Structural and biochemical characterisation of the de novo cysteine biosynthesis pathway in trypanosomatids.

KATE VICTORIA SOWERBY

Structural and biochemical characterisation of the *de novo* cysteine biosynthesis pathway in trypanosomatids.

Kate Victoria Sowerby

Submission for the Degree of Doctor of Philosophy (PhD)

Department of Chemistry

Durham University

August 2023

Abstract

Neglected tropical diseases (NTD) affect over a billion people throughout the world. Chagas disease, caused by the protozoan parasite *T. cruzi* and leishmaniasis, a result of over 15 species of *Leishmania*, are NTDs selected for control and elimination. Under a One Health approach *T. theileri* infections affecting cattle should also be eradicated. Current treatments for these diseases are inadequate and their use is associated with many severe side effects, often leading to incomplete treatment. These trypanosomatids rely on cysteine biosynthesis for the production of trypanothione, which is essential in allowing the parasite to survive within the host. The *de novo* pathway of cysteine biosynthesis requires the conversion of L-serine into O-acetyl-L-serine by serine acetyltransferase, and then the transformation of O-acetyl-L-serine into L-cysteine, which is catalysed by cysteine synthase. These enzymes present an attractive potential for drug development against *T. cruzi*, *Leishmania* spp. and *T. theileri*.

This work aimed to explore the *de novo* cysteine biosynthesis pathway further through exploration of the underlying molecular mechanisms. This was achieved through investigations into the proteins involved. Studies of *Tc*SAT were conducted and it was concluded that further work is required to produce soluble recombinant protein to be used in additional experiments. Biochemical and crystallographic studies of *Tc*CS, *Li*CS and *Tth*CS were conducted. Crystal structures of the three enzymes were determined at resolutions of 1.80 Å for *Tc*CS, 1.75 Å for *Li*CS and 2.75 Å for *Tth*CS. These three homodimeric structures show the same overall fold and support a common reaction mechanism. Interactions between the two proteins,

forming a cysteine synthase complex were explored with the characterisation of activity and initial structural studies undertaken. Through use of the cysteine synthase structures, several avenues of *de novo* drug design were explored. To further this aim, 14 structures of *Tc*CS and *Li*CS collectively were determined with fragments bound, resulting in the selection of a lead compound to develop further.

Overall, understanding of the *de novo* cysteine biosynthesis pathway in trypanosomatids has been explored and the foundations of drug design have been established.

Acknowledgements

I have many people to thank for this work, without whom this PhD would not have been possible.

Firstly, my thanks go to my supervisor, Ehmke – thank you for your continual support, advice and encouragement over the last four years.

Thank you to Ian and Deborah for keeping everything running, fixing all the things I broke and always helping me out. Thank you to all the members of Labs 229 and 234 past and present for your help and friendship, not to mention enthusiasm for tea times! Extra thanks go to Stef for her endless expertise in crystallography, Emma for never failing to answer my continual questions, Katy for many a margarita night after a hard day of science, Charlie, Dan, Abbey, Dori, Izzy, Davide and Charlotte for your encouragement and assistance, and to Emma and Yasmine for their friendship, advice and willingness to drink wine with me. Thanks also go to the GCRF Network, in particular Ariel, Ana Milena and Sabrina for their knowledge of all things trypanosomatid!

Thanks to the Mass Spectrometry, NMR services and Elizabeth Bromley for their assistance with this work.

Finally, thank you to my family and friends – especially my Mum and Dad – for their constant love and encouragement both generally and throughout this PhD.

Abbreviations

ÄKTA	FPLC system (Cytiva)
BCS	Basic Chemical Space
CBS	Cystathionine β-synthase
CC ½	Correlation coefficient ½
CCP4	Collaborative Computational Project 4
CGL	Cystathionine γ-lyase
CoA	Coenzyme A
Coot	Crystallography Object-Orientated Toolkit
CS	Cysteine synthase
CSC	Cysteine synthase complex
DIALS	Diffraction Integration for Advanced Light Sources
DLS	Diamond Light Source
DMSO	Dimethylsulphoxide
DNA	Deoxyribonucleic acid
EDTA	Ethylenediaminetetraacetic acid
GST	Glutathione-S-Transferase
HEPES	(4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid)
HRV3C	Human Rhinovirus 3C protease
IC_{50}	half maximal inhibitory concentration
IMAC	Immobilised Metal Affinity Chromatography
IPTG	Isopropyl ß-d-1-thiogalactopyranoside
JCSG	Joint Consortium for Structural Genomics
LB	Lysogeny Broth
LC-ESI	Liquid chromatography electron spray ionisation
LMB	Laboratory of Molecular Biology
LogP	Octanol-water partition coefficient
MALDI-TOF	Matrix-assisted laser desorption/ionisation-time of flight
MBP	Maltose Binding Protein
NADH	Nicotinamide Adenine Dinucleotide Hydrogen
NCS	Non-crystallographic symmetry
NEB	New England Biolabs
NMR	Nuclear Magnetic Resonance
NTD	Neglected Tropical Disease
NusA	N-utilisation substance
OAS	O-Acetyl-L-Serine
OD_{600}	Optical density 600 nm
PACT	pH, Anion and Cation
PDB	Protein Data Bank
PEG	Polyethylene Glycol
PGA	Poly-γ-glutamic acid
PLP	Pyridoxal-5-Phosphate
PTFE	Polytetrafluoroethylene
RMSD	Root mean squared deviation

SAT	Serine Acetyltransferase
SAXS	Small Angle X-ray Scattering
SDS	Sodium Dodecyl Sulfate
SDS-PAGE	Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis
SEC	Size Exclusion Chromatography
SG1	Shotgun 1
TEM	Transmission Electron Microscopy
UV	Ultraviolet

Contents

Declaration and Statement of Copyright

The work in this thesis was carried out in the Department of Chemistry, Durham University, between October 2019 and May 2023. All work is the author's own unless otherwise stated. This work has not previously been submitted for a degree at this or any other institution.

Peer-reviewed publications from this thesis include the following, at the time of writing:

Sowerby, K., Freitag-Pohl, S., Murillo, A. M., Silber, A. M., & Pohl, E. (2023). Cysteine synthase: multiple structures of a key enzyme in cysteine synthesis and a potential drug target for Chagas disease and leishmaniasis. *Acta Crystallographica Section D: Structural Biology*, 79(6), 518-530.

Chapter 1: Introduction

Neglected tropical diseases are a wide-ranging group of diseases which affect over a billion people in tropical and sub-tropical areas around the globe. There are approximately twenty NTDs which have been targeted by the WHO for control and elimination, including Chagas disease and Leishmaniases (WHO, 2020).

Chagas disease, also known as American trypanosomiasis, is caused by *Trypanosoma cruzi* and affects 6 to 8 million people. This disease is endemic in the Americas however due to climate change and migration has become a global concern for non-endemic regions, such as Europe, Australia, Canada and Japan (Lidani et al., 2019). Additionally, a closely related trypanosomatid, *Trypanosoma theileri*, is found globally and is capable of infecting livestock mammals including sheep and cattle (Fentahun, 2020; Kelly et al., 2017). Diseases resulting from *T. theileri* infection are not well examined due to the presumed limited economic effects. Although it is not entirely comprehended, *T. theileri* infection can occur due to coinfection with infections such as bovine leukaemia virus or *Theileria orientalis* (Brotánková et al., 2022; Suganuma et al., 2022). The One Health approach aims to achieve improved health outcomes for humans, animals and the environment. This approach can be applied to combat neglected tropical diseases (Peterson et al., 2021). Under this methodology, drugs against veterinary trypanosomatids such as *T. theileri* should be proactively developed (Chimera et al., 2021).

Trypanosoma cruzi and Chagas Disease

T. cruzi is a protozoan parasite that, like many trypanosomatids, has a two stage infection: acute and chronic phases with an intermediate stage in between. After an

incubation stage of 1 to 2 weeks, the acute stage begins (Hemmige et al., 2012). The acute stage lasts between 4 and 8 weeks and is generally asymptomatic or presents as varied non-specific symptoms such as prolonged fever, malaise, lymphadenomegaly, localised or generalised subcutaneous oedema. During the acute phase, mortality ranges from 5 to 10% (Pereira et al., 2013). In endemic areas, the disease is typically acquired in childhood (Lancet Neurology, 2009).

Following the acute stage, the intermediate stage occurs where parasitaemia falls and patients become asymptomatic with antibodies present to the parasite. Patients in the intermediate phase stay infected which can result in the emergence of the chronic phase months or decades later in 15-30% of patients (Hemmige et al., 2012).

Within the chronic phase, different clinical forms of infection are observed. Around 60% of people present with the indeterminate form which has no symptoms; 20-40% will develop cardiac and/or digestive forms which ultimately results in death. There are four main developments that occur during the cardiac form: parasite induced myocardial damage, immune-mediated myocardial damage, cardiac dysautonomia and cardiovascular anomalies (Pereira et al., 2013). The cardiac form can result in sudden cardiac death which is often undiagnosed (Nunes et al., 2018). The digestive form usually results in megasyndrome due to damage to the autonomic ganglia which results in denervation of the tubular structures of the gastrointestinal tract. Megaoesophagus, dysphagia, weight loss and chronic aspiration can also occur (Hemmige et al., 2012).

Leishmaniasis

More than 15 species of trypanosomatids from the *Leishmania* genus can cause Leishmaniasis. This disease is endemic in most tropical and sub-tropical regions of the world (Burza et al., 2018). Leishmaniasis has a broad range of manifestations that can vary in severity dependant on the *Leishmania* species causing the infection (Torres-Guerrero et al., 2017). *Leishmania infantum* infection can result in visceral leishmaniasis, post-kala azar (dermal leishmaniasis), cutaneous leishmaniasis and mucosal leishmaniasis (Mann et al., 2021). Visceral leishmaniasis is the most serious form of leishmaniasis and presents as fever, weight loss, enlarged spleen and liver. The mortality rate is high for untreated cases, with mortality up to 15% of patients even with treatment (Matlashewski, 2001). Post-kala azar occurs as a complication of visceral leishmaniasis and results in papular or nodular rashes, hyperkeratosis and liquefaction degeneration of basal cells (Zijlstra et al., 2003). Cutaneous leishmaniasis presents as ulcerative lesions on the face (Aoun et al., 2014). Although the least common form of *L. infantum* infection, mucosal leishmaniasis causes mucosal lesions in immunocompromised patients (Richter et al., 2011).

Diagnosis

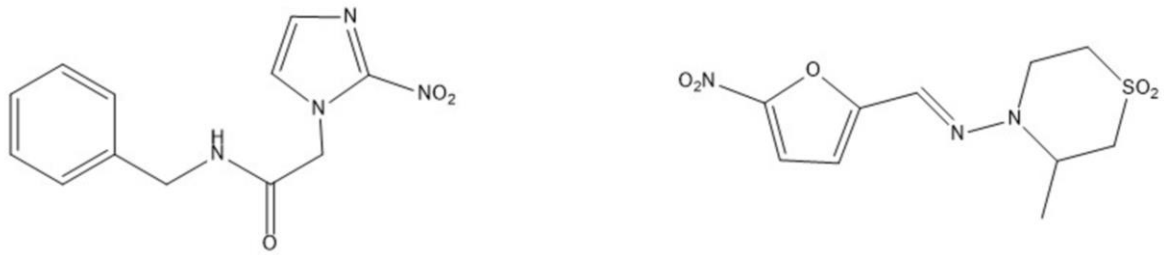
Diagnosis of Chagas disease is difficult, especially as during the acute phase diagnosis is based on the direct observation of *T. cruzi* parasites in the blood. Serological tests for *T. cruzi* antibodies are often negative and difficult to interpret. Further diagnostic methods are required for the diagnosis of parasite entry such as observation of inoculation chagoma or by parasitological diagnosis due to the high parasitaemia (Lescure et al., 2010). More recently, qPCR has become more widely used to identify the presence of *T. cruzi* DNA from whole blood samples (Mayta et

al., 2019). Chronic phase Chagas disease is diagnosed by parasitological, serological and molecular biology techniques (Pereira et al., 2013).

Diagnosis of leishmaniasis is similarly challenging, in part due to the generalised symptoms resembling other common diseases. Observation of parasites present in bone marrow or in splenic aspirates is the best method of diagnosis. Serological tests and molecular techniques are also used however there are complications with these methods regarding positive but asymptomatic patients and these approaches are unable to diagnose a relapse of the disease (Thakur et al., 2020).

Current Treatments

Current treatments for Chagas disease include only two drugs, benznidazole and nifurtimox, Figure 1-1, which were introduced in the late 1960s. For both drugs, the aim is to reduce the parasite load and to reduce transmission (Coura et al., 2011).



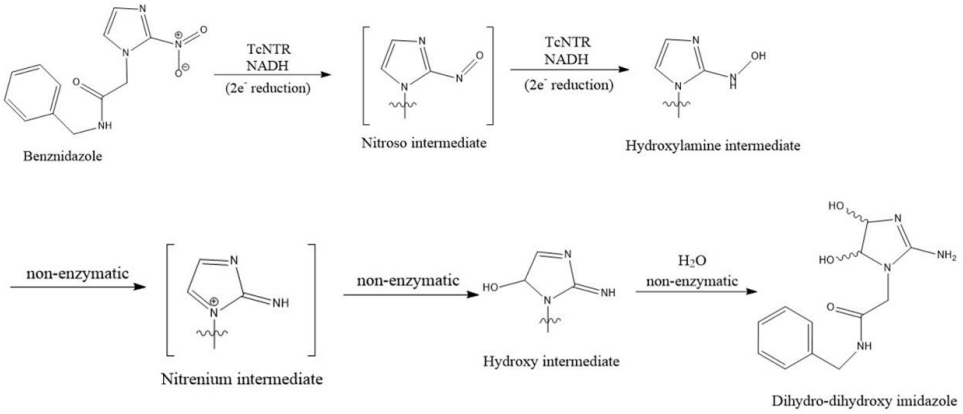
Benznidazole Nifurtimox

Figure 1-1 - Chemical structures of benznidazole (left) and nifurtimox (right).

Both are prodrugs, which are activated by trypanosomal nitroreductases to produce reactive metabolites, resulting in cellular toxicity and therefore parasite death,

Figure 1-2 (Boiani et al., 2010; Patterson et al., 2014). NADH-dependant trypanosomal type I nitroreductases are key to the activation of these drugs. Reduction of benzanidazole and nifurtimox lead to the formation of a hydroxylamine derivative that is converted to glyoxal through several intermediate stages. The glyoxal produced is cytotoxic and mutagenic, resulting in parasite death (Hall et al., 2012).

Benzanidazole activation:



Nifurtimox activation:

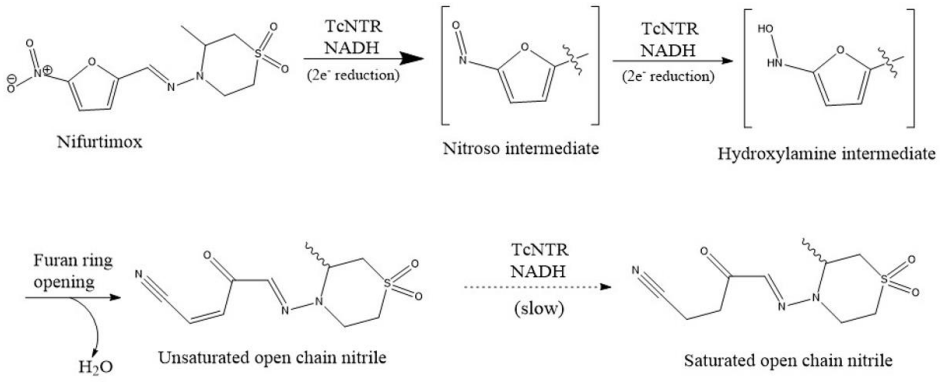
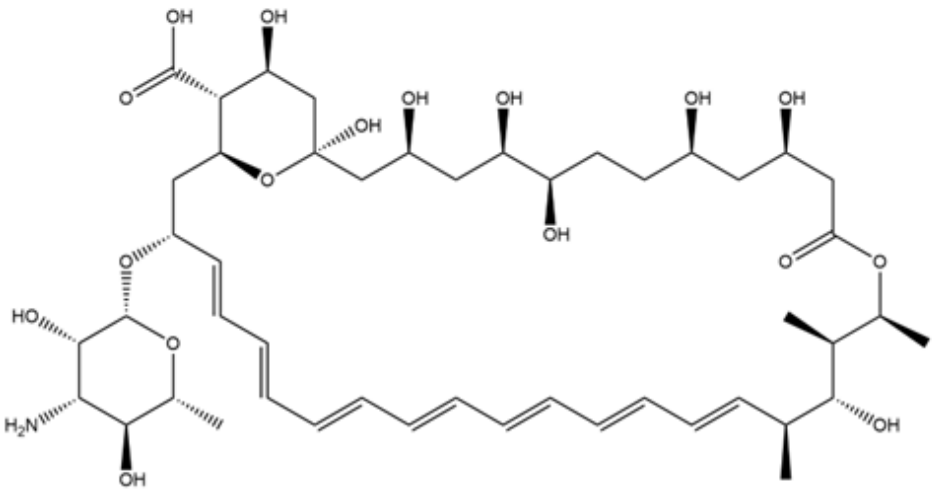


Figure 1-2 – Activation of benznidazole (above) and nifurtimox (below) by *T. cruzi* nitroreductase (*Tc*NTR) using NADH. Based on information from (Patterson et al., 2014)

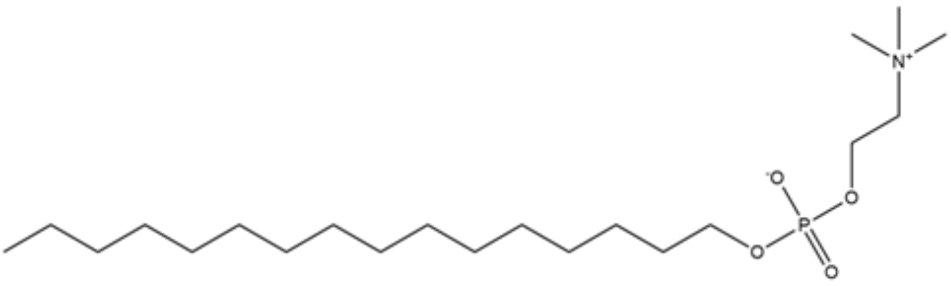
These treatments were previously thought to have only been effective during the acute phase, however following evidence that parasite persistence triggers the chronic pathology, treatments are now used in both phases. Nevertheless, these treatments have severe difficulties associated with their use (Kratz et al., 2018). Benznidazole has serious side effects such as rashes, epigastric pain and abdominal bloating yet generally has good adherence and tolerability with 77% of patients completing the full course (Olivera et al., 2015). Nifurtimox has been reported to have significantly higher side effects than benznidazole and a lower course

completion rate therefore benznidazole has become the preferred treatment for Chagas disease (Forsyth et al., 2016 (Jackson et al., 2020)).

The predominant first line treatment for leishmaniasis is pentavalent antimony derivatives, meglumine antimoniate and sodium stibogluconate. The structures and mode of action of these molecules remain mainly unknown (Frézard et al., 2009; Lindquist-Kleissler et al., 2023). These drugs are severely toxic, resulting in cardiotoxicity, pancreatitis and fever amongst other side effects and patients are required to be hospitalised during administration. Other drugs such as liposomal amphotericin B and miltefosine are available as second line treatments but these alternatives have associated problems such as prohibitive expense to most patients and toxicity (Moore et al., 2010), Figure 1-3. Amphotericin B inhibits *in vivo* replication of fungi and has been repurposed for the treatment of leishmaniasis (Adler-Moore et al., 2002). Despite the chemical structure being well known, the molecular mechanism remains elusive in *Leishmania*. It has been suggested that the mode of action involves the drug binding to sterols in the parasite membrane leading to pore formation and disruption of membranes (Saha et al., 1986; Singh et al., 2005). Contrastingly, the mechanism of miltefosine is well understood. The drug is known to inhibit synthesis of phosphatidylcholine and inhibits the cytochrome c oxidase, although other targets are suspected to be effected, such as calcium channels (Pinto-Martinez et al., 2018).



Amphotericin B



Miltefosine

Figure 1-3 - Chemical structures of Amphotericin B and Miltefosine.

In addition, resistance to these drugs has been reported, limiting their efficacy and suitability as a treatment (Masmoudi et al., 2013). Increasing resistance results in the use of higher doses causing increased side effects and reduced compliance for the same treatment results. Resistance to drug treatments for leishmaniasis has been emerging for the past 30 years and is widely reported (Croft et al., 2006).

Available treatments for both Chagas disease and leishmaniasis are inadequate due to their limited efficacy, frequent side effects and long treatment regimens (Francisco et al., 2020). Further, the mode of action of these drugs are generally poorly understood, greatly complicating any attempts to improve efficacy. Consequently, these neglected tropical diseases present great unmet medical challenges and the urgent need for novel treatments.

Life cycles and vector borne transmission

All the trypanosomatid family shares a complex lifecycle including an insect vector and a mammalian host. Chagas disease is mainly transmitted by triatomine insects, which are obligate haematophagous insects and when infected, spread *T. cruzi* through a bloodmeal (Vieira et al., 2018). Chagas disease is mainly transmitted through vectors, blood transfusion, oral, placental or congenital means. Triatomines are the main vector for *T. cruzi* (Coura, 2015). Whilst all species of the triatomine insects are capable of being *T. cruzi* vectors; the *Panstrongylus*, *Rhodnius* and *Triatoma* genera are the main vectors with *Triatoma infestans* being the most prevalent (Vieira et al., 2018). There are also secondary cases of transmission through handling or consumption of infected animals, organ transplants from previously infected patients and sexual transmission. Direct transmission to humans and other animals, especially marsupials, from *T. cruzi* reservoirs can occur. Over 100 mammal species are *T. cruzi* reservoirs including marsupials, rodents, bats, carnivores and primates (Coura, 2015). Triatomines are obligate hematophagous insects and acquire the *T. cruzi* trypomastigotes through a bloodmeal from an infected individual. The bloodstream trypomastigotes mature to metacyclic trypomastigotes and are excreted in the faeces which enter the mammalian

bloodstream through pre-existing lesions, bite sites or through mucosal tissues (Vieira et al., 2018).

L. infantum is transmitted through phlebotomine sandflies (Alten et al., 2016). In the Mediterranean basin alone, eight different species of sandflies were identified as *L. infantum* vectors including *Phlebotomus ariasi*, which has been identified as a predominant vector for *L. infantum* (Alten et al., 2016). The main reservoir *of L. infantum* is the domestic dog but other mammals such as bats, armadillos and sloths have also been identified as reservoirs (Alcover et al., 2020; Vulpiani et al., 2011).

T. theileri is mainly spread by Tabanidae flies yet also has the potential to be transmitted by dipteran vectors, such as mosquitos, and sandflies (Brotánková et al., 2022; Fentahun, 2020).

Vector control has been shown to be insufficient in the eradication of these parasites due to a multitude of factors, including the variety of reservoirs and an emergence in resistance to insecticides, and emphasises the need for affordable and effective drugs (Francis et al., 2021).

The life cycle of *T. cruzi* is well established and shown in Figure 1-4. In *T. cruzi* infection, the trypomastigotes, which are non-replicative, differentiate into epimastigotes (Figure 1-4, step 5), which proliferate and establish the infection in the triatomine insect (Figure 1-4, step 6). In the terminal region of the insect digestive tube, epimastigotes differentiate into infective, non-replicative metacyclic trypomastigotes (Figure 1-4, steps 7 and 8). During the next bloodmeal on a new mammalian host, metacyclic trypomastigotes are expelled together with the excreta

of the insect (Figure 1-4, step 1). These parasites invade the mammalian host through mucosa and/or small wounds and infect many types of nucleated cells including macrophages (Figure 1-4, steps 2 and 3). Once inside the cell, metacyclic trypomastigotes differentiate into the proliferative amastigotes which divide in the cell cytoplasm (Figure 1-4, steps 4). At high densities, amastigotes differentiate into bloodstream trypomastigotes which invade other cells, or reach the bloodstream where they can be taken up by a new triatomine during a bloodmeal (Figure 1-4, step 5) (Tyler et al., 2001).

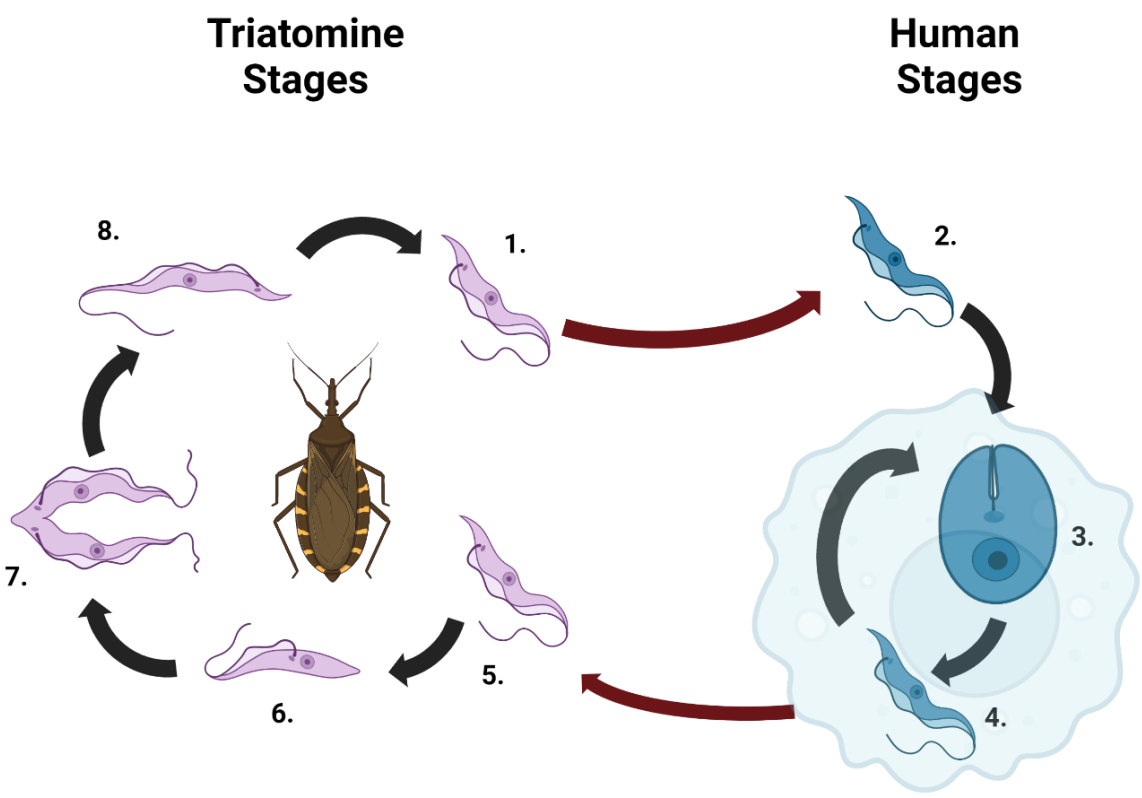


Figure 1-4 - *Trypanosoma cruzi* life cycle. Insect stages (1, 5 – 8) shown in lilac on left, human stages (2-4) shown in blue on right.

The *L. infantum* life cycle is similar with sand flies becoming infected with *L. infantum* when the sand fly takes up phagocytic cells bearing amastigotes during its bloodmeal. Inside the insect, amastigotes differentiate into replicative procyclic promastigotes. These forms stop dividing and differentiate into non-proliferative, infective metacyclic promastigotes which are expelled together with the saliva during a new bloodmeal on a new mammalian host. The parasites deposited into the bite site infect macrophages recruited to the wound. Within the macrophages, the promastigote differentiates into an amastigote (Sunter et al., 2021). The amastigotes replicate in a parasitophorous vacuole until lysing the cell. Once released into the extracellular medium, amastigotes will be able to infect other phagocytic cells, which can lead to a systemic infection (Andreani et al., 2015).

Redox control, trypanothione and cysteine in trypanosomatids

Due to the life cycles shared by trypanosomatids, redox control is essential to parasite survival upon invasion of host cells (Krauth-Siegel et al., 2008). In contrast to most organisms, the redox system used by trypanosomatids relies on trypanothione instead of glutathione. Trypanothione is produced from glutathione and spermidine (Krauth-Siegel et al., 2003). The trypanothione redox system is heavily involved in the parasite's ability to survive the oxidative bursts imposed intracellularly within the macrophages upon infection of human cells (Dumas et al., 1997). Cysteine is an essential component to this pathway. In addition to the roles of stability, catalytic activity and protein regulation, cysteine is a critical component in the synthesis of trypanothione (Canepa et al., 2009).

Further to the roles described above, cysteine is also used to form iron-sulphur clusters which are required to maintain fundamental processes in any organisms. Iron-sulphur clusters are vital components of electron transport chains, participate in substrate binding and activity and also are important in the regulation of gene expression (Johnson et al., 2005).

Cysteine is incorporated into glutathione which is in turn encompassed into trypanothione and mono-glutathionylspermidine, Figure 1-5.

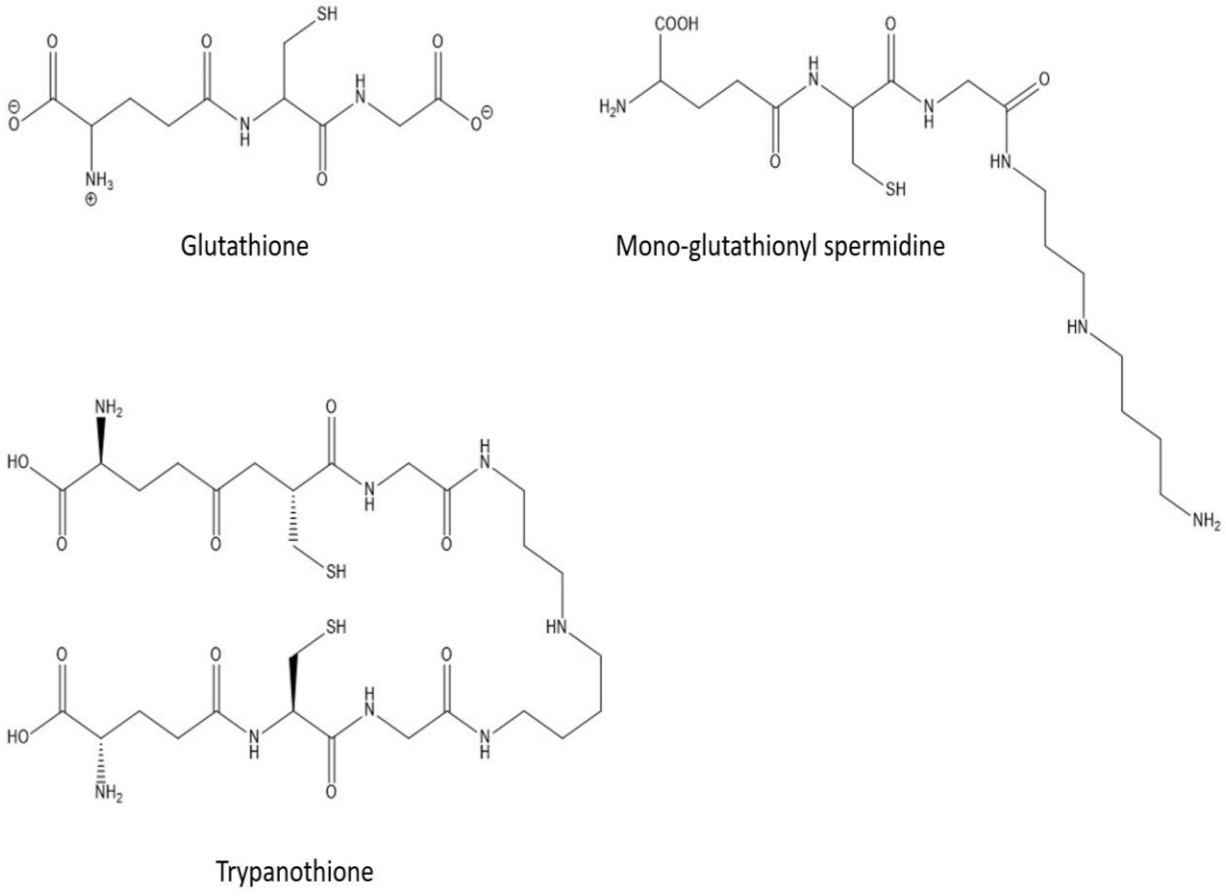


Figure 1-5 – Chemical structures of glutathione, mono-glutathionyl spermidine and trypanothione.

Mono-glutathionyl spermidine and trypanothione are used to maintain the intracellular redox state and allows for both long term and short term changes to occur depending on the concentration of the molecules present. This buffering system allows for the parasites to survive oxidative bursts that occur during host infection and to adapt to the different metabolic and environmental conditions that are required by the digenetic life cycle of trypanosomatids (Krauth-Siegel et al., 2008).

Cysteine biosynthesis and *de novo* pathway

Over the last decade, increasing evidence has emphasised the importance of cysteine biosynthesis. In trypanosomatids, cysteine synthesis can proceed through either the *de novo* pathway or the reverse trans-sulphuration pathway, Figure 1-6. In the *de novo* pathway, L-serine is modified to O-acetyl-L-serine (OAS) by serine acetyltransferase (SAT). O-acetyl-L-serine is then converted to L-cysteine in a reaction catalysed by cysteine synthase (CS) (Nozaki et al., 2001).

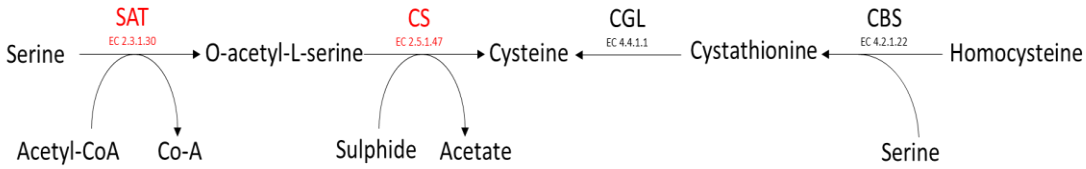


Figure 1-6 - Cysteine biosynthesis pathway in trypanosomatids. Enzymes involved in the *de novo* synthesis pathway are shown in red. SAT – serine acetyltransferase, CS – cysteine synthase, CGL – Cystathionine γ-lyase, CBS - Cystathionine β-synthase.

In *Trypanosoma cruzi,* proteomics showed in the amastigote stage CS is expressed around four times as much compared to the trypomastigote stage and around double the amount of CS protein expressed in the epimastigote stage has demonstrated that in the mammalian stages of infection, cysteine synthesis relies

heavily on the *de novo* pathway (Marciano et al., 2012). Additionally, *L. infantum* has been shown to rely on the trypanothione redox system to survive oxidative stress imposed by the mammalian immune system (Battista et al., 2020). Cysteine synthesis in *L. infantum* has not been previously studied in depth, though analysis of the genes encoding enzymes responsible for cysteine synthesis in *Leishmania major* has suggested that *L. infantum* also contains the two pathways of cysteine biosynthesis and therefore will be capable of *de novo* cysteine synthesis (Williams et al., 2009). *T. theileri* has not been studied at all in this regard, however the conservation of the genes in this pathway, not only for *T. cruzi*, *L. infantum* and *T. theileri*, but also for other related protozoan parasites such as *L. major* and *E. histolytica* clearly underlines the importance of these genes and the *de novo* pathway (Fyfe et al., 2012; Mori et al., 2015).

Serine Acetyltransferase

Serine acetyltransferase (SAT) in *T. cruzi* is responsible for the regulation of the sulphate incorporation, activation, and assimilation pathway. This is likely achieved by controlling the abundance of OAS present and is suggested to be part of a more complex mechanism to regulate thiol concentrations within the parasites. *Tc*SAT mRNA is constitutively expressed in both the insect and mammalian stages of *T. cruzi* which provides further evidence that the cysteine synthesis pathway has a regulatory role (Nozaki et al., 2001).

SAT belongs to the left-handed parallel β-helix family and is responsible for the catalysis of the production of OAS from acetyl CoA and serine (Oldham et al., 2022). There are 7 structures of apo SAT that have been experimentally determined from

Entamoeba histolytica, *Brucella abortus, Neisseria gonorrhoeae*, *Haemophilus influenzae*, *Vibrio cholerae*, *Escherichia coli* and *Yersinia pestis*. All of these organisms, excepting the eukaryotic *E. histolytica*, are bacterial prokaryotes and a phylogenetic tree is shown in Figure 1-7. A mechanism for the reaction of acetyl CoA and serine has been proposed based on *Hi*SAT, Figure 1-8.

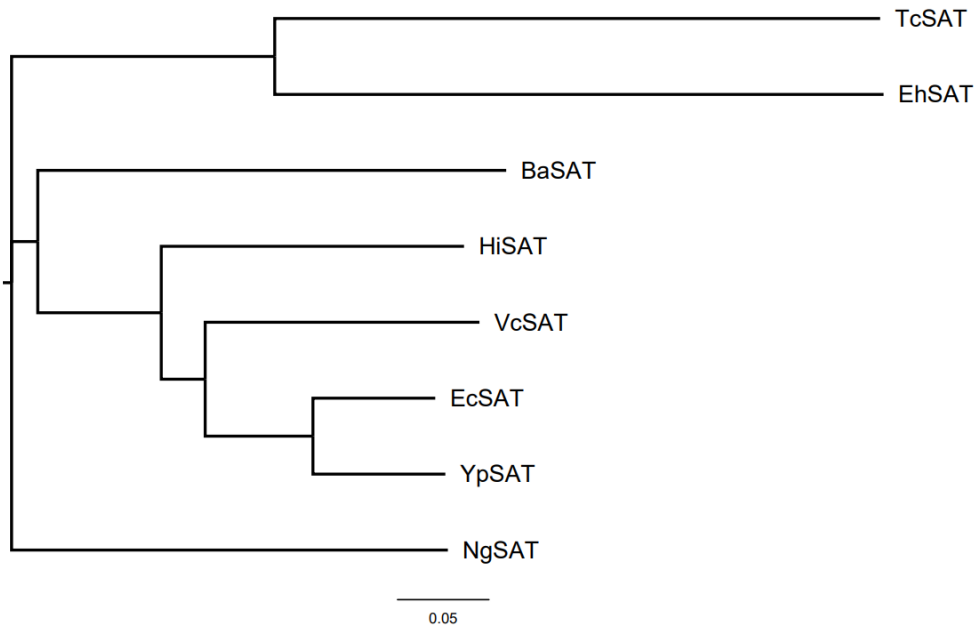


Figure 1-7 Phylogenetic tree of serine acetyltransferase (SAT) from *Trypanosoma cruzi* (Tc), *Entamoeba histolytica* (Eh), *Brucella abortus* (Ba), *Neisseria gonorrhoeae* (Ng), *Haemophilus influenzae* (Hi), *Vibrio cholerae* (Vc), *Escherichia coli* (Ec)and *Yersinia pestis* (Yp).

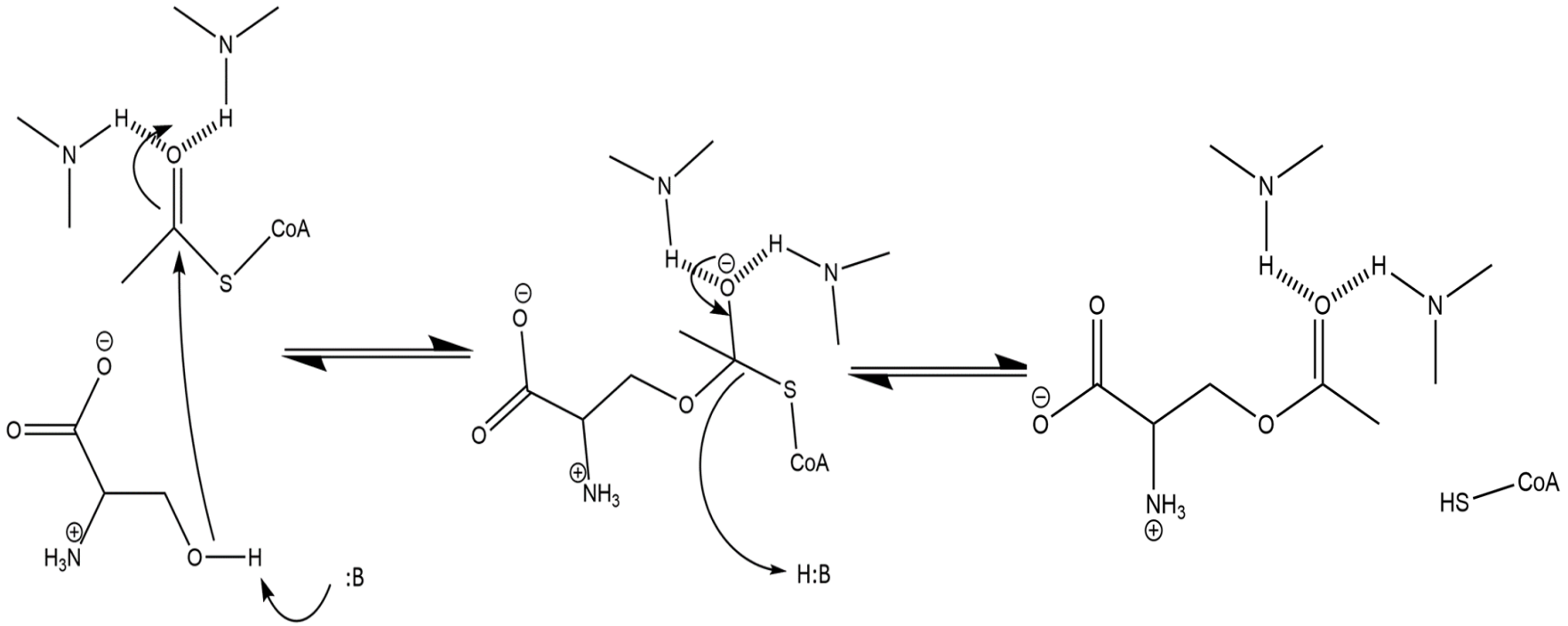


Figure 1-8 - Reaction mechanism of *Hi*SAT. From (Guan et al., 2008; Leu et al., 1994).

Sequence searches of the Protein Data Bank with the *Tc*SAT sequence found a 31% sequence identity to the *E. histolytica* serine acetyltransferase, which is unsurprising as *Eh*SAT is also a protozoan parasite (Nozaki et al., 2001). The structure of *Eh*SAT is shown in Figure 1-9. Unlike the bacterial SAT structures, *Eh*SAT is a trimer rather than a dimer of trimers. In all structures, the C-terminal left-handed parallel β-helix is highly conserved. Variation can be seen in the α helices which are responsible for the formation of complexes such as a trimer and then a hexamer as a dimer of trimers (Oldham et al., 2022). The C-terminus forms the active site for serine to bind between two adjacent monomers. A catalytic triad is formed from the imidazole group of His^{158} of one monomer, the carboxylic acid of Asp^{143} from the neighbouring monomer and the Oγ of L-serine (Pye et al., 2004).

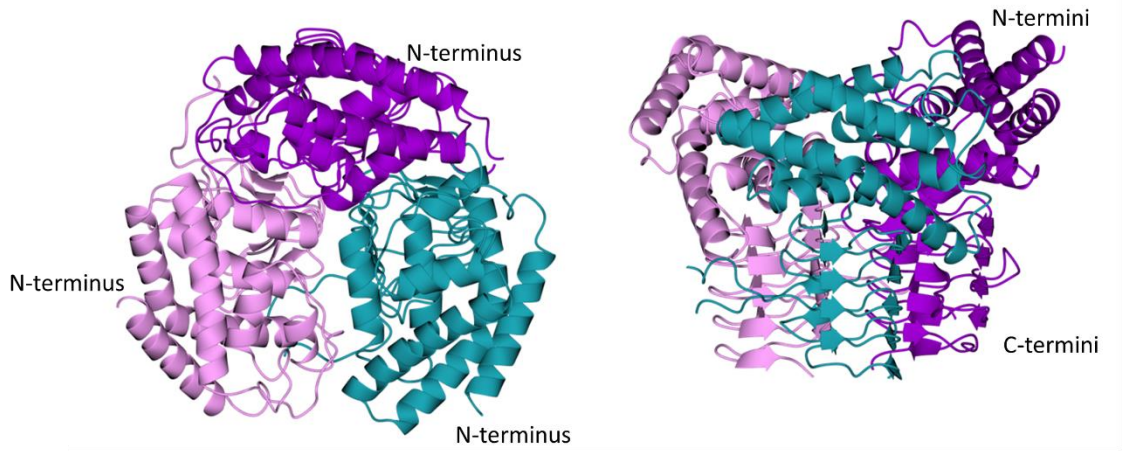


Figure 1-9 – Ribbon diagram of homotrimer of *Eh*SAT, top view (left) and side view (right). PDB 3P1B, (Kumar et al., 2011)

A literature search has not shown any human homolog of the SAT gene. SAT from *E. histolytica* was shown to be a regulated critical enzyme in the cysteine biosynthesis pathway, as it is in *T. cruzi* (Nozaki et al., 1999). *Eh*SAT was found to differ significantly from plant and bacterial SAT by being in a feedback loop which causes

inhibition of SAT at physiological concentrations of cysteine. *Eh*SAT has been shown to bind cysteine in the active site, the same place as the serine, Figure 1-10. This causes the cysteine to act as an inhibitor of *Eh*SAT (Kumar et al., 2011).

The other component of the serine to OAS conversion is acetyl CoA. The *Hi*SAT structure shows acetyl CoA binding between β helices between each monomer, Figure 1-11. *Hi*SAT shares a sequence identity of only 35% with *Tc*SAT, yet it can be assumed that the active site and molecular mechanism is conserved within eukaryotes suggesting that the binding sites would be similar for both enzymes.

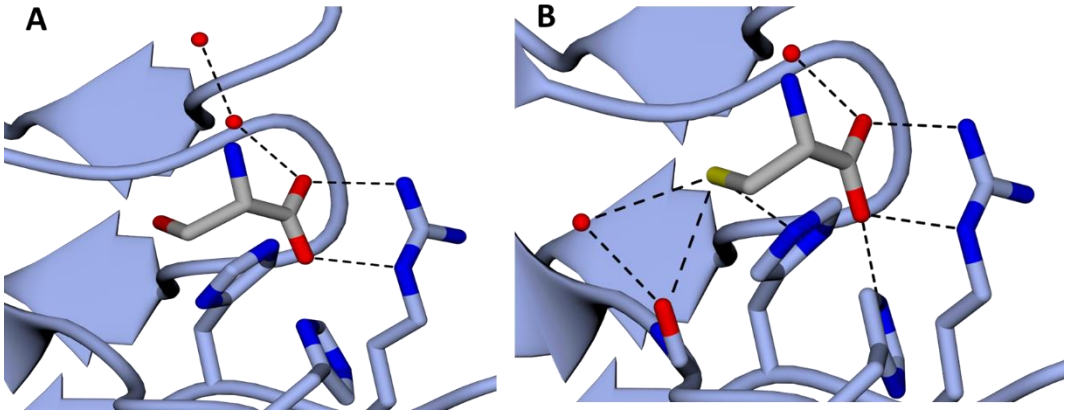


Figure 1-10 - Active site of *Eh*SAT with serine (A) and cysteine (B) bound. *Eh*SAT is shown in ribbon form with serine, cysteine and interacting amino acids shown in stick representation. Figure generated from PDB entries 3P47 and 3Q1X (Kumar et al., 2011).

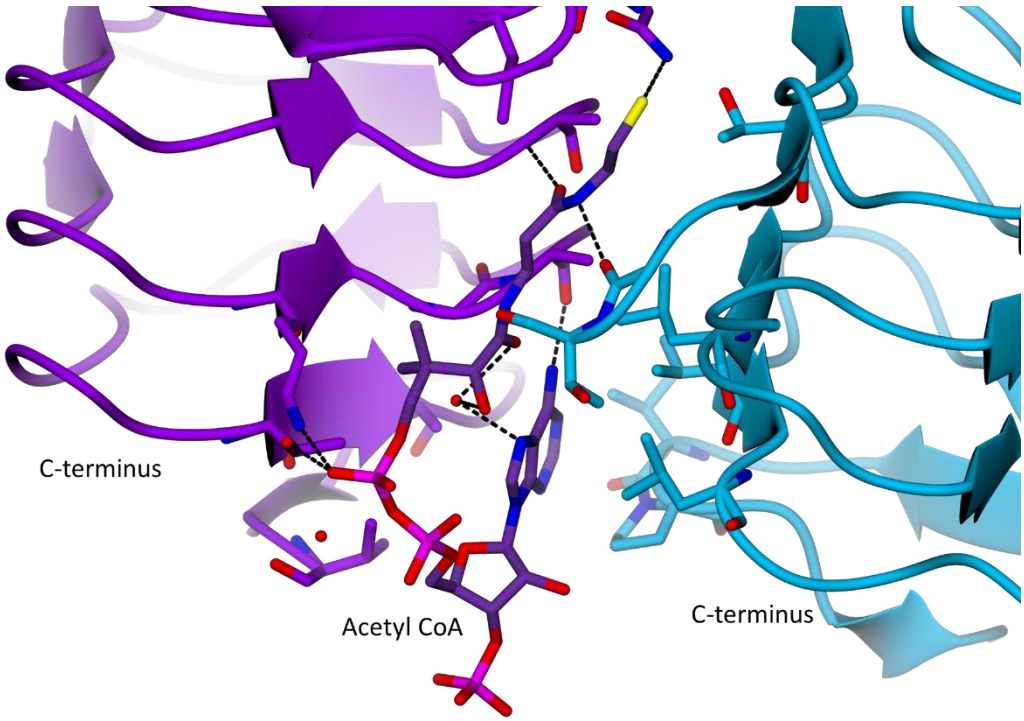


Figure 1-11 – Acetyl CoA, represented in stick form, bound to the *Hi*SAT trimer (Chain A in purple, chain B in teal) shown in ribbon form with amino acids from each monomer interacting with acetyl CoA shown in stick representation, PDB 1SST, (Olsen et al., 2004).

Residues His^{154}, Gln^{174}, Ala^{200}, Ala^{218}, Lys^{215} and Thr^{231} of one monomer and Gly^{180}, Thr^{181} and Arg^{238} of the adjacent monomer interact with the CoA. *Hi*SAT has also been shown to have serine and cysteine binding at the same location (Olsen et al.,

2004). This further supports the notion that regulation of SAT occurs through a negative feedback loop formed between SAT and cysteine.

Cysteine Synthase

Cysteine synthase (CS) belongs to the β-family of PLP-dependant enzymes and has been shown to catalyse β-elimination of the acetate group of O-acetylserine (produced by SAT) (Liang et al., 2019).

Pyridoxal 5'-phosphate (PLP) is a vitamin B6 derivative and a widely used co-factor for many enzymes in a variety of organisms (John, 1995), Figure 1-12. Approximately 4% of all classified enzyme activity is PLP-dependant (Liang et al., 2019). PLP-dependant enzymes (excepting glycogen phosphorylases) are most commonly used in amino acid synthesis, although they are also used in decarboxylation, deamination and racemisation (John, 1995).

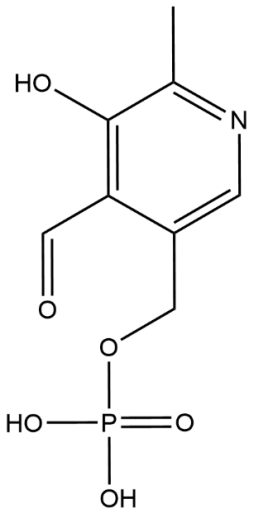


Figure 1-12 - Chemical structure of PLP

PLP possesses a heteroaromatic pyridine ring, one aldehyde group, a hydroxyl group and a phosphate group and these groups allow PLP to act as a cofactor. The aldehyde group forms an imine with a free amino group of the enzyme, called the

internal aldimine. The free amino group is typically a lysine residue. Another unprotonated primary amino group, from the substrate, can then attack the internal aldimine which makes the formation of the internal aldimine reversible. The heteroaromatic pyridine ring stabilises the carbanionic intermediate that is formed. In elimination reactions, such as the OAS to cysteine conversion, the intermediate then undergoes elimination with either α- or β-elimination occurring dependant on the leaving group at the β-position (Liang et al., 2019). This reaction mechanism for *Tc*CS is shown in Figure 1-13.

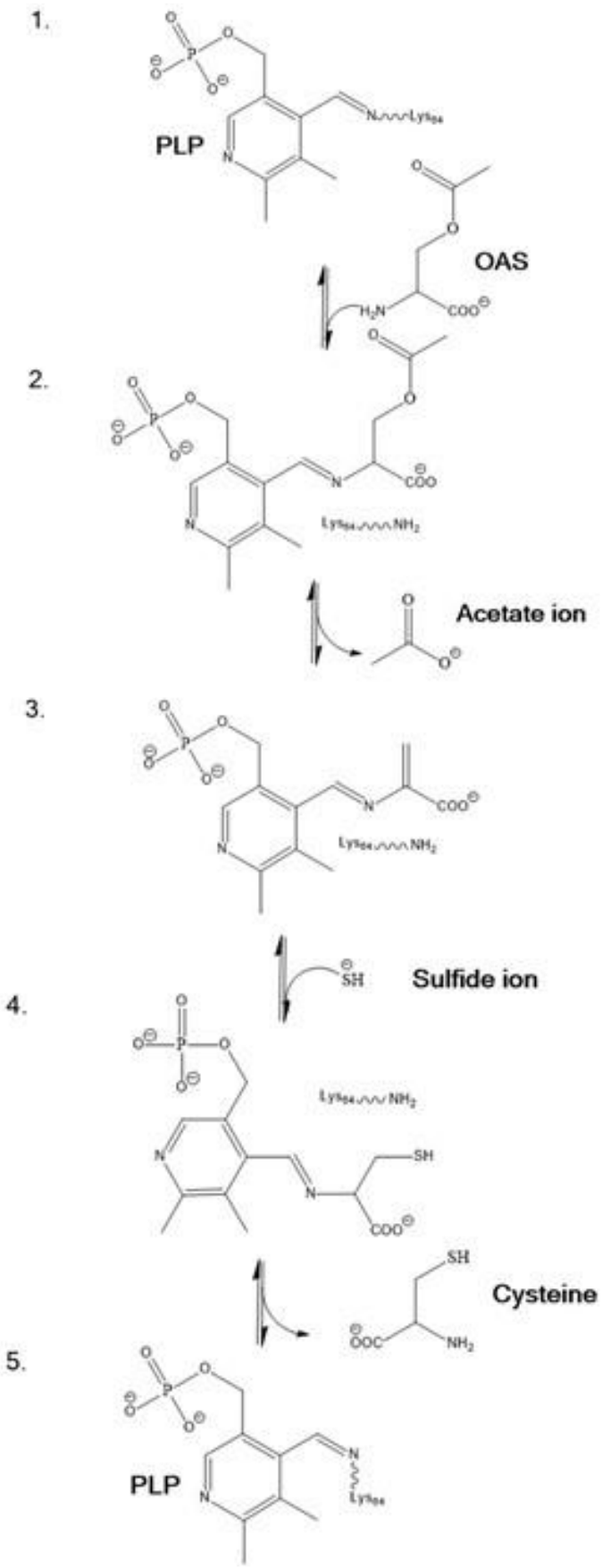


Figure 1-13 - Reaction mechanism of *Tc*CS with OAS and sulfide.

PLP-dependant enzymes are classified into 5 groups with CS belonging to the fold type II group. These enzymes have the Schiff base lysine closer to the N-terminus and the loop region for the PLP phosphate group closer to the C-terminus than in fold type I enzymes. Fold type II enzymes also have a conserved serine residue that interacts with the pyridine nitrogen (Liang et al., 2019). Fold type II enzymes are active as dimers, trimers or oligomers and many are under allosteric control.

PLP-dependant enzymes have previously been identified as potential drug targets for protozoan diseases (Kronenberger et al., 2014). *Leishmania major* cysteine synthase was predicted to be a good drug target due to the lack of human cysteine synthase or de novo cysteine synthesis pathway (Williams et al., 2009).

In *T. cruzi* studies, no cDNA for *Tc*CS was found suggesting that *Tc*CS mRNA may not be abundant in epimastigotes compared to other cDNAs such as *Tc*CBS. In vivo, *Tc*CS protein activity was detected at significant levels in both the epimastigote and the trypomastigote stages, with *Tc*CS shown to be more active in the trypomastigote stage (Nozaki et al., 2001).

Cysteine synthase has been studied in other trypanosomatids, particularly *Leishmania* species and a phylogenetic tree is shown in . *L. braziliensis* shares the two pathways of cysteine biosynthesis (Romero et al., 2015). Due to the similarities between *Lbr*CS and trypanosomatids, it can be suggested that the results in *Lbr*CS will be comparable to other trypanosomatids. *Lbr*CS was overexpressed in promastigotes, resulting in increased *Lbr*CS activity and significantly increased survival on exposure to H_2O_2 in the parasites (Romero et al., 2015). This finding is a highly suggestive indicator of the critical role of played by cysteine synthase in

preventing oxidative stress. Cysteine synthase in also found in another closely related parasite, *Trypanosoma rangeli*. *Tr*CS differs from other cysteine synthases as it has mutations in the PLP binding domain which has caused *Tr*CS to only possess two of the four lysine residues important for catalytic activity. *Tr*CS also lacks the canonical β8-β9 loop that is classic to cysteine synthase enzymes and is important for access to the active site. Further sequencing detected that *Tr*CS was not present in either epimastigote or trypomastigote forms of *T. rangeli* suggesting that this closely related parasite does not have a *de novo* cysteine biosynthesis pathway. This could be explained by the significantly lower thiol requirements found in *T. rangeli* compared to other trypanosomatid epimastigotes (Romero et al., 2014).

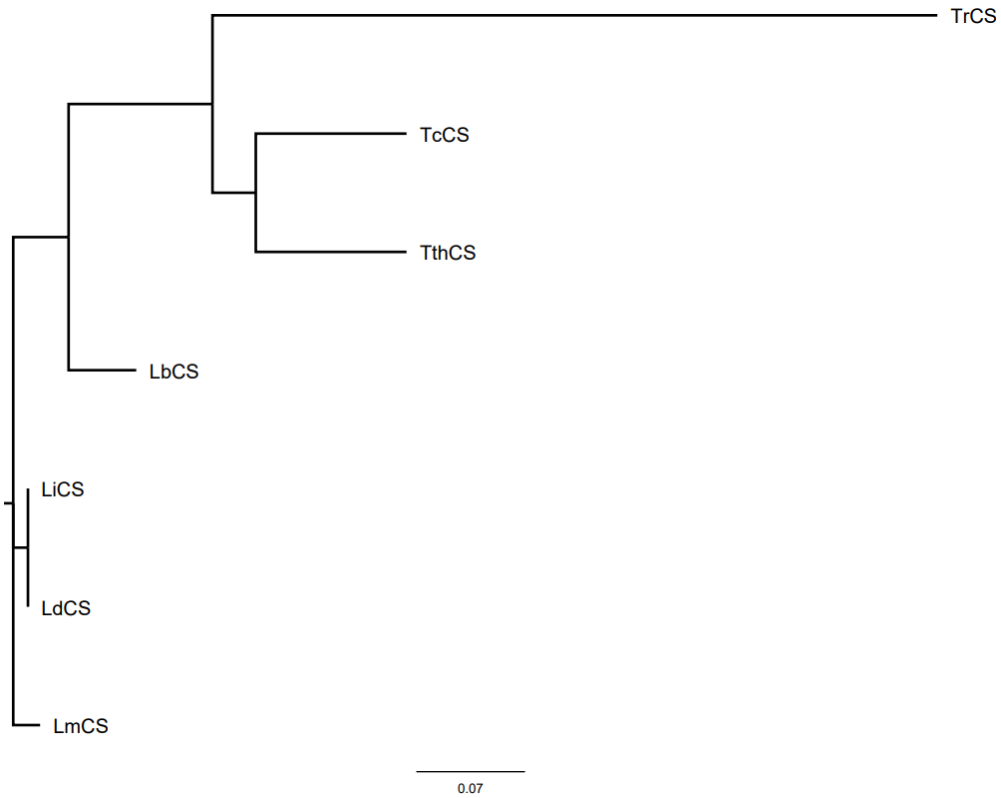


Figure 1-14 Phylogenetic tree of cysteine synthase (CS) from Trypanosoma cruzi (Tc), Trypanosoma rangeli (Tr), Trypanosoma theileri (Tth), Leishmania braziliensis (Tb), Leishmania influenzae (Li), Leishmania donovani (Ld) and Leishmania major (Lm).

The crystal structure of *Leishmania major* CS has been elucidated, Figure 1-15, and is typical of a cysteine synthase (Fyfe et al., 2012). The only other trypanosomatid cysteine synthase structure cysteine synthase determined is from *Leishmania donovani* and a superposition of the two structures is shown in Figure 1-16, (Raj et al., 2012).

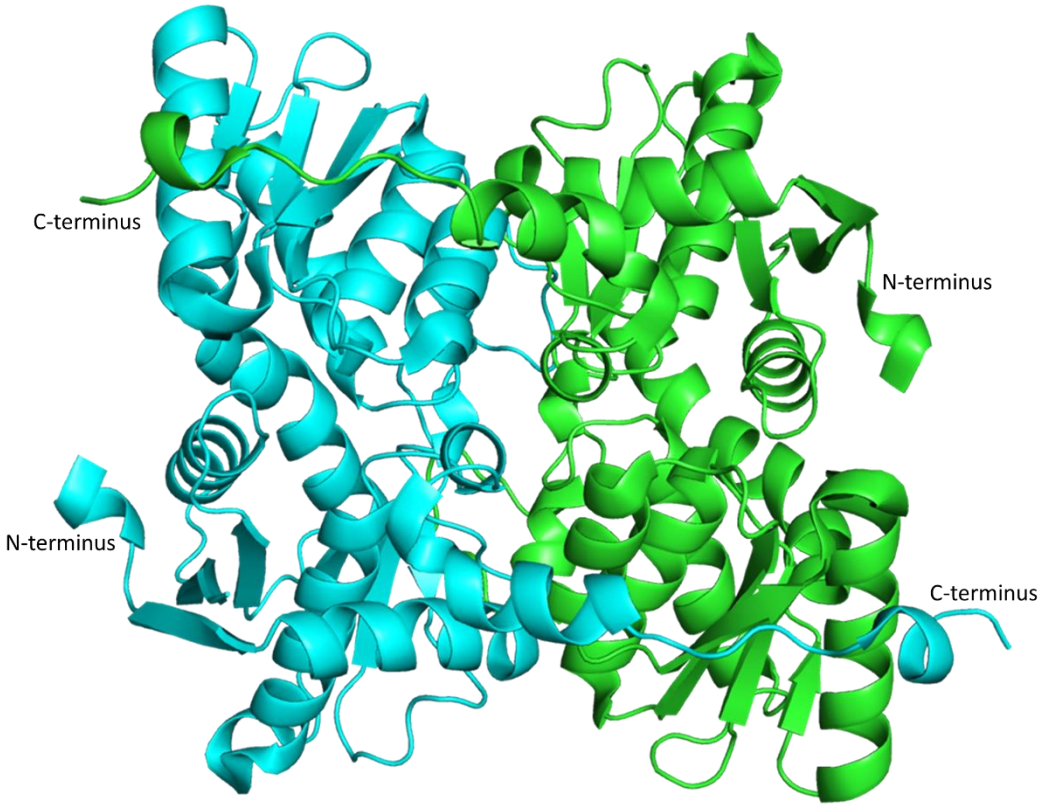


Figure 1-15 –Homodimer of *Leishmania major* cysteine synthase shown in ribbon representation, (Fyfe et al., 2012) Chain A is shown in blue, chain B in green.

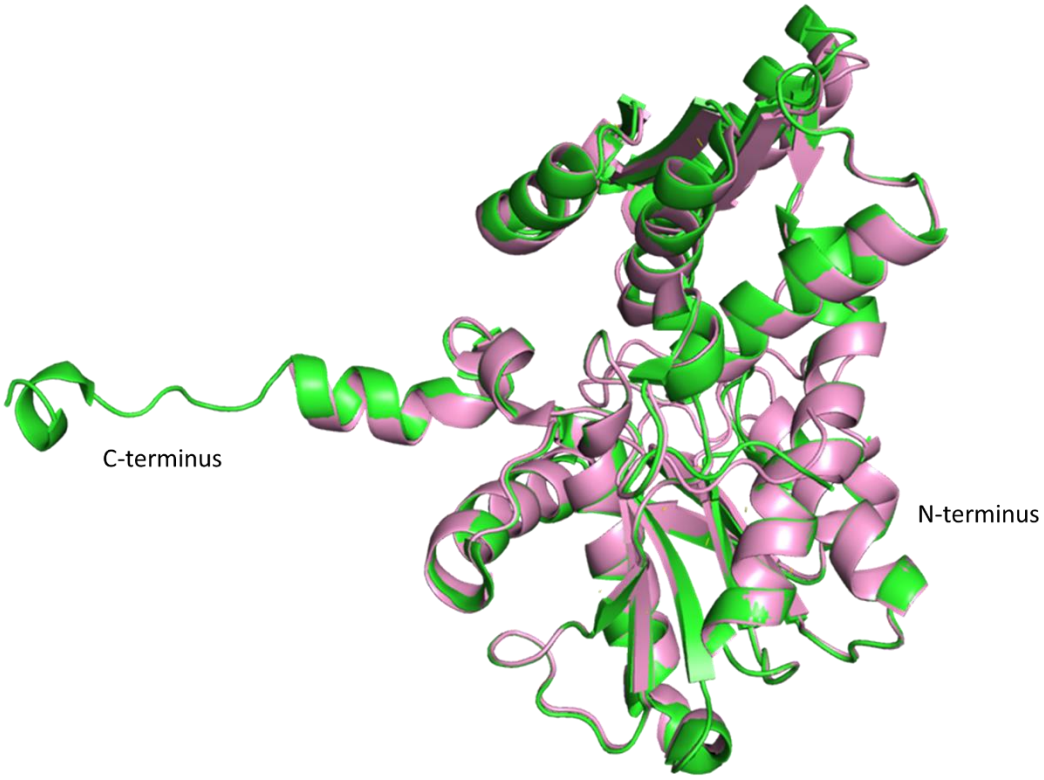


Figure 1-16 - Superposition of A subunits from *Lm*CS (green, Fyfe et al., 2012) and *Ld*CS (pink, Raj et al., 2012) in ribbon form.

These two structures are almost identical to one another with a sequence identity of 97% and also are similar to other cysteine synthase structures present in the PDB. From this it can be suggested that the structures of other trypanosomatid cysteine synthases will be alike.

Cysteine Synthase Complex

The cysteine synthase complex (CSC) was first discovered in *S. typhimurium* and is formed from SAT and CS (Kredich et al., 1969). Despite the CSC discovery occurring over 50 years ago, very little is known of the CSC. The high affinity interaction between SAT and CS is highly conserved through species however no structure has been solved for this complex (Benoni et al., 2017). A study in *E. coli* has suggested than the CSC forms through three stable intermediates (Wang et al., 2012). Size

exclusion chromatography and sedimentation ultracentrifugation experiments with CSC from soybeans indicated that CSC was composed of one SAT trimer with three CS dimers (Kumaran et al., 2009). However, another experiment using hydrogen/deuterium exchange mass spectrometry suggested that the composition of CSC from *E. coli* is a dimer of trimers of SAT with each trimer binding to one CS homodimer (Rosa et al., 2021).

CSC has been identified both a sensor and as a component of the regulatory system controlling cellular sulfur homeostasis in plants (Wirtz et al., 2006). Additionally, formation of CSC enhances SAT activity by increasing the production of OAS – the rate limiting step in the biosynthesis pathway (Kumaran et al., 2009).

Project Aims

Cysteine biosynthesis in trypanosomatids is a crucial survival mechanism allowing cysteine to be incorporated into the production of trypanothione. Trypanothione, in turn, allows the trypanosomatid to survive the oxidative bursts imposed by the host immune system upon invasion of the human host. This pathway provides a new opportunity to combat the neglected tropical diseases Chagas disease, leishmaniasis and *T. theileri* infection.

Studies of cysteine biosynthesis in trypanosomatids has revealed the *de novo* pathway composing of serine acetyltransferase and cysteine synthase (Nozaki et al., 2001). Serine acetyltransferase catalyses the conversion of serine to OAS and is likely to be regulated through a negative feedback mechanism involving cysteine (Kumar et al., 2011). 7 apo structure of SAT have been determined, all forming homotrimers with 6 of the 7 forming dimers of trimers. Cysteine synthase is

responsible for the conversion of OAS to cysteine using PLP as a cofactor to facilitate the reaction (Liang et al., 2019). Structures of many cysteine synthases from a variety of organisms including two *Leishmania* species have been determined and indicate that trypanosomatid CS structures will be homodimers. SAT and CS have been found to form a bienzymatic complex - the cysteine synthase complex, yet it has only been briefly studied (Benoni et al., 2017).

Serine acetyltransferase and cysteine synthase have previously been identified as essential enzymes and promising potential drug targets for treatment of Chagas disease, leishmaniasis and *T. theileri* infection (Silber Lab, unpublished).

The aims of this project were therefore to generate greater understanding of cysteine biosynthesis in trypanosomatids through structural and biochemical investigation into the two proteins critical to the pathway. Exploration of the molecular mechanisms in atomic detail was be critical to this aim. This study was undertaken through expression and purification of both SAT and CS. Following this success, structural studies were conducted primarily through X-ray crystallography. Biochemical characterisation of the activity of the proteins were conducted. These experiments aimed to provide a more complete view of cysteine production in trypanosomatids.

Further, this project aimed to provide the first insights into the cysteine synthase complex in trypanosomatids by producing the cysteine synthase complex, carrying out investigations of the biochemistry involved and using electron microscopy will begin to explore the structure of this complex.

Finally, by using the structures determined in this project, inhibition of the *de novo* pathway of cysteine biosynthesis will be explored. This will involve exploration into known inhibition methods, development of different techniques to identify molecules binding to the proteins and additionally, a structural based drug design approach to design *de novo* inhibitors of this pathway will be conducted.

Chapter 2: Materials and Methods

General

All chemicals used were purchased from Sigma-Aldrich, unless specified. Competent cells were purchased from New England Biolabs. Optical density at 600 nm (OD_{600}) measurements, DNA and protein concentration estimates were made using a DeNovix D5-11+ Spectrophotometer. Protein molecular weights, theoretical isoelectric points and estimated extinction coefficients were calculated using ExPASy (Gasteiger et al., 2005). Sequence analysis was performed using BLAST searches and alignments performed using Clustal Omega and aligned with protein sequences using ESPript 3 (Boratyn et al., 2013; Madeira et al., 2019; Robert & Gouet, 2014). All biochemical and biophysical data were processed and graphs produced using GraphPad Prism 9. NMR data were analysed using MestRe Nova (Willcott, 2009). All images were produced using Pymol and CCP4mg (McNicholas et al., 2011; Schrödinger LLC, 2015)

Buffer Table

Unless specified all buffer volumes are 1 L.

Table 2-1 - Buffers used.

Buffer	Ingredients
HisTrap Binding Buffer	50 mM Tris-HCl pH 7.5 500 mM NaCl 30 mM Imidazole 2 mM β-mercaptoethanol
HisTrap Elution Buffer	50 mM Tris-HCl pH 7.5 500 mM NaCl 500 mM Imidazole 2 mM β-mercaptoethanol
SDS Running Buffer (10x)	250 mM Tris Base 190 mM Glycine 35 mM SDS

SGC Buffer (3x)	60 mM Hepes pH 7.5 900 mM NaCl 30 % w/v glycerol
SEC Buffer (1x)	60 mM Hepes pH 7.5 900 mM NaCl
Lysogeny Broth	10 g Casein Digest peptone 5 g Sodium Chloride 5 g Yeast Extract
Mass Spectrometry Buffer	10 mM Ammonium Bicarbonate pH 7.5
IBP Buffer 1	50 mM Tris-HCl pH 8.5 500 mM NaCl 30 mM Imidazole 2 mM β-mercaptoethanol 0.8% Igepal®
IBP Buffer 2	50 mM Tris-HCl pH 8.5 500 mM NaCl 30 mM Imidazole 2 mM β-mercaptoethanol 1.5% Igepal®
IBP Buffer 3	50 mM Tris-HCl pH 8.5 1 M NaCl 2 mM β-mercaptoethanol
IBP Buffer 4	50 mM Tris-HCl pH 8.5
IBP Buffer 5	50 mM Tris-HCl pH 8.5 8 M Urea
Refolding Buffer	50 mM Tris-HCl pH 7.8 150 mM NaCl 150 mM $MgCl_2$
Fluorescence Assay Buffer	100 mM Hepes pH 7.5
Cysteine synthase activity buffer	50 mM Tris pH 7.5 2 mM β-mercaptoethanol
Serine Acetyltransferase Activity Buffer	100 mM Tris pH 7.5
NMR buffer	50 mM Tris pH 7.5 100 mM NaCl

Antibiotics

Ampicillin was used at a final concentration of 100 μgmL^{-1} where appropriate and was used in the selection of transformed cells. Kanamycin was used at a final concentration of 30 μgmL^{-1} where appropriate.

Bacterial strains and plasmids used

All *E. coli* expression strains and cloning vectors used are shown in Table 2-2 and Table 2-3. Plasmid maps can be found in the appendix (Figures S9-3-1, S9-3-3, S9-3-5, S9-3-7, S9-4-2, S9-4-4, S9-4-6). Cloning of the sequences into the selected vector was conducted by Genscript to create expression constructs. Plasmids were centrifuged at 6000 RPM for 1 min at 4°C. Following this, 20 µL of water was added and plasmids were vortexed for 1 min. Plasmids were diluted to 5 ngμL^{-1} in water for use in transformations.

Table 2-2 - Plasmids used.

Plasmid used	Protein gene contained	Tag encoded	Antibiotic resistance conferred	Expression system used
pJOE5751	*Trypanosoma cruzi* Cysteine Synthase	N-terminal 6xHis	Ampicillin	Arabinose
pJOE5751	*Trypanosoma cruzi* Serine Acetyltransferase	N-terminal 6xHis	Ampicillin	Arabinose
pET28a	*Trypanosoma cruzi* Serine Acetyltransferase	N-terminal 6xHis	Kanamycin	T7
pET50b	*Trypanosoma cruzi* Serine Acetyltransferase	N-terminal 6xHis, NusA	Kanamycin	T7
pGEX-4T-3	*Trypanosoma cruzi* Serine Acetyltransferase	N-terminal GST	Ampicillin	T7
pMAL-p5E	*Trypanosoma cruzi* Serine Acetyltransferase	N-terminal maltose	Ampicillin	T7
pET15b	*Leishmania infantum* cysteine synthase	N-terminal 6xHis	Ampicillin	T7
pET15b	*Trypanosoma theileri* cysteine synthase	N-terminal 6xHis	Ampicillin	T7

Transformation of recombinant plasmid into competent cells

Table 2-3 shows the bacterial strains used.

Table 2-3 - Bacterial strains used.

Strain Used	**Use**	**Genotype**
DH5α	Storage of plasmids	fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17
BL21	Expression of recombinant proteins	fhuA2 [lon] ompT gal [dcm] ΔhsdS
BL21 (DE3)	Expression of recombinant proteins	fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS λ DE3 = λ sBamHIo ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5
SHuffle T7 Express	Expression of recombinant proteins	*fhuA2 lacZ::T7 gene1* [lon] *ompT ahpC gal λatt:*::pNEB3-r1-*cDsbC* (Spec^{R}, *lacI^{q}*) Δ*trxB sulA11 R(mcr-73::miniTn10*--Tet^{S})2 [dcm] *R(zgb-210::Tn10* --Tet^{S}) *endA1 Δgor Δ(mcrC-mrr)114::IS10*

Lemo21(DE3)	Expression of recombinant proteins	*fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS/ pLemo(CamR) λ DE3 = λ sBamHIo ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5 pLemo = pACYC184-PrhaBAD-lysY*

All cells used were purchased from NEB. Transformations of plasmids into competent cells were all performed according to the manufacturer's instructions.

Glycerol Stock Preparation

Glycerol stocks were prepared by adding 500 µL of overnight culture to 500 µL of sterile 50% glycerol and stored at -80°C.

General Protein Expression

Small Scale expression tests

Glycerol stocks were used to generate two 100 mL cultures containing appropriate antibiotic, grown at 30°C (if SHuffle T7 Express cells) or 37°C (all other cell types), 150 RPM. Cultures were grown to OD_{600} 0.4-0.6 before induction of one culture with L-Rhamnose to a final concentration of 0.2% for pJOE5751 vector or 1 mM IPTG all other vectors. The other culture was grown without L-Rhamnose or IPTG. Induced and non-induced cultures were grown at various temperatures for an additional 16 hours and 1 mL samples were taken from each culture 2 hours, 4 hours and 16 hours after induction.

Solubility Tests

Glycerol stocks were used to generate two 100 mL cultures containing appropriate antibiotic, grown at 30°C, 150 RPM. Cultures were grown as described for small scale expression tests. Normalised samples were taken and analysed by whole cell solubility test SDS-PAGE.

Expression of recombinant proteins in *E. coli*

Proteins were expressed as fusion proteins containing an N-terminal Tag encoded by the vector. 25 mL cultures, containing appropriate antibiotics, were produced from transformed *E. coli* BL21 (DE3) or SHuffle T7 Express cells. Cultures were grown overnight at 30°C or 37°C (SHuffle T7 Express/BL21 (DE3)), 150 RPM. These cultures were used to initiate 1 L cultures with appropriate antibiotic added. These cultures were grown at 30°C or 37°C, 150 RPM to mid-log phase with OD_{600} measuring 0.4 to 0.6 and induced with 0.2% Rhamnose or 1 mM IPTG. Cells were then grown for 2, 4 or 16 hours. Cells were centrifuged at 4000 RPM, 4°C for 25 minutes, using a Beckman Avanti Hi-Speed centrifuge. The supernatant was discarded, and pellets were combined into 2 L pellets. The pellets were then frozen at -80°C. The pellet provided the starting point for protein purification.

Immobilised metal ion affinity chromatography

Before first use, the IMAC column was washed with 5 column volumes of MilliQ water then with 5 column volumes of HisTrap Elution Buffer. Finally, the column was washed with 10 column volumes of HisTrap Binding Buffer. The pelleted cells were resuspended in 30 mL of HisTrap Binding Buffer, unless otherwise stated, with protease inhibitor added. The resuspended cells were sonicated on ice for 2

minutes at 2 x 10% cycle and at 50% power. The resulting lysate was centrifuged using a Beckman Avanti Hi-Speed centrifuge at 20 000 RPM, 4°C for 50 minutes. The supernatant was then filtered through a 0.45 µm filter.

The supernatant was loaded onto a 5 mL His-Trap FF Column (Cytiva) using a peristaltic pump. The column was then attached to an ÄKTA Pure and eluted by fast protein liquid chromatography using a linear gradient of HisTrap Binding and Elution Buffers. The supernatant, column flow through and initial wash were collected and a 10 µL sample taken. The eluted fractions were collected, and samples were taken to be prepared for SDS-PAGE analysis. Samples were prepared by mixing 10 µL of protein with 10 µL of 2x SDS-PAGE dye. Samples were then heated for 10 minutes at 96°C. Finally, samples were briefly spun down before being loaded onto a gel for analysis.

Purification of insoluble protein

Purification from inclusion bodies

The pelleted cells were resuspended in 30 mL of IBP Buffer 1 with a protease inhibitor tablet added. The resuspended cells were sonicated on ice for 2 minutes at 2 x 10% cycle and at 50% power. The resulting lysate was centrifuged using a Beckman Avanti Hi-Speed centrifuge at 15 000 RPM, 4°C for 20 minutes. Samples of both the pellet and the supernatant were taken and the supernatant was discarded. The pellet was resuspended in 10 mL of IBP Buffer 2, sonicated on ice for 2 minutes at 2 x 10% cycle and at 50% power and then a further 20 mL of IBP Buffer 2 was added. The lysate was centrifuged at 15 000 RPM, 4°C for 20 minutes. This was repeated with IBP Buffer 3 and IBP Buffer 4. The resulting pellet was resuspended in

30 mL of IBP Buffer 5. After 30 minutes the lysate was centrifuged at 15 000 RPM, 4°C for 20 minutes and the supernatant was collected.

Protein Refolding and Tag Cleavage

The supernatant from the purification from inclusion bodies was placed in dialysis tubing and placed in Refolding Buffer with HRV 3C at 4°C overnight. The protein was then passed through an immobilised nickel column to remove cleaved tag and uncleaved protein from the sample using an ÄKTA pure.

Protein Dialysis

The eluted fractions containing protein were pooled and placed in dialysis tubing which was placed in 1x SGC buffer at 4°C overnight.

Protein concentration

Protein was placed in a Vivaspin^{TM}20 MWCO 5000 concentrator (Cytiva) and spun at 4000 RPM, 4°C until the desired concentration was reached.

Chemical cross linking of protein samples

Protein samples were cross linked using 0.1% glutaraldehyde for 15 minutes. After 15 minutes the reaction was quenched with 1 M Tris pH 7.5.

Size exclusion Chromatography

Gel filtration

Size exclusion chromatography was carried out using columns listed in Table 2-4 using an ÄKTA Pure FPLC. Protein was injected onto the column and eluted according to manufacturer's instructions. Protein was eluted and samples were stored at 4°C until required.

Table 2-4 - Gel filtration columns used.

Gel filtration column used	Used to separate proteins of size (Da)	Purchased from
HiLoad® 16/600 Superdex® 75 PG	3 000 – 70 000	Cytiva
Superdex® 75 Increase 10/300 GL	3 000 – 70 000	Cytiva
HiLoad® 16/600 Superdex® 200 PG	10 000 – 600 000	Cytiva
HiPrep™ 16/60 Sephacryl® S-300 HR	1 000 – 1 500 000	Cytiva
HiPrep™ 16/60 Sephacryl® S-400 HR	2 000 – 8 000 000	Cytiva
HiPrep™ 16/60 Sephacryl® S-500 HR	40 000 – 20 000 000	Cytiva

Estimation of molecular weight

An estimation of the molecular weight of the protein loaded on to the gel filtration column can be calculated by using a Gel Filtration Calibration Kit (Cytiva) according to the manufacturer's instruction. Blue dextran is loaded to determine the void volume of the column. The elution volumes of the recommended protein mix are recorded and the K_{av} is determined by the equation:

$$K_{av} = \frac{V_e - V_o}{V_c - V_0}$$

A calibration curve of K_{av} versus log molecular weight is created and the equation of the line is determined (y=mx+b). For the protein of interest, the elution volume is used to determine the K_{av} and the K_{av} is used to determine the log molecular weight with the following equation:

$$Log\ MW = \frac{b - K_{av}}{m}$$

The log molecular weight is raised to the power 10 and this provides an estimate of the molecular weight.

SDS-PAGE analysis of protein samples

SDS-PAGE gels were made using either an 8%, 10%, 12% or 14% resolving gel solution and stacking gel solution. Gels were run at 200 V for 60 minutes. Gels were stained with InstantBlue™ Protein Stain (Expedeon) and imaged using a BioRad Molecular Imager® GelDoc™ XR+ with ImageLab™ Software. 5 µL of a 10-250 kDa PageRuler™ Prestained Protein Ladder (Thermo Scientific) was loaded onto each gel.

Preparation of Samples for SDS-PAGE Analysis

Whole cell samples

For analysis of whole cells, 1 mL of culture was spun for 4 minutes at 14 500 RPM. The supernatant was discarded and the pellet resuspended in 15 µL 4xSDS-loading dye and 50 µL of buffer for T=0 and T=2 samples, 30 µL of 4xSDS-loading dye and 100 µL of buffer for T=4 samples and 100 µL of 4xSDS-loading dye and 300 µL of buffer for T=16 samples. Samples were boiled at 96°C for 20 minutes and then spun for 20 minutes at 14 500 RPM.

Whole cell solubility test

For analysis of whole cells from the solubility test, normalised samples of culture were taken and spun for 4 minutes at 14 500 RPM. The supernatant was discarded and the pellet resuspended in 30 µL 4xSDS-loading dye and 100 µL of buffer. Samples were boiled at 96°C for 20 minutes and then spun for 20 minutes at 14 500 RPM. Supernatant and pellet were separated and the pellet was resuspended in a

further 30 µL 4xSDS-loading dye and 100 µL of buffer before boiling at 96°C for 20 minutes.

Protein samples

10 µL of 2xSDS-loading dye was added to 10 µL of sample and the mixture was boiled at 96°C for 10 minutes. The samples were briefly centrifuged and 10 µL loaded onto the gel.

Mass Spectrometry

Protein was exchanged into Mass Spectrometry Buffer and concentrated to ~ 1 $mgmL^{-1}$ using a Vivaspin™ 500 MWCO 5000 spin column (Cytiva). 100 µL of sample was analysed by time-of-flight mass spectrometry (TOF-MS) using a QToF Premier Spectrometer (Walters) and spectra collected by the Mass Spectrometry Service in the Department of Chemistry, Durham University. LC-ESI mass spectrometry was performed by the Mass Spectrometry Service in the Department of Biology, Durham University.

Circular Dichroism

Protein was dialysed into SGC buffer and were CD measurements performed by Professor Elizabeth Bromley, Department of Physics, Durham University.

Crystal Structure Determination

General

All initial protein crystallisation screens were conducted using a vapour diffusion method with 96-well 2 lens sitting drop MRC plate. All crystallisation optimisation screens were conducted using a vapour diffusion method with 22 mm 24 well sitting

drop plates or 24 well hanging drop plates. Unless stated, proteins were dialysed into 1x SGC buffer and concentrated for crystallisation.

Crystallisation Experiments

Initial high-throughput screening was conducted using a Mosquito® Crystal (TTP Labtech). 96-condition screens tested were: BCS, Clear Strategy™ Screen I, Clear Strategy™ Screen II, JCSG-*plus* (Page et al., 2003), The LMB Crystallization Screen™, MIDAS*plus*™, Morpheus®, Morpheus® II, PACT *premier*™, PGA™ Eco, Structure Screen 1 + 2, ShotGun 1 (all from Molecular Dimensions). Screen conditions and protein samples were combined in two ratios for each screen: 1:1 (100 nL protein: 100 nL screen condition) and 2:1 (200 nL protein: 100 nL protein condition). Screens were then sealed and checked using Leica MZ16 Stereomicroscope.

Optimisation Screens

Optimisation screens were designed based on the original contents of the screen. Screens were designed, solutions were produced and filtered through a sterile 0.45 µm filter before storage at 4°C.

Sitting Drop Crystallisation Plates

500 µL of optimisation screen was added to the appropriate well. For 1:1 conditions, 1 µL of protein was added to 1 µL of screen. For 1:2 conditions, 1 µL of protein was added to 2 µL of screen. For 2:1 conditions, 2 µL of protein was added to 1 µL of screen. Plates were then sealed and viewed using a Leica MZ16 Stereomicroscope.

Hanging Drop Crystallisation Plates

The rims of each well were greased with silicon vacuum grease. 500 µL of optimisation screen was added to the wells. On 22 mm glass slides, a 1 µL and a 2

µL drop of protein was added. To each drop 1 µL of well solution was added. The glass slides were then inverted and placed on the wells. Plates were then sealed and viewed using a Leica MZ16 Stereomicroscope.

Protein-ligand cocrystallisation trials

Both sitting drop and hanging drop crystallisation plates were prepared and 500 µL of optimisation screen was added to the appropriate well. Protein was incubated with ten times the molar concentration of ligand for 1 hour. Both sitting drop and hanging drop trays were then set up as described above with the protein-ligand mixture.

Seeding of crystals

Seed stocks were created by manually crushing crystals and adding to a pre-cooled Eppendorf containing 100 µL of reservoir. A PFTE seed bead was added, and crystals were vortexed at 30 second intervals followed by 30 seconds on ice for 5 minutes (Bergfors, 2007). Sitting drop trays were set up using the Mosquito crystal (TTP Labtech) and Swiss Ci3 plates (Molecular Dimensions). 30 µL of reservoir were added to each well. Drops were set up with 100 nL of reservoir, 100 nL of protein and 20 nL of seed stock. Plates were sealed and left for crystals to grow.

Soaking of crystals with small molecules

The molecule library was defrosted and spun down at 1000 RPM, 4°C, 2 mins. Crystals were fished and placed into 1:1 drops of 1 µL of compound stock: 1 µL of reservoir for 10 minutes. Crystals were then cryoprotected and mounted.

Cryoprotection and Mounting of Protein Crystals

Crystals were cryoprotected with 25% glycerol (1 µL 50% glycerol:1 µL crystallisation reagent) prior to flash cooling in $N_{2(l)}$ for storage and transport to a synchrotron (Teng, 1990).

Diffraction Data Collection and Processing

General

Data were collected on the I-03, I-04 and I-24 beamlines at the Diamond Light Source and on the BioMAX beamline at MAX IV. All data were analysed using the CCP4i2 suite (Winn et al., 2011). Xia2/Dials was used to manually process the diffraction data (Winter et al., 2013). Molecular replacement was performed using PHASER to search for molecules in the asymmetric unit based on the Matthews factor (Matthews, 1968; Mccoy et al., 2007; Potterton et al., 2018). The electron density maps were then improved using Parrot (Cowtan, 2010). Buccaneer was then used to build the initial protein side chains and residues were assigned to a chain (Cowtan, 2006). Coot was used for manual model building incorporating water and ligands into the structure (Emsley et al., 2010). REFMAC was used to perform iterative cycles of refinement using local non-crystallographic symmetry restraints where applicable and refinement was monitored using R_{free} (Brünger, 1992; Headd et al., 2014; Murshudov et al., 2011; Usón, Pohl, et al., 1999).

TcCS

The *Tc*CS diffraction data were processed using Xia2/DIALS. Due to pseudosymmetry within the crystal (β is almost 90°), initial structure solution in the suggested $P2_12_12_1$ space group failed. The data were reprocessed in the $P2_1$ space group and

molecular replacement using *L. major* cysteine synthase as the search model (PDB ID 4AIR) (Fyfe et al., 2012), which shares a sequence identity of 78% with *Tc*CS, was successful. *Lm*CS was selected using the PDB sequence alignment tool (Berman et al., 2000) and aligned to the *Tc*CS sequence using CLUSTALW (Madeira et al., 2019). Based on the Matthews factor, four molecules in the asymmetric unit was expected and therefore the search for four monomers was conducted (Matthews, 1968; Mccoy et al., 2007; Potterton et al., 2018).

LiCS and TthCS

The refined structure of *Tc*CS was used as a search model during molecular replacement with PHASER for both *Li*CS and *Tth*CS (Mccoy et al., 2007). Due to the Matthews factor, the search was conducted using 2 molecules of the *Tc*CS structure in the asymmetric unit (Matthews, 1968).

Data processing of diffraction data collected from FragLite cocrystallisation experiments

For FragLites data collected, structures were solved as described above using either the *Tc*CS or *Li*CS model. During data processing, Friedel pairs were kept separated to allow an anomalous dispersion difference map to be generated.

Negative staining and transition electron microscopy

Crosslinked protein was applied to a 400 mesh carbon grid and stained with 1% uranyl acetate for 10 minutes. Images were taken using a Hitachi H7600 TEM at 100 kV.

^{19}F NMR

Ligand stocks were produced using NMR buffer to dissolve the compounds. Purified

TcCS was dialysed into NMR buffer overnight and samples for NMR were prepared with 1 mM ligand, 100 uM protein and 10% D_2O. Ligand samples were prepared to 1 mM with 10% D_2O. ^{19}F NMR was conducted by the Solution-State NMR Facility at Durham University.

Fluorescence Assays

Protein was dialysed into fluorescence assay buffer and stock solutions of amino acids were made up in fluorescence assay buffer. Ligands were added in quadruplicate to a 96 well plate at the concentrations 0, 0.1, 0.25, 0.5, 0.75, 1, 2, 3, 4, 5, 10 and 25 mM. Protein was added to a 96-well plate at 200 nM. Fluorescence assay buffer was added to a final volume of 100 µL. The plate was incubated at room temperature for 5 minutes. A spectrum fluorescence reading was taken following excitation at 412 nm with the spectrum reading every 10 nm between 450 nm and 650 nm on a BioTek™ Synergy™ H4 Hybrid Microplate Reader.

Activity Assays

500 µL Scale Cysteine Synthase Activity

Protein was dialysed into activity assay buffer and stock solutions of OAS and sodium sulfide were made up in activity assay buffer. The reaction was carried out in 1.5 mL Eppendorf tubes. OAS was added at 0, 0.1, 0.25, 0.5, 1, 2.5, 5, 7.5, 10, 15 and 20 mM. Sodium sulfide was added at 3 mM and protein was added at 150 nM. Activity assay buffer was added to a final volume of 500 µL. The Eppendorf tubes were sealed and the reaction proceeded at 37°C for 30 minutes. After, the tubes were unsealed and 500 µL of acetic acid and 500 µL of ninhydrin solution was added. The tubes were sealed and heated at 96°C for 10 minutes. After this, the 3.5

mL of ice-cold ethanol was added; 1 mL of the reaction mixture was used for absorbance readings, taken at 560 nm using a Spectrostar nano plate reader.

50 µL Scale Cysteine Synthase Activity

Protein was dialysed into activity assay buffer and stock solutions of OAS and sodium sulfide were made up in activity assay buffer. OAS was added to a 96 well PCR plate (Star Labs) at varying concentrations between 0 and 10 mM. Sodium sulfide was added at 3 mM and protein was added at 150 nM. Activity assay buffer was added to a final volume of 50 µL. The plate was sealed and incubated at 37°C for 30 minutes. After, 50 µL of acetic acid and 50 µL of ninhydrin solution was added. The plate was resealed and heated at 96°C for 10 minutes. After this, 100 µL of reaction was transferred to a 96 well flat bottom plate (Star Labs). Absorbance readings were taken at 560 nm using a BioTek™ Synergy™ H4 Hybrid Microplate Reader.

C-terminal SAT Peptides Effect on Activity

The cysteine synthase activity assay was conducted using the 50 µL scale with 1 mM OAS and peptides at 150 µM.

200 µL Scale Serine Acetyltransferase Activity Assay

Protein was dialysed into activity assay buffer and stock solutions of DTNB, $MgCl_2$, L-Serine and Acetyl CoA were made up in buffer. Final concentrations are as follows, 0.5 mM DTNB, 5 mM $MgCl_2$, 0.2 mM Acetyl CoA and 150 nM protein. L-serine was used at varying concentrations between 0 and 5 mM.

De novo cysteine biosynthesis pathway assay

Protein was dialysed into activity assay buffer and stock solutions were made up in activity assay buffer. Serine was added to a 96 well PCR plate (Star Labs) at varying concentrations between 0 and 10 mM. For all other reactants, final concentrations are as follows: 5 mM $MgCl_2$, 0.2 mM Acetyl CoA, 3 mM sodium sulfide and 150 nM protein. Activity assay buffer was added to a final volume of 50 µL. The plate was sealed and incubated at 37°C for 30 minutes. After, 50 µL of acetic acid and 50 µL of ninhydrin solution was added. The plate was resealed and heated at 96°C for 10 minutes. After this, 100 µL of reaction was transferred to a 96 well flat bottom plate (Star Labs). Absorbance readings were taken at 560 nm using a BioTek™ Synergy™ H4 Hybrid Microplate Reader.

Cysteine inhibition assay

The methodology established for the Serine Acetyltransferase Activity Assay was used. In addition, varying amounts of cysteine from 0 to 1 mM were added. Readings were taken after 30 and after 60 minutes.

Chapter 3: Investigations into recombinant expression of serine acetyltransferase from *Trypanosoma cruzi*

Introduction to recombinant protein expression

Production of recombinant proteins for structural biology is well established and often conducted in *E. coli* (Rosano et al., 2014). The ease of genetic manipulation, extensively studied system and inexpensive and rapid culture make *E. coli* the first choice expression host for recombinant proteins (Jia et al., 2016).

Choice of Expression Vector

An important consideration in recombinant protein expression is vector choice as this influences the promotor used, selection marker chosen and affinity tag included. There are many different expression vectors available in *E. coli* (Hunt, 2005). pET plasmids possess the T7 promotor system and is an extremely popular vector series for recombinant protein expression (Rosano et al., 2014). An alternative expression system to T7 is the pBAD expression system (Wegerer et al., 2008). Selection markers are typically antibiotic resistance genes and are used to prevent growth of plasmid-free cells (Rosano et al., 2014). Affinity tag selection can influence protein solubility, ligand binding and protein expression in addition to providing a purification method (Kimple et al., 2013).

Affinity chromatography provides a method of specific protein purification from a mixture of assorted proteins. This occurs based on specific but reversible interactions between two molecules (Urh et al., 2009). Typically, affinity tags such as poly-histidines, glutathione S-transferase (GST) and maltose binding protein (MBP) are chosen. Hexahistidine binds to immobilised nickel ions while GST binds to

glutathione fixed to chromatography resin and MBP binds to amylose (Lichty et al., 2005).

The pJOE5751 plasmid is derived from pBAD and therefore controlled by rhamnose, carries ampicillin resistance and adds an N-terminal six histidine tag to the protein of interest with a small number of additional residues (Cornish et al., 2022; Wegerer et al., 2008). Alternative vectors of interest are under control of the T7 expression system and include pET50b encoding an N-terminal NusA tag and two N-terminal 6xHis tags. NusA is a 55 kDa N-utilization substance and is used to confer stability and high solubility to fusion proteins. Additionally, pET50b also has a HRV 3C cleavage site for removal of the NusA tag (Costa et al., 2014).

The pGEX-4T-3 vector encodes an N-terminal GST tag which, like NusA, can promote protein solubility in *E. coli* and can also be used for purification (Harper et al., 2011). Finally, the pMAL-p5E vector encodes an N-terminal MBP vector which is also known to increase protein solubility whilst also being able to be used as a purification technique (Costa et al., 2014; P. Sun et al., 2011).

Choice of expression strain

There are many different strains of *E. coli* that can be chosen for the expression of recombinant protein. *E. coli* BL21 (DE3) is the most commonly used strain and usually provides an initial starting point for expression due to the chromosomally encoded bacterial T7 RNA polymerase, low cost and ability to reach high cell densities (Du et al., 2021; Fathi-Roudsari et al., 2015; Studier et al., 1986). Other strains such as SHuffle T7 Express can be used to aid in the production of active and correctly folded protein (Fathi-Roudsari et al., 2015; Lobstein et al., 2016). An

advantage of SHuffle T7 Express cells include the promotion of disulphide bond formation in the cytoplasm through constitutive expression of DsbC, a disulphide bond isomerase, which encourages correct folding of proteins with incorrect oxidation states. DsbC also functions as a chaperone protein to assist in protein folding (Lobstein et al., 2016).

In order to produce a high yield of soluble protein, temperature and time of induction are important factors to be optimised. When using overexpression conditions, proteins frequently form insoluble aggregates or inclusion bodies (Jhamb et al., 2012). This is due to several factors such as use of high temperatures, high inducer concentrations and expression under strong promoter sequences. This leads to a high translational rate of expression which can overwhelm the bacterial protein quality control system leading to partly or misfolded proteins undergoing aggregation to form inclusion bodies (Singh et al., 2015). Lowering the temperature of a culture will result in less protein formation due to reduced cell metabolism but this also leads to increased solubility of the protein of interest and therefore less inclusion body formation (Jhamb et al., 2012).

Overexpression and purification of serine acetyltransferase (SAT) from *T. cruzi* will allow for structural and biochemical studies to be conducted. During previous studies on protein biosynthesis with *T. cruzi* conducted by the Silber group, SAT was identified as a potential drug target for Chagas disease.

Results

Multiple sequence alignment analysis of *Trypanosoma cruzi* Serine Acetyltransferase

As a first step in studying this protein, analysis of multiple sequence alignments of *Tc*SAT was performed, Figure 3-1. As expected, *Tc*SAT shares high sequence identity with other homologous trypanosomatids - 59% and 56% with *T. grayi* and *T. theileri* respectively. The closest related sequence with an available structure is *Pv*SAT sharing 44% sequence identity however this structure is only available on the PDB (Kuzin et al., 2008). The closest structure that has been published is *Entamoeba histolytica* SAT. *Tc*SAT shares 42% sequence identity with *Eh*SAT isoform 1 (Kumar et al., 2011).

In the *Eh*SAT structure, the acetyl-CoA binding site is highly disordered and open so interacting residues can not be seen (Kumar et al., 2011). Although the *Leishmania major* SAT structure has not yet been determined, four residues, Cys^{376}, Pro^{383}, Ser^{393} and Gln^{391} have been identified as likely to be involved in acetyl CoA binding in *Lm*SAT (Williams et al., 2009). These residues correspond to residues 254, 261, 268 and 270 using the *Eh*SAT numbering in Figure 3-1. As a closely related trypanosomatids, this finding in *Lm*SAT is relevant to *Tc*SAT. The sequence alignment shows that only Cys^{376} is conserved through all sequences analysed. Pro^{383} is conserved in all sequences except *Eh*SAT. Ser^{393} is conserved between only *Lm*SAT and *Tc*SAT, and Gln^{391} is not conserved in any other sequences with mutations to Pro or Ser instead. The limited conservation of these residues suggest that *Tc*SAT is

unlikely to bind to acetyl CoA in the same manner predicted for *Lm*SAT as typically catalytic residues are all highly conserved.

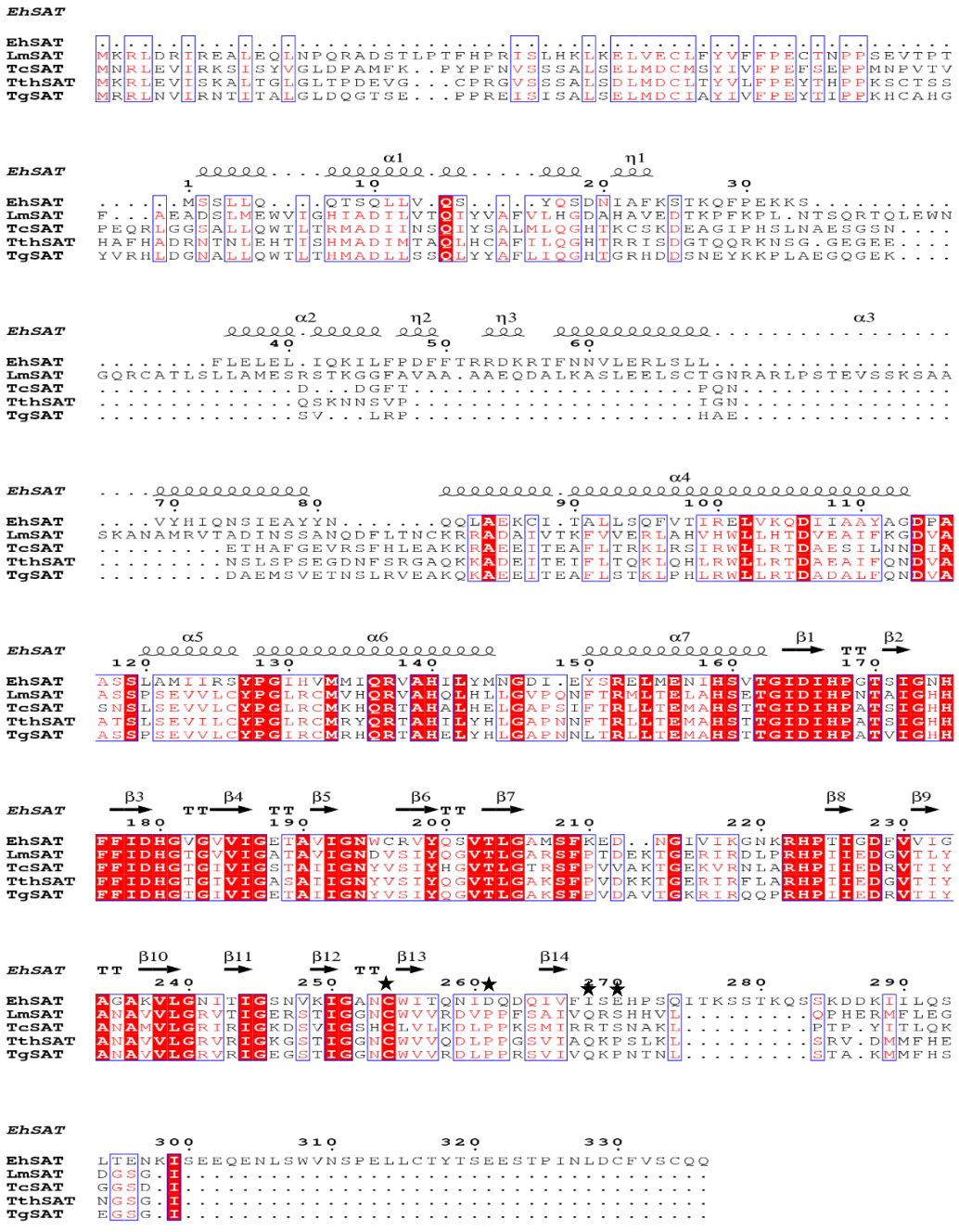


Figure 3-1 - Sequence alignment of *Tc*SAT, *Eh*SAT, *Lm*SAT, *Tth*SAT and *Tg*SAT using the *Eh*SAT structure (Kumar et al., 2011). *Tc*SAT – *Trypanosoma cruzi* SAT; *Eh*SAT – *Entamoeba histolytic*a SAT; *Lm*SAT – *Leishmania major* SAT; *Tth*SAT – *Trypanosoma theileri* SAT; *Tg*SAT – *Trypanosoma greyi* SAT. Stars denote residues implicated in acetyl CoA binding.

Expression tests of pJOE5751_*Tc*SAT

Expression of *Tc*SAT was initially performed using the pJOE5751 vector. The gene of interest was inserted into the vector using the BamHI and BsrGI restriction sites to encode a non-cleavable polyhistidine N-terminal tag. Based on the *Tc*SAT sequence shown in Figure S9-3-2, pJOE5751_*Tc*SAT has been calculated to be 39.9 kDa.

To test the expression of pJOE5751_*Tc*SAT, the plasmid was transformed into BL21 (DE3). Transformed cells were grown to mid-log phase before induction with 0.2% L-Rhamnose. Samples were taken at several time points and analysed by SDS-PAGE, Figure 3-2.

As shown by the arrow in Figure 3-2, there is a consistent band of expressed protein at approximately the calculated weight for *Tc*SAT. Due to the presence of this band, large scale expression and purification was conducted.

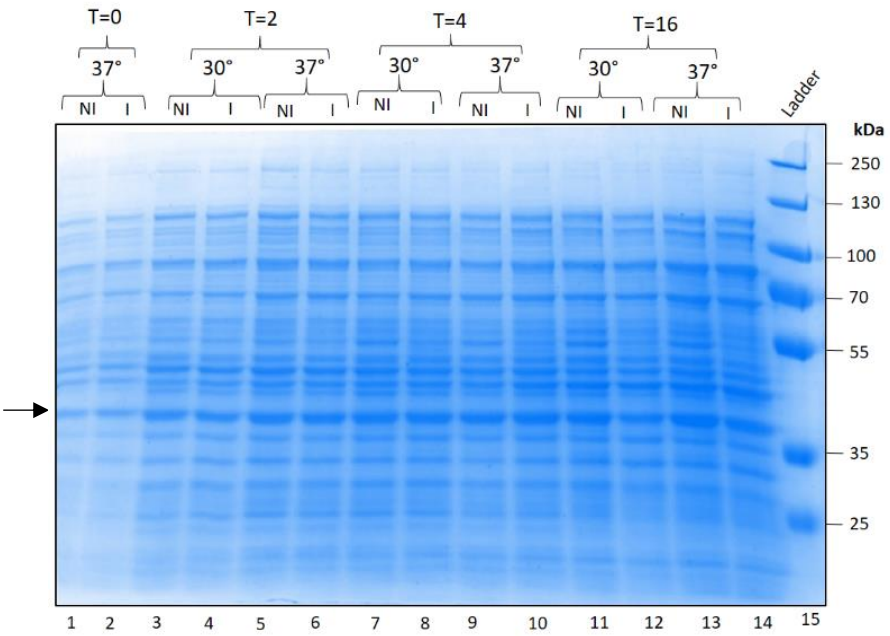


Figure 3-2 - SDS-PAGE of expression tests of pJOE5751_*Tc*SAT at 30°C and 37°C and at different time points. Non-induced (NI) vs induced with 0.2% L-Rhamnose (I) cells at time (T) = 0, 2, 4 and 16 hours after induction. Arrow indicates protein of interest.

Purification of pJOE5751_*Tc*SAT

*Tc*SAT was expressed in the pJOE5751 vector at 37°C for 16 hours and then purified by IMAC. Samples were taken and analysed by SDS-PAGE, Figure 3-3. As seen in lanes 6 to 13, the protein is not present in the eluted fractions, however a band of approximately the correct size is present in the pellet sample and so is likely that the protein is insoluble under these conditions. Due to the insolubility of the expressed protein, it was decided to alter the expression conditions.

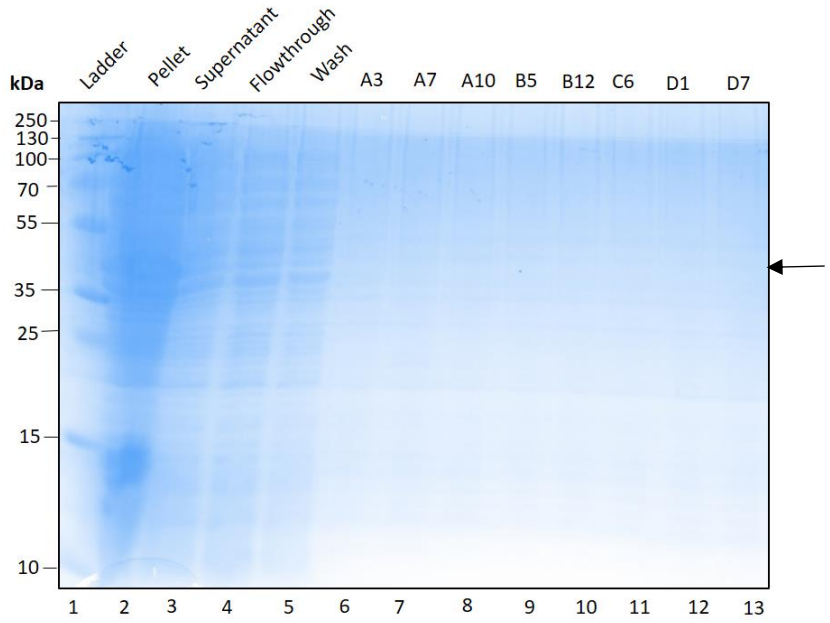


Figure 3-3 - SDS-PAGE showing purification of protein expressed at 37°C for 16 hours. Lanes 6 – 13 contain elution from the IMAC column. Arrow indicates expected migration of protein.

Purification of *Tc*SAT expressed at 25°C for 4 hours

Due to the insolubility of *Tc*SAT under previous conditions, the temperature was lowered to 25°C and the length of expression time was reduced 4 hours to reduce the chances of inclusion bodies forming. Samples were analysed by SDS-PAGE, Figure 3-4. Lanes 11 and 12 do show protein bands of approximately the correct

size for *Tc*SAT yet lane 4 shows a substantial band at the correct size, suggesting that the majority of the protein is insoluble, and expression could be further improved.

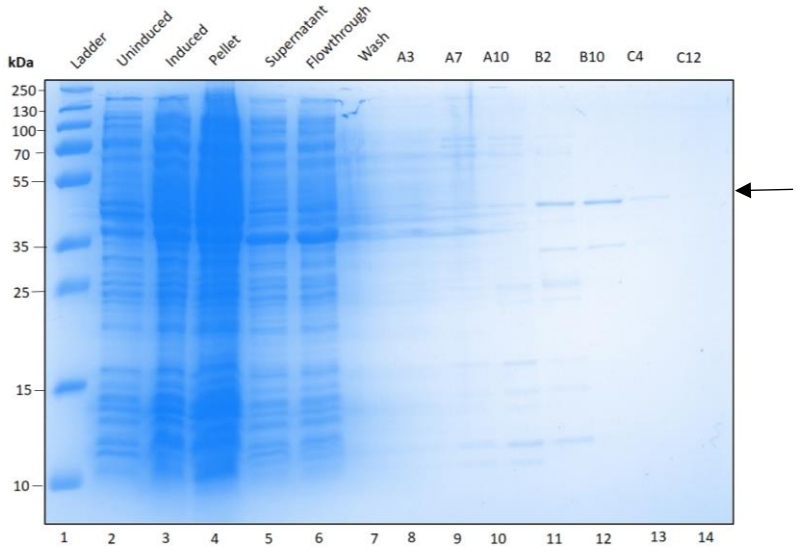


Figure 3-4 - SDS-PAGE showing purification of *Tc*SAT expressed at 25°C for 4 hours. Lanes 8 – 14 contain elution from the IMAC column. Arrow indicates expected migration of protein.

Expression tests of *Tc*SAT in SHuffle T7 Express

pJOE5751_*Tc*SAT was transformed into SHuffle T7 Express cells as these cells contain additional advantages, such as the consistent expression of chaperone proteins, to conventional BL21 (DE3) cells for the solubilising of proteins (Lobstein et al., 2016).

Expression tests of *Tc*SAT in SHuffle T7 Express cells were conducted to determine the optimum conditions for protein expression. Expression tests were conducted at 25°C and at 30°C, samples were taken and analysed by SDS-PAGE, Figure 3-5. As there appears to be a faint but increasing band, at the correct size for *Tc*SAT in lane 14, solubility tests were conducted at 30°C, expression for 16 hours to ensure that the protein was soluble in these conditions.

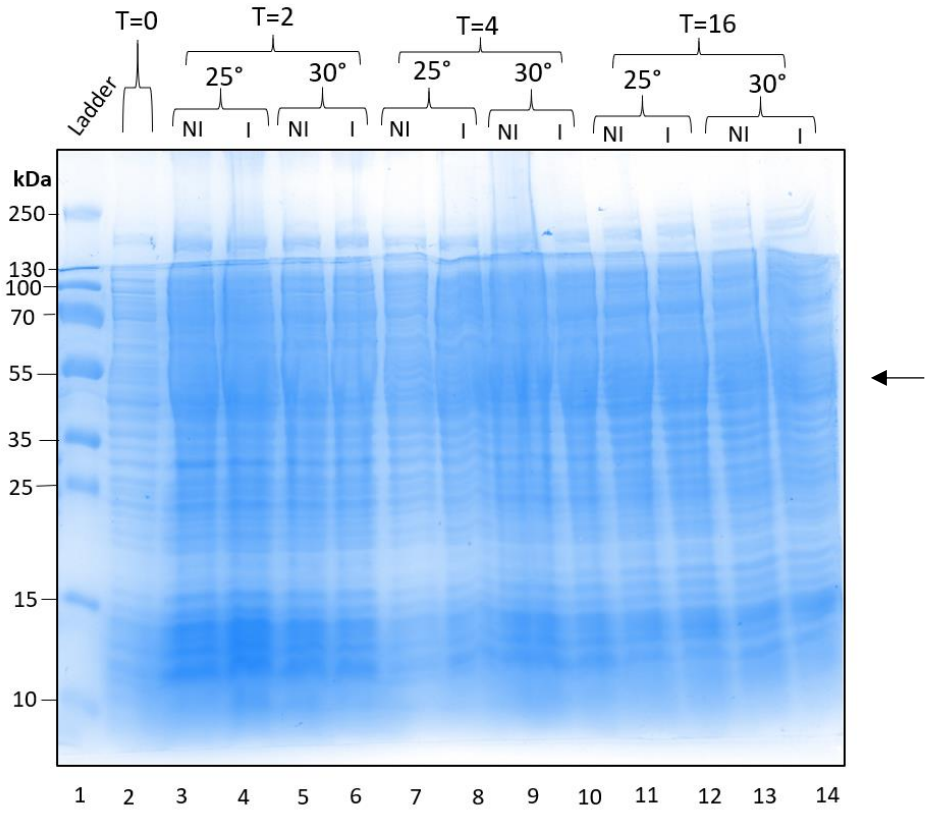


Figure 3-5 -SDS-PAGE showing expression tests of *Tc*SAT in SHuffle at 25°C and at 30°C. Non-induced (NI) vs induced with 0.2% L-Rhamnose (I) cells at time (T) = 0, 2, 4 and 16 hours after induction. Lane 1 shows protein ladder. The arrow indicates expected size of protein of interest.

_Tc_SAT Small Scale Solubility Test

Solubility tests of pJOE5751_*Tc*SAT in SHuffle T7 express cells were conducted at 30°C. Samples were taken and analysed by SDS-PAGE, shown in Figure 3-6.

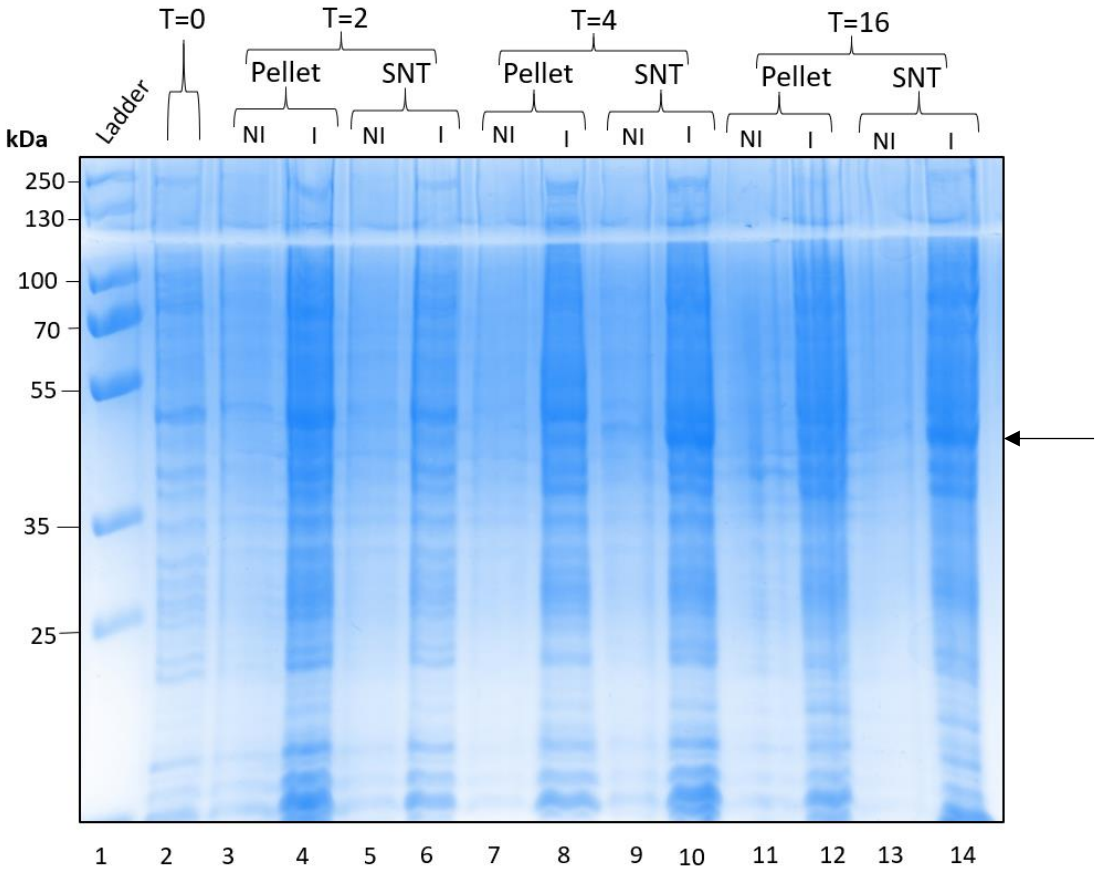


Figure 3-6 - SDS-PAGE showing TcSAT solubilty test, conducted at 30°C. Lane 1 contains protein ladder. Non-induced (NI) vs induced with 0.2% L-Rhamnose (I) cells at time (T) = 0, 2, 4 and 16 hours after induction (T) from pellet and supernatant (SNT). The arrow indicates the expected protein of interest.

There appears to be a band of approximately the correct size, present in the induced sample of the supernatant at 30°C, 16 hours after induction that is absent from the pellet sample at the same conditions. This implies that the protein will be soluble under these conditions. Therefore, *Tc*SAT was expressed on a large scale (12 L) using these conditions.

Purification of *Tc*SAT

*Tc*SAT, expressed under the conditions previously described, was purified and samples were taken and analysed by SDS-PAGE, Figure 3-7. Despite solubility tests indicating that the protein is soluble under these conditions, protein purification by IMAC does not yield soluble protein in sufficient amounts for structural studies. From Figure 3-7, it appears that the protein remains in the pellet and further changes to the expression conditions are required to produce soluble protein. These results show that expression and purification of *Tc*SAT in the pJOE5751 vector results in insoluble protein which cannot be purified by IMAC.

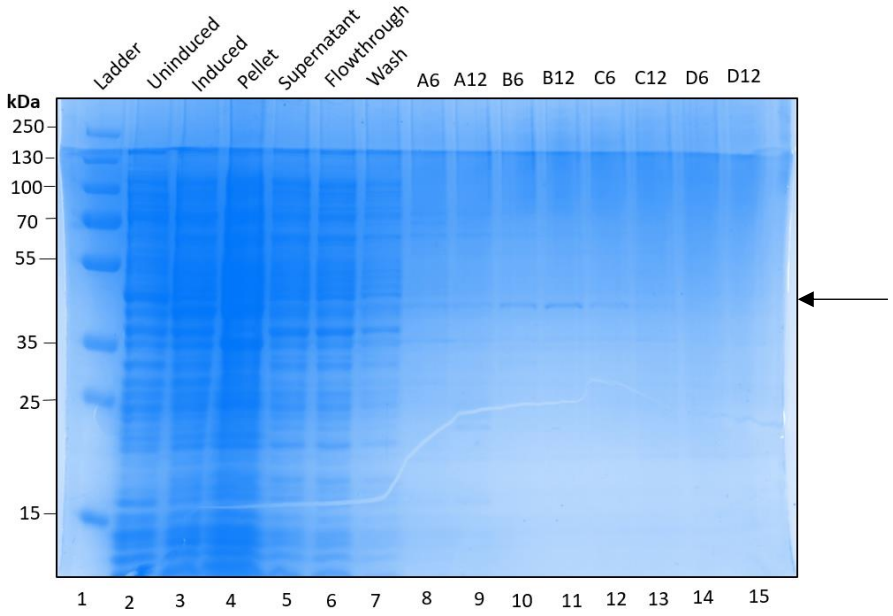


Figure 3-7 - SDS-PAGE showing purification of *Tc*SAT by IMAC with expression conditions of 30°C, 16 hours after induction. Arrow indicates expected migration of protein.

Alternative *Trypanosoma cruzi* Serine Acetyltransferase Vectors

With the aim of increasing the solubility of *Tc*SAT, three alternative vectors were tried. These vectors were pET50b, pGEX-4T-3 and pMAL-p5E, vector maps shown in Figures S9-3-3, S9-3-5, S9-3-7.

pGEX-4T-3_TcSAT

An N-terminal GST tag and thrombin cleavage site was added to *Tc*SAT through the pGEX-4T-3 vector using the EcoRI and NotI restriction sites. pGEX-4T-3_*Tc*SAT was transformed into SHuffle T7 Express cells before expression and solubility tests were carried out, Figure 3-8. Figure 3-8A shows pGEX-4T-3_*Tc*SAT is being expressed at all temperatures and times but the solubility gels indicate that the protein is split between the soluble and insoluble fractions at all times for both 30°C and 37°C.

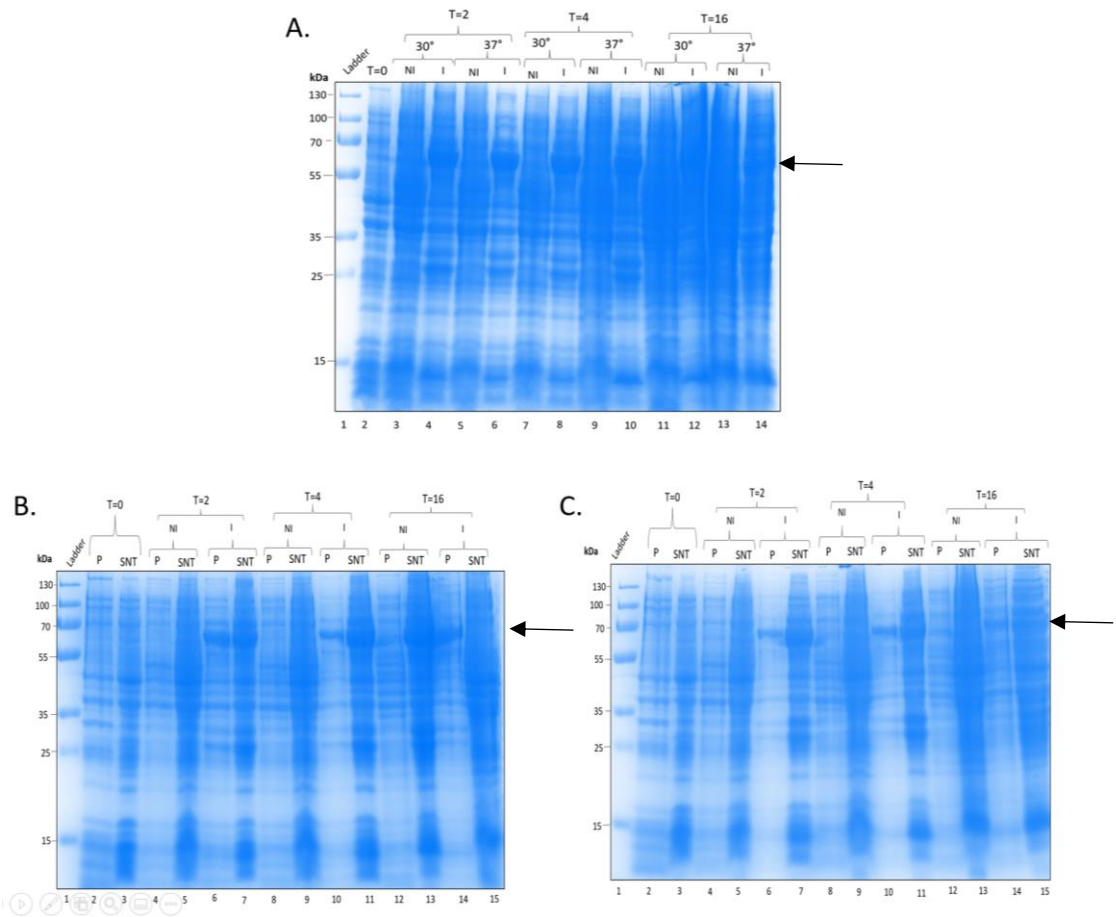


Figure 3-8 - GST::*Tc*SAT expression and solubility tests. Arrow indicates expected migration of protein. **A**) GST::*Tc*SAT expression tests at 30°C and 37°C. **B**) Solubility tests of GST::*Tc*SAT at 30°C. **C**) Solubility tests of GST::*Tc*SAT at 37°C. NI is non-induced cells and I is cells induced with 1 mM IPTG.

Due to the presence of bands indicating protein in both the soluble and insoluble fractions, it is unlikely there will be a significant increase in protein solubility using this vector however purification of the protein present in the soluble fraction may be possible and so purification will occur.

Purification of N-terminal GST tagged TcSAT

GST::*Tc*SAT was expressed at 30°C for 2 hours and then purified by GSTrap. The results of this purification were analysed by SDS-PAGE, Figure 3-9. Despite the results of the solubility tests, there is no *Tc*SAT protein present in the supernatant, whilst a large band of the correct size is present in the pellet, indicating that the

protein has remained insoluble. Due to the protein remaining insoluble, efforts to purify the protein in this vector were not continued.

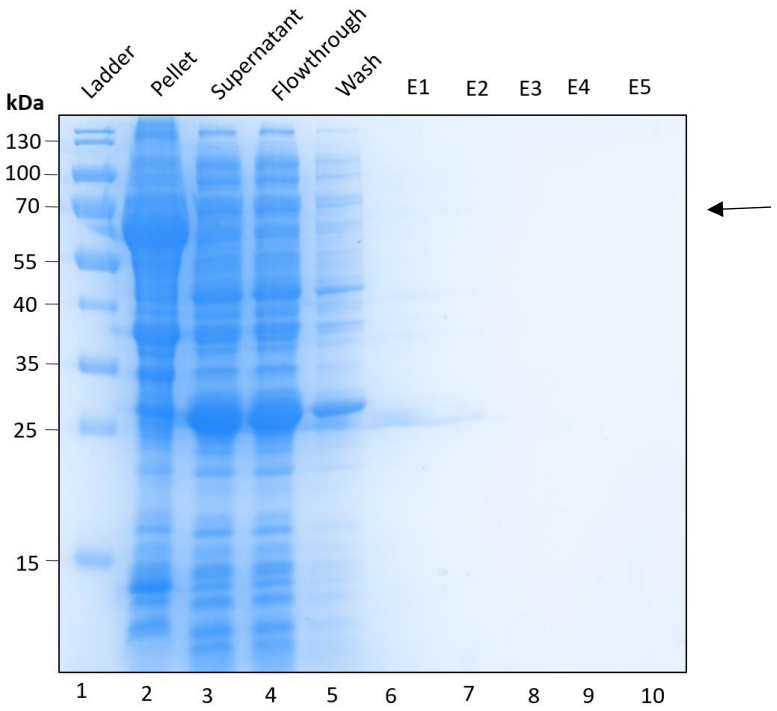


Figure 3-9 - GSTrap purification of GST::*Tc*SAT. The arrow indicates the size of the protein of interest. Lanes 6 – 10 contain the elution of the GSTrap.

2.3.1.5 Purification of N-terminal MBP tagged TcSAT

A MBP tag and enterokinase cleavage site were added to N-terminal of *Tc*SAT through the pMAL-p5E vector through the use of NdeI and EcoRI restriction sites. pMAL-p5E_*Tc*SAT was transformed into SHuffle T7 express cells; expression and solubility tests were conducted at 30°C and at 37°C, Figure 3-10. As can be seen in Figure 3-10A, expression is best at 30°C and comparable at 2 and at 4 hours. This is confirmed by the solubility tests in Figure 3-10B and Figure 3-10C, which indicate that although there is still some protein in the pellet at both 2 and 4 hours, the protein is mostly in the soluble fraction. Based on these results, expression of *Tc*SAT at 30°C for 2 hours would achieve optimal expression.

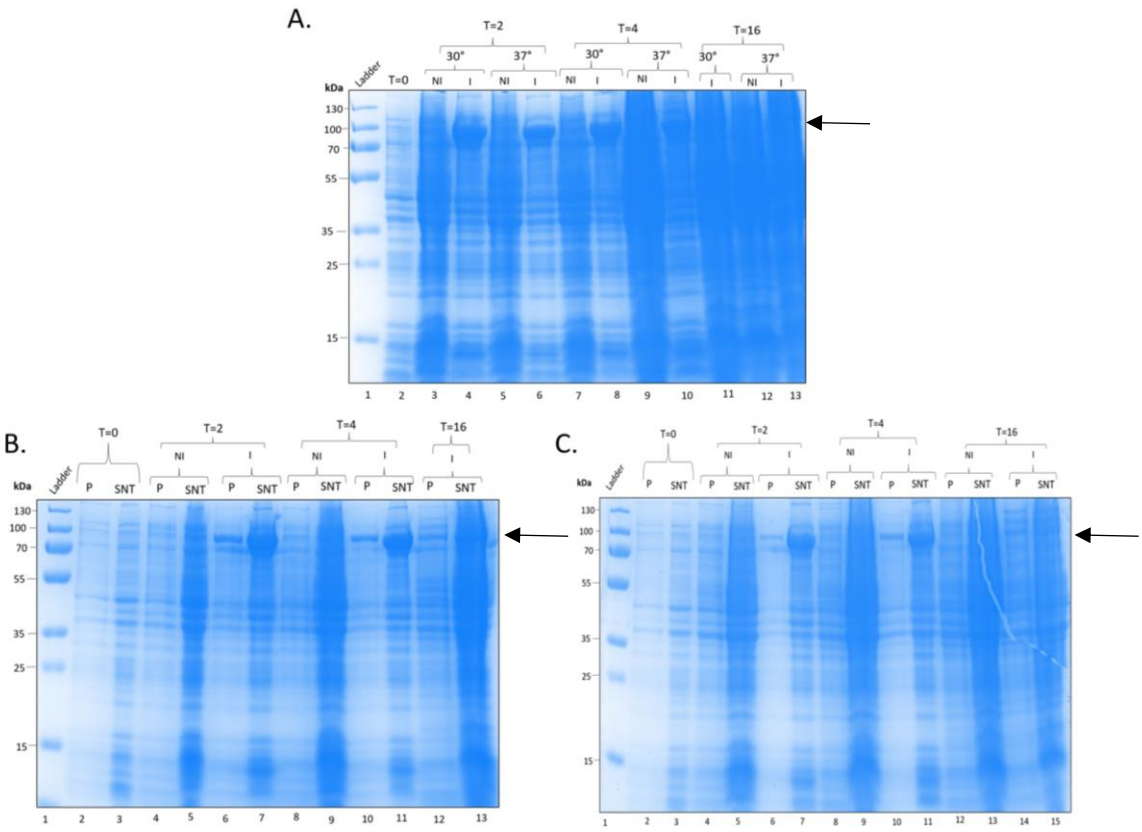


Figure 3-10 - MBP::*Tc*SAT expression and solubility tests. Arrow indicates expected migration of protein. **A**) MBP::*Tc*SAT expression tests at 30°C and 37°C. **B**) Solubility tests of MBP::*Tc*SAT at 30°C. **C)** Solubility tests of MBP::*Tc*SAT at 37°C. NI is non-induced cells and I is cells induced with 1 mM IPTG.

Purification of N-terminally NusA tagged TcSAT

For the pET50b construct, the restriction sites SmaI and SacI were used with the addition of a stop codon after the *Tc*SAT sequence. This resulted in an N-terminally tagged *Tc*SAT sequence with a NusA solubilising tag and polyhistidine tag for purification. Use of the chosen restriction sites maintains the HRV3C cleavage site allowing tag removal for downstream applications. pET50b_*Tc*SAT was transformed into SHuffle T7 express cells; expression and solubility tests were conducted shown in Figure 3-11.

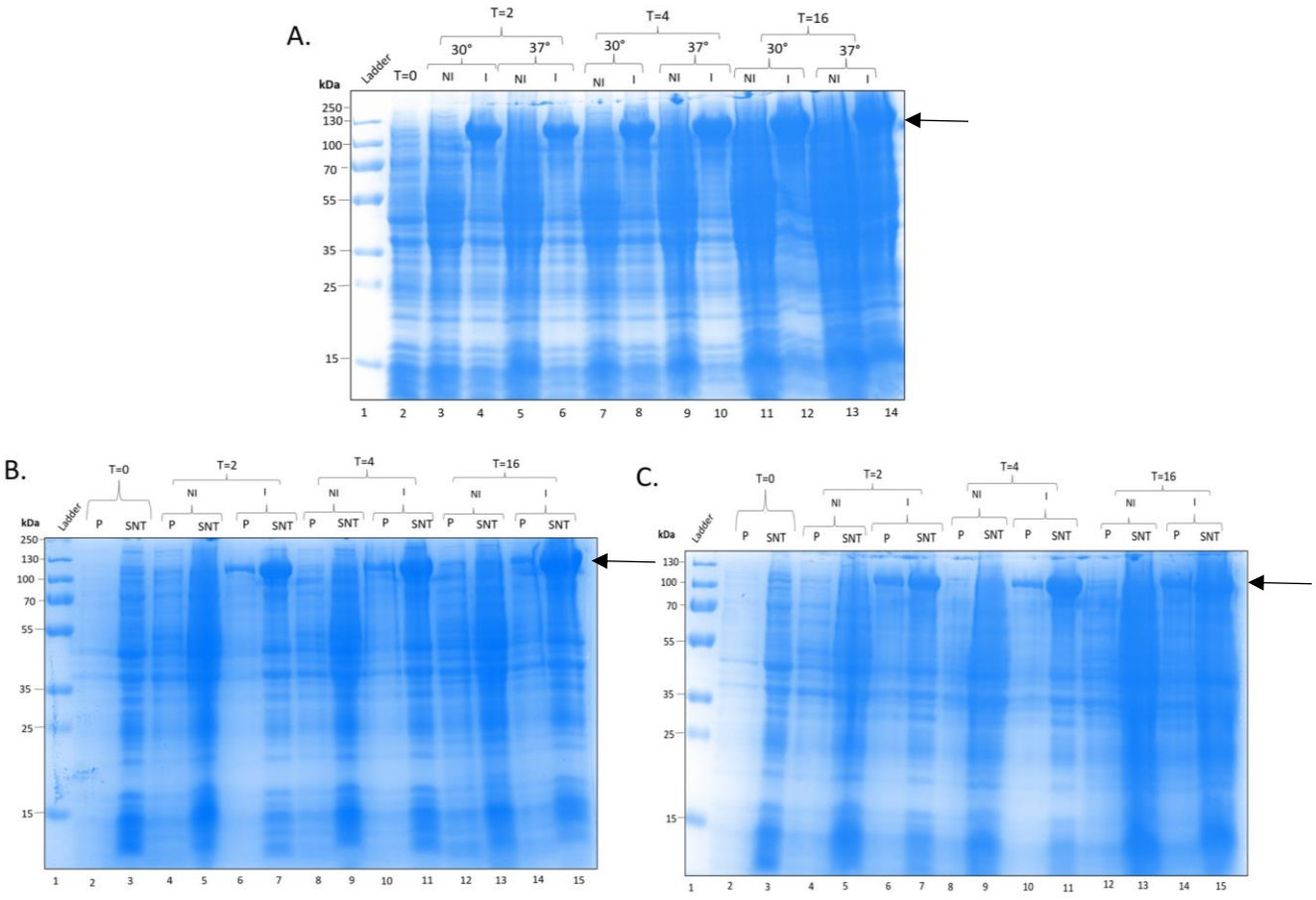


Figure 3-11 - NusA::SAT expression and solubility tests. Arrow indicates expected migration of protein. **A**) NusA::*Tc*SAT expression tests at 30°C and 37°C. **B**) Solubility tests of NusA::*Tc*SAT at 30°C. **C**) Solubility tests of NusA::*Tc*SAT at 37°C. NI is non-induced cells and I is cells induced with 1 mM IPTG.

As shown in the gels, for all times and both temperatures NusA::*Tc*SAT is being expressed and is soluble. Optimal expression conditions for the production of soluble protein appear to be expression of NusA::*Tc*SAT at 30°C for 2 hours.

Purification of N-terminally NusA tagged TcSAT

Following results of the expression and solubility tests, 2 L of NusA::*Tc*SAT was expressed in Shuffle T7 Express cells at 30°C for 2 hours. The protein was then purified using IMAC and analysed by SDS-PAGE, Figure 3-12. Some protein does appear to be soluble and able to be purified. Despite this, the majority of the protein remains in the pellet and therefore insoluble requiring further optimisation.

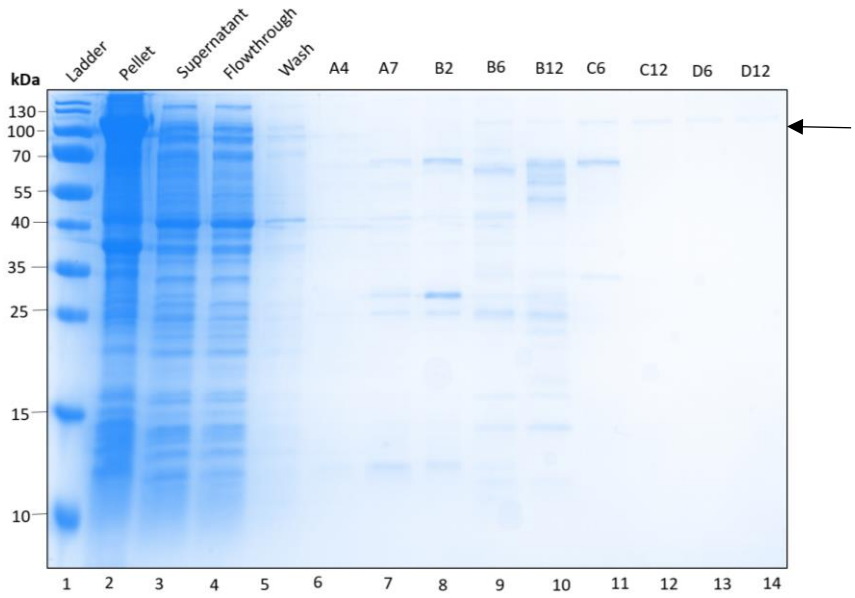


Figure 3-12 - SDS-PAGE showing NusA::*Tc*SAT purification. Arrow indicates expected migration of protein.

The effect of chemotropic agents as additives in the media

Due to the small amount of soluble protein expressed and purified with the NusA::*Tc*SAT vector in the SHuffle cells, additives to the growth media were evaluated to improve protein solubility. Whilst changes to the growth media were made, conditions for each expression remained 2 hours at 18°C. The protein from each expression was purified by HisTrap and then analysed by SDS-PAGE, Figure 3-13. Terrific broth was used as it can support a higher density of bacteria and therefore produce a higher yield of soluble protein, Figure 3-13A (Tartoff et al., 1987). The expression volume was reduced from 1 L to 500 mL, Figure 3-13B, as aeration can improve protein yield. 1% glucose was added to LB, Figure 3-13C, as glucose can repress the lac operon often activated by lactose present in the media before the induction agent is added. 4% glycerol was added to LB, Figure 3-13D, to provide an alternative carbon source. Although there may be slight improvements with the addition of 4% glycerol, none of the chemotropic agents are significantly

affecting the production of soluble protein. This suggests that the protein is inherently prone to aggregation during expression.

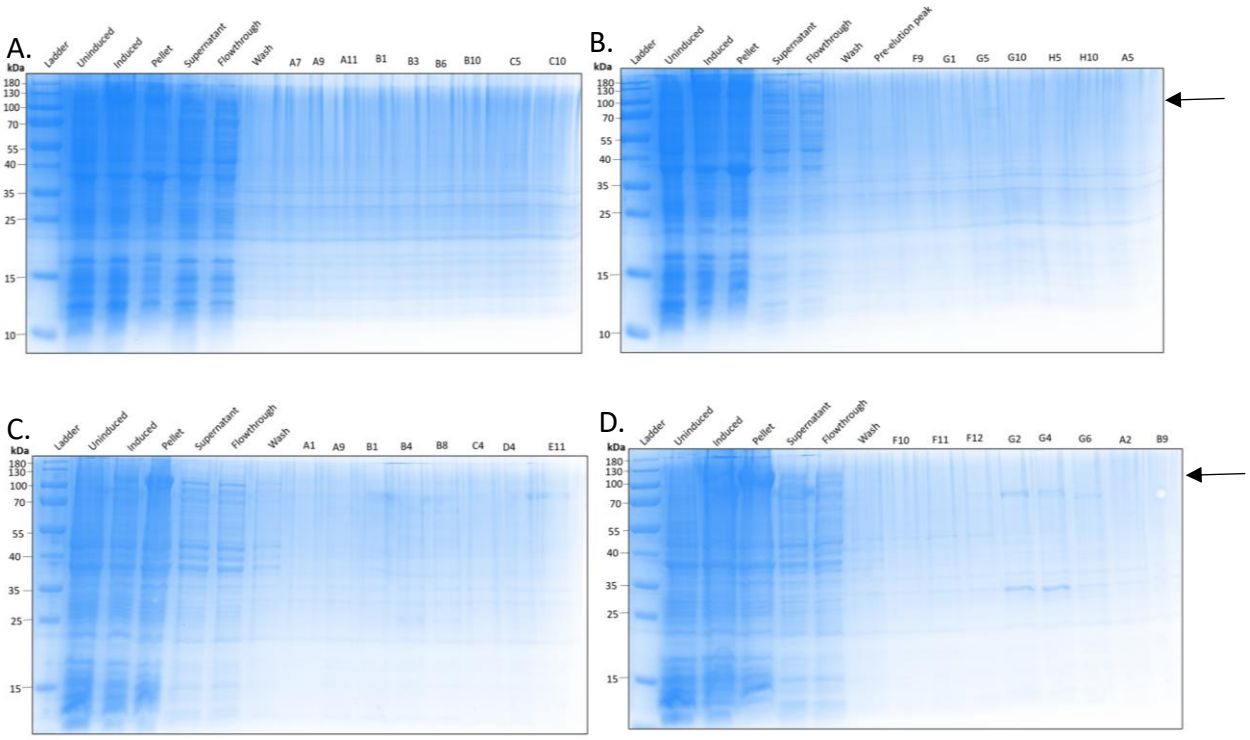


Figure 3-13 - SDS-PAGE from purifications of NusA::*Tc*SAT with changes to expression media **A)** with Terrific Broth. **B)** with a smaller volume of 500 mL **C)** including 1% glucose. **D)** including 4% glycerol. Arrows indicate expected migration of protein.

***Trypanosoma cruzi* Serine Acetyltransferase unfolding and refolding**

Due to the persistent insolubility of *Tc*SAT, purification of the protein from the inclusion bodies was performed. 2L of NusA::*Tc*SAT was expressed for 2 hours at 18°C and purification was performed and analysed by SDS-PAGE, Figure 3-14. It can be seen that the protein is present in the pellet at each purification stage and then is present in the final supernatant (lane 12) after the urea buffer is added.

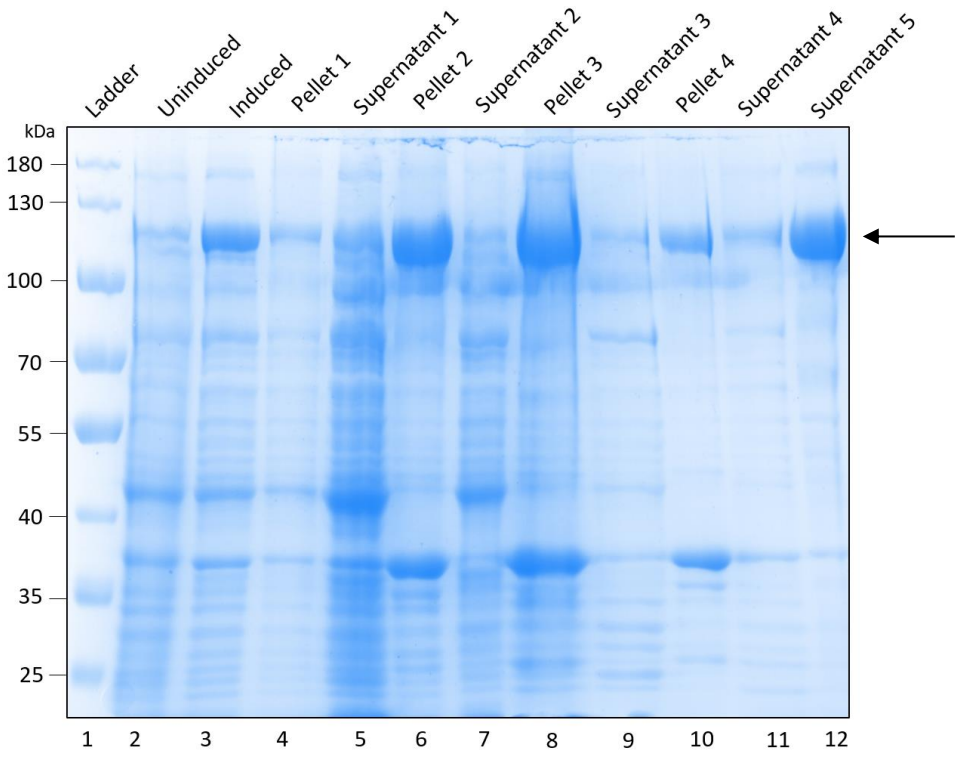


Figure 3-14 - SDS-PAGE showing Inclusion body purification of NusA::*Tc*SAT. Arrow indicates protein of interest.

The protein was successfully unfolded and high yields (~100 mg/L) were obtained. The denaturant was removed overnight by dialysis and the N-terminal His and NusA tags encoded by the pET50b vector were cleaved with HRV 3C. The cleaved protein was separated from tagged protein, the N-terminal tag and HRV 3C by gel filtration. This was analysed by SDS-PAGE, Figure 3-15. Cleaved protein is shown in the black box, lanes 4 and 5 show the tagged protein that has not been cleaved and lanes 14 and 15 show the protein tag. *Tc*SAT was successfully separated from NusA::*Tc*SAT and the NusA fusion tag. There are no impurities present in the resulting sample and so further experiments can be performed using this protein without the need for additional purification steps.

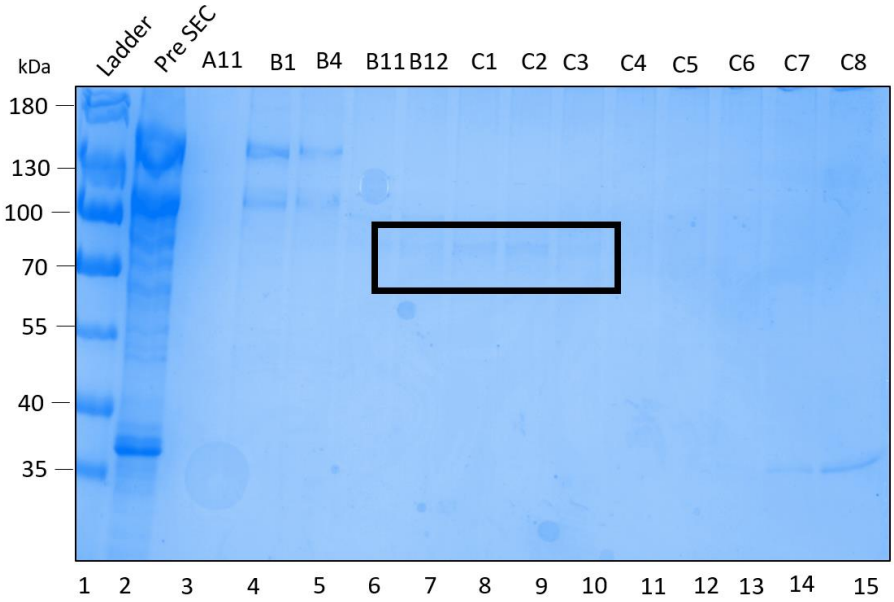


Figure 3-15 - SDS-PAGE showing *Tc*SAT separation after tag cleavage by size exclusion.

Circular Dichroism

In order to confirm that *Tc*SAT refolded successfully, circular dichroism was performed, Figure 3-16. The spectrum suggests that the protein is folded and

contains a combination of α-helices and β-sheets (Rodger et al., 2021).

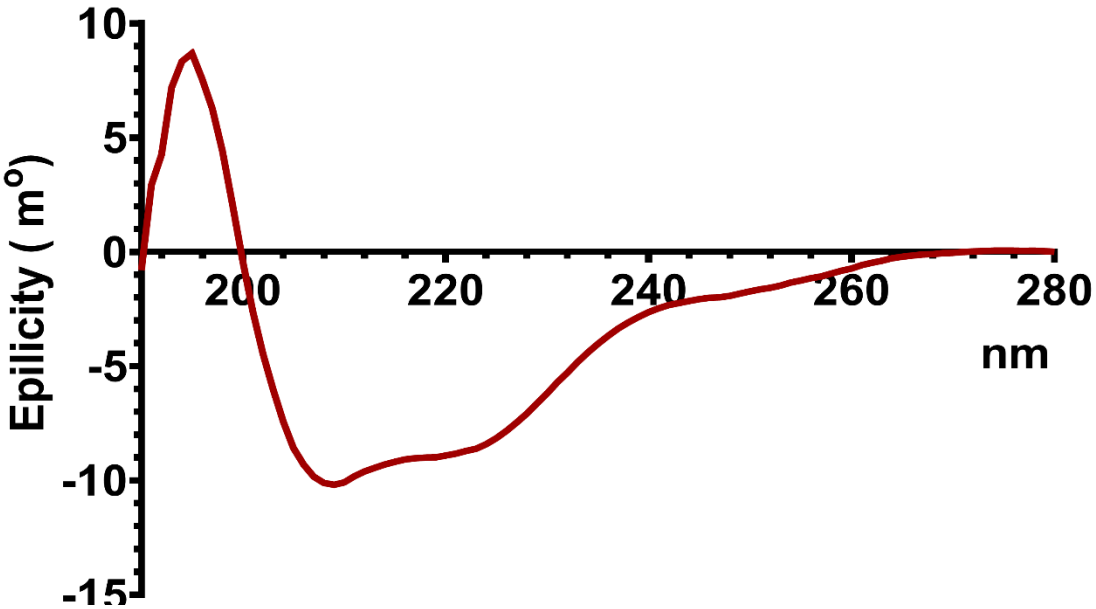


Figure 3-16 - Circular dichroism spectrum for *Tc*SAT

Analysis of the data using DICHROWEB, estimates that the protein is 16% α-helices, 34% β-sheets and 50% random coil (Andrade et al., 1993; Lobley et al., 2002; Miles et al., 2022; Whitmore et al., 2004, 2008). This high percentage of random coil suggests that the protein may still be disordered. Refolded *Tc*SAT was tested for activity on a 50 µL, 100 µL and 500 µL scale with DTNB however no activity was detected.

***Trypanosoma cruzi* Serine Acetyltransferase crystallisation experiments**

The cleaved *Tc*SAT protein was concentrated to 5.7 $mgmL^{-1}$ and a sample was loaded onto SDS-PAGE for analysis. The protein was used to set up 5 high throughput crystallisation screens – Morpheus I, Morpheus II, Shotgun I, MIDAS + and BCS. 8

conditions from these screens produced crystals, shown in Table 3-1, with example images shown in Figure 3-17.

Table 3-1 - High throughput screen conditions resulting in *Tc*SAT crystals

Screen	Condition
Shotgun I	0.2 M Ammonium chloride, 20 % w/v PEG 3350
Shotgun I	0.2 M Ammonium sulfate, 0.1 M Sodium HEPES pH 7.5, 25 % w/v PEG 3350
Shotgun I	0.2 M Ammonium formate 20 % w/v PEG 3350
BCS	0.05 M Ammonium sulfate, 0.05 M Lithium sulfate, 0.1 M Bis-Tris propane pH 8.5, 28 % v/v PEG Smear Broad
Morpheus I	0.09 M NPS, 0.1 M Buffer System 1 pH 6.5, 30% v/v Precipitant Mix 1
Morpheus I	0.09 M NPS, 0.1 M Buffer System 1, pH 6.5, 30% v/v Precipitant Mix 3
MIDAS+	0.5 M Ammonium phosphate monobasic, 12.5% w/v Poly(acrylic acid sodium salt) 2100
MIDAS+	0.1 M Potassium/sodium phosphate pH 7, 35% v/v Polypropylene glycol 400

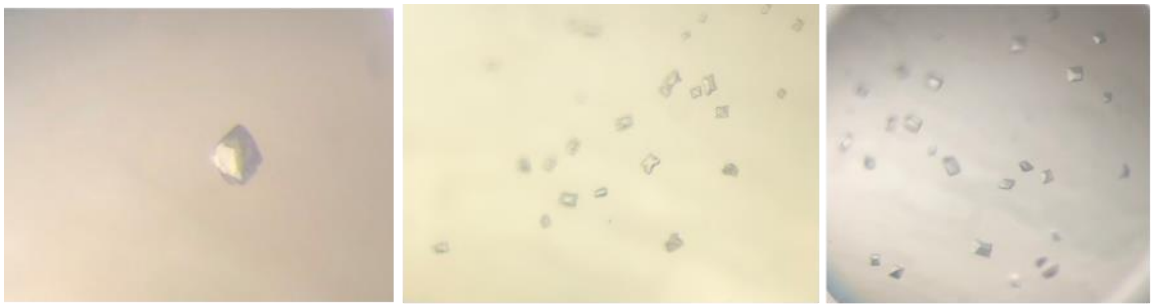


Figure 3-17 - Typical crystals produced from refolded *Tc*SAT

Crystals from each condition were cryoprotected with 50% glycerol before freezing in $N_{2(l)}$. These crystals were sent to DLS and diffraction experiments were performed. Data collected showed that these crystals were not protein but multicrystalline salt and therefore alternative methods are required in order to determine the *Tc*SAT structure.

Conclusions

A wide range of attempts were made to express and purify *Tc*SAT in *E. coli* to generate soluble protein however the protein was consistently present in the insoluble fraction. Based on this work, it can be concluded that the production of soluble recombinant *Tc*SAT using an *E. coli* expression system is unlikely but expression in mammalian or yeast cell expression systems may yield soluble protein. Mammalian cells can express complex and large recombinant proteins, perform post translational modifications, and have good protein folding abilities (Gray, 1997; Tripathi et al., 2019). Yeast cells also have enhanced protein folding abilities compared to *E. coli* and unlike mammalian cells have high product yield (Baghban et al., 2019; Tripathi et al., 2019).

Although the protein was successfully extracted from inclusion bodies and circular dichroism confirms that the protein was refolded, the high percentage of random coil seen in the results suggest that the protein was not completely refolded and the active conformation was not achieved. The denaturant was removed using dialysis however other methods to remove the denaturing element could be tried to prevent misfolding as rapid decreases in denaturant concentration can lead to misfolding and this could be combatted by trying refolding using microfluid chips (Yamaguchi et al., 2014).

While the misfolded protein was used to set up high throughput crystal screens, all conditions produced salt crystals. This indicates that if soluble protein or correctly folded protein is obtained, the conditions identified are likely to still produce salt crystals and so alternative conditions should be tried.

Chapter 4: Exploring the enzymatic mechanism of cysteine synthase through structural and biochemical studies

Introduction into X-ray crystallography

X-ray crystallography is a well-established method of protein structure determination. Macromolecular crystallography includes the crystallisation of nucleic acids, proteins, and large complexes such as viruses and ribosomes (McPherson et al., 2013). X-ray crystallography aims to produce a 3D molecular structure from a single crystal upon exposure to an X-ray beam (Smyth et al., 2000).

To produce crystals suitable for use in crystallographic experiments, reliably producible soluble protein at a high yield is required (Smyth et al., 2000). The protein must be purified to homogeneity to produce crystals capable of diffracting to a high resolution (Dessau et al., 2011).

Crystals are formed from protein upon a particular level of supersaturation being achieved resulting in nucleation (Timofeev et al., 2023). The formation of protein crystals begins with nucleation. The process of nucleation is the least understood step of the crystallographic process (Smyth et al., 2000). Nucleation represents a first-order phase transition and defines many features of the crystals grown including number, size, distribution and polymorphism of the crystal (McPherson & Gavira, 2013, Vekilov & Vorontsova, 2014). Following nucleation, crystals grow in a rate limiting step (Smyth et al., 2000).

Most macromolecular crystals are grown by solvent vapour diffusion due to the advantages of simplicity and relative affordability that this method offers. This technique relies on vapour from a drop containing a protein-precipitant mixture

diffusing into the reservoir until the partial vapour pressure over the surface of the solution and the drop is equivalent. This increases the protein concentration and leads to supersaturation of the solution and to the formation of crystals or amorphous precipitate, Figure 4-1 (Timofeev et al., 2023). This occurs in sitting drop or hanging drop format. Sitting drop results in the protein sitting directly on the surface of the crystallisation plate, above the crystallisation conditions in the well whereas hanging drop suspends the protein on a coverslip above the well solution avoiding contact with the crystallisation plate (King et al., 2009). Advantages of the sitting drop format include no limitation on drop size, faster than hanging drop, economic and easy to automate. However, the sitting drop vapour diffusion method is limited as the handling of crystals grown in this way can be difficult due to adherence to the crystallisation plate. In contrast, the hanging drop method results in easier handling of crystals, but the drop volume is limited, the cost is higher, and automation is difficult. Despite this, vapour diffusion remains the technique of choice for crystal growth (Adachi et al., 2003).

Supersaturating protein concentration and therefore the formation of protein crystals can be achieved using precipitating reagents which decrease the protein solubility resulting in favourable conditions for protein crystallisation. As crystallisation is incredibly varied, typically a multitude of solutions are tested with a variety of chemically diverse compounds (Skarina et al., 2014). Typically, commercially available and varied screens are used. The precipitants used are wide ranging and can be grouped into four overarching categories: salts, organic solvents, long chain polymers and low-molecular weight polymers. Salt concentration is also a key component of crystallisation conditions as, at sufficiently high concentration

macromolecules will associate with each other to satisfy electrostatic requirements resulting in crystal formation. Additionally, at sufficiently low salt concentration macromolecular solubility cannot be maintained due to a lack of cation availability which can also result in crystal formation (McPherson & Gavira, 2013).

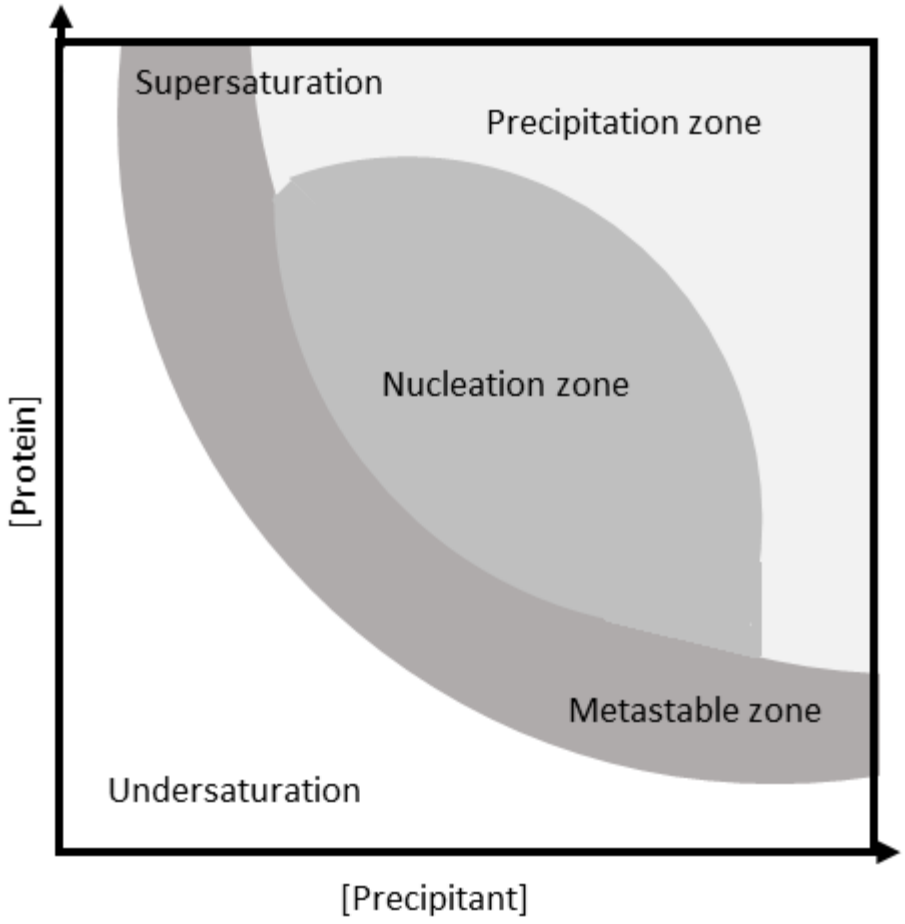


Figure 4-1 - Idealised phase diagram

Once grown, protein crystals are cooled to below 120 K to prevent deterioration of the crystal upon exposure to high intensity X-rays (Haas, 1968). Cryoprotectants depress the freezing point of water and aid in the prevention of ice formation and therefore crystal destruction during flash-cooling (Pflugrath, 2015). Cryoprotectants

are used to prevent damage from both X-rays and that can occur from cooling. There are various types of cryoprotectant that are split into two broad categories: penetrating and non-penetrating. Penetrating cryoprotectants include glycerol, glucose and low molecular weight PEG; these molecules penetrate the crystal lattice and cause dehydration. High molecular weight PEG and oils such as paraffin oil are non-penetrating and instead coat the crystal lattice (Jang et al., 2022). Glycerol is a typically used cryoprotectant due to its antifreeze properties; glycerol is inserted into the water hydrogen bond network which helps prevent ice formation (Vera et al., 2011).

Macromolecular crystallography relies on the scattering of X-rays by electrons in the sample of interest (Wlodawer et al., 2013). Today, synchrotrons are used to produce this X-ray radiation as when charged particles are accelerated, electromagnetic radiation is emitted (Dauter et al., 2010). The several orders of magnitude of increased brightness produced by synchrotron radiation enables higher resolution structures to be obtained over a significantly shorter time comparatively with a home source (Helliwell et al., 2015). Typically, a relative flux of 12 keV photons are used (O'Neill et al., 2002). Synchrotron radiation has resulted in higher resolution structure determination, increased use of multiple-wavelength anomalous dispersion to solve the crystallographic phase problem, study of larger molecular weight molecules, smaller crystals being able to be used and in time resolved studies being possible (Helliwell et al., 2015).

The accurate collection of X-ray diffraction data is critical for structure determination. This is achieved using hybrid photon counting detectors such as the

DETECTRIS Pilatus which reduces noise and has a larger dynamic range compared to earlier detectors meaning that both strong and weak intensity reflections can be captured (Broennimann et al., 2006; Krause et al., 2020). Additionally, EIGER detectors have begun to be used; these pixel-array detectors have smaller pixel size, higher frame rate and readout resulting in less dead time during data collection (Casanas et al., 2016).

In order to determine the structure of a macromolecule, both the intensity and phase of each reflection within the diffraction data needs to be known. This is needed to calculate the electron density map that is then used for structure solution. The need for both intensity and phase of the data gives rise to the phase problem which is that the intensities of the data are easily and accurately measured with modern detectors however the phases cannot be measured (Hauptman, 1991).

This problem can be overcome with several methods. If data is collected to a high enough resolution, typically near atomic resolution, direct methods can be used (Frazão et al., 1999; Sheldrick et al., 2012). This involves calculating the phases by using the mathematical relationship to structure factors (Sheldrick, 1990). Phases are selected for a selection of strong reflections and in turn, the other phases are calculated from them. This is usually not applicable for most proteins as structure factors cannot be measured at atomic resolution (Usón & Sheldrick, 1999). Additionally, the phase relationships become weaker as the number of atoms in the structure grows, making this method unsuitable for most proteins (Cowtan, 2003). Isomorphous replacement uses heavy atom substitution to solve the phase problem. This classical solution was used to determine the first structures of

myoglobin and of haemoglobin (Green et al., 1954; Kendrew et al., 1960). Heavy atoms such as Cu and Se result in changes compared to typical atoms, such as to Patterson functions, that can be used to locate the heavy atoms and therefore calculate phases based on the differences in data collected from the native structure and from the heavy atom derivatives. Typically, the heavy atoms are soaked into the crystals (Taylor, 2003). Multiwavelength anomalous dispersion (MAD) became a popular alternative to isomorphous replacement due to tuneable X-ray beamlines. Whilst scattering is usually independent of wavelength, at the absorption edges scattering changes rapidly with wavelength. The absorption edge is the point for each element that there is an increase in the absorption of X-ray photons which occurs when the energy of the photon corresponds to the energy shell of the atom. The anomalous scattering allows phase data to be calculated in a similar way as isomorphous replacement. Anomalous scattering at the absorption edge has a phase shift causing the Friedel pairs to no longer be equivalent and this can provide further phase information such as the location of heavy atom coordinates (Cowtan, 2003).

Molecular replacement is the main ways that modern crystallography solves the phase problem. This method uses the coordinates of a known structure to provide initial estimates of the phases of a new and unknown structure. Typically, models are selected on high sequence identity between the two proteins and completeness of the known structure (Evans et al., 2008). Due to the widespread incorporation of AlphaFold into structural biology, high accuracy models based on the sequence of the unknown protein can be produced for molecular replacement (Barbarin-Bocahu et al., 2022).

The iterative process of agreement between a structural model and the experimental data is termed refinement. The R-factor describes the agreement between the amplitudes of the structure factors which are calculated from the crystallographic model and the structure factors determined from the diffraction data. Optimisation of the model is conducted to minimise the R-factor, however this can lead to overfitting. To combat this, R_{free} and R_{work} are used. A small selection of reflections, typically 5%, are omitted from the refinement process to create an independent value (Brünger, 1992). R_{work} is then the value given to the other reflections that are used in the refinement. A large difference between R_{free} and R_{work} is indicative of overfitting or of an incomplete model (Shabalin et al., 2018). B-factors are used describe the decrease of X-ray scattering due to thermal motion and disorder. B-factors are calculated for each atom and can be used to estimate the flexibility of atoms, side chains or regions of the structure and are refined together with the positional coordinates for each atom. The resolution of the structure is correlated with the B-factors obtained; lower resolution structures, typically 3 – 5 Å, results in higher B-factors, in the range of 100 – 200 $Å^2$. Relatively high local B-factors can indicate disorder in that region of the model, this is typically seen with flexible side chains, low occupancy or partially occupied ligands (Z. Sun et al., 2019) . Ramachandran plots are used to show the distribution of the torsion angles of the protein backbone and produce reports of the number of residues that belong to allowed, favoured or outlier regions based on these angles. This validation method can be used to determine the quality of a model and models with no outliers are considered optimal for structures (Sobolev et al., 2020).

Results

Multiple sequence alignment analysis

Analysis of the multiple sequence alignment of the CS sequences from *Trypanosoma cruzi, Leishmania infantum and Trypanosoma theileri* was conducted, Figure 4-2. There is 74% sequence identity between *Tc*CS and *Li*CS, and 80% with *Tth*CS. This shared identity is to be expected as these organisms are all trypanosomatids and shows that all three proteins are closely related. There is also high sequence identity (74%) between *Tc*CS, *L. major* CS and *L. donovani* CS, the only other two trypanosomatid CS structures determined. Additional sequence analysis reveals that all three CS sequences contain the four canonical lysine residues that have been identified as critical for catalysis – Lys^{41}, Lys^{52}, Lys^{68} and Lys^{200}. Additionally, these sequences also contain the highly conserved consensus sequence for PLP binding (PXXSVKDR) seen in other cysteine synthases (Romero et al., 2015). The conservation of the canonically critical lysine residues and the PLP binding motif suggest that all three proteins will bind to PLP and should be capable of producing cysteine from OAS.

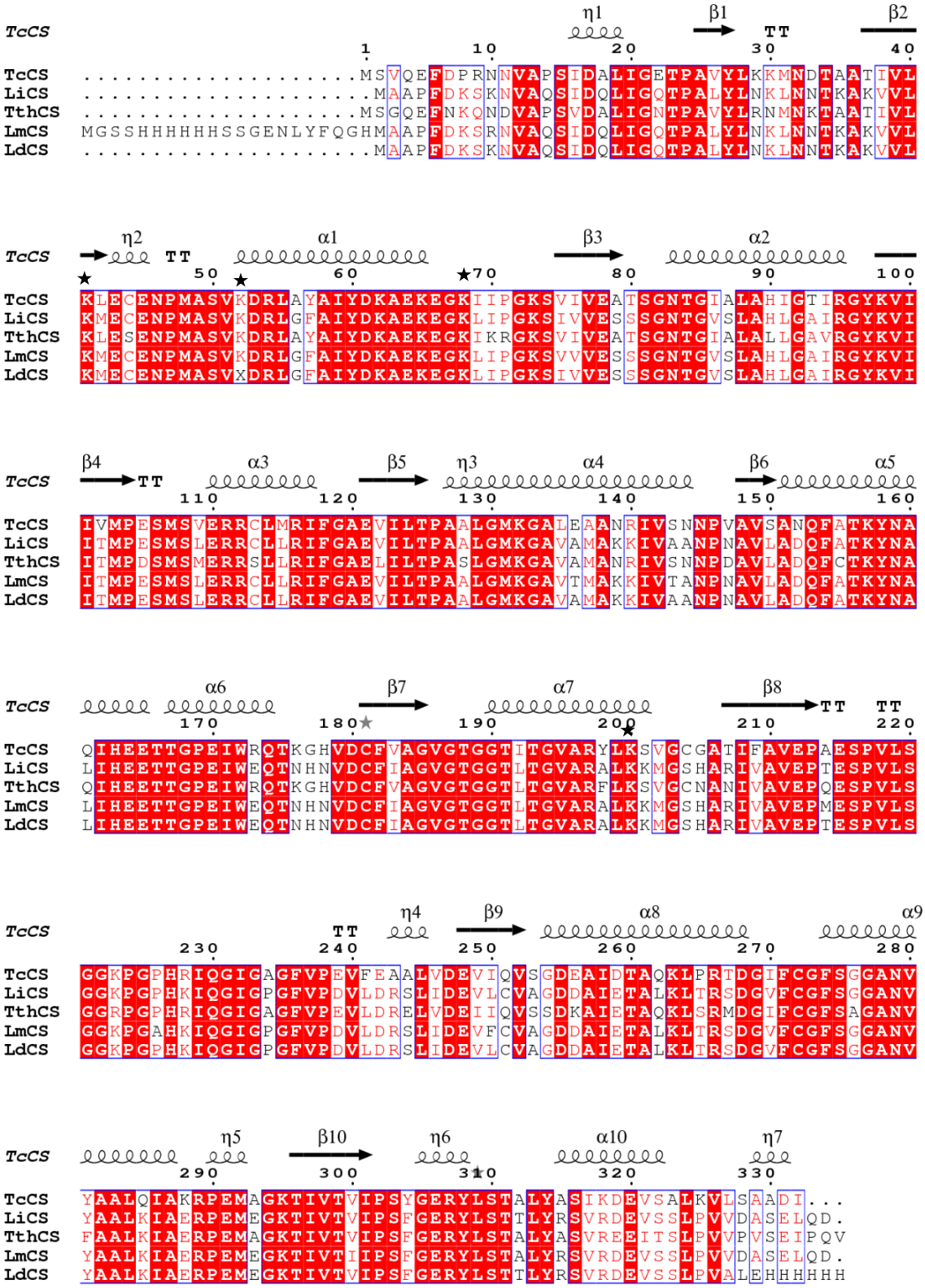


Figure 4-2 - Sequence alignment cysteine synthase from *Trypanosoma cruzi* (*Tc*CS), *Leishmania infantum* (*Li*CS), *Trypanosoma theileri* (*Tth*CS), *Leishmania major* (*Lm*CS) and *Leishmania donovani* (*Ld*CS) with structure from *Tc*CS (Sowerby et al., 2023). Stars indicate catalytically relevant residues.

Protein expression and purification

For expression and purification of the protein, a *Tc*CS expression vector was designed, inserting the *Tc*CS sequence into the pJOE5751 vector using the BamHI and BsrGI restriction sites to add an N-terminal polyhistidine tag (Wegerer et al., 2008). This protein construct was calculated to have a molecular weight of 36 608 Da. pJOE5751_*Tc*CS was transformed into *E. coli* BL21 and expression tests were performed to determine ideal expression conditions, shown in Figure S9-4-7. It was concluded that expression at 37°C for 16 hours was the optimum condition for *Tc*CS expression. Expression vectors of *Li*CS and *Tth*CS were then designed. The sequence was inserted into the pET15b vector using the NdeI and XhoI sites to allow for the addition of a N-terminal polyhistidine tag. pET15b_*Li*CS was calculated to be 37 380 Da whilst pET15b_*Tth*CS was 37 734 Da. Plasmid maps for all three constructs are shown in Figures S9-4-2, S9-4-4 and S9-4-6. As before, expression tests were performed and the optimum expression conditions for both proteins was determined to be 30°C for 16 hours.

All three proteins were expressed and purified by IMAC. Samples were taken and analysed by SDS-PAGE, Figure 4-3.

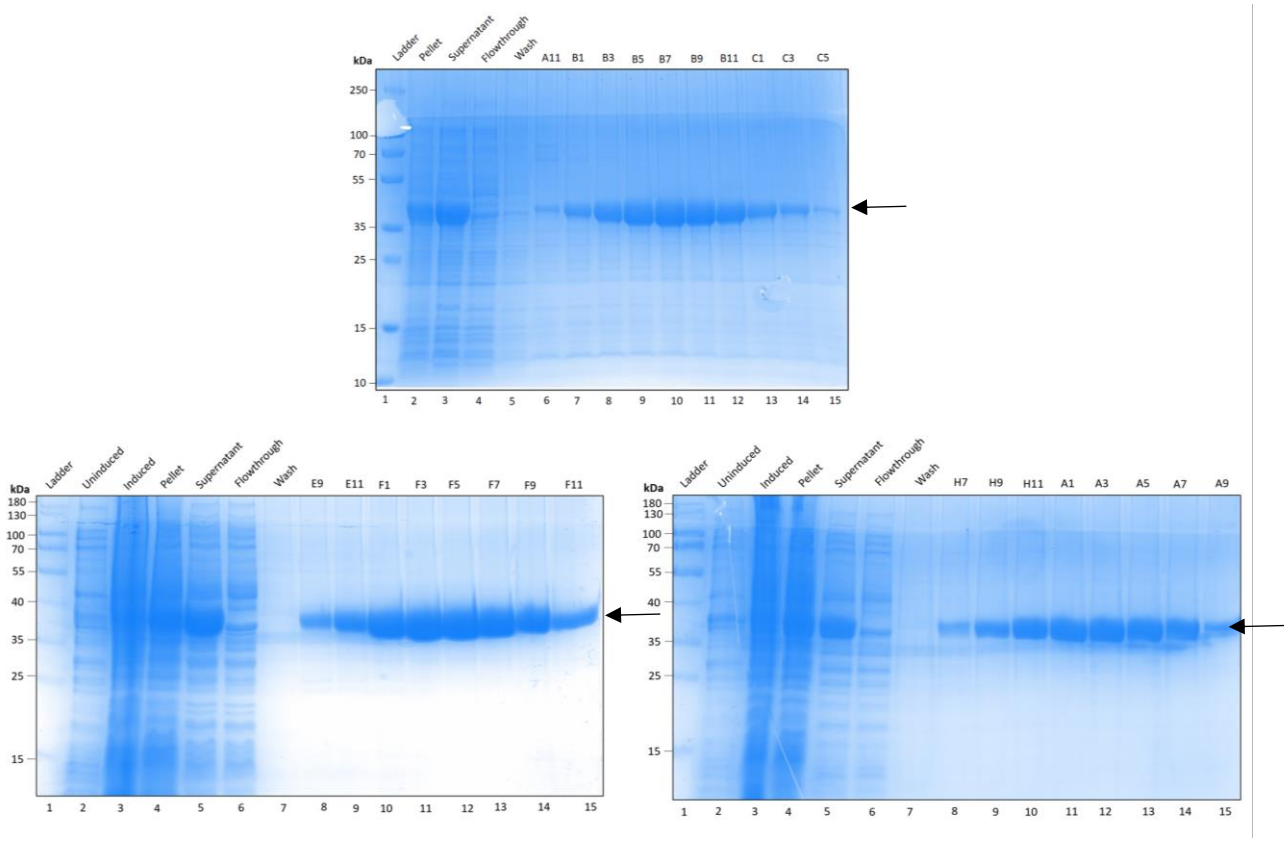


Figure 4-3 - Purification of *Tc*CS (above), *Li*CS (left) and *Tth*CS (right). Arrows indicate expected migration of protein.

All three gels show the successful purification of soluble protein of approximately the correct size in large yields (approximately 10 mg per L) and with no impurities present. These factors indicate that the protein expression and purification is sufficient and therefore further experiments can be conducted without additional purification stages.

Protein Characterisation

LC-ESI mass spectrometry was conducted on all three proteins, Figure 4-4. The predicted mass of *Tc*CS is 36 608 Da, the peak of 36 670 Da can be attributed to *Tc*CS within the error of the experiment. The peak of 36 840 Da in the *Tc*CS spectrum corresponds to the mass of *Tc*CS with a covalently bound PLP, which is 232 Da. For both the *Li*CS and *Tth*CS spectra, the main peak is within the experimental error of

the correct protein size. Additionally, both spectra have a peak of an additional 82 Da which can be attributed to an additional ion such as sodium acetate.

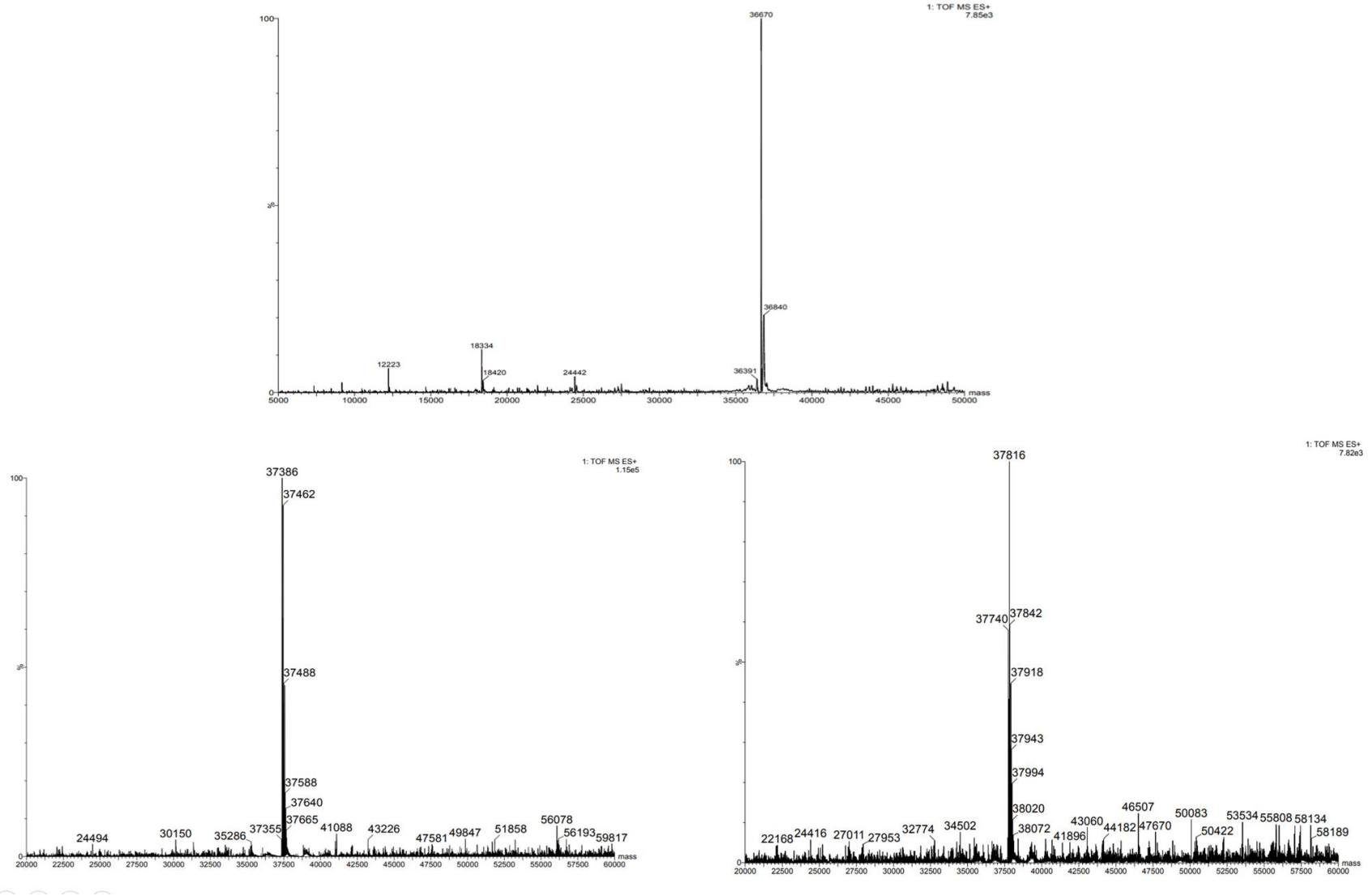


Figure 4-4 - Mass spectra of *Tc*CS (above), *Li*CS (left) and *Tth*CS (right).

Gel filtration experiments were conducted on *Li*CS and *Tc*CS proteins to provide insight into the stoichiometric state in solution. Size exclusion chromatography was performed, Figure 4-5. SDS-PAGE gels show that the protein from these fractions is of the correct size for each protein, *Tc*CS shown in Figure S9-4-10.

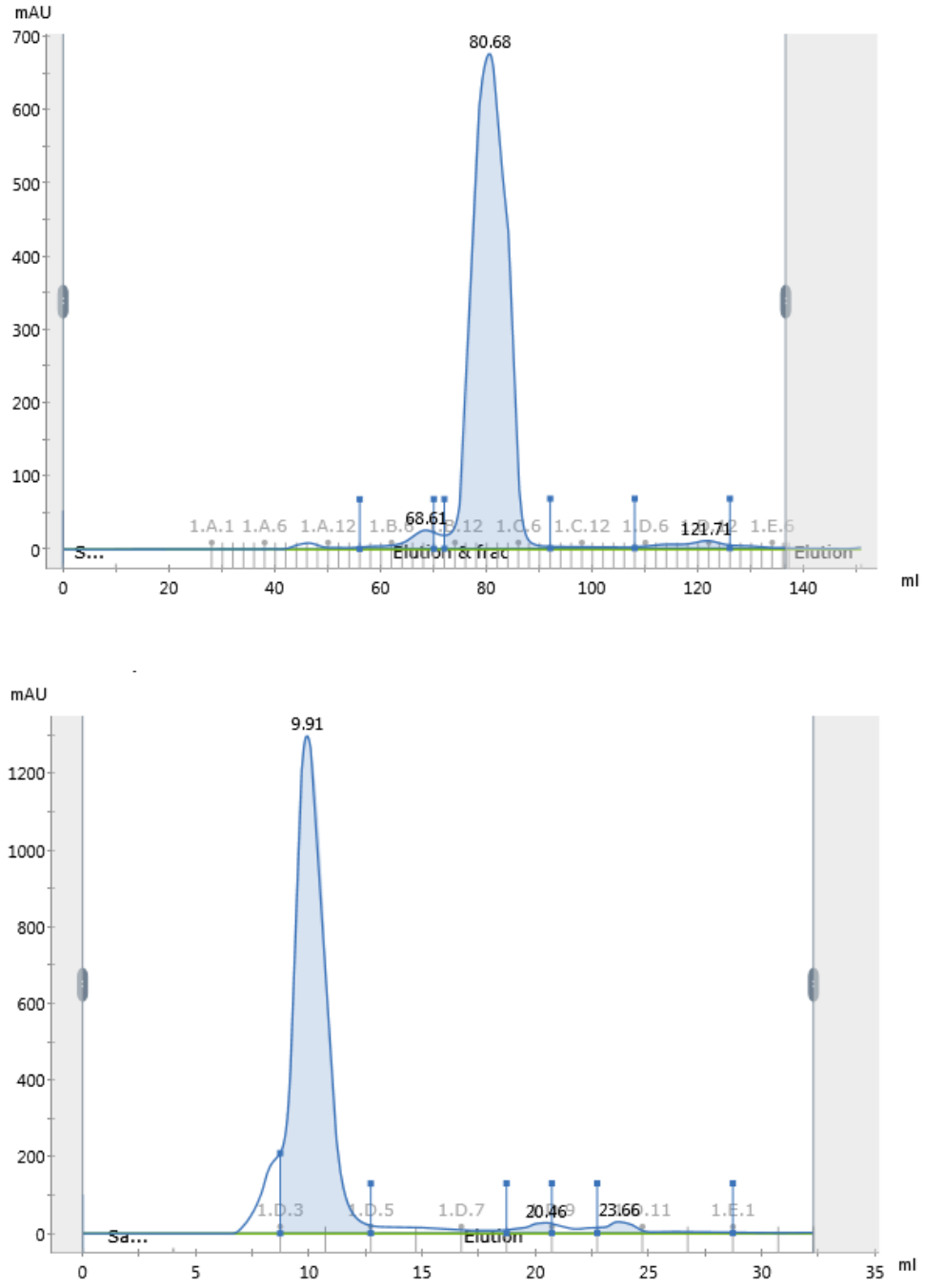


Figure 4-5 - Gel filtration chromatograms of *Li*CS (above) and *Tc*CS (below).

From the UV traces above and column calibration, a single peak at the correct elution volume for both *Li*CS and *Tc*CS to be dimers is seen. Due to the similarities between *Li*CS, *Tc*CS and *Tth*CS, and the experimental results all three proteins are suggested to be dimers in solution.

Determining enzymatic activity of cysteine synthase from *T. cruzi, L. infantum* and *T. theileri*

Cysteine production is typically detected using a well-established ninhydrin based assay (Gaitonde, 1967). To test the activities of *Tc*CS, *Li*CS and TthCS, this assay was converted into a 96 well format to establish a high throughput method (Sowerby et al., 2023). Cysteine production was determined for each recombinant enzyme with increasing OAS concentration and a constant sulphide concentration, Figure 4-6. From Figure 4-6 it can be seen that the activity of the cysteine synthases is dependent on increasing OAS concentration. The K_m and V_{max} values for each protein, Table 4-1, show that all three proteins have a similar activity particularly at low (sub mM) concentrations of OAS. *Tc*CS and *Tth*CS activity remains equivalent within the error of the experiment.

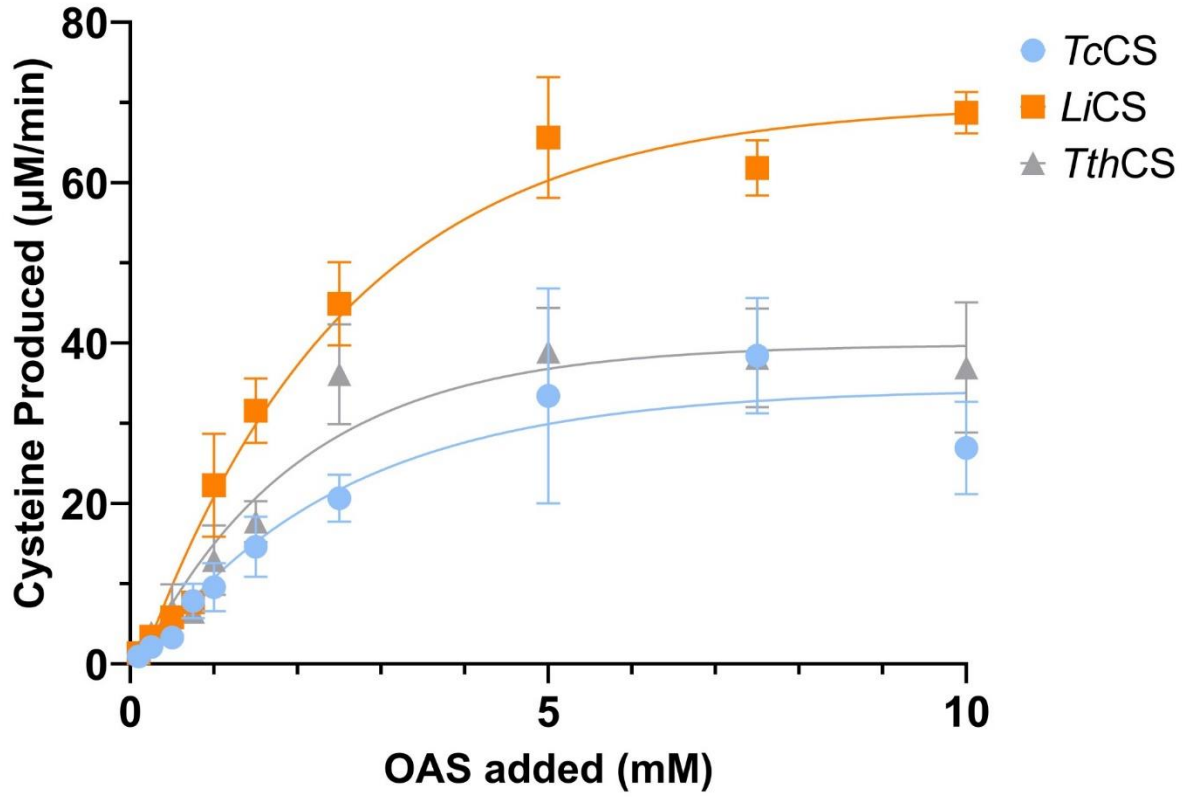


Figure 4-6 - Enzymatic Activities of *Tc*CS (blue), *Li*CS (orange) and *Tth*CS (grey). For each protein, n=6 with a minimum of 4 data points used for each measurement.

Table 4-1 - K_m and V_{max} values for *Tc*CS, *Li*CS and *Tth*CS calculated from Figure 4-6.

	K_m (mM)	**V_{max} (µM min^{-1})**
***Tc*CS**	3.1	43
***Li*CS**	3.5	97
TthCS	2.4	50

Due to the high sequence similarity between *Tc*CS and *Tth*CS, they can be expected to share a similar cysteine dependence. *Li*CS produces cysteine at significantly higher concentrations compared to *Tc*CS and *Tth*CS, indicating that under the assay conditions *Li*CS is more active with higher concentrations of OAS.

Structural studies of cysteine synthase

After establishing activity, structural studies were conducted. All three proteins were dialysed into 1x SGC buffer and concentrated to 3.8 $mgmL^{-1}$, 12.0 $mgmL^{-1}$ and 12.7 $mgmL^{-1}$ for *Tc*CS, *Li*CS and *Tth*CS respectively. High throughput crystallisation screens were set up for each protein. These screens included BCS, Clear Strategy™ Screen I, Clear Strategy™ Screen II, JCSG-plus, MIDASplus™, Morpheus®, Morpheus® II, PACT premier™, Structure Screen 1 + 2, ShotGun 1.

For *Tc*CS, initial crystals were formed in conditions containing ammonium sulphate. *Tc*CS crystals, Figure 4-7, were promising but required optimisation to produce larger crystals.

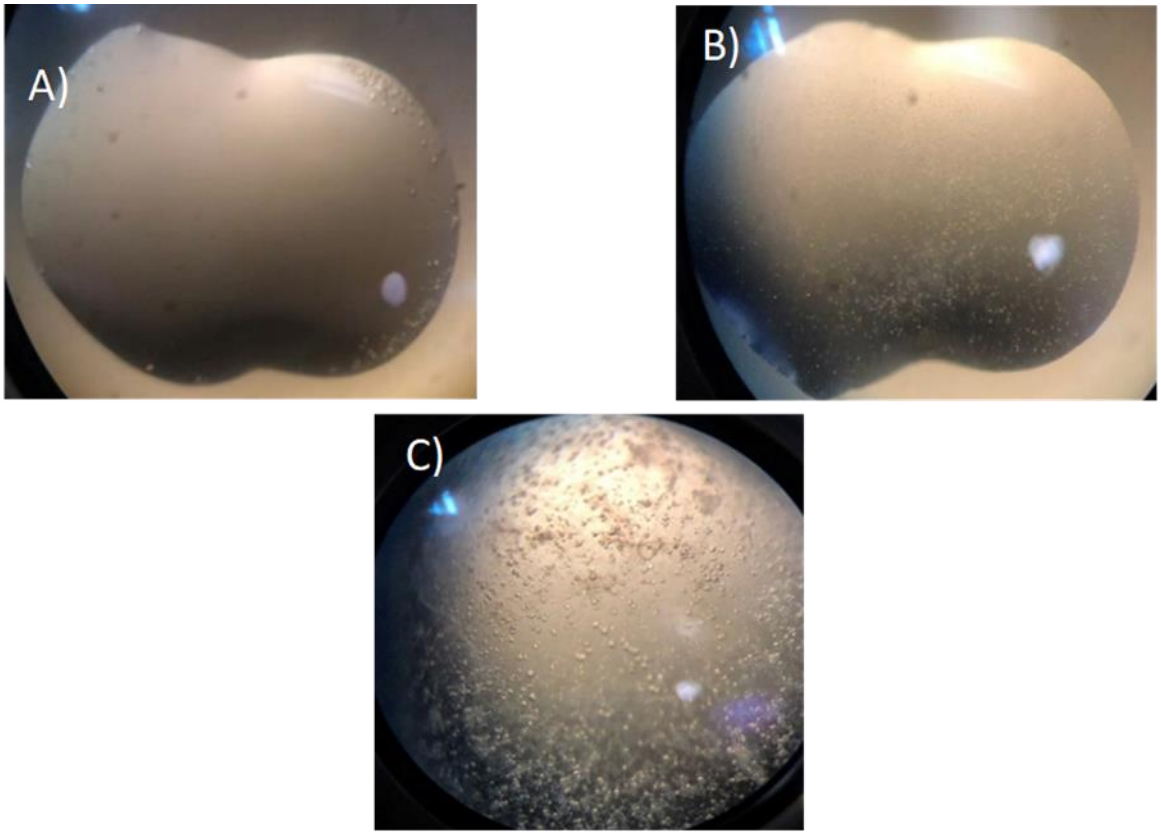


Figure 4-7 - Typical *Tc*CS crystals produced from conditions in **Error! Reference source not found.**. A – *Tc*CS crystals produced with conditions of 2 M Ammonium Sulphate, 0.1 M Bis-Tris pH 5.5. B – *Tc*CS crystals produced with conditions of 0.2 M Ammonium Sulphate, 0.1 M MES pH 6.5, 30% w/v PEG5000 MME. C – *Tc*CS crystals produced with conditions of 2 M Ammonium Sulphate, 0.1 M HEPES pH 7.5.

Optimised crystallisation screens were designed by varying the crystallisation conditions shown in **Error! Reference source not found.**. Crystal screens were set up in both hanging drop and sitting drop trays. Crystals formed by vapour diffusion in hanging drop trays, Figure 4-8. These larger crystals are more suitable for X-ray crystallography. The best *Tc*CS crystals were obtained by 2.4 M ammonium sulfate, 0.1 M bis-Tris pH 6.0.

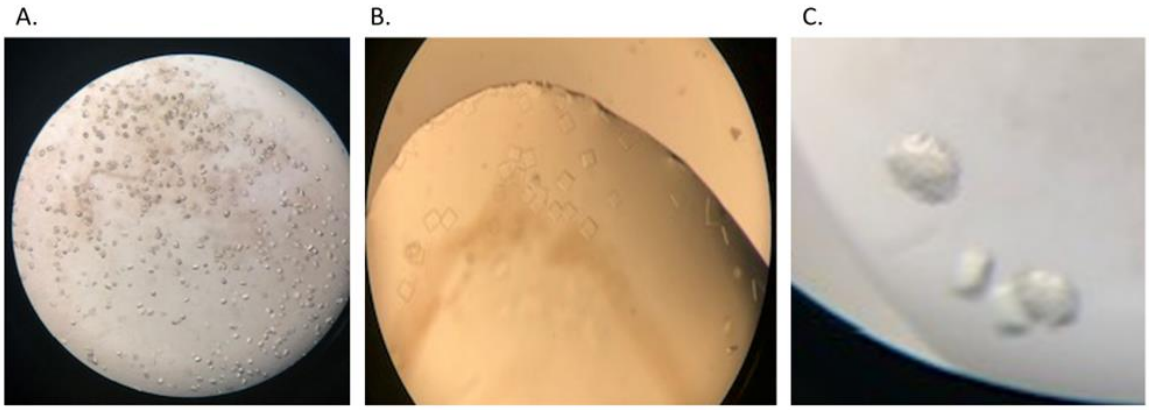


Figure 4-8 - Typical *Tc*CS crystals after optimisation.

For *Li*CS, a selection of initial crystal conditions are shown in Table 4-2 and typical crystals are shown in Figure 4-9. A full list of conditions forming crystals is shown in the appendix, Table S9-4-1. The initial crystals were small (~10 µm) and needle-like.

Table 4-2 - Conditions for initial *Li*CS crystal formation.

Conditions
0.1 M Bis-Tris pH 5.5 25 % w/v PEG 3350
0.2 M Sodium chloride 0.1 M BIS-Tris 5.5 25 % w/v PEG 3350
1.6 M Sodium citrate tribasic dihydrate pH 6.5

Figure 4-9 - Typical crystals from initial *Li*CS screen.

The optimisation of the initial crystals was conducted by setting up sitting drops in 24-well plates. Typical *Li*CS crystals are shown in Figure 4-10 and although still needle like, optimisation produced larger (~100 µm) and thicker crystals. The best crystals of *Li*CS were obtained using 0.2 M Sodium chloride, 0.1 M Bis-Tris pH 5.8, 30% (w/v) PEG3350.

Figure 4-10 - Typical *Li*CS crystals after optimisation.

For *Tth*CS, the same screening process was performed and conditions producing crystals are shown in Table 4-3 and crystals produced in Figure 4-13. The crystals produced from the initial screens were of suitable size and quality to be used in diffraction experiments without optimisation.

Table 4-3 - Conditions of *Tth*CS crystals

0.2 M Calcium chloride dihydrate 0.1 M Tris 8.0 20 % w/v PEG 6000
4.0 M Sodium formate
60 % v/v T-mate pH 7.0
2.7 M Ammonium sulfate 0.1 M Sodium cacodylate pH 6.5
2.7 M Ammonium sulfate 0.1 M Tris pH 7.5

For all three proteins, when suitable crystals were obtained, crystals were mounted in cryoloops, cryoprotected with 25% (v/v) glycerol before flash-cooling in $N_{2(l)}$ (Teng, 1990). High resolution data were collected, Figure 4-11, Figure 4-12 and Figure 4-13. The approximate size of the crystals were *Tc*CS - 120 µm, *Li*CS - 300 µm and *Tth*CS – 100 µm. Data were collected at 0.9762 Å using the IO3 and I24 beamlines at Diamond Light Source.

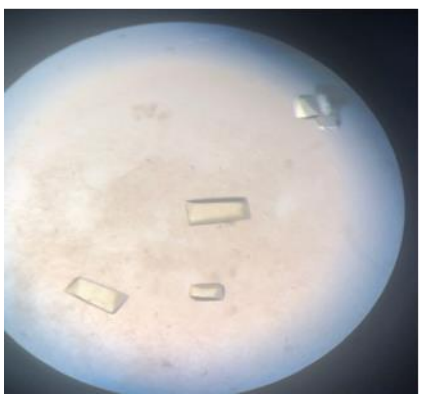
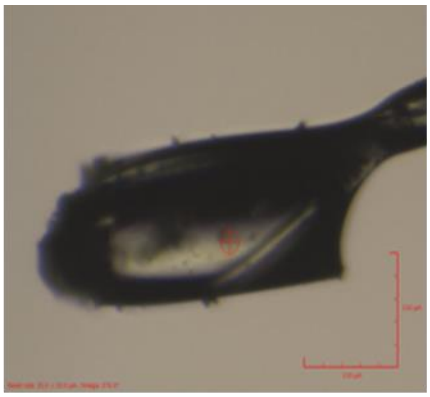
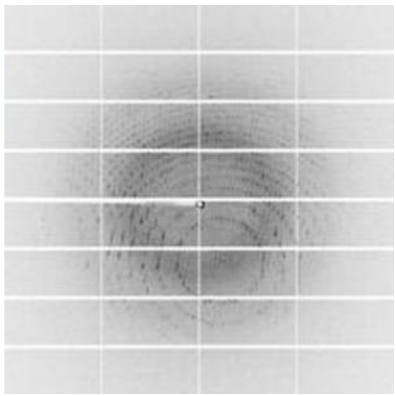




Figure 4-11 - *Tc*CS crystals in drop (left), single crystal in loop (center) and diffraction pattern produced (right).

Figure 4-12 - *Li*CS crystals in drop (left), single crystal in loop (centre) and diffraction pattern produced (right).

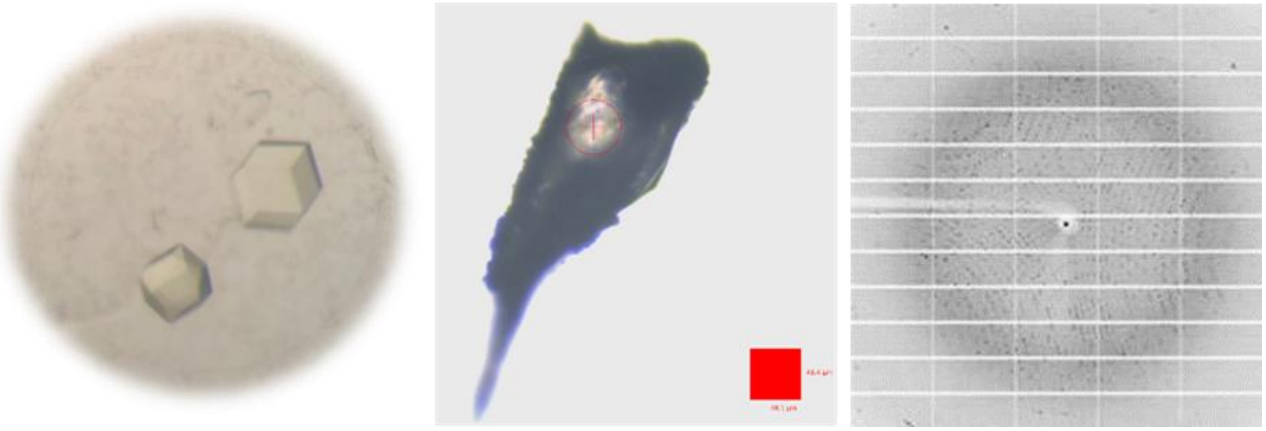


Figure 4-13 - *Tth*CS crystals in drop (left) single crystal in loop (centre) and diffraction pattern produced (right).

4 data sets were collected from 3 different *Tc*CS crystals on the I-03 Diamond beamline. All data sets were processed using XIA2/DIALS and statistics for the best dataset are shown in Table 4-4 (Winter, 2010). *Tc*CS data required manual processing as auto processing assigned the incorrect space group of $P2_1\,2_1\,2_1$ to the data. The *L. major* CS shares a 78% sequence identity with *Tc*CS and the structure (PDB code 4AIR) was used as the search model for *Tc*CS (Fyfe et al., 2012). During

molecular replacement 4 molecules in the asymmetric unit were searched using PHASER for based on the Matthews coefficient (Matthews, 1968). The best crystal diffracted to a resolution of 1.8 Å and was successfully processed in the space group P 2_1; this space group typically has α=β=γ=90°, however in this data set the angles are α=γ=90°, β=89.5 due to the presence of pseudo-orthorhombic symmetry in the crystal.

Table 4-4 - Data collection statistics for *Tc*CS, *Li*CS and *Tth*CS.

Dataset	***Tc*CS**	***Li*CS**	**TthCS**
	Data Collection		
Beamline	I03 DIAMOND	I24 DIAMOND	I24 DIAMOND
Detector	Eiger2 XE 16M	Pilatus3 6M	Pilatus3 6M
Wavelength / Å	0.9762	0.9762	0.9762
Exposure Time / seconds	0.05	0.02	0.02
Transmission /%	100	50	50
Degrees Collected /°	360	180	180
Image Slices /°	0.1	0.05	0.05
Resolution Range / Å	54.94 - 1.80 (1.83 - 1.80)	87.77 – 1.75 (1.78 – 1.75)	76.47 – 2.75 (2.80 – 2.75)
Space Group	*P* 2_1	*P* 2_1 2_1 2_1	*I* 2 3
a, b, c, / Å	54.9, 66.6, 167.4	48.9, 87.8, 138.0	187.4, 187.4, 187.4
αβγ / °	90, 89.5, 90	90, 90, 90	90, 90, 90

Molecules/ASU	4	2	2
R_{merge}	0.054 (0.216)	0.099 (0.566)	0.213 (0.774)
R_{pim}	0.038 (0.160)	0.031 (0.228)	0.029 (0.104)
CC (1/2)	0.9961 (0.9059)	0.991 (0.805)	0.997 (0.962)
Multiplicity	3.37 (3.15)	11.8 (7.8)	56.3 (57.3)
Completeness / %	98.27 (96.34)	99.6 (97.5)	100 (100)
I / σ(I)	12.93 (1.01)	11.96 (1.02)	23.50 (1.81)
Unique reflections	110250 (5376)	60590 (2908)	28493 (1434)
Refinement			
No. of reflections used	110107	60520	28475
R_{work}	0.191	0.164	0.148
R_{free}	0.217	0.205	0.209
Protein atoms	9394	4615	4857
Asymmetric unit contents	2 homodimers	homodimer	homodimer
Ligands	4 PLP, 1 OAS, 2 glycerol, 1 ribose	-	2 PLP
RMS deviations			
Bond lengths (Å)	0.01	0.01	0.01
Bond Angles (°)	1.51	1.80	2.65
Mean B Values			
Protein	26.6	26.4	71.5

Water	35.7	34.5	62.4
Ligands	34.3	35.3	85.0
Ramachandran plot (%)			
Favoured	98.2	97.8	95.4
Allowed	1.8	1.8	4.1
PDB codes	8B9Y	8B9M	8B9W

***Tc*CS structure**

The *Tc*CS structure has been determined to 1.80 Å, with no outliers present in the Ramachandran plot. Excellent R-factors for a structure of this resolution were achieved with R_{work} at 19.1% and R_{free} at 21.7% demonstrating a well refined structure at high resolution. The average B-factors for the protein, ligands and water molecules are all within the same range between 26 and 36 $Å^2$. The model contains 4 chains in the asymmetric unit resulting in two homodimers and conserves the overall fold expected for cysteine synthase enzymes. The overall fold of all four protein chains is essentially identical, with r.m.s.d.s between individual monomers ranging from 0.30 to 0.83 Å when α carbons of all residues were compared.

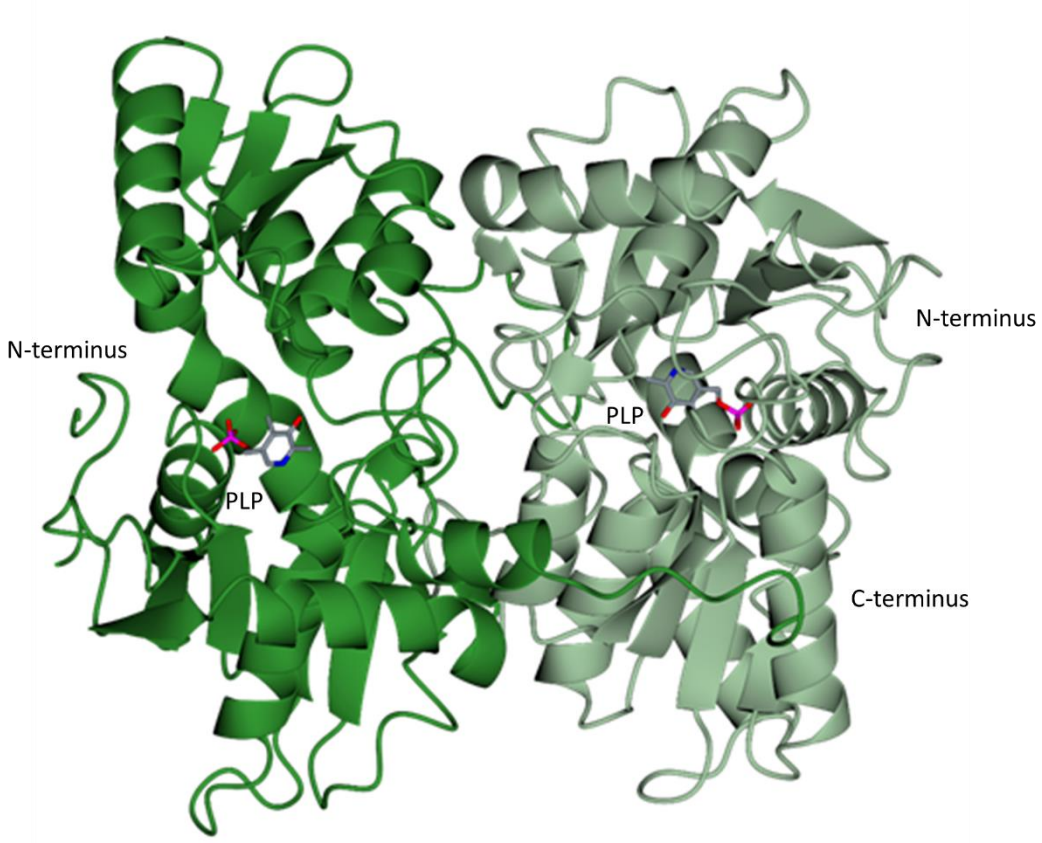


Figure 4-14 – Homodimeric *Tc*CS structure shown in ribbon diagram. Chain A in dark green, chain B in light green.

The *Tc*CS model includes 344 residues in each chain; the density for residues begins at Val^{3} so previous residues, such as the polyhistidine tag were not included in the model. Poorly ordered residues were not placed in the electron density; these include residues 221–230 in chain A, 315–325 in chain B and 319–325 in chain D. As expected, given the high level of sequence identity, *Tc*CS shares a similar fold with other cysteine synthases, with the exception of the C-terminus. The C-terminal tail of each chain is formed by a flexible region that stretches across to the partner subunit of each dimer,

Analysis of the structure with PISA (Krissinel et al., 2007), shows that *Tc*CS is a dimer confirming the size exclusion experiment in solution. Chain A has a buried surface of

2862 A^2 whilst Chain B has a buried surface of 2774 A^2. Contrastingly, Chain C has a buried surface of 3037 A^2 and Chain D has a buried surface of 2982 A^2. For both subunits, a large amount of the key interactions of the dimer are formed from the C-terminal tails; these flexible regions are varied, Figure 4-15. Within the two dimers of *Tc*CS the AB dimer interface forms from 23.4% of residues (73 residues) from Chain A and 22.5% of residues (71 residues) of Chain B. There is a greater amount of interface for the CD dimer, with the C chain using 24.8% of residues (82 residues) and the D chain using 24.3% of residues (78 residues) to form the dimer interface. The A, C and D chain C-terminal tails extend to the partner subunit however the B chain C-terminal tail folds back into the B chain, entering the active site, Figure 4-15. These two dimers show that there is significant flexibility of the C-terminus.

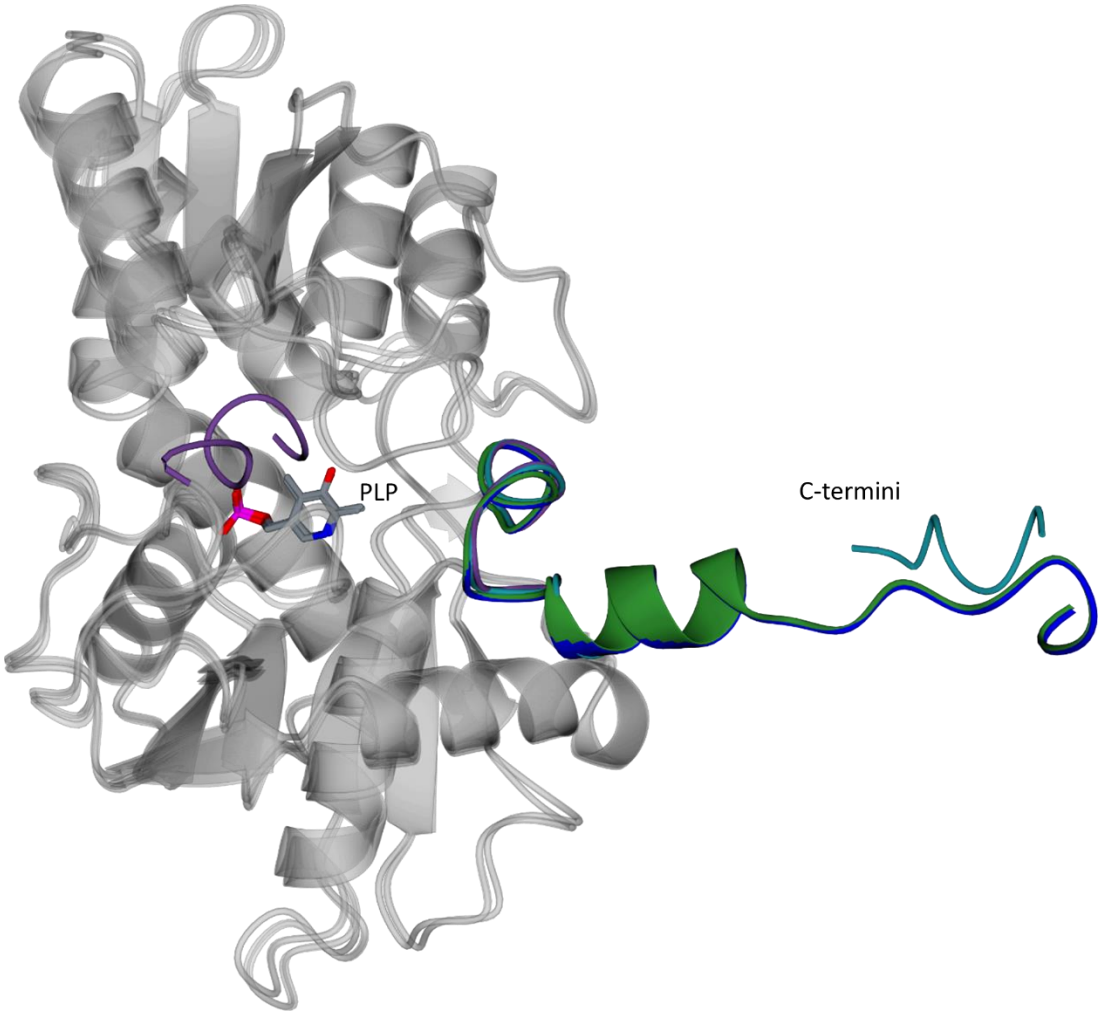


Figure 4-15 - Superposition of *Tc*CS chains in ribbon form showing flexibility in C-terminus. Chain A in dark green, chain B in purple, chain C in dark blue, chain D in teal. PLP shown in stick representation.

Investigation into the active site of TcCS

In the active site of each *Tc*CS subunit, additional electron density fitting to PLP was found. One molecule of PLP was found to be covalently bound to Lys^{52} in each subunit. Additionally, Thr^{188}, Gly^{189}, Thr^{191} and Ser^{275} all form hydrogen bonds to the cofactor, Figure 4-16. The importance of these residues can be seen as they are highly conserved throughout the CS sequences. The residues involved in the binding of PLP were identified during bioinformatic analysis.

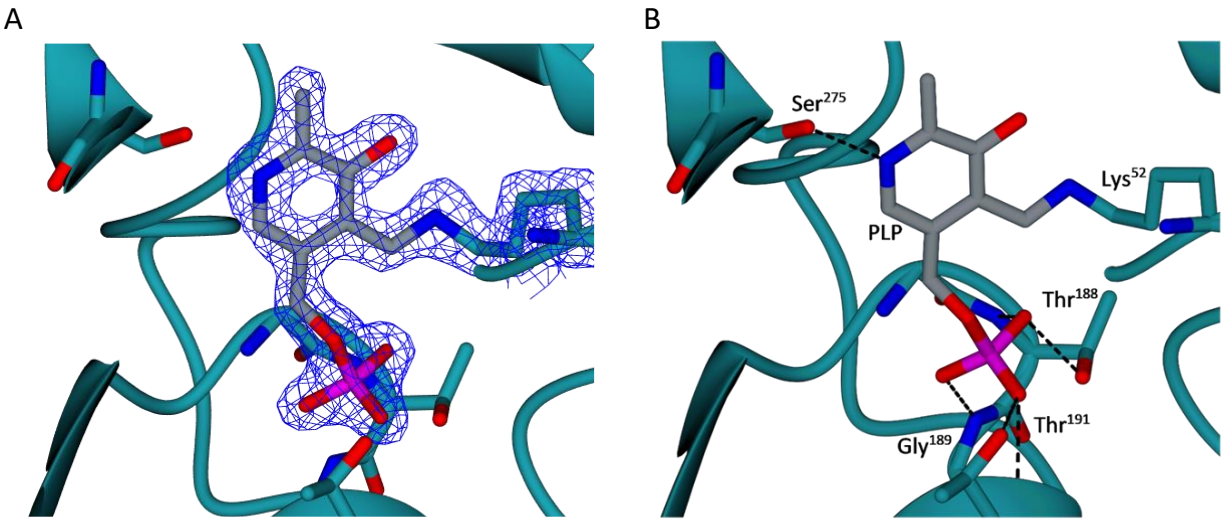


Figure 4-16 - Close up of *Tc*CS active site with PLP covalently bound. A- $2F_o$-F_c density for PLP and Lys^{52} shown in blue mesh at 1 σ. B – labelled ribbon diagram of *Tc*CS active site. Hydrogen bonding is indicated by dashed lines. H atoms are omitted for clarity.

In the active site of subunit A, additional density to the PLP was present. This density is fitted by O-acetyl-l-serine, an essential substrate of the cysteine biosynthesis reaction, Figure 4-17. The capture of this intermediate allies with the canonical reaction mechanism for cysteine synthase. Hydrogen bonds between Lys^{52}, Asn^{83} and OAS are formed. The reactive N atom of the OAS is 5.7 Å from the O atom of PLP that together form the Schiff base. The presence of OAS in the active site of *Tc*CS shows the structural changes that occur in the presence of reaction intermediates. The position of OAS within the active site of *Tc*CS shows significant variability, which is consistent with results from other structures, for example that of O-acetylserine sulfhydrylase from *Haemophilus influenzae* in complex with OAS (PDB 5DBE;Singh et al., 2015. The placement of OAS in this position is consistent with unbiased electron density; however, to complete the next step of the reaction the PLP needs to react with the amino group of the OAS to form the intermediate

Schiff base. This step would require a rearrangement of the lysine side chain as well as of the substrate.

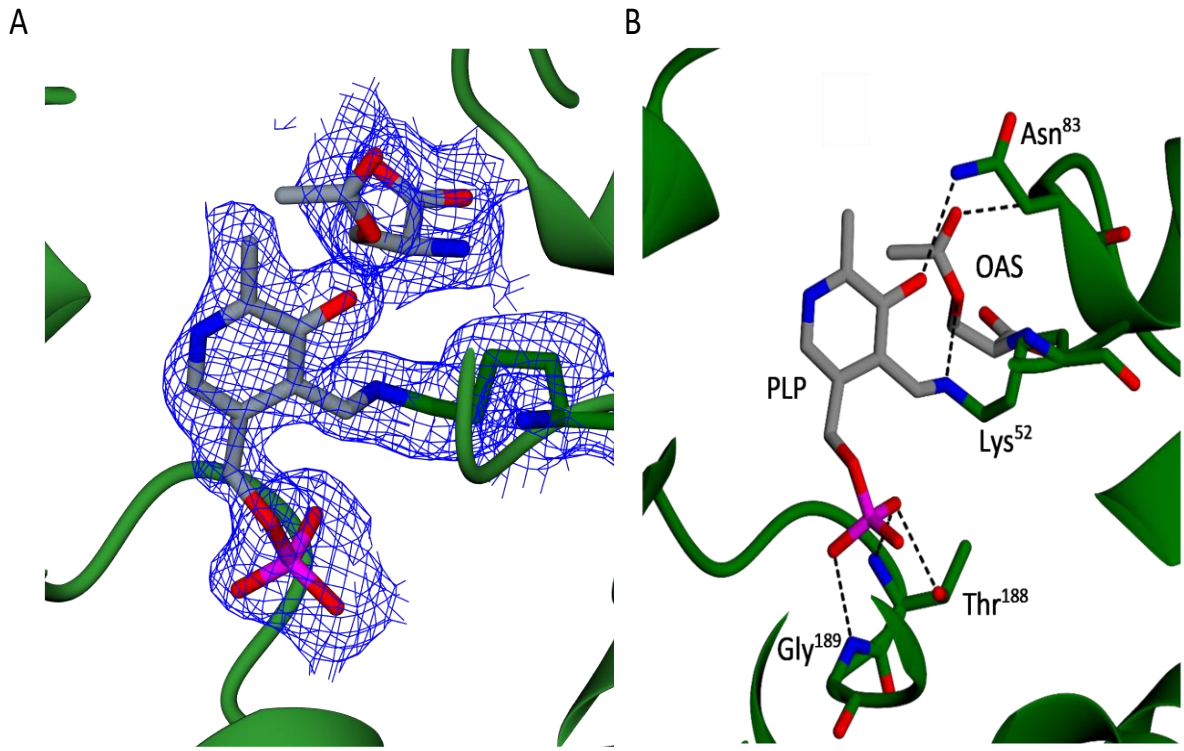


Figure 4-17 - Closeup of the active site of *Tc*CS with PLP and OAS bound. A- $2F_o$-F_c density for PLP, OAS and Lys^{52} shown in blue mesh at 1 σ. B – labelled ribbon diagram of *Tc*CS active site. Hydrogen bonding is indicated by dashed lines. H atoms are omitted for clarity.

A loop formed of residues 221–237 covers the active site in the absence of OAS in the active site. In the presence of OAS this loop shows increased flexibility. Although further studies would be required to confirm, the absence of this loop over the active site could allow substrates such as hydrogen sulphide to interact with the OAS and PLP, allowing the reaction to proceed. This shows that cysteine synthase has both an open and a closed form. The contrast between the two structures shows the structural changes that can occur during different stages of the reaction cycle.

Molecules at the dimer interface of TcCS

Both of the dimer interfaces of *Tc*CS had small molecules bound. Based on the unbiased electron density at the AB interface, ribose was determined to be bound, Figure 4-18A and B. Residues Glu^{57}, Asn^{58} and Pro^{59} form hydrogen bonds to ribose. At the CD interface, glycerol was found, Figure 4-18C and D. Hydrogen bonds are formed between Cys^{56}, Glu^{57}, Asn^{58} and the glycerol molecule.

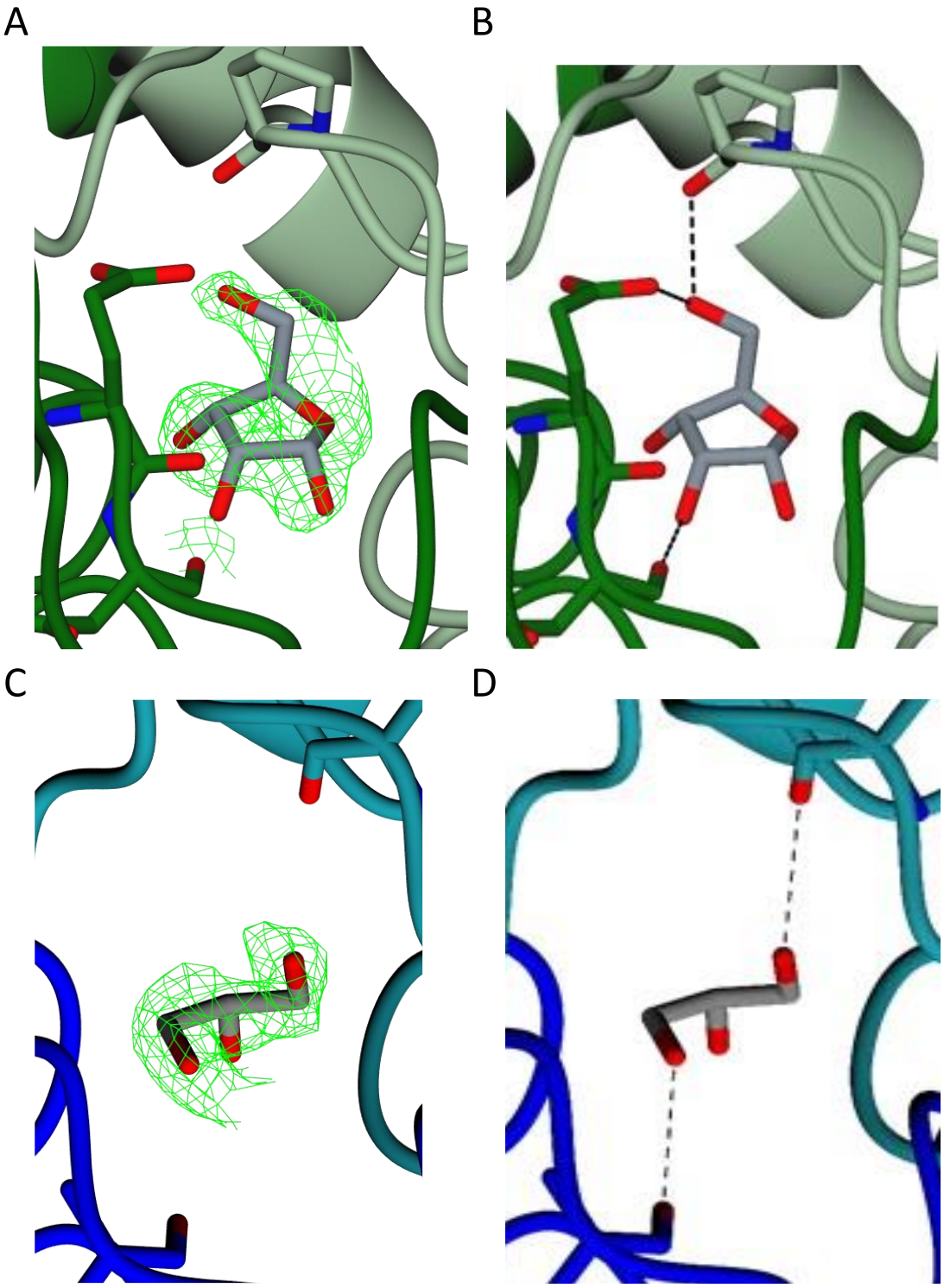


Figure 4-18 - Molecules at the *Tc*CS dimer interfaces. A. Ribose at the AB interface with a F_o-F_c density map shown in green mesh at 1 σ. B – Hydrogen bonding pattern of ribose with *Tc*CS. C - Glycerol at the CD interface with a F_o-F_c density map shown in green mesh at 2 σ. D – Hydrogen bonding pattern of glycerol with *Tc*CS. Hydrogen bonding is indicated by dashed lines. H atoms are omitted for clarity.

Ribose was not added to the growth media suggesting that during protein expression this molecule was scavenged from the *E. coli* cells. Ribose is an abundant sugar with great cellular importance due to its role in RNA and DNA and therefore would be present in large quantities within the cell during protein production.

Glycerol was present in the buffer after purification and continued to be used in crystallisation conditions and as the cryoprotectant. Both are hydrophilic molecules and the presence of two different, hydrophilic molecules which form hydrogen bonds with the same residues indicate a hydrophilic pocket formed by the dimer interface. This could be evidence of an allosteric binding site at the dimer interface. Additionally, this represents a promising starting point for the development of a protein-protein interaction (PPI) inhibitor for the interruption of dimer formation.

_Li_CS structure

The *Li*CS structure was solved at 1.75 Å resolution, Table 4-4 and Figure 4-19. Briefly, the R factors for this model are R_{work} of 16.4% and R_{free} of 20.5%. The average B-factors for the protein, ligands and water are within the range of 26 – 35 $Å^2$ and 99.8% of the residues are within favoured or allowed areas of the Ramachandran plot. The asymmetric unit contains one homodimer. Each *Li*CS monomer adopts a compact structure composed of 9 α-helices and 12 β-sheets with an extended C-terminal tail. With the exception of β1, all β-sheets run parallel. The active site of *Li*CS is found at the centre of each subunit, between α1 and α5. The C-terminal tail of each chain is formed by a flexible region that stretches across to the partner subunit of each dimer.

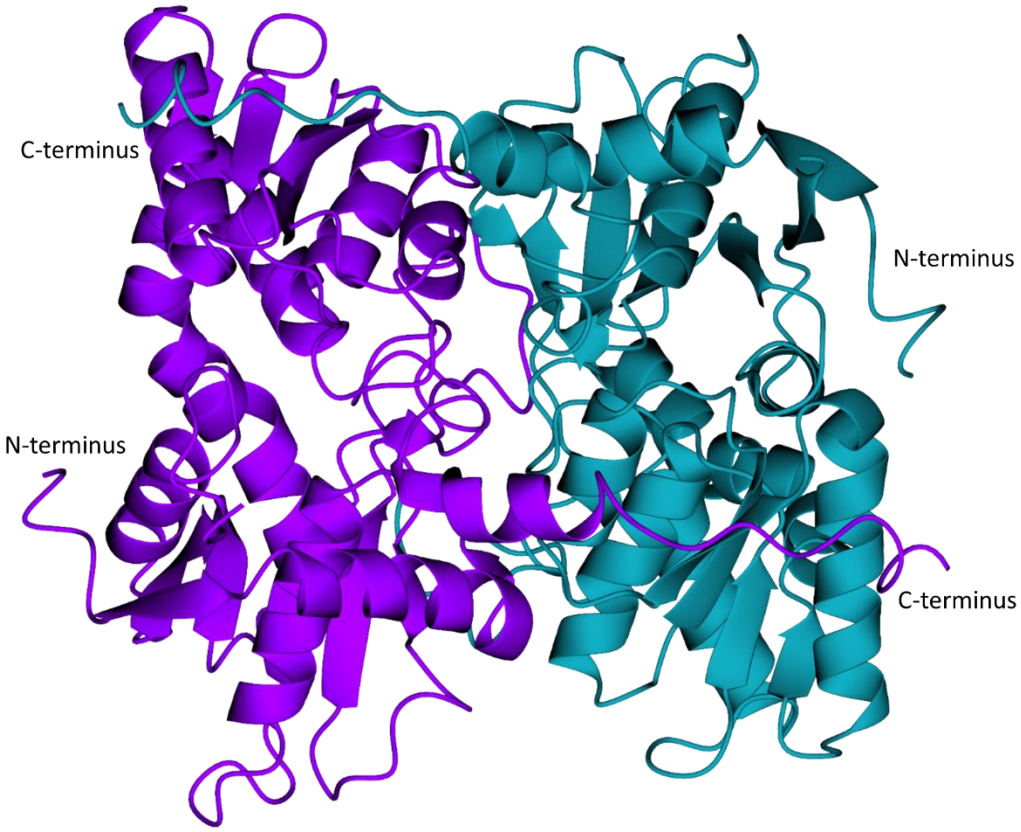


Figure 4-19 - Ribbon diagram of *Li*CS dimer. Chain A in purple, chain B in teal.

Analysis with PISA shows that *Li*CS has a large dimer interface with 27.0% of chain A and 26.6% of chain B residues involved in forming the dimer interface (Krissinel et al., 2007). This extensive interface has a buried surface of approximately 3246 A^2.

Although the purified protein was yellow in colour during purification, suggesting the presence of the cofactor PLP; the crystals were not coloured suggesting that the PLP was lost during crystallisation. As there is no PLP identified at the active site lysine of *Li*CS, this conformation corresponds to the apo form. The apo form of *Li*CS shows key changes that occur in the structure in the absence of PLP, such as the absence of a flexible loop covering the active site.

*Tth*CS structure

To explore the mechanism of cysteine synthesis in further atomic detail, the structure of *Tth*CS was solved at 2.75 Å resolution, Table 4-4 and Figure 4-20. The structure has an R_{work} of 14.8% and an R_{free} of 20.9% with 99.5% of the residues within the favoured or allowed sections of the Ramachandran plot. The average B-factors for this model are between 62 and 85 $Å^2$ for the protein, ligands and water molecules. The *Tth*CS structure has one dimer per asymmetric unit and adopts the same overall fold as the apo *Li*CS and holo *Tc*CS structure. The two monomers have almost identical conformations, with an r.m.s.d. of 0.74 $Å^2$ between α carbons of all residues in the chain.

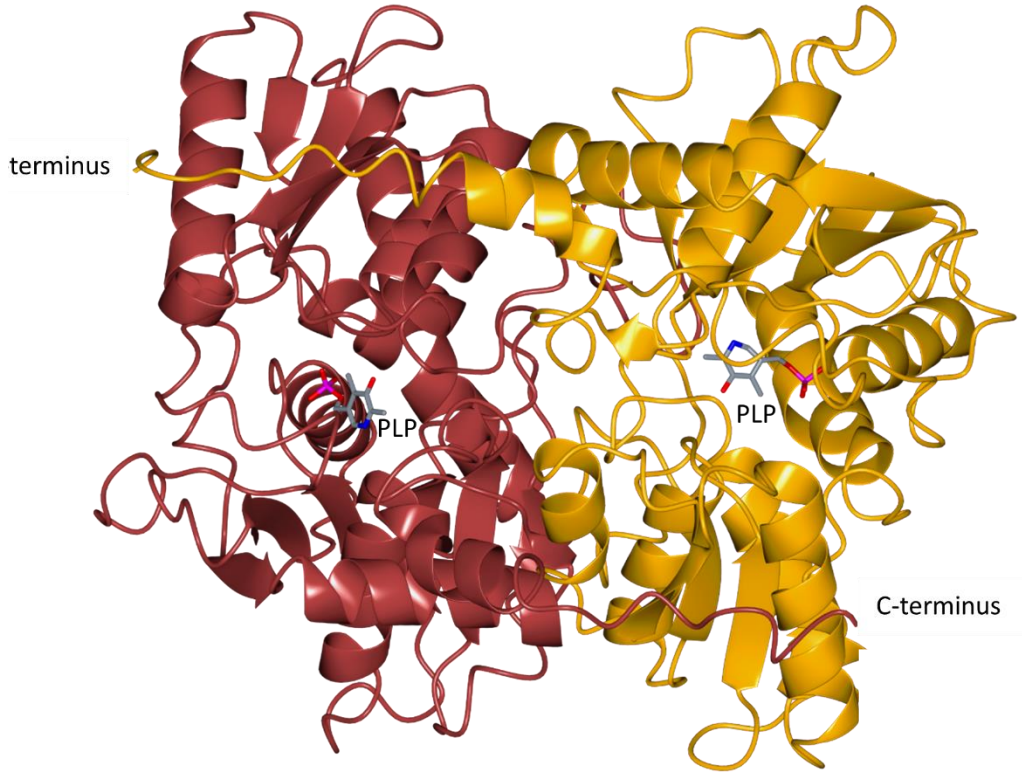


Figure 4-20 - Ribbon diagram of *Tth*CS dimer. Chain A in yellow, Chain B in brown. PLP is shown in stick representation.

Unsurprisingly, the *Tth*CS dimer also reveals an extensive dimer interface with buried surface areas of 3053 $Å^2$ for Chain A and 3099 $Å^2$ for Chain B. As 25% of the protein is involved in dimer interactions, it clearly indicates that the enzyme is a functional homodimer (Krissinel & Henrick, 2007).

In the active site of cysteine synthase, the PLP cofactor is covalently bound to the active-site lysine, Figure 4-21. In each chain of *Tth*CS continuous electron density was found which fits to the structure of PLP. In this structure the covalently bound state of PLP has been captured.

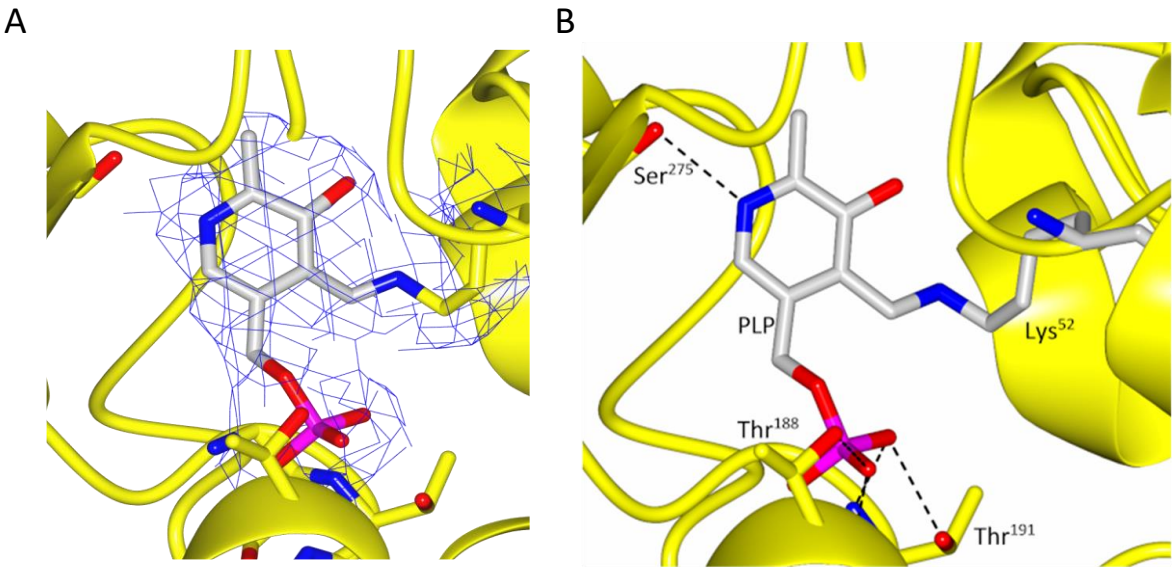


Figure 4-21 - Close up of *Tth*CS active site. A- $2F_o$-F_c density for PLP and Lys^{52} shown in blue mesh at 1 σ. B – labelled ribbon diagram of *Tth*CS active site. Hydrogen bonding is indicated by dashed lines. H atoms are omitted for clarity.

Both subunits of the dimer have PLP covalently bound at Lys^{52}, forming a Schiff base which establishes conserved hydrogen bonds to Thr^{188}, Thr^{191} and Ser^{275}. As seen in the multiple sequence alignment of the proteins, the indicated residues are conserved. The binding of PLP confirms the involvement of the residues predicted by bioinformatic analysis.

Overall, the three CS structures share a similar fold, superposition shown in Figure 4-22 with r.m.s.d.s in the range of 0.6 – 1.1 Å when α carbons of all residues were used. From the experiments described above, all three structures are functional dimers in solution, which is consistent with other CS structures studied. In these structures, key parallels can be seen in the active site of each protein suggesting that the reaction mechanism will be consistent with other CS proteins and each other. Further, this implies that any inhibition affecting the active site of one protein

would be capable of translating to the other two proteins, and potentially wider to other CSs. Additionally, due to the similarities between *Tc*CS, *Li*CS and *Tth*CS, it can be assumed that findings in one enzyme can be transferred to the other two to provide deeper understanding of the molecular interactions occurring.

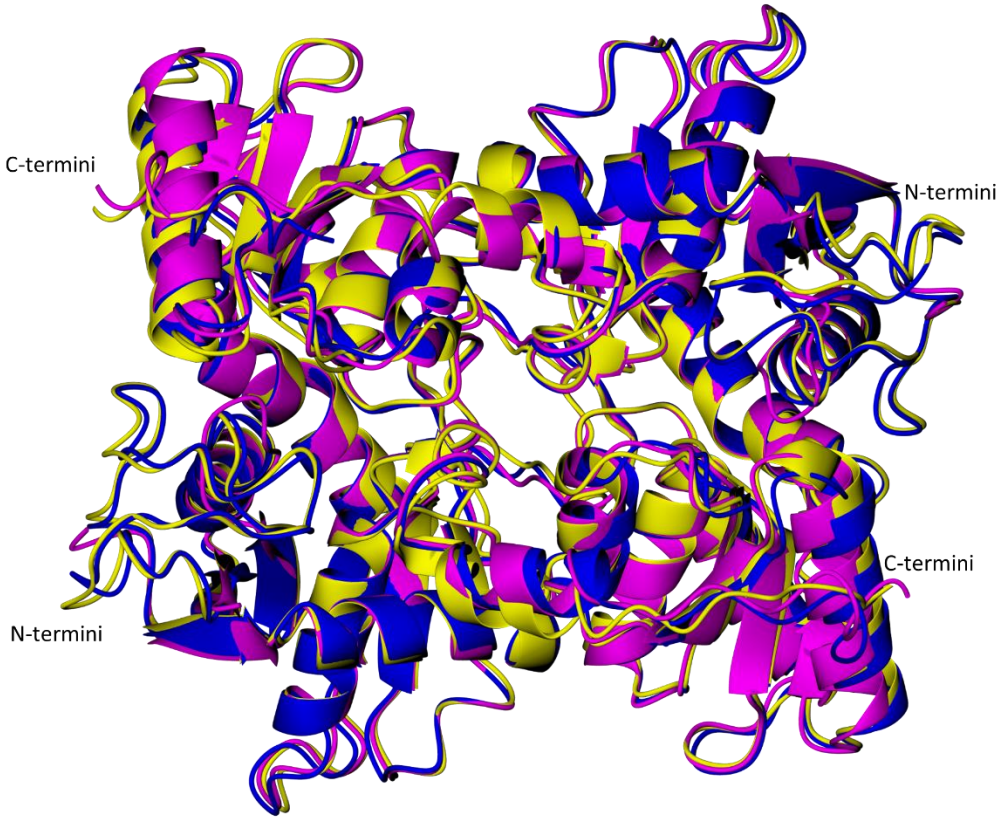


Figure 4-22 – Ribbon diagram of least squares superposition of *Tc*CS (dark blue), *Li*CS (pink) and *Tth*CS (yellow).

The covalently bonded PLP in the active site of both *Tc*CS and *Tth*CS represent the holo forms of the enzyme whilst the lack of PLP in *LiCS* provides the apo form of CS. The presence of PLP enables the catalysis of OAS to L-cysteine and although absent in the *Li*CS structure, it is expected that the lack of PLP is a result of crystallisation as the protein is active. Both the *Li*CS structure and Chain A of the *Tc*CS structure are in the open form whereas the three other chains of *Tc*CS and the *Tth*CS structure are

seen in the closed form. The flexible loop formed of residues 221-237 in *Tc*CS and 221-233 in *Tth*CS are responsible for the open/closed form designation. This flexible loop is not present in the electron density of the open form structures.

The three CS structures elucidated in this work all represent discrete stages in the cysteine synthesis reaction route. *Li*CS, as the apo form, shows the protein without the cofactor whereas *Tth*CS, as the holo form, proves the residues responsible for PLP binding. *Tc*CS reveals a reactionary intermediate of the cysteine synthesis pathway due to the presence of OAS within the active site. This reveals the residues involved in OAS binding and additionally the subtle but important structural changes that occur during the reaction. Together the three structures give an important insight into the cysteine synthesis of trypanosomatids at a molecular level.

Conclusions

Recombinant *Tc*CS, *Li*CS and *Tth*CS have been expressed and purified in large amounts. The three proteins have undergone biochemical characterisation to show that they are of the expected size and are expected to form homodimers in solution. Biochemical analysis shows that all three proteins are active and capable of producing cysteine at increasing levels of OAS.

For all proteins studied, crystals were successfully produced and diffraction experiments conducted. Three high resolution structures of cysteine synthase from a range of trypanosomatids have been determined. Each structure provides a different stage of the reaction mechanism with *Li*CS providing the apo structure,

*Tth*CS the holo structure and *Tc*CS showing both a holo structure and a reactionary intermediate.

In addition to determining detailed information about the active site of cysteine synthase, further binding sites were identified. The elucidation of a hydrophilic pocket at the interface of each dimer reveals a potential starting point for the exploration of protein-protein interaction inhibitors through *in silico* studies. An interruption to the extensive dimer interface is likely to lead to the exposure of hydrophobic surfaces resulting in the aggregation of the protein and therefore prevention of cysteine catalysis.

When the structures of *Tc*CS, *Li*CS, *Tth*CS and *Lm*CS are compared, overall the canonical cysteine synthase fold is conserved. The r.m.s.d.s for *Tc*CS, *Li*CS and *Tth*CS are all within the range of 0.6 – 1.1 Å, shown in appendix, Table S9-4-2. The α carbons of residues 18-320 from *Tc*CS, 24-326 from *Li*CS, 25-327 from *Tth*CS and 18-325 from *Lm*CS were used. Upon closer examination of each monomer and in comparison to the previously determined *L. major* CS that was used for molecular replacement during structure solution for the *Tc*CS structure, the r.m.s.d.s vary from 0.30 – 1.1 Å, shown in appendix, Table S9-4-2. This suggests that globally the structures are very similar and when each monomer is compared this similarity is maintained. The A and C monomer of *Tc*CS have r.m.s.ds of only 0.30 Å, showing that the monomers are very similar. As the C-terminus was not included in the comparisons this is to be expected. The largest r.m.s.d of 1.09 Å is between the *Tth*CS A monomer and the *Li*CS B monomer. This is likely a result of the flexible loop being present in the *Li*CS monomer but absent in the *Tth*CS structure.

Overall, these structures provide a basis for further structural drug design work. Further the similarities in these structures suggest that effects on one enzyme can be carried over to the others.

Chapter 5: Examining the structure and function of the Cysteine Synthase Complex from *Trypanosoma cruzi*

Introduction

The presence of a cysteine synthase complex was first identified in *S. typhimurium* by ultracentrifugation and was shown to be able to carry out both steps of the *de novo* cysteine synthesis pathway (Kredich et al., 1969). Studies of this enzymatic complex have been conducted in *E. coli*, and in plants such as *A. thaliana*, soybeans and spinach, in addition to studies performed with *S. typhimurium* (Bogdanova et al., 1997; Droux et al., 1998; Kredich et al., 1969; Kumaran et al., 2009). Despite the range of studies into this enzymatic complex, it is poorly understood, and findings are varied.

The cysteine synthase complex in *E. coli* is suggested to be formed between a serine acetyltransferase as a dimer of homotrimers and two dimers of homodimeric cysteine synthase (Wang et al., 2012). This is further supported by SAXS data consistent with this model (Rosa et al., 2019). This complex has been shown to have three stable intermediate forms, with different isomerisation of SAT when the complex forms. The same study suggested that SAT interacts with CS in a non-allosteric manner involving the C-terminus of SAT (Wang et al., 2012).

In *S. typhimurium,* it is proposed that the CSC is composed of a 1:2 ratio of SAT to CS. This suggestion is based on the molecular weights obtained and calculations with the molecular weights from recently determined crystal structures suggest that it is likely that the complex forms from a hexamer of SAT and then two dimers of CS (Hara et al., 1990; Kredich et al., 1969; Kumaran, Yi, Krishnan, & Jez, 2009).

CSC in soybeans has been shown to differ from the suggestions for *E. coli* and *S. typhimurium* with the complex composing of a single SAT trimer with three CS dimers. This composition has been shown to have three distinct binding events, as each dimer of cysteine synthase associates with the serine acetyltransferase trimer (Kumaran et al., 2009).

Previous works have shown that CSC functions by altering the activities of each enzyme in the complex, rather than the more common protein complex mechanism of channelling metabolites. Formation of the CSC in soybeans results in a two-fold increase in SAT activity and in an inhibition of CS activity (Kumaran et al., 2009).

Currently, although the cysteine synthase complex has been identified through sequence comparison in *Leishmania major* and the enzyme complex purified, no studies of the structure or function of the cysteine synthase complex have been conducted in trypanosomatids (Williams et al., 2009).

Results

Protein Expression and Purification

Due to difficulties in stable and soluble expression and purification of *Tc*SAT alone, it was decided to coexpress the *Tc*SAT and *Tc*CS proteins in order to produce CSC. Initially co-expression through transformation with *Tc*SAT with competent *Tc*CS expressing *E. coli* cells was undertaken. This did not result in soluble *Tc*SAT production. In order to form the CSC, the genes encoding *Tc*CS and *TcSAT* were cloned into a pETDuet1 vector using the sites BamHI and HindIII for *Tc*CS and BlgII and BlpI for *Tc*SAT. Due to the design of this vector, *Tc*CS will be produced first then *Tc*SAT will follow. As *Tc*CS is very soluble, stable, and highly expressed, it could act as a chaperone to *Tc*SAT; the plasmid was constructed so that any *Tc*SAT produced would bind to available *Tc*CS. Additionally stop codons were inserted into the 3' terminus of the sequences to prevent any C-terminal additions. The plasmid map for this construct is shown in Figure 5-1.

Figure 5-1 - pETDuet1.*Tc*CS_*Tc*SAT vector map

pETDuet1.*Tc*CS_*Tc*SAT was expressed at 30°C overnight, purified and analysed by SDS-PAGE shown in Figure 5-2.

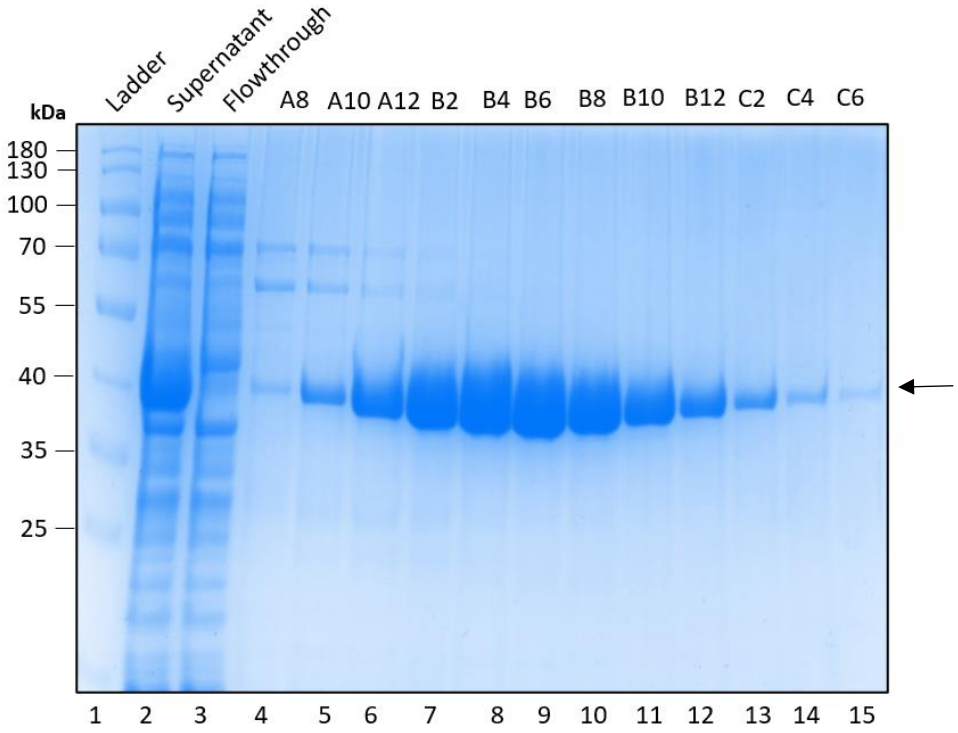


Figure 5-2 - SDS PAGE gel showing HisTrap purification of *Tc*CSC. Lane 1 shows protein ladder, lane 2 shows the supernatant, lane 3 shows the flow through of column, lanes 4 – 15 show elution of column. Arrow indicates expected migration of *Tc*CS and *Tc*SAT.

The prominent bands of protein around 40 kDa indicate the successful production of soluble protein. The monomeric forms of CS and SAT are both around 38 kDa so would be indistinguishable in the gel. These bands were analysed by ESI mass spectrometry, but protein identification was inconclusive for *Tc*SAT. The resultant protein was pooled and concentrated. Buffer exchange was performed, and the protein was loaded onto a Sephacryl 300 gel filtration column for further purification and to gain insight into the overall size of the complex and therefore the composition. This experiment will also show if any CSC has been successfully formed, as alone *Tc*CS is a dimer in solution.

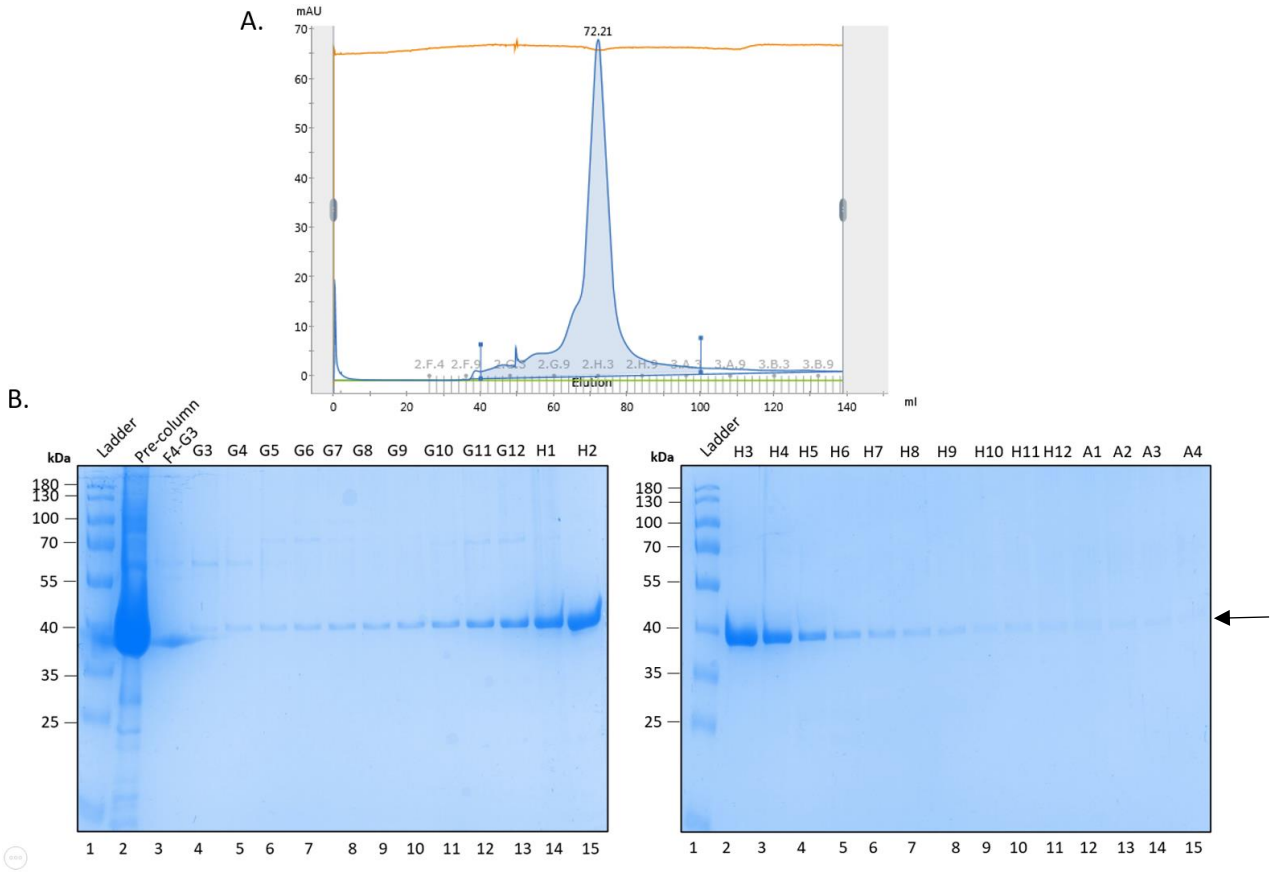


Figure 5-3- Separation of *Tc*CSC protein complex using a Sephacryl 300 gel filtration column. A - UV trace of gel filtration experiment. B - SDS PAGE of fractions collected from gel filtration. The black box indicates bands of interest at 70 kDa. For both gels, lane 1 shows the ladder. For the left gel, lane 2 shows concentrated protein sample before size exclusion was performed. lanes 3-15 and lanes 2 – 15 from the right gel, show elution from the size exclusion column. Arrow indicates expected migration of *Tc*CS and *Tc*SAT.

From the UV trace in Figure 5-3, it can be seen that a main peak eluting at 72 mL is present with some protein appearing to be constantly eluted throughout the experiment. The main peak can be seen in the UV trace can be attributed to *Tc*CS as a significant peak around 75 mL of elution would be expected due the separation range of a Sephacryl 300 gel filtration column. From the SDS-PAGE gels in Figure 5-3, it can be seen that the most prominent band is eluted throughout the gel filtration experiment. This band is indicative of *Tc*CS but the presence of *Tc*CS throughout

suggesting that *Tc*CSC has successfully been produced. If only *Tc*CS was expressed and purified then all the *Tc*CS would be eluted in lanes 13-15 of Figure 5-3B and lanes 1-2 of Figure 5-3C. Due to the constant elution of *Tc*CS and the presence of the *Tc*SAT bands it can be concluded that CSC has been produced. This experiment also suggests that although the complex may be initially bound together, as the gel filtration proceeds, the equilibrium of $\mathrm{SAT} + \mathrm{CS} \rightleftharpoons \mathrm{CSC}$ causes the complex to dissociate as the uncomplexed *Tc*CS and *Tc*SAT moves down the column to be eluted consistently.

Chemical cross-linking of the CSC

In order to prevent the disassociation of CSC during gel filtration, crosslinking with glutaraldehyde was performed. CSC was purified using IMAC, and resultant protein fractions pooled and dialysed. CSC was crosslinked to form covalent bonds between the CSC components. This allows for the separation of the CSC from uncomplexed SAT and CS whilst maintaining the complex. After crosslinking, gel filtration was performed, as shown in Figure 5-4.

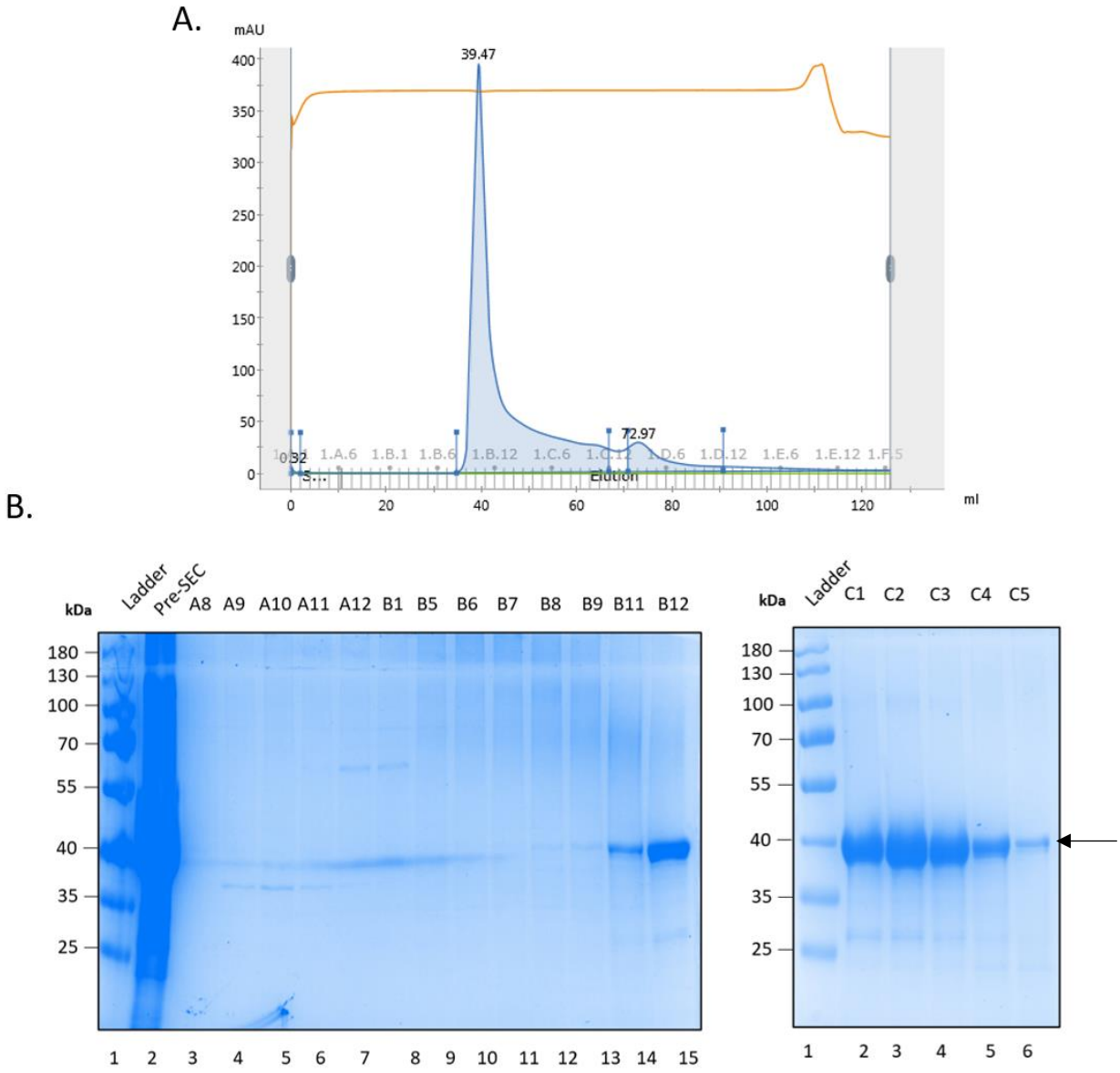


Figure 5-4- Separation of cross linked *Tc*CSC using Sephacryl 300 gel filtration column. A - UV trace from gel filtration B - SDS PAGE showing protein fraction from gel filtration. For both gels lane 1 shows the ladder. For the left gel, lane 2 shows concentrated protein sample before size exclusion was performed. lanes 3-15 and lanes 2 – 6 from the right gel, show elution from the size exclusion column. Arrow indicates expected migration of *Tc*CS and *Tc*SAT.

The UV trace shown in Figure 5-4, shows that cross linking CSC has caused the main elution peak to shift right dramatically. The main peak is now eluted at 40 mL, which is the void volume of a Sephacryl 300 gel filtration column and suggested the protein eluted is larger than the resolution range of the column and a gel filtration column with a higher molecular weight separation range should be used. There is a

small peak remaining at 72 mL showing that some *Tc*CS remains unlinked and is separated from the complex. The SDS-PAGE gels in Figure 5-4B show bands corresponding to the size of *Tc*CS and *Tc*SAT can be seen in lanes 2-4 of Figure 5-4B right. These samples were taken from the main peak and suggest that crosslinking was successful and maintained the integrity of the complex.

Transmission Electron Microscopy of CSC

To determine structural information on the CSC, negative staining was conducted, and transition electron microscopy was performed. The images obtained are shown in Figure 5-5 and in appendix, Figure S9-5-1. Images A-C show aggregation of protein and an excess of sample on the grid. Images D-F show a lower concentration of protein over the grid, with less aggregation and the protein in a range of orientations. As the protein seems prone to aggregation, the conditions used for negative staining and TEM could be optimised. From these images it is clear that protein samples are present and visible through negative staining.

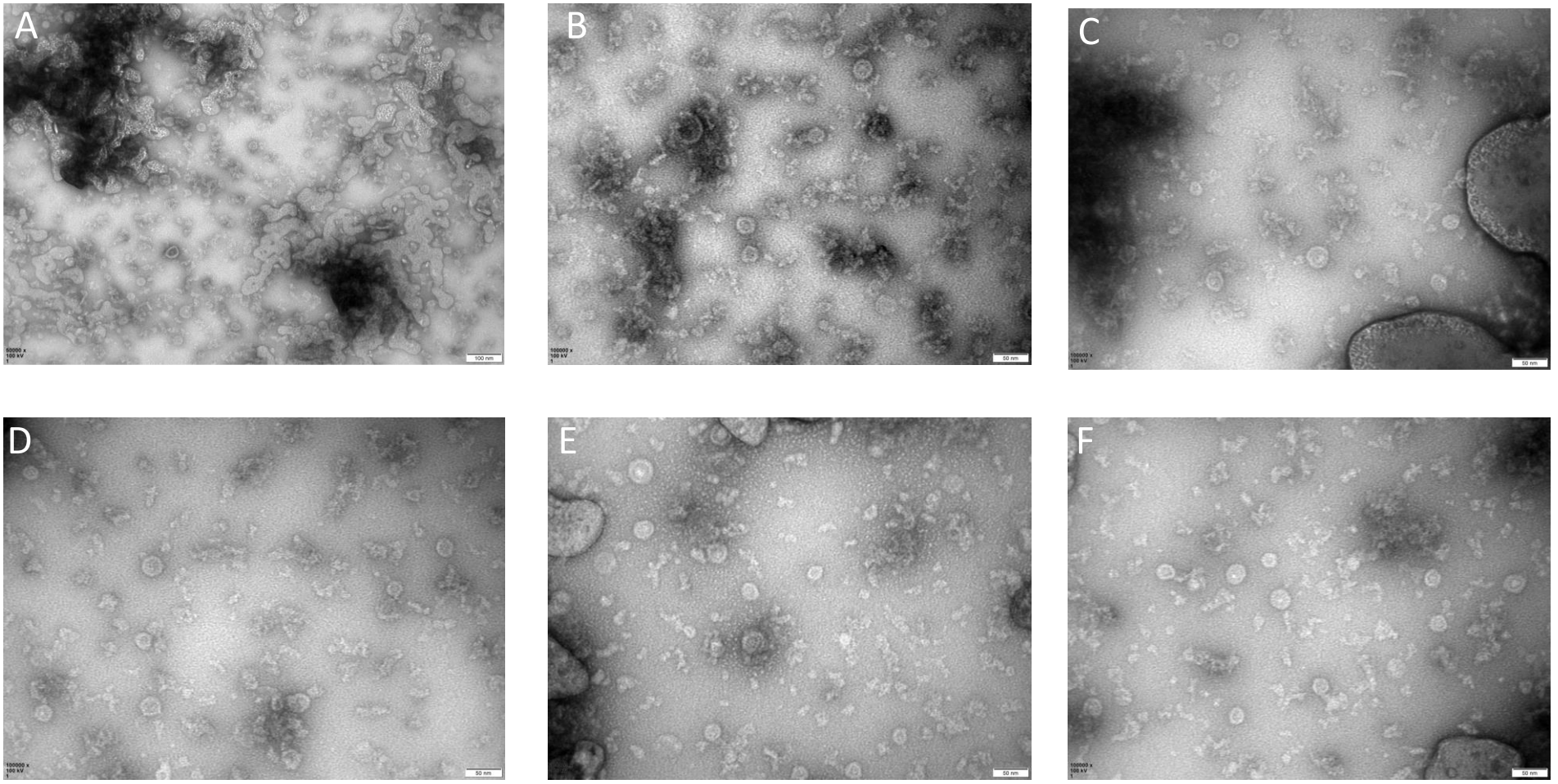


Figure 5-5- Sample of Transmission Electron Microscopy images of *Tc*CSC after negative staining. *Tc*CSC stained with 1% uranyl acetate and imaged at 100 kV. Scale shown for reference.

Upon close inspection of the images in Figure 5-5, several interesting particles can be seen, Figure 5-6. The left and right images appear to be mostly linear with larger protrusions at the top and bottom. The central image shows three projections from the central circle.

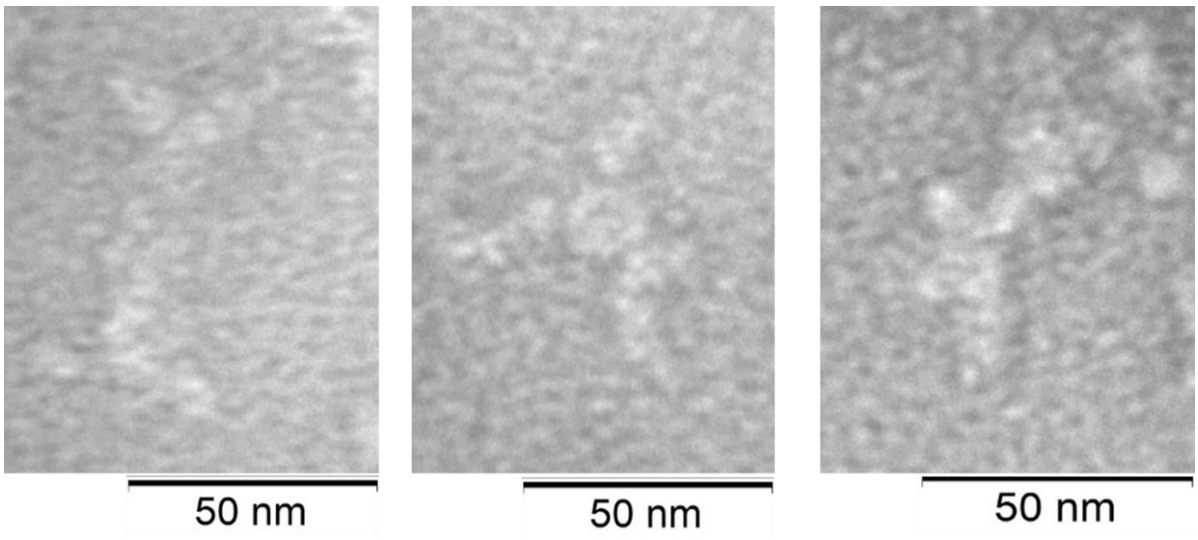


Figure 5-6 - Close ups of selected particles from images of *Tc*CSC negative staining, Figure 5-5.

These TEM images show the presence of single particles from CSC crosslinking experiments. Analysis of the particles shows that although there is aggregation present in the sample, there are also particles adopting different conformations. In particular, the centre image could be a SAT trimer surrounded by three CS dimers supporting the hypothesis suggested from soybean CSC. This suggests that further work on the structure of CSC is worthy of pursuit.

Biochemical characterisation of *Tc*CSC

To test if the *Tc*CSC was reconstituted correctly and capable of completing the *de novo* cysteine synthesis pathway, the activities of *Tc*CSC were tested. Firstly the conversion of serine into OAS, and then OAS into cysteine was tested.

Production of OAS from Serine

To determine SAT activity in the CSC, the amount of OAS produced from varying serine conditions was determined through the amount of CoA produced. CoA production was measured using a DTNB based assay (Riddles et al., 1983). *Tc*CS was confirmed to have no SAT activity and included as a negative control. As shown in Figure 5-7, CSC is active and capable of producing OAS indicating the successful production of the *Tc*SAT component of the complex.

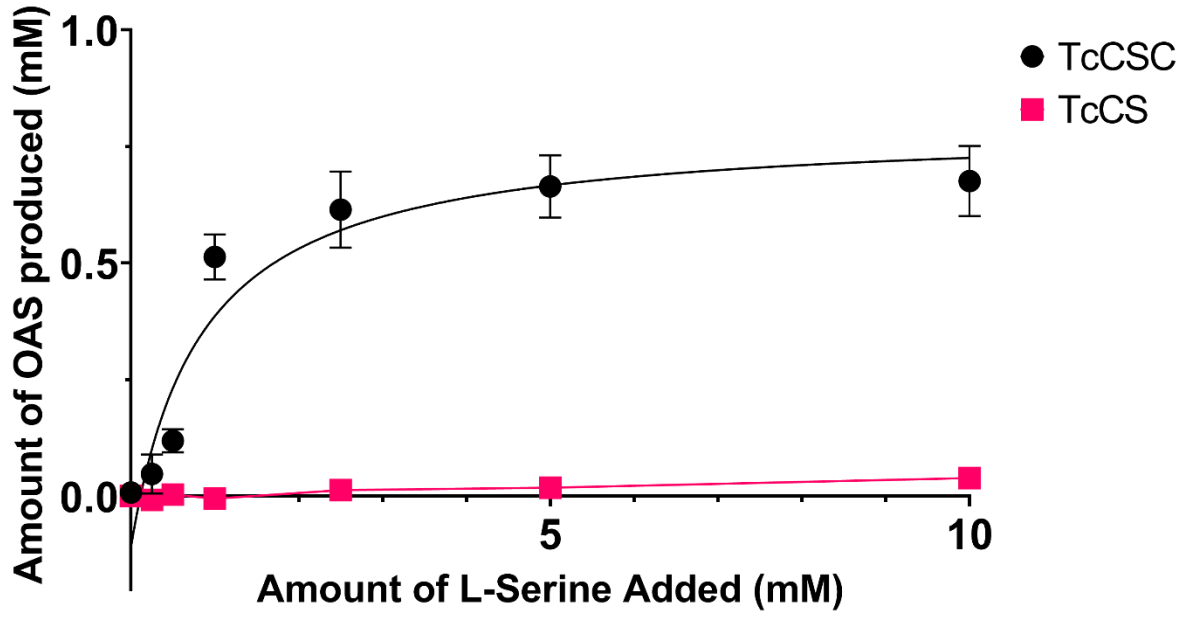


Figure 5-7 – Amount of OAS produced by *Tc*CSC (in black) and *Tc*CS (in pink) with increasing serine concentration.

Production of cysteine from OAS

To confirm CS activity in CSC, the amount of cysteine produced from varying OAS concentrations were measured using ninhydrin and compared to *Tc*CS. *Tc*CS was included as a positive control as previous studies have confirmed activity of *Tc*CS (Sowerby et al., 2023). As shown in Figure 5-8, CSC is active and capable of producing cysteine at a comparable rate to *Tc*CS indicating that CS activity is maintained in the complex.

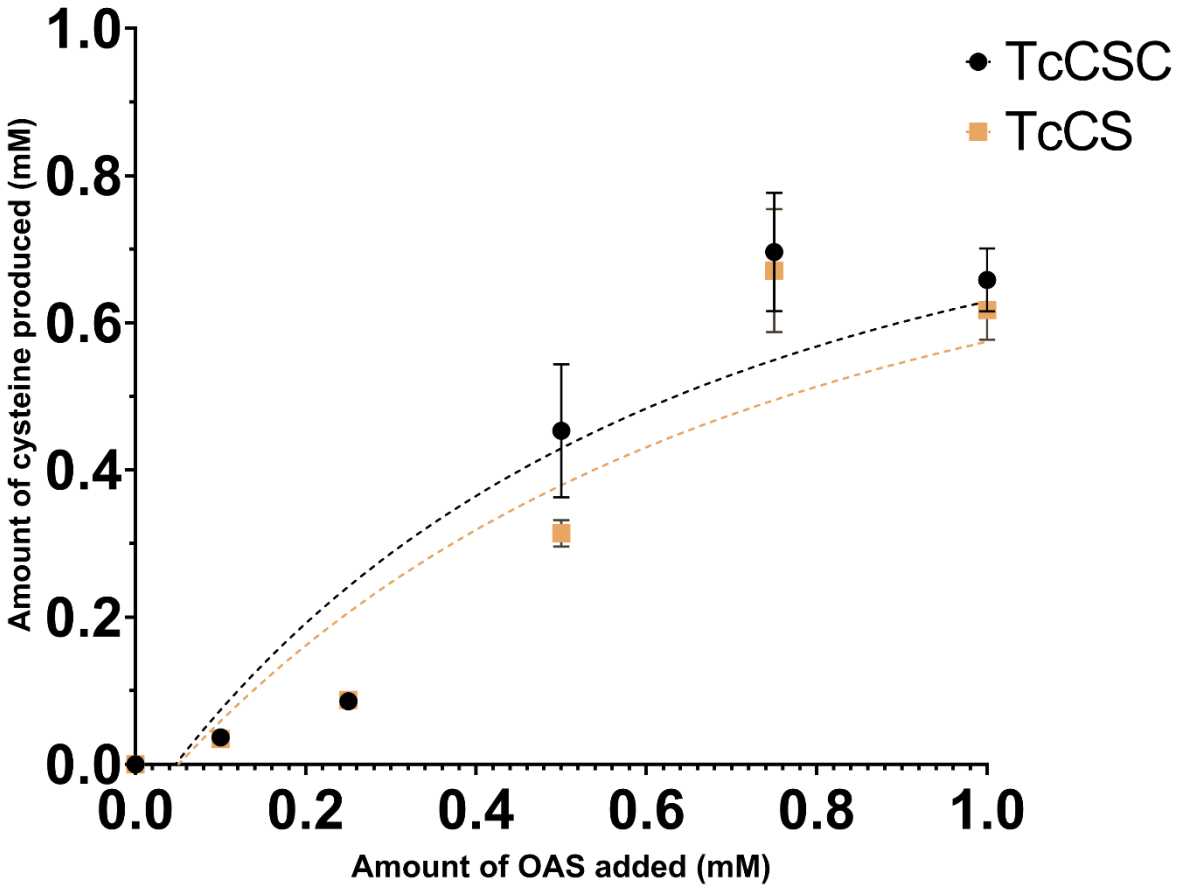


Figure 5-8 – Amount of cysteine produced with increasing concentrations of OAS by *Tc*CS (gold) and *Tc*CSC (black).

Production of Cysteine from Serine

After confirming activity of CSC at each individual step, the ability of CSC to produce cysteine from serine was assessed. As shown in Figure 5-9, CSC is active and is capable of conducting the entire *de novo* cysteine synthesis pathway. *Tc*CS was included as a negative control although at high concentrations of L-serine it appears that *Tc*CS has some activity, this is likely to be an artifact of the assay used.

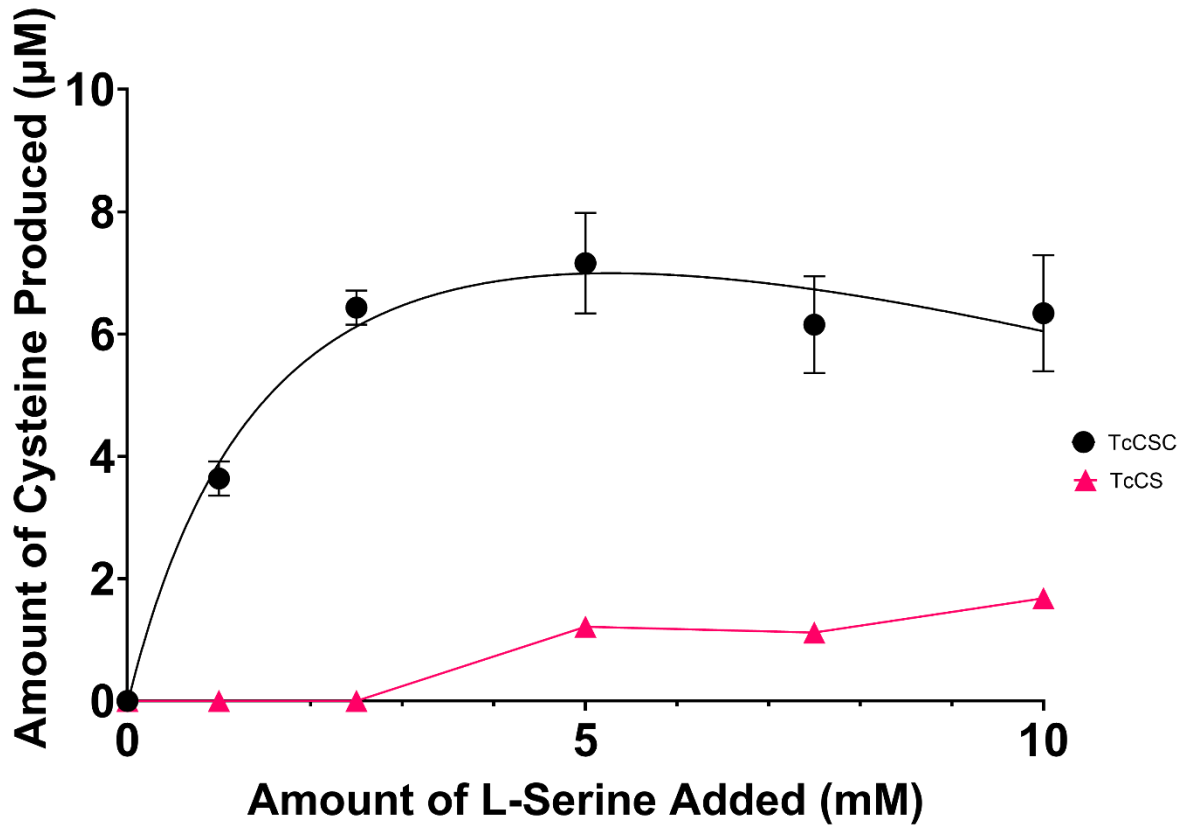


Figure 5-9 – Amount of cysteine produced with increasing serine concentration for *Tc*CSC (black) and *Tc*CS (pink).

Inhibition of TcCSC by cysteine

The activity of CSC to convert L-serine to OAS was studied in the presence of increasing cysteine concentrations. The change in absorbance of DTNB and

therefore the amount of CoA produced was measured after 30 mins and after 1 hour, shown in Figure 5-10. After 30 minutes, there is increases in absorbance at lower concentrations of cysteine, suggesting that at higher concentrations the cysteine is inhibiting activity. As low as 100 µM cysteine shows a significant effect at 30 mins. These results are consistent after 1 hour.

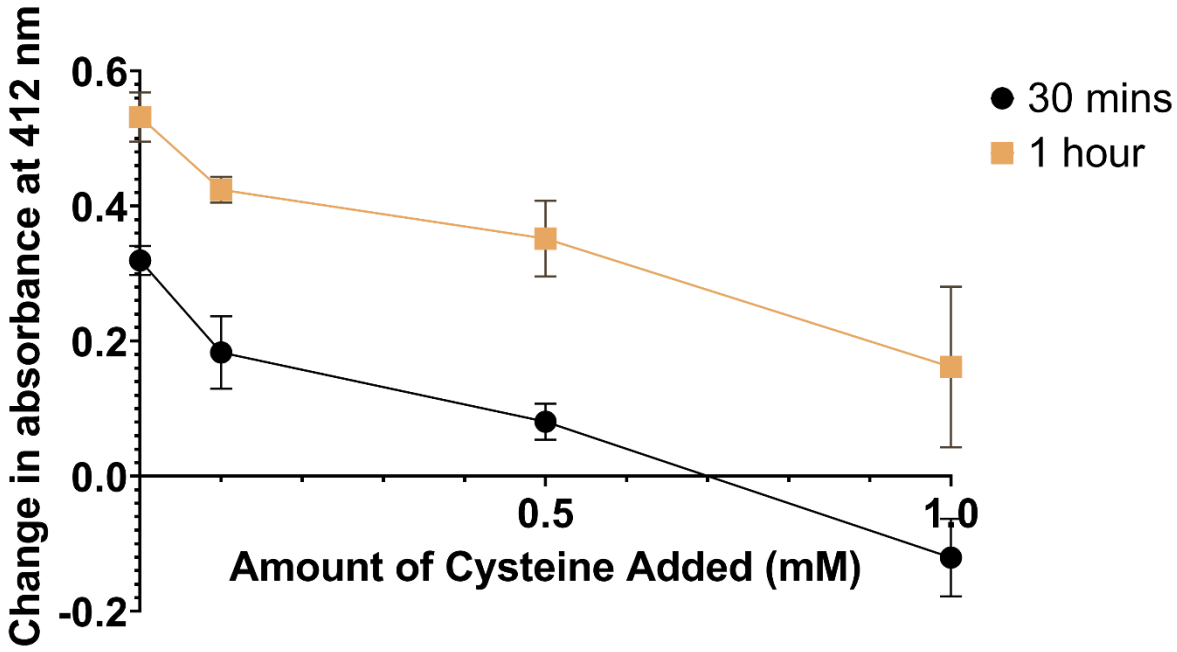


Figure 5-10 – Change in absorbance of DTNB at 412 nm with increasing amounts of cysteine added to *Tc*CSC after 30 mins (black) and after 1 hour (gold).

Overall, an active CSC complex has been produced and is inhibited by cysteine supporting the hypothesis of that the CSC activity is regulated by a negative feedback loop.

Conclusions

For the first time, expression of the *Trypanosoma cruzi* cysteine synthase complex was successfully established in *E. coli* through the use of a pETDuet vector. After cross linking, this protein complex was used for initial structural studies into this complex through negative staining Electron Microscopy. The images generated show particles in a range of orientations and this works lays the foundation for cryo-electron microscopy studies needed to obtain a high resolution structure of this complex.

Further, the CSC produced is active and capable of completing the entire *de novo* cysteine biosynthesis pathway. Each stage of the *de novo* cysteine synthesis pathway was studied for the CSC before assays into the entire pathway were conducted. These activity studies successfully show that the cysteine synthase complex is capable of producing both OAS and cysteine from the constituents of the pathway.

Additionally, CSC is indicated to be involved in larger thiol control within the cell and to also to be controlled by negative feedback. In this work, it has been shown that CSC from *T. cruzi* is inhibited by cysteine, suggesting that cysteine biosynthesis in *T. cruzi* and trypanosomes largely is controlled by a feedback mechanism.

When compared to CSC from other organisms, very little is known of the overall CSC structure as no transition electron microscopy or cryo-electron microscopy studies have been successfully conducted.

Contrastingly to *Tc*CSC, soybean CSC has been shown to reduce CS activity compared to CS alone (Kumaran et al., 2009). *Tc*CSC maintained the same level of CS

activity as *Tc*CS suggesting either a loss of CSC inhibition or a constant dependence on cysteine production in *T. cruzi*.

In summation, an active and functional CSC from *T. cruzi* has been expressed and purified for the first time. Assays to determine the function of this complex have been established with initial insights to support a negative feedback loop being obtained. Investigations into the structure of this bienzyme protein complex has been established.

Chapter 6: Using a structure based drug design approach to create an inhibitor for cysteine synthase.

Introduction to structure based drug design

Structure based drug design (SBDD) provides the most powerful and efficient process of drug discovery. There are usually four phases of drug developement: discovery, development, clinical trials, and registry. SBDD has advantages over traditional drug discovery as the process is a more efficient, rapid and specific (Batool et al., 2019). SBDD requires the structure of a potential target to be determined, typically through X-ray crystallography, NMR or homology modelling. Structures determined by X-ray crystallography are more commonly used due to advantages such as high resolution information, visibility of ordered water molecules within the structure and the size of structures available range from only a few amino acids to nearly 1 MDa (Anderson, 2003).

Following the determination of a structure, further experiments are performed to identify small molecules that bind to the protein. This is typically performed through molecular docking or *de novo* design (Batool et al., 2019). Molecular docking can be used as a high throughput method that can be used to evaluate large libraries to identify molecules that are likely to bind to the protein which can then be validated through *in vitro* assays (Śledź et al., 2018). *De novo* drug design refers to the production of new compounds from small molecules - typically, these small molecules binding to the protein are identified and then developed (Batool et al., 2019). One method of *de novo* drug design that has been developed is fragment-based drug design. This involves discovery of small molecules that typically bind the

protein weakly and then developing these molecules into more potent binders through optimisation (Li, 2020).

A range of techniques can be employed to determine binding of a molecule to protein and in this work exploitation of molecules shown to inhibit other cysteine synthases, NMR, and fragment-based drug discovery will be explored further.

Molecules inhibiting homologous cysteine synthases

Based on previous work on the cysteine synthase complex, interaction of the C-terminus of SAT has been shown to decrease the activity of CS in *E. coli* (Droux et al., 1998). Exploiting this has led to the development of small peptides based on the C-terminus of corresponding SAT sequences. These peptides range from 4-mers to 10-mers. One study from *L. donovani* shows that the cysteine synthase interacts with peptides designed from the C-terminal *Ld*SAT sequence (Raj et al., 2012).

Peptide-protein inhibition studies with *Hi*SAT based peptides and *Hi*CS revealed that the final five amino acids were responsible for SAT/CS interaction in *H. influenzae*. This resulted in attempts to identify CS inhibitors using pentapeptides based on the *Hi*SAT sequence (Salsi et al., 2010). The most promising of these results were developed into small molecules using *in silico* techniques (Pieroni et al., 2016). A structure of the most promising of these molecules, UPAR415 bound to *St*CS was determined, Figure 6-1.

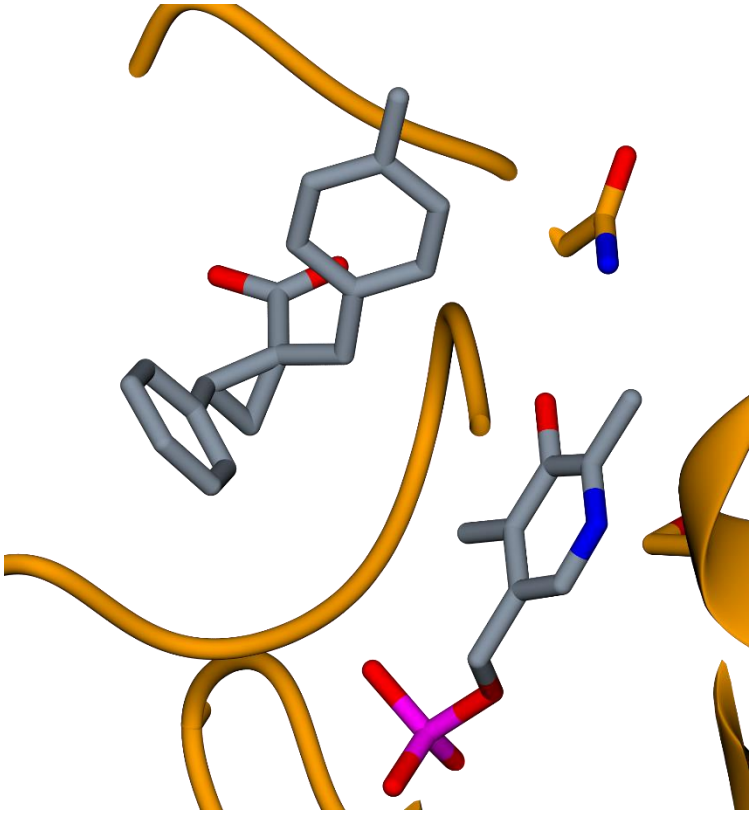


Figure 6-1 - UPAR415 in complex with *Hi*CS (PDB: 6Z4N) (Annunziato et al., 2021).

UPAR415 was shown to have a K_d in the nanomolar range and in combination with colistin showed inhibitory activity against several pathogens including methicillin resistant *S. aureus*, *E. coli* and *K. pneumoniae* (Annunziato et al., 2021).

^{19}F NMR as a hit discovery tool

NMR provides a valuable tool for screening the binding of protein and ligands, typically requiring small amounts of unlabelled protein (Pellecchia et al., 2002). Using fluorinated ligands, ^{19}F NMR can be performed and this technique provides an approach for fragment-based drug design (FBDD) (Norton et al., 2016).

^{19}F is a naturally 100% abundant isotope with a spin of ½ and typically is not found in proteins or the buffered solutions used with them (Dalvit et al., 2019). In general, molecules with a single fluorine atom or with equivalent fluorine atoms are used in these studies. This results in a single fluorine signal from the ligand, and this allows for screening of chemical mixtures as the analysis is simplified. Fluorine has large chemical shift anisotropy which makes the difference in bound and unbound state signals very large, especially when high magnetic fields are used as this increases the signal to noise ratio (Dalvit et al., 2003; Yu et al., 2013). This enables weak-affinity molecules to be tested. Using high magnetic fields lead to very broad line widths for ligand-protein complexes which makes the use of fluorine direct detection methods impractical (Dalvit et al., 2003).

This is overcome by using a Carr-Purcell-Meiboom-Gill (CPMG) pulse sequence. This pulse sequence produces an echo which decays exponentially as a function of T_2 relaxation time. T_2 relaxation time is inversely related to the molecular weight and magnetic field strength of a molecule so when a ligand is bound to a protein, the T_2 relaxation time is shorter than it would be for the ligand alone (Buchholza et al., 2021). This causes a broader signal to be produced when the ligand and protein are interacting as seen in Figure 6-2.

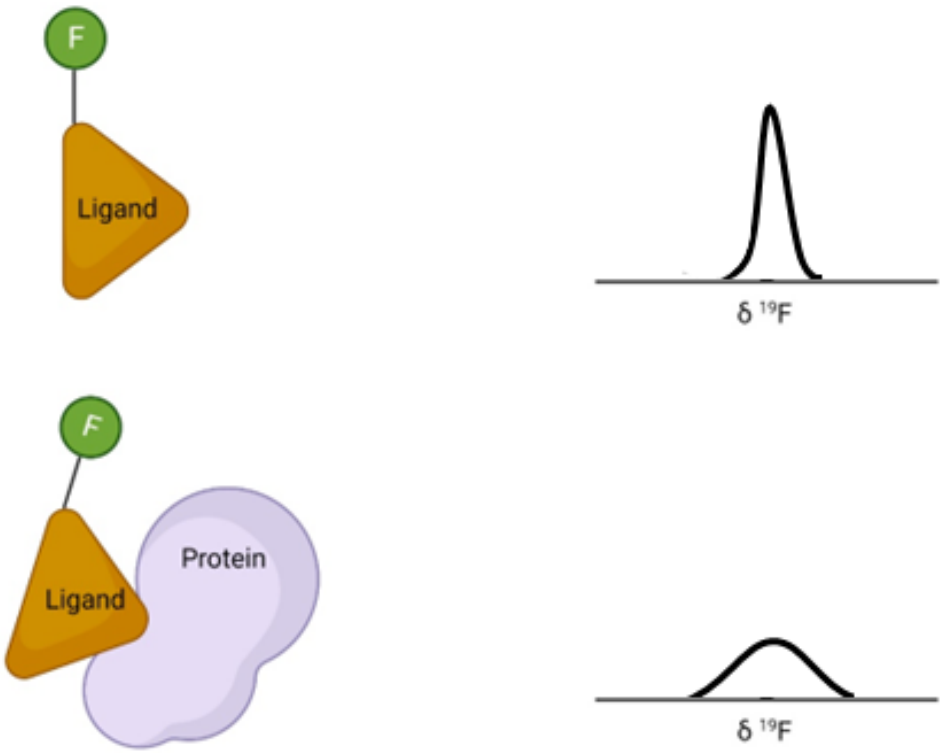


Figure 6-2 – Effect on ^{19}F NMR signal when ligand is free vs bound to the protein. Produced using information from (Dalvit et al., 2012).

In addition to the expected signal broadening, occasionally a change in the chemical shift can occur with protein-ligand interaction (Dalvit et al., 2019). Overall, line broadening is a more accurate indicator of interaction that chemical shift (Tengel et al., 2004).

Fragment-Based Drug Design

FBDD is based on the idea of small molecules – fragments – that bind to a specific target to identify binding sites on a protein. These fragments then provide a starting point for the generation of drug leads (Bollag et al., 2012; Perera et al., 2017; Shuker et al., 1996). The process typically develops from the screening of a small library of low molecular weight compounds that are large enough to provide interactions with the protein of interest whilst being small enough to limit unfavourable interactions such as steric clashes (Kirsch et al., 2019). One such library is the Fraglite library which has the added advantage of each fragment possessing a heavy halogen atom (Wood et al., 2019). The incorporation of halogens results in anomalous scattering which allows the unambiguous detection of the atom and can lead to the rapid identification of hot spots for binding (Bauman et al., 2015). Brominated fragments have also led to increased hit rates such as 24% compared to an overall rate of 4% (Bauman et al., 2012).

Typically, these fragments are soaked into previously grown crystals, diffraction data collected, and hits identified (Chilingaryan et al., 2012). Following fragment hit identification, fragment to lead optimisation is performed. The three main methods of optimisation are linking, merging and growing (Kirsch et al., 2019). Fragment linking involves the connecting of two fragments that bind at distinct locations in the protein. Successful linking involves one fragment interacting through polar interactions and the other via van der waals interactions. This approach is challenging as typically fragments do not bind to the protein in suitable positions for linking (Bancet et al., 2020). Fragment merging uses two or more fragments binding to the same location and through medicinal chemistry combines them together

(Hudson et al., 2012). Fragment growing is the most widely used method of hit optimisation (Li, 2020). This process aims to increase the size of the molecule by attaching various additional functional groups (Kirsch et al., 2019).

The previously described recombinant protein *Tc*CS and the structures of *Tc*CS and *Li*CS will be used to explore different methods of structural based drug design.

Results

Exploiting naturally binding molecules to investigate *Tc*CS inhibitors

Initial binding experiments

To be able to determine if molecules are binding to the protein, a binding assay is required. An assay using the fluorescence of PLP was designed based on a ligand binding to PLP in the active site results in the emission intensity altering. When excited at 412 nm with, the PLP emission maximum occurs at 490 nm (Cook et al., 1996). If binding occurs, the intensity of the emission will increase at the maximum except in the case of OAS, where the emission maximum will decrease in fluorescence intensity (Banerjee et al., 2011). To test binding of molecules to *Tc*CS, cysteine, serine, OAS and isoleucine were tested, Figure 6-2.

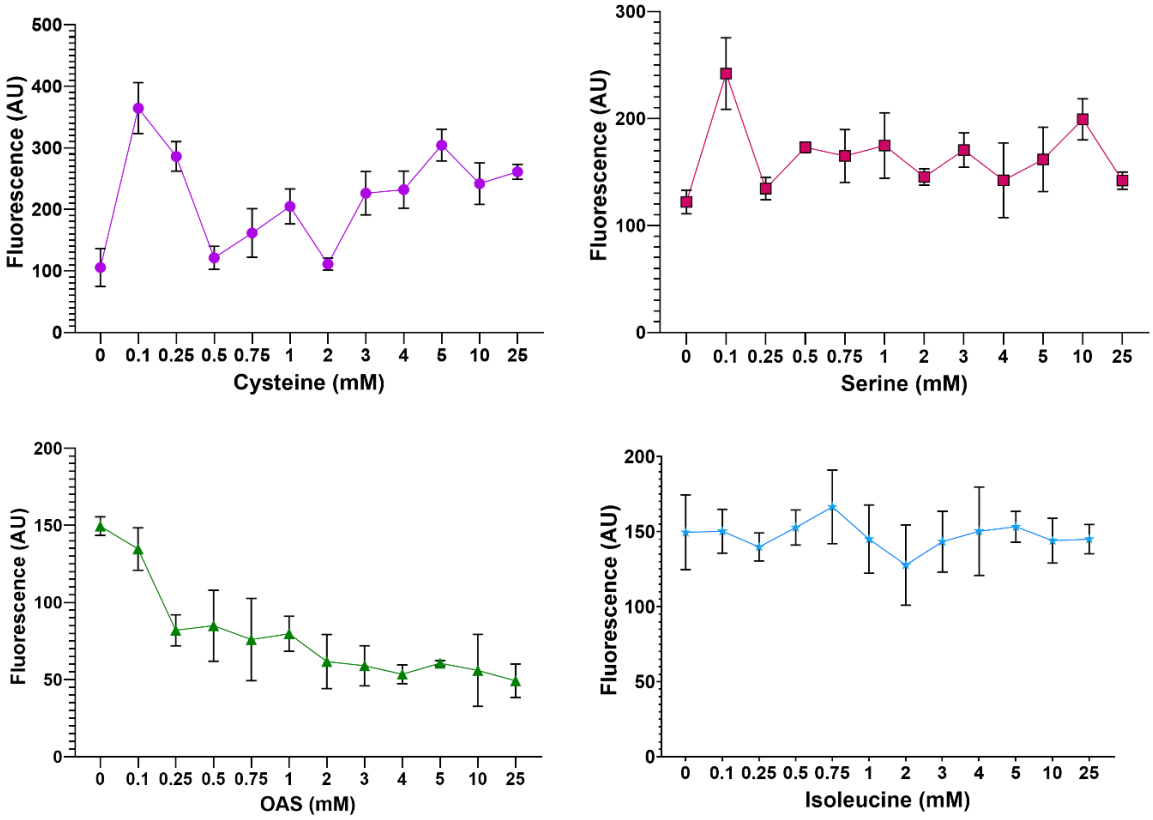


Figure 6-2 - Fluorescence at 490 nm of *Tc*CS excited at 412 nm with cysteine, serine, OAS and isoleucine. Standard error was calculated with n=3, and a minimum of 2 data points used per measurement.

For both cysteine and serine, as the concentration of amino acid increases the general trend of the fluorescence increases however the significance of these results are dubious due to the large range of the results. Nevertheless, this suggests that both molecules may be binding to *Tc*CS around the PLP. For OAS, a significant decrease in fluorescence intensity can be seen with the addition of as little as 0.25 mM OAS. Additions of 0.25 mM to 25 mM OAS has decreased the fluorescence intensity even further. This shows that OAS is binding to *Tc*CS in the active site as expected. As isoleucine concentration increases the fluorescence intensity remains constant, this indicates that the isoleucine is not binding to *Tc*CS in the active site. This establishes OAS as a positive control for active site binding with the potential to

use cysteine and serine as well. The most significant change in fluorescence is seen with OAS, however due to OAS resulting in a decrease in fluorescence intensity and the ready availability of cysteine and serine, cysteine or serine will be used as positive controls. Isoleucine has been established as a negative control for binding in the active site as the fluorescence intensity does not change significantly.

To determine if the molecules influence the protein activity and therefore are capable of inhibition, the established *Tc*CS activity assay was used (Sowerby et al., 2023).

Using custom designed C-terminal SAT peptides to determine the effect on *Tc*CS

Previous work with CS shows an interaction with C-terminal peptides from SAT and typically uses peptides of 10 amino acids in length (Huang et al., 2005). To determine the effect of SAT C-terminal peptides on *Tc*CS, 96 peptides were designed. It was decided to use homologous sequences of SAT to design peptides based on the *Tc*SAT C-terminal sequence with mutations based on the different residues present in the other sequences. Sequence alignments with *Tc*SAT were conducted and from the resulting sequences, four sequences were selected based on highest sequence identity to *Tc*SAT. A further four SAT sequences were selected by highest sequence identity between TcSAT and homologues with known crystallographic structures of SAT, Figure 6-3.

When designing peptides, it was decided to use a length between 10 and 15 amino acids so that the peptides were large enough to conserve any structural elements, such as an α-helix, that may be integral to their function but also small enough for

easier synthesis. In total, 112 peptides were designed and then 96 were selected for production based on ease of synthesis.

The first 5 peptides were designed using the original *Tc*SAT C-terminal tail sequence to mimic the expected biological response with length increasing by one residue to 15 residues, Figure 6-3. The next 73 peptides were designed by making single mutations to the *Tc*SAT sequence and increasing peptide length from 10 amino acids to 15. The mutations introduced to the TcSAT C-terminal sequence were based on a sequence alignment, Figure 6-3.

8 peptides were designed based on SAT sequences with high sequence identity with lengths of 10 and 15 amino acids without mutations. 10 peptides were designed based on sequences with high sequence identity to *Tc*SAT from available crystallographic structures with lengths of 10 and 15 amino acids without mutations. Briefly, to allow for easier synthesis changes to the sequences such as the exchange of initial residues was conducted. Included in these alterations is the modification of isoleucine to methionine, glutamic acid to glutamine or valine and valine to serine. Finally, the peptide order was randomised before synthesis and are listed in the appendix, Table S9-6-1.

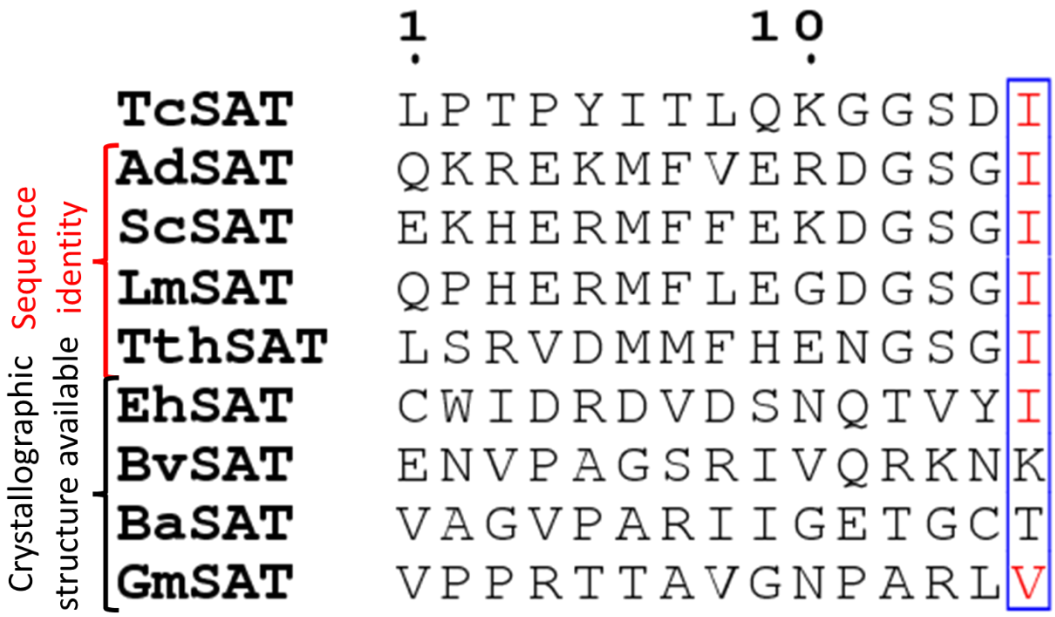


Figure 6-3 - Sequence alignment of *Tc*SAT C-terminal 15 amino acids with other homologous SAT.

Based on synthesis criteria provided by Belytics, 96 peptides were synthesized by Belyntic GmbH in plate format. The purified peptides were analysed, lyophilized, and weighed into separate vials. These peptides were dissolved in water and made up to a concentration of either 5 mM or 500 mM dependant on the amount of peptide synthesised. The final 96 peptides produced are shown in the appendix, Table S9-6-1. Not all peptides were successfully synthesised, peptide 1F failed to be produced and 10 peptides (1A, 1B, 2H, 6D, 9D, 9E, 11E, 11G and 12A) were generated in small amounts (<2 mg).

Determining peptide binding to *Tc*CS

To determine if the peptides are binding in the active site of the protein and therefore interacting with PLP, fluorescence assays were performed. As seen previously, interaction between PLP and additional molecules, such as OAS or

cysteine, causes the fluorescence of the PLP to be altered when excited at 412 nm. A fluorescence spectra, shown in Figure 6-4, was measured between 450 nm and 600 nm for all peptides. As the fluorescence maximum for PLP is 490 nm when excited at 412 nm, shown in Figure 6-5, decisions on the peptides will be based upon the 490 nm readings.

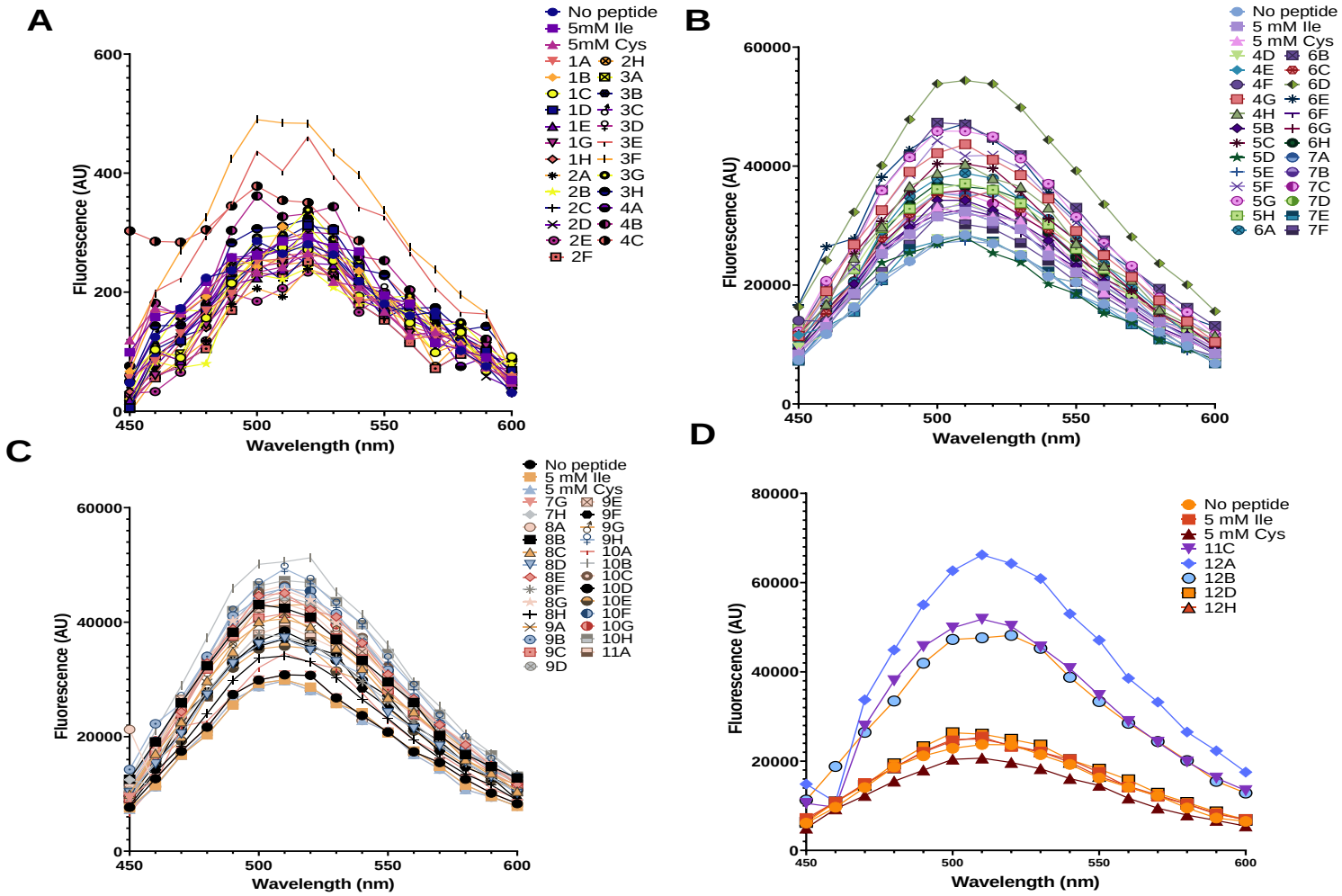


Figure 6-4 - Fluorescence spectra of *Tc*CS when excited at 412 nm with custom SAT peptides A – peptides 1A to 4C, B – peptides 4D to 7F, C – peptides 7G to 11A, D – peptides 11C, 12A, 12B, 12D and 12H.

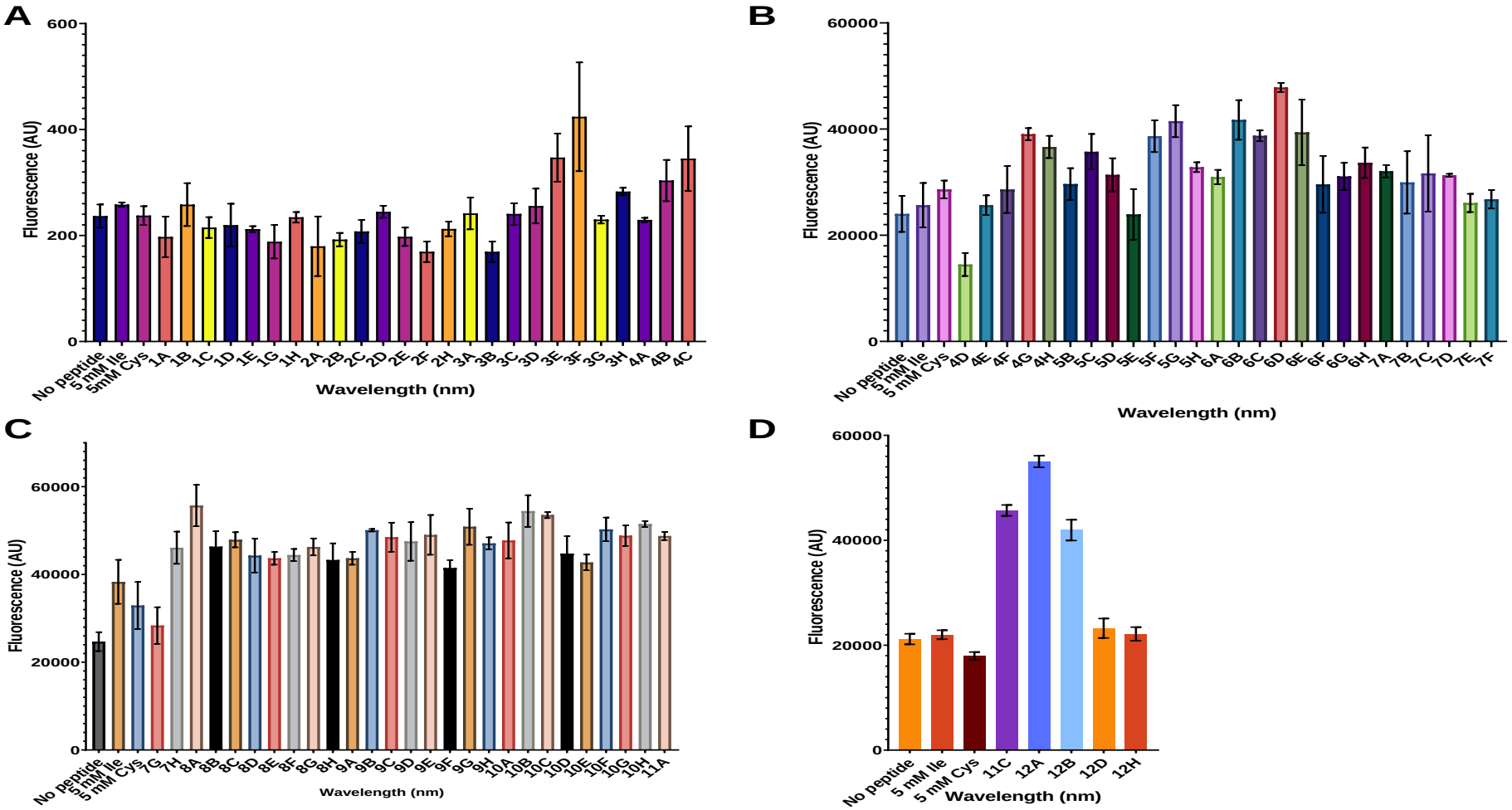


Figure 6-5 - Fluorescence of *Tc*CS at 490 nm when excited at 412 nm with custom SAT peptides A – peptides 1A to 4C, B – peptides 4D to 7F, C – peptides 7G to 11A, D – peptides 11C, 12A, 12B, 12D and 12H. Standard error was calculated with n=3, and a minimum of 2 data points used per measurement.

Analysis of these results when compared to the readings of *Tc*CS without peptide reveal that there are 52 peptides that show a significant change in fluorescence intensity of *Tc*CS, listed in the appendix, Table S9-6-2. Generally, an increase in fluorescence is indicative of binding, however as in the case of OAS, a decrease in fluorescence can also occur upon binding. These changes indicate that these peptides are binding at the active site of *Tc*CS. Of these 52 peptides, 48 are based on the *Tc*SAT sequence with mutations present, with all lengths of peptides present. The other 4 peptides are from proteins with high sequence similarity.

Determining the effect of peptides on the activity of *Tc*CS

Based on literature involving *Eh*CS, it is expected that the SAT C-terminal based peptides will result in inhibition of cysteine synthase (Kumar et al., 2011). The activity assay for measuring cysteine synthase was adapted into a high throughput 96 well plate method (Gaitonde, 1967; Sowerby et al., 2023). The absorbance of ninhydrin at 560 nm was measured and converted to amount of cysteine produced.

Figure 6-6 shows the effect on the amount of cysteine produced with the custom SAT peptides.

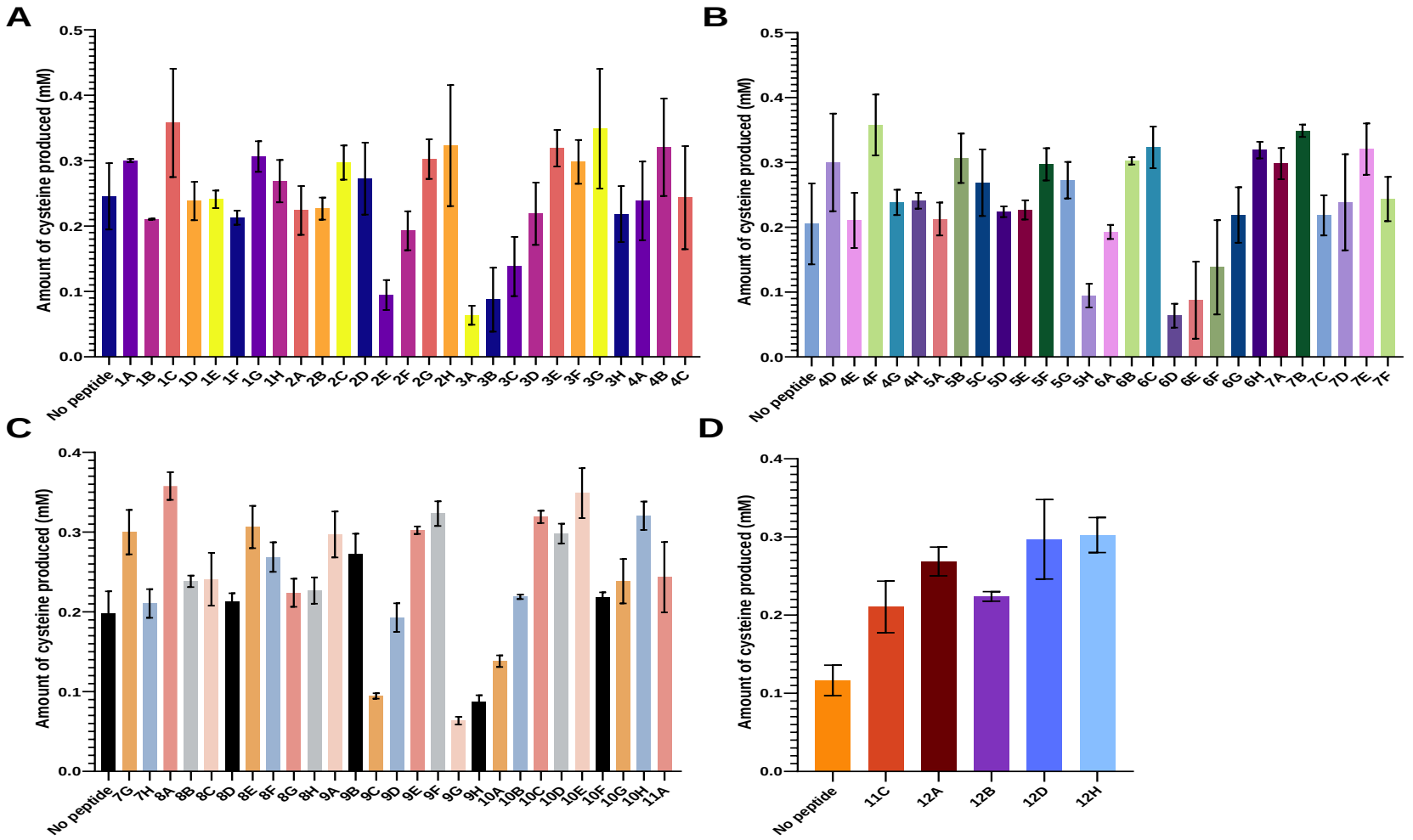


Figure 6-6 - Activity of *Tc*CS with custom SAT peptides A – peptides 1A to 4C, B – peptides 4D to 7F, C – peptides 7G to 11A, D – peptides 11C, 12A, 12B, 12D and 12H. Standard error was calculated with n=3, and a minimum of 2 data points used per measurement.

There are 36 peptides that show a significant effect upon the amount of cysteine produced, listed in the appendix, Table S9-6-2. All peptide lengths from 10 to 15 amino acids are included. 31 of the peptides are based on the *Tc*SAT sequence with mutations. The remaining 5 peptides are from the C-terminal sequences of homologous SATs.

Of the 36 peptides 10 were shown to decrease activity, listed in the appendix, Table S9-6-2. These peptides are all based on the *Tc*SAT sequence with single mutations and range in lengths between 10 and 15 amino acids. The remaining 26 peptides were shown to increase the activity of *Tc*CS. Most of these peptides are also based on the *Tc*SAT sequence with mutations and a range of lengths, however some of the peptides based on the sequences of other organisms were shown to increase activity.

Overall, just over a third of peptides based on the C-terminal sequence of *Tc*SAT and of homologues show a significant effect on the activity of cysteine synthase.

Soaking of peptides into *Tc*CS crystals.

Based on the binding and activity results of the peptides with *Tc*CS, several peptides with significant results were selected for soaking into *Tc*CS crystals, shown in Table 6-1. The peptides selected were based on significantly affecting the activity or fluorescence intensity with *Tc*CS. It was decided to soak peptides into established *Tc*CS crystals in order to determine where the peptides affect the *Tc*CS as some peptides shown to affect activity were not shown to bind in the active site, suggesting an additional binding location.

Table 6-1 - Peptides selected for soaking to *Tc*CS crystals. Peptides in blue have been shown to decrease activity while peptides in black result in an increase in *Tc*CS activity. Peptides in green were found to bind to the active site of *Tc*CS.

2E	TcSAT10_5	ITL**H**KGGSDI
3A	TcSAT10_9	ITLQ**G**GGSDI
3B	TcSAT10_10	ITLQK**N**GSDI
3C	TcSAT10_11	ITLQK**D**GSDI
5H	TcSAT12_8	PYITLQ**R**GGSDI
6D	TcSAT12_12	PYITLQKGGS**G**I
6E	TcSAT13_1	TPYI**M**LQKGGSDI
6H	TcSAT13_4	TPYIT**V**QKGGSDI
8A	TcSAT14_1	PTPYI**M**LQKGGSDI
8E	TcSAT14_5	PTPYITL**H**KGGSDI
9A	TcSAT14_9	PTPYITLQ**G**GGSDI
9B	TcSAT14_10	PTPYITLQK**N**GSDI
9C	TcSAT14_11	PTPYITLQK**D**GSDI
9D	TcSAT14_12	PTPYITLQKGGS**G**I
9E	TcSAT15_1	LPTPYI**M**LQKGGSDI
9F	TcSAT15_2	LPTPYI**F**LQKGGSDI
9G	TcSAT15_3	LPTPYIT**F**QKGGSDI
9H	TcSAT15_4	LPTPYIT**V**QKGGSDI
10A	TcSAT15_5	LPTPYITL**H**KGGSDI
10C	TcSAT15_7	LPTPYITLQ**E**GGSDI
10D	TcSAT15_8	LPTPYITLQ**R**GGSDI
10E	TcSAT15_9	LPTPYITLQ**G**GGSDI
10H	TcSAT15_12	LPTPYITLQKGGS**G**I
11C	LmSAT_10_2	**I**FLEGDGSGI
12A	ThSAT_10 _2	**I**MFHENGSGI
12B	TthSAT_15_3	**E**KHERMFFEKDGSGI

32 *Tc*CS crystals were used in soaking experiments with the peptides. One peptide was soaked per crystal and crystals were frozen for diffraction experiments (Teng, 1990). Of the 32 crystals, 5 diffracted to 2.2 Å or higher. Data were processed and the original *Tc*CS structure was used in structure solution (Sowerby et al., 2023).

Data collected from a *Tc*CS crystal soaked with the peptide 9G, showed additional electron density to the protein structure, unbiased electron density is shown in Figure 6-7. Whilst this density was initially thought to represent the 9G peptide, the

structure showed that this density fitted instead to a C-terminus of the CS, Figure 6-8. This was also the case in the other 4 data sets.

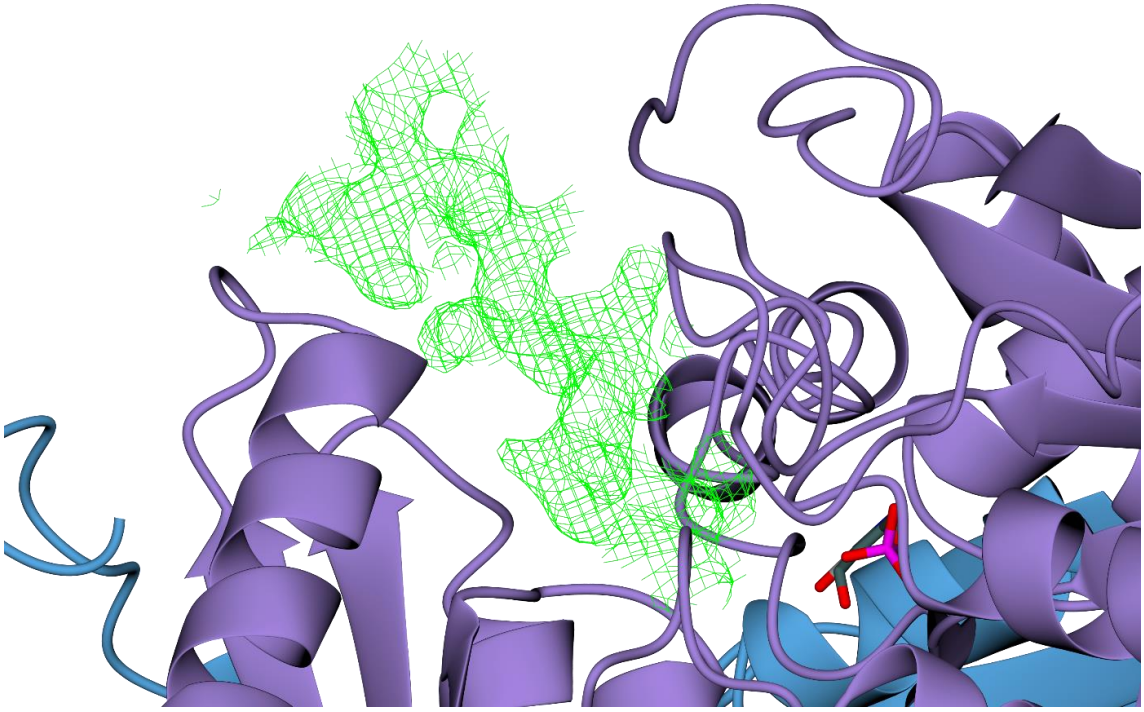


Figure 6-7 - Unbiased electron density in *Tc*CS structure at 1σ after one round of refinement. *Tc*CS shown in ribbon diagram with chain A in blue and chain B in purple. The F_o-F_c electron density map shown in green at 1σ.

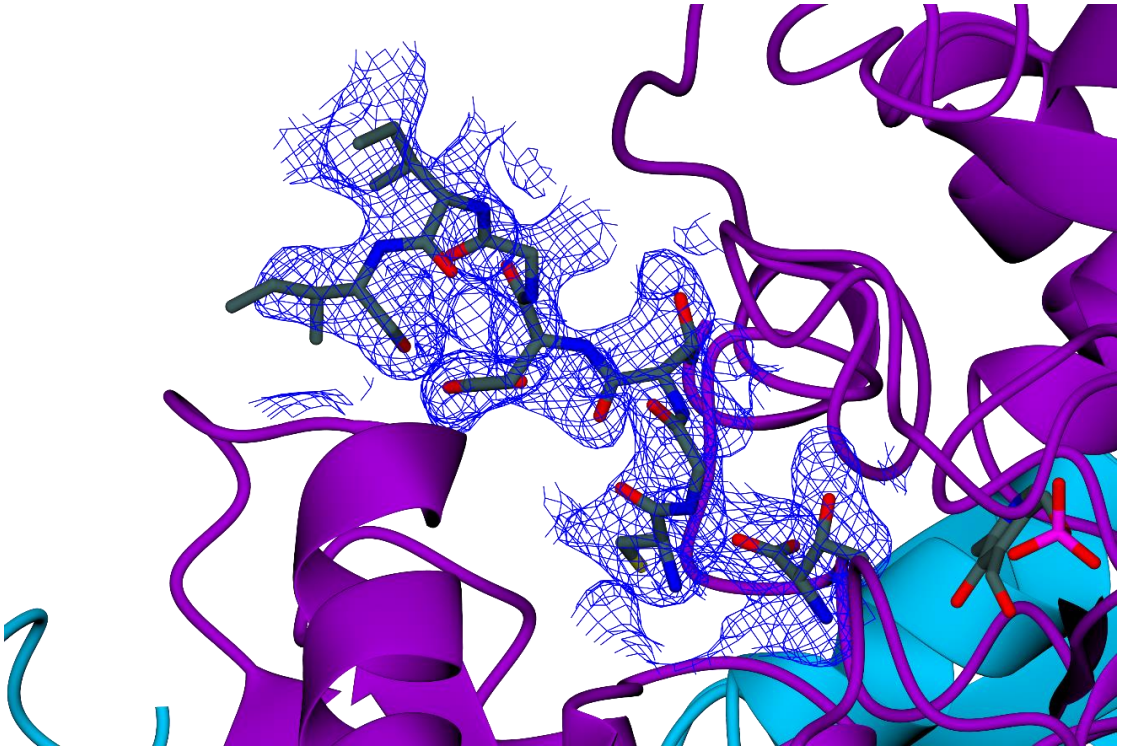


Figure 6-8 - *Tc*CS in ribbon form with C-terminus shown in stick representation. Chain A is in blue with chain B in purple. 2Fo-Fc map shown is at 1σ.

In summary, 32 crystals were soaked with peptides and diffraction experiments were performed. This resulted in 5 data sets, however the molecular details of the peptides binding and effecting *Tc*CS remains to be determined.

Comparison with other C-terminal peptide structures

Structural determination of C-terminal peptides bound to CS have been seen twice in *H. influenzae* as a pentamer from HiSAT (PDB 3IQG) and the final four residues of a decamer from *Hi*SAT (PDB 1Y7L) and a tetramer (PDB 3T4P) in *L. donovani* (Huang et al., 2005; Raj et al., 2012; Salsi et al., 2010).

When compared with other structures containing peptides or fragments of peptides bound to homologous crystals, determination of a C-terminus was unexpected as the electron density was in same areas of peptide binding from other structures,

Figure 6-9. This may be a result of crystallisation as in this study and in studies with *H. influenzae,* activity has clearly been shown to be effected.

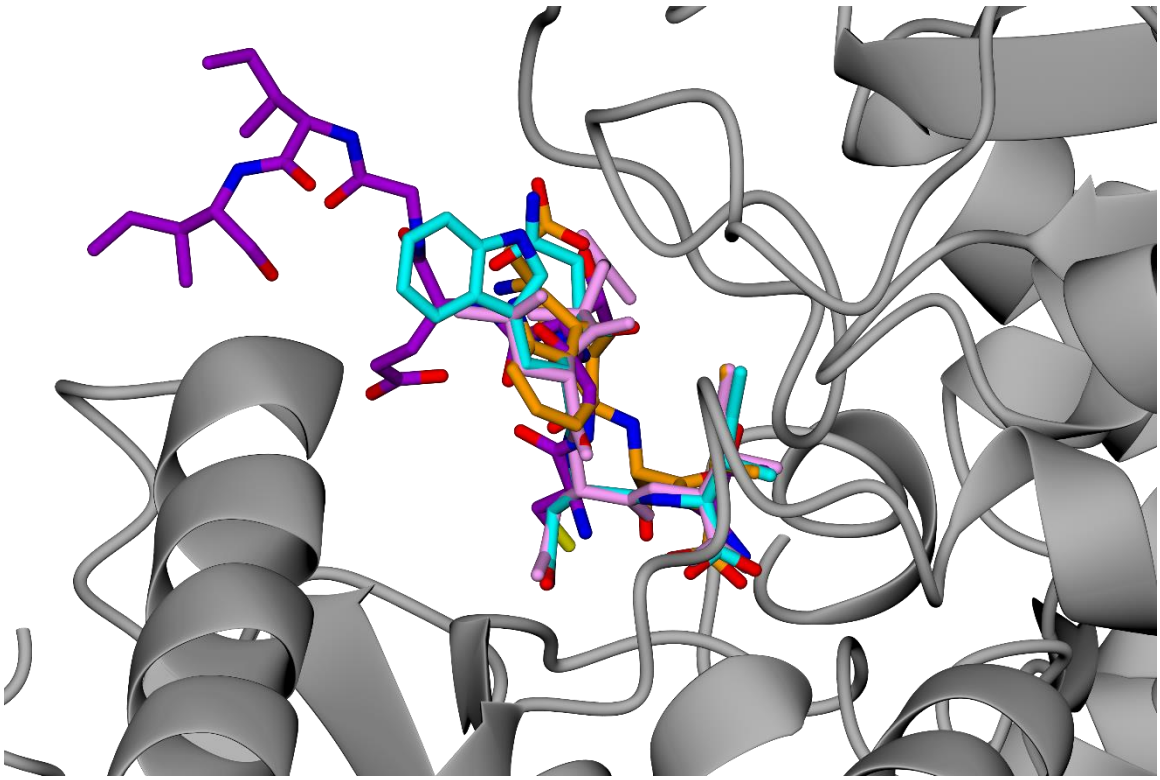


Figure 6-9 - *Tc*CS shown in ribbon form with C-terminus in purple, with superposition of peptides from 3IQG in blue, 1Y7L in pink and 3T4P in orange.

Analysis of C-terminal peptides with *Tc*CS

The ability to bind to the active site and the effect on cysteine production of *Tc*CS in the presence of 96 custom made peptides was measured. Of these, 52 peptides were shown to affect the PLP in the active site, implying binding and 36 were shown to effect cysteine production; either increasing or decreasing activity. 22 peptides were shown to both bind and effect activity. Although some peptides do not appear to bind in the active site they are still affecting activity and suggest that these peptides are binding to an allosteric site. The full sequence of these peptides is shown in Table 6-2.

Table 6-2 – Custom SAT Peptides shown to affect activity but not by binding in the active site. Residues in red indicate a mutation from the original TcSAT sequence.

2E	*Tc*SAT10_5	ITLHKGGSDI
3A	*Tc*SAT10_9	ITLQGGGSDI
3B	*Tc*SAT10_10	ITLQKNGSDI
3C	*Tc*SAT10_11	ITLQKDGSDI
4F	*Tc*SAT11_10	YITLQKNGSDI
5B	*Tc*SAT12_2	PYIFLQKGGSDI
6B	*Tc*SAT12_10	PYITLQKNGSDI
6C	*Tc*SAT12_11	PYITLQKDGSDI
6D	*Tc*SAT12_12	PYITLQKGGSGI
7B	*Tc*SAT13_6	TPYITLEKGGSDI
7E	*Tc*SAT13_9	TPYITLQGGGSDI
7G	*Tc*SAT13_11	TPYITLQKDGSDI
12D	n*Bv*SAT_15	ENVPAGSRIVQRKNK
12H	n*Gm*SAT_15	VPPRTTAVGNPARLV

Effect on mutations on peptide behaviour

The alteration of glycine to asparagine in peptides 3B, 4F, 6B and of glycine to aspartic acid in peptides 3C, 6C and 7G show both increases and decreases to activity. Mutation of the glutamine residue, an amide, to histidine, a basic amino acid in 2E confers a decrease in activity whereas mutation of the glutamine to glutamic acid, an acidic amino acid, in 7B causes an increase in activity. This

suggests that in this position, the charge of the amino acid has a significant effect on activity.

Despite changing the lysine residue to a glycine in both peptides 3A and 7E, these peptides cause converse effects with 3A decreasing activity and 7E increasing activity. The similarity in these sequences, particularly with the same mutation, effecting the activity in such opposing ways suggest that the length of sequence is critical to activity with shorter peptides resulting in a decrease in activity.

Altering the tyrosine residue in 5B to a phenylalanine therefore changing a small nucleophilic residue for a larger aromatic residue increases in activity. Further, altering the acidic aspartic acid to the small neutral glycine residue in 6D decreases the activity. As shown in Salsi et al. (2010), when the penultimate residue is a glutamine, the residue contributes significantly to binding. The increase in activity caused by peptides 12D and 12H may be due to the increased length of peptide.

Overall, these results suggests that smaller and basic amino acids in a short peptide may produce a potent inhibitor of *Trypanosoma cruzi* cysteine synthase.

Comparison of *Tc*CS studies with UPAR415, a previously developed peptide based inhibitor

Further evidence that this a promising avenue of exploration is seen in previous studies, such a work with CS from *S. enterica* serovar *Typhimurium* (*St*CS) to find an inhibitor against this bacterial pathogen (Annunziato et al., 2021). *Tc*CS and *St*CS share a sequence identity of only 47%, which is to be expected as StCS is a bacterial protein rather than a trypanosomatid protein.

A superposition of *Tc*CS and *St*CS reveals an r.m.s.d. of 1.2 A^2 over α carbons of 283 residues. The unexpectedly high r.m.s.d. is likely to be due to the highly conserved fold of CS; the inhibitor determined in the *St*CS structure does fits to the active site of *Tc*CS with the location of the PLP matching and with the UPAR415 inhibitor located in the same place as the OAS molecule determined in the *Tc*CS structure.

Using ^{19}F NMR to determine small molecule binding to *Tc*CS

To establish conditions producing successful ^{19}F NMR spectra with *Tc*CS, proof of principle experiments were conducted with available fluorinated molecules. The overall aim of these experiments was to optimise conditions for successful ^{19}F NMR spectra collection with the *Tc*CS protein. 5 fluorinated small molecules were selected to be used in ^{19}F NMR experiments to determine binding to *Tc*CS, Figure 6-10. As fluorinated phenylalanine was the closest available molecule to known binding amino acids such as cysteine and serine, 3 compounds based on phenylalanine containing at least one fluorine atom, 2F, 34F and 35F were obtained from available libraries. 2 compounds based on fragments from the Fragile library, Fr4 and Fr23 were chosen and synthesised based on other experimental work.

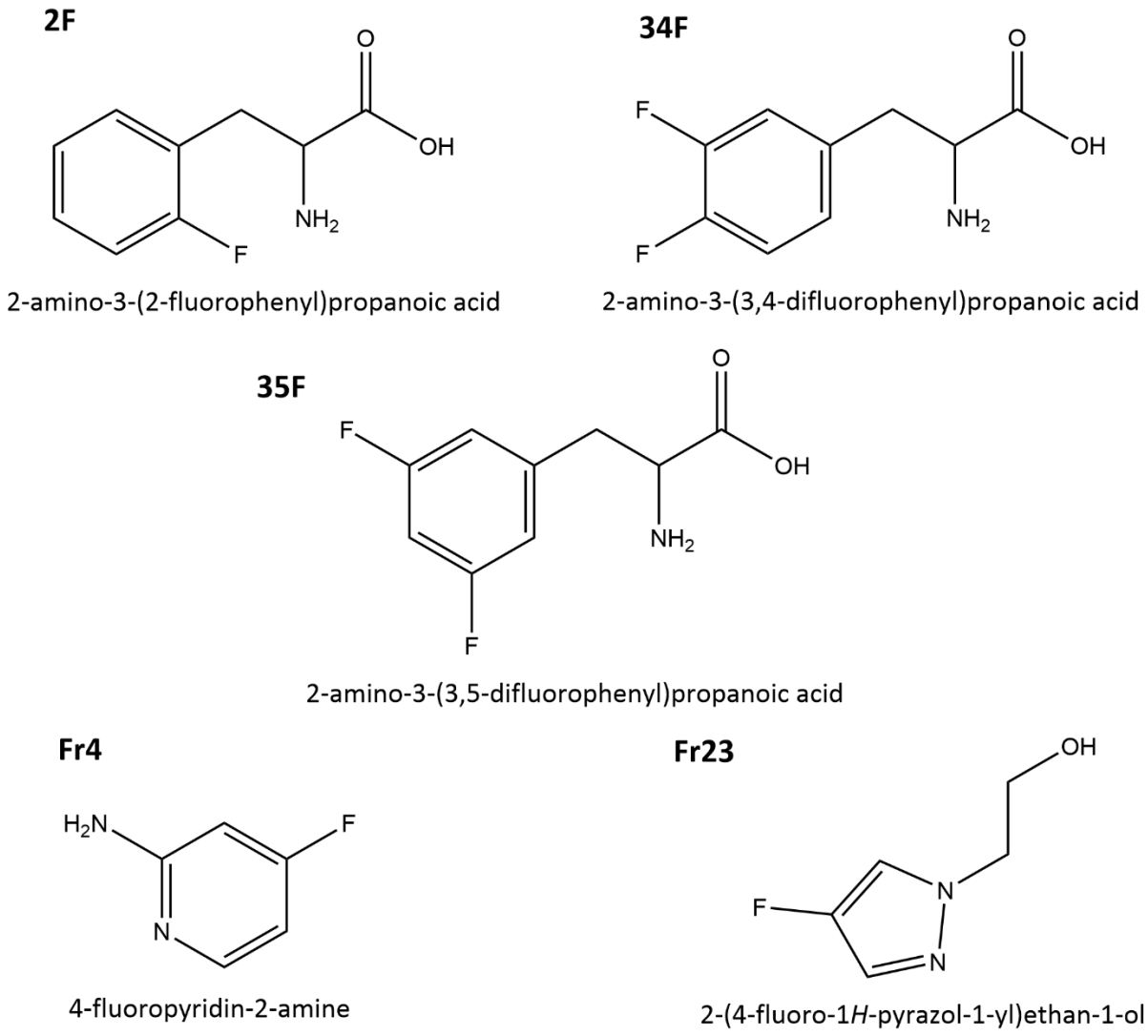


Figure 6-10 - Compounds selected for use in ^{19}F NMR experiments.

Spectra of the compound alone and then the compound with protein were collected for each sample, excluding Fr23 where only data for the compound was collected. The stacked spectra for 2F with and without *Tc*CS is shown in Figure 6-11, spectra of 34F and 35F are shown in appendix, Figures S9-6-1 and S9-6-2. The Fr4 fragment alone resulted in a broad line width that made this unsuitable for the experiment as after the T_2 filter application, no signal could be seen.

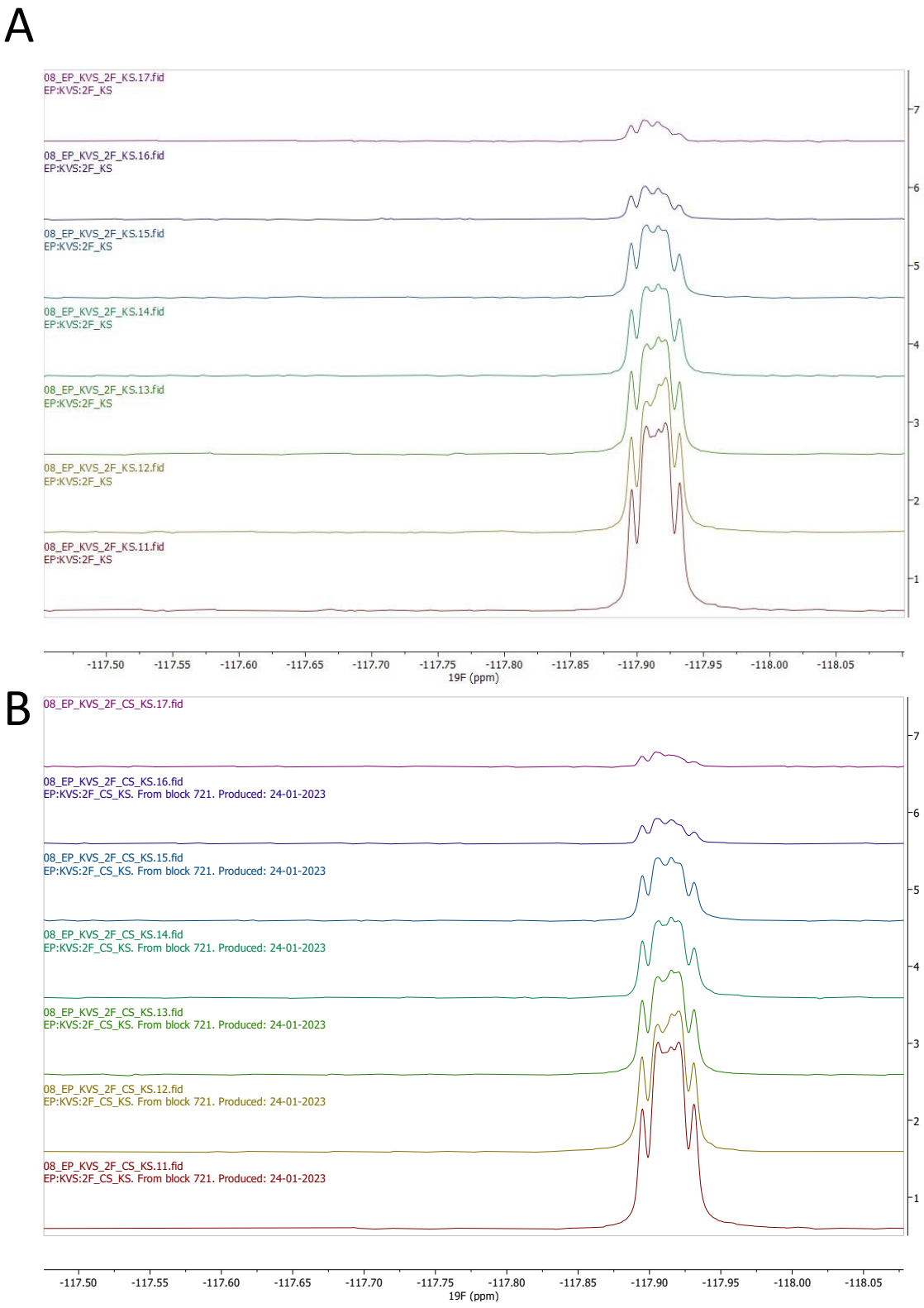


Figure 6-11 - ^{19}F NMR spectra of 2F without (A) and with (B) *Tc*CS. Spectra from different filter timepoints are stacked, 1 – 4 ms, 2 – 160 ms, 3 – 320 ms, 4 – 480 ms, 5 – 640 ms, 6 – 1280 ms, 7 – 1680 ms.

From the stacked spectra, no change in chemical shift is observed. To identify a change in line width for 2F, 34F and 35F the absolute integral was measured against the T_2 filter time applied to the sample and compared, Figure 6-12. Binding is indicated by a difference in signal decay.

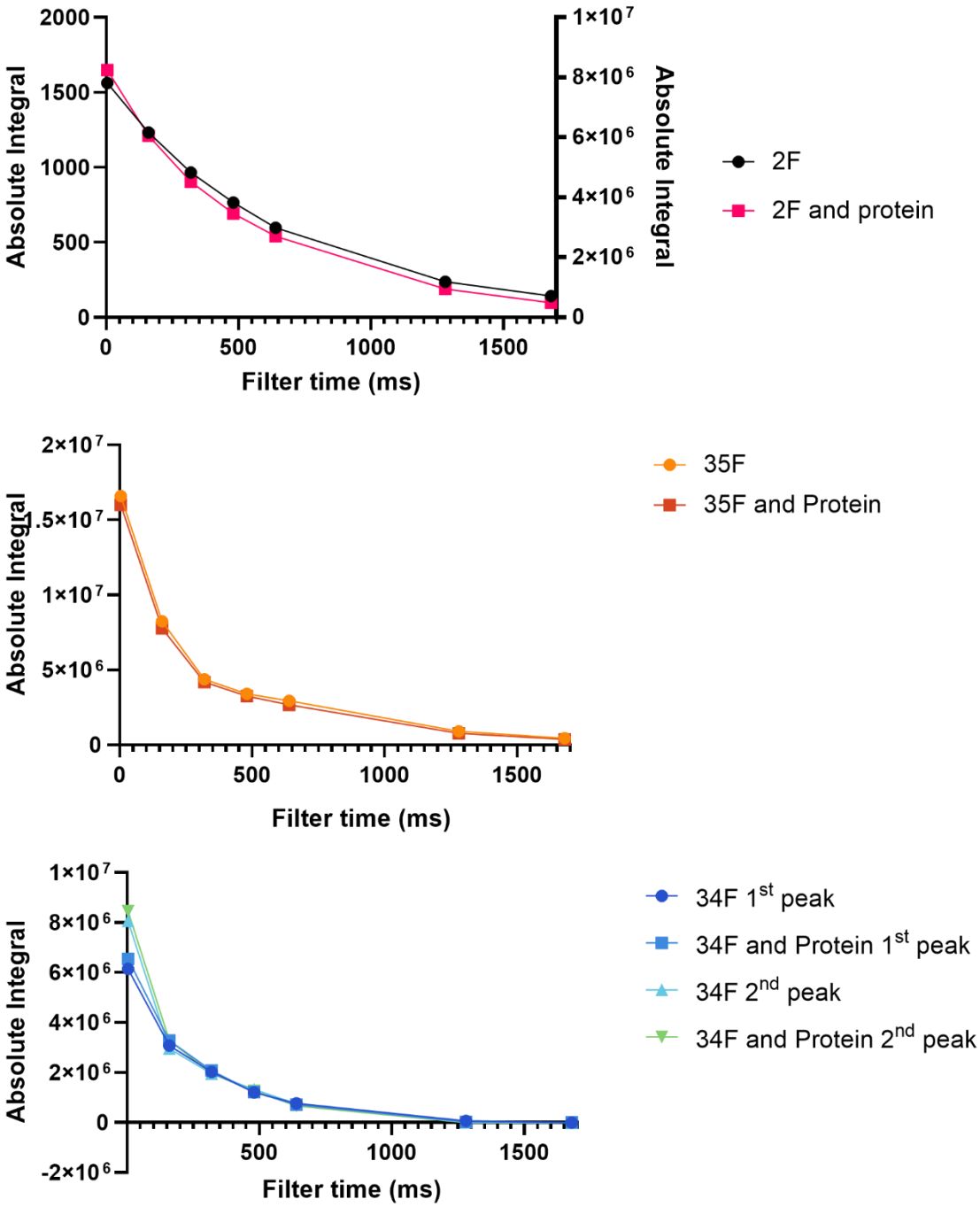


Figure 6-12 - Signal decay of fluorinated compound vs fluorinated compound and protein for 2F, 35F and 34F.

The 2F dataset shows a small difference between the compound alone and in the presence of the protein however as these samples were run only once, significance cannot be judged from the result presented. The 35F dataset clearly shows that

there is no change occurring in the presence of the protein. The 34F dataset contains information for the 1st and 2nd peak due to two inequivalent fluorine atoms in the molecule generating two peaks. Both peaks do not show a change occurring when *Tc*CS is present.

Overall, successful conditions for the collection of ^{19}F NMR data with *Tc*CS was established. The NMR spectra collected indicates that the three molecules do not bind and therefore three small fluorinated molecules have been identified as negative controls for future experiments. This suggests that there is either no binding or that if binding is occurring it is at too small a proportion compared to the unbound compound to be detected.

Using fragment-based drug design strategies for *de novo* synthesis of inhibitors

Initial soaking with the Fraglite library

Recombinant protein of *Tc*CS and *Li*CS were produced and crystals were grown, as previously described. Stock solutions were prepared from the Fraglite library (Wood et al., 2019). Crystals of *Tc*CS and *Li*CS were soaked for 10 minutes with each of the 31 compounds in the Fraglite library. For each crystal, diffraction data were collected, manually processed and structures determined. Statistics for *Tc*CS structures and *Li*CS structures are shown in the appendix, Table S9-6-3 and Table S9-6-4. The electron density in the fully refined structure was analysed for evidence of fragments bound to the structure. To assist in determining the locations of the fragments, the anomalous difference density was used to resolve the locations of the halogen atoms found in the fragments.

***Tc*CS fragments**

For *Tc*CS, six compounds were successfully soaked into the crystals with 28 positions found, Table 6-3. For all six structures, the protein crystallised in the $P2_1$ group, the crystals diffracted to 2.3 Å or above with a completeness of 100% except for the Fraglite 17 data set which had a completeness of 98%. 97% or above of the protein lies in the favoured areas of the Ramachandran plot. With the fully refined structures, ligands were unambiguously determined in the unbiased F_o-F_c map before further validation using the anomalous dispersion difference map calculated.

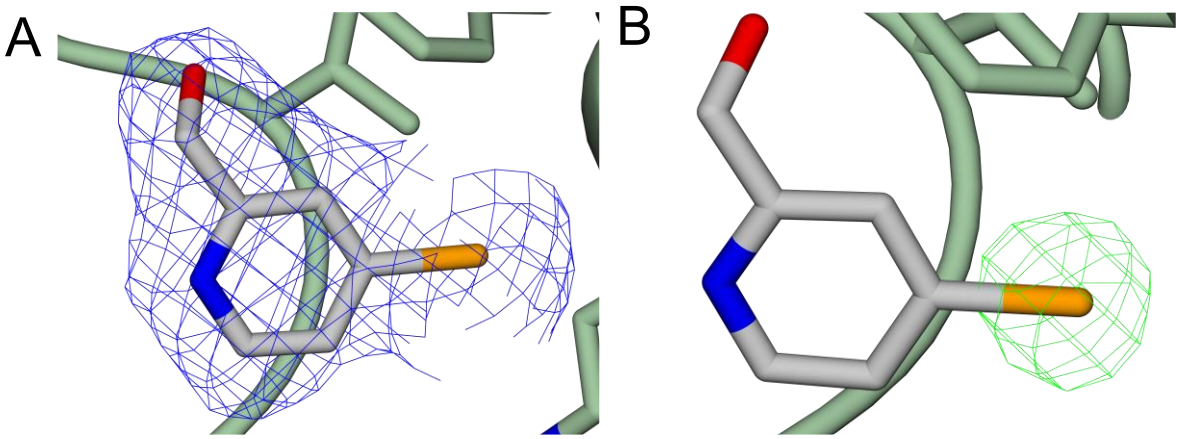


Figure 6-13 - FragLite 18 in *Tc*CS showing density. *Tc*CS is shown in ribbon format with neighbouring residues and FragLite shown in stick format. Atom colour coding: O – red, N – blue, S – yellow, Br – orange. A – $2F_o$-F_c density (shown in blue) for FragLite, shown at 1 σ. B – anomalous difference density (shown in green) for FragLite, shown at 3 σ.

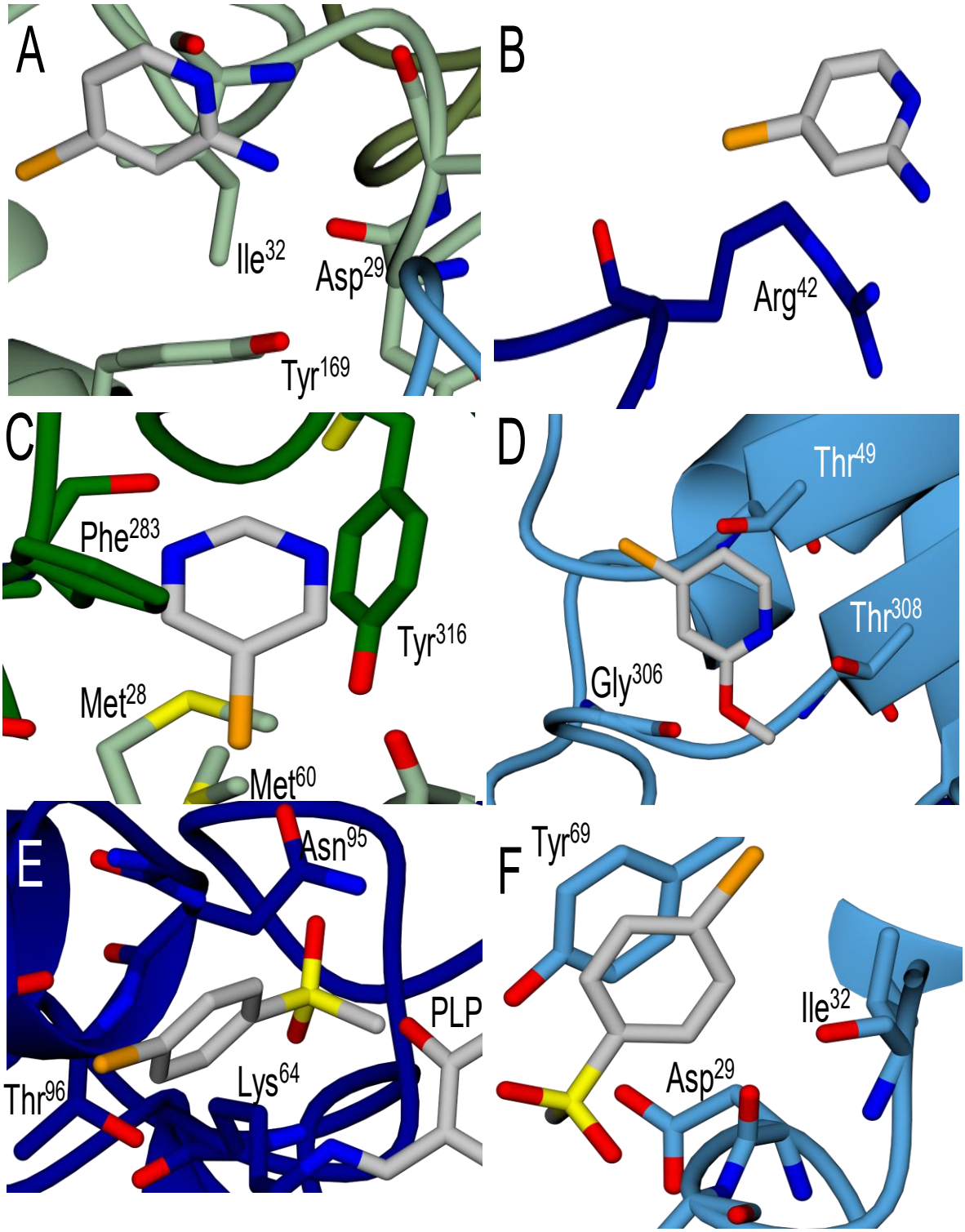
All six compounds contained bromine and anomalous dispersion difference maps were produced to determine the position of the heavy atom in the structure. For all 6 structures, the overall B factors for the protein are between 54 and 66 $Å^2$ with the average B-factors for the ligands between 68 and 84 $Å^2$. The slightly higher B-factors seen in the ligands compared to the protein are to be expected as whilst the ligands are clearly present some movement is to be expected.

Table 6-3 - Fragments bound to *Tc*CS structures.

Crystal System	**Fragment (Total number in asymmetric unit)**	**Interacting residues (number of fragments involved)**	**Figure 6-14**
*Tc*CS	4 (5)	Asp^{29}, Ile^{32}, Tyr^{69} (3)	A
		Arg^{42} (2)	B

*Tc*CS	13 (4)	Met^{28}, Met^{60}, Phe^{283}, Tyr^{316}	C
*Tc*CS	15 (3)	Thr^{49}, Gly^{306}, Thr^{308}	D
*Tc*CS	17 (4)	Lys^{64}, Asn^{95}, Thr^{96}, Gly^{234}, PLP	E
		(3)	
		Asp^{29}, Ile^{32}, Tyr^{69}	F
		(1)	
*Tc*CS	18 (8)	Ile^{32}, Arg^{66h}, Tyr^{69}, Thr^{178}	G
		(4)	
		Ser^{162}, Ala^{163}, Asn^{164}	H
		(2)	
		Trp^{184}, Arg^{185}(1)	I
		Thr^{92}, Asn^{95}, Gln^{165} (1)	J
*Tc*CS	23 (4)	Ile^{32}, Arg^{66}, Tyr^{69}	K

Fragments were typically found in all four chains of the asymmetric unit; images of each fragment are shown in Figure 6-14. Images of fragments in density are shown in the appendix, Figure S9-6-5. The fragments found in *Tc*CS were bound in several different places and therefore have interactions with different residues in the structure.


A
Ile32
Asp29
Tyr169
B
Arg42
C
Phe283
Tyr316
Met28
Met60
D
Thr49
Thr308
Gly306
E
Asn95
PLP
Thr96
Lys64
F
Tyr69
Ile32
Asp29

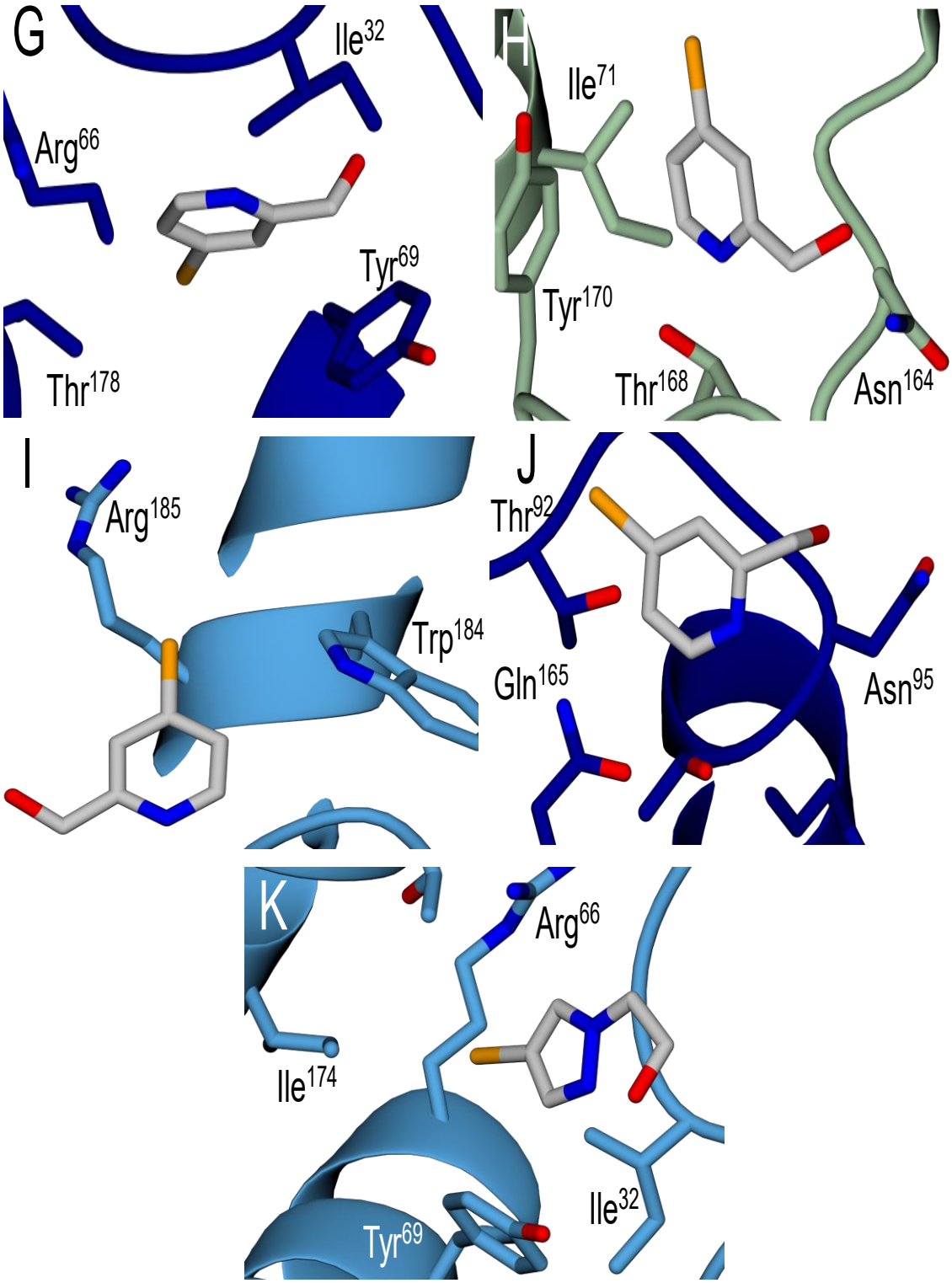


Figure 6-14 - FragLites bound to *Tc*CS. *Tc*CS shown in ribbon form with chain A in light green and chain B in dark green, chain C in dark blue and chain D in light blue. FragLites and interacting residues are shown in stick form. Atom colour coding: O – red, N – blue, S – yellow, Br – orange.

Fragment 4 was bound in two locations, with three fragments interacting with the Ala^{30} site and two fragments interacting with the Arg^{42} residue. A single site was identified for Fragment 13 with all 4 fragments bound to the same residues. Three sites were identified for Fragment 15 however hydrogen bonds were only formed between one of these fragments and residues of the protein. Fragment 17 was identified in two positions, the first with three fragments is the Asn^{95} site, also forming hydrogen bonds with PLP and Lys^{64}. The second site involves Ile^{32} and one fragment was discovered here. For Fragment 18, four binding sites were elucidated. The first site had four fragments binding to the Ile^{32} site, the second had two fragments in the Ala^{163} site. The other two fragments were bound in two distinct locations. Four molecules of Fragment 23 were found bound to Ala^{30} in the *Tc*CS structure. The determination of 6 fragments bound to *Tc*CS establish a foundation for further FBDD work.

_Li_CS fragments

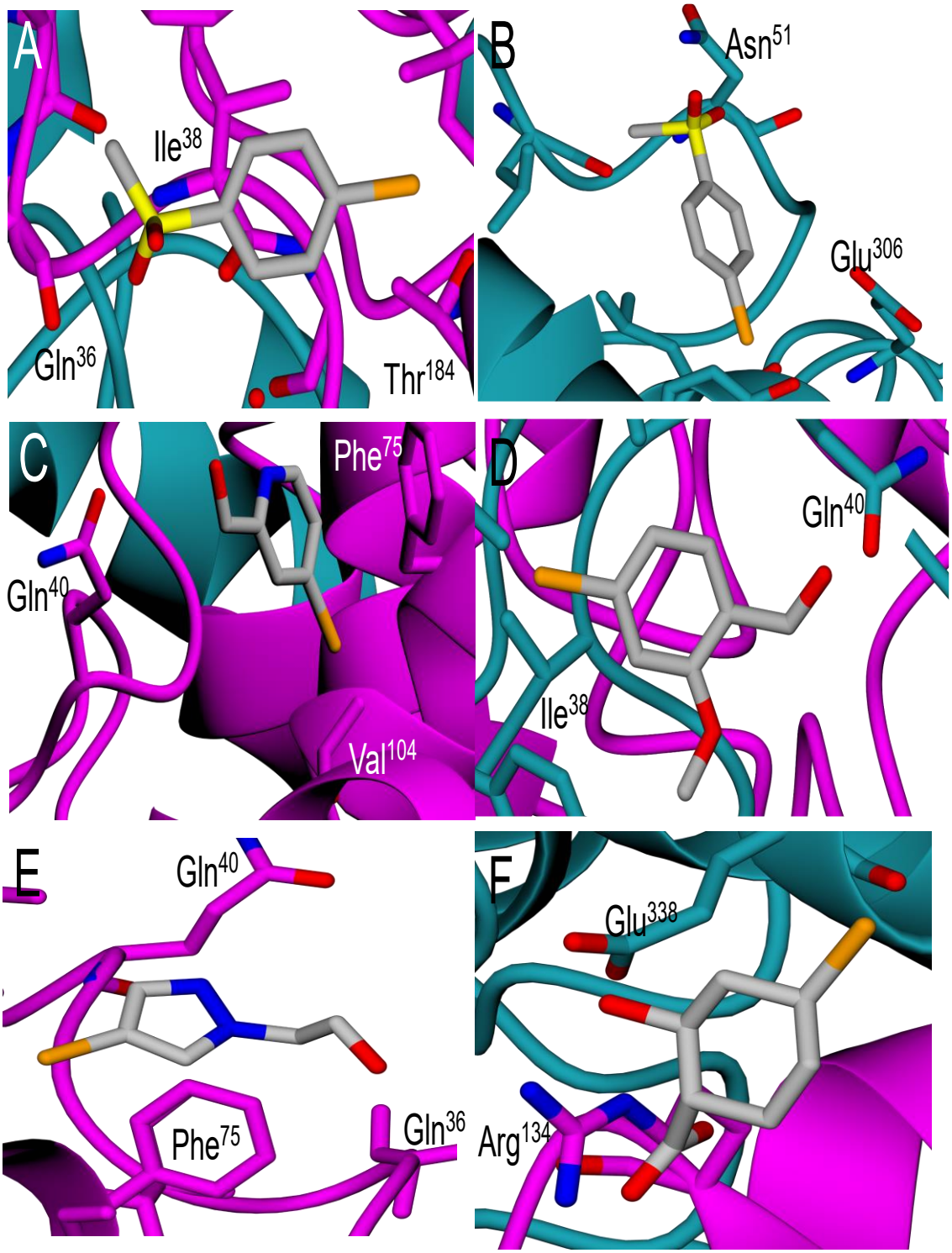
To complement the data for *Tc*CS, the same experiments were repeated with *Li*CS. 8 compounds were successfully soaked with 20 positions found in the structure, Table 6-4. All 8 crystals were in the $P2_1\ 2_1\ 2_1$ space group and diffracted to 2.6 Å or higher with a completeness of 98% or above. 97% or above of the protein lies in the favoured areas of the Ramachandran plot. The 8 compounds bound to *Li*CS contained bromine and anomalous dispersion difference maps were produced to determine the position of the heavy atom within the structure. For all 8 structures, the overall B factors for the protein are between 27 and 59 $Å^2$ while the average B-factors for the ligands is between 34 and 101 $Å^2$.

Table 6-4 - FragLites found bound to *Li*CS.

Crystal System	Fragment (Number in asymmetric unit)	Interacting residues	Figure 6-15
*Li*CS	17 (4)	Gln^{36}, Ile^{38}, Thr^{184} (2)	A
		Asn^{51}, Glu^{306} (2)	B
*Li*CS	18 (2)	Gln^{40}, Phe^{75}, Val^{104}	C
*Li*CS	22 (2)	Ile^{38}, Gln^{40}	D
*Li*CS	23 (2)	Gln^{36}, Gln^{40}, Phe^{75}	E
*Li*CS	24 (3)	Arg^{134} (B), Glu^{338} (A) (1)	F
		Asp^{35}, Ile^{38}, Phe^{75} (2)	G
*Li*CS	28 (3)	Gln^{36}, Gln^{40}, Phe^{75}, Thr^{184} (2)	H
		Glu^{63} (A and B) (1)	I
LiCS	29 (2)	Ile^{38}, Gln^{40}, Phe^{75}	J
LiCS	31 (2)	Ile^{38}, Gln^{40}, Arg^{72}, Phe^{75}, Thr^{184}, Glu^{188}	K

Fragments in the *Li*CS structures were found in both chains of the asymmetric unit as well as in three other sites and are shown in Figure 6-15. Images of fragments in

density are shown in the appendix, Figure S9-6-6.



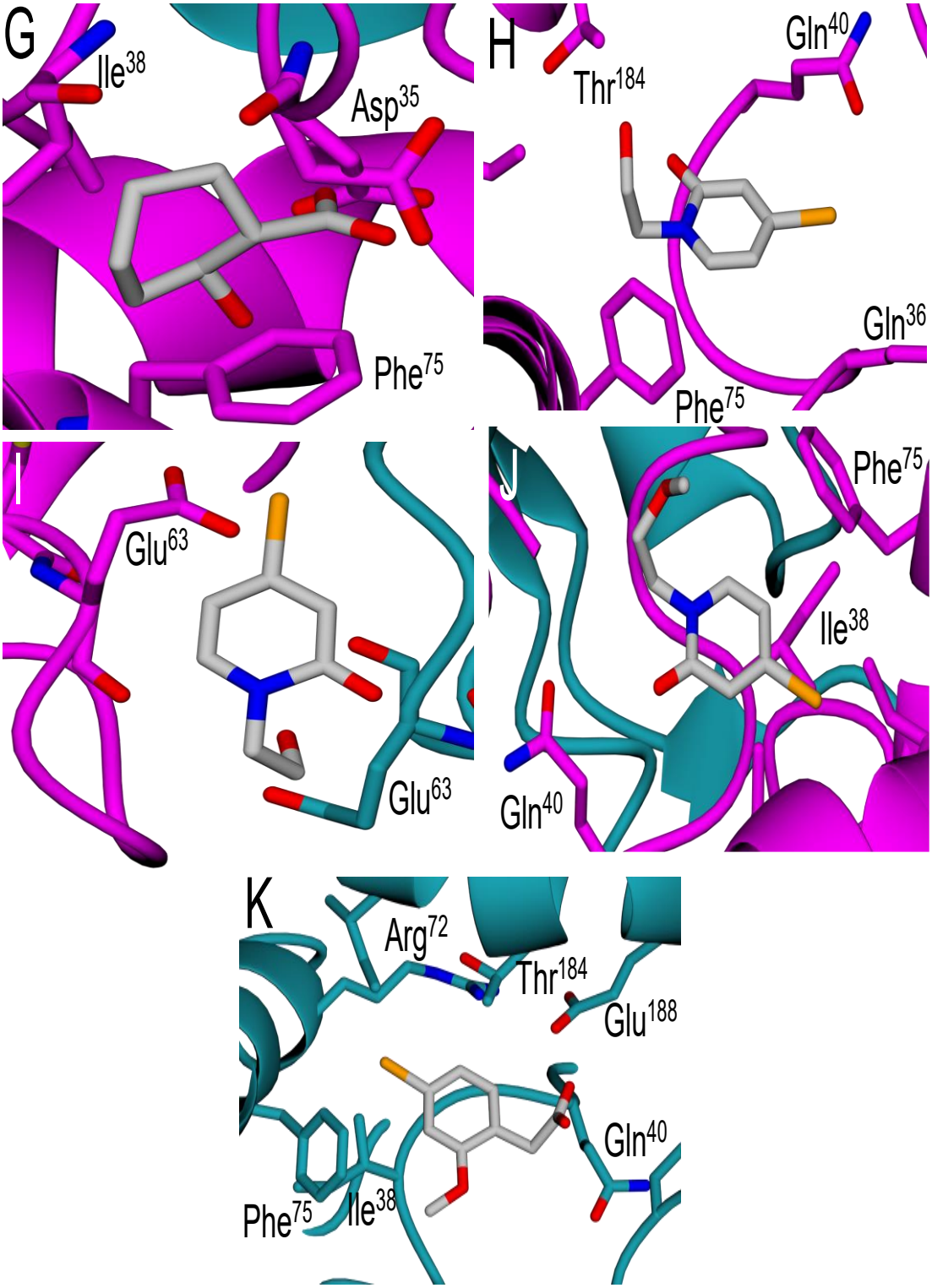


Figure 6-15 - *Li*CS structures with FragLites bound. *Li*CS shown in ribbon format with chain A in pink and chain B in blue. FragLites and interacting residues are shown in stick form. Atom colour coding: O – red, N – blue, S – yellow, Br – orange.

*Li*CS fragments were clustered in three main binding sites. A copy of all fragments are found in the Gln^{40} site in each chain. Two molecules of fragment 17 were also found at Asn^{51}. Fragments 23 and 28 were discovered at the dimer interface. Discovery of a range of FragLites bound to *Li*CS in several positions provides several potential sites for further development.

Overview of FragLites bound to structures

11 different fragments were determined to be bound to the structures of *Li*CS and *Tc*CS over 14 data sets. A superposition of these structures are shown in Figure 6-16. Through this 11 different binding sites were identified, with the pocket involving the residues Ile^{32} (*Tc*CS) and Gln^{40} (*Li*CS) the most frequently occurring. There are 3 fragments for which the C-Br bond has broken and so the anomalous difference density is not present. This is to be expected due to the labile nature of a C-Br bond however the fragment is still described as bound due to the clear electron density present.

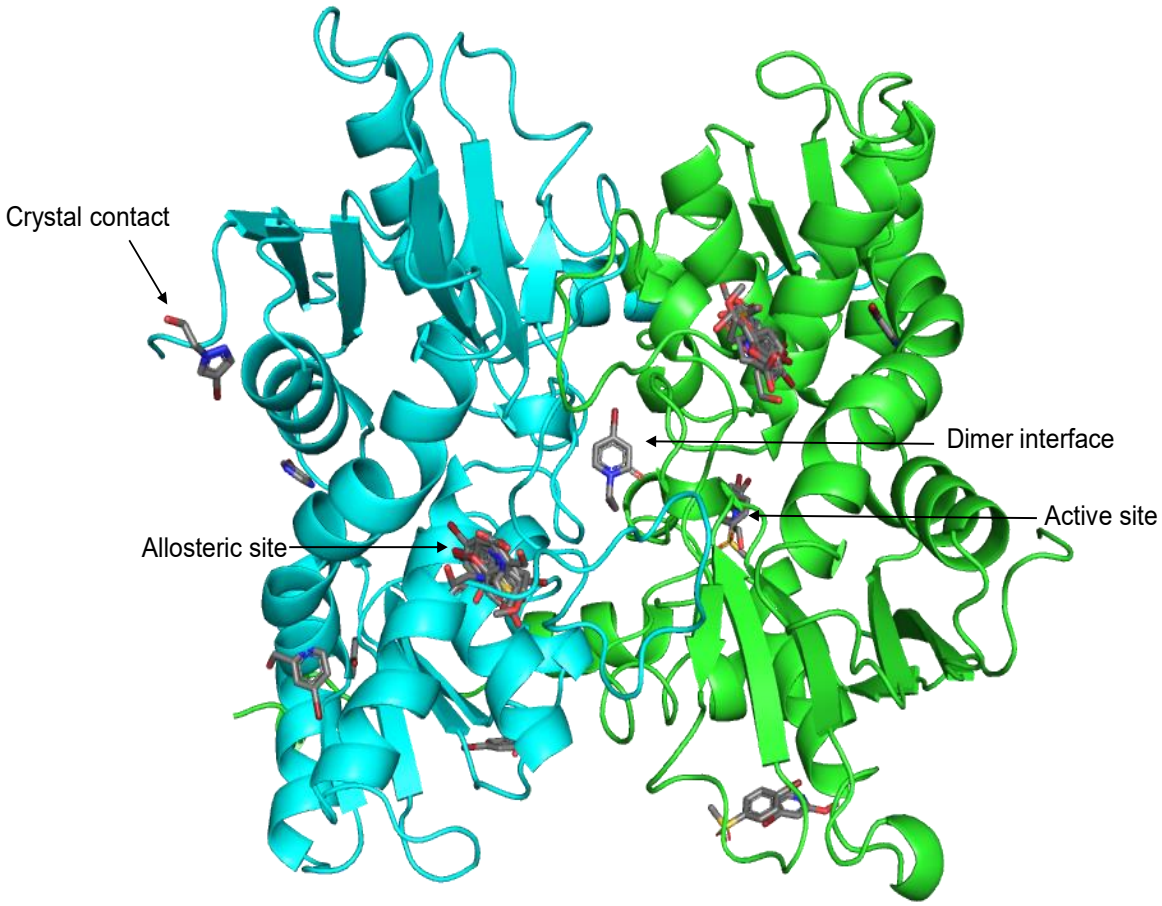


Figure 6-16 - Superposition of FragLite structures with protein shown in ribbon form and fragments shown in stick form. Chain A in blue, chain B in green.

The fragments determined in the structures can be roughly grouped into four categories: active site, allosteric site, dimer interface and crystal contact. The fragments found in the active site provide a promising starting point for the development of competitive inhibitors, particularly the copy of Fragment 17 in the active site of *Tc*CS which forms hydrogen bonds with the PLP. Fragments determined in potential allosteric sites reveal binding hotspots with the proteins and also represent the capacity to be expanded into non-competitive regulators of cysteine synthase. One copy of Fragment 24 is found at the dimer interface of the two monomers. This fragment presents an intriguing basis for the development of

protein-protein interaction inhibitors, especially when considered with other molecules, such as ribose, determined to be bound to the dimer interface. Some fragments, such as Fragment 22 in *Li*CS, represent crystal contacts and therefore should not be pursued further.

Due to the positioning of the fragments determined the molecules are not suitable for fragment linking due to the significant distance between them. In light of this, fragment growing methods were chosen to be pursued.

Effect of FragLites on Activity

Following the successful determination of fragments bound to *Tc*CS and *Li*CS, the activity of *Tc*CS with the fragments was examined to determine if an effect on activity is seen, Figure 6-17.

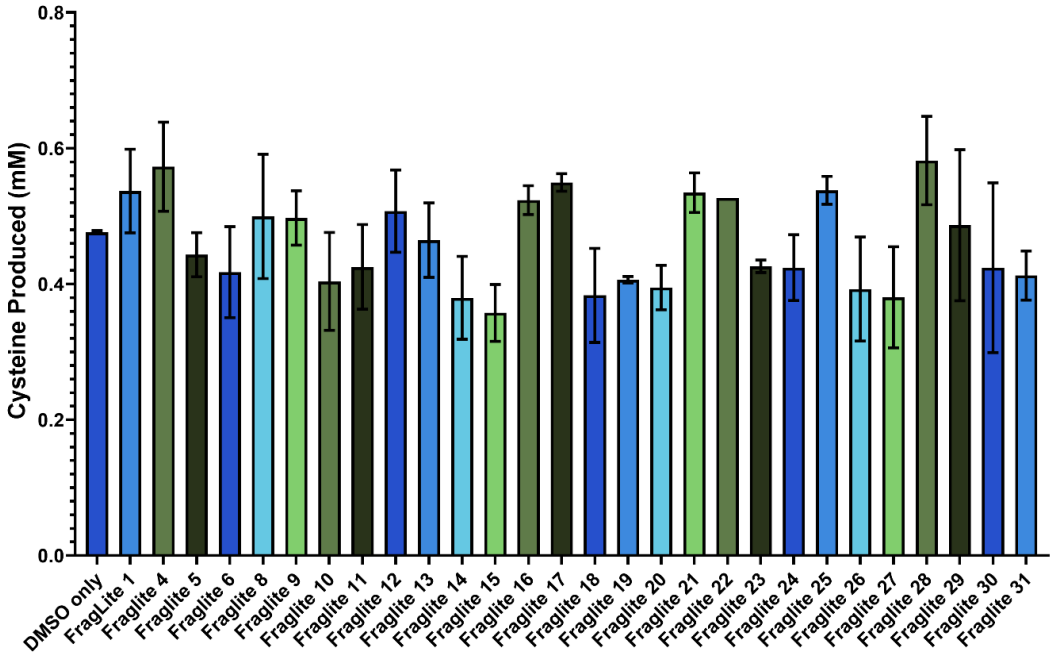


Figure 6-17 - Effect on *Tc*CS activity with Fraglites. n=3 with at least 2 data points used per measurement.

As shown in Figure 6-17, no fragment has a sufficiently significant reduction in activity to be explored as a potential inhibitor. Whilst this result was expected due to the expected low binding constant and confirms that investigations into fragment linking or fragment growing is required.

Exploration of the DSi-poised library with *Tc*CS crystals

Following the success of soaking the crystals with the Fraglite library, it was decided to scale up to test a larger library of compounds using the XChem facility at Diamond Light Source. The DSi-poised library was selected as it contains a large number of compounds previously identified as allowing for rapid synthesis following a fragment hit. The library contains 988 compounds, of which 170 were soaked into *Tc*CS crystals. Crystals were harvested and transferred to sitting drop trays before soaking and collection for diffraction experiments. Data were collected automatically from 170 crystals. From those crystals, 85 diffracted to below 4 Å, but after manual data processing only 72 were processed in the correct space group. 20 data sets were selected by PanDDA as containing potential ligands yet upon manual inspection no ligands were found (Pearce et al., 2017). Although a higher throughput technique than manual soaks, this method did not result in any fragments clearly bound to the crystal. This is likely to be a result of the lower resolution data collected which in turn could be attributed to automatic data collection and extensive crystal handling.

Development of inhibitor compounds based on FragLite work

In order to explore the allosteric binding site identified by the FragLite screenings, the fragment cluster containing Fraglite 17, Figure 6-18, was chosen as a starting

point for exploration of this pocket due to its presence in a cluster of bound fragments.

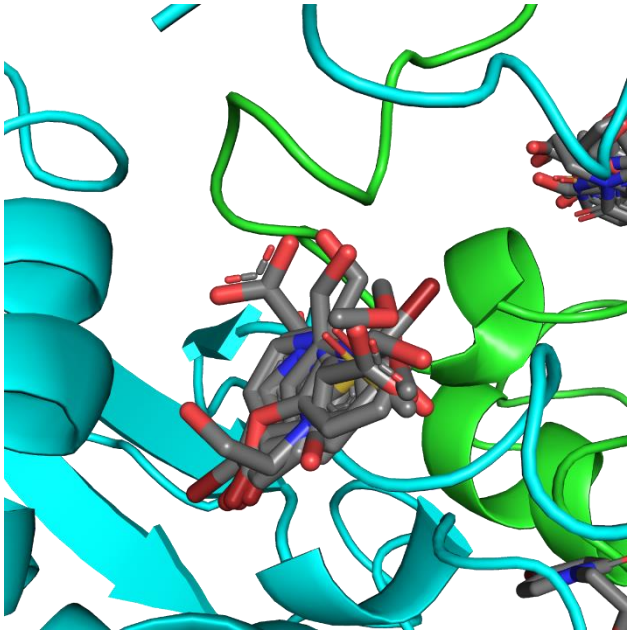


Figure 6-18 - Fragment cluster at Gln^{40} including Fraglite 17. Chain A in blue, chain B in green.

AI-lead fragment growing was conducted by collaborators in the Cole group at Newcastle University (Bieniek et al., 2022). Briefly, AI-lead software was used to add different chemical groups to the benzene ring of FragLite 17 and a *in silico* library of >1000 molecules was produced. These molecules were docked into the protein structure, scored and the best 150 results were provided. Each molecule was manually inspected, a selection of molecules shown in Figure 6-19.

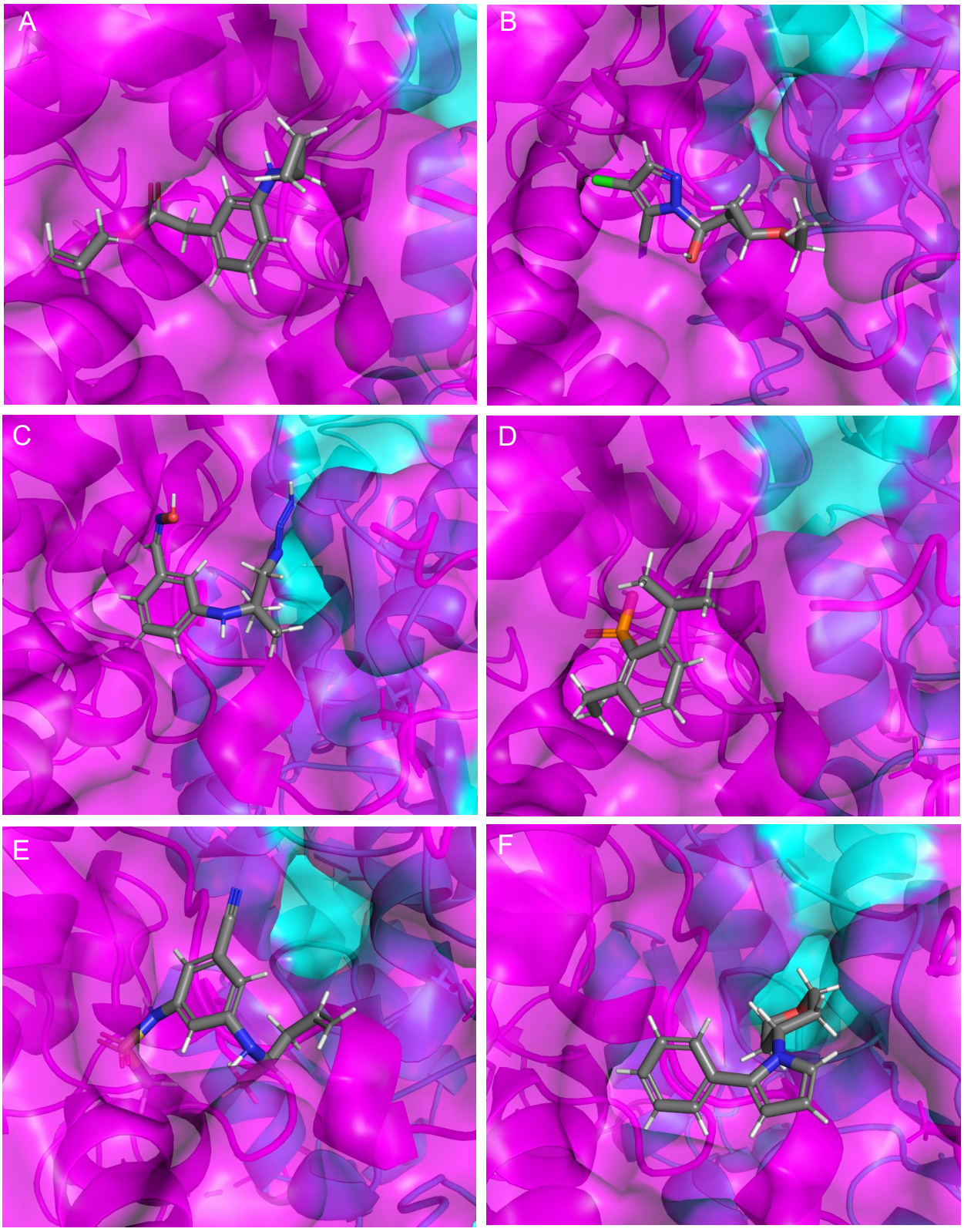


Figure 6-19 - Typical examples of molecules produced in AI lead fragment growing.

In Figure 6-19A, the compound has a severe clash between the branch coming from the benzene ring and the protein molecule, making it unsuitable as an inhibitor. In Figure 6-19B, the molecule has a less severe clash, however the C≡N bond is inserted into the surface of the protein. This clash will rule out this molecule and the C≡N bond is likely to be reactive. The Figure 6-19C compound displays an improved fit to the pocket but there is an azide bond and a nitric oxide bond on this compound. Both groups are highly reactive and likely to result in off target reactions occurring with the protein and with other molecules present. The Figure 6-19D compound is less reactive although there is a small clash occurring between the oxygen molecule and the surface of the protein. This molecule is also relatively small compared to others and more like a fragment; as the fragments has no effect on activity, this compound will be discarded. The Figure 6-19E compound fits the pocket well however has sulfonamide and cyanide groups making it very reactive and likely to cause unspecific reactions. The Figure 6-19F compound fits to the pocket nicely, would be reasonable to synthesise and does not contain any highly reactive groups so would be a good candidate molecule to pursue further.

Enamine Molecules

In addition to synthesis of the AI grown molecules, the best 2000 molecules from the previously generated library was searched against the Enaminie REAL database (Gorgulla et al., 2020). This database contains commercially available molecules with established synthesis routes. This search resulted in 57 potential compounds and after manual inspection, 16 molecules were selected for further analysis. Typically lead compounds have an octanol-water partition coefficient (logP) value of between 1 and 3 with a molecular weight between 100 and 250 Da. When

designing from fragments, no more than 3 hydrogen bond acceptors, donors and rotatable bonds is preferred (Mahapatra et al., 2022). The ideal logP value for oral and intestinal absorption is between 1.35 and 1.8 (Bhal, 2011). Of the 16 molecules, 9 molecules have a predicted logP between 1 and 3, Figure 6-20A and B respectively. Of the 16 molecules, only 4 do not fall within the 100 – 250 Da range, Figure 6-20C and D. All 16 molecules have fewer than 3 hydrogen bond donors; 4 molecules have more than 3 hydrogen bond acceptors, Figure 6-20E. With these criteria, there is only one molecule that satisfies all the ideal requirements, compound 37, Figure 6-20F.

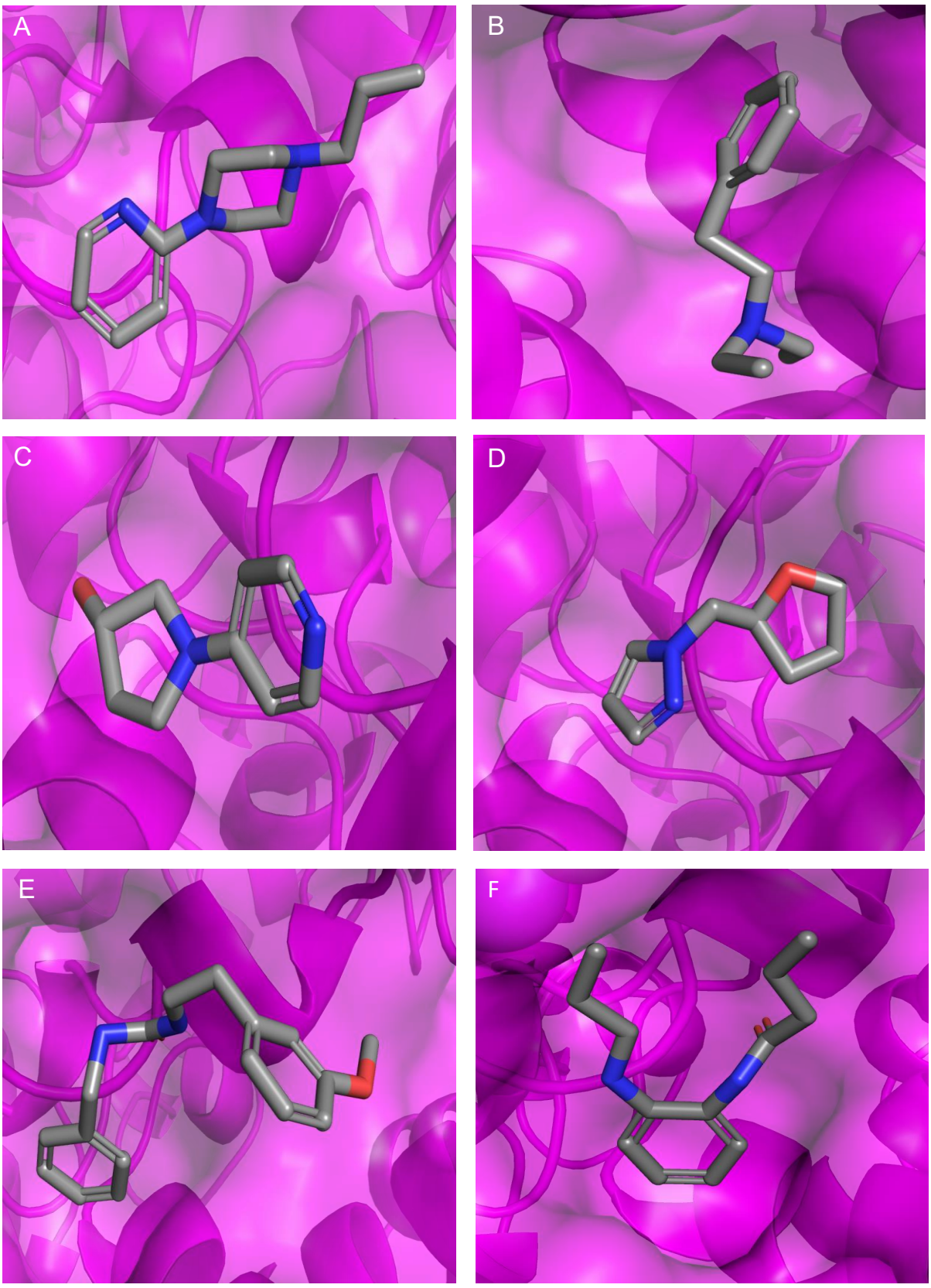


Figure 6-20 - Compounds identified from Enamine database.

Based on a good fit to the binding pocket in addition to good chemical properties such as an ideal logP of 1.47, a molecular weight of 138 Da, 2 hydrogen bond acceptors and one hydrogen bond donator, compound 37 has been identified as the most promising compound.

Conclusions

Using the previously determined structures of *Tc*CS and *Li*CS, several methods of drug design have been explored. These methods include using previously identified inhibitors such as the C-terminus of SAT to affect CS activity, establishing ^{19}F NMR methodology for the screening of fluorinated compounds with *Tc*CS, and using fragment-based drug design.

Design and synthesis of 96 peptides based on the C-terminal sequence of *Tc*SAT and other homologs was undertaken. Using this library, exploration of natural inhibition of *Tc*CS through interaction with *Tc*SAT based peptides yielded results showing both an increase and a decrease in activity of *Tc*CS. Following this, structural studies were undertaken to determine the location of peptide binding within *Tc*CS however this remains unclear. Fluorescence studies suggest that peptide binding may occur outside of the active site but this remains to be determined. From these results, it was determined that shorter and more basic peptides were more likely to decrease the activity of *Tc*CS. This method is supported by other CS studies using the same approach to provide a basis for inhibitor development.

Experimental methodology for ^{19}F NMR has been established with *Tc*CS with proof-of-principle experiments being successfully conducted. Conditions of the experiment were optimised, and three different compounds were identified to be used as negative controls in future experiments.

De novo drug synthesis was explored using a fragment-based drug design approach with both *Li*CS and *Tc*CS. The high sequence identity of 78% between these two proteins allows for findings from one protein to be translated to the other. The

fragment screening resulted in 14 structures and the identification 11 distinct fragments. The fragments determined are broadly grouped into four categories: active site, allosteric site, dimer interface and crystal contact. With the aim of exploration of the allosteric site, Fragment 17 was selected for use in an AI-lead fragment growing programme. This generated a library of over 1000 potential compounds to be explored further.

This molecular library was used in exploration of the Enamine REAL which yielded 16 promising molecules (Gorgulla et al., 2020). After detailed examination of fit to pocket, clashes with protein and chemical properties such as logP, compound 37 was selected for further investigation.

This work provides a promising basis for determining the binding of molecules to *Tc*CS and the discovery of an inhibitor for cysteine synthase.

Chapter 7: Conclusions and Further Work

Over a billion people are affected by neglected tropical diseases globally. Chagas disease and leishmaniasis have been identified for control and elimination by WHO due to high unmet medical need (WHO, 2020). Current treatments for these diseases are insufficient due to limited efficacy, numerous side effects and emerging resistance in addition to poor understanding of the mode of action (Croft et al., 2006; Francisco et al., 2020).

Chagas disease and leishmaniasis are the result of infection by the parasites *Trypanosoma cruzi* and *Leishmania* species respectively (Hemmige et al., 2012; Mann et al., 2021). In these pathogens, redox control is performed by trypanothione (Krauth-Siegel et al., 2003). There is extensive incorporation of cysteine during trypanothione synthesis and therefore cysteine biosynthesis within these parasites is critical to the production of this essential molecule (Canepa et al., 2009). The *de novo* pathway of cystine biosynthesis involves serine acetyltransferase and cysteine synthase to catalyse the conversion of serine to cysteine through an intermediate, O-acetyl serine (Nozaki et al., 2001).

Serine acetyltransferase and cysteine synthase have been previously identified as important potential drug targets for the treatment of Chagas disease and this finding can be extended to visceral leishmaniasis and to *Trypanosoma theileri* infection. This work aimed to unravel the molecular basis of cysteine production in trypanosomatids by determining multiple structures of CS from protozoan parasites. The secondary aim was to lay the foundation for future drug development, particularly structure based drug design. To achieve these aims, investigations will

be conducted into cysteine synthase, serine acetyltransferase and any interactions between the two enzymes through biochemical, biophysical and structural characterisation of the *de novo* cysteine biosynthesis pathway.

Investigation into *Trypanosoma cruzi* serine acetyltransferase

Serine acetyltransferase is responsible for the conversion of serine to OAS using acetyl CoA in cysteine biosynthesis. In this work, recombinant expression of *Tc*SAT was undertaken in *E. coli*. Despite the use of four different plasmids encoding different N-terminal fusion tags and the testing of a variety of expression hosts and expression conditions, *Tc*SAT remained insoluble. One potential explanation for this is that when expressed in *E. coli*, the protein is misfolded as *E. coli* is not the natural host, resulting in the formation of insoluble aggregates. An additional reason for the insoluble nature of *Tc*SAT is that the protein could be intrinsically hydrophobic or has exposed hydrophobic regions resulting in the production of aggregates.

Following the varied but unsuccessful attempts to obtain soluble *Tc*SAT, the protein was extracted from the insoluble fraction by unfolding and refolding. Circular dichroism suggested that the protein was refolded into the expected mixture of α-helices and β-sheets, however studies of the activity of refolded *Tc*SAT showed that the protein was not active. A refolded but inactive protein may be due to the lack of a potential co-factor, such as magnesium or potentially the presence of other proteins to form a complex. Additionally, some proteins require further post translational modifications to be activated such as glycosylation that are not performed in *E. coli* cells. The refolded protein was used in crystallisation trials yet no promising conditions were identified.

Determining the structure of cysteine synthase

Cysteine synthase catalyses the conversion of OAS to cysteine using a sulfide source and a PLP co-factor. In order to determine the structures of cysteine synthase from *T. cruzi, L. infantum and T. theileri*, all three proteins were expressed recombinantly in *E. coli* and successfully purified. Biochemical characterisation revealed that all three proteins were catalytically active and capable of completing the conversion reaction. For the proteins, crystals were produced through high throughput screening and for *Tc*CS and *Li*CS, crystals were optimised. Structures of *Tc*CS, *Li*CS and *Tth*CS were determined at 1.8 Å, 1.7 Å and 2.8 Å respectively and refined to excellent crystallographic statistics. Each structure represents different stages of the reaction with *Li*CS providing the apo form, *Tth*CS the holo form with the presence of the covalently bound PLP and *Tc*CS providing the structures of both the holo form and of a reaction intermediate with the presence of OAS in the active site. Overall, the three structures are shown to be very similar and share the canonical cysteine synthase fold. These structures provide evidence that in trypanosomatids, the molecular mechanism of cysteine synthesis occurs as expected. Due to the high similarity of the structures, any further findings using the structures can be applied to the other two proteins and beyond to the entire family of trypanosomatid cysteine synthases. This approach allows for a much wider structure based drug design campaign and increases the chances of success by using and comparing several proteins.

Investigation into the cysteine synthase complex

Previously, the cysteine synthase complex was identified in *L. major* through sequence comparison and purification of the complex, however no detailed studies

have been performed on trypanosomatids (Williams et al., 2009). In this work the successful expression and purification of *Tc*CSC is described through use of a pETDuet vector to co-express the complex. The CSC was crosslinked to maintain the complex and structural studies were undertaken. The complex was negatively stained, and transmission electron microscopy was performed. The TEM images resulted in the detection of protein present on the grids. To elucidate more information on this complex, such as the constitution of the complex requires further investigation. The activity of the CSC was studied with *Tc*CSC shown to be capable of completing both the individual stages of the reaction and the entire *de novo* synthesis pathway. This finding deviates from other CSC results with *Tc*CS activity remaining comparable to *Tc*CSC for the second stage of the reaction. Furthermore, this work provides the critical first steps to obtain a high resolution structure of this complex and to elucidate details of the complex on a molecular level.

Drug design using the cysteine synthase structures

Following the successful determination of three cysteine synthase structures, the structures were used in a range of drug design methods. To take advantage of the natural inhibition reported for homologous CS with molecules based on the C-terminus of the corresponding SAT sequence, 96 peptides based on the sequences of *Tc*SAT and of high sequence identity and crystallographic homologs. Studying the interaction of C-terminal *Tc*SAT peptides will further enhance the understanding of CSC interactions. Binding to the active site and the effect on activity were determined. The peptides showed significant increases and decreases in activity during these studies. Based on the mutations causing the most changes to activity,

it can be concluded that a smaller peptide containing basic amino acids is more effective at reducing activity. For the homolog *St*CS, this approach has been used to develop an effective inhibitor, UPAR415 (Marchetti et al., 2021). This provides additional support to the strategy of using C-terminal peptides to investigate inhibition of CS and suggests that further development of the *Tc*SAT based peptides shown to decrease *Tc*CS activity holds the potential to become potent inhibitors.

To explore small molecule binding to *Tc*CS and provide proof-of-princple experiments, a ^{19}F NMR based methodology was developed using available fluorinated molecules. This approach can be used to determine the binding of protein and potential ligands as fluorine is not typically found within proteins. The conditions for the experiment were optimised and negative controls established with three molecules. Under these experimental conditions, phenylalanine based fluorinated molecules are unlikely to prove good binders to *Tc*CS in *in vitro* situations. Further, characterisation of these molecules with *Tc*CS can be used in future experiments to provide a comparison to other molecules demonstrating successful binding.

In a *de novo* based approach to drug design, a fragment-based drug design campaign was undertaken using the FragLite library with *Tc*CS and *Li*CS (Wood et al., 2019). This approach resulted in 14 structures over the two proteins with 11 fragments bound. These fragments were then used in an AI-based fragment growing programme to produce a library of over 2000 potential compounds. The top 150 *de novo* molecules were manually inspected. Of these 150 compounds, several were eliminated due to unreasonable chemistries that would make

synthesis improbable. These results were used to improve the AI pipeline, for example by removing molecules with unsuitable chemistries such as azide groups or molecules clashing with the protein.

The library of 2000 compounds was used to search the Enamine REAL database resulting in the selection of 57 molecules. Following inspection of these 57 molecules, 16 compounds were inspected and selected for further exploration. From these 16 compounds, compound 37 (Z1662989717 from the Enamine REAL database) was determined to be the most promising candidate for further exploration.

Future work

The main achievement of this project is a deeper understanding of the *de novo* biosynthesis pathway in trypanosomatids. One way this has been achieved is the determination of three cysteine synthase structures. These high resolution structures have led to the exploration of inhibitors. Through a variety of methods, several molecules have been identified and have led to two main pathways of exploration. The first of these is a short, basic peptide based on the *Tc*SAT sequence. This motif showed the largest reduction in activity with *Tc*CS and represents a great potential to be developed further into a potent inhibitor. The other route to investigate is the molecules that have resulted from fragment screening. Following the 11 successfully determined fragments, molecules have been developed, with a promising candidate identified.

To enhance this work, further libraries of fragments such as the SpotXplorer or MiniFrags libraries could be explored though X-ray screening. This would provide

more information on the already determined binding sites and could result in additional binding pockets being determined and may lead to closer fragments that can be linked to provide an inhibitor. To complement determination of new fragments and to provide deeper understanding of the interactions between cysteine synthase and the fragments already determined to be bound, biophysical techniques should be explored. One such technique to be explored is isothermal titration calorimetry to confirm binding between protein and fragments and further to provide detailed information about the binding concentrations between the two.

Once the molecules have been obtained, *in vitro* testing using the established activity assay will be undertaken. Additional characterisation, such as studying the interactions between the proteins and molecules through spectral shift to obtain affinity constants, complemented by temperature related intensity change studies to measure the strength of the interaction. For both pathways, characterisation of the interactions with the molecules and *Tc*CS requires exploration with the establishment of binding constants. IC_{50} will be obtained for each molecule before the most promising of these compounds will undergo *in vivo* testing in *T. cruzi* and in infected mouse models.

To continue investigations into the CSC, further structural analysis is required. Small angle X-ray scattering could provide an indication of the stoichiometry of the complex and overall estimation of the secondary structure elements. To progress in the search for a high resolution structure of CSC, cryo-electron microscopy should be explored. The complex is of a suitable size for these studies and a variety of

molecules have been seen in the negative staining images indicating that an array of orientations are formed.

In summation, the work presented here forms a strong foundation for the development of inhibitors to prevent cysteine biosynthesis in trypanosomatids aiding in the treatment of several neglected diseases.

Chapter 8: References

Adachi, H., Takano, K., Morikawa, M., Kanaya, S., Yoshimura, M., Mori, Y., & Sasaki, T. (2003). Biological Crystallography Application of a two-liquid system to sitting-drop vapour-diffusion protein crystallization. *Acta Cryst*, *59*, 194–196.

Adler-Moore, J., & Proffitt, R. T. (2002). AmBisome: liposomal formulation, structure, mechanism of action and pre-clinical experience. *The Journal of Antimicrobial Chemotherapy*, *49*, 21–30.

Alcover, M. M., Ribas, A., Guillén, M. C., Berenguer, D., Tomás-Pérez, M., Riera, C., & Fisa, R. (2020). Wild mammals as potential silent reservoirs of *Leishmania infantum* in a Mediterranean area. *Preventive Veterinary Medicine*, *175*.

Alten, B., Maia, C., Odete Afonso, M., Campino, L., Jiménez, M., González, E., Molina, R., Laure Bañuls, A., Prudhomme, J., Vergnes, B., Toty, C., Cassan, C., Rahola, N., Thierry, M., Sereno, D., Bongiorno, G., Bianchi, R., Khoury, C., Tsirigotakis, N., … Gradoni, L. (2016). Seasonal Dynamics of Phlebotomine Sand Fly Species Proven Vectors of Mediterranean Leishmaniasis Caused by *Leishmania infantum*. *PLoS Neglected Tropical Diseases*, *10*(2), 1–22.

Anderson, A. C. (2003). The Process of Structure-Based Drug Design speed at which drug leads can be identified and evaluated *in silico*. *Chemistry & Biology*, *10*, 787–797.

Andrade, M. A., Chacón, P., Merelo, J. J., & Morán, F. (1993). Evaluation of secondary structure of proteins from UV circular dichroism spectra using an unsupervised learning neural network. *Protein Engineering*, *6*(4), 383–390.

Andreani, G., Ouellet, M., Menasria, R., Gomez, A. M., Barat, C., & Tremblay, M. J. (2015). *Leishmania infantum* Amastigotes Trigger a Subpopulation of Human B Cells with an Immunoregulatory Phenotype. *PLOS Neglected Tropical Diseases*, *9*(2), e0003543.

Annunziato, G., Spadini, C., Franko, N., Storici, P., Demitri, N., Pieroni, M., Flisi, S., Rosati, L., Iannarelli, M., Marchetti, M., Magalhaes, J., Bettati, S., Mozzarelli, A., Cabassi, C. S., Campanini, B., & Costantino, G. (2021). Investigational Studies on a Hit Compound Cyclopropane–Carboxylic Acid Derivative Targeting *O* -Acetylserine Sulfhydrylase as a Colistin Adjuvant. *ACS Infectious Diseases*, *7*(2), 281–292.

Aoun, K., & Bouratbine, A. (2014). Cutaneous Leishmaniasis in North Africa: a review. *Parasite*, *21*(14).

Baghban, R., Farajnia, S., Rajabibazl, M., Ghasemi, Y., Mafi, A. A., Hoseinpoor, R., Rahbarnia, L., & Aria, M. (2019). Yeast Expression Systems: Overview and Recent Advances. *Frontiers in Bioengineering and Biotechnology*, *7*(December).

Bancet, A., Raingeval, C., Lomberget, T., Le Borgne, M., Guichou, J.-F. O., & Krimm, I. (2020). Fragment Linking Strategies for Structure-Based Drug Design. *J. Med. Chem*, *63*, 11420–11435.

Banerjee, S., Ekka, M. K., & Kumaran, S. (2011). Comparative thermodynamic studies on substrate and product binding of O-Acetylserine Sulfhydrylase reveals two different ligand recognition modes. *BMC Biochemistry*, *12*(1).

Barbarin-Bocahu, I., & Graille, M. (2022). The X-ray crystallography phase problem solved thanks to AlphaFold and RoseTTAFold models: a case-study report . *Acta Crystallographica Section D : Structural Biology*, *78*(4).

Batool, M., Ahmad, B., & Choi, S. (2019). Molecular Sciences A Structure-Based Drug Discovery Paradigm. *International Journal of Molecular Sciences*, *20*(11).

Battista, T., Colotti, G., Ilari, A., & Fiorillo, A. (2020). Targeting Trypanothione Reductase, a Key Enzyme in the Redox Trypanosomatid Metabolism, to Develop New Drugs against Leishmaniasis and Trypanosomiases. *Molecules (Basel, Switzerland)*, *25*(8), 1924.

Bauman, J. D., Joe, J., Harrison, E. K., & Arnold, E. (2015). Rapid experimental SAD phasing and hot-spot identification with halogenated fragments. *IUCrJ*.

Bauman, J. D., Patel, D., & Arnold, E. (2012). Fragment Screening and HIV Therapeutics. *Topics in Current Chemistry*, *317*, 181.

Benoni, R., De Bei, O., Paredi, G., Hayes, C. S., Franko, N., Mozzarelli, A., Bettati, S., & Campanini, B. (2017). Modulation of *Escherichia coli* serine acetyltransferase catalytic activity in the cysteine synthase complex. *FEBS Letters*, *591*(9), 1212–1224.

Bergfors, T. (2007). Succeeding with seeding: some practical advice. *Evolving Methods for Macromolecular Crystallography*.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., & Bourne, P. E. (2000). The Protein Data Bank. *Nucleic Acids Research,* 28(1), 235-242.

Bhal, S. K. (2011). *LogP - Making Sense of the Value*. Advanced Chemistry Development Inc. (ACD/Labs).

Bieniek, M. K., Cree, B., Pirie, R., Horton, J. T., Tatum, N. J., Cole, D. J., & Cree, B. ✉. (2022). An open-source molecular builder and free energy preparation workflow. *Nature Communications Chemistry*, *5*(136).

Bogdanova, N., & Hell, R. (1997). Cysteine synthesis in plants: protein-protein interactions of serine acetyltransferase from *Arabidopsis thaliana*. *The Plant Journal*, *11*(2), 251–262.

Boiani, M., Piacenza, L., Hernández, P., Boiani, L., Cerecetto, H., González, M., & Denicola, A. (2010). Mode of action of nifurtimox and N-oxide-containing heterocycles against *Trypanosoma cruzi*: is oxidative stress involved? *Biochemical Pharmacology*, *79*(12), 1736–1745.

Bollag, G., Tsai, J., Zhang, J., Zhang, C., Ibrahim, P., Nolop, K., & Hirth, P. (2012). Vemurafenib: the first drug approved for BRAF-mutant cancer. *Nature Reviews Drug Discovery 2012 11:11*, *11*(11), 873–886.

Boratyn, G. M., Camacho, C., Cooper, P. S., Coulouris, G., Fong, A., Ma, N., Madden, T. L., Matten, W. T., McGinnis, S. D., Merezhuk, Y., Raytselis, Y., Sayers, E. W., Tao, T., Ye, J., & Zaretskaya, I. (2013). BLAST: a more efficient report with usability improvements. *Nucleic Acids Research*, *41*(Web Server issue), W29.

Broennimann, Ch., Eikenberry, E. F., Henrich, B., Horisberger, R., Huelsen, G., Pohl, E., Schmitt, B., Schulze-Briese, C., Suzuki, M., Tomizaki, T., Toyokawa, H., &

Wagner, A. (2006). The PILATUS 1M detector. *Journal of Synchrotron Radiation*, *13*(2), 120–130.

Brotánková, A., Fialová, M., Čepička, I., Brzoňová, J., & Svobodová, M. (2022). Trypanosomes of the *Trypanosoma theileri* Group: Phylogeny and New Potential Vectors. *Microorganisms*, *10*(2).

Brünger, A. T. (1992). Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature 1992 355:6359*, *355*(6359), 472–475.

Burza, S., Croft, S. L., & Boelaert, M. (2018). Leishmaniasis. *The Lancet*, *392*(10151), 951–970.

Canepa, G. E., Bouvier, L. A., Miranda, M. R., Uttaro, A. D., & Pereira, C. A. (2009). tCharacterization of Trypanosoma cruzi L-cysteine transport mechanisms and their adaptive regulation. *FEMS Microbiology Letters*, *292*(1), 27–32.

Caroline R. Buchholza, & William C. K. Pomerantz. (2021). 19F NMR viewed through two different lenses: ligand-observed and protein-observed 19F NMR applications for fragment-based drug discovery. *RSC Chemical Biolog*, *2*(1312).

Casanas, A., Warshamanage, R., Finke, A. D., Panepucci, E., Olieric, V., Nöll, A., Tampé, R., Brandstetter, S., Förster, A., Mueller, M., Schulze-Briese, C., Bunk, O., & Wang, M. (2016). EIGER detector: Application in macromolecular crystallography: *Acta Crystallographica Section D: Structural Biology*, *72*(9), 1036–1048.

Chilingaryan, Z., Yin, Z., & Oakley, A. J. (2012). Fragment-Based Screening by Protein Crystallography: Successes and Pitfalls. *Int. J. Mol. Sci*, *13*, 12857–12879.

Chimera, E. T., Fosgate, G. T., Etter, E. M. C., Boulangé, A., Vorster, I., & Neves, L. (2021). A one health investigation of pathogenic trypanosomes of cattle in Malawi. *Preventive Veterinary Medicine*, *188*, 105255.

Cook, P. F., Tai, C. H., Hwang, C. C., Woehl, E. U., Dunn, M. F., & Schnackerz, K. D. (1996). Substitution of pyridoxal 5'-phosphate in the O-acetylserine sulfhydrylase from *Salmonella typhimurium* by cofactor analogs provides a test of the mechanism proposed for formation of the α-aminoacrylate intermediate. *Journal of Biological Chemistry*, *271*(42), 25842–25849.

Cornish, K. A. S., Lange, J., Aevarsson, A., & Pohl, E. (2022). CPR-C4 is a highly conserved novel protease from the Candidate Phyla Radiation with remote structural homology to human vasohibins. *Journal of Biological Chemistry*, *298*(5), 101919.

Costa, S., Almeida, A., Castro, A., & Domingues, L. (2014). Fusion tags for protein solubility, purification, and immunogenicity in *Escherichia coli*: The novel Fh8 system. *Frontiers in Microbiology*, *5*(FEB), 63.

Coura, J. R. (2015). The main sceneries of chagas disease transmission. The vectors, blood and oral transmissions - A comprehensive review. *Memorias Do Instituto Oswaldo Cruz*, *110*(3), 277–282.

Coura, J. R., & Borges-Pereira, J. (2011). Chronic phase of Chagas disease: Why should it be treated? A comprehensive review. *Memorias Do Instituto Oswaldo Cruz*, *106*(6), 641–645.

Cowtan, K. (2003). Phase Problem in X-ray Crystallography, and Its Solution. *ELS*, *5*.

Cowtan, K. (2006). The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallographica Section D: Biological Crystallography*, *62*(9), 1002–1011.

Cowtan, K. (2010). Recent developments in classical density modification. *Acta Crystallographica Section D: Biological Crystallography*, *66*(4), 470–478.

Croft, S. L., Sundar, S., & Fairlamb, A. (2006). Drug Resistance in Leishmaniasis. *Clinical Microbiology Reviews*, *19*(1), 111–126.

Dalvit, C., Fagerness, P. E., Hadden, D. T. A., Sarver, R. W., & Stockman, B. J. (2003). Fluorine-NMR experiments for high-throughput screening: Theoretical aspects, practical considerations, and range of applicability. *Journal of the American Chemical Society*, *125*(25), 7696–7703.

Dalvit, C., & Vulpetti, A. (2012). Technical and practical aspects of ^{19}F NMR-based screening: Toward sensitive high-throughput screening with rapid deconvolution. *Magnetic Resonance in Chemistry*, *50*(9), 592–597.

Dalvit, C., & Vulpetti, A. (2019). Ligand-Based Fluorine NMR Screening: Principles and Applications in Drug Discovery Projects. *Journal of Medicinal Chemistry*, *62*(5), 2218–2244.

Dauter, Z., Jaskolski, M., & Wlodawer, A. (2010). Impact of synchrotron radiation on macromolecular crystallography: a personal view. *Journal of Synchrotron Radiation*, *17*(Pt 4), 433.

Dessau, M. A., & Modis, Y. (2011). Protein Crystallization for X-ray Crystallography. *Journal of Visualized Experiments*.

Droux, M., Ruffet, M., Douce, R., & Job, D. (1998). Interactions between serine acetyltransferase and O-acetylserine (thiol) lyase in higher plants--structural and kinetic properties of the free and bound enzymes. *European Journal of Biochemistry*, *255*(1).

Du, F., Liu, Y.-Q., Xu, Y.-S., Li, Z.-J., Wang, Y.-Z., Zhang, Z.-X., & Sun, X.-M. (2021). Regulating the T7 RNA polymerase expression in *E. coli* BL21 (DE3) to provide more host options for recombinant protein production. *Microbial Cell Factories*, *20*(1), 189.

Dumas, C., Ouellette, M., Tovar, J., Cunningham, M. L., Fairlamb, A. H., Tamar, S., Olivier, M., & Papadopoulou, B. (1997). Disruption of the trypanothione reductase gene of *Leishmania* decreases its ability to survive oxidative stress in macrophages. *The EMBO Journal*, *16*(10), 2590.

Emsley, P., Lohkamp, B., Scott, W. G., & Cowtan, K. (2010). Features and development of Coot. *Acta Crystallographica Section D: Biological Crystallography*, *66*(4), 486–501.

Evans, P., & Mccoy, A. (2008). Biological Crystallography An introduction to molecular replacement. *Acta Cryst D*, *64*, 1–10.

Fathi-Roudsari, M., Akhavian-Tehrani, A., & Maghsoudi, N. (2015). Comparison of Three *Escherichia coli* Strains in Recombinant Production of Reteplase. *Avicenna Journal of Medical Biotechnology*, *8*(1), 16–22.

Fentahun, T. (2020). Molecular detection of *Trypanosoma theileri* and a new Trypanocidal drug, a review. *Journal of Life Science and Biomedicine*, *10*(3), 29–43.

Forsyth, C. J., Hernandez, S., Olmedo, W., Abuhamidah, A., Traina, M. I., Sanchez, D. R., Soverow, J., & Meymandi, S. K. (2016). Safety Profile of Nifurtimox for Treatment of Chagas Disease in the United States. *Clinical Infectious Diseases.*

Francis, S., Frank, C., Buchanan, L., Green, S., Stennett-Brown, R., Gordon-Strachan, G., Rubio-Palis, Y., Grant, C., Alexander-Lindo, R. L., Nwokocha, C., Robinson, D., & Delgoda, R. (2021). Challenges in the control of neglected insect vector diseases of human importance in the Anglo-Caribbean. *One Health*, *13*.

Francisco, A. F., Jayawardhana, S., Olmo, F., Lewis, M. D., Wilkinson, S. R., Taylor, M. C., & Kelly, J. M. (2020). Challenges in Chagas Disease Drug Development. *Molecules*, *25*(12).

Frazão, C., Sieker, L., Sheldrick, G., Lamzin, V., LeGall, J., & Carrondo, M. A. (1999). *Ab initio* structure solution of a dimeric cytochrome c3 from *Desulfovibrio gigas* containing disulfide bridges. *Journal of Biological Inorganic Chemistry* , *4*(2), 162–165.

Frézard, F., Demicheli, C., & Ribeiro, R. R. (2009). Pentavalent Antimonials: New Perspectives for Old Drugs. *Molecules*, *14*, 2317–2336.

Fyfe, P. K., Westrop, G. D., Ramos, T., Müller, S., Coombs, G. H., & Hunter, W. N. (2012). Structure of *Leishmania major* cysteine synthase. *Acta*

Crystallographica. Section F, Structural Biology and Crystallization Communications, *68*, 738–743.

Gaitonde, M. K. (1967). A Spectrophotometric Method for the Direct Determination of Cysteine in the Presence of Other Naturally Occurring Amino Acids. *Biochem. J*, *104*, 627.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). Protein Identification and Analysis Tools on the ExPASy Server. In *The Proteomics Protocols Handbook* (pp. 571–607). Humana Press.

Gorgulla, C., Boeszoermenyi, A., Wang, Z.-F., Fischer, P. D., Coote, P. W., Das, K. M. P., Malets, Y. S., Radchenko, D. S., Moroz, Y. S., Scott, D. A., Fackeldey, K., Hoffmann, M., Iavniuk, I., Wagner, G., & Arthanari, H. (2020). An open-source drug discovery platform enables ultra-large virtual screens. *Nature*, *580*, 663.

Gray, D. (1997). Overview of Protein Expression by Mammalian Cells. *Current Protocols in Protein Science*, *10*(1), 591.

Green, D. W., Ingram, V. M., & Perutz, M. F. (1954). The structure of haemoglobin - IV. Sign determination by the isomorphous replacement method. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, *225*(1162), 287–307.

Guan, R., Roderick, S. L., Huang, B., & Cook, P. F. (2008). Roles of Histidines 154 and 189 and Aspartate 139 in the Active Site of Serine Acetyltransferase from *Haemophilus influenzae*. *Biochemistry*, *47*(24), 6322.

Haas, D. J. (1968). X-ray studies on lysozyme crystals at –50°C. *IUCr*, *24*(604), 604–604.

Hall, B. S., & Wilkinson, S. R. (2012). Activation of benznidazole by trypanosomal type I nitroreductases results in glyoxal formation. *Antimicrobial Agents and Chemotherapy*, *56*(1), 115–123.

Hara, S., Payne, M. A., Schnackerzj, K. D., & Cook, P. F. (1990). A rapid purification procedure and computer-assisted sulfide ion selective electrode assay for O-acetylserine sulfhydrylase from *salmonella typhimurium*. *Protein Expression and Purification*, *1*(1), 70–76.

Harper, S., & Speicher, D. W. (2011). Purification of proteins fused to glutathione S-transferase. *Methods in Molecular Biology (Clifton, N.J.)*, *681*, 259–280.

Hauptman, H. A. (1991). The phase problem of X-ray crystallography. *Reports on Progress in Physics*, *54*(11), 1427.

Headd, J. J., Echols, N., Afonine, P. V., Moriarty, N. W., Gildea, R. J., & Adams, P. D. (2014). Flexible torsion-angle noncrystallographic symmetry restraints for improved macromolecular structure refinement. *Acta Crystallographica Section D: Biological Crystallography*, *70*(5), 1346–1356.

Helliwell, J. R., & Mitchell, E. P. (2015). Synchrotron radiation macromolecular crystallography: Science and spin-offs. *IUCrJ*, *2*(Pt 2), 283–291.

Hemmige, V., Tanowitz, H., & Sethi, A. (2012). *Trypanosoma cruzi* infection: a review with emphasis on cutaneous manifestations. *Int J Dermatol*, *51*(5), 501–508.

Huang, B., Vetting, M. W., & Roderick, S. L. (2005). The Active Site of O-Acetylserine Sulfhydrylase Is the Anchor Point for Bienzyme Complex Formation with Serine Acetyltransferase. *Journal of Bacteriology*, *187*(9), 3201.

Hudson, S. A., McLean, K. J., Surade, S., Yang, Y. Q., Leys, D., Ciulli, A., Munro, A. W., & Abell, C. (2012). Application of fragment screening and merging to the discovery of inhibitors of the *Mycobacterium tuberculosis* cytochrome P450 CYP121. *Angewandte Chemie - International Edition*, *51*(37), 9311–9316.

Hunt, I. (2005). From gene to protein: a review of new and enabling technologies for multi-parallel protein expression. *Protein Expression and Purification*, *40*(1), 1–22.

Jang, K., Kim, H. G., Hlaing, S. H. S., Kang, M., Choe, H. W., & Kim, Y. J. (2022). A Short Review on Cryoprotectants for 3D Protein Structure Analysis. *Crystals*, *12*(2), 138.

Jhamb, K., & Sahoo, D. K. (2012). Production of soluble recombinant proteins in *Escherichia coli*: Effects of process conditions and chaperone co-expression on cell growth and production of xylanase. *Bioresource Technology*, *123*, 135–143.

Jia, B., & Jeon, C. O. (2016). High-throughput recombinant protein expression in *Escherichia coli*: current status and future perspectives. *Open Biology*, *6*(160196).

John, R. A. (1995). Pyridoxal phosphate-dependent enzymes. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology*, *1248*(2), 81–96.

Kelly, S., Ivens, A., Mott, G. A., O'Neill, E., Emms, D., Macleod, O., Voorheis, P., Tyler, K., Clark, M., Matthews, J., Matthews, K., & Carrington, M. (2017). An alternative strategy for trypanosome survival in the mammalian bloodstream revealed through genome and transcriptome analysis of the ubiquitous bovine parasite *Trypanosoma* (megatrypanum) *theileri*. *Genome Biology and Evolution*, *9*(8), 2093–2109.

Kendrew, J. C., Dickerson, R. E., Strandberg, B. E., Hart, R. G., Davies, D. R., Phillips, D. C., & Shore, V. C. (1960). Structure of myoglobin: A three-dimensional Fourier synthesis at 2 A. resolution. *Nature*, *185*(4711), 422–427.

Kimple, M. E., Brill, A. L., & Pasker, R. L. (2013). Overview of Affinity Tags for Protein Purification. *Current Protocols in Protein Science*, *73*(1), 9.9.1-9.9.23.

King, G. J., Chen, K. E., Robin, G., Forwood, J. K., Heras, B., Thakur, A. S., Kobe, B., Blomberg, S. P., & Martin, J. L. (2009). Interaction between Plate Make and Protein in Protein Crystallisation Screening. *PLoS ONE*, *4*(11).

Kirsch, P., Hartman, A. M., Hirsch, A. K. H., Empting, M., A., De, A. (2019). Concepts and Core Principles of Fragment-Based Drug Design. *Molecules.*

Kratz, J. M., Bournissen, G., Forsyth, C. J., Sosa-Estani, S., & Müller Kratz, J. (2018). Clinical and pharmacological profile of benznidazole for treatment of Chagas disease. *Expert Review of Clinical Pharmacology.*

Krause, L., Tolborg, K., Grønbech, T. B. E., Sugimoto, K., Iversen, B. B., & Overgaard, J. (2020). Accurate high-resolution single-crystal diffraction data from a Pilatus3 X CdTe detector. *Journal of Applied Crystallography*, *53*(Pt 3), 635–649.

Krauth-Siegel, R. L., & Comini, M. A. (2008). Redox control in trypanosomatids, parasitic protozoa with trypanothione-based thiol metabolism. *Biochimica et Biophysica Acta (BBA) - General Subjects*, *1780*(11), 1236–1248.

Krauth-Siegel, R. L., Meiering, S. K., & Schmidt, H. (2003). The Parasite-Specific Trypanothione Metabolism of *Trypanosoma* and *Leishmania*. *Biological Chemistry*, *384*(4), 539–549.

Kredich, N. M., Becker, M. A., & Tomkins, G. M. (1969). Purification and Characterization of a cysteine synthetase, a Bifunctional Protein Complex, from *Salmonella typhimurium*. *Journal of Biological Chemistry*, *244*(9), 2428–2439.

Krissinel, E., & Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. *Journal of Molecular Biology*, *372*, 774–797.

Kronenberger, T., Lindner, J., Meissner, K. A., Zimbres, F. M., Coronado, M. A., Sauer, F. M., Schettert, I., & Wrenger, C. (2014). Vitamin B6-Dependent Enzymes in the Human Malaria Parasite *Plasmodium falciparum*: A Druggable Target? *BioMed Research International*, 1(11).

Kumar, S., Raj, I., Nagpal, I., Subbarao, N., & Gourinath, S. (2011). Structural and biochemical studies of serine acetyltransferase reveal why the parasite *Entamoeba histolytica* cannot form a cysteine synthase complex. *Journal of Biological Chemistry*, *286*(14), 12533–12541.

Kumaran, S., Yi, H., Krishnan, H. B., & Jez, J. M. (2009). Assembly of the Cysteine Synthase Complex and the Regulatory Role of Protein-Protein Interactions. *The Journal of Biological Chemistry*, *284*(15), 10268–10275.

Kuzin, A. P., Chen, Y., Seetharaman, J., Sahdev, S., Wang, D., Mao, L., Cunningham, K., Maglaqui, M., Xiao, R., Liu, J., Everett, J. K., Acton, T. B., Rost, B., Montelione, G. T., Hunt, J. F., Tong, L., & Northeast Structural Genomics Consortium (NESG). (2008). *RCSB PDB - 3F1X: Three dimensional structure of the serine acetyltransferase from Bacteroides vulgatus, NORTHEAST STRUCTURAL GENOMICS CONSORTIUM TARGET BVR62.*

Lancet Neurology. (2009). Leading Edge Chagas disease: the forgotten American neuroinfection. *Lancet Neurology, The*, *8*.

Lescure, F. X., Le Loup, G., Freilij, H., Develoux, M., Paris, L., Brutus, L., & Pialoux, G. (2010). Chagas disease: changes in knowledge and management. *The Lancet Infectious Diseases*, *10*(8), 556–570.

Leu, L.-S., & Cook, P. F. (1994). Kinetic Mechanism of Serine Transacetylase from *Salmonella typhimurium*. *Biochemistry*, *33*, 2667–2671.

Li, Q. (2020). Application of Fragment-Based Drug Discovery to Versatile Targets. *Frontiers in Molecular Biosciences*, *7*(180).

Liang, J., Han, Q., Tan, Y., Ding, H., & Li, J. (2019). Current advances on structure-function relationships of pyridoxal 5'-phosphate-dependent enzymes.*Frontiers in Molecular Biosciences,* 6(MAR).

Lichty, J. J., Malecki, J. L., Agnew, H. D., Michelson-Horowitz, D. J., & Tan, S. (2005). Comparison of affinity tags for protein purification. *Protein Expression and Purification*, *41*(1), 98–105.

Lidani, K. C. F., Andrade, F. A., Bavia, L., Damasceno, F. S., Beltrame, M. H., Messias-Reason, I. J., & Sandri, T. L. (2019). Chagas disease: From discovery to a worldwide health problem. *Journal of Physical Oceanography*, *49*(6), 1–13.

Lindquist-Kleissler, B., & Johnstone, T. C. (2023). Models of the putative antimony(v)–diolate motifs in antileishmanial pentavalent antimonial drugs. *Dalton Transactions*, *52*(27), 9229–9237.

Lobley, A., Whitmore, L., & Wallace, B. A. (2002). DICHROWEB: an interactive website for the analysis of protein secondary structure from circular dichroism spectra. *Bioinformatics (Oxford, England)*, *18*(1), 211–212.

Lobstein, J., Emrich, C. A., Jeans, C., Faulkner, M., Riggs, P., & Berkmen, M. (2016). SHuffle, a novel *Escherichia coli* protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm. *Microbial Cell Factories*, *15*(1), 1–16.

Madeira, F., Park, Y. M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A. R. N., Potter, S. C., Finn, R. D., & Lopez, R. (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Research*, *47*(W1), W636–W641.

Mahapatra, M. K., & Karuppasamy, M. (2022). Fundamental considerations in drug design. *Computer Aided Drug Design (CADD): From Ligand-Based Methods to Structure-Based Approaches*, 17.

Mann, S., Frasca, K., Scherrer, S., Henao-Martínez, A. F., Newman, S., Ramanan, P., & Suarez, J. A. (2021). A Review of Leishmaniasis: Current Knowledge and Future Directions. *Current Tropical Medicine Reports*, *8*(2), 121–132.

Marchetti, M., De Angelis, F. S., Annunziato, G., Costantino, G., Pieroni, M., Ronda, L., Mozzarelli, A., Campanini, B., Cannistraro, S., Bizzarri, A. R., & Bettati, S. (2021). A Competitive O-Acetylserine Sulfhydrylase Inhibitor Modulates the Formation of Cysteine Synthase Complex. *Catalysts, 11*(6), 700.

Marciano, D., Santana, M., & Nowicki, C. (2012). Functional characterization of enzymes involved in cysteine biosynthesis and H_2S production in *Trypanosoma cruzi. Molecular and Biochemical Parasitology*, *185*(2), 114–120.

Masmoudi, A., Hariz, W., Marrekchi, S., Amouri, M., & Turki, H. (2013). Old World cutaneous leishmaniasis: diagnosis and treatment. *Journal of Dermatological Case Reports*, *7*(2), 31–41.

Matlashewski, G. (2001). *Leishmania* infection and virulence. *Med Microbiol Immunol* , *190*, 37–42.

Matthews, B. W. (1968). Solvent content of protein crystals. *Journal of Molecular Biology*, *33*(2), 491–497.

Mayta, H., Romero, Y. K., Pando, A., Verastegui, M., Tinajeros, F., Bozo, R., Henderson-Frost, J., Colanzi, R., Flores, J., Lerner, R., Bern, C., & Gilman, R. H. (2019). Improved DNA extraction technique from clot for the diagnosis of Chagas disease. *PLoS Neglected Tropical Diseases*, *13*(1).

Mccoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., & Read, R. J. (2007). Phaser crystallographic software. *J. Appl. Cryst*, *40*, 658–674.

McNicholas, S., Potterton, E., Wilson, K. S., & Noble, M. E. M. (2011). Presenting your structures: The CCP4mg molecular-graphics software. *Acta Crystallographica Section D: Biological Crystallography*, *67*(4), 386–394.

McPherson, A., & Gavira, J. A. (2013). Introduction to protein crystallization. *Acta Crystallographica Section F*, *70*, 2–20.

Miles, A. J., Ramalli, S. G., & Wallace, B. A. (2022). DichroWeb, a website for calculating protein secondary structure from circular dichroism spectroscopic data. *Protein Science : A Publication of the Protein Society*, *31*(1), 37–46.

Moore, E. M., & Lockwood, D. N. (2010). Treatment of Visceral Leishmaniasis. *Journal of Global Infectious Diseases*, *2*(2), 151.

Mori, M., Jeelani, G., Masuda, Y., Sakai, K., Tsukui, K., Waluyo, D., Watanabe, Y., Nonaka, K., & Matsumoto, A. (2015). Identification of natural inhibitors of *Entamoeba histolytica* cysteine synthase from microbial secondary metabolites. *Frontiers in Microbiology, 6.*

Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F., & Vagin, A. A. (2011). REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D: Biological Crystallography*, *67*(4), 355–367.

Norton, R. S., Leung, E. W. W., Chandrashekaran, I. R., & Macraild, C. A. (2016). Applications of ^{19}F-NMR in Fragment-Based Drug Discovery. *Molecules* .

Nozaki, T., Asai, T., Sanchez, L. B., Kobayashi, S., Nakazawa, M., & Takeuchi, T. (1999). Characterization of the Gene Encoding Serine Acetyltransferase, a Regulated Enzyme of Cysteine Biosynthesis from the Protist Parasites *Entamoeba histolytica* and *Entamoeba dispar*. *Journal of Biological Chemsitry*, 274(45).

Nozaki, T., Shigeta, Y., Saito-Nakano, Y., Imada, M., & Kruger, W. D. (2001). Characterization of transsulfuration and cysteine biosynthetic pathways in the protozoan hemoflagellate, *Trypanosoma cruzi*. *Journal of Biological Chemistry*, *276*(9), 6516–6523.

Nunes, M. C. P., Beaton, A., Acquatella, H., Bern, C., Bolger, A. F., Echeverría, L. E., Dutra, W. O., Gascon, J., Morillo, C. A., Oliveira-Filho, J., Ribeiro, A. L. P., & Marin-Neto, J. A. (2018). Chagas Cardiomyopathy: An Update of Current Clinical Knowledge and Management: A Scientific Statement From the American Heart Association. *Circulation*, *138*(12), e169–e209.

Oldham, K. E. A., Prentice, E. J., Summers, E. L., & Hicks, J. L. (2022). Serine acetyltransferase from *Neisseria gonorrhoeae*; structural and biochemical basis of inhibition. *Biochemical Journal*, *479*(1), 57.

Olivera, M. J., Cucunubá, Z. M., Álvarez, C. A., & Nicholls, R. S. (2015). Safety Profile of Nifurtimox and Treatment Interruption for Chronic Chagas Disease in Colombian Adults. *Am. J. Trop. Med. Hyg*, *93*(6), 1224–1230.

Olsen, L. R., Huang, B., Vetting, M. W., & Roderick, S. L. (2004). Structure of Serine Acetyltransferase in Complexes with CoA and Its Cysteine Feedback Inhibitor . *Biochemistry*, *43*, 6013–6019.

O'Neill, P., Stevens, D. L., & Garman, E. F. (2002). Physical and chemical considerations of damage induced in protein crystals by synchrotron radiation: a radiation chemical perspective. *Journal of Synchrotron Radiation*, *9*(Pt 6), 329–332.

Page, R., Grzechnik, S. K., Canaves, J. M., Spraggon, G., Kreusch, A., Kuhn, P., Stevens, R. C., & Lesley, S. A. (2003). Shotgun crystallization strategy for structural genomics: An optimized two-tiered crystallization screen against the Thermotoga maritima proteome. *Acta Crystallographica - Section D Biological Crystallography*, *59*(6), 1028–1037.

Patterson, S., & Wyllie, S. (2014). Nitro drugs for the treatment of trypanosomatid diseases: past, present, and future prospects. *Trends in Parasitology*, *30*(6), 289–298.

Pearce, N. M., Bradley, A. R., Krojer, T., Marsden, B. D., Deane, C. M., & Delft, F. Von. (2017). Partial-occupancy binders identified by the Pan-Dataset Density Analysis method offer new chemical opportunities and reveal cryptic binding sites. *Structural Dynamics*, *4*(3).

Pellecchia, M., Sem, D. S., & Wüthrich, K. (2002). NMR in drug discovery. *Nature Reviews Drug Discovery 2002 1:3*, *1*(3), 211–219.

Pereira, P. C. M., & Navarro, E. C. (2013). Challenges and perspectives of Chagas disease: a review. *The Journal of Venomous Animals and Toxins Including Tropical Diseases*, *19*(1).

Perera, T. P. S., Jovcheva, E., Mevellec, L., Vialard, J., De Lange, D., Verhulst, T., Paulussen, C., Van De Ven, K., King, P., Freyne, E., Rees, D. C., Squires, M., Saxty, G., Page, M., Murray, C. W., Gilissen, R., Ward, G., Thompson, N. T., Newell, D. R., … Lorenzi, M. V. (2017). Discovery & pharmacological characterization of JNJ-42756493 (Erdafitinib), a functionally selective small-molecule FGFR family inhibitor. *Molecular Cancer Therapeutics*, *16*(6), 1010–1020.

Peterson, J. K., Bakuza, J., & Standley, C. J. (2021). One health and neglected tropical diseases—multisectoral solutions to endemic challenges. *Tropical Medicine and Infectious Disease*, *6*(1), 6–8.

Pflugrath, J. W. (2015). Practical macromolecular cryocrystallography. *Acta Crystallographica. Section F, Structural Biology Communications*, *71*(Pt 6), 622.

Pieroni, M., Annunziato, G., Beato, C., Wouters, R., Benoni, R., Campanini, B., Pertinhez, T. A., Bettati, S., Mozzarelli, A., & Costantino, G. (2016). Rational Design, Synthesis, and Preliminary Structure−Activity Relationships of α-Substituted-2-Phenylcyclopropane Carboxylic Acids as Inhibitors *of Salmonella typhimurium* O-Acetylserine Sulfhydrylase. *Journal of Medicinal Chemistry*, 59(6), 2567-2578.

Pinto-Martinez, A. K., Rodriguez-Durán, J., Serrano-Martin, X., Hernandez-Rodriguez, V., & Benaim, G. (2018). Mechanism of Action of Miltefosine on *Leishmania donovani* Involves the Impairment of Acidocalcisome Function and the Activation of the Sphingosine-Dependent Plasma Membrane Ca2+ Channel. *Antimicrobial Agents and Chemotherapy*, *62*(1).

Potterton, L., Agirre, J., Ballard, C., Cowtan, K., Dodson, E., Evans, P. R., Jenkins, H. T., Keegan, R., Krissinel, E., Stevenson, K., Lebedev, A., McNicholas, S. J., Nicholls, R. A., Noble, M., Pannu, N. S., Roth, C., Sheldrick, G., Skubak, P., Turkenburg, J., … Wojdyr, M. (2018). CCP4i2: The new graphical user interface to the CCP 4 program suite. *Acta Crystallographica Section D: Structural Biology*, *74*, 68–84.

Pye, V. E., Tingey, A. P., Robson, R. L., & Moody, P. C. E. (2004). The structure and mechanism of serine acetyltransferase from *Escherichia coli*. *Journal of Biological Chemistry*, *279*(39), 40729–40736.

Raj, I., Kumar, S., & Gourinath, S. (2012). The narrow active-site cleft of O-acetylserine sulfhydrylase from Leishmania donovani allows complex formation with serine acetyltransferases with a range of C-terminal sequences. *Acta Crystallographica Section D: Biological Crystallography*, *68*(8), 909–919.

Richter, J., Hanus, I., Häussinger, D., Löscher, T., & Harms, G. (2011). Mucosal Leishmania infantum infection. *Parasitology Research*, *109*(3), 959–962.

Riddles, P. W., Blakeley, R. L., & Zerner, B. (1983). Reassessment of Ellman's reagent. *Methods in Enzymology*, *91*(C), 49–60.

Robert, X., & Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Research*, *42*, W320.

Rodger, A., & Marshall, D. (2021). Beginner's Guide to circular dichroism. *The Biochemist*, *43*(2), 58–64.

Romero, I., Téllez, J., Eiko Yamanaka, L., Steindel, M., José Romanha, A., & Carlos Grisard, E. (2014). Transsulfuration is an active pathway for cysteine biosynthesis in *Trypanosoma rangeli,* Parasites and Vectors, 197, 7(1).

Romero, I., Téllez, J., Romanha, A. J., Steindel, M., & Grisard, E. C. (2015). Upregulation of Cysteine Synthase and Cystathionine β-Synthase Contributes to *Leishmania braziliensis* Survival under Oxidative Stress. *Antimicrobial Agents and Chemotherapy*, *59*(8), 4770–4781.

Rosa, B., Dickinson, E. R., Marchetti, M., Campanini, B., Pioselli, B., Bettati, S., & Dyrberg Rand, K. (2021). Revealing the Dynamic Allosteric Changes Required for Formation of the Cysteine Synthase Complex by Hydrogen-Deuterium Exchange MS. *Mol Cell Proteomics*, *20*, 100098.

Rosa, B., Marchetti, M., Paredi, G., Amenitsch, H., Franko, N., Benoni, R., Giabbai, B., De Marino, M. G., Mozzarelli, A., Ronda, L., Storici, P., Campanini, B., & Bettati, S. (2019). Combination of SAXS and Protein Painting Discloses the Three-Dimensional Organization of the Bacterial Cysteine Synthase Complex, a Potential Target for Enhancers of Antibiotic Action. *International Journal of Molecular Sciences*, *20*(20).

Rosano, G. L., & Ceccarelli, E. A. (2014). *Recombinant protein expression in Escherichia coli: advances and challenges*.

Saha, A. K., Mukherjee, T., & Bhaduri, A. (1986). Mechanism of action of amphotericin B on *Leishmania donovani* promastigotes. *Molecular and Biochemical Parasitology*, *19*(3), 195–200.

Salsi, E., Alexander Bayden S., Francesca Spyrakis, Amadasi, A., Campanini, B., Bettati, S., Dodatko, T., Cozzini, P., Kellogg, G. E., Cook, P. F., Roderick, S. L., & Mozzarelli, A. (2010). Design of O-Acetylserine Sulfhydrylase Inhibitors by Mimicking Nature. *J. Med. Chem*, *53*, 345.

Schrödinger LLC. (2015). *The PyMOL Molecular Graphics System, Version~1.8*.

Shabalin, I. G., Porebski, P. J., & Minor, W. (2018). Refining the macromolecular model-achieving the best agreement with the data from X-ray diffraction experiment. *Crystallogr Rev*, *24*(4), 236–262.

Sheldrick, G. M. (1990). Phase annealing in SHELX-90: direct methods for larger structures. *Acta Crystallographica A*, *46*(6), 467–473.

Sheldrick, G. M., Gilmore, C. J., Hauptman, H. A., Weeks, C. M., Miller, R., & Usón, I. (2012). *Ab initio* phasing. *Acta Crystallographica F*, 413–432.

Shuker, S. B., Hajduk, P. J., Meadows, R. P., & Fesik, S. W. (1996). Discovering High-Affinity Ligands for Proteins: SAR by NMR. *Science*, *274*(5292), 1531–1534.

Singh, A., Upadhyay, V., Upadhyay, A. K., Singh, S. M., & Panda, A. K. (2015). Protein recovery from inclusion bodies of *Escherichia coli* using mild solubilization process. *Microbial Cell Factories*, *14*(1), 41.

Singh, S., & Sivakumar, R. (2005). Challenges and new discoveries in the treatment of leishmaniasis. *Article in Journal of Infection and Chemotherapy*.

Skarina, T., Xu, X., Evdokimova, E., & Savchenko, A. (2014). High-throughput crystallization screening. *Methods in Molecular Biology*, *1140*, 159–168.

Śledź, P., & Caflisch, A. (2018). Protein structure-based drug design: from docking to molecular dynamics. *Current Opinion in Structural Biology*, *48*, 93–102.

Smyth, M. S., & Martin, J. H. J. (2000). X-Ray crystallography. *Molecular Pathology*, *53*(1), 8.

Sobolev, O. V., Afonine, P. V., Moriarty, N. W., Hekkelman, M. L., Joosten, R. P., Perrakis, A., & Adams, P. D. (2020). A global Ramachandran score identifies protein structures with unlikely stereochemistry. *Structure*, *28*(11), 1249.

Sowerby, K., Freitag-Pohl, S., Murillo, A. M., Silber, A. M., & Pohl, E. (2023). Cysteine synthase: Multiple structures of a key enzyme in cysteine synthesis and a potential drug target for Chagas disease and Leishmaniasis. *Acta Crystallographica Section D Structural Biology*, *79*(6), 1–46.

Studier, F. W., & Moffatt, B. A. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *Journal of Molecular Biology*, *189*(1), 113–130.

Suganuma, K., Kayano, M., Kida, K., Gröhn, Y. T., Miura, R., Ohari, Y., Mizushima, D., & Inoue, N. (2022). Genetic and seasonal variations of *Trypanosoma theileri* and the association of *Trypanosoma theileri* infection with dairy cattle productivity in Northern Japan. *Parasitology International*, *86*, 102476.

Sun, P., Tropea, J. E., & Waugh, D. S. (2011). Enhancing the Solubility of Recombinant Proteins in *Escherichia coli* by Using Hexahistidine-Tagged Maltose-Binding Protein as a Fusion Partner. *Methods in Molecular Biology*, *705*, 259–274.

Sun, Z., Liu, Q., Qu, G., Feng, Y., & Reetz, M. T. (2019). Utility of B-Factors in Protein Science: Interpreting Rigidity, Flexibility, and Internal Motion and Engineering Thermostability. *Chemical Reviews*.

Sunter, J., & Gull, K. (2021). Shape, form, function and *Leishmania* pathogenicity: from textbook descriptions to biological understanding. *Open Biology*, *7*(9), 170165.

Tartoff, K. D., & Hobbs, C. A. (1987). Improved Media for Growing Plasmid and Cosmid Clones. *Bethesda Res. Lab. Focus*, *9*(12).

Taylor, G. (2003). The phase problem. *Acta Crystallographica Section D*, *59*(11), 1881–1890.

Teng, T.-Y. (1990). Mounting of crystals for macromolecular crystallography in a free-standing thin film. *Journal of Applied Crystallography*, *23*(5), 387–391.

Tengel, T., Fex, T., Emtenäs, H., Almqvist, F., Sethson, I., & Kihlberg, J. (2004). Use of ^{19}F NMR spectroscopy to screen chemical libraries for ligands that bind to proteins. *Organic and Biomolecular Chemistry*, *2*(5), 725–731.

Thakur, S., Joshi, J., & Kaur, S. (2020). Leishmaniasis diagnosis: an update on the use of parasitological, immunological and molecular methods. *J Parasit Dis*, *44*(2), 253–272.

Timofeev, V., & Samygina, V. (2023). Protein Crystallography: Achievements and Challenges. *Crystals*, *13*(1).

Torres-Guerrero, E., Quintanilla-Cedillo, M. R., Ruiz-Esmenjaud, J., & Arenas, R. (2017). Leishmaniasis: a review. *F1000 Research*, *6*.

Tripathi, N. K., & Shrivastava, A. (2019). Recent Developments in Bioprocessing of Recombinant Proteins: Expression Hosts and Process Development. *Frontiers in Bioengineering and Biotechnology*, *7*(December).

Tyler, K. M., & Engman, D. M. (2001). The life cycle of *Trypanosoma cruzi* revisited. *International Journal for Parasitology*, *31*(5–6), 472–481.

Urh, M., Simpson, D., & Zhao, K. (2009). Chapter 26 Affinity Chromatography. General Methods. *Methods in Enzymology*, *463*(C), 417–438.

Usón, I., Pohl, E., Schneider, T. R., Dauter, Z., Schmidt, A., Fritz, H. J., & Sheldrick, G. M. (1999). 1.7 Å Structure of the stabilized REI(v) mutant T39K. Application of local NCS restraints. *Acta Crystallographica Section D: Biological Crystallography*, *55*(6), 1158–1167.

Usón, I., & Sheldrick, G. M. (1999). Advances in direct methods for protein crystallography. *Current Opinion in Structural Biology*, *9*(5), 643–648.

Vekilov, P. G., & Vorontsova, M. A. (2014). Nucleation precursors in protein crystallization. *Acta Crystallographica Section F*, *70*, 271–282.

Vera, L., Czarny, B., Georgiadis, D., Dive, V., & Stura, E. A. (2011). Practical Use of Glycerol in Protein Crystallization. *Crystal Growth & Design*, *11*(7), 2755–2762.

Vieira, C. B., Praça, Y. R., Bentes, K. L. da S., Santiago, P. B., Silva, S. M. M., Silva, G. D. S., Motta, F. N., Bastos, I. M. D., de Santana, J. M., & de Araújo, C. N. (2018). Triatomines: Trypanosomatids, Bacteria, and Viruses Potential Vectors? *Frontiers in Cellular and Infection Microbiology*, *8*(November), 405.

Vulpiani, M. P., Iannetti, L., Paganico, D., Iannino, F., & Ferri, N. (2011). Methods of Control of the *Leishmania infantum* Dog Reservoir: State of the Art. *Veterinary Medicine International*, *2011*, 13.

Wang, T., & Leyh, T. S. (2012). Three-stage assembly of the cysteine synthase complex from *Escherichia coli*. *Journal of Biological Chemistry*, *287*(6), 4360–4367.

Wegerer, A., Sun, T., & Altenbuchner, J. (2008). Optimization of an *E. coli* L-rhamnose-inducible expression vector: test of various genetic module combinations. *BMC Biotechnology*, 2(8).

Whitmore, L., & Wallace, B. A. (2004). DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Research*, *32*(Web Server issue).

Whitmore, L., & Wallace, B. A. (2008). Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases. *Biopolymers*, *89*(5), 392–400.

WHO. (2020). Ending the neglect to attain the Sustainable Development Goals: a road map for neglected tropical diseases 2021–2030. Geneva: World Health Organization (https://www.who.int/neglected_diseases/Revised-DraftNTD-Roadmap-23Apr2020.pdf). In *Word Health Organization*.

Willcott, M. R. (2009). MestRe Nova. *Journal of the American Chemical Society*, *131*(36), 13180–13180.

Williams, R., Westrop, G. D. D., & Coombs, G. H. H. (2009). Two pathways for cysteine biosynthesis in *Leishmania major*. *Biochemical Journal*, *420*(3), 451–462.

Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., & Wilson, K. S. (2011). Overview of the CCP4 suite and current developments. In *Acta Crystallographica Section D: Biological Crystallography* (Vol. 67, Issue 4, pp. 235–242). International Union of Crystallography.

Winter, G. (2010). xia2: an expert system for macromolecular crystallography data reduction. *Journal of Applied Crystallography*, *43*(1), 186–190.

Winter, G., Lobley, C. M. C., & Prince, S. M. (2013). Decision making in xia2. *Acta Crystallographica Section D: Biological Crystallography*, *69*(7), 1260–1273.

Wirtz, M., & Hell, R. (2006). Functional analysis of the cysteine synthase protein complex from plants: Structural, biochemical and regulatory properties. *Journal of Plant Physiology*, *163*(3), 273–286.

Wlodawer, A., Minor, W., Dauter, Z., & Jaskolski, M. (2013). Protein crystallography for aspiring crystallographers . *FEBS Journal*, *280*(22), 5705–5736.

Wood, D. J., Daniel Lopez-Fernandez, J., Knight, L. E., Al-Khawaldeh, I., Gai, C., Lin, S., Martin, M. P., Miller, D. C., Cano, C. C., Endicott, J. A., Hardcastle, I. R., Noble, M. E. M. M., Waring, M. J., Lopez-Fernandez, J. D., Knight, L. E., Al-Khawaldeh, I., Gai, C., Lin, S., Martin, M. P., … Waring, M. J. (2019). FragLites, Minimal,

Halogenated Fragments Displaying Pharmacophore Doublets. An Efficient Approach to Druggability Assessment and Hit Generation. *Journal of Medicinal Chemistry*, *62*(7), 3741–3752.

Yamaguchi, H., & Miyazaki, M. (2014). Refolding Techniques for Recovering Biologically Active Recombinant Proteins from Inclusion Bodies. *Biomolecules*, *4*(1), 235.

Yu, J.-X., Hallac, R. R., Chiguru, S., & Mason, R. P. (2013). New Frontiers and Developing Applications in ^{19} F NMR. *Prog Nucl Magn Reson Spectrosc*, *70*, 25–49.

Zijlstra, E. E., Musa, A. M., Khalil, E. A. G., El Hassan, I. M., & El-Hassan, A. M. (2003). Post-kala-azar dermal leishmaniasis. *Lancet Infectious Diseases*, *3*(2), 87–98.

Chapter 9: Appendix

Chapter 3 Appendices

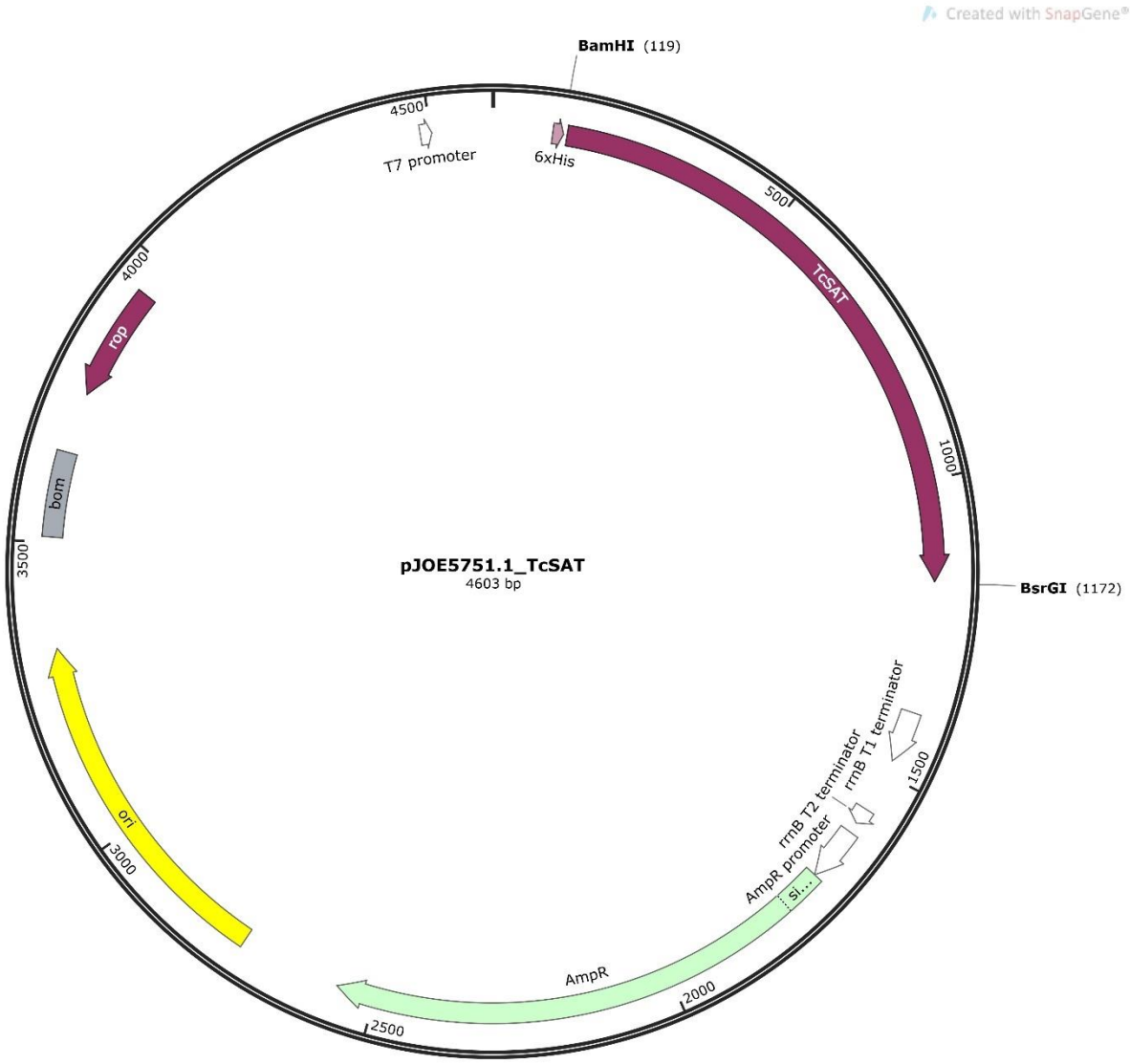


Figure S9-3-1 Plasmid map for *Tc*SAT in the pJOE5751 vector.

MGSSHHHHHHSSMNKEILAVVEAVSNEKALPREKIFEALESALATATKKKYEQEID
VRVQIDRKSGDFDTFRRWLVVDEVTQPTKEITLEAARYEDESLNLGDYVEDQIESV
TFDRITTQTAKQVIVQKVREAERAMVVDQFREHEGEIITGVVKKVNRDNISLDLGN
NAEAVILREDMLPRENFRPGDRVRGVLYSVRPEARGAQLFVTRSKPEMLIELFRIE
VPEIGEEVIEIKAAARDPGSRAKIAVKTNDKRIDPVGACVGMRGARVQAVSTELGG
ERIDIVLWDDNPAQFVINAMAPADVASIVVDEDKHTMDIAVEAGNLAQAIGRNGQN
VRLASQLSGWELNVMTVDDLQAKHQAEAHAAIDTFTKYLDIDEDFATVLVEEGFST
LEELAYVPMKELLEIEGLDEPTVEALRERAKNALATIAQAQEESLGDNKPADDLLN
LEGVDRDLAFKLAARGVCTLEDLAEQGIDDLADIEGLTDEKAGALIMAARNICWFG
DEATSGSGHHHHHHSAALEVLFQGPGMNRLEVIRKSISYVGLDPAMFKPYPFNVSS
SALSELMDCMSYIVFPEFSEPPMNPVTVPEQRLGGSALLQWTLTRMADIINSQIYS
ALMLQGHTKCSKDEAGIPHSLNAESGSNDDGFTPQNETHAFGEVRSFHLEAKKRAE
EITEAFLTRKLRSIRWLLRTDAESILNNDIASNSLSEVVLCYPGLRCMKHQRTAHA
LHELGAPSIFTRLLTEMAHSTTGIDIHPATSIGHHFFIDHGTGIVIGSTAIIGNYV
SIYHGVTLGTRSFPVVAKTGEKVRNLARHPIIEDRVTIYANAMVLGRIRIGKDSVI
GSHCLVLKDLPPKSMIRRTSNAKLPTPYITLQKGGSDIELALVPRGSKETAAAKFE
RQHMDSSTSAALEA

Figure S9-3-2 *Tc*SAT protein sequence in one letter amino acid code including tag and pET50b residues.

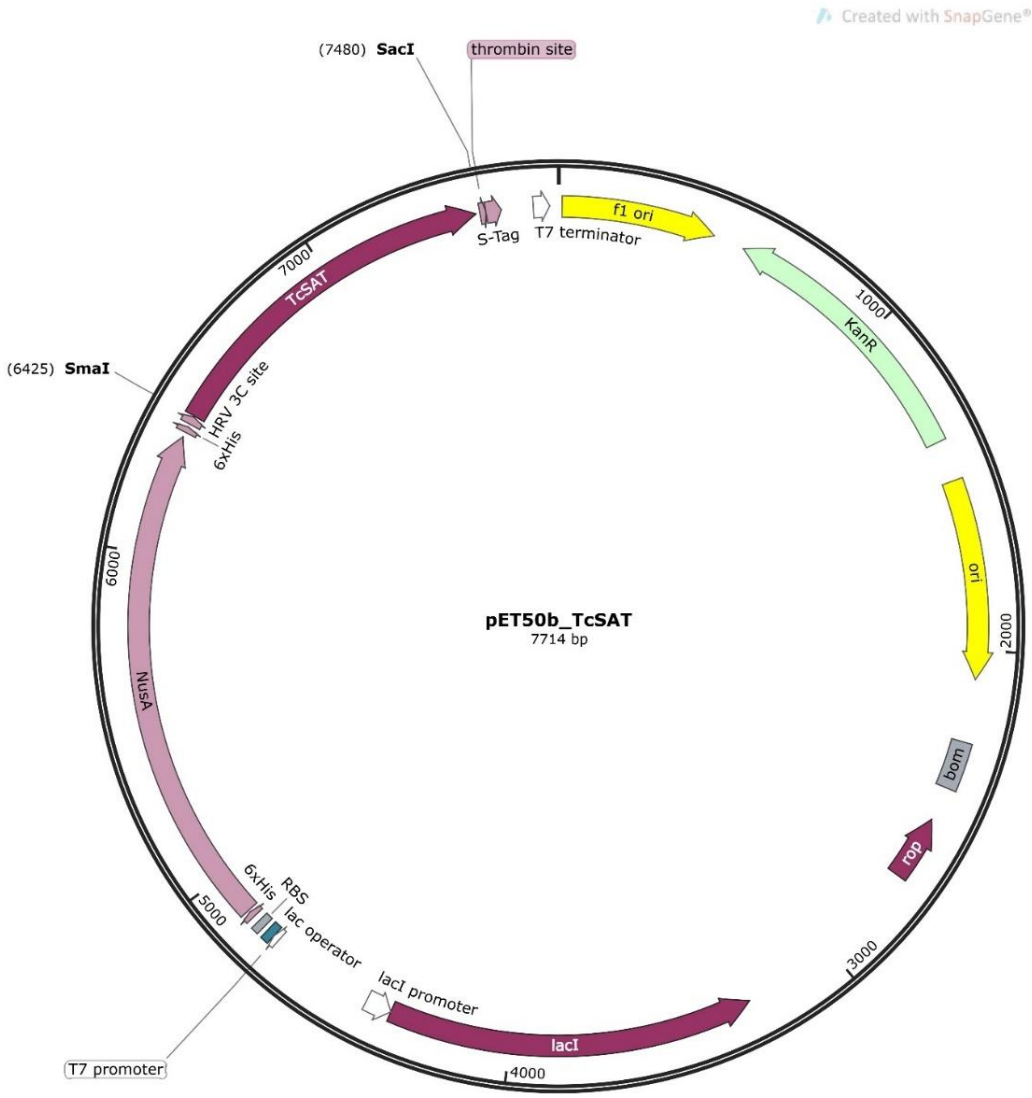


Figure S9-3-3 Plasmid map for *Tc*SAT in the pET50b vector.

MSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFELGLEFPNLP
YYIDGDVKLTQSMAIIRYIADKHNMLGGCPKERAEISMLEGAVLDIRYGVSRIAYS
KDFETLKVDFLSKLPEMLKMFEDRLCHKTYLNGDHVTHPDFMLYDALDVVLYMDPM
CLDAFPKLVCFKKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGGGDHPPKSDLVPR
GSPNRLEVIRKSISYVGLDPAMFKPYPFNVSSSALSELMDCMSYIVFPEFSEPPMN
PVTVPEQRLGGSALLQWTLTRMADIINSQIYSALMLQGHTKCSKDEAGIPHSLNAE
SGSNDDGFTPQNETHAFGEVRSFHLEAKKRAEEITEAFLTRKLRSIRWLLRTDAES
ILNNDIASNSLSEVVLCYPGLRCMKHQRTAHALHELGAPSIFTRLLTEMAHSTTGI
DIHPATSIGHHFFIDHGTGIVIGSTAIIGNYVSIYHGVTLGTRSFPVVAKTGEKVR
NLARHPIIEDRVTIYANAMVLGRIRIGKDSVIGSHCLVLKDLPPKSMIRRTSNAKL
PTPYITLQKGGSDI

Figure S9-3-4 *Tc*SAT protein sequence in one letter amino acid code with tag and residues from pGEX-4T-3 vector.

Figure S9-3-5 Plasmid map for *Tc*SAT in the pGEX-4T-3 vector.

MKIKTGARILALSALTTMMFSASALAKIEEGKLVIWINGDKGYNGLAEVGKKFEKD
TGIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHDRFGGYAQSGLLAEITPDKAFQ
DKLYPFTWDAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEEIPALDKELKAK
GKSALMFNLQEPYFTWPLIAADGGYAFKYENGKYDIKDVGVDNAGAKAGLTFLVDL
IKNKHMNADTDYSIAEAAFNKGETAMTINGPWAWSNIDTSKVNYGVTVLPTFKGQP
SKPFVGVLSAGINAASPNKELAKEFLENYLLTDEGLEAVNKDKPLGAVALKSYEEE
LVKDPRIAATMENAQKGEIMPNIPQMSAFWYAVRTAVINAASGRQTVDEALKDAQT
NSSSNNNNNNNNNLGDDDDKVPHMMNRLEVIRKSISYVGLDPAMFKPYPFNVSSS
ALSELMDCMSYIVFPEFSEPPMNPVTVPEQRLGGSALLQWTLTRMADIINSQIYSA
LMLQGHTKCSKDEAGIPHSLNAESGSNDDGFTPQNETHAFGEVRSFHLEAKKRAEE
ITEAFLTRKLRSIRWLLRTDAESILNNDIASNSLSEVVLCYPGLRCMKHQRTAHAL
HELGAPSIFTRLLTEMAHSTTGIDIHPATSIGHHFFIDHGTGIVIGSTAIIGNYVS
IYHGVTLGTRSFPVVAKTGEKVRNLARHPIIEDRVTIYANAMVLGRIRIGKDSVIG
SHCLVLKDLPPKSMIRRTSNAKLPTPYITLQKGGSDI

Figure S9-3-6 *Tc*SAT with tag with pMAL-p5E vector.

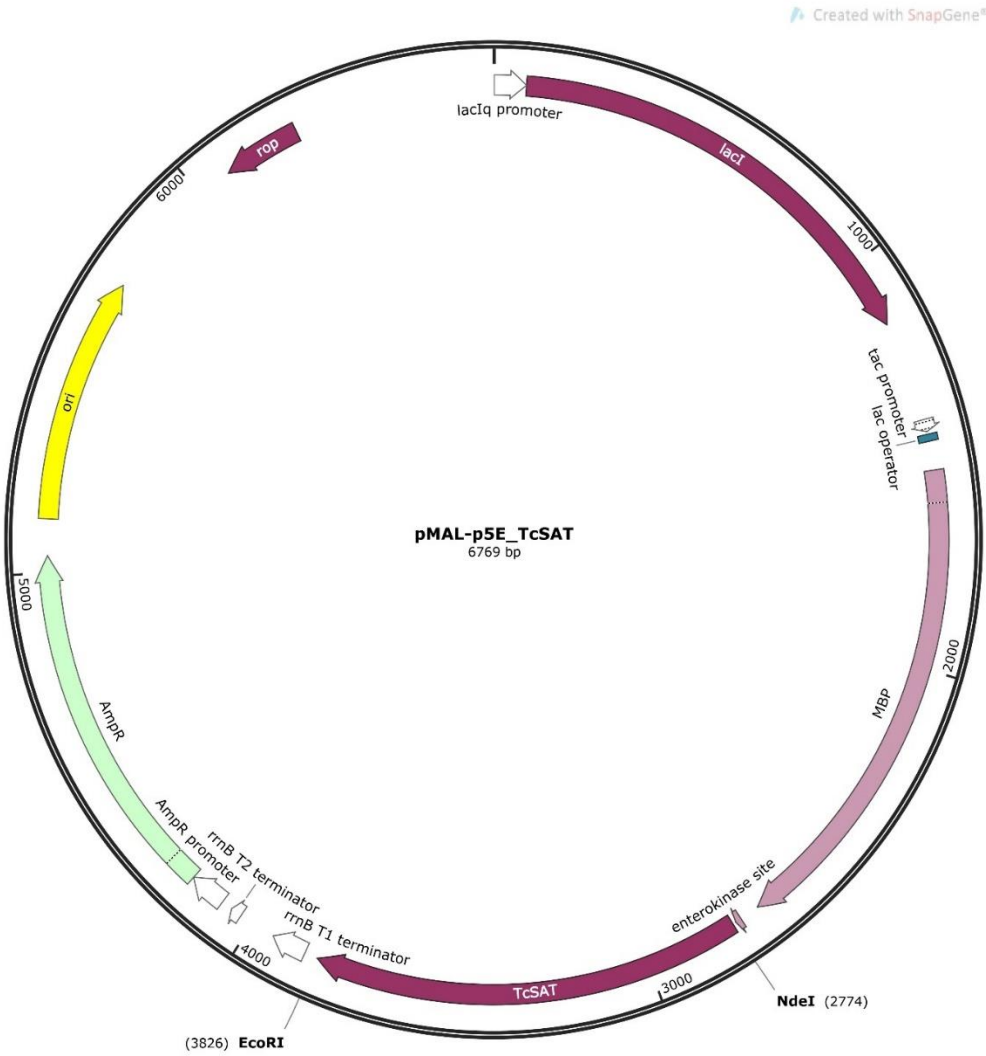


Figure S9-3-7 Plasmid map for *Tc*SAT in the pMAL-p5E vector.

Chapter 4 Appendices

MTMITHHHHHHGSSVQEFDPRNNVAPSMDALIGETPAVYLKRMNDTAATIVLKLEC
ENPMASVKDRLAYAIYDKAEKEGKIIPGKSVIVEATSGNTGIALAHIGTIRGYKVI
IVMPESMSIERRCLMRIFGAEVILTPAALGMKGALEAVNRIVSNNPDAVSANQFAT
KYNAQIHEETTGPEIWRQTKGHVDCFVAGVGTGGTITGVARYLKSVGCGATIFAVE
PAESPVLSGGKPGPHRIQGIGAGFVPEVFEAALVDEVIQVSGDEAIDTAQKLPRTD
GIFCGFSGGANVYAALQIAKRPEMAGKTIVTVIPSYGERYLSTALYSSIKDEVFAL
KVLSAADI

Figure S9-4-1 *Tc*CS including His Tag and pJoe5751 residues.

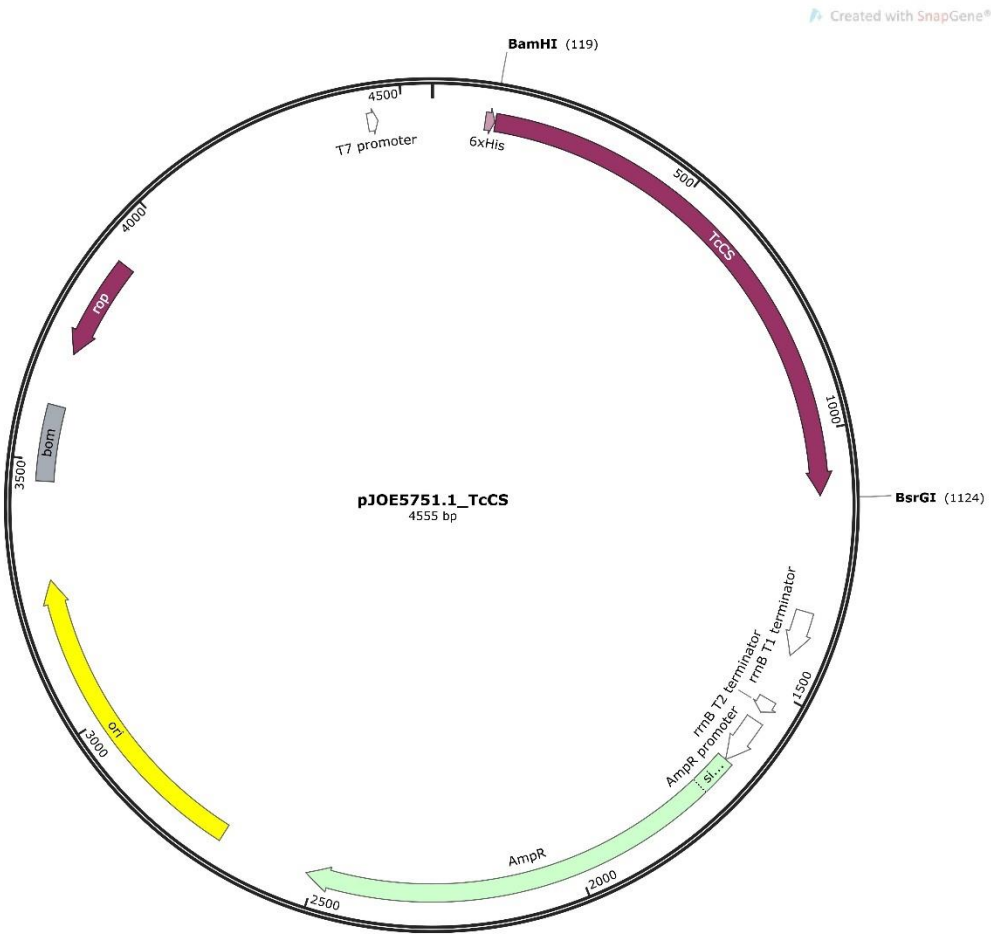


Figure S9-4-2 Plasmid map of *Tc*CS in pJOET5751 vector

MGSSHHHHHHSSGLVPRGSMAAPFDKSKNVAQSIDQLIGQTPALYLNKLNNTKAKV
VLKMECENPMASVKDRLGFAIYDKAEKEGKLIPGKSIVVESSSGNTGVSLAHLGAI
RGYKVIITMPESMSLERRCLLRIFGAEVILTPAALGMKGAVAMAKKIVAANPNAVL
ADQFATKYNALIHEETTGPEIWEQTNHNVDCFIAGVGTGGTLTGVARALKKMGSHA
RIVAVEPTESPVLSGGKPGPHKIQGIGPGFVPDVLDRSLIDEVLCVAGDDAIETAL
KLTRSDGVFCGFSGGANVYAALKIAERPEMEGKTIVTVIPSFGERYLSTTLYRSVR
DEVSSLPVVDASELQD

Figure S9-4-3 *Li*CS including His Tag and pET15b residues.

Figure S9-4-4 Plasmid map of *Li*CS in pET15b vector.

MGSSHHHHHHSSGLVPRGSMSGQEFNKQNDVAPSVDALIGNTPAVYLRNMNKTAAT
IVLKLESENPMASVKDRLAYAIYDKAEKEGKIKRGKSVIVEATSGNTGIALALLGA
VRGYKVIITMPDSMSMERRSLLRIFGAELILTPASLGMKGAVAMANRIVSNNPDAV
LADQFCTKYNAQIHEETTGPEIWRQTKGHVDCFVAGVGTGGTLTGVARFLKSVGCN
ANIVAVEPQESPVLSGGRPGPHRIQGIGAGFVPEVLDRELVDEIIQVSSDKAIETA
QKLSRMDGIFCGFSAGANVFAALKIAERPEMAGKTIVTVIPSFGERYLSTALYASV
REEITSLPVVPVSEIPQV

Figure S9-4-5 *Tth*CS sequence including His Tag and pET15b residues.

Figure S9-4-6 Plasmid map of *Tth*CS in pET15b vector.

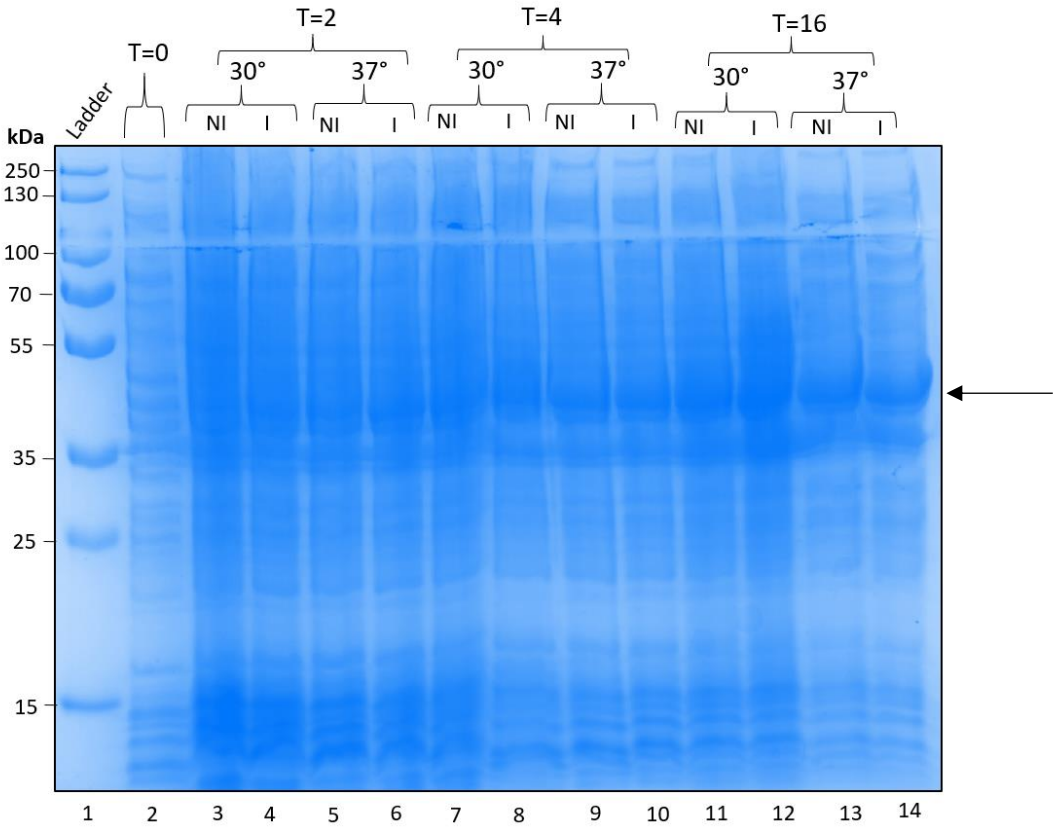


Figure S9-4-7 Expression tests of pJOE5751_*Tc*CS at 30°C and 37°C and at different time points. Non-induced (NI) vs induced with 0.2% L-Rhamnose (I) cells at time (T) = 0, 2, 4 and 16 hours after induction. Lane 1 shows protein ladder. The arrow indicates protein of interest.

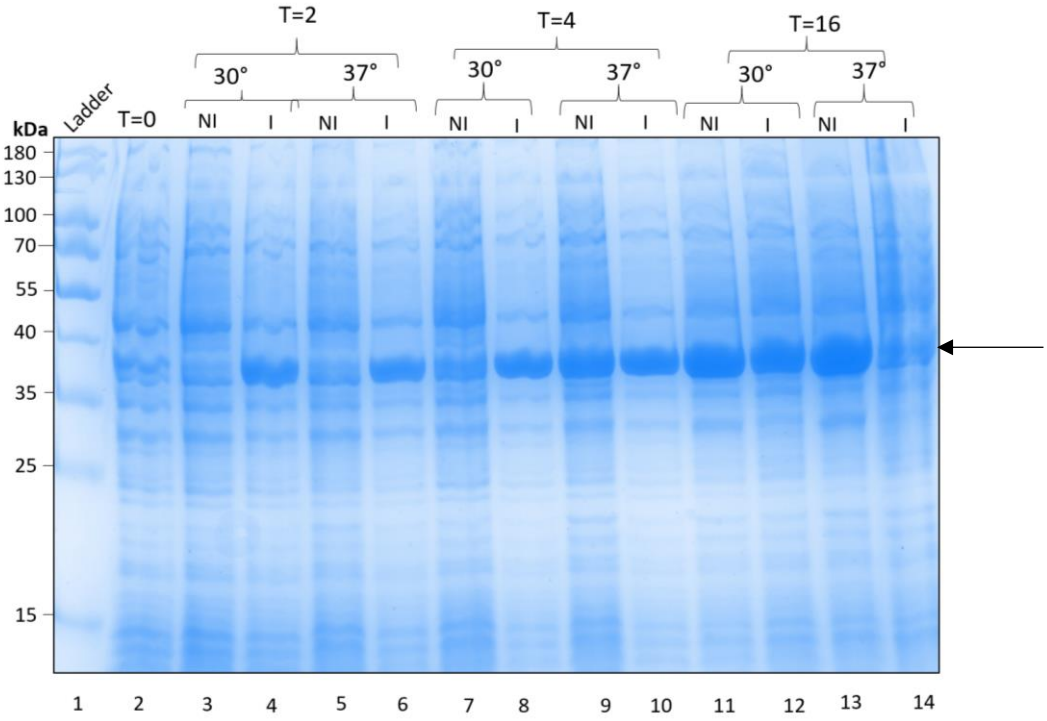


Figure S9-4-8 Expression tests of LiCS at 30°C and 37°C and at different time points. Non-induced (NI) vs induced with 250 mM IPTG (I) cells at time (T) = 0, 2, 4 and 16 hours after induction. Lane 1 shows protein ladder. The arrow indicates protein of interest.

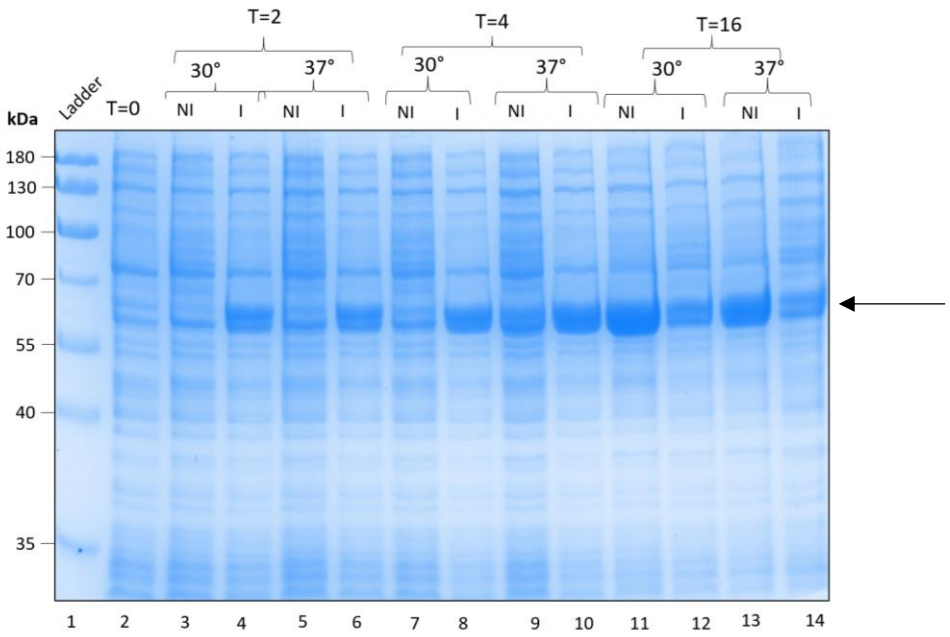


Figure S9-4-9 Expression tests of *Tth*CS at 30°C and 37°C and at different time points. Non-induced (NI) vs induced with 250 mM IPTG (I) cells at time (T) = 0, 2, 4 and 16 hours after induction. Lane 1 shows protein ladder. The arrow indicates protein of interest.

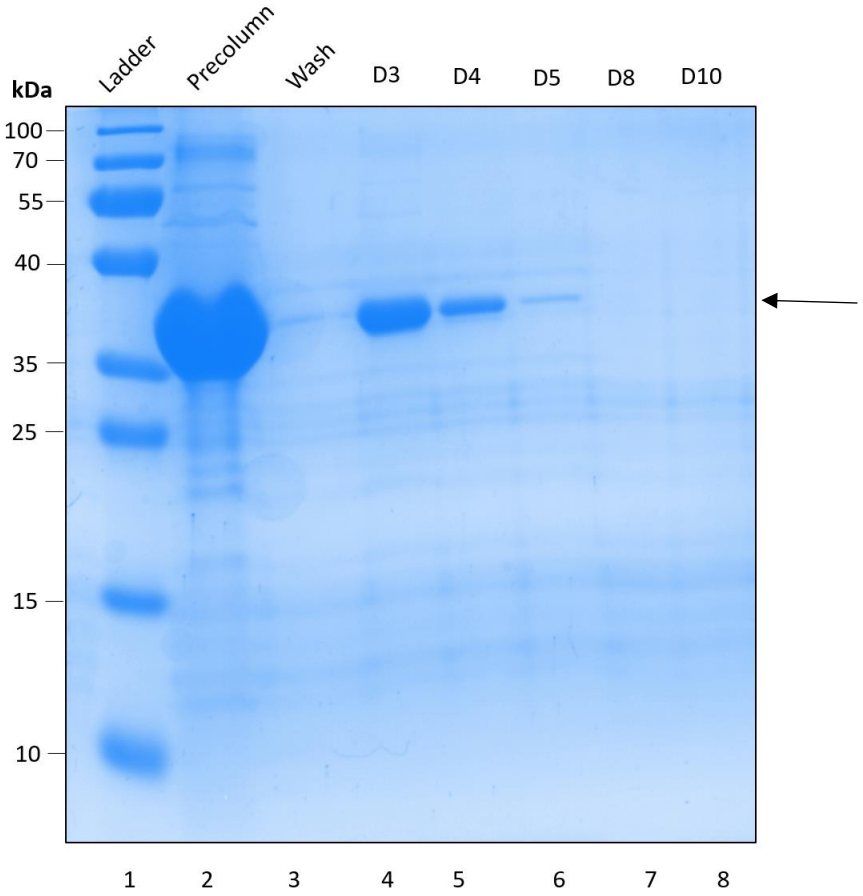


Figure S9-4-10 SDS-PAGE of *Tc*CS gel filtration. Lane 1 shows ladder, lane 2 shows protein sample before gel filtration, lane 3 shows column wash, lanes 4-8 show elution from the column. Arrow indicates expected migration of protein.

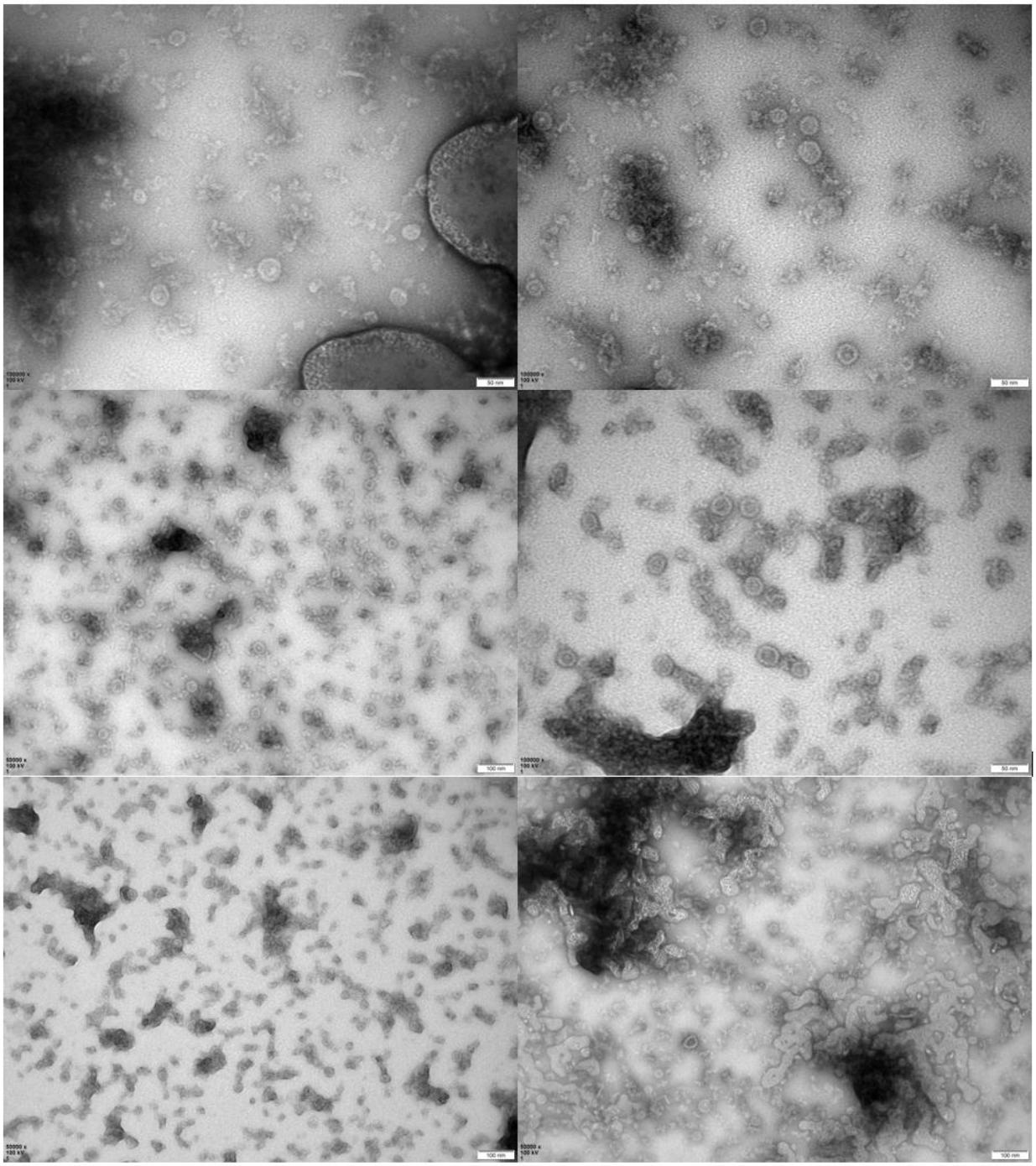
Table *S9-4-1 Li*CS crystallisation conditions:

Conditions
0.1 M Bis-Tris pH 5.5 25 % w/v PEG 3350
0.2 M Sodium acetate trihydrate 0.1 M Sodium cacodylate 6.5 30 % w/v PEG 8000
0.1 M SPG 5.0 25 % w/v PEG 1500
0.1 M MIB 4.0 25 % w/v PEG 1500
0.1 M MIB 5.0 25 % w/v PEG 1500
0.1 M MMT 5.0 25 % w/v PEG 1500
0.1 M MMT 6.0 25 % w/v PEG 1500
0.2 M Sodium fluoride 0.1 M Bis-Tris propane pH 6.5 20 % w/v PEG 3350
0.2 M Lithium sulfate 0.1 M ADA pH 6.5 30 % v/v PEG Smear Medium
0.05 M Magnesium sulfate heptahydrate 0.1 M HEPES pH 7.5 28 % v/v PEG Smear Medium
0.2 M Ammonium acetate 0.1 M Sodium citrate pH 5.6 30 % w/v PEG 4000
0.15 M Potassium thiocyanate 0.1M Sodium acetate pH 5.5 18% w/v PEG 3350
0.2 M Potassium thiocyanate 0.1 M Sodium acetate pH 5.5 25 % w/v PEG 2000 MME
0.2 M Potassium thiocyanate 0.1 M Sodium cacodylate 6.5 25 % w/v PEG 2000 MME
1.6 M Sodium citrate tribasic dihydrate pH 6.5
0.1 M BIS-Tris pH 5.5 25 % w/v PEG 3350
0.2 M Sodium chloride 0.1 M BIS-Tris 5.5 25 % w/v PEG 3350
C4 0.09 M NPS 0.1 M Buffer System 1 pH 6.5 37.5% v/v Precipitant Mix 4

Table S9-4-2 showing the r.m.s.d.s of each monomer of *Tc*CS, *Li*CS, *Tth*CS and *Lm*CS. In this comparison, α carbons of residues 18-320 from *Tc*CS, 24-326 from *Li*CS, 25-327 from *Tth*CS and 18-325 from *Lm*CS were used.

	*Tc*CS A	*Tc*CS B	*Tc*CS C	*Tc*CS D	*Li*CS A	*Li*CS B	*Tth*CS A	*Tth*CS B	*Lm*CS A	*Lm*CS B
*Tc*CS A		0.54	0.30	0.50	0.70	0.78	0.63	0.56	0.74	0.71
*Tc*CS B	282		0.60	0.17	0.84	1.02	0.64	0.71	0.95	0.94
*Tc*CS C	287	298		0.54	0.78	0.73	0.69	0.63	0.80	0.77
*Tc*CS D	282	303	298		0.83	0.94	0.62	0.68	0.74	0.92
*Li*CS A	274	270	275	270		0.72	0.74	0.72	0.60	0.58
*Li*CS B	275	271	274	269	277		1.09	0.97	0.39	0.35
*Tth*CS A	281	303	297	303	268	269		0.74	1.04	1.02
*Tth*CS B	281	303	297	303	269	270	268		0.93	0.89
*Lm*CS A	275	271	277	271	276	274	267	270		0.57
*Lm*CS B	262	258	264	258	263	261	254	257	280	

Chapter 5 Appendices



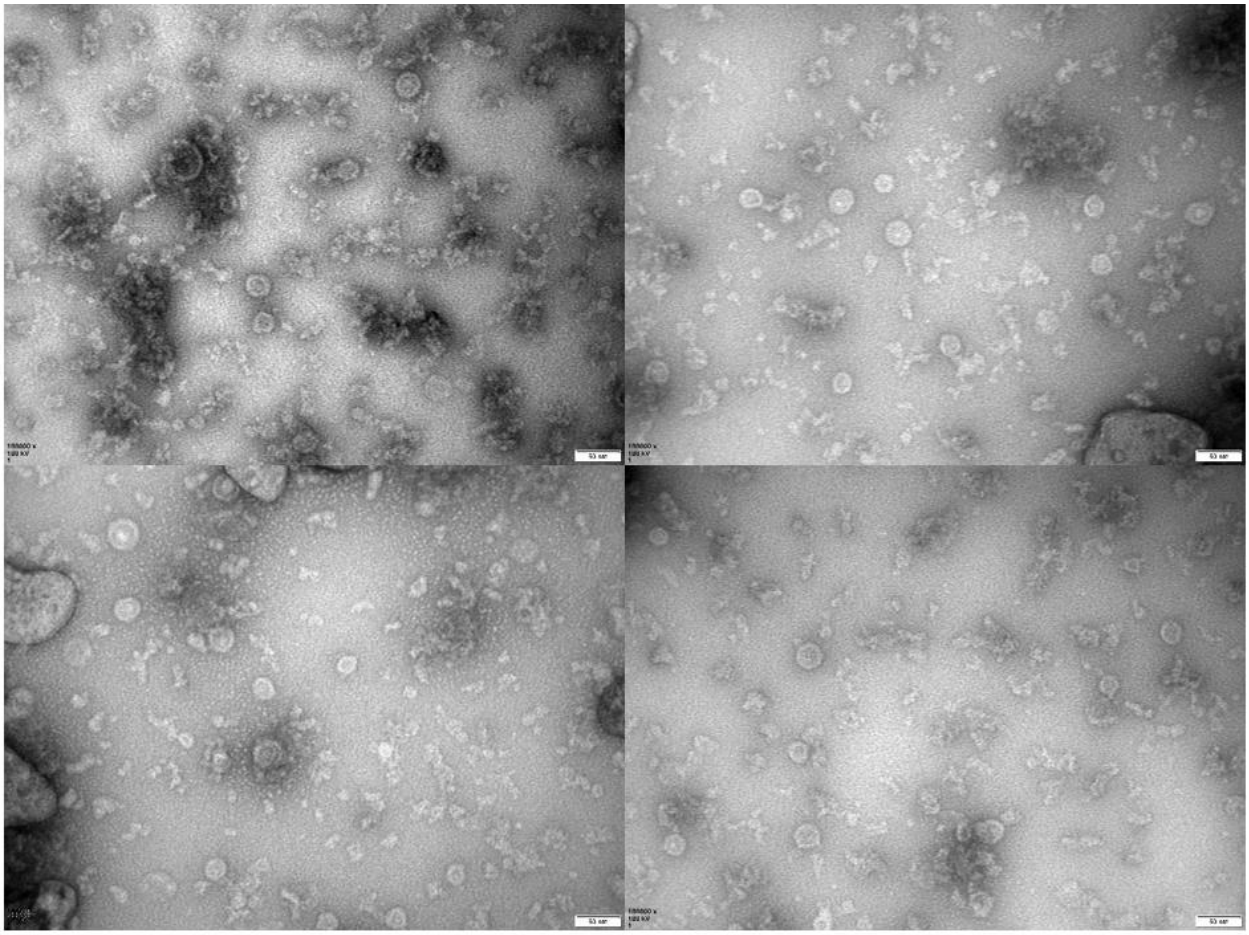


Figure S9-5-1 TEM negative stain images of *Tc*CSC. These images show different spatial points on the grid when stained with 1% uranyl acetate.

Chapter 6 Appendices

Table S9-6-1 – Custom SAT Peptides

Position	Peptide name	Sequence
1A	nVcSAT_10	SLDMDQQFNG
1B	VcSAT_15_2	ESDKPSLDMDQQFNG
1C	nTcSAT10	ITLQKGGSDI
1D	nTcSAT11	YITLQKGGSDI
1E	nTcSAT12	PYITLQKGGSDI
1F	nTcSAT13	TPYITLQKGGSDI
1G	nTcSAT14	PTPYITLQKGGSDI
1H	nTcSAT15	LPTPYITLQKGGSDI
2A	TcSAT10_1	I**M**LQKGGSDI
2B	TcSAT10_2	I**F**LQKGGSDI
2C	TcSAT10_3	IT**F**QKGGSDI
2D	TcSAT10_4	IT**V**QKGGSDI
2E	TcSAT10_5	ITL**H**KGGSDI
2F	TcSAT10_6	ITL**E**KGGSDI
2G	TcSAT10_7	ITLQ**E**GGSDI
2H	TcSAT10_8	ITLQ**R**GGSDI
3A	TcSAT10_9	ITLQ**G**GGSDI
3B	TcSAT10_10	ITLQK**N**GSDI
3C	TcSAT10_11	ITLQK**D**GSDI
3D	TcSAT10_12	ITLQKGGS**G**I
3E	TcSAT11_1	YI**M**LQKGGSDI
3F	TcSAT11_2	YI**F**LQKGGSDI
3G	TcSAT11_3	YIT**F**QKGGSDI
3H	TcSAT11_4	YIT**V**QKGGSDI
4A	TcSAT11_5	YITL**H**KGGSDI
4B	TcSAT11_6	YITL**E**KGGSDI
4C	TcSAT11_7	YITLQ**E**GGSDI
4D	TcSAT11_8	YITLQ**R**GGSDI
4E	TcSAT11_9	YITLQ**G**GGSDI
4F	TcSAT11_10	YITLQK**N**GSDI
4G	TcSAT11_11	YITLQK**D**GSDI
4H	TcSAT11_12	YITLQKGGS**G**I
5A	TcSAT12_1	PYI**M**LQKGGSDI
5B	TcSAT12_2	PYI**F**LQKGGSDI
5C	TcSAT12_3	PYIT**F**QKGGSDI
5D	TcSAT12_4	PYIT**V**QKGGSDI
5E	TcSAT12_5	PYITL**H**KGGSDI
5F	TcSAT12_6	PYITL**E**KGGSDI
5G	TcSAT12_7	PYITLQ**E**GGSDI
5H	TcSAT12_8	PYITLQ**R**GGSDI
6A	TcSAT12_9	PYITLQ**G**GGSDI
6B	TcSAT12_10	PYITLQK**N**GSDI
6C	TcSAT12_11	PYITLQK**D**GSDI
6D	TcSAT12_12	PYITLQKGGS**G**I
6E	TcSAT13_1	TPYI**M**LQKGGSDI
6F	TcSAT13_2	TPYI**F**LQKGGSDI
6G	TcSAT13_3	TPYIT**F**QKGGSDI
6H	TcSAT13_4	TPYIT**V**QKGGSDI
7A	TcSAT13_5	TPYITL**H**KGGSDI

7B	TcSAT13_6	TPYITL**E**KGGSDI
7C	TcSAT13_7	TPYITLQ**E**GGSDI
7D	TcSAT13_8	TPYITLQ**R**GGSDI
7E	TcSAT13_9	TPYITLQ**G**GGSDI
7F	TcSAT13_10	TPYITLQK**N**GSDI
7G	TcSAT13_11	TPYITLQK**D**GSDI
7H	TcSAT13_12	TPYITLQKGGS**G**I
8A	TcSAT14_1	PTPYI**M**LQKGGSDI
8B	TcSAT14_2	PTPYI**F**LQKGGSDI
8C	TcSAT14_3	PTPYIT**F**QKGGSDI
8D	TcSAT14_4	PTPYIT**V**QKGGSDI
8E	TcSAT14_5	PTPYITL**H**KGGSDI
8F	TcSAT14_6	PTPYITL**E**KGGSDI
8G	TcSAT14_7	PTPYITLQ**E**GGSDI
8H	TcSAT14_8	PTPYITLQ**R**GGSDI
9A	TcSAT14_9	PTPYITLQ**G**GGSDI
9B	TcSAT14_10	PTPYITLQK**N**GSDI
9C	TcSAT14_11	PTPYITLQK**D**GSDI
9D	TcSAT14_12	PTPYITLQKGGS**G**I
9E	TcSAT15_1	LPTPYI**M**LQKGGSDI
9F	TcSAT15_2	LPTPYI**F**LQKGGSDI
9G	TcSAT15_3	LPTPYIT**F**QKGGSDI
9H	TcSAT15_4	LPTPYIT**V**QKGGSDI
10A	TcSAT15_5	LPTPYITL**H**KGGSDI
10B	TcSAT15_6	LPTPYITL**E**KGGSDI
10C	TcSAT15_7	LPTPYITLQ**E**GGSDI
10D	TcSAT15_8	LPTPYITLQ**R**GGSDI

10E	TcSAT15_9	LPTPYITLQ**G**GGSDI
10F	TcSAT15_10	LPTPYITLQK**N**GSDI
10G	TcSAT15_11	LPTPYITLQK**D**GSDI
10H	TcSAT15_12	LPTPYITLQKGGS**G**I
11A	AdSAT_10_2	**I**FVERDGSGI
11B	ScSAT_10_2	**I**FFEKDGSGI
11C	LmSAT_10_2	**I**FLEGDGSGI
11D	TthSAT_15_3	LSRVD**II**FHENGSGI
11E	AdSAT_15_3	**E**KREIFVERDGSGI
11F	LmSAT_15_2	**E**PHERIFLEGDGSGI
11G	nEhSAT1_10	DVDSNQTVYI
11H	nEhSAT1_15	CWIDRDVDSNQTVYI
12A	ThSAT_10 _2	**I**MFHENGSGI
12B	TthSAT_15_3	**E**KHERMFFEKDGSGI
12C	nBvSAT_10	GSRIVQRKNK
12D	nBvSAT_15	ENVPAGSRIVQRKNK
12E	nBaSAT_10	ARIIGETGCT
12F	nBaSAT_15	VAGVPARIIGETGCT
12G	nGmSAT_10	TAVGNPARLV
12H	nGmSAT_15	VPPRTTAVGNPARLV

Table S9-6-2 – Custom SAT Peptides binding and effect on *Tc*CS activity

Peptides binding in active site	**Peptides effecting *Tc*CS Activity**	**Peptides decreasing *Tc*CS activity**	**Peptides increasing *Tc*CS activity**
2B, 2F, 3B, 3E, 3F, 3H, 4B – 4D, 4H, 5C, 5D, 5F – 6E, 6G – 7A, 7D, 7H – 11A, 11C, 12A, 12B	2E, 3A – 3C, 4F, 5B, 5H, 6B – 6D, 6H – 7B, 7E, 7G, 8A, 8B, 8E, 8F, 9A – 9C, 9E – 10A, 10C – 10E, 10H, 11C, 12A, 12B, 12D, 12H	2E, 3A – 3C, 5H, 6D, 9C, 9G – 10A	4F, 5B, 6B, 6C, 6H – 7B, 7E, 7G, 8A, 8B, 8E, 8F, 9A, 9B, 9E, 9F, 10C – 10E, 10H, 11C, 12A, 12B, 12D, 12H

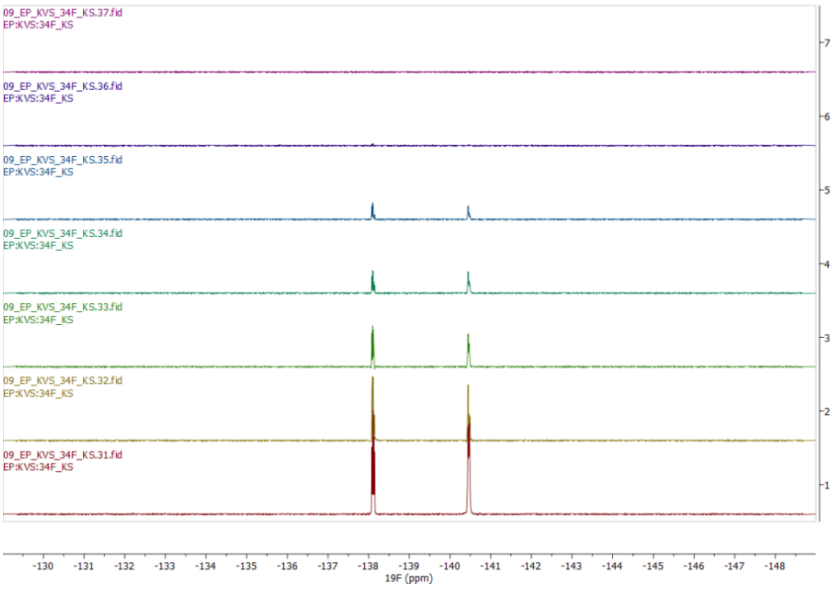


Figure S9-6-1 NMR spectra of 34F without *Tc*CS.

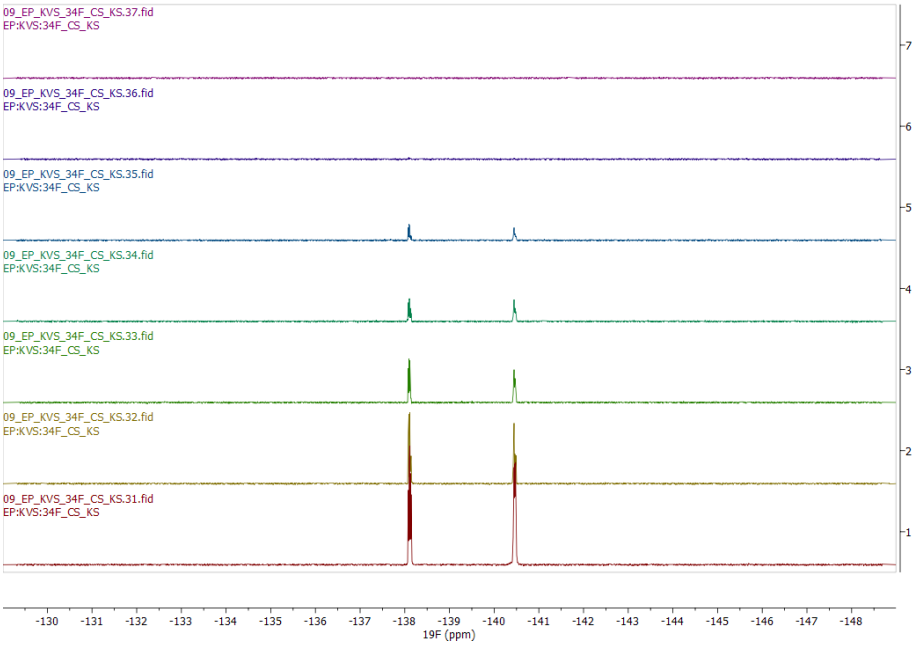


Figure S9-6-2 NMR spectra of 34F with *Tc*CS.

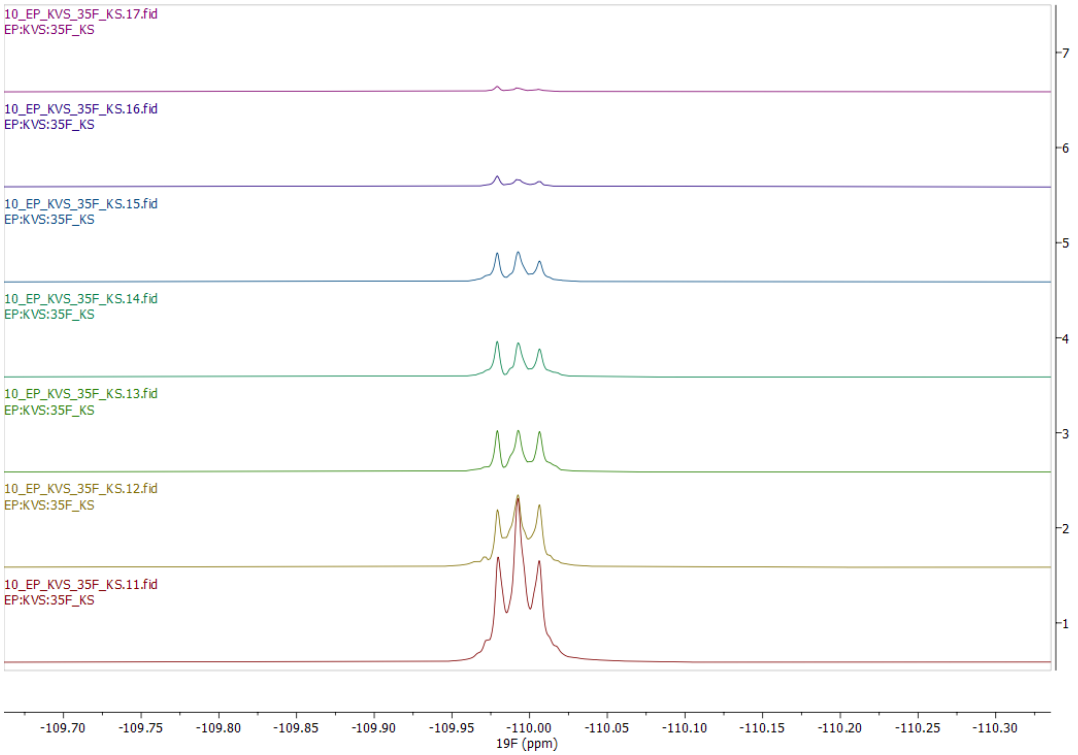


Figure S9-6-3 NMR spectra of 35F without *Tc*CS.

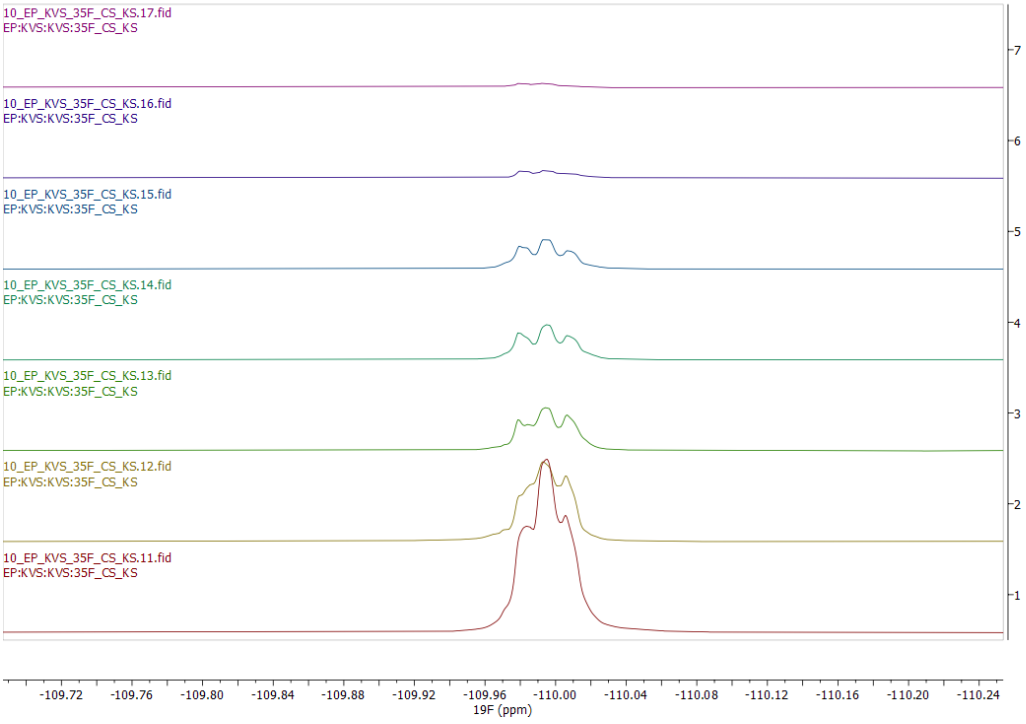


Figure S9-6-4 NMR spectra of 35F with *Tc*CS.

Table S9-6-3 *Tc*CS data collection statistics.

	*Tc*CS F4	*Tc*CS F13	*Tc*CS F15	*Tc*CS F17	*Tc*CS F18	*Tc*CS F23
Data Collection						
Beamline	I04 DLS	I04 DLS	I04 DLS	I04 DLS	I04 DLS	I04 DLS
Detector	Eiger2 XE 16M	Eiger2 XE 16M	Eiger2 XE 16M	Eiger2 XE 16M	Eiger2 XE 16M	Eiger2 XE 16M
Temperature (K)	100	100	100	100	100	100
Wavelength (Å)	0.9194	0.8984	0.8984	0.8984	0.8984	0.8984
Exposure time (s)	0.01	0.05	0.05	0.05	0.05	0.05
Transmission (%)	100	100	100	100	100	100
Rotation range (°)	360	360	360	360	360	360
Image slices (°)	0.1	0.1	0.1	0.1	0.1	0.1
Resolution range (Å)	55.3 – 2.2	43.8 – 2.3	43.8 – 2.2	43.7 – 2.1	63.5 – 2.3	46.6 – 2.3
Space group	$P2_1$	$P2_1$	$P2_1$	$P2_1$	$P2_1$	$P2_1$

a b c (Å)	55.3, 68.6, 171.3	55.4, 68.9, 170.9	55.3, 68.4, 171.2	55.2, 68.2, 170.8	55.5, 68.5, 171.2	55.5, 68.8, 171.1
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90
R_{merge}	0.057 (0.791)	0.067 (0.698)	0.062 (0.626)	0.067 (0.546)	0.060 (0.451)	0.066 (0.841)
R_{pim}	0.025 (0.339)	0.030 (0.310)	0.029 (0.285)	0.031 (0.263)	0.028 (0.201)	0.033 (0.424)
$CC_{1/2}$	0.999 (0.816)	0.998 (0.751)	0.999 (0.831)	0.998 (0.858)	0.998 (0.905)	0.999 (0.644)
Multiplicity	6.3 (6.4)	5.7 (6)	5.2 (5.5)	5.4 (5.2)	5.6 (5.8)	4.6 (4.8)
Completeness (%)	100 (100)	100 (100)	100 (100)	98.6 (97.8)	100 (100)	100 (100)
I/σ(I)	16.2 (1.0)	13.6 (1)	13.9 (1)	12.2 (1)	14.3 (1)	12.5 (1)
Unique reflections	63663 (3087)	61236 (3059)	63386 (0.831)	70100 (3412)	55084 (2720)	56876 (2784)
Refinement						
No. of refs used	63647	61222	63371	70085	55069	56862
R_{work}	0.198	0.214	0.216	0.212	0.215	0.212
R_{free}	0.238	0.255	0.259	0.258	0.268	0.256
Protein atoms	9408	9522	9505	9537	9643	9525

Ligands	85	88	87	104	132	96
RMS deviations						
Bond lengths (Å)	0.0081	0.0071	0.0071	0.0077	0.0068	0.0067
Bond angles (°)	1.5	1.4	1.4	1.5	1.4	1.4
Mean b values (Å^2)						
Protein	58	61	59	54	64	66
water	45	49	48	43	50	51
ligand	83	76	83	84	77	68
Ramachandran plot (%)						
Favoured	97	98	98	98	97	98
Allowed	2	2	2	2	3	2

Values in brackets refer to outer resolution shell.

Table S9-6-4 - *Li*CS data collection statistics.

	*Li*CS F17	*Li*CS F18	*Li*CS F22	*Li*CS F23	*Li*CS F24	*Li*CS F28	*Li*CS F29	*Li*CS F31
Data Collection								
Beamline	I04 DLS	I04 DLS	I04 DLS	I04 DLS	I04 DLS	I04 DLS	I04 DLS	I04 DLS
Detector	Eiger2 XE 16M	Eiger2 XE 16M	Eiger2 XE 16M	Eiger2 XE 16M	Eiger2 XE 16M	Eiger2 XE 16M	Eiger2 XE 16M	Eiger2 XE 16M
Temperature (K)	100	100	100	100	100	100	100	100
Wavelength (Å)	0.9184	0.9184	0.9184	0.9184	0.9184	0.9184	0.9184	0.9184
Exposure time (s)	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Transmission (%)	50	50	50	50	50	50	50	50
Rotation range (°)	360	360	360	360	360	360	360	360
Image slices (°)	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1

Resolution range (Å)	54.0 – 1.8	72.1 – 2.6	73.5 – 2.2	71.7 – 2.5	54.0 – 1.8	54.0 – 1.9	53.8 – 1.9	73.9 – 2.5
Space group	$P2_1\ 2_1\ 2_1$	$P2_1\ 2_1\ 2_1$	$P2_1\ 2_1\ 2_1$	$P2_1\ 2_1\ 2_1$	$P2_1\ 2_1\ 2_1$	$P2_1\ 2_1\ 2_1$	$P2_1\ 2_1\ 2_1$	$P2_1\ 2_1\ 2_1$
a b c (Å)	48.7, 87.6, 136.9	48.4, 86.0, 132.3	48.6, 87.2, 136.2	48.3, 85.6, 131.7	48.8, 87.7, 137.2	48.7, 87.6, 137.0	48.4, 87.2, 136.7	48.7, 87.6, 137.7
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90
R_{merge}	0.127 (0.939)	0.206 (1.373)	0.167 (0.997)	0.144 (0.695)	0.106 (0.919)	0.165 (1.668)	0.097 (1.151)	0.106 (0.332)
R_{pim}	0.037 (0.412)	0.062 (0.390)	0.049 (0.424)	0.044 (0.338)	0.030 (0.346)	0.047 (0.549)	0.028 (0.360)	0.031 (0.122)
$CC_{1/2}$	0.999 (0.580)	0.996 (0.812)	0.997 (0.613)	0.996 (0.761)	0.999 (0.692)	0.998 (0.604)	0.999 (0.633)	0.998 (0.937)
Multiplicity	12 (5.9)	12.5 (13.2)	11.7 (6.3)	10.3 (4.5)	12.2 (7.3)	12.9 (9.4)	13.0 (10.3)	12.2 (8.0)
Completeness (%)	99.9 (97)	100 (100)	99.1 (89.6)	98.6 (83.9)	97.7 (77.8)	100.0 (98.9)	98.9 (86.0)	99.7 (92.6)
I/σ(I)	12 (1)	7.4 (1)	9.9 (1.1)	10.0 (1)	13.7 (1)	11.2 (1)	14.9 (1)	18.1 (3.8)
Unique reflections	56988 (2742)	18356 (912)	29539 (1305)	18844 (768)	55892 (2228)	47908 (2302)	47441 (2023)	21077 (962)

Refinement								
No. of refs used	56914	18306	29480	18793	55821	47837	47378	28240
R_{work}	0.182	0.221	0.195	0.197	0.181	0.191	0.190	0.167
R_{free}	0.211	0.307	0.241	0.272	0.208	0.220	0.225	0.203
Protein atoms	4533	4478	4622	4615	4582	4615	4589	4615
Ligands	60	18	22	18	31	33	24	25
RMS deviations								
Bond lengths (Å)	0.0074	0.0046	0.0069	0.0062	0.0079	0.0069	0.00637	0.0083
Bond angles (°)	1.4	1.2	1.4	1.4	1.4	1.4	1.4	1.5
Mean b values ($Å^2$)								
Protein	27	59	44	52	29	34	37	36
water	30	45	39	40	32	60	41	32
ligand	62	91	101	85	60	34	36	70

Ramachandran plot (%)								
Favoured	98	97	98	98	99	99	98	97
Allowed	2	3	2	1	1	1	2	3

Values in brackets refer to outer resolution shell.

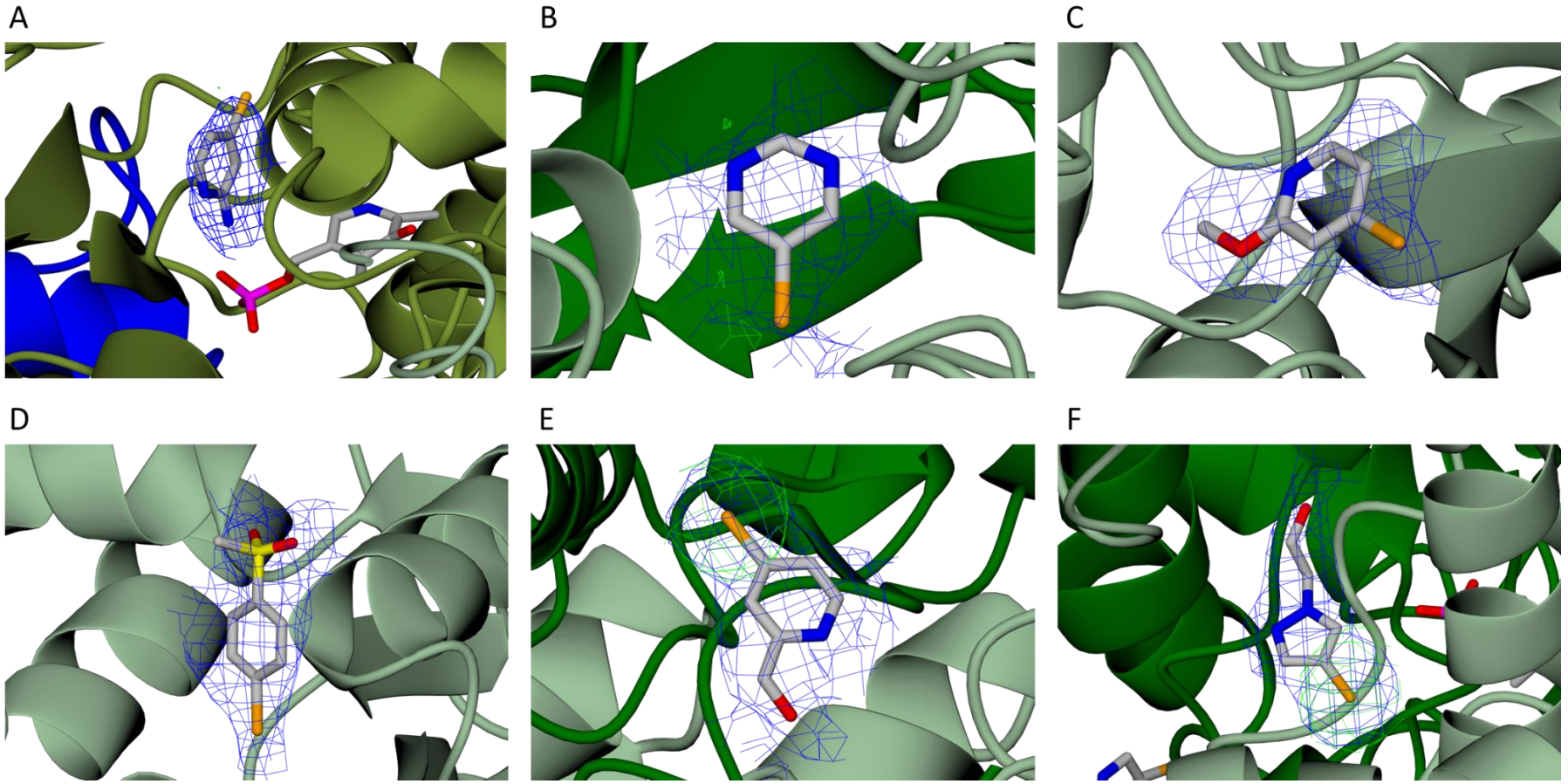
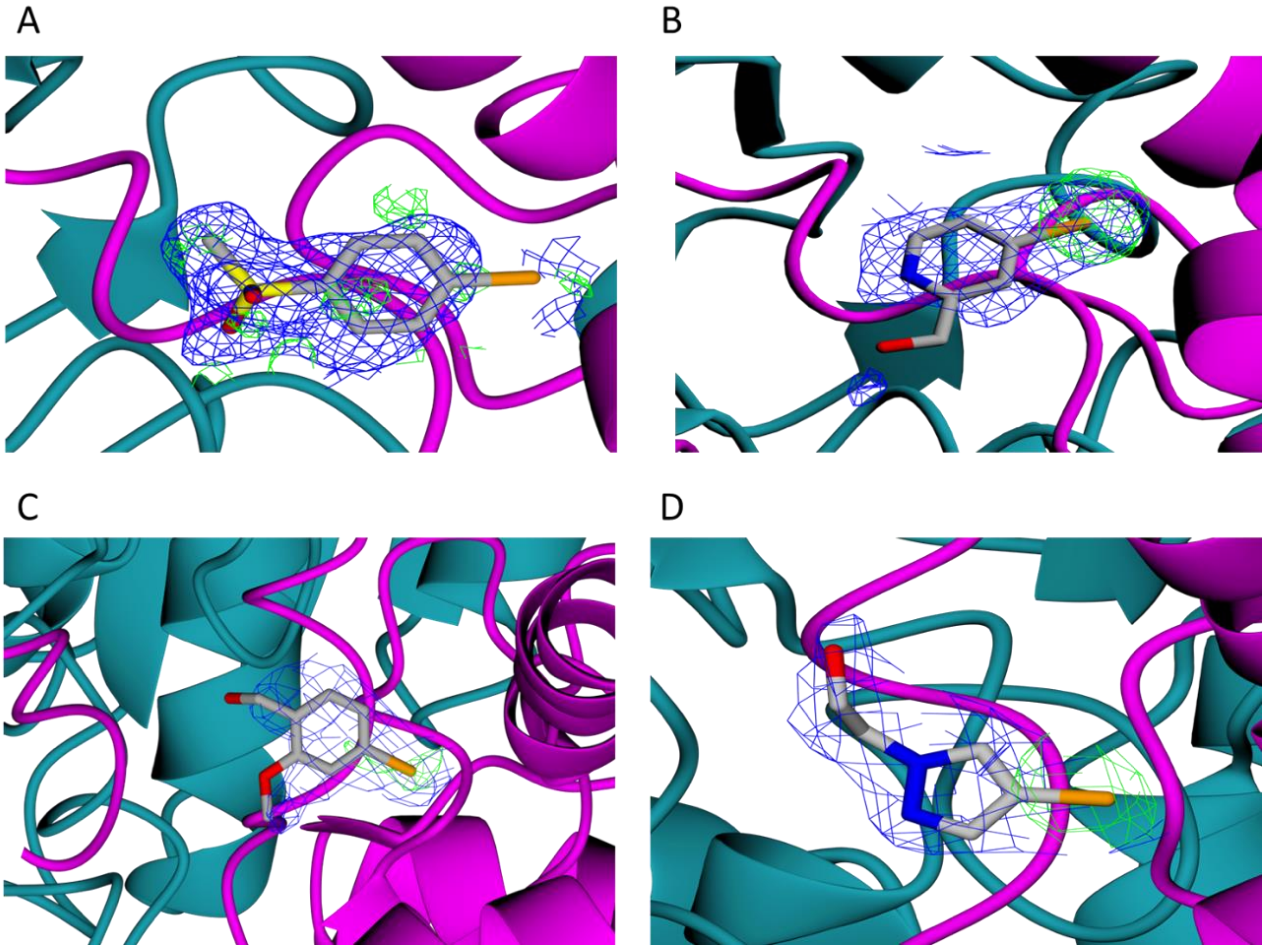


Figure S9-6-5 *Tc*CS fragments in $2F_o$-F_c density at 1 σ in blue. Anomalous difference F_o-F_c map shown at 3 σ in green.


A
B
C
D

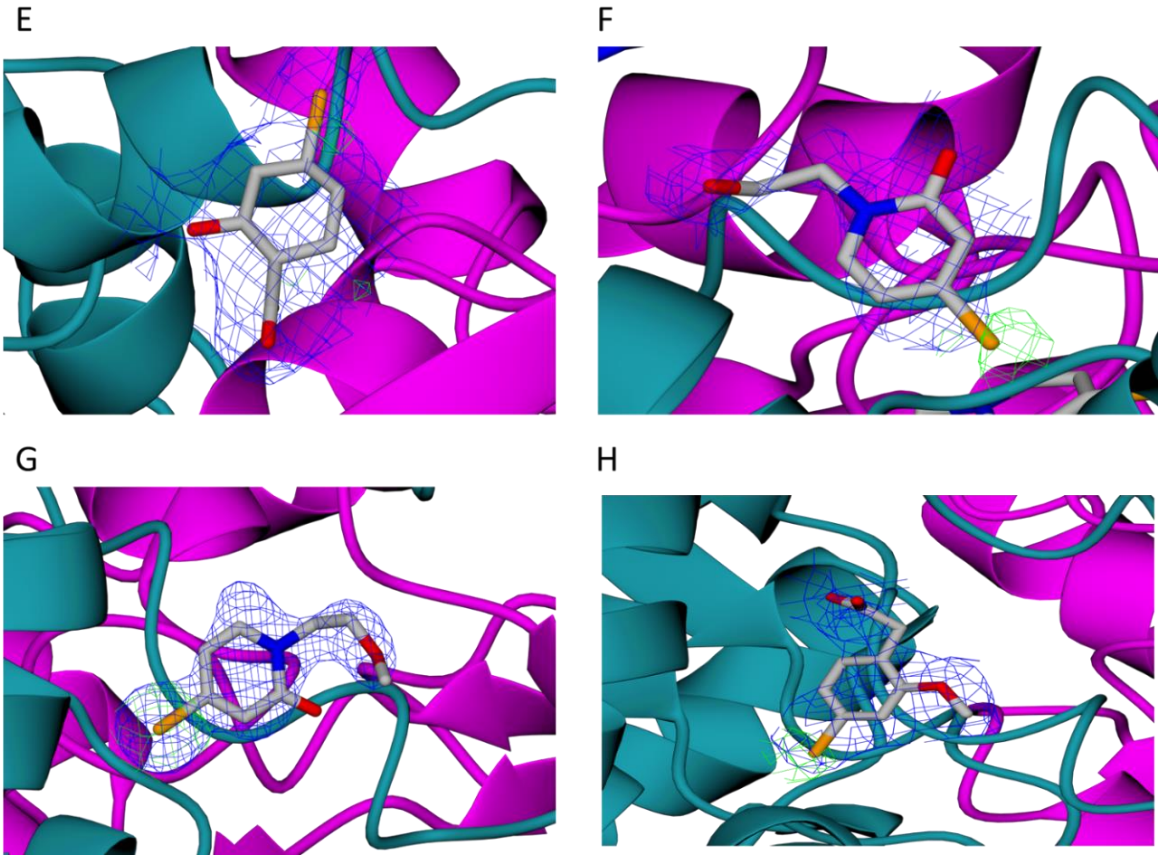


Figure S9-6-6 *Li*CS fragments in $2F_o$-F_c density at 1 σ in blue. Anomalous difference F_o-F_c map shown at 3 σ in green.