Towards X-ray structural analysis of *Thermus thermophilus* sulphite reductase: Purification characterisation and crystallisation

M. Sc. Thesis
Michael D. O'Donnell

Supervisor: Prof. Tewfik Soulimane

A thesis submitted to the University of Limerick in candidature for the degree of Master of Science. Submitted to the University of Limerick 2017

Abstract

The process of sulfur remediation is an ancient process credited with the origins of life, through the interconversion of sulfur molecule. The sulfite reductases play an essential role, allowing an organism to reduce sulfite, building a range of sulfur based molecules such as the amino acids and vitamins. The sulfite reductases can be divided into different groups. This depends on the number of subunits and whether the sulfite reductase operates via an assimilatory or a dissimilatory mode. While there have been a number of structures solved and a mechanism for this family proposed, the overall understanding of the function of this class of enzymes is still scarce.

Here, the sulfite reductase of *Thermus thermophilus* was studied, with the protein previously cloned BL21* *E. coli* cells. Initial work revolved around increasing the yield, through the optimization of the lysis and purification protocol leading to homogenous population of the sample required for structural studies. Research was also performed on the optimisation of the crystallisation conditions for the protein. Crystals obtained have been used for the structure determination by X-ray crystallography. The crystal structure has been solved to 2.38 Å resolution.

Over the course of the study the yield was increased to 24 mg/L (a 400% increase). The conversion of the dimeric species to the monomer was also achieved. The isolation of the monomer produced much higher resolution crystals than the mixed population samples with the best crystals being acquired from using 30 mg/mL of sulfite reductase with 12.5 mM sodium sulfite as the additive.

Declaration

I hereby declare that this is the work of my own investigations and that this thesis has not been submitted previously in this form or any other form to this or any other university in candidature for any higher decree

_____

Michael O'Donnell

Acknowledgements

First and foremost, I would like to thank my supervisor, Prof Tewfik Soulimane, for his constant support throughout the project for his guidance and support. His tireless efforts have helped me massively throughout this project. I would like to Dr. Mohamed Noor for training me in the tricks of the trade, for his constant aid throughout the project and for his work on this study. I would like to thank Dr. Edel Durack, who started this study and created the construct and designed the initial procedure. I would like to thank Dr. Jakki Cooney and Dr Todd Kagawa for aiding me in the lab. Onto this I would like to thank all the members of the L2001 lab from throughout my time here, especially Kevin, Lorraine and Daragh.

I am much obliged to the Irish Research Council for funding my research, throughout the last three years.

I am grateful to my friends, both old and new who have supported me throughout this time. Especially those who kept me from insanity, whether it was Summer, Trevor, Thomas, Markus, Jackson, Alex, Luis, Lenka, Jesus, Valerie, Lenny, Pavan, Kunal, Lea or Ander. Outside of this, there has to be a special mention to Enrico, en leu to him including me on his.

I am gratefully to my family for helping me relax when I was not working and without whom I surely would have not this far

Table of contents

Abstract ii

Declaration iv

Acknowledgements vi

Table of figures x

List of abbreviations xii

1.0 Introduction 1

1.1 Life in anaerobic environments and the role of sulfur 1

1.2 Sulfite reductases 6

1.2.1 The role of siroheme 15

1.2.2 The Fe-S cluster 17

1.3 *Thermus thermophilus* 18

1.4 Crystallisation 19

1.5 Aims of this study 23

2.0 Material and methods 24

2.1 Materials 24

2.1.1 General equipment 24

2.1.2 Reagents used in this study 24

2.1.3 Crystal screens 24

2.2 Methods 25

2.2.1 Protein expression trials by auto-induction 25

2.2.2 SDS-PAGE 25

2.2.3 Western blot 26

2.2.4 Lysis of cells for SUR purification 27

2.2.5 Immobilized metal affinity chromatography 27

2.2.6 Size exclusion chromatography 28

2.2.8 Bicinchoninic acid assay 28

2.2.7 Protein crystallisation 28

2.2.8 Bioinformatics 29

3.0 Results and discussion 30

3.1 Bioinformatical analysis of the target 30

3.1.1 General bioinformatics 30

3.1.2 Phyogenetics and alignment 33
3.2 Purification and achieving homogeneity 34
3.2.1 General purification 34
3.2.2 Modifications to lysis 38
3.2.3 Modifications to size exclusion protocol 41
3.2.4 Conclusion on purification 45
3.3 Crystal Trials 46
3.3.1 Initial crystallisation 46
3.3.2 Optimization and the additive screen 47
3.3.3 The use of substrates in order to optimize crystallisation 55
3.3.4 Crystallisation conclusion 55
3.4 Crystallography 57
4.0 Conclusion and future work 62
5.0 Bibliography 64

Table of figures

Figure 1.1: Conversion of sulphur containing molecules in hydrothermal environments.....2
Figure 1.2: Common sulfite pathways.....3
Figure 1.3: Basic overview of the sulfur cycle.....5
Figure 1.4: β-subunit of *E. coli* sulfite reductase.....10
Figure 1.5: Electron transfer and reduction of sulfite.....12
Figure 1.6: SUR from maize complexed with ferredoxin.....14
Figure 1.7: Siroheme structure.....16
Figure 1.8: Active site structures of the β-subunit of *E. coli* SUR.....17
Figure 1.9: Protein crystallisation phase diagram.....20
Figure 3.1: Blast domain results for SUR from *Thermus thermophilus*.....31
Figure 3.2: Phylogenetic tree of selected SUR homologs from a range of different bacteria.33
Figure 3.3: SDS-PAGE analysis of IMAC samples.....36
Figure 3.4: Western blot carried out after SDS-PAGE analysis.....36
Figure 3.5: Initial SEC of the SUR protein after purification.....37
Figure 3.6: SDS-PAGE analysis of SEC samples.....38
Figure 3.7: Graphical representation of the results acquired for the lysis optimization.....40
Figure 3.8: Re-chromatography of monomer peak (A) and dimer peak (B).....42
Figure 3.9: Chromatography of peak before and after incubation with SO3.....44
Figure 3.10: Initial hits from the index screen.....46
Figure 3.11: Crystal hits from the additive screen.....48
Figure 3.12: Results of additive screen.....49
Figure 3.13: Results of the additive screen with 30 mg/mL SUR.....53
Figure 3.14: Additive screen optimization results.....54
Figure 3.15: Crystals acquired from interaction with substrates.....55
Figure 3.16: X-ray diffraction pattern of *T. thermophilus* SUR.....57
Figure 3.17: Structure of *T. thermophilus* monomer.....58
Figure 3.18: Active site of *T. thermophilus* SUR.....59
Figure 3.19: Active site of *T. thermophilus* SUR.....59

List of abbreviations

ADP – Adenosine diphosphate

AMP – Adenosine monophosphate

APS – Ammonium persulphate

ATCC – American Type Culture Collection

ATP – Adenosine triphosphate

DNA – Deoxyribonucleic acid

EDTA – Ethylenediaminetetraacetic acid

HPLC – High-Performance Liquid Chromatography

IMAC – Immobilised Metal Affinity Chromatography

IPTG – Isopropyl β-D-1-thiogalactopyranoside

MW – Molecular Weight

NADH – Nicotinamide adenine dinucleotide

NADPH - Nicotinamide adenine dinucleotide phosphate

NAG- N-acetyl-D-glucosamine

NAM - N-acetylmuramic acid

ORF – Open Reading Frame

PCR – Polymerase Chain Reaction

PDB – Protein Data Bank

RNA – Ribonucleic acid

SDS – Sodium Dodecyl Sulphate

SDS-PAGE – Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis

SEC – Size Exclusion Chromatography

SUR – Sulfite Reductase

TAE – Tris-acetatae-EDTA

TBS – Tris Buffered Saline

TEMED – N,N,N', N' - tetramethylethylenediamine

1.0 Introduction

1.1 Life in anaerobic environments and the role of sulfur

In this world of ours, there is one thing we can be said regardless of our views or of our beliefs; that life is amazing. Whether this is in the troves of our highly hospitable troposphere, in the seemly inhospitable depths of the darkest deep sea trench or like in the case of the organism under investigation *Thermus thermophilus* in centre of a volcanic hot spring, life always finds a way to survive; not just survive it thrives, with billions of organisms in existence and more being formed every single day. In fact, life is so prevalent that the Drake equation, the equation set out by Dr. Drake in the 1960s to predict the likelihood of life, doesn't tell us just that there is life on other planets in the galaxy we are in, but tells us that statically, there are at least 10 other planets in our galaxy alone processing intelligent life (Bracewell, 1976). This does raise the question of how all of this is possible. How can any living organism do this? How does a bacterium survive in an atmosphere that for centuries, we had believed was inhospitable; such as a hot spring, where our flesh would burn or in an anoxic atmosphere where we as a species would merely suffocate and die (Bertoldo et al, 2002)? There is no magic behind this, no divine miracle. They simply adapt. Sometimes this involves adapting proteins involved in membrane stability (van de Vossenberg et al., 1998). Sometimes this is achieved by having enzymes with higher activity at higher temperatures, higher salt concentrations or higher atmospheric pressures. Sometimes, this process of adaption involves utilizing a new substrate or a new electron donator. But in all cases, they adapt.

One of the main ways life has successfully adapted has been through the use of the elements and molecules readily available to it. This is especially the case when the usual components of life are missing. The use of sulfur based molecules is an example of this. The ability to utilize sulfur in the respiratory chain instead or oxygen has allowed a large number of bacterial and archaea species survive in what would be otherwise inhospitable environments, where anoxic or oxygen limiting conditions occur (Jorgensen & Postgate, 1982). Even outside of this substantial evolutionary advantage, being able to process sulfur, in a number of different forms and oxidation states, is needed for the cell to perform a large number of processes. Among these are protein construction (being a key component of the amino acids cysteine and methionine) (Lehninger et al., 2008), ensuring protein stability (disulphide bonds and through sulfation increasing protein-protein interactions (Lehninger et al., 2008)),

and the establishment of a viable energy store (Willey et al., 2013). Due to this, one of the most important processes in nature has to be the sulfur cycle.

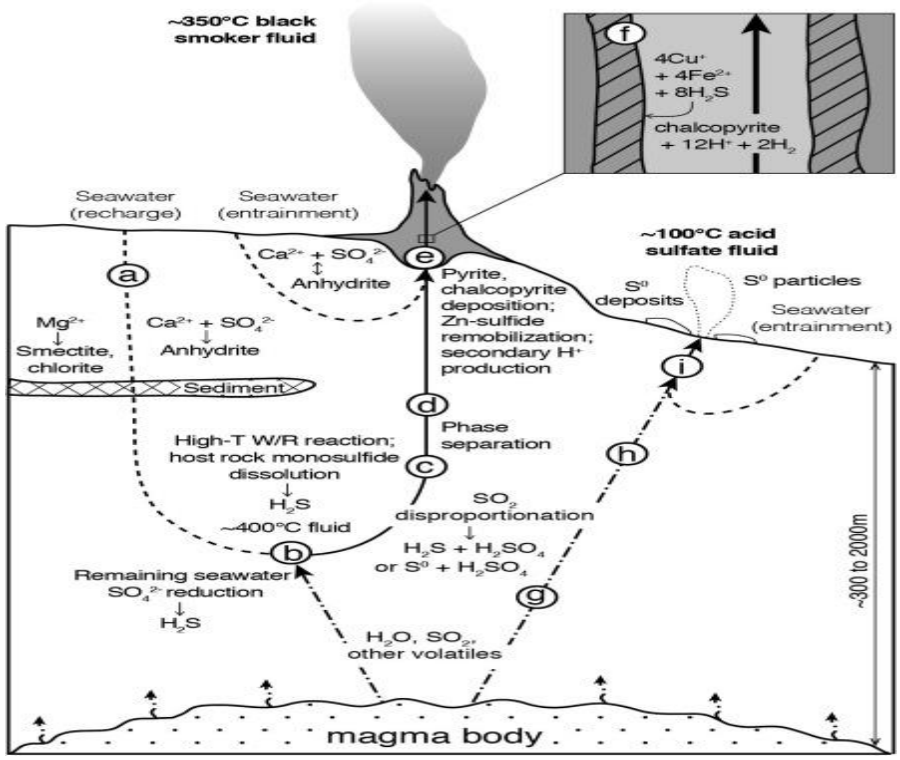

Figure 1.1: ***Conversion of sulphur containing molecules in hydrothermal environments***

Sample non-microbial based sulfur cycle, showing the different conversions observed in a black smoker (McDermott et al., 2015)

Prior to discussing the microbial involvement in the sulfur cycle, it is important to note that the sulfur cycle itself does not actually need an organism to proceed. This makes it rather unlike a large number of other cycles, such as the carbon cycle or the nitrogen cycle. In both these cases a large number of different enzymes are needed to interconvert the major molecules and without microbial involvement, they cannot occur (Mathews et al., 1999). The sulfur cycle is different due to the major components of the sulfur cycle being a lot less structural complex. This means they are statistically more likely to occur without a guiding force. Due to this, the cycle has been found to occur as long as the right conditions of pH and oxidation-reduction are present, even in a sterile environment. This theory explains why the sulfur cycle occurred for millions of years prior to the origin of life. A lot of current data

suggests that this cycle may have been the primary driving force behind the origin of life, supplying a readily accessible energy gradient (Kasting et al., 1989). The process in question, originally made possible by the heat and turbidity (due to a large concentration of a number of chemical species) of the early oceans, can still be seen, in a number of locations, around the planet today. This is evident in the cases of thermal springs and the areas surrounding white and black smokers with a basic example of this shown in fig 1.1 (McDermott et al., 2015). In these areas, comparable to what our planet experienced in its early stages, all parts of the cycle can be seen occurring and a number of more complex and higher energy sulfur based molecules can be readily found (Wachtershauser, 1998). Even today on the sea floor, a rather extreme environment, a readily accessible supply of sulfur can be found. This time in the form of anhydrite and pyrite, both found regularly mixed through shale. Sulfur molecules are periodically added from these into the cycle through to erosion and leaching by sea water from the rock bed (Johnston et al., 2006).

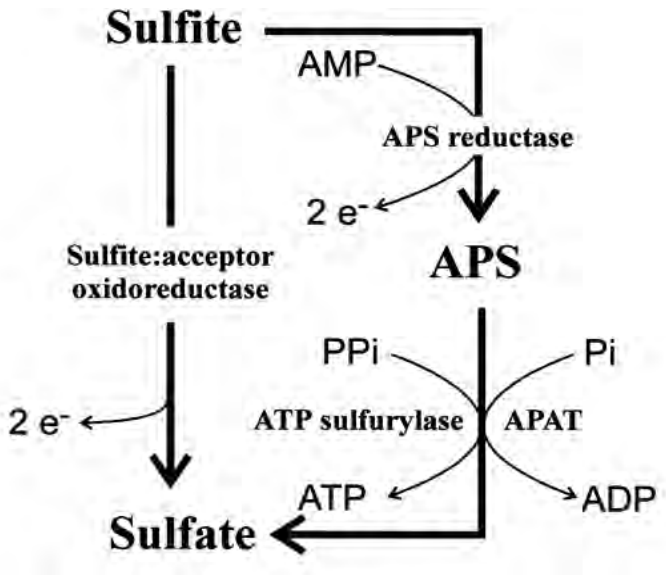

Figure 1.2: ***Common sulfite pathways***
Potential sulfite pathways in the cell (Kappler & Dahl, 2001)

This massive presence of sulfur based molecules in the oceans is relatively low compared to what would have been seen in the early oceans. High temperatures and acidic conditions promoted these conditions. This high level of sulfur along with the presence of all the other needed chemical species, such as carbon and nitrogen sources available in the pre-life stew, probably explains how early earth's early existence as an inhabitable planet. This perhaps explains why all organisms have enzymes to utilise sulfur molecules in one way or another

(Kasting et al., 1989). These sulfur sources are as readily available now as they were then for organisms to utilise, with organic and inorganic sulfur molecules available in soil and in aqueous environments. There are also a small number of sulfur based molecules occurring at low concentrations in the atmosphere.

Outside of this naturally occurring part of the cycle, microbes play an important role in the sulfur cycle. This is through the production of larger, more significantly complex and higher energy molecules such as the amino acids and a number of vitamins. There are a number of different layers to this cycle with numerous microbes reducing sulfate for their own uses, whether to produce amino acids or for energy, and then a group of larger organisms utilising these as a food source and get their supply of sulfur based molecules through this. Plants have also developed methods to achieve this. Occasionally they do this themselves. On other occasions they form symbiotic relationships with different microorganisms. The importance of this cycle to microbial life is evident with sulfur being the 10th most abundant element in the universe but the 6th most popular element in microbial life showing the abundance of its use (Klotz, 2011).

In the majority of cases, the starting point for the microbial sulfur cycle is the processing of sulfide or the processing of thiosulfate. These have been found to be the two most readily available inorganic forms of sulfur to microorganisms on the planet (Bickle et al., 1994). This reduction occurs normally via a mechanism known as assimilatory sulfate reduction (as opposed to dissimilatory sulfate reduction which happens when sulfate is processed in an anoxic environment) (Willey et al., 2013). In the case of assimilatory reduction, the sulfate molecule is usually used to generate a range of different sulfate containing organic compounds such as polysaccharides, phenols and steroids. These molecules are then stored in the cell until they can be later further broken down (Schiff & Fankhauser, 1981). The process of dissimilarity sulfate reduction on the other hand can be hard to define. This process both describes the conversion of sulfate to sulfide for energy gain and any set of pathways used by the cell to take up, release or convert sulfate to any sulfur intermediate as long as these can be broken down to sulfite. These of course exclude cases covered by the assimilatory reduction (Kappler & Dahl, 2001).

Multiple examples of both of the above reaction types occur within a number of different organisms throughout eukaryotes, prokaryotes and archaea. An example of interest here, showing the importance of the cycle, occurs in bacterial and archaea species, when an

organism uses sulfate instead of oxygen as the terminal electron donator (Kasting et al., 1989). This process of anaerobic respiration in recent history has become relatively common occurring in a large range of bacteria and other microorganisms. In these cases, sulfide acts as the electron source for anoxygenic photosynthetic organisms as well as for chemolithoautotrophs such as in the *Chlorobi* and *Thiobacillus* species (Willey et al., 2013). This allows organisms to survive in areas where there would struggle due to a lack of oxygen or a total absence (Schiff & Fankhauser, 1981).

This leads to one of the final species in this whole cycle, the molecule of interest of the study and one of the most important intermediate species in the cycle, sulfite or SO_3^{2-}, a piece of the dissimilatory sulfur pathway (simplified in fig 1.3). This molecule can be reduced to form the violate dimethylsulfide, the starting block of the anabolic cycle. From here, the cell can build a number of different molecules such as dimethylsulfonopropionate (used by a number of bacterioplankton species as a method of energy storage), sulfur stores, sulfonated polysaccharides, certain amino acids, vitamins and a number of S based lipids (Hell et al., 2008).

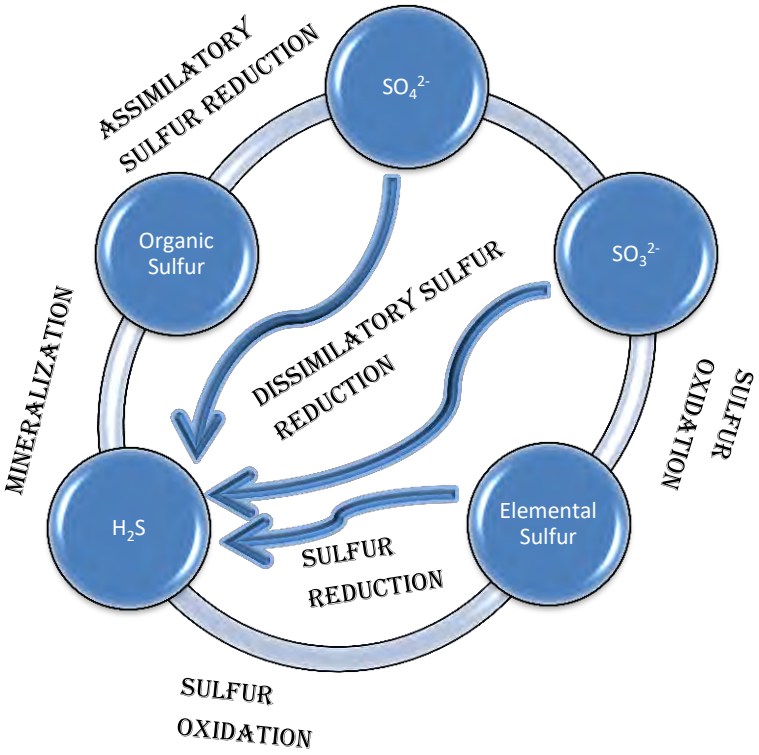

Figure 1.3: ***Basic overview of the sulfur cycle***
Overall sulfur cycle showing all the potential processes involved and the terminology used to describe each (Willey et al.,2013)

1.2 Sulfite reductases

After briefly examining the sulfur cycle as a whole, it quickly becomes apparent why the cell needs sulfur, especially considering sulfur is needed for a range of different processes within the cell (Klotz, 2011). There are a vast range of biological compounds containing sulfur, including the vitamins biotin and thiamine; the cofactors S-adenosyl-L-methionine, coenzyme A, the molybdenum cofactor (MoCo), and lipoic acid; the chloroplast lipid sufloquinovosyl diacylglycerol; the amino acids cysteine and methionine and many secondary compounds such as lipids, steroids and polysaccharides (Hell et al., 2008; Leustek, 2002; Schiff & Fankhauser, 1981). While there may be a range of different biological molecules common to all organisms and there may be a single global sulfur cycle, sulfur processing is, of course, like a lot of things in nature, not a one size fits all solution by any stretch of the imagination (Hell et al., 2008). There are a number of enzymes involved in this, with a number occurring at very high frequencies throughout nature and others unique to certain populations of microbes. Outside of this, there is even a level of diversity among the same enzyme in different organisms within this cycle. For example the protein in this study, the sulfite reductases show a large variation from organism to organism, in regards to amino acid sequence, number of subunits and oligomer state (Crane et al., 1995; Hell et al., 2008; Kim et al., 2016). This of course does make sense, as not every cell is exposed to the same environment. Hence they need to adapt. If this was not the case, all life would occur as some large unicellular being with a small number of proteins occupying a single small niche. Onto this, the fact is there is competition means that it pays to be different.

Sulfur of course has many uses within the cell and the organisms have only a limited amount of each of the different forms of sulfur (with sulfur existing in various oxidation states in Earth's biosphere such as S^{0} (elemental sulfur), S^{6+} (Sulfate), S^{4+} (Sulfite) and S^{2-} (Sulfide) (Amend & Shock, 2001)) available to the cell. Therefore it makes sense that an organism would have developed a subset of proteins designed to be able to compete with other organisms to make sure the cell gets enough of each. Analysis of this concept, explains why there is so much variation between what should be proteins within the family, with everything from the enzyme's affinity to the kinetics being defined by very small alterations in structure or amino acid sequence. Studying the case of the sulfite reductases in particular, this family of proteins can probably be regarded as being from 2 different sub families of the protein. Members in each case either being classified as an assimilatory or a dissimilatory sulfite reductase. This depends on a number of areas such as the biological function,

spectroscopic characteristics and the catalytic properties of the protein (Crane & Getzoff, 1996).

Prior to investigations into what is currently known about the sulfite reductases, it is necessary to consider the substrate. In this case it is usually sulfite. Sulfite itself is an anion, since it contains a negative charge, yet unlike a large number of traditional anionic molecules, this charge is not fixed. It can occur on any two out of the three oxygen atoms in the structure at any one time with the charge resonating between the three possible confirmations at any one time (Gillespie, 1992). This means there will be a S=O and two S-O bonds present at any one time and this combined with the lone pair of electrons on the sulfur atom. This implies that this molecule should occur as a trigonal pyramidal molecule similar to the VSPER standard of ammonia, and of course this is the case (Powers & Olson, 1980).

While there are cases where sulfite is only processed in one way in certain cell lines, especially in the cases of a number of eukaryotic cells, sulfite as a whole can be reduced in many bacteria and Archaea in both a dissimilatory and an assimilatory way (Kappler, 2011). In the case of the dissimilatory method of sulfite reduction, the cell normally utilizes the same subset enzymes as in sulfate assimilation, revolving around ATP sulfurylase, APS reductase, and sulfite reductase in a form of a cascade. These firstly creating sulfite and then proceed to reduce it. In the case of assimilatory reduction, this tends to be much more simplistic due to the presence of oxygen unlike dissimilatory reduction which can be complicated due to it occurring in oxygen limiting environments (H. D. Peck, 1959). All the proteins involved in these pathways work in a comparable way, resulting in the same end product. The overall structures of each family has been found to be very different (Wachtershauser, 1998). These do raise the question of why the cell does not just store the sulfite instead of using a combination of different proteins to process sulfite either by catabolic or anabolic means. This is because sulfite cannot be stored in large quantities in the cell (Niknahad & O'Brien, 2008). While in small amounts, sulfite can be beneficial to the cell, with the mitochondria in eukaryotic cells showing a high affinity for it using it as an energetic substance, but this is only at low micro molar levels. At higher levels it proves toxic for the cell, due to the build-up of reactive oxygen species (ROS) and other free radicals. Due to this, there is the need to either break down sulfite to something the cell can deal with or build it into something the cell can safely store (Goubern et al., 2007). While this may seem like a lot of hassle, it is an unavoidable process being the only way that a cell can build the S containing amino acids,

meaning the only way to acquire the essential amino acid cysteine, without consuming it from another source or acquiring it from the breakdown of another protein. This process occurs via an organic synthesis from a combination of sulfite and O-acetylserine (Quentmeier et al., 2000).

It is possible to write a few pages if not a full book (there are a few examples of these) describing all the enzymes that utilise sulfur or even just the enzymes involved in these 2 pathways. A few more chapters could be written explaining each of their structures and how these enzymes work. Regardless of this, the main point discussed in this piece is not the cycle but the sulfite reductase family itself. As a whole, family members of the sulfite reductases act as a major source of sulfite remediation reducing sulfite to hydrogen sulphide (Crane et al., 1995), with this set of enzymes catalysing the final complex SO_3 six-electron reduction in accordance to the equation below to hydrogen sulfate and water

$$\mathbf{SO_3^{2-} + 6e- + 8H^+ \rightarrow H_2S + 3H_2O}$$ (Crane & Getzoff, 1996)

It is also possible for the protein to also preform assimilatory nitrite reduction, catalysing the six-electron reduction of nitrite to ammonia shown below:

$$\mathbf{NO_2- +6e- +7H^+ \rightarrow NH_3 +2H_2O}$$ (Wagner et al., 1998)

This 6 electron reaction in both cases is made possible by a number of items of interest in the protein's structure, such as the Fe-S cluster (McRee et al., 1986) and the presence of the siroheme (Murphy et al., 1974). Both of these will discussed later on in this piece.

Not all sulfite reductases are the same and like the pathways can be classified as either assimilatory or dissimilatory, depending on a number of factors. What, though in this case, is the difference? To begin with, assimilatory sulfite reductases are normally found to only exist in bacteria, fungi and plants with none, as of yet, ever being reported in an animal species (H. Peck & Lissolo, 1988). Dissimilatory, on the other hand, have been so far observed in organisms found either to be quite morphologically diverse or quite metabolically diverse sulfate-reducing bacteria as well as in a number of thermophilic species (Fauque, 1995). While these two subfamilies of sulfite reductases are exceptional similar, they differ in the end function. Assimilatory sulfite reductases generate sulfide, so that the cell can incorporate it into sulfur based molecules. The dissimilatory sulfite reductases reduces sulfite in order to utilize it as the terminal redox coupling of the respiratory cycle (Crane & Getzoff, 1996). Therefore it has been clearly seen that both these types of sulfite reductase are required in

order to allow for both biosynthetic assimilation of sulfur and for the dissimilation of oxyanions (such as sulfate) in order for the cell to use them for energy conservation (LeGall & Fauque, 1988). While the two distinct forms of the enzyme have a number of structurally and functionally differences, there remains a lot in common. This is due to the common aim in both cases being to reduce sulfite. This can be used to explain why both the dissimilatory and the assimilatory sulfite reductase contain the siroheme iron-sulfur cluster, made up of a reduced polyphyrin ring (the siroheme) connected to the [Fe_4S_4] cluster by a cysteine thiolate (with both this cluster and this siroheme described later on in the piece) being a crucial component of sulfite reduction (Crane et al., 1995; Crane et al., 1997; Kim et al., 2016; Murphy & Siegel, 1973).

The assimilatory sulfite reductases subfamily, have never to date been reported in animals, but have been found in bacteria, fungi and plants (H. Peck & Lissolo, 1988). The reason for this, in a very simple sense, is that the organisms in question here are capable of sulfate uptake and consequently need a method to reduce it in order to incorporate them into larger molecules. It has been found that both S containing amino acids, cysteine and methionine are manufactured in the cell through the assimilatory sulfite reductase containing pathway (Hell et al., 2008). The beginning of this pathway (shown in fig 1.2 in the APS containing side of the schematic) commences through the adenylation of sulfate. This has been taken up by the cell, and is used to generate adenosine 5'-phosphosulfate (APS) (Kappler & Dahl, 2001). This in turn is directly reduced to the substrate of interest here, sulfite, through the reaction of this APS by the enzyme APS reductase (Hell et al., 2008) or through the intermediate step of phosphorylation by APS kinase followed by a reduction of this product (Crane & Getzoff, 1996). Here is where the enzyme of interest finally comes into play in order to convert sulfite to sulphide. This can be then built into the amino acid precursors of O-acetylserine or O-acetylhomoserine, or used to build a range of different S containing compounds (Kappler & Dahl, 2001).

While the assimilatory reaction tends to occur in the presence of oxygen, the dissimilatory sulfite reductases, tends to operate under strict anaerobic conditions, playing an important role within the biogeochemical sulfur cycle (H. D. Peck, 1959). These proteins, observed as both single and double subunit proteins, have been identified in a range of different organisms occurring in a wide range of different environments (Crane et al., 1995; Nakayama et al., 2000). The reaction occurring here requires the transfer of six electrons provided from an electron donator. In the case of chemotrophic organisms this normally comes in the form of nicotinamide adenine dinucleotide phosphate (NADPH) and in phototrophic organisms as ferredoxin. As in the case of the assimilatory sulfite reductases, the dissimilatory version results in the complete reduction of the sulfite to sulphide (Hell et al., 2008).

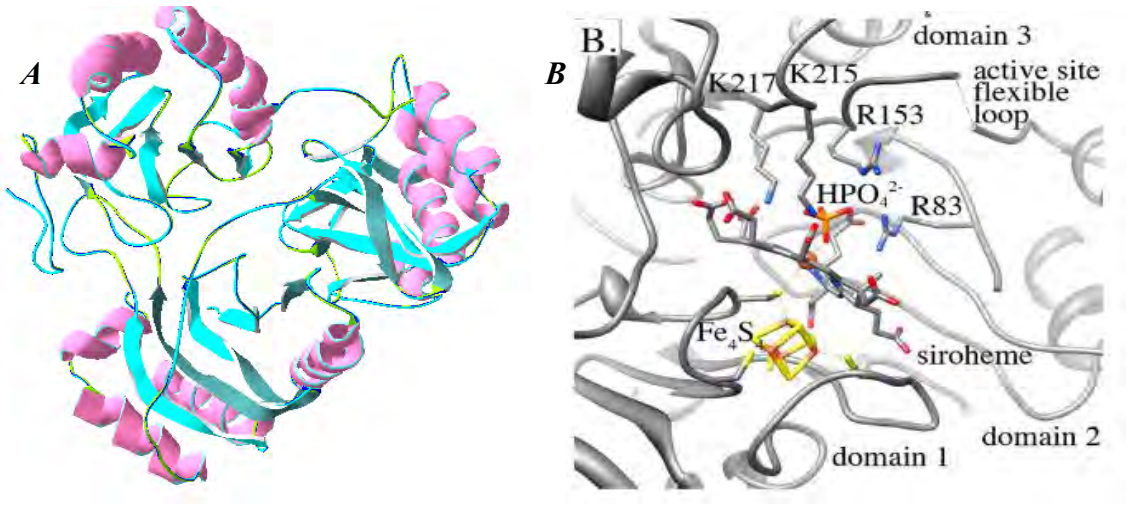

Figure 1.4: *β-subunit of E. coli sulfite reductase*
Crystal structure of the β-subunit of the E.coli sulfite reductase to a resolution of 1.6 Å (PDB ID:4HTR), showing the whole structure (A) and active site, showing the Fe-S cluster in yellow, the siroheme and the key amino acid residues interacting with a protonated phosphate molecule (B) (Smith & Stroupe, 2012)

In the case of this subfamily of proteins, a lot of what is understood about their interactions comes from the work done by Brian Crane and Elizabeth Getzoff. This is from both their early work and eventual solution of the crystal structure of the 2 subunit SUR of *E. coli (*Crane *&* Getzoff*, 1996;* Crane *et al., 1995)*. This structure, as one of the earliest if not the earliest member of this family's structure to be published, allowed us to see for the first time how exactly a SUR like protein might work. What Crane and Getzoff eventually saw was a complex, containing 8 x 66 kDa flavoprotein subunits (now referred to commonly as the α-subunit) and 464 kDa hemoprotein subunits (now commonly referred to as the β-subunit, shown in fig 1.4) (Crane et al., 1995). When these were finally isolated, it was found that

these two subunits had massively different structure and functionality. The α-subunit was found to accept electrons from NADPH. The β-subunit was found to then accept these electrons from the α-subunit, catalysing the eventual reduction of the substrate, sulfite or nitrite, in accordance to the equations set out previously (Crane & Getzoff, 1996).

The α-subunit has since been found to only occur in a subclass of the sulfite reductases. Some propose that this is a recent addition to aid in the electron transfer to the active site of the protein, explaining its absence in a large number of cases. On the other hand, the β-subunit has been found throughout the family and in this case is the *Thermus thermophilus*, unlike the α-subunit, has a homolog.

The β-subunit of the protein is the only place where sulfite reduction actually occurs, with the cofactors required to perform this and the amino acids involved in this process all being bound at the interface of three of the domains located in this subunit (Crane et al., 1997). This interface, composed of a number of β-sheets flanked by α-helices, is where the siroheme and the Fe-S cluster both are located (Crane & Getzoff, 1996), bound to a number of highly conserved cysteine residues. Outside of this, there is also a phosphate anion interacting directly with Arg residues located at positions 83 and 153 on the backbone and the Lys residues located at positions 215 and 217. These residues interact with the substrate whether it is sulfite or nitrite (Smith & Stroupe, 2012). All of the features listed here are shown in fig 1.6 and between these 3 major contributors, the 6 electron transfer, which facilitates this reduction occurring, occurs and this is what is shown in fig 1.5.

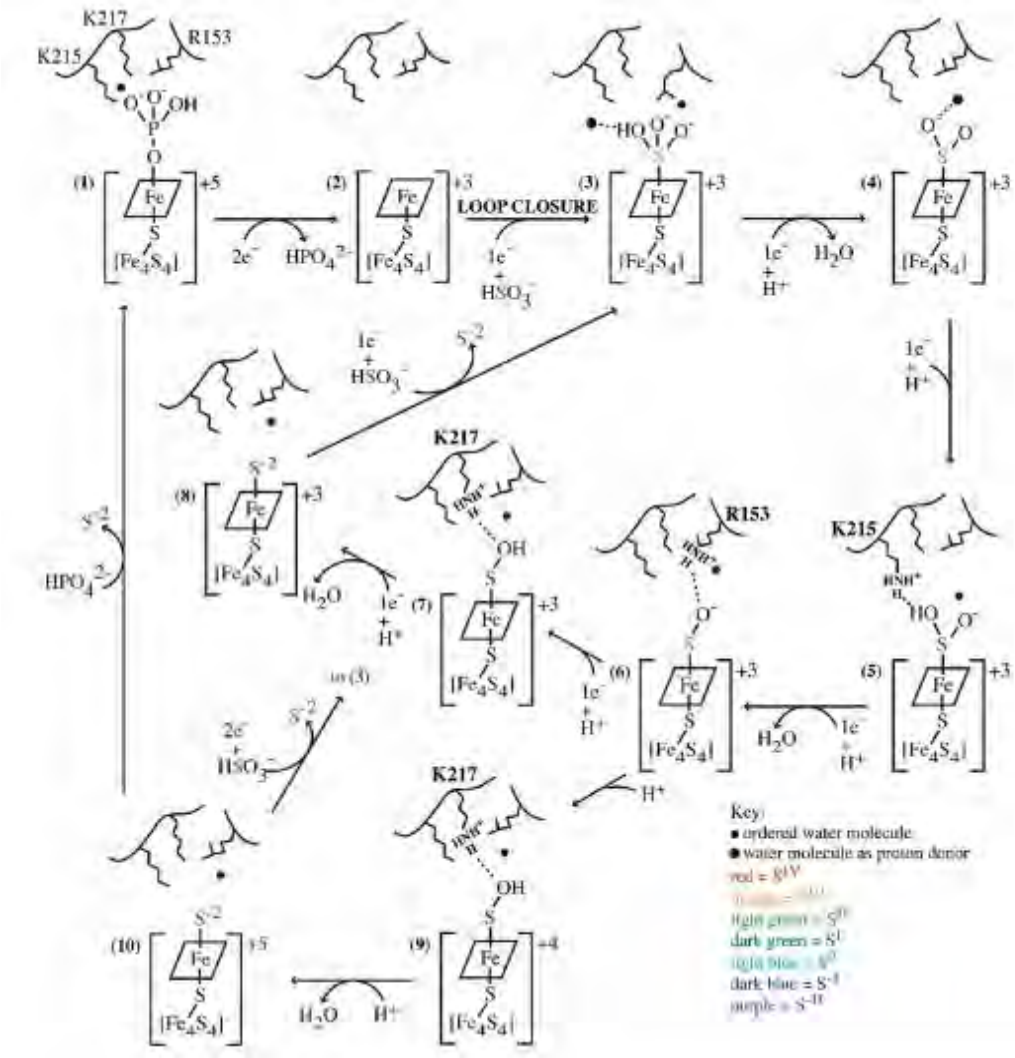

Figure 1.5: ***Electron transfer and reduction of sulfite***
Reduction of sulfite as proposed to occur in the β-subunit of E. coli Sulfite Reductase's active site, showing the interactions of the substrate with siroheme and residues K217, K215 and R153 (Smith &Stroupe, 2012)

So how does the reaction proceed? To start with, this proposed six electron reduction occurs with the electrons moving through a number of different co-factors within the active site, initiating the reduction with the system being reduced and the HPO_4^{2-}at the active site being released (Smith & Stroupe, 2012). This step is proposed to allow for the binding of the protonated substrate HSO_4^-. This now binding in place of the phosphate group interacting with Arg residues 153 and Lys residues 215 and 217 and the siroheme to lock it into the active site (Nakayama et al., 2000). Before reduction can occur, this molecule is first protonated, through the transfer of a H^+ from a water molecule (shown as a black dot in fig 1.5) located in the active site, to the protonated substrate. This is followed by a second

dehydration event with the substrate being protonated by a second water molecule in the active site. The Lys at position 215 then donates a proton to promote the second dehydration event. This leaves the protein with one oxygen to remove. It does this by first having this final oxygen group protonated through the donation of a proton from the arginine at position 153. Once fully protonated by the 215 lysine, the final dehydration event occurs and the S^{2-} is released. This allows for either the phosphate group to return to the active site or for the binding of another molecule of substrate (Smith & Stroupe, 2012).

While a lot of work has been done recently to figure out the full mechanism of the sulfite reductases, work on the structures of these proteins are still rather limited. Current focus has been on a number of different plant sulfite reductases. This is as result of food supply rapidly becoming a global issue, driven by the use of certain plant extracts as biofuels and a rapidly growing global population. An example is the sulfite reductase of maize. Maize, being the plant from which we get corn, is a stable food sources in a large number of areas across the world, used as food, animal feed and for biofuel. The sulfite reductase of maize, published in 2010, contained an active site and a siroheme/Fe-S cluster binding pattern comparable with other solved SUR structures. While this is expected, what made it unusual is the fact that the SUR here had a number of similarities to what was seen in this project with the SUR of *Thermus thermophilus*. Unlike the *E.coli* member of sulfite reductase family, the maize sulfite reductase seems to only contain a single peptide, with the protein of interest believed to occur with only a single subunit. It lacked the α-subunit and occurred as a 65 kDa monomer, in agreement to what was seen in this project (Nakayama et al., 2000). It still contained the key aspects of the SUR family. The set of 4 cysteine residues, all of whom are associated with the Fe-S centre but one of whom is also associated with the siroheme, occurred in the same location as the *E. coli* SUR β-subunit (Crane et al., 1995; Nakayama et al., 2000). The conserved lysine and arginine residues also remained at the same positions in the active site (Nakayama et al., 2000). Unlike *E. coli* and a number of other reported sulfite reductases, it was seen as a monomer (Kim et al., 2016). This was also the case for the protein of interest. There are a number of other areas where this protein could potentially be similar. Ferredoxin, the electron donator for the maize version of the protein, has been reported in *Thermus thermophilus,* suggesting this could be the electron donator in both cases (Kim et al., 2016).

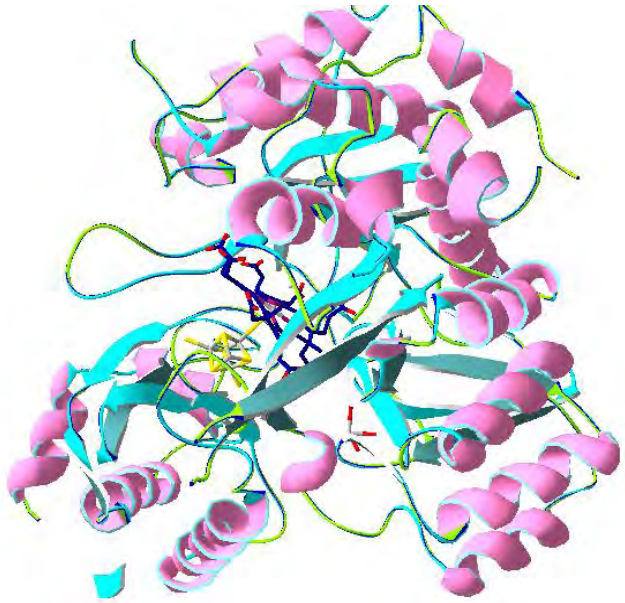

Figure 1.6: ***SUR from maize complexed with ferredoxin***
Crystal structure of the SUR from maize, complexed with ferredoxin with a resolution of 2.2 Å (PDB ID: 5H8Y) (Kim et al., 2016)

Of course these are more than 2 structures available for the sulfite reductases in the protein database. Analysis of the database reveals 4 other structures, 2 of whom contain the α-subunit like *E. coli* and 2 which do not, similar to maize (*Zea may*). With only 6 structures classified in this family, it is possible to see the scope for work in this area

Table 1.1: *Proteins classified as Sulfite Reductases found in the Protein Data Bank*

Protein	Organism	Presence of α-subunit	Citation
5H8Y	*Zea mays*	No	(Kim et al., 2016)
2AOP	*E. coli*	Yes	(Crane et al., 1997)
3MM5	*Archaeoglobus fulgidus*	Yes	(Parey et al., 2010)
2V4J	*Desulfovibrio vulgaris*	Yes	(Oliveira et al., 2008)
4RKM	*Wolinella succinogenes*	No	(Hermann et al., 2015)
3VLX	*Nicotiana tabacum*	No	(Nakanoet al., 2012)

Outside of the sulfite reductases, this family of proteins also contains the assimilatory Nitrite reductases. These catalyse the reduction of nitrite to ammonia (Crane et al., 1997). This part of the overall family seems to have diversified from what was originally a SUR. This historical protein seems to have gained a higher affinity for nitrite over sulfite, while retaining an affinity for sulfite. Observations from the tobacco plant seem to suggest that this is a reversible mutation, with the tobacco nitrite reductase being easily mutated into a sulfite reductase. This also opens up the possibility that the sulfite reductases may have evolved in some cases from nitrite reductases. Despite being a plausible explanation, it is unlikely with more organisms possessing this style of sulfite reductase than nitrite reductase (Nakano et al., 2012).

The nitrite reductases occur in plants, fungi and bacteria and overall have a very similar structure to the sulfite reductases containing the siroheme, the Fe-S centre as well as all the other aspects of the active site (Kleinhofs et al., 1989). While the assimilatory version of these enzymes are rather like the protein family of interest here, the bacterial versions of the dissimilatory nitrite reductases are a diverse group of enzyme. Some of these dissimilatory enzymes are even responsible for the regeneration of NAD+ during anaerobic growth. Despite facilitating a similar reaction, none show homology with the sulfite reductases (Cole, 1989). This means that they have evolved independently to enable the same process.

In the case of these dissimilatory nitrite reductases and in the sulfite reductases, the family contains both the Siroheme and the Fe-S centre at their active site. These will now be examined individually.

1.2.1 The role of siroheme

The sulfite reductase family of proteins possess rather an unusual ligand in the form of the siroheme. This component functions through the acceptance of an electron from either a FAD species, in the case of the two subunit proteins, facilitated by the α-subunit (Vega & Garrett, 1975) or, in the case of the single subunit proteins from nicotinamide adenine dinucleotide phosphate (NADPH) (in chemotrophic organisms) or from ferredoxin (in phototrophic organisms). These electrons, following this, are transferred to the substrate allowing for the dehydration reaction to occur (Brychkova et al., 2012; Hell et al., 2008). The structure of this siroheme group, seen in fig 1.7,closely resembles a heme group, with both being synthesized from an uroporphyrinogen III, a heme-like group and a precursor of the vitamin B_{12} (Hansen

et al., 1997). The relationship between the siroheme and the uroporphyrinogen was found to be important for the sulfite reductases. For *Thermus thermophilus*, the protein was found surrounded in the genome with proteins responsible for the modification of uroporphyrinogen III. It is not the shape of this compound or the compound itself that makes the siroheme unusual as a cofactor in this system. It is how the siroheme binds to the protein that makes it unusual, with one of the 4 cysteines involved in the Fe-S cluster also acting as an axial ligand for the siroheme, rather than a number of Cys interacting with the Fe atom, contrary to what is normally seen in other Fe containing systems (Crane et al., 1995; Murphy et al., 1974; Nakayama et al., 2000).

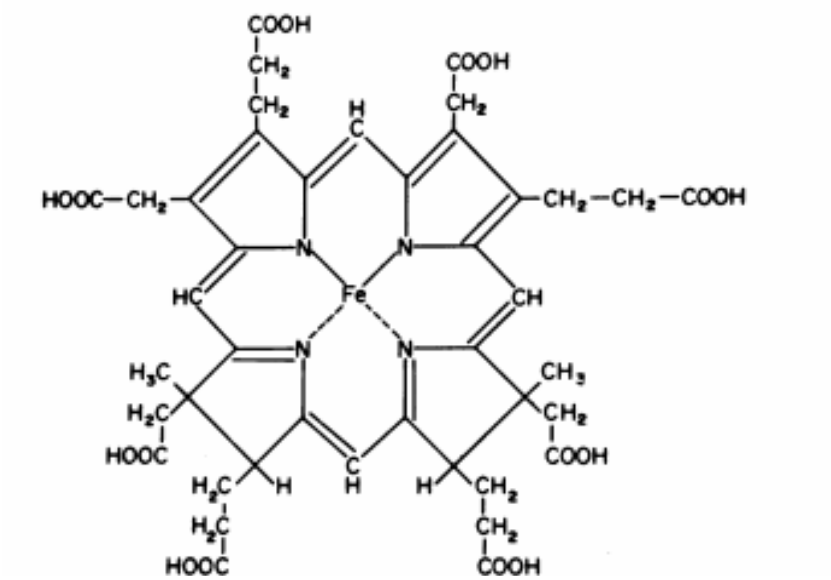

Figure 1.7: *Siroheme structure*

Chemical structure of Siroheme found in sulfite reductase active site, showing all major groups and central N-bound iron residue (Murphy et al., 1974)

One of the main functions of this siroheme is in the process of substrate capture. It alongside arginine and lysine residues are responsible for the capture of both nitrile and sulfite, in their protonated states (Smith & Stroupe, 2012). When this is compared with the capture mechanism of nitrite (fig 1.8), it becomes clear why there is a cross selectivity of anions in the cases of nitrite and sulfite reductases. The only difference between the capture of nitrite and the capture of sulfite, in the *E. coli* system, is the 3-D conformation of the side chain of the Arg 153, which is involved in proton donation (Nakayama et al., 2000). Outside of these, the Fe-S centre is also present and aids in the reduction mechanism through the transfer of electrons (Smith & Stroupe, 2012).

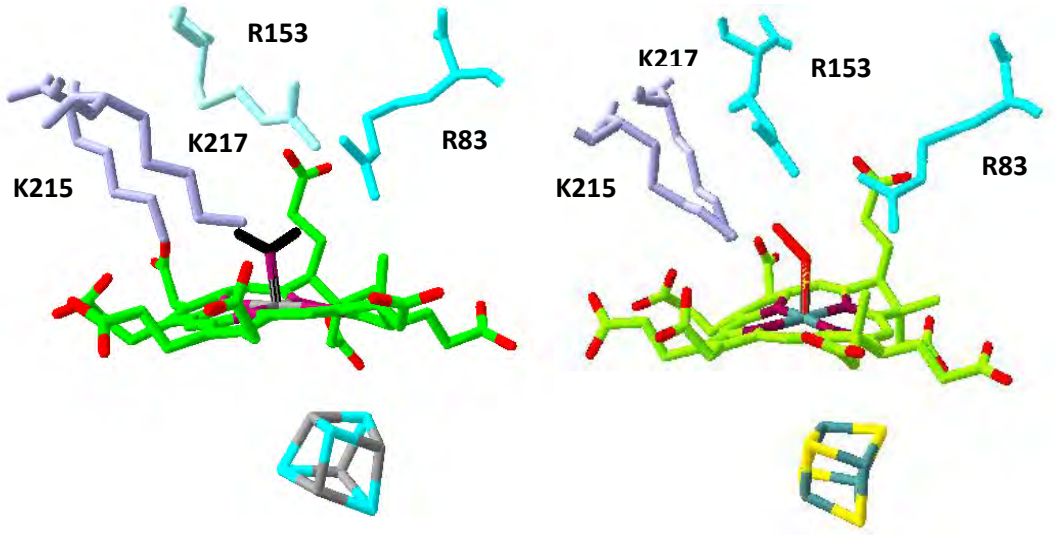

Figure 1.8: ***Active site structures of the β-subunit of E. coli SUR***
Active site of the sulfite reductase of E. coli in complex with nitrite (PDB: 3GEO) (left) and sulfite (PDB: 7GEP) (right) as depicted from Nakayama et al. in their paper on sulfite reductases (Nakayama et al., 2000)

1.2.2 The Fe-S cluster

The final piece of the puzzle in the case of the sulfite reductases is the iron sulfur (Fe-S) cluster. This facilitates the overall reduction of the substrate in all these systems, similar to the siroheme (McRee et al., 1986). This complex, present on the β-subunit of the *E. coli* sulfite reductase and in the single subunit of the maize sulfite reductase, occurs as a Fe_4S_4 cluster. In all cases it has been found occurring at the 3 domain interface at the surface of the protein (Crane & Getzoff, 1996). This complex, responsible in aiding in the 6-electron reduction of SO_3^{2-} to S^{2-}, is hinged on a highly conserved 4 Cys grouping, one of whom also interacts with the siroheme (Murphy et al., 1974). The high level of conservation shows the importance of this cluster with it acting as an intermediator for electron transfer. These 4 Cys residues have been found occurring in all known examples of sulfite reductases (it is important to note that one of these also acts as the ligand to the siroheme, and thus the protein probably could not act as a sulfite reductase as much due to the lack of a siroheme regardless of the lack of this cluster) (Nakayama et al., 2000). This cluster is believed to be the acceptor for the electrons in all cases whether from the α-subunit, NADPH or from ferrodoxin. As of yet, it has not yet been established how this process occurs, and while it is now known how the electrons interact once they reach the centre, there is still work needed in this area (Murphy et al., 1974; Nakayama et al., 2000; Smith & Stroupe, 2012).

1.3 *Thermus thermophilus*

The sulfite reductase in question does not come from maize or *E. coli* and does come from a thermophilic Gram-negative, non-motile, nonsporulating eubacteria called *Thermus thermophilus* HB8 (ATCC 27634). This has been characterised though the immotile, rod like morphology (Degryse et al., 1978). This oxidase positive thermophile, originally classified as *Flavobacterium thermophilum*, was initially isolated by Tairo Oshima and Kazutomo Imahori in a thermal vent in Izu, Japan, where it was identified as a facultative anaerobe (Oshima & Imahori, 1974). Since then it has been regularly used as a model organism for a number of proteins such as the cytochrome *c* oxidases (ba_3 and caa_3), due to its proteins' high level of thermos-stability (Noor & Soulimane, 2012).

It has been reported that *Thermus thermophilus* has an optimal growth range of between 65 and 72 ^{o}C, a maximum growth temperature of 85 ^{o}C and a minimum growth temperature of 47 ^{o}C, showing the large range of temperature over which proteins from this organism are expected to operate (Oshima & Imahori, 1974). Outside of the strain initially purified, another strain of scientific interest has been since identified, having been given the notation HB27. Unlike HB8, this strain is strictly aerobic, due to the fact that it lacks the highly unstable plasmid containing the nitrate respiration genes which allows HB8 survive in anaerobic conditions (Noor & Soulimane, 2012).

In terms of importance to the study, the fact that this organism has previously been worked with in the lab, and its sulfite reductase has been found within a cluster with the enzymes needed to produce the siroheme and sulfur processing enzymes, has led to us investigating it. Additionally, this protein is still significantly different that the β-subunit of the sulfite reductase of *E. coli* to make this protein structurally interesting.

1.4 Crystallisation

One of the most important fields in biochemistry and the major focus of this whole project is structural biochemistry. This rapidly growing and expanding field is constantly changing with new developments and new discoveries moving the field towards the stage where it can be a validation tool for computational prediction of crystal structure (if the current issues in this field are ever rectified). This field has improved significantly in the last number of years. Diffraction data collection has gone from taking a time period of a year to now, being possible to do over the space of a few weeks if not days. This progress has been further enhanced with the development of a range of new algorithms which have made processing data a lot quicker and also massively by the development of computers (McPherson, 1990). The overall process can be used to give a large amount of details about the inner workings of the protein, showing how the protein interacts with its substrate as well as inhibitors and whether it has different configurations in the presence of different chemical and biological species. This allows any individual to figure out what co-factors are present, which amino acids they interact with different, and how the amino acids side chain conformations change in 3-D space when in the presence of each of these different chemical and biological species.

The process of crystallisation itself is not unique to proteins. It can occur with nearly all chemical species when the right conditions (concentration, pH, ect) are present. Needless to say, the mechanism of crystallisation in all of these systems will proceed in a similar way with 2 very discrete steps; nucleation and crystal growth. The first of these steps, nucleation, involves the phase transition of the molecule of interest. In the case of a protein or another chemical species, they transition from the state of disorder in which it may or may not naturally occur to a state where long ranged order is present (McPherson, 2004).

In proteins for nucleation to occur, the solution containing the protein of interest must first enter a state of supersaturation. This allows the system enter the non-equilibrium conditions needed for nucleation to occur. Unlike a large number of chemical species, where crystallisation will occur spontaneously, normally this process in protein systems does require a helping hand to achieve its aim. The key aspect of these systems is the need to bring the protein slowly out of solution at a rate slow enough that the protein can form crystalline structures without it forming a precipitate. This process is normally achieved using one of a number of different precipitating agents such as salts (for instance ammonium sulfate) or high molecular weight hydroscopic molecules like polyethylene glycols (PEGs). These will slowly

remove water from the system, in order to increase the concentration of the protein in solution. This results in the system reaching the state of supersaturation where the protein molecules begin to aggregate and due to the controlled rate, the nuclei (the minimum quantity of a new phase capable to survive independently) of the crystals have enough time to start to form. Following this is the stage of crystal growth where the crystals finally increase in size to a scale where it is possible to use them to acquire a diffraction pattern.

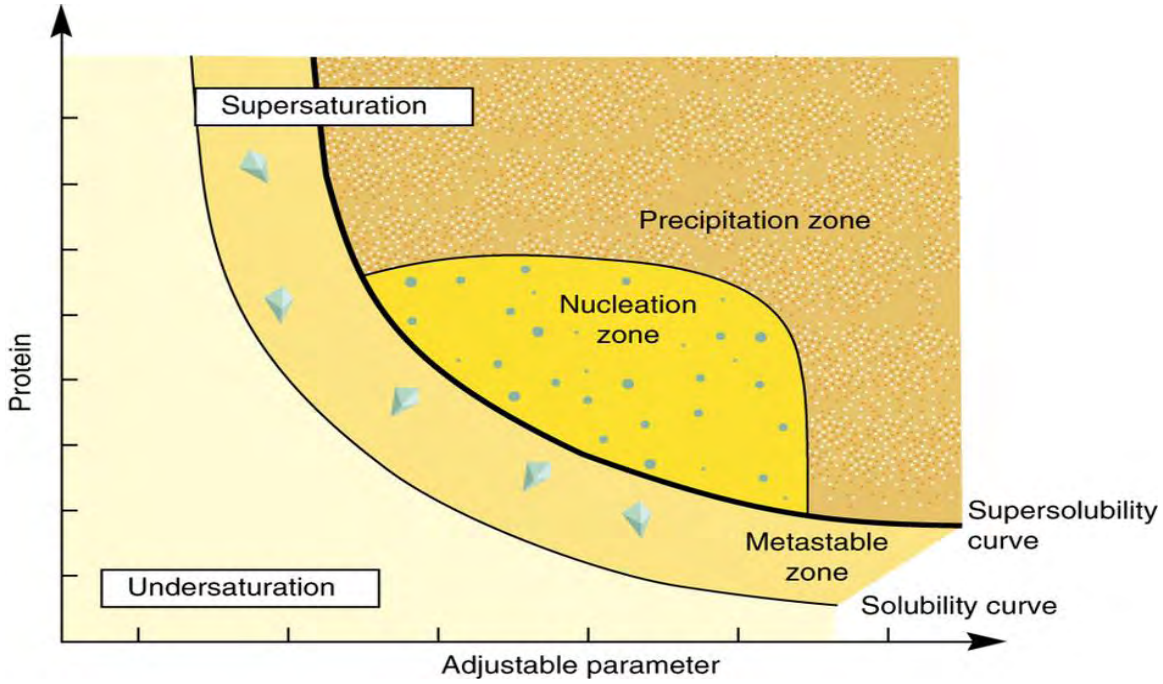

Figure 1.9: ***Protein crystallisation phase diagram***
Protein crystal phase diagram with the adjustable parameters covering everything from temperature and pH to the presence of a precipitant and the additive concentration on the x-axis and protein concentration on the y-axis (Khurshid et al., 2014)

The whole process of crystallisation can be most efficiently described using a phase diagram. One of whom is shown in fig 1.9. This process is defined by the concentration of the species to be crystalized as a factor of its total solubility in the solution. This concentration tends to fall into one of 4 key zones: an under-saturation zone, a metastable zone, a nucleation zone and finally a precipitation zone.

The first of these zones and the one most likely to occur in the case of most protein crystallisation experiments is the precipitation zone. What occurs in this zone is one of the most simple to understand as this is where due to the kinetics, the protein will precipitate so

rapidly, that the protein does not have the chance to reorder itself and cannot achieve long range order. Therefore, all that is observed in this concentration zone of the system is the protein achieving a level of chaos where no nucleation can occur and instead precipitation can and does occur.

The next most likely zone to be observed is the polar opposite of the previous example and is referred to as the under-saturation zone. Here there is not enough protein present for the protein to come out of solution and consequently nucleation cannot occur, or if it does occur, the time period it exists for is minimal before the protein re-enters solution.

This then allows us to move onto the final two zones; the most important zones in any crystallisation experiment: the nucleation zone and the metastable zone. The most logical of these to commence with is the nucleation zone. This is where the origin of all crystal in a sample occurs, with nucleation occurring spontaneously and after that nothing else. This allows the system to move into the most important zone, the metastable zone. The metastable zone is where crystal growth occurs. At this stage, the process of nucleation has ceased to occur. So if the objective is achieve large crystals, it is necessary to alter the ratio of time the solution spends in the nucleation zone versus the metastable zone, with large periods in the nucleation zone resulting in a large number of small crystals, while a long period in the metastable zone resulting in a small amount of large crystals (Chayen, 2004).

The process of human driven crystallisation has been present for thousands of years, yet the process of crystallizing proteins is a relatively new technology. Records suggest that this first occurred in the 19th century and originally acted as a method of protein purification. Following this came a stage where it was utilized as a method to show that a protein was pure while continuing as a method to purify a protein of interest from a mixture of other proteins (McPherson, 2004). This all changed with the discovery of X-ray diffraction. This allowed for the determination of a crystal structure of any chemical species by simply exposing crystals of said species to high power X-rays. This simple process for chemical species, which tend to have reasonably simple structures, proved to be a simple and rapid analytical technique to identify APIs, but in the case of proteins, the process proved anything but rapid and simple. The larger structure in proteins means that a lot more data is needed to determine a structure. Accordingly when working with proteins, the most critical point is getting a single crystal large and stable enough that it can be diffracted to a suitable resolution. There was an additional issue in these cases as a result of their size and complexity which makes the

molecules rather fragile and heat labile. Due to this, being struck with high energy beams tended to damage the protein, prior to the point good quality data. As a result of all these issues, a number of methods have been designed with the sole aim of achieving good quality protein crystals. This has at this stage become an art form (Chayen, 2004). While there are a number of methods depending on whether the protein occurs in the membrane or not and other parameters like this, the two most common methods for the setup of protein crystals are the sitting drop and hanging drop methods, with both based on vapour diffusion.

The first, and the simpler and more widely use of these methods, is the sitting drop method. This, working on the principal of vapour diffusion, generates crystals by using a gradient driven diffusion to bring the solution into a state of supersaturation at a controlled pace. This overall process works very simply on the fact that the solution in the reservoir of the crystal plate has been mixed in equal volumes with the protein sample, on the pedestal. Due to the fact these 2 solutions have been mixed, there is now a lower concentration of the reagent in the drop (what we call the mixture of solution and sample on the pedestal) when compared to the reservoir and from here the system attempts to reach equilibrium. It does this by drawing water from the drop into the reservoir. Due to this water evaporation, the concentration of protein in the drop increases which leads to nucleation at the metastable zone. Equilibration is finally reached when the reagent concentration in the drop is approximately the same as that in the reservoir. At this stage the whole process ends and if it has been successful, there should be protein crystals (Hampton 2010).

This leads nicely into the next method, the hanging drop method of vapour diffusion. Recently, this has become quite a popular method in protein crystallisation experiments. This procedure works on a similar principal to the sitting drop method, driving the process through a concentration gradient, resulting in water leaving the drop, increasing the overall concentration of the protein to be crystalized. The difference this time is with the positioning of the drop. Here it is held in place through the drop's interaction with a layer of oil on the cover slip. Overall all that is off interest is that the difference in salt concentration between the reservoir and the drop allows the water directly move through the drop in order to achieve the level of supersaturation needed (Hampton 2016).

This process has also been simplified with the fact over the course of previous experiments in this field; it has been found that certain conditions are more likely to produce protein crystals than others. In order to make things easier, these have now been commercially reproduced as

screens in order to identify what conditions give crystals at a relatively high level of throughput by allowing for the testing of a large amount of different conditions, of different additives, of different salts and of different precipitation agents all at once (Cudney et al., 1994; Hörer et al., 2013; Resch et al., 1995; Trakhanov & Quiocho, 1995). As a whole all of this just means, that today it is easier than ever to get high quality crystals, yet there is still no fool-proof method yet available to achieve this.

1.5 Aims of this study

It can now be seen that the protein family of interest, the sulfite reductases, while well studied in recent times, are still lacking in information in a large number of areas, with only a small number of crystal structures published as of yet. This study aimed to achieve a greater level of understanding of the structure-function relationship of family as a whole, through the acquisition of the3-D structure of the *Thermus thermophilus* protein. Therefore, various issues such as protein yield, homogeneity and determination of the oligomeric state have to be addressed. Hence the first focus of this study was to produce and purify this novel member of the sulfite reductase family in a way that it is of high purity, homogenous and at workable high yields. Once obtained, the concentrated sample will be crystallized with the ultimate objective of solving the crystal structure.

2.0 Material and methods

2.1 Materials

2.1.1 General equipment

Table 2.1: *Major Equipment used over the course of this project*

Equipment	Manufacturer	Place of origin
ÄKTAexplorer HPLC System	GE Healthcare	Buckinghamshire, UK
ÄKTAprime Plus HPLC System	GE Healthcare	Buckinghamshire, UK
Evolution 201 UV-Visible Spectrophotometer	Thermo Scientific	Dublin, Ireland
Hettich Rotanta 460R centrifuge	DJB Labcare Ltd.	Newport Pagnell, UK
Innova 40R benchtop incubator shaker	Eppendorf	Cambridge, UK
Minispin Microcentrifuge	Eppendorf	Cambridge, UK
ND-1000 Spectrophotometer	Shimadzu	Milton Keys, UK
Ni-Sepharose 6 Fast Flow Resin	GE Healthcare	Uppsala, Sweden
PAGE Apparatus	Medical Supply Company	Dublin, Ireland
Soniprep 150 ultrasonic disintegrator	Wolflabs MSE	York, UK
Sorvall RC 6+ Centrifuge	Thermo Scientific	Dublin, Ireland
Superdex 16/600	GE Healthcare	Buckinghamshire, UK
XK 16	GE Healthcare	Buckinghamshire, UK

2.1.2 Reagents used in this study

Chemicals unless otherwise stated were purchased from Sigma Aldrich (Dublin Ireland)

2.1.3 Crystal screens

During the course of this project, 2 major commercial screens were used, the Index Screen and the Additive Screen, both commercially available from Hampton

2.2 Methods

2.2.1 Protein expression trials by auto-induction

Protein expression was carried out using the pET 22b vector (containing the gene of interest), in *E. coli* BL21*. All protein production in this study was carried out via the Studier method of auto-induction, with induction lasting 48 h and following the conditions listed below.

Auto-induction was carried out via the method set out for ZYP-5052 auto-induction media in Studier's 2005 paper (Studier, 2005). This process was started via setting up an O/N culture at 180 RPM and 37 ^{o}C with a 1% inoculation from a glycerol stock. This was used to inoculate ZY broth (10 g/L tryptone, 5 g/L yeast extract), supplemented with 50 mL/L 20 P (141.96 g/L Na_2HPO_4, 136.09 g/L $KH_2PO_4$and 66.07 g/L $(NH_4)_2SO_4$), 20 mL/L 5052 (250 g/L glycerol, 25g/L glucose and 100 g/L α-lactose) and 1 mL/L 1 M $MgSO_4$at a 0.5 % v/v inoculation.

Ampicillin was added to the culture, to a final concentration of 100 µg/mL, to ensure selectivity for the cells of interest. Cultures were grown for 3 h at 150 RPM and 37 ^{o}C to allow for growth. The culture is then induced at 150 RPM, 25 ^{o}C for 45 h. This culture was then spun down for 20 min at 8,000g to isolate the cell pellet.

2.2.2 SDS-PAGE

Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) analysis was performed throughout this study to test both for the presence and purity of proteins using a 16% resolving gel prepared using the standard method (Laemmli, 1970). The resolving gel was prepared as follows:

Table 2.2: *SDS-PAGE 16% resolving gel*

Reagent	Concentration	Volume
Acrylamide	40%	3 mL
Tris-HCl pH 8.8	1.5 M	1.95 mL
APS	10%	75 µL
SDS	10%	75 µL
TEMED		3 µL
Water		2.4 mL

This gel was poured into the sealed gel plate and allowed to polymerise under a layer of milliQ water to ensure that the gel did not dry out. The stacking gel, 4.5% in all cases, was prepared and poured once the resolving gel had polymerised, and was prepared using the following concentrations:

Table 2.3: *SDS-PAGE 16% stacking gel*

Reagent	Concentration	Volume
Acrylamide	40%	0.32 mL
Tris-HCl pH 6.8	1 M	0.32 mL
APS	10%	25 µL
SDS	10%	25 µL
TEMED		3 µL
Water		1.8 mL

Gels were placed in the rig and the process of electrophoresis was then preformed in 1x SDS running buffer, prepared from a 5x stock (15.14 g/L Trizma base, 72 g/L glycine, 5 g/L SDS), with all samples being prepared using a 5x SDS based loading buffer (0.25 % bromophenol blue, 0.5 M dithiothreitol (DTT), 50 % glycerol& 10 % sodium dodecyl sulfate (SDS) dissolved in 0.25 M Tris-HCl pH 6.8). Whole cell samples were mixed in a 1:1 ratio with the 5 X loading buffer and were heated to 95 ^{o}C for 5 min in a heating block to lyse the cell and to lower the overall viscosity allowing them to be loaded into the gel. Protein samples were not heated were incubated at room temperature for 5 min with the loading buffer at a 1X concentration prior to being loaded into the gel. Gels were run at 25 mA per gel for 70 min, and following this is stained with Instant Blue (Expendeon, Cambridgeshire, UK) for 45 min and then de-stained in milliQ water overnight

2.2.3 Western blot

All western blots were performed using the semi-dry method. In this method, the SDS gel, run prior to this, is sandwiched between two layers of blotting paper, with the gel placed in the centre on a layer of blotting membrane made from nitrocellulose. Each layer was firstly soaked in milliQ H_2O for 15 min, soaked in transfer buffer for 5 min and then played on apparatus as a layer of 3 blotting sheets, followed by the membrane, then the gel and then the 3 further layers of blotting paper. The protein was transferred from the gel to the membrane at a rate of 80 mA for 70 min.

Following the transfer of the protein from the gel to the membrane, the membrane was then washed for 3 x 5 min runs using 15 mL TBS-T each time. Following this, the membrane was blocked using 5 % skimmed milk in TBS-T for 1 h with gentle shaking at RT. This was followed by a further three wash steps with TBS-T to wash off any milk residue and then this membrane was incubated with antibody solution for 1 h before a final wash step with TBS-T. The membrane was then allowed to react with TMB, giving a red colour if there was a poly-his to react with.

2.2.4 Lysis of cells for SUR purification

The cultures were centrifuged at 8,000 g for 20 min at 4 ^{o}C, with the supernatant being discarded and the pellet re-suspended in 5 mL of a 10 mM Tris-HCl pH 8.8, 30 mM NaCl buffer for every 1 g of cells. Once homogenisation was achieved 300 mg of lysozyme was then added along with 3 mM EDTA and this solution was kept stirring at RT for 1 h before being placed in the -80 ^{o}C freezer for a minimum of 8 h, for the purpose of preforming a freeze-thaw. This frozen solution was thawed at 30 ^{o}C, submerged in a water bath. 3 mM of MgCl was added alongside DNase and this was incubated at 4 ^{o}C for 1 hour before centrifugation at 22,000 g for 1 h to remove whole cell and membrane fractions. The concentration of NaCl in the supernatant was brought up to 300 mM and 10 mM imidazole was added.

Following further investigation into this process a number of changes were made to the overall process with the changes discussed later in results and discussions section

2.2.5 Immobilized metal affinity chromatography

A XK 16 column was packed with Ni Sepharose 6 Fast Flow resin and the resin was regenerated using a 25 mL of a 0.5 M NaCl, 20 mM Tris-HCl pH 7.4, 50 mM EDTA solution to strip the column as per the manufacturer's protocol. The column was then washed with 50 ml of milliQ H_2O, recharged with 20 mL of 0.1 M $NiSO_4$ and cleaned extensively with milliQ water. Two column volumes (CV, 10 mL) of binding buffer with 1 M imidazole followed by two CV of the binding buffer itself were ran through the column to create a state of equilibrium at a rate of 1.5 mL/min, proceeded by the sample. Elution was then preformed stepwise using the same base components as the binding buffer but changing the concentration of Imidazole.

Table 2.4: *Immobilised metal affinity chromatography conditions for the sulfite reductase of Thermus thermophilus*

IMAC Run	Binding Buffer Components	Elution Concentrations
SOR	0.5 M NaCl 50 mM Tris-HCl pH 7.6 10 mM Imidazole	**Washes** (50 mM, 70 mM), **Elution** (100 mM, 250 mM, 500 mM)

2.2.6 Size exclusion chromatography

Size exclusion chromatography was performed on a Superdex 16/600 column at 1 mL/min in 50 mM Tris-HCl pH 7.6 buffer. Further alterations are discussed in the results part of the thesis.

2.2.8 Bicinchoninic acid assay

The bicinchoinic acid (BCA) assay was performed in all cases in accordance to the manufacturer's instructions. Standards of 0.025 mg/mL, 0.125 mg/mL, 0.25 mg/mL, 0.5 mg/mL, 0.75 mg/mL, 1 mg/mL and 1.5 mg/mL were prepared via serial dilution. 30 µL of each was pipetted into a 96 well plate. 30 µL of each protein sample was then added into the plate along with 30 µL of a 1:10, a 1:20 and a 1:100 dilution of each sample.

The BCA working reagent was freshly prepared by mixing solution A and solution B in a ratio of 50:1. 210 µL of this was added to each of the wells. A blank solution was prepared by adding 210 µL of this to 30 µL of the buffer used to carry out the dilutions. The plate was sealed and incubated at 37 ^{o}C for 30 min.

Once the incubation had been concluded, the standards were tested at 562 nm to establish a standard curve. The reading at 562 nm for each of the samples and their dilutions were then tested against this in order to calculate the concentrations.

2.2.7 Protein crystallisation

Initial screen

Initial Screens were prepared using either 100 mM HEPES pH 7.0, 2% PEG and 15% Tacsimate or 0.2 M Ammonium citrate pH 7.0 and 20% PEG as the reservoir solution, with 40 mg/mL of SUR in 50 mM Tris-HCl pH 7.6. 250 µL was placed in the well and a mixture of 1 µL protein and 1 µL reservoir solution were mixed on the pedestal

Index screen

Index Screen was carried out as per Hampton manual

Additive screen

The additive screen was prepared in all cases using a 3.5M NaCl, 100 mM Tris-HCl pH 7.5 base solution and then following what was set out in the Hampton manual, scaling down to achieve a 1 µL drop

Acquiring images of the crystal

The images of the crystals were acquired by using a Canon Ixus 170 with a microscope due to the lack of an imaginer.

2.2.8 Bioinformatics

Homology and domain identification

In all cases homology was searched for using the Blast suite of programmes. Domains from the structure were firstly identified with Blast P. Verification and further information was acquired using Inter Pro Scan.

Phylogenetic modelling and alignment

Alignment and phylogenetic was performed using the Muscle sequence alignment tool from EMBL-EBI using Fasta sequences acquired from Blast P. The results were visualised with Treeview.

3-D structure rendering

All visualations of 3-D structures were generated using Swiss Pdb Viewer Deep View v4.1

3.0 Results and discussion

3.1 Bioinformatical analysis of the target

3.1.1 General bioinformatics

Before commencement of the structural investigation into the protein or even the process of optimizing its purification, it was preferable that a greater understanding of its genomic context, its probable function and its possible similarities to other proteins were acquired. In the case of the protein of interest here, TTHA0672 (Kegg database), this process started with the latter of these steps, with the protein compared to the NCBI database using the Blast search algorithm. This relatively basic programme utilizes the BLOSUM (Blocks Substitution Matrix) matrix. The algorithm calculates how close two proteins are to each other in an evolutionary sense. This is done through the statistical analysis of the possibility of a single amino acid being substituted for another in a biological system. When, this was utilized on the protein sequence of TTHA0672, the top result came from the organism of interest *Thermus thermophilus* with a 100 % identity, consistent with what was expected.

There were a number of issues with these initial results with the proteins with the highest levels of homogeneity labelled as a ferredoxin nitrate reductases and not a sulfite reductases (SUR). On further examination, it was seen that the ferredoxin-nitrite reductases can also be classified as homologs of the sulfite reductase, β-subunit. While the reason for this isn't clear at first, when this was investigated, it is revealed that sulfite reductases and ferredoxin-nitrite reductase are related with both possessing the ability to reduce sulfite and nitrite and in reality the two types only differing due to the affinity they have for the 2 substrates (Crane & Getzoff, 1996; Murphy et al., 1974). Since the programme only has the ability to compare proteins with proteins previously deposited onto the system, it is possible to understand why this mislabelling occurs. This possibility also implies that the specification of the enzyme for the selectivity of nitrite or for sulfite may be a recent alteration.

Outside of this, the results were as expected with the closest matches being from *Thermus parvatensis* (95% identity), *Thermus filiformis* (92% identity), *Thermus aquaticus* (92% identity) and *Thermus scotoductus* (89% identity), all of whom are coming from *Thermus* species. After this, the first proteins observed being the nitrite reductase of *Roseivirga spongicola* (identity 44%), and the sulfite reductases of *Nitrospira defluvii* (identity 42%) and *Synechococcus* sp. PCC 7002 (identity 42%), all of whom are water based organisms. This is

expected with thermophilic bacteria being water based. As for the model protein, and the first sulfite reductase identified and crystalized, the sulfite reductase of *E. coli,* this does not compare favourably with the protein in question here. Firstly no protein was found in *Thermus thermophilus* to resemble the α-subunit. The *E.coli* β-subunit, then, only had a 24% identity when compared with the *T. thermophilus's* SUR. This implies that the protein of interest here does not contain 2 subunits but there is just a single subunit. This is not as odd as it sounds, with a number of sulfite reductases found to contain only a single peptide.

From the data, it is still not clear if the protein was a sulfite or nitrite reductase, meaning it was necessary to compare the domains to the two species. The data accessible to Blast is not complete relying on data submitted from other structures. Despite this, it is possible to see from fig 3.1, that there are a number of identifiable domains

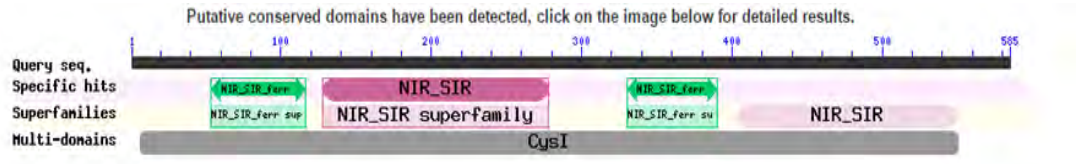

Figure 3.1: ***Blast domain results for SUR from Thermus thermophilus***
Blast P data or the protein of interest, showing the presence of a number of domains

The firstdomainis the nitrite/sulfite reductase ferrodoxin-like domain. This domain,identified in the SiR-HP subunit of the two subunit SiRs,is present in all single subunit sulfite reductases, and it is noted that it occurs with a level of symmetry with two copies of the domain within the protein. Allignment of this along with other ferredoxin like domains shows a large level of conservation throughout a number of key residues.

Following this is the 4Fe-4S domain. This has been shown in the *E. coli* SiR to act as a ligand to the siroheme.In this domainthere is an iron bound to cysteine at position 433, 439, 478 and 482 (Crane et al., 1997), and it is from one of these that the siroheme molecule is also associated. This is found in the SiR-HP subunit, all four of these Cys residues are highlyconserved in these structures (Crane et al., 1995).

To finally identify, whether this is probably a sulfite or a nitite reductase, it is necessary to study the proteins encoded by the genes surrounding the protein of interest in this case. In terms of genomic context, looking at the TTHA0672 of *Thermus thermophilus,* with the help

of the NCBI database, a number of sulfur based proteins were found to surround the protein of interest. This cluster, containing ATP Sulfurylase (a upstream protein in the metabolism of sulfur), phosphoadenosine phosphosulfate reductase(this operates in the opposite direction, this time building sulfite up into a regulatory molecule (Schomburg & Salzmann, 1990)), and Sirohemesynthase (a uroporphyrin-III methyltransferase, a protein which forms uroporphyrin-III, a precursor of siroheme), contains everything expected to surround a gene encoding for a sulfite reductase.

It is possible to back up this theory by comparing it similar proteins, such as the *Thermus parvatiensis* homologs of the *Thermus thermophilus* protein, an unclassified ferredoxin nitrite reductase. Being a ferredoxin nitrite reductase, this protein is an example of a protein which may be either a sulfite reductase or a nitrite reductase. Due to the fact it is directly downstream from Sulfate adenylyl transferase, which produces APS (which is reduced to sulfite, the substrate of the Sulfite reductases), and that there is a thiazole synthase (which uses S based substrates to form the Vitamin thiazole (B1) (Kriek et al., 2007)), one can safely assume that this is sulfite reductase, solely based on the cluster in which it is found. Outside of the *Thermus* family, there is still a similar picture for the sulfite reductases with *Marinithermus hydrothermalis* for example, occurring in a similar cluster, implying that this protein is probably a sulfite reductase. All of this work put together, backs up the overall hypothesis that this is a sulfite reductase.

This is slightly different when one delves into the 2 subunit version of the protein, focusing on the sulfite reductase of*E. coli*here due to the 3-D structure for both subunits being accessiblewith one of these having homology with our protein of intrest. Examining the genomic context of the β-subuntit one can see that the α-subunit is directly downstream fromit and as a whole it occurs with another sulfur related protein, in this case phosphoadenosine phosphosulfate reductase, which produces sulfite as a by-product. Where this cluster does differ from what was seen previously is the fact that the usual siroheme processing proteins are absent. There are two further flavoproteins in this area but after that there is a lack of a number of the other sulfur processing proteins seen around the one subunit proteins. Despite this, these proteins are in the genome of *E. coli*, just not in the cluster. This may explain the difference from what has previously been seen. These enzymes are all present and probably interact in the same cycles and just occur in other clusters. More out of a point of interest than of importance, it has been found that the two subunits of the sulfite

reductase here are found beside the *E. coli* clustered regularly interspaced short palindromic repeats proteins which in reality act as a defence system for *E. coli*.

3.1.2 Phyogenetics and alignment

It is important to understand the difference between the protein of interest and the nitrite and sulfite reductases of *E. coli* on a phylogenetic basis. Compared with the protein of interest, the sulfite reductase of *Thermus thermophilus*, the proteins in *E. coli* have been shown to be more closely related to one another than the two sulfite reductases are. This is contrary to what was expected. The analysis, carried out using the β-subunit of the *E. coli* sulfite reductase and the sulfite and nitrite reductases of a number of intermediary species, using the programme Muscle, also showed that the *E. coli* nitrite and sulfite reductases were more closely related to one another than they were to any other sulfite or nitrate reductase. This combined with the nitrate reductase of *Roseivirga spongicola* being more closely related to the protein of interest than the sulfite reductase of *M. maris*, illustrates the point that the split between sulfite and nitrite reductases occurred a lot more recently than the split between the sulfite reductases of different organisms. This explains why there is such a cross selectivity of substrates and similar mechanisms and active sites.

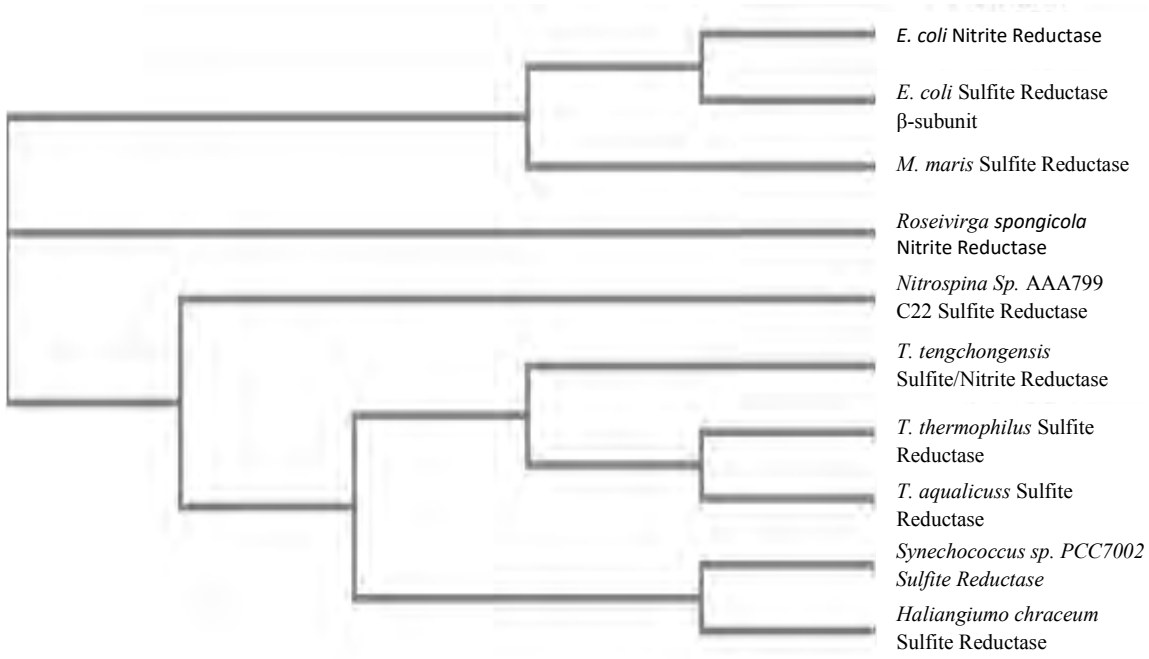

Figure 3.2: ***Phylogenetic tree of selected SUR homologs from a range of different bacteria***

3.2 Purification and achieving homogeneity

Prior to commencement of this study, the gene from the protein of interest had previously been cloned into BL21*. A basic procedure for the expression of the protein through auto induction and purification using immobilised metal affinity chromatography and size exclusion chromatography had also been set out by Dr. Edel Durack. Furthermore, there had been a number of commercial screens preformed on the protein at a concentration of 40 mg/ml giving initial hits with the index screen for the 100 mM HEPES pH 7.0, 2% PEG and 15% Tacsimate and the 0.2 M Ammonium citrate pH 7.0 and 20% PEG conditions, but this was only the case with one production run and was not at this stage reproducible.

3.2.1 General purification

For crystallisation trials, milligram quantities of the protein are needed in a highly pure and homogenous state. The conditions, proteins experience within the cell are quite different to the conditions they see while in buffers, making this more difficult. To achieve this level of purity and homogeneity, a number of chromatography steps were needed, including immobilized metal affinity chromatography (IMAC) and originally several size exclusion chromatography (SEC) steps, quite a common trait in the purification of any proteins. While each chromatography step will increase the purity of the protein, this does not come without disadvantages. For instance, when a large number of chromatography steps are performed, a lower percentage yield is achieved. Additional steps in the procedure increase the total time period needed for the purification to occur and this increases the chance of the protein being denatured or damaged. In conclusion, it is necessary to keep the number of steps as small as possible to ensure the maximum yield and maximum stability but enough steps must be present to ensure a high level of high purity.

In the case of the SUR from *Thermus thermophilus*, initial processing was started with a lysozyme-freeze thaw combination to ensure lysis followed by IMAC and SEC. This is a very common method to lysis cells, being simple, cheap and being a two-step process. This method results in minimal stress to the protein. In the first stage of this method the cell pellet is incubated with lysozyme, a protein found in the tears and the whites of eggs which breaks up the bacterial cell walls by catalysing the hydrolysis of the 1,4β-bonds between the N-acetylmuramic acid (NAM) and N-acetyl-D-glucosamine (NAG) residues, acting as a defence mechanism where it is employed. This is complimented with the freeze thaw method, which uses the mechanical stress of the growth of ice crystals, ripping the cell membrane apart to break up the cells.

Once the solution from this lysis step had been clarified, via centrifugation, the protein solution was further purified using immobilised metal affinity chromatography. IMAC, a form of affinity chromatography, works by utilization the fact that certain ligands have an affinity for certain proteins and can be used to reversibly bind them. In the case of IMAC, the ligand is a metal (normally nickel, but for higher specificity this can be cobalt), and the protein which it binds tends to be one containing a His tag, made up of 6 or more histidine residues. These histidine tags, which have a negative charge, are attracted to the positively charge metal centres on the column and selectively bind. Once the solution has been ran through the column, the proteins which do not bind to the column are washed out, using 30 mM, 50 mM and 100 mM imidazole washes. Imidazole is then added to the column at various concentrations (150 mM, 250 mM, 350 mM and 500 mM) to allow for elution. Imidazole competes with the bound proteins, eluting proteins of lower binding capacity first and those with higher affinities off at higher concentrations. On that account these interaction can then be used as a very efficient way of purifying proteins.

In this case, the clarified solution containing the protein of interest, as well as a large number of other proteins, was the loaded onto an IMAC column. A step wise gradient was employed to elute the SUR of *Thermus thermophilus*. Here by increasing the imidazole concentration in a controlled manner, it was possible to remove the impurities which were more weakly bound to the resin. This resulted in a relatively pure protein sample after only one chromatography step. The process was also greatly helped by the fact that the protein was coloured, which allowed for us to follow it throughout the purification process and allowed us to see where the highest concentrations of the protein were coming.

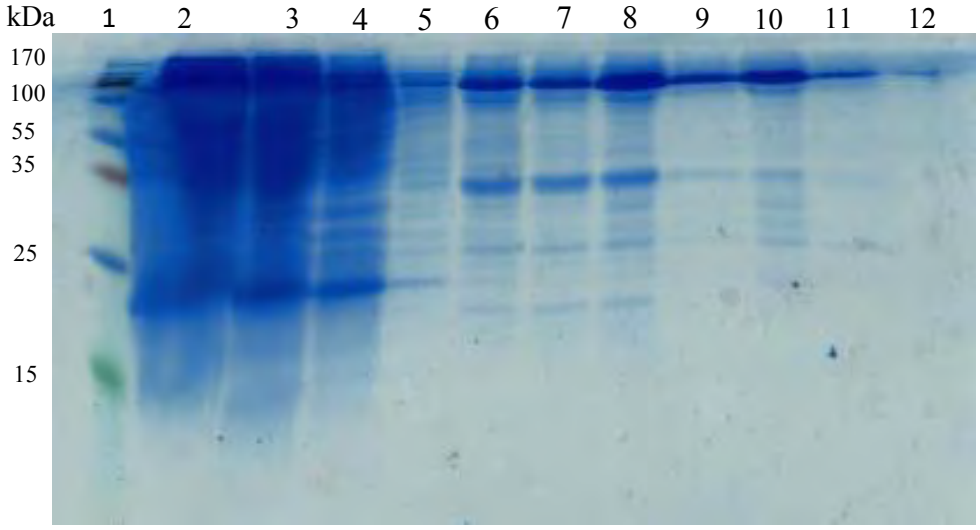

Figure 3.3: *SDS-PAGE analysis of IMAC samples*
SDS gel showing samples acquired with whole cell pellet (2), lysate (3), flow through (4), 50 mM elution (5), 70 mM elution (6), 100 mM elution (7 & 8), 250 mM elution (9, 10 & 11), and 500 mM (12)

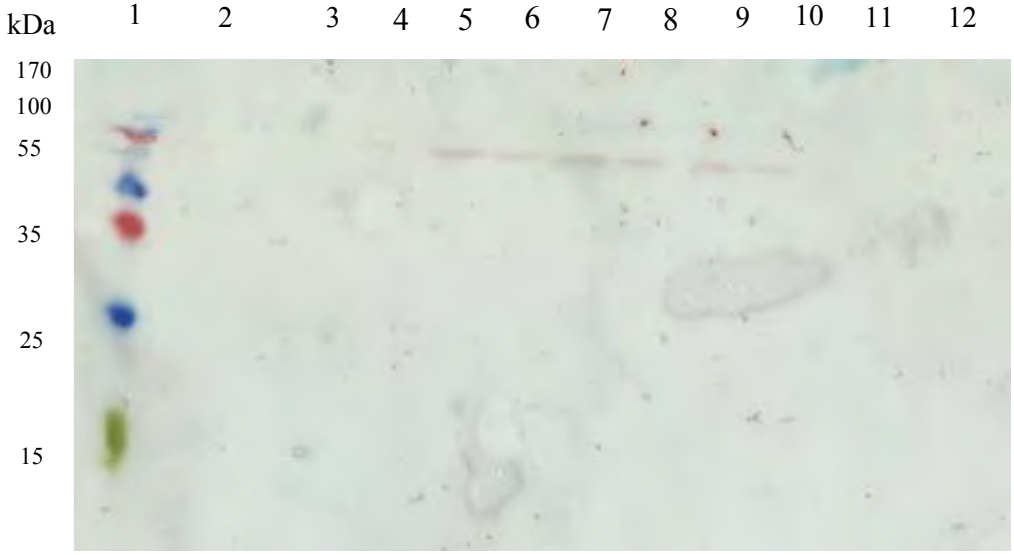

Figure 3.4: Western blot carried out after *SDS-PAGE analysis*
Western blot of the SDS gel shown in fig 3.3 with whole cell pellet (2), lysate (3), flow through (4), 50 mM elution (5), 70 mM elution (6), 100 mM elution (7 & 8), 250 mM elution (9, 10 & 11), 500 mM (12)

The protein fractions, after being tested via SDS-PAGE and through the use of western blotting, were concentrated and subsequently ran on a gel filtration column. This showed three clear peaks. One slightly occurred slightly after the void volume and has been attributed to being the result of aggregate. Two other peaks, following column calibration and data from the crystal structure, have been proposed to be a monomer and dimer. All samples from this process were analysed via SDS-PAGE, to show that 1; they were all the same protein and 2; that all samples were pure. In this case all proteins were seen to have the expected size of 66 kDa, as can be seen in fig 3.6, it does occur over a range of different volumes, implying that the protein is forming a number of different conformations.

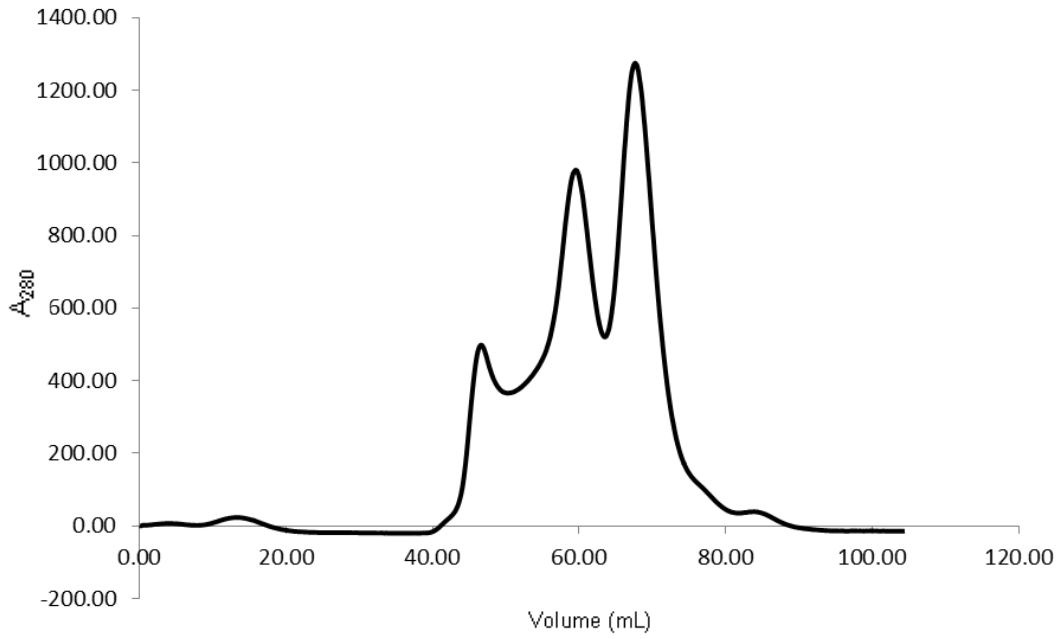

Figure 3.5: ***Initial SEC of the SUR protein after purification***

SEC of SUR protein after IMAC, showing 3 peaks with aggregate peak at circa 42 mL, dimer at circa 60 mL and monomer at 70 mL, all of which was confirmed via column calibration

Following SEC, the amount of protein was quantified using the BCA assay with Bovine Serum Albumin (BSA) as the standard at a number of different concentrations (0.025 mg/ml, 0.125 mg/ml, 0.25 mg/ml, 0.5 mg/ml, 0.75 mg/ml, 1 mg/ml and 1.5 mg/ml), with a number of dilutions made to ensure the protein falls within this bracket. Unlike traditional spectrometry, this method does not rely on the number of aromatic amino acids and does not vary from protein to protein, allowing for the accurate determination of the protein concentration. This was then either aliquoted and frozen using liquid nitrogen or used for crystal trials.

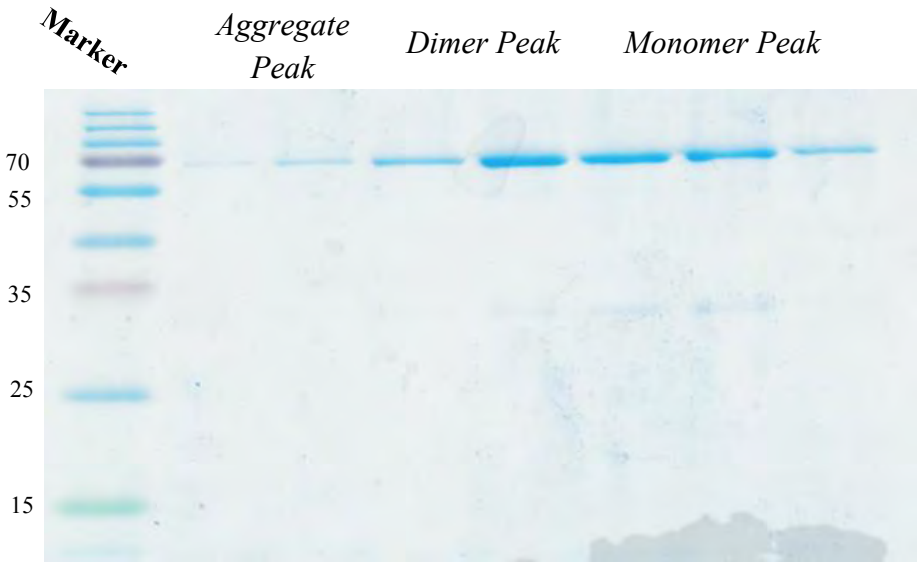

Figure 3.6: ***SDS-PAGE analysis of SEC samples***
SDS gel showing samples acquired from size exclusion chromatography showing 2 samples from the aggregate peak, 2 samples from the dimer peak and three samples from the monomer peak

3.2.2 Modifications to lysis

With three peaks and a relatively low amount of protein being produced at this stage of the study, there was a range of different issues still to be addressed. This resulting variation between production runs, made it difficult to directly compare crystallisation trials. In each case, a number of different production runs needed to cover all the conditions. Attempts were made originally to increase the amount of protein produced per run, with the initial quantity reported as 19 mg per 10 L of culture. Through use of the BCA assay, it had been established that this 19 mg was divided out in a ratio of 4.2 mg : 6.6 mg : 8.2 mg. This meant that only 8.2 mg of the total amount of protein which had been isolated was in the monomer form. This was later identified as the only population which would crystalize. This meant that the usable protein ratio was 0.8 mg/L, with the culture taking 3 days to grow and a further 3 to process the protein to give less than 1 mg/L of usable protein. Hence something substantial had to be done.

Investigating the finer details of this procedure, the first point of note was the amount of cells which were not being lyzed by the lysosome-freeze/thaw combination and therefore going to waste following the clarification of the solution. Work was done to identify a method to lyse

these remaining cells. To commence this, the process of sonication was tested as a method of lysis. This method, where sound waves are used at different frequencies to disrupt the cell wall and hence further lyse the cells, has been employed in a large number of cases to achieve efficient cell lysis. While there is some batch to batch variation in regards to this the amount of protein acquired this way form a 5 L of culture, this additional step moved the yield of the sulfite reductase to 26 mg. This resulted in a jump of over 100%, due to a much larger cell lysis with the protein occurring in a 4.7 : 9.3 : 12 ratio, illustrating what a simpler and directly comparable ratio. This was also accompanied by a shift on the amount of crystallisable protein from 0.8 mg/L to 2.4 mg/L which relates to a 200 % jump. While there is always a risk that sonication can damage the protein being worked with, there are cases in which it doesn't, such as in the case of the protein of interest.

With sonication being a disliked method due to its severity and the risk to the protein due to this harshness, an additional freeze thaw step was also attempted to increase the level of lysis. This is a much gentler step and would not lyse as many cells but the risk of damaging the protein should be significantly less. As an additional step, this resulted in the ratio of the protein changing with 7.88 mg of protein being found in what is believed to be aggregated state of the protein. This was actually unexpected but considering that the solution has been heated twice, this may have been the reason for this. Despite this, 2.5 mg of the protein have been found in the dimer fraction and 4.9 mg being in the crystallisable monomer state. This was a 0.12 mg/L increase at the expense of this higher level of aggregation.

The third method employed revolved around the SUR in question being from a thermophilic bacterium, hence expected to be thermally stable. This could be combined with host strain in this case *E. coli* is a mesophilic bacterium, growing best at 37 ^{o}C. Owing to this, the assumption was made that by heating the solution to 45 ^{o}C; it should have been able to lyse the *E. coli* cells. This should denature the host proteins without affecting the protein of interest, as this should be stable at this temperature. This did not prove to be as efficient as the process of sonication. The reason was believed to be due to the His tag of some of these proteins becoming denatured at the higher temperature. This tag is not a part of the original protein, and therefore might not have possessed the same level of thermos-stability. Regardless, the amount of protein isolated only amounted to 12 mg from 5 L, occurring in the ratio of 4:2:6. This was slightly more efficient than the freeze thaw method, but found to be not as efficient as sonication. This results in a ratio of monomer per L of culture of 1.2 mg/L.

It was now possible to conclude that the best method for lysis in this procedure, taking into account both the amount of time needed to preform and the amount of usable protein following purification, was a lysozyme-freeze thaw step followed by sonication.

Table 3.1: *Summary of lysis optimization*

Method	Mg	mg/L
Original Method	4.1	0.8
With Heating	6	1.2
With Additional Freeze Thaw	4.9	0.92
With Sonication	12	2.4

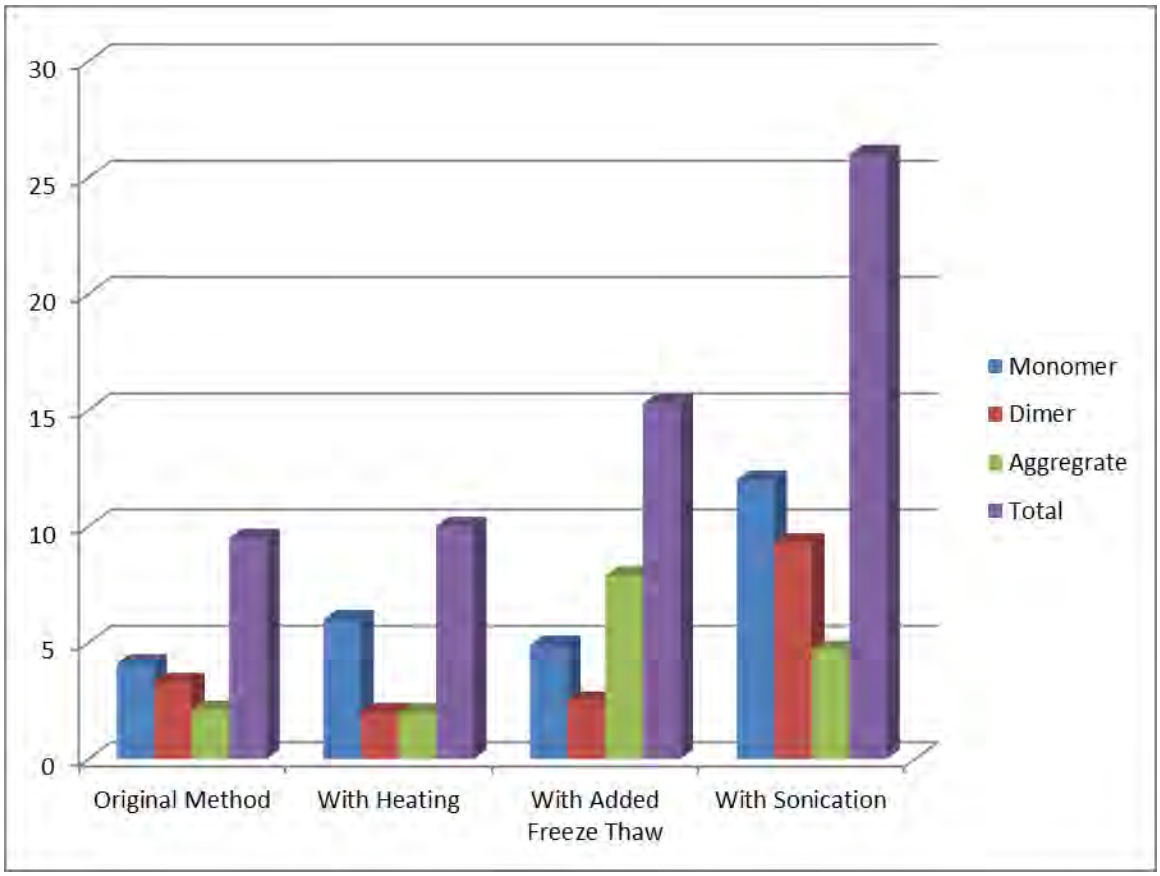

Figure 3.7: ***Graphical representation of the results acquired for the lysis optimization***

Graphical representation of the results acquired for the lysis optimization, with the quantity of monomer per 5 L of culture in blue, of dimer in red, of aggregate in green and the total amount of protein in purple

3.2.3 Modifications to size exclusion protocol

Once lysis had been modified, the aim moved to identifying a process to interconvert the protein populations or isolate the protein as a single population and therefore achieve homogeneity. Initial of attempts to achieve a consistency quantity of protein from the same population was performed via the process of re-chromatographing the peaks. This did result in a large loss of protein which was not present in the required populations.

The low yields directly affected initial work in the area of crystallisation, resulting in it being exceptionally slow due to slightly over half the protein produced being usable. The importance of homogeneity cannot be overstated here, with multiple populations increasing the statistics of disorder retarding crystal growth, with mixed population samples here not producing any crystal hits.

Original work in this study, aimed to resolve this by simply rerunning the separate protein peaks multiple times until homogeneity was achieved. It did achieve this and furthermore achieved crystals of usable quality

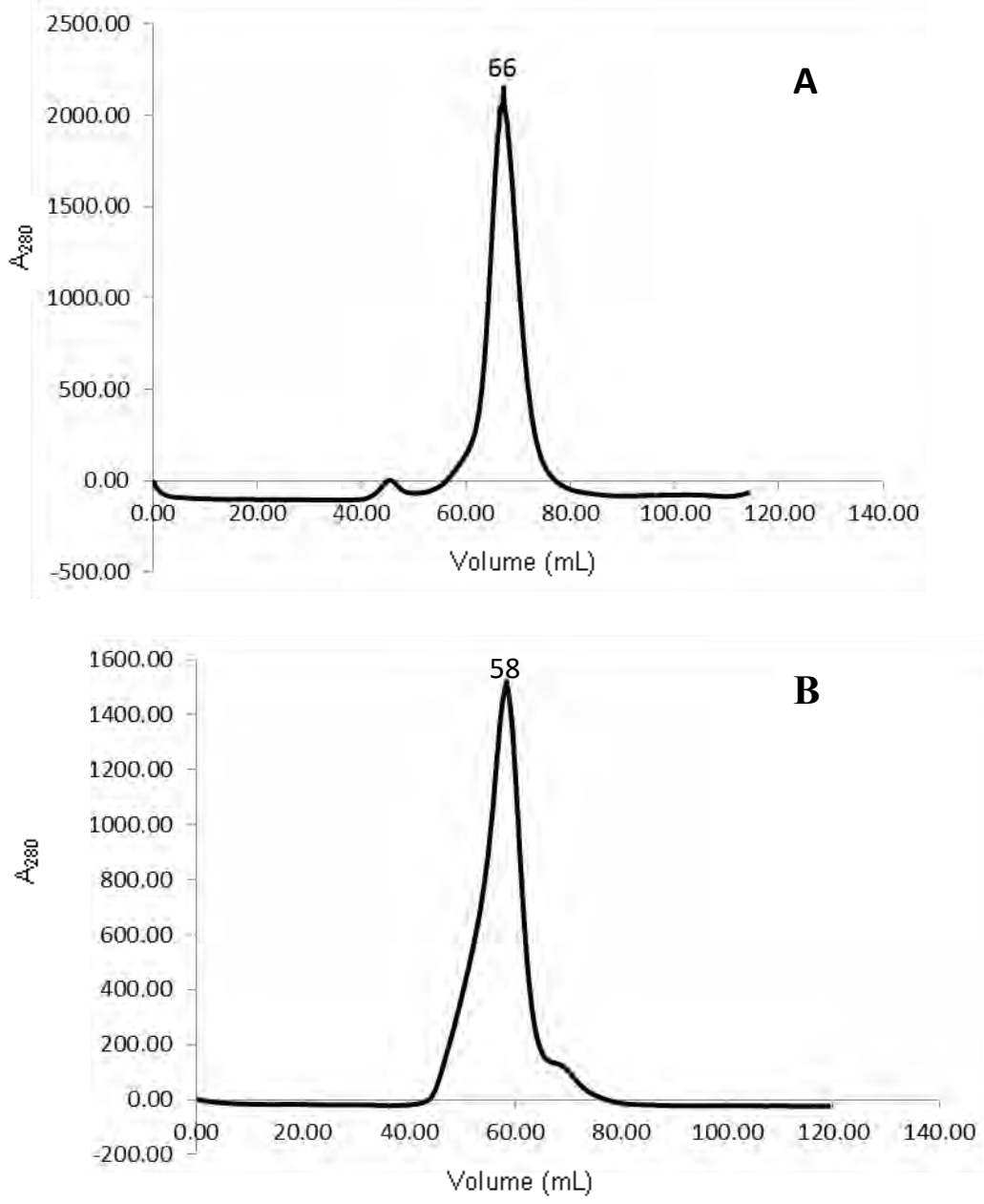

Figure 3.8: Re-*chromatography of monomer peak (A) and dimer peak (B)*

The chromatography from this was seen in fig 3.8. The process of re-chromatography found to be relatively successful. It did produced good quality crystal in the case of the monomeric peak, but in the case of the dimer peak, crystals acquired were not of usable quality. These crystals occurred as small and tended to occur in a range of different shapes. Initial beliefs were that these were crystals of the dimer form. This was later quashed with later

observations suggesting that these were not actual dimer crystals but rather dimer which had converted to monomer at high protein concentrations.

This does raise an interesting point from this study as it implies that the dimer population can be converted to the monomer population. This whole idea actually came to light through running the IMAC 250 mM Imidazole fraction through SEC, with the protein being accidently over concentrated due to a miscalculation of how much protein which was present. The reason for this was the switch from the original method of lysis to the method including sonication had been more efficient than originally suggested. This miscalculation created a stage where the monomer peak was massively amplified in comparison with the dimer peak (exceeding the loading capacity of the column with an A_{280} of 4000 units). It was suggested that this conversion had only occurred due to the amount of imidazole present or from the process of sonication. Attempts to reproduce this using sonication and through running samples spiked with 250 mM and samples eluted at 250 mM all proved unsuccessful. Due to this, all aspects of the procedure were re-examined, and it was decided to investigate whether it had been the higher concentrations which had caused the conversion between the 2 populations. Upon intensive testing, this did prove the case and there was now a method to further increase the overall yield.

There were still a number of issues, the first of which was the fact that the protein in question at the concentrations needed to achieve this, overloaded the UV detector. This meant that the chromatography in question was of very low quality. It was therefore often hard to see if the conversation had been complete or whether there was a second population present without rerunning some of the sample. Secondly, despite the majority of the protein being converted, a second chromatography step was often needed to remove a large dimer shoulder which occasionally occurred.

Regardless, all of this had shown that it was populations to convert into one another, and now it was now known that the monomer can be acquired from the dimer. The study aimed for a method to ensure complete conversion. Initially this was attempted by trying to concentrate the protein beyond the 35 mg/mL where initial results indicated that this was where the protein conversion started to occur (a lower concentration that the crystal plates had been previously set up at, meaning the crystals seen in the dimer plates probably came from monomers). Concentration of the protein beyond this stage was a tedious process with precipitation beginning to occur as low as 43 mg/mL.

Following attempts to convert populations via changes in concentration, an attempt was made to interconvert the "dimer" and "monomer" peaks, by exposing the mixed population to the probable substrate sulfite. This method had previously been deployed in a semi-successful manner for a member of the group who had been trying to achieve something similar for membrane proteins. The initial work on this process was performed by splitting a sample of IMAC purified protein in a 0.5 M NaCl, 50 mM Tris-HCl pH 7.6, 250 mM Imidazole buffer into 2 equal lots. One of these was injected as is and one of these was incubated with 100 mM sodium sulfite (the potential substrate) on ice for 30 min. The results of this, shown in the figure below, showed that the sulfite had a considerable effect, with after just half an hour of incubation with the substrate, the ratio of the peaks changing dramatically within this prep. The protein went from having three peaks, to a stage that it only has the aggregate peak and the monomer peak, with a slight shoulder for the dimer peak. This showed that the protein could be easily converted from the dimer to the monomer and all that was needed was an incubation on ice in the presence of the substrate.

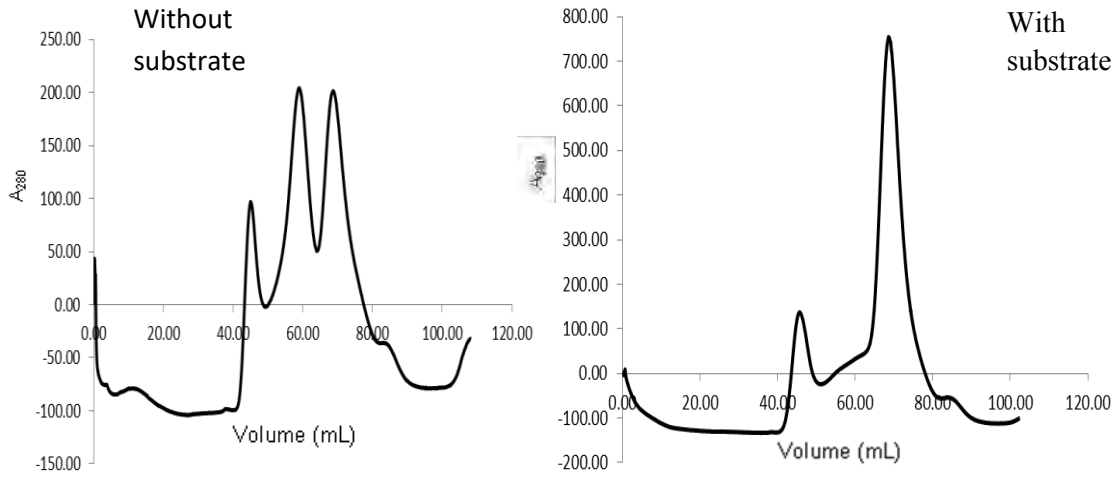

Figure 3.9: ***Chromatography of peak before and after incubation with SO_3***
Size exclusion chromatography of the IMAC purified protein before (left) and after (right) a 30 min incubation with 100 mM Sodium sulfite on ice

It was possible to argue at this stage, that there are a number of reasons why the protein may have changed from one population to another, such as the effect of the cold or by a random occurrence, and not necessarily through the effect of the substrate. It was also unknown if it was necessary to perform this on ice. Due to this, a number of combinations were attempted to see if these had an effect, with the protein incubated on ice with no substrate, with substrate at room temperature, and on ice with the substrate present. Investigation into the

period of incubation were also carried out with a 1 h incubation tested instead of 30 min to see if this removed the shoulder (i.e. to achieve total conversion). The results were simple; the incubation with sulfite was the only way the protein was found to occur in one population. This was regardless of if the sample was kept on ice. Without the substrate no conversion occurred. This showed that it was the substrate and not the ice driving the conversion. Additionally, the 1 hour incubation did prove more successful than the half hour incubation, showing full conversion. On a side note, the crystals produced from this method seem to be off superior quality

3.2.4 Conclusion on purification

Following this work, it was concluded that the best way to proceed in order to achieve a high yield of high quality, homogeneous protein, was to modify the original purification procedure to include a sonication step to improve the quantity of isolated protein and with the addition of a SO_3^{2-} species at 100 mM and the incubation of this protein with this substrate on ice for 1 h prior to injection on a size exclusion column

3.3 Crystallisation of target protein

3.3.1 Initial crystallisation

Prior to the commencement of the study, the protein had of interest had been crystallized, but only on a single occasion using a sample containing all 3 populations in the presence of 100 mM HEPES pH 7.0, 2% PEG and 15% Tacsimate for one plate and in the presence of 0.2 M Ammonium citrate pH 7.0 and 20% PEG on a second. Following this, screens and recreations of the initial conditions had failed to produce any successful hits. This still supplied a strong starting point, with conditions which had previously worked to work around. To commence this part of the study, plates were prepared using a mixed population of protein, from all three populations. This combination of populations did not produce any crystal hits. This agreed with results acquired by the outgoing postdoc over the past number of months. These conditions were still unsuccessful, when a single peak was utilized, illustrating that there must be something else at play, with the growth, lysis and purification procedure of the protein all being the exact same as when the crystallisation was successful.

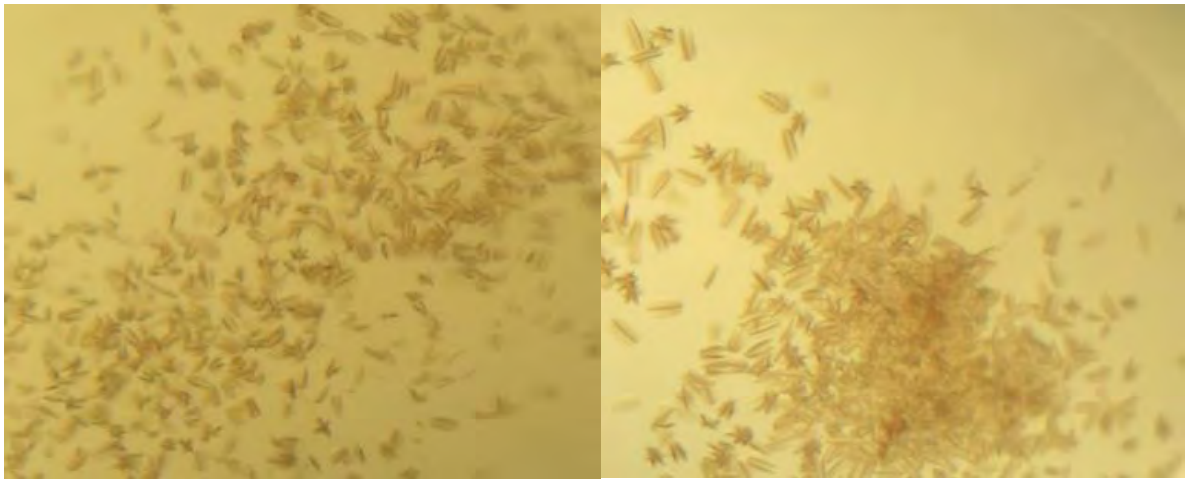

Figure 3.10: ***Initial hits from the index screen***
Initial hits acquired from the index screen using 40 mg/ml of purified SUR

The original crystallisation conditions were revisited and the conditions which had been successful were re-examined and due to this. Facts behind the original purification run emerged such as the fact that 5 M NaCl had been added dropwise to the purified protein during concentration of the protein post SEC due to the fact the protein had precipitated out of solution. This was identified as a potential avenue of investigation.

The protein drops were now set up on crystal plates using different concentrations of sodium chloride in addition to the base conditions of 100 mM HEPES pH 7.0, 2% PEG and 15%

Tacsimate and 0.2 M Ammonium citrate pH 7.0 and 20% PEG. The investigation did show something unusual, though in line with what was expected from the knowledge acquired from the initial prep that higher concentrations of sodium chloride resulted in crystal hits. The original index screen was re-investigated. This time the screen was re-performed at 40 mg/ml of protein in 50 mM Tris-HCl pH 7.6. The results of this were unusual, in the fact that nether 100 mM HEPES pH 7.0, 2% PEG and 15% Tacsimate nor 0.2 M Ammonium citrate pH 7.0 and 20% PEG producing crystals. Crystals were found to be present in conditions 9 (3.0 M NaCl, 100 mM BIS-TRIS pH 5.5), 10 (3.0 M NaCl, 100 mM BIS-TRIS pH 6.5), 11 (3.0 M NaCl, 100 mM HEPES pH 7.5) and 12 (3.0 M NaCl, 100 mM Tris-HCl pH 8.5), all of which contained 3.0 M NaCl. This implied that it was the NaCl, and not the actual PEG, Tacsimate or Ammonium citrate which resulted in the formation of crystals, showing that all that was needed to ensure crystal growth was the right pH and the right concentration of NaCl.

After examination of the plates from the index screen, the results were verified using both BIS-TRIS and MOPS at pHs of 5.5 and 6.5 and using Tris-HCl at pHs of 7.5 and 8.5, with crystal hits appearing in all conditions, with the higher pHs producing crystals at a more rapid rate, with a large number of crystals appearing overnight, while at the lower pHs of 5.5 and 6.5 generated crystals after 2-3 days. In all cases, the crystals took approximately a week to reach their optimal size.

3.3.2 Optimization and the additive screen

There were still a number of issues with these crystals, with the crystals being difficult to work with, both with them showing a level of instability in the cryo protectant (which was at this stage 40% sucrose) and having soft edges. The diffraction of these crystals were only marginally better than what was seen from the 100 mM HEPES pH 7.0, 2% PEG and 15% Tacsimate condition or the 0.2 M Ammonium citrate pH 7.0 and 20% PEG condition, both of which were also set up at 40 mg/mL. Initially investigations were done to identify whether the temperature the plates were set up at (with plates being set up at 4 ^{o}C, 16 ^{o}C and 20 ^{o}C) or the concentration of the protein (with 10 mg/mL, 20 mg/mL, 30 mg/mL and 40 mg/mL all tested with varying concentrations of NaCl) in the droplet would have an effect.

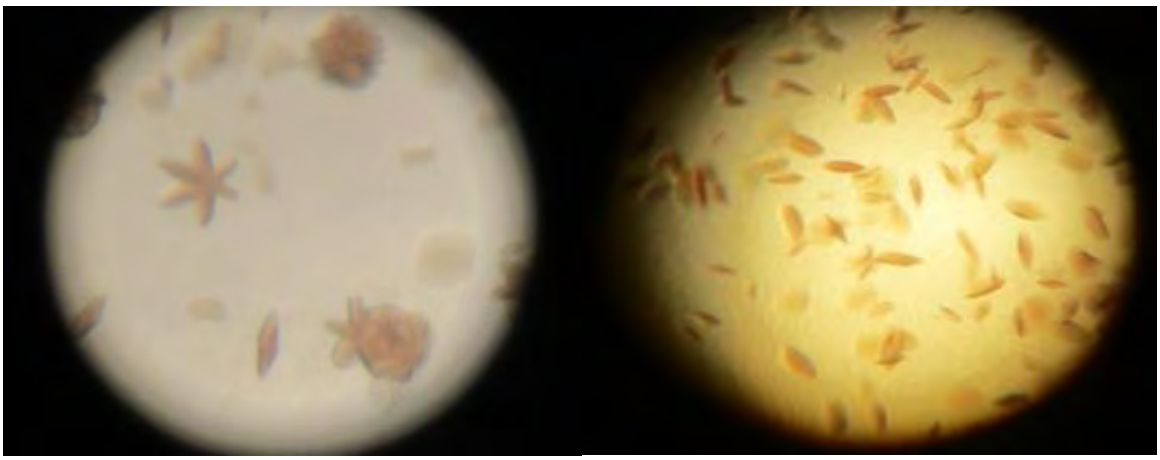

Figure 3.11: Crystal *hits from the additive screen*

Crystal *hits from the additive screen using 3.5 M NaCl, 100 mM Tris-HCl pH 7.5 and 40 mg/ml of purified SUR*

At this stage plates had been via the sitting drop method, hence a single hanging drop plate was set up to see if this had an effect. The crystals in all cases seemed to have a similar form and size, and still seemed to have the same issue with softness and difficulty of handling. It was decided that some form of additive was needed to improve the crystal quality. However, there was a minor, reproducible improvement to the crystals seen when it came to investigating the effect of the NaCl concentration on the formation of crystals, with the best crystals forming with an NaCl concentration of 3.5 M, so all plates from this point on were set up with 3.5 M NaCl, 100 mM Tris-HCl pH 7.5 (See fig 3.11).

Following this set of results, an additive screen was needed to achieve better quality crystals. This screen has been designed to give as wide a range of possible conditions as possible with everything from volatile substances to divalent salts and a large number of viscous substances. It does not cover the potential substrates of the protein of interest, sulfite or nitrite (Cudney et al., 1994; Hörer et al., 2013; Resch et al., 2013; Sousa, 1995; Trakhanov & Quiocho, 1995). Despite lacking the substrate, the screen itself should have supplied a better quality crystal and if not would at least supply a good idea of what was needed to achieve better protein crystals and this did prove to be the case. These screens were set up at concentrations of 20 mg/mL, 30 mg/mL and 40 mg/mL of the monomeric sulfite reductase, using the base condition of 3.5 M NaCl, 100 mM Tris-HCl pH 7.5, resulting in the testing of 288 different conditions, giving a total of 46 different additives which produced high quality crystals, with 16 of these conditions working at all 3 concentrations of sulfite reductase (Table 3.2).

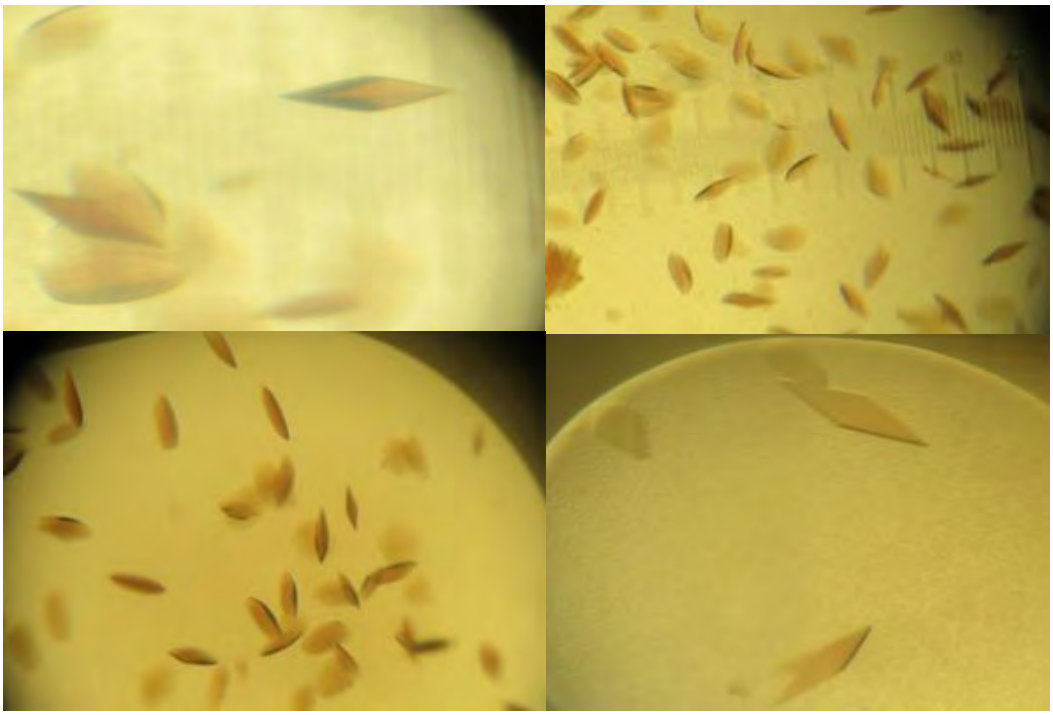

Figure 3.12: ***Results of additive screen***

Crystals from the top left, clockwise, 40 mg/mL index screen with 0.5 M Sodium fluoride as an additive, 30 mg/mL index screen with 0.1 M Taurine as an additive, 20 mg/mL screen with 0.1 M Strontium chloride hexahydrate as an additive and finally 20 mg/mL screen with 0.1 M Phenol as an additive

There were a total of 46 potential additives from this screen which supplied high quality crystals. Following inspection of all of these hits and through consultation with Dr. Mohamed Noor and Prof. Tewfik Soulimane, plates were set up according to the conditions in table 3.3, continuing with the base condition of 3.5 M NaCl, 100 mM Tris pH 7.5.

Table 3.2: *Results of additive screens for samples set up with 3.5 M NaCl, 100 mM Tris-HCl pH 7.5 and varying concentrations of protein (20, 30 or 40 mg/mL) and various additives, using usable crystal hits in green and micro crystals and precipitation in red (Conditions which showed only negative results are excluded, for full list of additives please refer to Additive Screen by Hampton Research)*

	20 mg/mL	30 mg/mL	40 mg/mL
0.1 M Barium chloride dihydrate			
0.1 M Calcium chloride dihydrate			
0.1 M Magnesium chloride hexahydrate			
0.1 M Strontium chloride hexahydrate			
0.1 M Chromium (III) chloride hexahydrate			
2.0 M Sodium Chloride			
0.5 M Sodium Fluoride			
1.0 M Caesium chloride			
0.1 M L-Proline			
0.1 M Phenol			
0.1 M Sodium bromide			
30% w/v 6-Aminohexanoic acid			
0.3 M Glycyl-glycyl-glycine			
0.1 M Taurine			
0.1 M Spermidine			
0.1 M Spermine tetrahydrochloride			
0.1 M Hexamine cobalt (III) chloride			
0.1 M Sarcosine			
0.1 M Trimethylamine hydrochloride			
0.1 M Urea			
0.01 M GSH (L-Glutathione reduced) 0.01 M GSSG (L-Glutathione oxidized)			
0.1 M Ethylenediaminetetraacetic acid (EDTA) disodium salt dehydrate			
5% w/v Polyvinylpyrrolidone K15			
30% w/v Sucrose			
40% v/v Pentaerythritolethoxylate (3/4 EO/OH)			
10% w/v Polyethylene glycol 3,350			

12% w/v myo-Inosito			
30% w/v D-(+)-Trehalose dehydrate			
30% v/v Ethylene glycol			
2.0 M NDSB-201			
2.0 M NDSB-211			
2.0 M NDSB-221			
1.0 M NDSB-256			
5% w/v n-dodecyl-N,N-dimethylamine-N-oxide, (LDAO, DDAO)			
5% w/v n-Dodecyl-b-D-maltoside			
30% w/v Trimethylamine N-oxide dihydrate			
30% w/v 1,6-Hexanediol			
50% v/v Jeffamine ® M-600 ® pH 7.0			
40% v/v (±)-1,3-Butanediol			
30% v/v 1,4 Dioxane			
30% v/v 2-Propanol			
40% v/v tert-Butanol			
40% v/v 1,3-Propanediol			
40% v/v Formanmide			
40% v/v 1-Propanol			
0.25% v/v Dichloromethane			
7% v/v 1-Butanol			

Table 3.3: *Conditions used to set up crystal plates, with one row of plate set up of each using 4 µL drops (2 µL protein, 1.6 µL reservoir solution, 0.4 µL additive)*

Condition	Concentration of Protein	Additive
1	30 mg/mL	0.1 M Barium chloride dihydrate
2	30 mg/mL	0.1 M Calcium chloride dihydrate
3	20 mg/mL	0.1 M Strontium chloride hexahydrate
4	30 mg/mL	0.5 M Sodium fluoride
5	30 mg/mL	0.1 M L-Proline
6	30 mg/mL	0.1 M Phenol
7	30 mg/mL	0.1 M Taurine
8	20 mg/mL	0.1 M Sarcosine
9	30 mg/mL	0.1 M Urea
10	30 mg/mL	5 % w/v Polyvinylpyrrolidone K15
11	30 mg/mL	2.0 M NDSB-221
12	30 mg/mL	30% v/v 1,4 Dioxane
13	30 mg/mL	30% v/v 2-Propanol

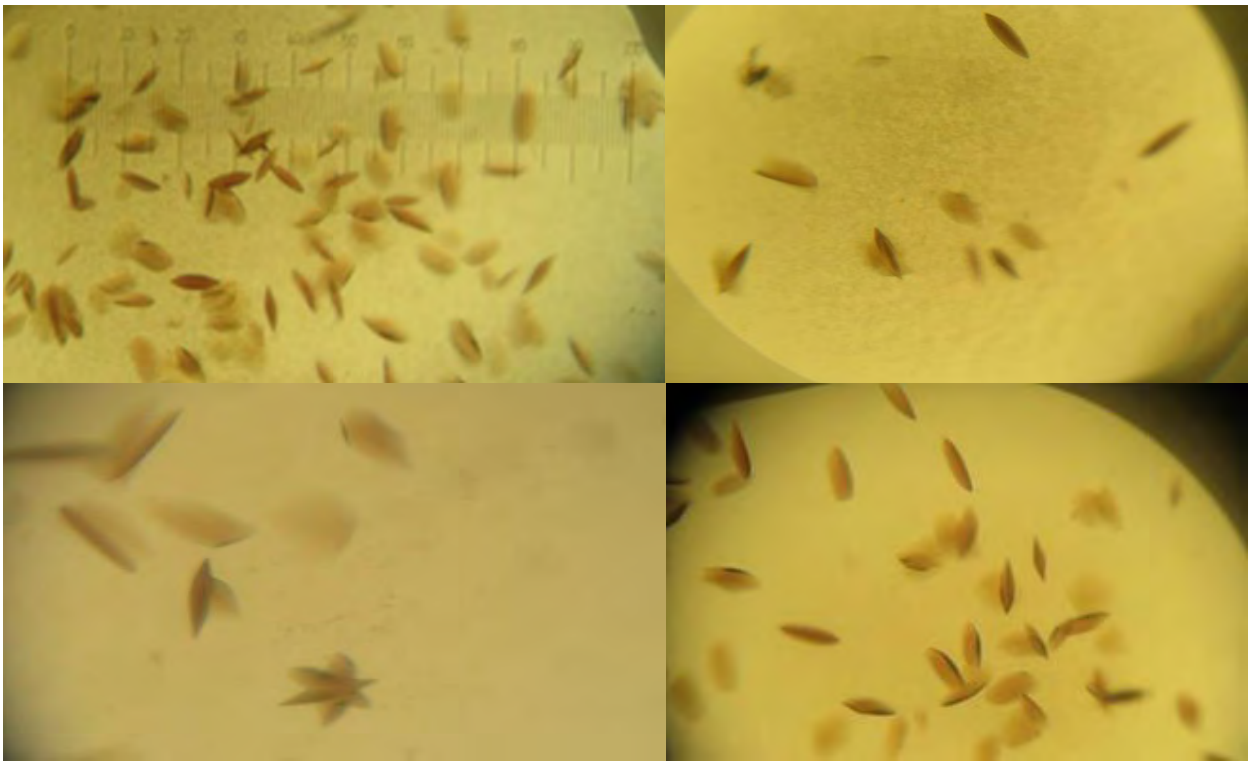

Figure 3.13: ***Results of the additive screen with 30 mg/mL SUR***

Crystals acquired from optimization of the additive screen with the pictures from the top left, clockwise being, 30 mg/mL plate with 0.1 M Calcium chloride dihydrate as an additive, 30 mg/mL plate with 0.1 M Barium chloride dihydrate as an additive, 30 mg/mL plate with 0.1 M Phenol as an additive and finally 30 mg/mL plate with 30% v/v 1,4 Dioxane as an additive

From these plates, a number of conditions stood out. 5 in particular;5% w/v Polyvinylpyrrolidone K15, 2.0 M NDSB-221, 0.1 M Urea, 0.1 M Sarcosine and0.1 M Strontium chloride hexahydrate. These plates were re-set up using 30 mg/mL for 5% w/v Polyvinylpyrrolidone K15, 2.0 M NDSB-221, 0.1 M Urea or 0.1 M Sarcosine as the additive, and at 20 mg/mL, crystals for 0.1 M Strontium chloride hexahydrate.

For these 5 plates, there was the added advantage of the sulfite reductase now being used coming from the production runs including the 1 h incubation with sulfite. The crystals acquired from this were off better quality and of higher diffraction quality.

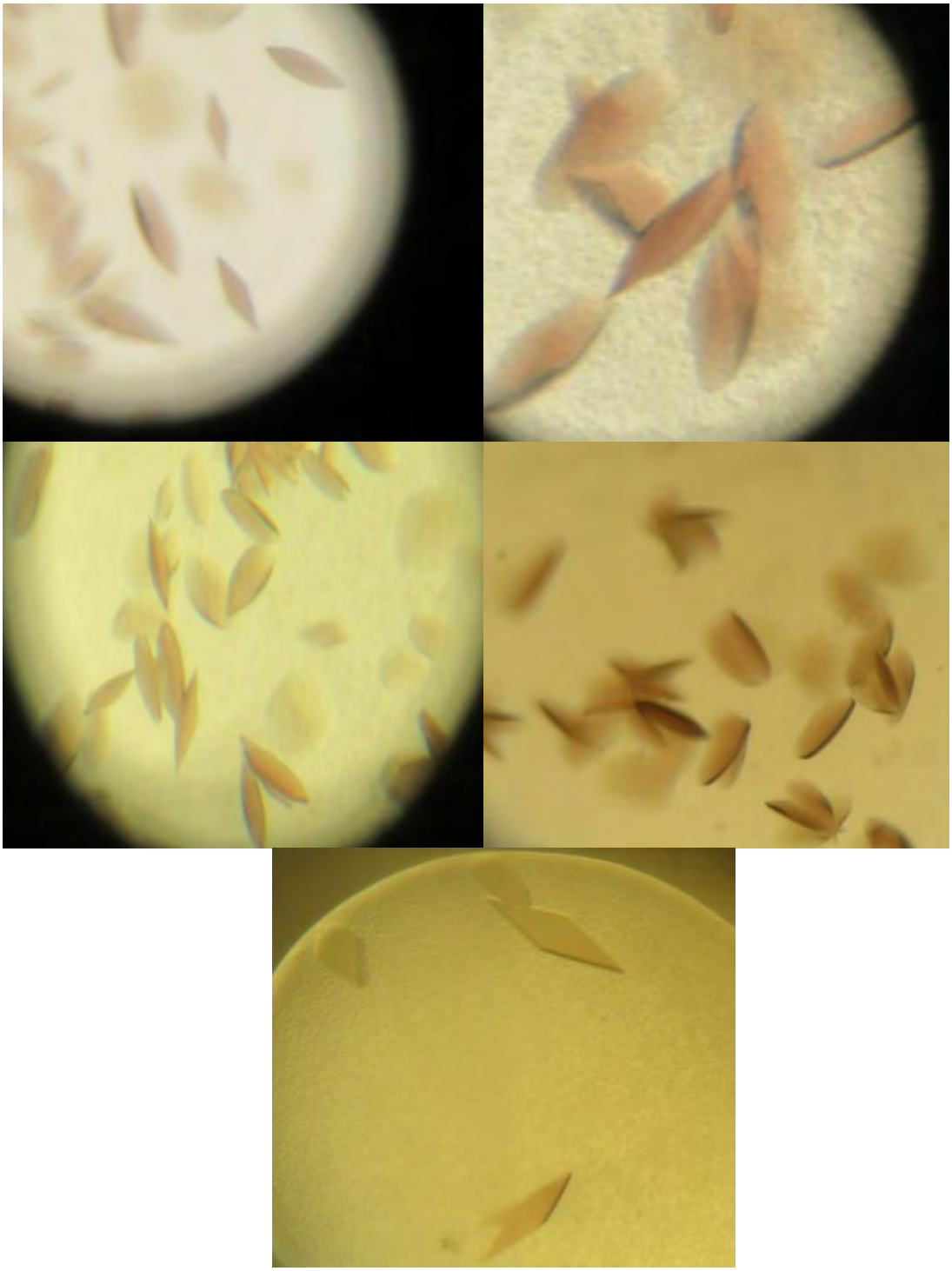

Figure 3.14: ***Additive screen optimization results***

Crystals acquired from optimization of the additive screen with the pictures from the top left, 30 mg/mL plate with 5% w/v Polyvinylpyrrolidone K15.as an additive, top right 30 mg/mL plate with 2.0 M NDSB-221 as an additive, centre left 30 mg/mL plate with 0.1 M Urea as an additive, centre right 30 mg/mL plate with 0.1 M Sarcosine as an additive and bottom, 20 mg/mL plate with 0.1 M Strontium chloride hexahydrate as the additive

3.3.3 The use of substrates in order to optimize crystallisation

Attempts were also made to crystalize the protein with the potential substrates. The protein was crystalized with a number of different forms of the potential substrate sulfite and nitrite, such as the calcium, sodium and potassium salts. Each of these were set up on 2 plates, one using concentrations of between 10 and 250 mM and one between 1 mM and 20 mM, with high quality crystals being achieved with Calcium nitrite and Sodium sulfite, both at 25 mM. Of all of these conditions, the addition of 25 mM sodium sulfite produced the best crystals.

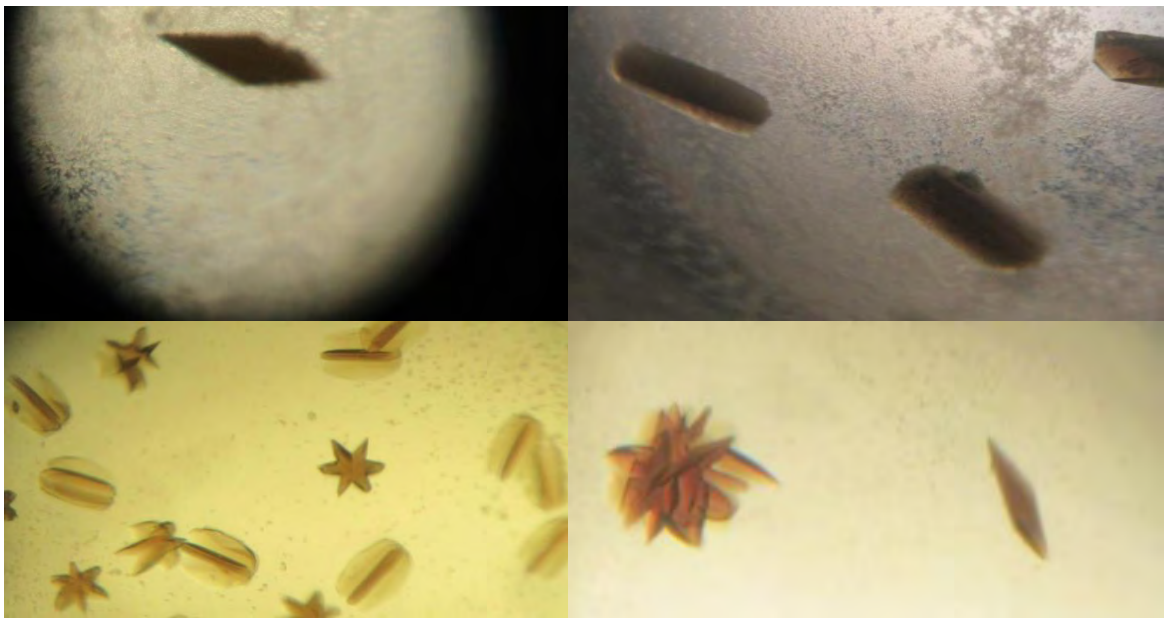

Figure 3.15: ***Crystals acquired from interaction with substrates***

Crystals acquired from optimization of the plates with the pictures from the top left, 30 mg/mL plate with 0.1 M Sodium sulfite.as an additive, top right 30 mg/mL plate with 0.1 M Calcium nitrite as an additive, bottom left 30 mg/mL plate with 0.1 M Potassium sulfite as an additive, bottom right 40 mg/mL plate with 0.1 M Sodium sulfite

3.3.4 Crystallisation conclusion

Throughout the course of this project, it is possible to see that the process of crystallisation has been optimized with the initial conditions proving somewhat unsuccessful prior to the addition of an additional chemical species, in the form of NaCl. Relatively poor crystals were still produced here; only refracting to relatively low resolutions. It can be concluded that the originally diffractable crystals came about more as a happy mistake than a breakthrough, to such an extent that if had the original conditions been kept it is possible that there may have never been another crystal hit. The process of conversion to the method used here was a long

and painful one with 384 conditions prepared from commercial screens, in addition to the 96 conditions set up for both the dimer peak and all three peaks combined (none supplying a usable hit but did show poor crystals under the NaCl conditions, which we later explained was through a partial conversion of the monomer)

Table 3.4: *Break down of screens set up of monomer peak from commercial kits*

#	Crystallisation Screen	No. of Conditions
1	Index Screen	96
2	Additive Screen (20 mg/mL)	96
3	Additive Screen (30 mg/mL)	96
4	Additive Screen (40 mg/mL)	96
	Total	384

Following the use of the commercial screens, the conditions were further optimized, by alternating the pH of the solution, by alternating the protein concentration and through the additional of different chemical species such as the substrates. From this, the final condition used for the process of crystallisation was a protein concentration of 30 mg/mL with 12.5mM Sodium sulfite.as an additive and using 3.5 M NaCl, 100 mM Tris-HCl pH 7.5 as the base condition. Overall, it was seen that the biggest contributing factor came in the form of the incubation with the sulfite. This improved the resolution of the structure massively, with a resolution of 2.3 Å, down from initial resolution of over 4 Å down. This proved to be an added bonus onto the additional usable protein. This may have been due to the protein now naturally occurring in one population, or a lower level of damage due to lower levels of concentration and chromatography needed.

3.4 Crystallography

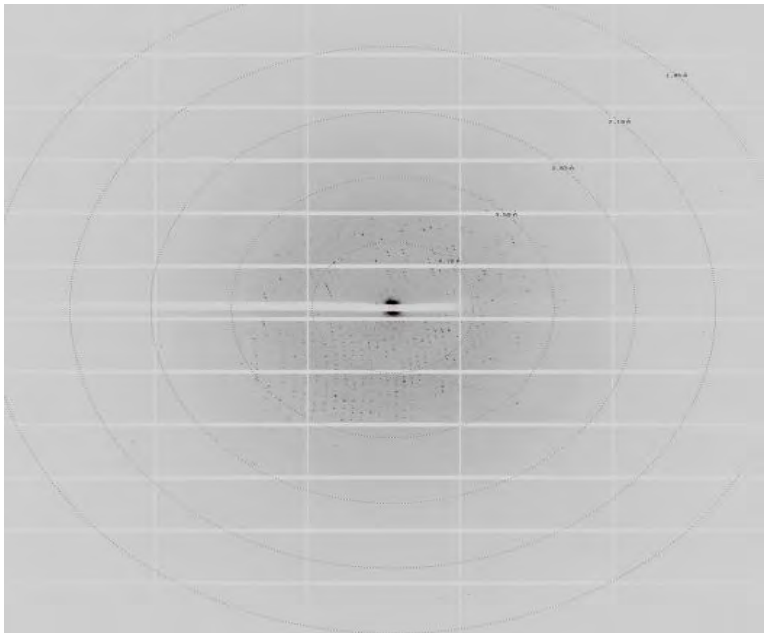

Figure 3.16: ***X-ray diffraction pattern of T. thermophilus SUR***

Diffraction pattern achieved from sulfite reductase crystals grown in 3.5 M NaCl, 100 mM Tris-HCl pH 7.5with 12.5 mM Sodium sulfite.as an additive

With the crystals optimized at a number of different stages throughout this study, it was finally made possible to analyse the data acquired and use this in order to produce a crystal structure. In each case, the protein crystals acquired were shot, generating a diffraction pattern like the pattern shown in fig 3.16, which could be then analysed to give us a 3-D structure. By combining these datasets, with the datasets originally acquired from the initial work, it was possible to start to see the outline of a crystal structure. With the addition of the datasets from the crystals prepared with different additives and with the potential substrates of this particular sulfite reductase, a structure can finally be seen, like the one in fig 3.17. This shows a number of important components required for the activity of the sulfite reductase, such as the siroheme and the Fe-S cluster.

Using the protein's diffraction pattern in this case and the pdb (protein data bank) file generated from this data, it can be seen that this molecule presents as a monomer. This is reinforced by data acquired from SEC.

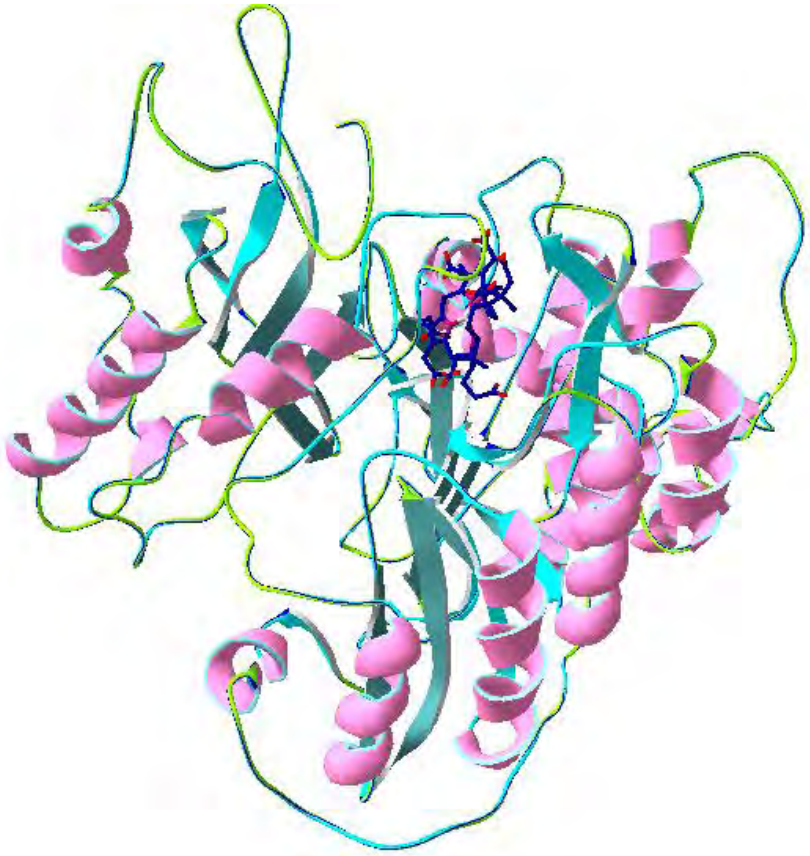

Figure 3.17: ***Structure of T. thermophilus monomer***

Crystal structure prediction for the sulfite reductase of Thermus thermophilus

Analysis of this data as a whole, it can be observed that the data acquired on this protein does agree with the data previously acquired from other proteins in this family, with the siroheme and the Fe-S cluster occurring within the clef shown in fig 3.18. As expected, these co-factors occur in the area with the usual Lys and Arg domains occurring with

The positioning of the siroheme and the Fe-S cluster here is consistent with what was seen in other sulfite reductases and can be directly compared with the β-subunit of sulfite reductase from *E. coli*, with the siroheme positioned across from 2 Lys and 2 Arg. Similarities are clearly visible with the SUR of *E. coli* (see fig 1.7) with the active site the Lys residues at points 174 and 176 on the protein backbone as opposed to positions 215 and 217 in the *E. coli* version but maintaining similar spacing on the backbone (Crane & Getzoff, 1996). The Arg residues of interest occur at positions 97 and 133 on the backbone instead of the 83 and 153 in *E. coli* but occurs in the same area when comparing the 3-D configuration of the active site, shown in fig 3.17 for *Thermus thermophilus* and in fig 1.7 for *E. coli* (Crane et al., 1995).

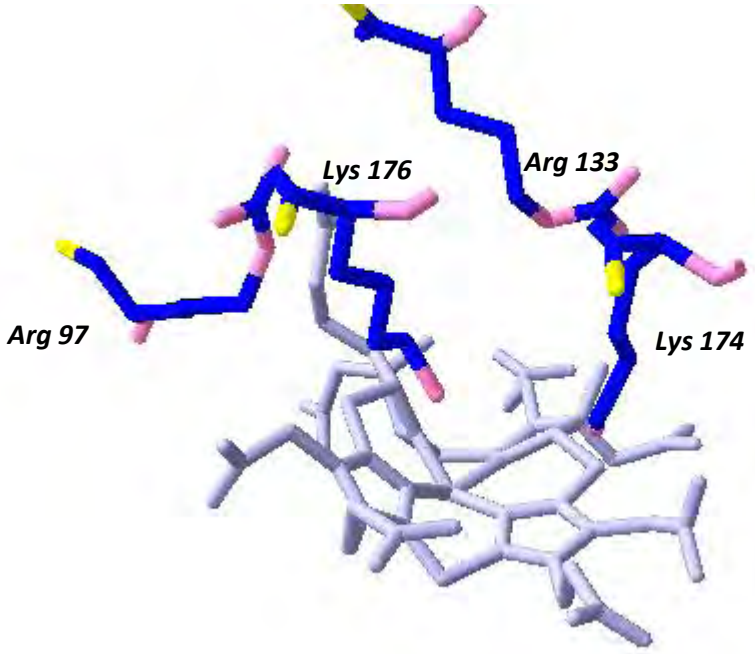

Figure 3.18: ***Active site of T. thermophilus SUR***

Active site of the sulfite reductase of Thermus thermophilus showing the positioning of siroheme (red), the Fe-S structure and the key residues in crystal structure

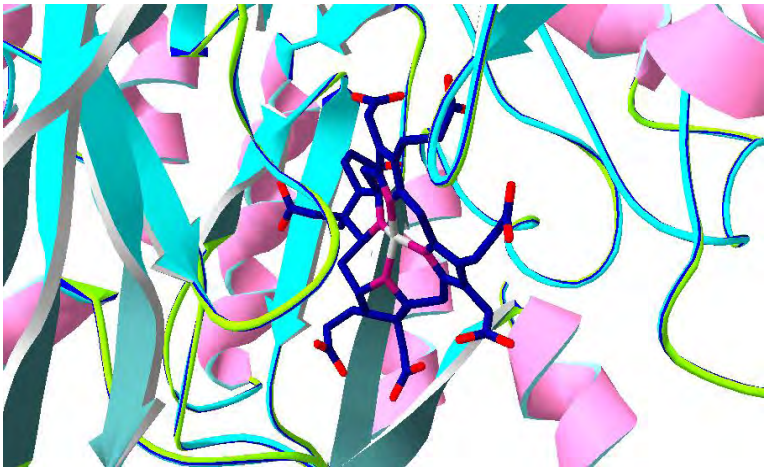

Figure 3.19: ***Active site of T. thermophilus SUR***

Positioning of siroheme (in blue) in crystal structure of the sulfite reductase of Thermus thermophilus

This active site is opened on the surface and can be reached easily by the substrate as can be seen in fig 3.18. Alongside the presence of all the residues needed for sulfite reduction, it is possible to identify this as having a similar mechanism as the *E. coli* active site. Reinforced by how these residues operate in unison with the siroheme for the reduction of sulfite, it can be concluded that this enzyme would operate via a similar mechanism to what is shown in fig 1.5. It is proposed that the substrate with interact with the siroheme and the residues Lys 176, Lys 174, Arg 133 and Arg 97 using a similar mechanism with the protonation occurring via Arg 133 and Lys 174. The Fe-S cluster would also still remain as a method for electron transfer.

This structure, occurring with a C 2 2 21 space group and a unit cell of 131.27 Å X 424.19 Å X 84.52 Å, as shown in table 3.5, has been solved to the statistical levels shown below.

Table 3.5: *Statistical data for SUR crystal structure*

	SUR *
Wavelength	1.738
Resolution range	78.52 - 2.38 (2.465 - 2.38)
Space group	C 2 2 21
Unit cell	
a,b,c (Å)	131.27 424.19 84.52
α, β, γ	90, 90, 90
Total reflections	1230515 (116165)
Unique reflections	94726 (9324)
Multiplicity	13.0 (12.5)
Completeness (%)	0.99 (0.99)
Mean I/sigma(I)	8.96 (0.45)
Wilson B-factor	57.90
R-merge	0.2136 (5.606)
R-meas	0.2224 (5.846)
R-pim	0.06156 (1.648)
CC1/2	0.998 (0.559)
CC*	1 (0.847)
Reflections used in refinement	94726 (8912)
Reflections used for R-free	4813 (467)
R-work	0.2413 (0.5582)
R-free	0.2941 (0.5926)
CC(work)	0.949 (0.640)
CC(free)	0.918 (0.573)

Number of non-hydrogen atoms	8634
macromolecules	8440
ligands	150
solvent	44
Protein residues	1081
RMS(bonds)	0.016
RMS(angles)	1.39
Ramachandran favoured (%)	84
Ramachandran allowed (%)	11
Ramachandran outliers (%)	4.8
Rotamer outliers (%)	15
Clashscore	13.81
Average B-factor	94.56
macromolecules	93.88
ligands	143.59
solvent	59.29

Statistics for the highest-resolution shell are shown in parentheses.

4.0 Conclusion and future work

Over the course of this study, it was shown that the protein could be isolated at a higher yield, without resulting in damage to the protein, through adding one additional stage to the lysis procedure in the form of sonication. This resulted in a 200% increase in protein isolated with the amount of usable protein produced per L increasing from 0.8 mg/L to 24 mg/L. This came with no increase in the percentage of the protein in an aggregated state. Also no additional damage to the protein was observed. Hence it was adopted into the purification method. The original method of purification was further altered with the use of additional steps to bring the protein from 2 populations to 1. This was done in 2 ways. Rising the concentration of the protein and incubation with 100 mM sodium sulfite both supplied similar results here, with the later method becoming the method of choice due to the fact it produced complete conversion after an 1 h incubation on ice with this substrate. This combined with the previous adjustment to the procedure resulted in approximately a 400 % in protein production. The time needed for purification was also decreased, with one less chromatography step than the next fastest method.

The change in crystal conditions was also significant with the condition being changed from the 40 mg/mL sulfite reductase (mixed populations in 50 mM Tris-HCl pH 7.6), 100 mM HEPES pH 7.0, 2% PEG and 15% Tacsimate condition and the 40 mg/mL sulfite reductase (mixed populations in 50 mM Tris-HCl pH 7.6)0.2 M Ammonium citrate pH 7.0 and 20% PEG condition in the original protocol. The additional datasets were acquired in a range of conditions revolving around a protein concentration of 20-40 mg/mL (monomeric in 50 mM Tris-HCl pH 7.6) with 3.5 M NaCl, 100 mM Tris-HCl pH 7.5 as the reservoir solution. This was further refined using a range of different additives. Work on the initial conditions concluded that the initial hits were due to the NaCl used to bring the protein back into solution after precipitation and not from the other components of the solution. Due to this, the screening was reopened and it was found that the best crystals were acquired at 30 mg/mL sulfite reductase purified in 50 mM Tris-HCl pH 7.6 with 12.5 mM sodium sulfite as an additive with 3.5 M NaCl, 100 mM Tris-HCl pH 7.5 as the reservoir solution.

The crystal data acquired showed that the protein resembles the β-subunit of the model protein for this family, the sulfite reductase of *E. coli*. Importantly, as a result of a combination of the data from SEC calibration and from the crystal data, it is proposed that the

protein of interest here is a monomer. The protein also showed the presence of both the Fe-S cluster and the siroheme, which are important pieces of information as these are both needed for the process of sulfite reduction. Alongside this, Arg residues occur at position 97 and 133 and Lys residues at positions 174 and 176 on the protein backbone. This positioning implies that the mechanism of sulfite reduction in the protein is comparable to what was proposed for the *E. coli* sulfite reductase shown in fig 1.5. This would make sense considering the homology between the proteins.

From this point, there are still a number of areas still to be considered, such as methods aiding the further enhancement of the crystal structure, through the fine tuning of the conditions. There is still work outstanding in order to figure out the enzyme kinetics and what is the electron donor in the case of this enzyme. Current work by the study leans towards this being ferredoxin the electron donator for the maize sulfite reductase. This reinforced by the fact that ferredoxin is produced by the organism in question. Hence plates are being set up around this, with ferrodoxin present. An additive screen will also be carried out using the ideal conditions, containing sodium sulfite as an additive on all conditions

5.0 Bibliography

Amend, J. P., & Shock, E. L. (2001). Energetics of overall metabolic reactions of thermophilic and hyperthermophilic Archaea and Bacteria. *FEMS microbiology reviews, 25*(2), 175-243.

Bertoldo, C., Grote, R., & Antranikian, G. (2002). Extremophiles: Life in Extreme Environments. *Encyclopedia of Environmental Microbiology*.

Bickle, M., Alt, J., & Teagle, D. (1994). Sulphur transport and sulphur isotope fractionations in ocean floor hydrothermal systems. *Mineral. Mag. A, 58*, 88-89.

Bracewell, R. N. (1976). *Refinements to Drake's equation.* Paper presented at the Anaheim International Astronautical Federation Congress.

Brychkova, G., Yarmolinsky, D., Ventura, Y., & Sagi, M. (2012). A novel in-gel assay and an improved kinetic assay for determining in vitro sulfite reductase activity in plants. *Plant and Cell Physiology, 53*(8), 1507-1516.

Chayen, N. E. (2004). Turning protein crystallisation from an art into a science. *Current opinion in structural biology, 14*(5), 577-583.

Cole, J. (1989). Physiology, biochemistry, and genetics of nitrite reduction by Escherichia coli. *Molecular and genetic aspects of nitrate assimilation.*, 229-243.

Crane, B. R., & Getzoff, E. D. (1996). The relationship between structure and function for the sulfite reductases. *Current opinion in structural biology, 6*(6), 744-756.

Crane, B. R., Siegel, L. M., & Getzoff, E. D. (1995). Sulfite reductase structure at 1.6 Å: evolution and catalysis for reduction of inorganic anions. *Science, 270*(5233), 59-67.

Crane, B. R., Siegel, L. M., & Getzoff, E. D. (1997). Probing the catalytic mechanism of sulfite reductase by X-ray crystallography: structures of the Escherichia coli hemoprotein in complex with substrates, inhibitors, intermediates, and products. *Biochemistry, 36*(40), 12120-12137.

Cudney, R., Patel, S., Weisgraber, K., Newhouse, Y., & McPherson, A. (1994). Screening and optimization strategies for macromolecular crystal growth. *Acta Crystallographica Section D: Biological Crystallography, 50*(4), 414-423.

Degryse, E., Glansdorff, N., & Piérard, A. (1978). A comparative analysis of extreme thermophilic bacteria belonging to the genus Thermus. *Archives of Microbiology, 117*(2), 189-196.

Fauque, G. D. (1995). Ecology of sulfate-reducing bacteria. In *Sulfate-Reducing Bacteria* (pp. 217-241): Springer.

Gillespie, R. J. (1992). Multiple bonds and the VSEPR model. *J. Chem. Educ, 69*(2), 116.

Goubern, M., Andriamihaja, M., Nübel, T., Blachier, F., & Bouillaud, F. (2007). Sulfide, the first inorganic substrate for human cells. *The FASEB Journal, 21*(8), 1699-1706.

Hansen, J., Muldbjerg, M., Chérest, H., & Surdin-Kerjan, Y. (1997). Siroheme biosynthesis in Saccharomyces cerevisiae requires the products of both the MET1 and MET8 genes. *FEBS letters, 401*(1), 20-24.

Hell, R., Dahl, C., Knaff, D. B., & Leustek, T. (2008). *Sulfur metabolism in phototrophic organisms* (Vol. 27): Springer Science & Business Media.

Hermann, B., Kern, M., La Pietra, L., Simon, J., & Einsle, O. (2015). The octahaem MccA is a haem c-copper sulfite reductase. *Nature, 520*(7549), 706-709.

Hörer, S., Reinert, D., Ostmann, K., Hoevels, Y., & Nar, H. (2013). Crystal-contact engineering to obtain a crystal form of the Kelch domain of human Keap1 suitable for ligand-soaking experiments. *Acta Crystallographica Section F: Structural Biology and Crystallisation Communications, 69*(6), 592-596.

Johnston, D. T., Poulton, S. W., Fralick, P. W., Wing, B. A., Canfield, D. E., & Farquhar, J. (2006). Evolution of the oceanic sulfur cycle at the end of the Paleoproterozoic. *Geochimica et Cosmochimica Acta, 70*(23), 5723-5739.

Jorgensen, B., & Postgate, J. (1982). Ecology of the bacteria of the sulphur cycle with special reference to anoxic-oxic interface environments [and discussion]. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 298*(1093), 543-561.

Kappler, U. (2011). Bacterial sulfite-oxidizing enzymes. *Biochimica et Biophysica Acta (BBA)-Bioenergetics, 1807*(1), 1-10.

Kappler, U., & Dahl, C. (2001). Enzymology and molecular biology of prokaryotic sulfite oxidation. *FEMS microbiology letters, 203*(1), 1-9.

Kasting, J. F., Zahnle, K., Pinto, J., & Young, A. (1989). Sulfur, ultraviolet radiation, and the early evolution of life. *Origins of Life and Evolution of the Biosphere, 19*(2), 95-108.

Khurshid, S., Saridakis, E., Govada, L., & Chayen, N. E. (2014). Porous nucleating agents for protein crystallisation. *Nature protocols, 9*(7), 1621-1633.

Kim, J. Y., Nakayama, M., Toyota, H., Kurisu, G., & Hase, T. (2016). Structural and mutational studies of an electron transfer complex of maize sulfite reductase and ferredoxin. *Journal of biochemistry*, mvw016.

Kleinhofs, A., Warner, R., Lawrence, J., Melzer, J., Jeter, J., Kudrna, D., et al. (1989). Molecular genetics of nitrate reductase in barley. *Molecular and genetic aspects of nitrate assimilation.*, 197-211.

Klotz, M. (2011). BryantDA and Hanson TE (2011) The microbial sulfurcycle. *The microbial sulfur cycle*, 5.

Kriek, M., Martins, F., Leonardi, R., Fairhurst, S. A., Lowe, D. J., & Roach, P. L. (2007). Thiazole synthase from Escherichia coli an investigation of the substrates and purified proteins required for activity in vitro. *Journal of Biological Chemistry, 282*(24), 17413-17423.

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *nature, 227*, 680-685.

LeGall, J., & Fauque, G. (1988). Dissimilatory reduction of sulfur compounds. *Biology of anaerobic microorganisms*, 587-639.

Lehninger, A., Nelson, D. L., & Cox, M. M. (2008). Lehninger Principles of Biochemistry & eBook. (WH Freeman: New York, NY, USA).

Leustek, T. (2002). Sulfate metabolism. *The Arabidopsis Book*, e0017.

Mathews, C., van Holde, K., Appling, D., & Anthony-Cahill, S. (1999). Biochemistry.(4thedn). (Prentice Hall, Redwood City, CA).

McDermott, J. M., Ono, S., Tivey, M. K., Seewald, J. S., Shanks, W. C., & Solow, A. R. (2015). Identification of sulfur sources and isotopic equilibria in submarine hot-springs using multiple sulfur isotopes. *Geochimica et Cosmochimica Acta, 160*, 169-187.

McPHERSON, A. (1990). Review Current approaches to macromolecular crystallisation. In *EJB reviews 1990* (pp. 49-71): Springer.

McPherson, A. (2004). Introduction to protein crystallisation. *Methods, 34*(3), 254-265.

McRee, D., Richardson, D., Richardson, J., & Siegel, L. (1986). The heme and Fe4S4 cluster in the crystallographic structure of Escherichia coli sulfite reductase. *Journal of Biological Chemistry, 261*(22), 10277-10281.

Murphy, M. J., & Siegel, L. M. (1973). Siroheme and sirohydrochlorin the basis for a new type of porphyrin-related prosthetic group common to both assimilatory and

dissimilatory sulfite reductases. *Journal of Biological Chemistry, 248*(19), 6911-6919.
Murphy, M. J., Siegel, L. M., Tove, S. R., & Kamin, H. (1974). Siroheme: a new prosthetic group participating in six-electron reduction reactions catalyzed by both sulfite and nitrite reductases. *Proceedings of the National Academy of Sciences, 71*(3), 612-616.
Nakano, S., Takahashi, M., Sakamoto, A., Morikawa, H., & Katayanagi, K. (2012). X-Ray Crystal Structure of a Mutant Assimilatory Nitrite Reductase That Shows Sulfite Reductase-Like Activity. *Chemistry & biodiversity, 9*(9), 1989-1999.
Nakayama, M., Akashi, T., & Hase, T. (2000). Plant sulfite reductase: molecular structure, catalytic function and interaction with ferredoxin. *Journal of inorganic biochemistry, 82*(1), 27-32.
Niknahad, H., & O'Brien, P. J. (2008). Mechanism of sulfite cytotoxicity in isolated rat hepatocytes. *Chemico-biological interactions, 174*(3), 147-154.
Noor, M. R., & Soulimane, T. (2012). Bioenergetics at extreme temperature: Thermus thermophilus ba 3-and caa 3-type cytochrome c oxidases. *Biochimica et Biophysica Acta (BBA)-Bioenergetics, 1817*(4), 638-649.
Oliveira, T. F., Vonrhein, C., Matias, P. M., Venceslau, S. S., Pereira, I. A., & Archer, M. (2008). The crystal structure of Desulfovibrio vulgaris dissimilatory sulfite reductase bound to DsrC provides novel insights into the mechanism of sulfate respiration. *Journal of Biological Chemistry, 283*(49), 34141-34149.
Oshima, T., & Imahori, K. (1974). Description of Thermus thermophilus (Yoshida and Oshima) comb. nov., a nonsporulating thermophilic bacterium from a Japanese thermal spa. *International Journal of Systematic Bacteriology, 24*(1), 102-112.
Parey, K., Warkentin, E., Kroneck, P. M., & Ermler, U. (2010). Reaction cycle of the dissimilatory sulfite reductase from Archaeoglobus fulgidus. *Biochemistry, 49*(41), 8912-8921.
Peck, H., & Lissolo, T. (1988). Assimilatory and dissimilatory sulphate reduction: enzymology and bioenergetics. *The nitrogen and sulphur cycles. Cambridge University Press, Cambridge*, 99-132.
Peck, H. D. (1959). The ATP-dependent reduction of sulfate with hydrogen in extracts of Desulfovibrio desulfuricans. *Proceedings of the National Academy of Sciences, 45*(5), 701-708.
Powers, D., & Olson, H. (1980). Determination of S–O bond order in sulfur dioxide and dimethyl sulfite using a low-energy particle-accelerator technique. *The Journal of Chemical Physics, 73*(5), 2271-2279.
Quentmeier, A., Kraft, R., Kostka, S., Klockenkämper, R., & Friedrich, C. G. (2000). Characterization of a new type of sulfite dehydrogenase from Paracoccus pantotrophus GB17. *Archives of microbiology, 173*(2), 117-125.
Resch, M., Göpel, Y., Görke, B., & Ficner, R. (2013). Crystallisation and preliminary X-ray diffraction analysis of YhbJ from Escherichia coli, a key protein involved in the GlmYZ sRNA regulatory cascade. *Acta Crystallographica Section F: Structural Biology and Crystallisation Communications, 69*(2), 109-114.
Schiff, J., & Fankhauser, H. (1981). Assimilatory sulfate reduction. In *Biology of inorganic nitrogen and sulfur* (pp. 153-168): Springer.
Schomburg, D., & Salzmann, M. (1990). Uroporphyrinogen-III synthase. In *Enzyme Handbook 1* (pp. 873-876): Springer.

Smith, K. W., & Stroupe, M. E. (2012). Mutational analysis of sulfite reductase hemoprotein reveals the mechanism for coordinated electron and proton transfer. *Biochemistry, 51*(49), 9857-9868.

Sousa, R. (1995). Use of glycerol, polyols and other protein structure stabilizing agents in protein crystallisation. *Acta Crystallographica Section D: Biological Crystallography, 51*(3), 271-277.

Studier, F. W. (2005). Protein production by auto-induction in high-density shaking cultures. *Protein expression and purification, 41*(1), 207-234.

Trakhanov, S., & Quiocho, F. A. (1995). Influence of divalent cations in protein crystallisation. *Protein Science, 4*(9), 1914-1919.

van de Vossenberg, J. L., Driessen, A. J., Zillig, W., & Konings, W. N. (1998). Bioenergetics and cytoplasmic membrane stability of the extremely acidophilic, thermophilic archaeon Picrophilus oshimae. *Extremophiles, 2*(2), 67-74.

Vega, J. M., & Garrett, R. (1975). Siroheme: a prosthetic group of the Neurospora crassa assimilatory nitrite reductase. *Journal of Biological Chemistry, 250*(20), 7980-7989.

Wachtershauser, G. (1998). The case for a hyperthermophilic, chemolithoautotrophic origin of life in an iron-sulfur world. *Thermophiles: the keys to molecular evolution and the origin of life*, 47-57.

Wagner, M., Roger, A. J., Flax, J. L., Brusseau, G. A., & Stahl, D. A. (1998). Phylogeny of dissimilatory sulfite reductases supports an early origin of sulfate respiration. *Journal of bacteriology, 180*(11), 2975-2982.

Willey, J., Sherwood, L., & Woolverton, C. (2013). *Prescott's microbiology*: McGraw-Hill Higher Education.