauss, E. (2017). Simultaneous quantification of coenzyme A and its salvage pathway intermediates in in vitro and whole cell-sourced samples. *RSC advances*, *7*(32), 19717-19724.

Gout, I. (2018). Coenzyme A, protein CoAlation and redox regulation in mammalian cells. *Biochemical Society Transactions*, *46*(3), 721-728.

Gout, I. (2019). Coenzyme A: a protective thiol in bacterial antioxidant defence. *Biochemical Society Transactions*, *47*(1), 469-476.

Graeslund, S. & Hammarstroem, M. (2013). Affinity fusions for protein purification. Downstream Ind. *Biotechnology*. 191-199.

Grininger, M. (2014). Perspectives on the evolution, assembly and conformational dynamics of fatty acid synthase type I (FAS I) systems. *Current Opinion in Structural Biology*, *25*, 49-56.

Gudkova, D., Panasyuk, G., Nemazanyy, I., Zhyvoloup, A., Monteil, P., Filonenko, V., & Gout, I. (2012). EDC4 interacts with and regulates the dephospho-CoA kinase activity of CoA synthase. *FEBS Letters*, *586*(20), 3590-3595.

Henderson, R. (2013). Avoiding the pitfalls of single particle cryo-electron microscopy: Einstein from noise. *Proceedings of the National Academy of Sciences*, *110*(45), 18037-18041.

Ho, C. M., Li, X., Lai, M., Terwilliger, T. C., Beck, J. R., Wohlschlegel, J., & Zhou, Z. H. (2020). Bottom-up structural proteomics: cryoEM of protein complexes enriched from the cellular milieu. *Nature Methods*, *17*(1), 79-85.

Horie, S., Isobe, M., & Suga, T. (1986). Changes in CoA pools in hepatic peroxisomes of the rat, under various conditions. *Journal of Biochemistry*, *99*(5), 1345-1352.

Hörtnagel, K., Prokisch, H., & Meitinger, T. (2003). An isoform of hPANK2, deficient in pantothenate kinase-associated neurodegeneration, localizes to mitochondria. *Human Molecular Genetics*, *12*(3), 321-327.

Huang, X., Holden, H. M., & Raushel, F. M. (2001). Channeling of substrates and intermediates in enzyme-catalyzed reactions. *Annual Review of Biochemistry*, *70*(1), 149-180.

Hyde, C. C., Ahmed, S. A., Padlan, E. A., Miles, E. W., & Davies, D. R. (1988). Three-dimensional structure of the tryptophan synthase alpha 2 beta 2 multienzyme complex from Salmonella typhimurium. *Journal of Biological Chemistry*, *263*(33), 17857-17871.

Idell-Wenger, J. A., Grotyohann, L. W., & Neely, J. R. (1978). Coenzyme A and carnitine distribution in normal and ischemic hearts. *Journal of Biological Chemistry*, *253*(12), 4310-4318.

Jackowski, S., & Rock, C. O. (1984). Metabolism of 4'-phosphopantetheine in *Escherichia coli*. *Journal of Bacteriology*, *158*(1), 115-120.

Jackowski, S., & Rock, C. O. (1981). Regulation of coenzyme A biosynthesis. *Journal of Bacteriology*, *148*(3), 926-932.

Jenni, S., Leibundgut, M., Maier, T., & Ban, N. (2006). Architecture of a fungal fatty acid synthase at 5 A resolution. *Science*, *311*(5765), 1263-1267.

Kastritis, P. L., O'Reilly, F. J., Bock, T., Li, Y., Rogon, M. Z., Buczak, K., & Gavin, A. C. (2017). Capturing protein communities by structural proteomics in a thermophilic eukaryote. *Molecular Systems Biology*, *13*(7), 936.

Kupke, T., Uebele, M., Schmid, D., Jung, G., Blaesse, M., & Steinbacher, S. (2000). Molecular characterization of lantibiotic-synthesizing enzyme EpiD reveals a function for bacterial Dfp proteins in coenzyme A biosynthesis. *Journal of Biological Chemistry*, *275*(41), 31838-31846.

Leonardi, R., & Jackowski, S. (2007). Biosynthesis of pantothenic acid and coenzyme A. *EcoSal Plus*, *2*(2).

Leonardi, R., Zhang, Y. M., Rock, C. O., & Jackowski, S. (2005). Coenzyme A: back in action. *Progress in Lipid Research*, *44*(2-3), 125-153.

Lin, C. C., Kitagawa, M., Tang, X., Hou, M. H., Wu, J., Qu, D. C., & Chi, J. T. (2018). CoA synthase regulates mitotic fidelity via CBP-mediated acetylation. *Nature Communications*, *9*(1), 1-14.

Lomakin, I. B., Xiong, Y., & Steitz, T. A. (2007). The crystal structure of yeast fatty acid synthase, a cellular machine with eight active sites working together. *Cell*, *129*(2), 319-332.

Maco, B., Ross, I. L., Landsberg, M. J., Mouradov, D., Saunders, N. F., Hankamer, B., & Kobe, B. (2011). Proteomic and electron microscopy survey of large assemblies in macrophage cytoplasm. *Molecular & Cellular Proteomics*, *10*(6), M111.008763.

Maier, T., Jenni, S., & Ban, N. (2006). Architecture of mammalian fatty acid synthase at 4.5 A resolution. *Science*, *311*(5765), 1258-1262.

Maier, T., Leibundgut, M., & Ban, N. (2008). The crystal structure of a mammalian fatty acid synthase. *Science*, *321*(5894), 1315-1322.

Mendes, V., Green, S. R., Evans, J. C., Hess, J., Blaszczyk, M., Spry, C., & Blundell, T. L. (2021). Inhibiting Mycobacterium tuberculosis CoaBC by targeting an allosteric site. *Nature Communications*, *12*(1), 1-12.

Miller, J. R., Ohren, J., Sarver, R. W., Mueller, W. T., Dreu, P. D., Case, H., & Thanabal, V. (2007). Phosphopantetheine adenylyltransferase from *Escherichia coli*: investigation of the kinetic mechanism and role in regulation of coenzyme A biosynthesis. *Journal of Bacteriology*, *189*(22), 8196-8205.

Milne, J. L., Wu, X., Borgnia, M. J., Lengyel, J. S., Brooks, B. R., Shi, D., & Subramaniam, S. (2006). Molecular structure of a 9-MDa icosahedral pyruvate dehydrogenase subcomplex containing the E2 and E3 enzymes using cryoelectron microscopy. *Journal of Biological Chemistry*, *281*(7), 4364-4370.

Moolman, W. J., de Villiers, M., & Strauss, E. (2014). Recent advances in targeting coenzyme A biosynthesis and utilization for antimicrobial drug development. *Biochemical Society Transactions*, *42*(4), 1080-1086.

Nemazanyy, I., Panasyuk, G., Zhyvoloup, A., Panayotou, G., Gout, I. T., & Filonenko, V. (2004). Specific interaction between S6K1 and CoA synthase: a potential link between the mTOR/S6K pathway, CoA biosynthesis and energy metabolism. *FEBS Letters*, *578*(3), 357-362.

Nogales, E. (2016). The development of cryo-EM into a mainstream structural biology technique. *Nature Methods*, *13*(1), 24-27.

Olzhausen, J., Schübbe, S., & Schüller, H. J. (2009). Genetic analysis of coenzyme A biosynthesis in the yeast Saccharomyces cerevisiae: identification of a conditional mutation in the pantothenate kinase gene CAB1. *Current Genetics*, *55*(2), 163-173.

Olzhausen, J., Moritz, T., Neetz, T., & Schüller, H. J. (2013). Molecular characterization of the heteromeric coenzyme A-synthesizing protein complex (CoA-SPC) in the yeast *Saccharomyces cerevisiae*. *FEMS Yeast Research*, *13*(6), 565-573.

Ovádi, J., Huang, Y., & Spivey, H. O. (1994). Binding of malate dehydrogenase and NADH channelling to complex I. *Journal of Molecular Recognition*, *7*(4), 265-272.

Pan, P., & Dunn, M. F. (1996). β-site covalent reactions trigger transitions between open and closed conformations of the tryptophan synthase bienzyme complex. *Biochemistry*, *35*(15), 5002-5013.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, *25*(13), 1605-1612.

Punjani, A., Rubinstein, J. L., Fleet, D. J., & Brubaker, M. A. (2017). cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nature Methods*, *14*(3), 290-296.

Rhee, H. W., Zou, P., Udeshi, N. D., Martell, J. D., Mootha, V. K., Carr, S. A., & Ting, A. Y. (2013). Proteomic mapping of mitochondria in living cells via spatially restricted enzymatic tagging. *Science*, *339*(6125), 1328-1331.

Robishaw, J. D., Berkich, D., & Neely, J. R. (1982). Rate-limiting step and control of coenzyme A synthesis in cardiac muscle. *Journal of Biological Chemistry*, *257*(18), 10967-10972.

Rock, C. O., Calder, R. B., Karim, M. A., & Jackowski, S. (2000). Pantothenate kinase regulation of the intracellular concentration of coenzyme A. *Journal of Biological Chemistry*, *275*(2), 1377-1383.

Roh, S. H., Hryc, C. F., Jeong, H. H., Fei, X., Jakana, J., Lorimer, G. H., & Chiu, W. (2017). Subunit conformational variation within individual GroEL oligomers resolved by Cryo-EM. Proceedings of the National Academy of Sciences, 114(31), 8259-8264.

Ruiz, A., González, A., Muñoz, I., Serrano, R., Abrie, J. A., Strauss, E., & Ariño, J. (2009). Moonlighting proteins Hal3 and Vhs3 form a heteromeric PPCDC with Ykl088w in yeast CoA biosynthesis. *Nature Chemical Biology*, *5*(12), 920-928.

Scheres, S. H. (2012). RELION: implementation of a Bayesian approach to cryo-EM structure determination. *Journal of Structural Biology*, *180*(3), 519-530.

Scheres, S. H. (2012). A Bayesian view on cryo-EM structure determination. *Journal of Molecular Biology*, *415*(2), 406-418.

Scheres, S. H. (2016). Processing of structurally heterogeneous cryo-EM data in RELION. *Methods in Enzymology*, *579*, 125-157.

Schmidt, C. Q., Herbert, A. P., Hocking, H. G., Uhrin, D., & Barlow, P. N. (2008). Translational mini-review series on complement factor H: structural and functional correlations for factor H. *Clinical & Experimental Immunology*, *151*(1), 14-24.

Schmidt, C. Q., Herbert, A. P., Mertens, H. D., Guariento, M., Soares, D. C., Uhrin, D., & Barlow, P. N. (2010). The central portion of factor H (modules 10–15) is compact and contains a structurally deviant CCP module. *Journal of Molecular Biology*, *395*(1), 105-122.

Smith, B. J. (1984). SDS polyacrylamide gel electrophoresis of proteins. In *Proteins* (pp. 41-55). Humana Press.

Smith, C. M., & Savage Jr, C. R. (1980). Regulation of coenzyme A biosynthesis by glucagon and glucocorticoid in adult rat liver parenchymal cells. *Biochemical Journal*, *188*(1), 175-184.

Song, W. J., & Jackowski, S. (1992). Cloning, sequencing, and expression of the pantothenate kinase (coaA) gene of *Escherichia coli*. *Journal of Bacteriology*, *174*(20), 6411-6417.

Spry, C., Kirk, K., & Saliba, K. J. (2008). Coenzyme A biosynthesis: an antimicrobial drug target. *FEMS Microbiology Reviews*, *32*(1), 56-106.

Strauss, E. (2010). Coenzyme A biosynthesis and enzymology. *Comprehensive Natural Products II*, *7*, 351-410.

Szymanski, E. P., & Kerscher, O. (2013). Budding yeast protein extraction and purification for the study of function, interactions, and post-translational modifications. *JoVE (Journal of Visualized Experiments)*, (80), e50921.

Tahiliani, A. G., & Tahiliani, A. (1989). Dependence of mitochondrial coenzyme A uptake on the membrane electrical gradient. *Journal of Biological Chemistry*, *264*(31), 18426-18432.

Tubbs, P. K., & Garland, P. B. (1964). Variations in tissue contents of coenzyme A thio esters and possible metabolic implications. *Biochemical Journal*, *93*(3), 550.

Vallari, D. S., & Rock, C. O. (1987). Isolation and characterization of temperature-sensitive pantothenate kinase (*coaA*) mutants of *Escherichia coli*. *Journal of Bacteriology*, *169*(12), 5795-5800.

van der Westhuyzen, R., Hammons, J. C., Meier, J. L., Dahesh, S., Moolman, W. J., Pelly, S. C., & Strauss, E. (2012). The antibiotic CJ-15,801 is an antimetabolite that hijacks and then inhibits CoA biosynthesis. *Chemistry & Biology*, *19*(5), 559-571.

Verbeke, E. J., Mallam, A. L., Drew, K., Marcotte, E. M., & Taylor, D. W. (2018). Classification of single particles from human cell extract reveals distinct structures. *Cell Reports*, *24*(1), 259-268.

Verbeke, E. J., Zhou, Y., Horton, A. P., Mallam, A. L., Taylor, D. W., & Marcotte, E. M. (2020). Separating distinct structures of multiple macromolecular assemblies from cryo-EM projections. *Journal of Structural Biology*, *209*(1), 107416.

Westerhoff, H. V., & Welch, G. R. (1992). Enzyme organization and the direction of metabolic flow: physicochemical considerations. *Current Topics in Cellular Regulation*, *33*, 361-390.

White, S. W., Zheng, J., Zhang, Y. M., & Rock, C. O. (2005). The structural biology of type II fatty acid biosynthesis. *Annual Reviews in Biochemistry*, *74*, 791-831.

Williamson, J. R., & Corkey, B. E. (1979). [23] Assay of citric acid cycle intermediates and related compounds—Update with tissue metabolite levels and Intracellular Distribution. In *Methods in Enzymology* (Vol. 55, pp. 200-222). Academic Press.

Wriggers, W., & Chacón, P. (2001). Using Situs for the registration of protein structures with low-resolution bead models from X-ray solution scattering. *Journal of Applied Crystallography*, *34*(6), 773-776.

Yanagida, M. (2002). Functional proteomics; current achievements. *Journal of Chromatography B*, *771*(1-2), 89-106.

Yang, K., Eyobo, Y., Brand, L. A., Martynowski, D., Tomchick, D., Strauss, E., & Zhang, H. (2006). Crystal structure of a type III pantothenate kinase: insight into the mechanism of an essential coenzyme A biosynthetic enzyme universally distributed in bacteria. *Journal of Bacteriology*, *188*(15), 5532-5540.

Yu, H., Braun, P., Yıldırım, M. A., Lemmens, I., Venkatesan, K., Sahalie, J., & Vidal, M. (2008). High-quality binary protein interaction map of the yeast interactome network. *Science*, *322*(5898), 104-110.

Yun, M., Park, C. G., Kim, J. Y., Rock, C. O., Jackowski, S., & Park, H. W. (2000). Structural basis for the feedback regulation of *Escherichia coli* pantothenate kinase by coenzyme A. *Journal of Biological Chemistry*, *275*(36), 28093-28099.

Zhang, Y., Guo, X., Xiong, L., Yu, L., Li, Z., Guo, Q., & Lin, N. (2014). Comprehensive analysis of microRNA-regulated protein interaction network reveals the tumor suppressive role of microRNA-149 in human hepatocellular carcinoma via targeting AKT-mTOR pathway. *Molecular Cancer*, 13(1), 1-15.

Zhao, X., Li, G., & Liang, S. (2013). Several affinity tags commonly used in chromatographic purification. *Journal of Analytical Methods in Chemistry*, 2013:581093.

Zhou, B., Westaway, S. K., Levinson, B., Johnson, M. A., Gitschier, J., & Hayflick, S. J. (2001). A novel pantothenate kinase gene (PANK2) is defective in Hallervorden-Spatz syndrome. *Nature Genetics*, *28*(4), 345-349.

Zhyvoloup, A., Nemazanyy, I., Panasyuk, G., Valovka, T., Fenton, T., Rebholz, H., & Gout, I. T. (2003). Subcellular localization and regulation of coenzyme A synthase. *Journal of Biological Chemistry*, *278*(50), 50316-50321.

Zivanov, J., Nakane, T., Forsberg, B. O., Kimanius, D., Hagen, W. J., Lindahl, E., & Scheres, S. H. (2018). New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife*, *7*, e42166.