Sofia Soares Costa

Bachelor degree in Biochemistry

Structural studies on enzymes involved in microbial hydrogen sulfide metabolism

Dissertation presented to obtain the Master degree in Biochemistry for Health

Supervisor: Dr. José A. Brito, ITQB NOVA

Co-Supervisor: Dr. João B. Vicente, ITQB NOVA

Instituto de Tecnologia Química e Biológica António Xavier

ITQB NOVA

Oeiras
October 2021

Sofia Soares Costa

Bachelor degree in Biochemistry

Structural studies on enzymes involved in microbial hydrogen sulfide metabolism

Dissertation presented to obtain the Master degree in Biochemistry for Health

Supervisor: Dr. José A. Brito, ITQB NOVA

Co-Supervisor: Dr. João B. Vicente, ITQB NOVA

President of the Jury: Dr. Teresa Catarino

Jury Members:

Dr. Teresa Santos Silva
Dr. Ana Varela Coelho
Dr. José Artur Brito (Supervisor)

Instituto de Tecnologia Química e Biológica António Xavier

ITQB NOVA

Oeiras
October 2021

2021

Structural studies on enzymes involved in microbial hydrogen sulfide metabolism

Sofia Soares Costa

Structural studies on enzymes involved in microbial hydrogen sulfide metabolism

"*Nothing in life is to be feared, it is only to be understood.*

Now is the time to understand more, so that we may fear less."

– Dr. Marie Salomea Skłodowska Curie

(1867 – 1934)

Acknowledgments

First and foremost, I would like to thank my Supervisor Dr. José Artur Brito for the incredible past three years of mentorship and for the ones that are yet to come. This great adventure started in January 2019 when I stepped on MPX lab with little to no knowledge about X-ray crystallography. From the very beginning, you were thoughtful, patient, welcoming and a great teacher. Thank you for all the laughs, lessons, guidance and the best "almoçaradas". For the late nights in remote synchrotron X-ray data collection. For the early mornings in data processing, phasing and refinement… Actually, not really thankful for those… (just kidding, we are self-proclaimed workaholics). For all the hard work applying for several funding opportunities and for helping me in my scholarship applications, all of it, to keep me working in MPX lab beside you. I am also thankful for the silly jokes, the light and fun work environment, for the scares that surely kept me aware of my surroundings, for the massive support and encouragement and for the justified worries about my skateboarding progress. Thank you.

A special thanks to my Co-Supervisor Dr João Vicente. Thank you for helping me, managing your time and being available to answer my questions. I am thankful for all the guidance in this massive project.

I want to thank my PI Dr. Margarida Archer for giving me the incredible opportunity to work along with this fascinating, young and dynamic team. Three years ago, you accepted me into your lab and challenge me to do my very best and give X-ray crystallography a try. Now we celebrate together our successes, always with a delicious cake on the side.

A warm thank you to my lab colleagues and PhD students Diogo Athayde, José Rodrigues and Márcia Alves, and Dr. Federico Issoglio. A very special thanks to the newest PhD student Vanessa Almeida, who worked hard to secure the PhD fellowship and is my role model for perseverance.

I would like to express my profound gratitude to our collaborators from Indiana University Bloomington, Dr. David P. Giedroc and Dr. Brenna Walsh who were relentless for the past four years and gave us all the support and enlightenment we needed.

I would also like to thank all members of the Macromolecular Crystallography Unit, coordinated by Prof. Dr. Maria Arménia Carrondo, who arranged weekly insightful seminars that I had the pleasure to be part of.

A big hug and special thanks to my Master colleagues and friends that supported me in this journey.

To my best friend, Inês Didier, the first person I dedicate this work to. Always by my side on the good and bad days. The one that I confide in my troubles. The one that I do the craziest things with. The one who watches me when I fall and is there to help me get back on my feet. The one who taught me skateboarding and that life is more than work. My best person. Thank you dearly. I wish you the best of luck in the world.

To my great friends Yuri Pires, Rita, Alice and Ana Mota. The greatest friends one can ask and wish for.

I want to thank ITQB NOVA for providing me with this unique opportunity.

We acknowledge the European Synchrotron Radiation Facility (ESRF) for provision of beam time on ID30A-3 and ID23-2, and the Berlin Electron Storage Ring Society for Synchrotron Radiation (BESSY II) facility for provision of beam time on MX 14.1.

And finally, last but not least, a very special thanks to my parents Margarida Soares and António Costa, for the unconditional support and for teaching me to follow my dreams and pursue what makes me happy. I dedicate this work to them.

With a great journey comes great achievements.

And what a journey it was.

Thank you all.

RESUMO

O sulfureto de hidrogénio (H_2S) é uma molécula primitiva presente desde a atmosfera primordial da Terra. Organismos de todos os Domínios da Vida evoluíram desde cedo para usar esta molécula nas atividades fisiológicas. Sendo inicialmente reconhecido como uma potente toxina respiratória, estudos recentes revelaram o envolvimento do H_2S em processos como neurotransmissão, sinalização e inflamação. Assim, uma regulação estrita da concentração intracelular de H_2S é importante para qualquer organismo.

Staphylococcus aureus e *Enterococcus faecalis* são dois patogéneos importantes para humanos e possuem vias metabólicas regulatórias dos níveis de H_2S. Estas vias dependem da desintoxicação de H_2S por atividade enzimática, e uma vez que "estrutura é função", para compreender estas atividades é necessário conhecer a estrutura destas enzimas. Este trabalho foca-se em estudos estruturais por Cristalografia de Raios-X, uma técnica de biologia estrutural para a determinação da estrutura tridimensional de macromoléculas, aplicados a duas enzimas importantes na homeostase do H_2S nestes dois organismos: CstB de *S. aureus* e CoAPR de *E. faecalis*.

A enzima *Sa*CstB apresenta três atividades catalíticas: *persulfide dioxygenase*, *persulfide transferase* e *thiosulfate transferase*, e uma massa molecular teórica de ~52 kDa. Estudos recentes demonstraram o envolvimento de duas cisteínas, Cys-201 e Cys-408, e um centro férrico não-hémico na atividade da enzima. No âmbito desta Dissertação, determinámos as estruturas cristalográficas da *Sa*CstB nativa e das variantes C201S, C408S e C201S/C408S, a 2.69, 3.19, 2.4 e 2.2 Å de resolução, respetivamente.

A enzima *Ef*CoAPR apresenta uma massa molecular teórica de ~62 kDa e atividade de *coenzyme A persulfide reductase*. A enzima tem duas cisteínas catalíticas, Cys-42 e Cys-508, e dois cofatores, FAD e CoA. A estrutura tridimensional da *Ef*CoAPR foi determinada a 2.05 Å de resolução e revelou a existência de um túnel que conecta os dois resíduos catalíticos. A flexibilidade da CoA neste túnel é proposta como cataliticamente relevante.

Palavras-chave: sulfureto, metabolismo do enxofre, cristalografia de raios-X, estrutura tridimensional, *Sa*CstB, *Ef*CoAPR

ABSTRACT

Hydrogen sulfide (H_2S) is an ancient molecule present since Earth's primordial atmosphere. Organisms from all Domains of Life early evolved to use this molecule in their physiologic activities. H_2S was first recognized as a potent respiratory toxin, however, recent studies revealed its involvement in processes such as neurotransmission, signaling and inflammation. Therefore, a tight regulation of H_2S intracellular concentration is paramount in all organisms.

Staphylococcus aureus and *Enterococcus faecalis* are two major human pathogenic bacteria with regulatory metabolic pathways that control H_2S levels. Hence, these pathways rely on enzymes for H_2S-detoxification, and, since "structure is function", to understand these activities it is necessary to know these enzymes' structures. The present work focuses on structural studies through X-ray crystallography, a structural biology technique to determine the 3-dimensional structure of macromolecules, specifically of two important enzymes for H_2S homeostasis in these two organisms: the CstB from *S. aureus* and the CoAPR from *E. faecalis*.

*Sa*CstB enzyme has three catalytic activities: persulfide dioxygenase, persulfide transferase and thiosulfate transferase activities, and a theoretical molecular mass of ~52 kDa. Recent studies revealed the role of two conserved catalytic cysteines, Cys-201 and Cys-408, and a nonheme iron center in the enzymatic activity. In the present Dissertation, we solved the crystallographic structures of native *Sa*CstB and variants C201S, C408S and C201S/C408S to 2.69, 3.19, 2.4 and 2.2 Å resolution, respectively.

*Ef*CoAPR enzyme has a theoretical molecular mass of ~62 kDa and coenzyme A persulfide reductase activity. The enzyme has two catalytic cysteines Cys-42 and Cys-508 and two cofactors, FAD and CoA. The 3-dimensional structure of *Ef*CoAPR was determined to 2.05 Å resolution and revealed the presence of a tunnel connecting these two catalytic residues. CoA flexibility in this tunnel is proposed to be relevant for catalysis.

Key words: sulfide, sulfur metabolism, X-ray crystallography, 3-dimensional structure, *Sa*CstB, *Ef*CoAPR

List of Contents

Chapter I – Introduction

Chapter II – *Sa*CstB

List of Figures

List of Tables

List of Equations

List of Abbreviations

Å	Angström
ADP	Adenosine diphosphate
*Af*Npsr	NAD-dependent persulfide reductase from *Archaeoglobus fulgidus*
ATP	Adenosine Triphosphate
B. anthracis	*Bacillus anthracis*
*Ba*CoADR-RHD	Coenzyme A disulfide reductase-rhodanese fusion protein from *Bacillus anthracis*
BIS-TRIS	2-[bis(2-hydroxyethyl) amino]-2-(hydroxymethyl) propane-1,3-diol
°C	Degree Celsius
CA-MRSA	Community-associated methicillin-resistant *Staphylococcus aureus*
CBS	Cystathione-β-synthase
$CC_{1/2}$	Correlation coefficient $_{1/2}$
CCP4	Collaborative Computational Project number 4
CDR	Coenzyme A disulfide reductase
CDR-RHD	Coenzyme A disulfide reductase-rhodanese fusion protein
CN^-	Cyanide
CO	Carbon monoxide
CoA	Coenzyme A
cPDO-PT	coupled Persulfide dioxygenase-persulfide activity
Cryo-EM	Cryogenic electron microscopy
cst	copper-sensing operon repressor-like sulfur transferase
CysSS	Cysteine persulfide
Da	Dalton
DETAPAC	2-[bis[2-[bis(carboxymethyl) amino] ethyl] amino] acetic acid
DNA	Deoxyribonucleic acid
E. coli	*Escherichia coli*
E. faecalis	*Enterococcus faecalis*
EDTA	2-[2-[bis(carboxymethyl) amino] ethyl-(carboxymethyl) amino] acetic acid
*Ef*CoAPR	Coenzyme A persulfide reductase from *Enterococcus faecalis*
EPR	Electron paramagnetic resonance spectroscopy
Equiv	Equivalents

ESRF	European Synchrotron Radiation Facility
et al.	*et alii* (and other people)
eV	electronvolt
FAD	Flavin adenine dinucleotide
g	*g*-force
GSH	Reduced glutathione
GSSH	Glutathione persulfide
H_2O_2	Hydrogen peroxide
H_2S	Hydrogen sulfide
HA-MRSA	Healthcare-associated methicillin-resistant *Staphylococcus aureus*
HCN	Hydrogen cyanide
hETHE1	Human ethylmalonic encephalopathy protein 1
His-tag	Polyhistidine-tag
HPLC	High-performance liquid chromatography
HS^-	Hydrosulfide
HSO^{4-}	Hydrogen sulfate
IDO1	Indoleamine 2,3-Dioxygenase
IMAC	Immobilized metal affinity chromatography
K	Kelvin
LMW	Low molecular weight
M	Molar
MAD	Multiple anomalous scattering
MES	2-morpholin-4-yl ethanesulfonic acid
MIR	Multiple isomorphic replacement
MR	Molecular replacement
MRSA	Methicillin-resistant *Staphylococcus aureus*
MWCO	Molecular weight cut-off
NAD	Nicotinamide adenine dinucleotide
NCS	Non-crystallographic symmetry
NMDA	(2R)-2-(methylamino) butanedioic acid
NMR	Nuclear Magnetic Resonance
NO	Nitric oxide
$O_2^{\cdot}$	Superoxide radical
PAP	3'-Phosphate-adenosine-5'-diphosphate

PDB	Protein Data Bank
PDO	Persulfide dioxygenase
PEG	Polyethylene glycol
ppm	parts per million
PT	Persulfide transferase
RHD	Pseudorhodanese homology domain
Rhod	Rhodanese
RMSD	Root-mean-square deviation
RNA	Ribonucleic acid
ROS	Reactive oxygen species
rpm	revolutions per minute
R-SSH	Thiol persulfides
SAXS	Small-angle X-ray scattering
S. aureus	*Staphylococcus aureus*
S. loihica	*Shewanella loihica*
S^{2-}	Sulfide
$S_2O_3^{2-}$	Thiosulfate
*Sa*CstB	CstB from *Staphylococcus aureus*
SAD	Single anomalous scattering
SBD-F	Ammonium 7-fluoro-2,1,3-benzoxadiazole-4-sulfonate
SCN^-	Thiocyanate
SDS-PAGE	Sodium dodecyl sulfate–polyacrylamide gel electrophoresis
SIR	Single isomorphic replacement
*Sl*PV-4 Npsr	FAD-dependent pyridine nucleotide-disulfide oxidoreductase from *Shewanella loihica*
SO_3^{2-}	Sulfite
SO_4^{2-}	Sulfate
SQRs	Sulfide:quinone oxidoreductases
TCEP	3-[bis(2-carboxyethyl) phosphanyl] propanoic acid
TRIS	2-amino-2-(hydroxymethyl) propane-1,3-diol
TST	Thiosulfate transferase
USA	United States of America
UV	Ultraviolet
WT	Wild type

Amino acids

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

Chapter I

Introduction

1. Introduction

1.1. Hydrogen Sulfide – History and Origin

Hydrogen sulfide (H_2S) is a gaseous molecule at atmospheric temperature, very flammable and irritative [1]. With a vapour density of 1.19 $Kg \cdot m^{-3}$, H_2S is highly soluble in aqueous solution and with a biological membrane permeability coefficient comparable to those of molecular oxygen and carbon monoxide [2]. H_2S has pKa values of 7.04 and 11.96 [1] (Figure 1.1), easily permeating through biological membranes and accumulating inside the cells as hydrosulfide (HS^-) [2].

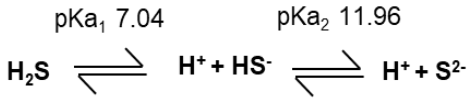


Figure 1.1 – Reaction scheme for the equilibrium of hydrogen sulfide (H_2S), hydrosulfide (HS^-) and sulfide (S^{2-}) in solution with the respective pKa values [1].

Some of the first documented references to H_2S are from the 3rd century, referred to as "Divine Water" or "Water of Sulphur" in the scriptures of the alchemist Zosimus of Panopolis [3]. In 1780, Carl Scheele published a book titled "Chemical Observations and Experiments on Air and Fire" [4] describing the discovery of H_2S, among other molecules, and later, in 1798, French chemist Claude Berthollet [5] discovered the molecular composition of the gas, naming it hydrogen sulfide.

H_2S is present in organisms from all three kingdoms of life. In some archaea and bacteria, H_2S enters the anoxygenic photosynthetic pathway as an electron donor. In mammals, and, particularly, in humans, it was first recognized as a potent respiratory toxin with no relevant physiological role identified for years [2]. Interestingly, the first study on the beneficial role of H_2S in health was reported in 1996 in mammalian neuronal tissues. This study revealed an unknown biosynthetic pathway where H_2S was produced endogenously in the brain through the activity of cystathione-β-synthase (CBS), one of the reasons for a higher concentration of H_2S in the brain tissues when compared to other tissues/organs [6]. Later on, in 2002, Wang[6] recognized H_2S as an important "gasotransmitter", pairing up with CO and NO. More recent studies identified fundamental functions of H_2S in neurotransmission [7], modulation of the immune system, anti-inflammatory response [8] and smooth muscles' tone [9], vasorelaxation [10], cellular signaling and sensing [11], stress response and homeostatic mechanisms [2].

H_2S is also present in the environment, as a side-product of petroleum, natural gas, coal and oil industries, paperboard factories, industrial sulfur production, volcanic activity and

bacterial metabolism. Concerning its environmental impacts, H_2S can cause detrimental effects on almost all organisms, depending on its concentration. The anthropogenic activity around H_2S exploration affects mostly the livestock, fish and plant species near industrial sites [3,12].

1.2. H_2S toxicologic effects

After carbon monoxide (CO), H_2S is the second most common cause of death by gas inhalation from occupancy exposure. Lethal human exposure is recurrent in oil and gas industries [3,13,14]. The exposure-response curve for H_2S reveals a small margin of safety, with fatal exposure more associated with H_2S concentration than exposure time (Figure 1.2) [15].

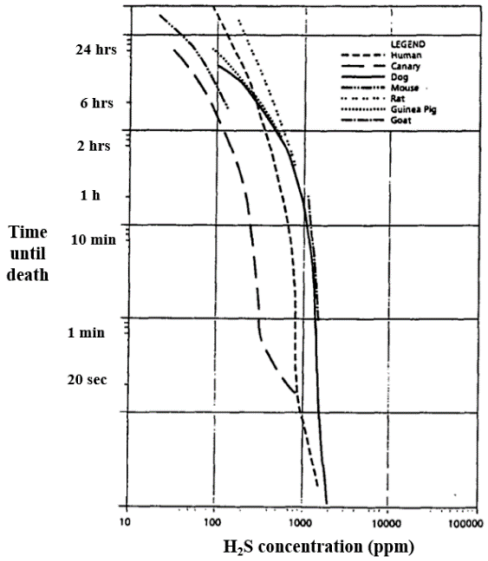


Figure 1.2 – Exposure-response curves of H_2S in humans, canaries, dogs, mouses, rats, guinea pigs and goats. Time until death as function of H_2S concentration (in ppm). Adapted from Guidotti, T. L. (1996) [15] and an unpublished report by Dr. Robert Rogers (1990) for the Alberta Energy Resources Conservation Board.

Contrary to most poisons, H_2S intoxication has a characteristic symptomatology [3,16], such as sudden loss of consciousness/"knockout", pulmonary edema, conjunctivitis, and olfactory dysfunction preceding complete olfactory paralysis. Victims of H_2S poisoning often report a sudden collapse, described as a "switch being turned off". This symptom is common for a 500 ppm H_2S concentration exposure and can be fatal within 4 to 8 hours of constant exposure as a consequence of respiratory paralysis and cellular anoxia [16].

H_2S also irritates mucous membranes in the lungs and eye epithelium, which combined with low solubility and high permeability, increases its prevalence in these tissues. Depression of the nervous system, olfactory paralysis and pulmonary edema are common symptoms of H_2S

exposure to concentrations higher than 250 ppm, while conjunctivitis and lung irritation are reported in cases of 20 ppm exposure. Superficial inflammation of the cornea with chromatic distortion and visual alterations are often reversible. Noteworthy, the characteristic rotten-egg like odor of H_2S can only be perceived to a maximum of 5 ppm, as the olfactory fatigue due to the intense smell dismisses it as a warning sign of H_2S toxic exposure. In severe cases, the loss of ability to perceive odor is irreversible. Non-specific secondary symptoms might include headache, short-term cognitive changes, seizures, nausea and vomiting. Chronic effects associated with H_2S exposure are still not clear [3,16]. Long-term effects of H_2S poisoning may include hyposmia (decreased sense of smell), dyssomnia (sleep disorders), phantosmia (sense of odors that are not present) and neurotoxic sequels [14,16].

As an example, we herein refer to a clinical case from 1987 [13], where a 24-year-old offshore oil worker exposed to a high level of H_2S (14,000 ppm) lead to severe and persistent neurological impairment. The case report revealed immediate severe dysfunction of the cognitive function, memory, visual perception, coordination, intelligence, corneal reflexes and reaction time. These injuries were consistent with cerebral cortex, brain stem, basal ganglia, hippocampus and diencephalic damage in similar case reports [13,17]. In 2020, the Occupational Safety and Health Administration agency of the United States Department of Labor reported three fatalities due to H_2S exposure [18]. Current treatment for H_2S poisoning includes hyperbaric oxygen therapy and nitrite administration, although evidence suggests that the latter is only effective if administrated immediately after exposure and a late administration may delay the treatment course [16]. A promising new study testing the efficacy of bicarbonate and glucose administration revealed a significant reduction in coma duration and mortality by 37 and 50 %, respectively, although these studies are still in early phase trials [16]. Teratogenic, carcinogenic, genotoxic and reproductive effects of H_2S remain unclear [3].

H_2S is rapidly eliminated from the organism through three main pathways: oxidation to sulfate (SO_4^{2-}), methylation and subsequent elimination through the urine, and reaction with metalloproteins or persulfidated proteins [3,19].

1.3. H_2S cytotoxicity

ATP synthesis is mediated through enzymatic complexes localized in the inner mitochondrial membrane through the flow of electrons down the electrochemical gradient. Inhibition of any of these complexes would disrupt the electron flow and energy production, increase NADH/NAD+ ratio (reductive stress), and lead to the formation of reactive oxygen species (ROS) [20].

The primary biochemical effect of H_2S exposure is cytochrome *c* oxidase inhibition, an enzyme of the electron transport chain in the mitochondria [19,21]. This inhibition causes the accumulation of superoxide radical ($O_2^{\bullet}$) due to the incomplete reduction of oxygen to water. A quick dismutase conversion of this radical to hydrogen peroxide (H_2O_2) is imperative to avoid cell damage [22]. Although cytochrome *c* oxidase activity is a sensitive biomarker for H_2S exposure in tissues as it is for hydrogen cyanide (HCN), the same treatment for HCN poisoning is not successful for H_2S intoxication [19,21,22]. Therefore, contrary to what was previously thought, a different mechanism for H_2S-induced cytotoxicity is speculated. Unlike HCN, H_2S is a reducing agent and, in combination with the reductive stress from NADH accumulation, can reduce Fe^{3+} to Fe^{2+} and release it from ferritin proteins and other metalloproteins, such as the above-mentioned cytochrome *c* oxidase. This reaction disrupts iron homeostasis, ATP production, increases the formation of ROS and affects numerous enzyme-dependent processes. Some ferric iron chelators, such as deferoxamine, can prevent cell death; however, treatment with extracellular iron-chelating agents, such as EDTA and DETAPAC, is not effective [22].

1.4. H_2S beneficial role

As previously stated, H_2S also plays important beneficial roles in health, assuming the organism's homeostasis is not perturbed. Several physiological mechanisms are, among others, associated with the action of H_2S in different tissues, modulating the immune response, anti-inflammatory response, smooth muscles' tone, vasorelaxation, neurotransmission, cellular signaling, and stress response [7-11].

In 2019, Nelp et al. [23] demonstrated that the redox reaction of H_2S with the ferric heme of human Indoleamine 2,3-Dioxygenase (IDO1) activates this enzyme, which is responsible for suppressing the immune response, mediating inflammatory and autoimmune reactions. The vascular effects of H_2S in synergy with nitric oxide (NO) were first proposed in 1997 [24] and demonstrated in 2001 by Zhao et al. [10]. In both *in vitro* and *in vivo* experiments with rats, H_2S-induced relaxation of blood vessels and consequent decrease in blood pressure was successfully shown. H_2S is currently considered the third gasotransmitter, after nitric oxide (NO) and carbon monoxide (CO) [25]. The effects of H_2S in neurotransmission were consistent with long-term induction of hippocampal response with the activation of N-methyl-D-aspartate (NMDA), synaptic transmission and protection of neurons against oxidative stress through modulation of gene expression. Anomalies in these functions are associated with the development of neurodegenerative diseases, such as Parkinson, Huntington and Alzheimer, hence, H_2S is proposed to act as a potent neuroprotector and neuromodulator [26]. H_2S can also function, in some

organisms, as a substrate for cysteine biosynthesis, energy source and electron donor to phototrophic and photosynthetic bacteria via H_2S deoxidation [27].

1.5. H_2S in bacteria

Taking into consideration all of the above-mentioned information, we can infer that the mechanisms behind H_2S detoxification in humans are complex, intricate, and mostly driven by enzymatic activity. On that account, bacterial systems are often used to study the enzymatic mechanisms behind environmental stress-responses, frequently in symbiosis with a host. Getting a better understanding of those enzymes would unravel the basic mechanics behind H_2S homeostasis.

Purple and green sulfur bacteria, and some members of purple nonsulfur bacteria, can oxidize H_2S to SO_4^{2-} through either photosynthesis under anaerobic conditions or chemosynthesis under aerobic conditions [28] (Figure 1.3).

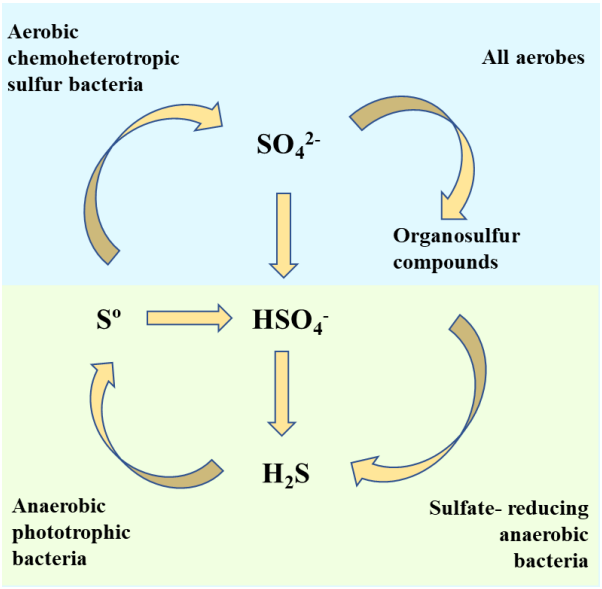


Figure 1.3 – Sulfur cycle evidencing hydrogen sulfide formation and degradation pathways for different organisms in anaerobic (**light green**) and aerobic (**light blue**) conditions. Molecules represented in the figure: hydrogen sulfide (H_2S), elemental sulfur (S^o), hydrogen sulfate (HSO_4^-) and sulfate (SO_4^{2-}). Adapted from Pfennig (1975) [28].

The simplest organisms catalysing complete sulfur oxidation and reduction are *Desulfovibrio*, *Chromatium* and *Chlorobium* bacteria. SO_4^{2-} is stable in aerobic and anaerobic conditions, while H_2S is only stable under anaerobic conditions due to quick auto-oxidation to SO_4^{2-} in the presence of oxygen. Dissimilatory sulfate reduction is achieved under anaerobic conditions in SO_4^{2-} reducing bacteria. Only phototrophic purple and green sulfur bacteria are

capable of oxidising H_2S to SO_4^{2-} to generate ATP and use H_2S as an electron donor in the pentose phosphate cycle [28,29]. In chemoheterotrophs, such as the nosocomial pathogen *Staphylococcus aureus*, in aerobic conditions H_2S oxidation is an important source of energy its cellular levels are tightly regulated through enzymatic activity. The difference in H_2S-tolerance levels is attributed to natural selection of these species according to their natural habitat [28,29].

Recent studies [30] reported that high concentrations of bacterial H_2S, e.g., in *S. aureus*, are associated to higher resistance against host leukocyte-mediated immune response, both *in vitro* and *in vivo*. Although, the mechanisms behind this bacterial resistance are unknown, it is theorized that H_2S can prevent oxidative damage by up-regulating superoxide dismutase and catalase enzymes [30,31]. Moreover, it is suggested that H_2S plays a role in protecting bacteria against the host's rapid innate response, hence, inhibition of bacterial H_2S may be an important antimicrobial therapeutic alternative for patients with inadequate immune functions [30]. Bacteria upregulation of H_2S/RSS ratio in response to β-lactam antibiotics might suggest underlying mechanisms of antibiotic resistance. Shatalin et al. found that Gram-negative and Gram-positive bacteria were more susceptible to a range of antibiotics after deletion or inhibition of enzymes that produce H_2S, suggesting that H_2S was important for antibiotic tolerance [31].

This work focuses on the structural elucidation by X-ray crystallography, of the enzymes: **CstB** from ***Staphylococcus aureus, Sa*CstB** (***Chapter II***), and **Coenzyme A persulfide reductase** from ***Enterococcus faecalis, Ef*CoAPR** (***Chapter III***). These enzymes play important roles in the regulation of H_2S levels in these human pathogens.

1.6. X-ray crystallography

Over the past decades, the number of protein structures determined by X-ray crystallography and deposited in the Protein Data Bank (PDB) has increased exponentially [32] since the first structure for myoglobin to 6 Å resolution published in 1958 [33]. While other techniques are starting to get their share of success, e.g., cryogenic electron microscopy (cryo-EM), X-ray crystallography is still the prevalent technique to determine the structure of molecules and macromolecules, such as proteins, DNA and RNA, with almost no limits in terms of complexity [32]. X-ray crystallography field continues to grant Nobel prizes and, currently, over 160,000 X-ray macromolecular structures are published in the PDB [34]. The development of powerful algorithms, computer programs and the increase of computational performance is intrinsic to the success of the methodology. These advancements were crucial to the development of modern scientific knowledge on biological functions and interaction of macromolecules.

Nonetheless, as previously stated, several other powerful techniques can be used to obtain three-dimensional information from macromolecules, namely NMR, SAXS and cryo-EM. Nuclear Magnetic Resonance (NMR) spectroscopy studies the response of atomic nuclei spins to external magnetic field variations induced by radiofrequency pulses. This technique has the great advantage to solve at an atomic level the structure of molecules and macromolecules in near-physiological conditions [35,36]. However, in NMR, the size of the molecule is a limitation, macromolecules with molecular weight above 35 kDa are difficult to solve, and almost unachievable without isotopic labeling [37]. Cryo-EM and SAXS have also the advantage to solve macromolecular structures in the native state. Cryo-EM is based on electron scattering signals that are amplified and converted to 2-dimensional images of the macromolecule. These images are organized in 2-dimensional classes depending on their orientation and merged to produce a 3-dimensional structure. This technique is most successful to study molecules of large size (e.g., whole viruses, fibrils and protein complexes) [35,36]. SAXS uses low-resolution information acquired from small-angle X-ray scattering (0.1-10º) to analyze the macromolecule shape, oligomeric state, molecular weight, protein-protein interaction, and protein folding, among others [36]. Due to its resolution limitation (no higher than 5 Å), it is often used as a first step to characterize the overall shape of a protein or complex, providing low(er) resolution information on the envelope of the sample under study [38].

1.6.1. History and crystal geometry

Three main events are considered the founding pillars of X-ray crystallography. Firstly, the discovery of X-rays in 1895 by Wilhelm Röntgen [39], secondly, the first theoretical proposal that X-rays wavelength was similar to atomic spacing, and, last but not less important, the first observation of X-ray diffraction in crystals by Max von Laue and colleagues [40,41]. Thus, the scientific ground was set by the Bragg's (William Henry, the father, and William Lawrence, the son), after mathematically explaining and introducing X-ray diffraction as a tool for crystal structure determination [41,42]. Since then, if the crystal is suitable, diffraction data has good quality, and if the "phase-problem" can be solved, it is possible to calculate electron density maps, build and refine a model of the macromolecule at atomic, or even sub-atomic, resolution. On the downside, one of the drawbacks of this technique is the crystal itself: one cannot do X-ray Crystallography without a crystal, and if the molecule does not form well-ordered 3-dimensional crystals, no structural model can be thereof derived.

In Crystallography, crystals are, defined as 3-dimensional arrangements of periodic repetitions of well-ordered molecules, stabilized mostly through non-covalent interactions between neighbor molecules and the solvent [43]. Most structural studies focus on macromolecules such as DNA, RNA and proteins due to their biological relevance. The main structural unit that repeats in space through translation vectors and reproduces the entire crystal lattice is denominated **unit cell**, with dimensions defined by six parameters, three axes (*a, b, c*) and three angles (α, β, γ). In the crystal lattice, equivalent planes (lattice planes) are unambiguously defined by integer numbers, denominated Miller indices (*h, k, l*), that uncover the number of intersections the lattice planes have with the crystal unit axes *a*, *b* and *c*, respectively [44]. Aside from translational symmetry, the crystallographic lattices can also present other symmetry correlations that define the crystal structure such as inversions, reflections, screw axes and symmetry axes [44,45].

The **asymmetric unit** is the smallest element that can reconstruct the unit cell through symmetry operations [46]. The number of molecules (or asymmetric units) in the unit cell depends on the crystallographic symmetry operators defining it. In the 3-dimensional space, 7 symmetry lattices, also called **crystal systems** are possible based on symmetry operations and unit cell geometry, as depicted in Table 1.1.

Table 1.1 – Crystal systems and Bravais Lattices [44,45].

Crystal system	Lattice symmetry	Conditions imposed on unit cell geometry
Triclinic	C_i	$a \neq b \neq c$ and $\alpha \neq \beta \neq \Upsilon \neq 90^o$
Monoclinic	C_{2h}	$a \neq b \neq c$ and $\alpha = \Upsilon = 90^o$, $\beta \neq 90^o$
Orthorhombic	D_{2h}	$a \neq b \neq c$ and $\alpha = \beta = \Upsilon = 90^o$
Tetragonal	D_{4h}	$a = b$ and $\alpha = \beta = \Upsilon = 90^o$
Trigonal	D_{3d}	$a = b$ and $\alpha = \beta = 90^o$, $\Upsilon = 120^o$; or $a = b = c$ and $\alpha = \beta = \Upsilon \neq 90^o$
Hexagonal	D_{6h}	$a = b$ and $\alpha = \beta = 90^o$, $\Upsilon = 120^o$
Cubic	O_h	$a = b = c$ and $\alpha = \beta = \Upsilon = 90^o$

In Crystallography, the symmetry of the lattice (**Laue class**) classifies the crystal system, and only a restricted set of symmetry point groups is allowed, denominated crystal classes, depending on the type of molecules in the crystal, e.g., a chiral molecule can only display 1 out of the 11 enantiomorphic point groups. Overall, only one combination of the 32 crystallographic point groups and 14 Bravais lattices is possible to define the space group of a specific crystal. Furthermore, it is also possible for two crystals of the same molecule to have different space groups and unit cell dimensions. Thus, only the structural description of the asymmetric unit and the crystal space group is required to construct the complete crystal structure [44].

X-rays are photons of energy above ultraviolet radiation and below gamma radiation on the electromagnetic spectrum, with an energy range from 0.12 to 120 keV and wavelength from 100 to 0.1 Å [44]. Waves are described with three parameters, wavelength, amplitude and phase. In crystallography, the most energetic X-rays are used due to their short wavelength (2 to 0.5 Å), near-atomic spacing and ability to penetrate the crystals [44]. Max von Laue observed in X-ray diffraction experiments that only when Bragg's law is obeyed and constructive interference occurs, an X-ray beam is scattered by the electrons of the molecules in the crystal and a diffraction pattern is recorded. The quality of this diffraction pattern ultimately defines the quality and resolution of the dataset. Following Bragg's law [42] (Equation 1 and Figure 1.4), d is the interplanar distance between parallel planes in the crystal lattice, θ is the reflection angle, λ is the beam wavelength and n is any positive integer [44,45].

$$2d \sin(\theta) = n \lambda \qquad \text{(Equation 1: Bragg's law)} \ [42]$$

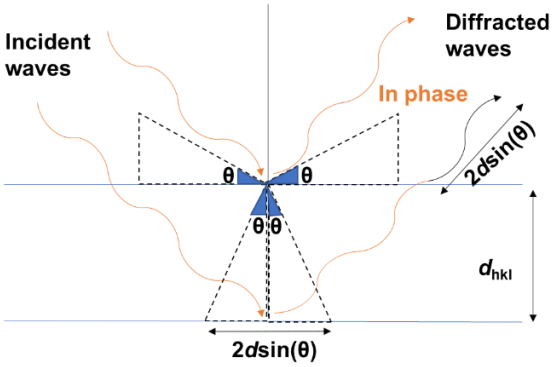


Figure 1.4 – Demonstration of Bragg's law. Constructive diffraction of two waves in phase (**orange**) by parallel lattice planes (**blue lines**) at distance *d*, with a path difference of 2*d*sin(θ). Adapted from Drenth, J. (2003) [45].

A 2-dimensional diffraction pattern is constructed in the reciprocal space, where a reciprocal lattice point (reflection) is unequivocally associated with a family of lattice planes in the real space (3-dimensional crystal) with specific Miller indices. Therefore, reflections measured at higher diffraction angles, represent shorter interplanar distances and will encode structural information of higher resolution [44,45].

In short, when the diffracted X-rays result from successive lattice planes with a path difference multiple of the wavelength, the amplitude of each diffracted wave is summed and produce a reflection of proportional intensity [45]. Considering that only the electrons in the molecules are responsible for diffraction, a diffraction pattern should reveal, indirectly, the distribution of these electrons, creating an electron density map of the atoms of all molecules of the crystal. The diffraction pattern only encodes information in the reciprocal space and needs to be correlated to the real space information to solve the position of each atom in the crystal structure. However, when the diffraction pattern is recorded, the phase information of the diffracted waves is lost, leading to the so-called "phase-problem", as the phase information is essential to calculate the electron density map. Currently, phase information can be retrieved using direct methods (by applying Sayre's equation), or indirect methods, such as molecular replacement (MR), single or multiple anomalous scattering (SAD/MAD), or single or multiple isomorphic replacement (SIR/MIR). After phasing through one of the above-mentioned methods, the electron density maps for the macromolecule are obtained. These maps are interpreted according to previous knowledge on that molecule, e.g., in the case of proteins, information on the protein sequence, amino acids' structure, stereochemistry, dihedral angles (Ramachandran plot), covalent and non-covalent interactions, steric effects, among others. After iterative cycles of model building and refinement, the information is converged to a final structure [44]. *See section 1.6.2.3 for more information.*

1.6.2. X-ray crystallization procedure pipeline

1.6.2.1. Crystallization experiment

In order to get crystals, one needs to give a "stimulus" to the molecules (in this Thesis: protein) in solution to come out of solution and go to a crystalline state. Unfortunately, the crystallization procedure is mostly empiric and iterative, therefore a complete assessment of all possible variables would be unachievable. However, empirical data gathered over the years made possible the construction of commercially available screening kits containing the most promising conditions to achieve protein crystallization [32,46]. Hence, an initial set of crystallization conditions, such as protein buffer, precipitating agents (precipitants), temperature, protein concentration, etc., are tested in a controlled environment to maximize the probability of success and the reproducibility of the method [44,46].

In the vapour-diffusion techniques, drops with protein and crystallization solution (precipitant) are set up on a closed system. All closed systems tend to the equilibrium and the water from the drop will evaporate until the precipitant concentration is equal to the reservoir. This process will cause a variation in both precipitant and protein concentrations that can be followed in a phase diagram (Figure 1.5) [47].

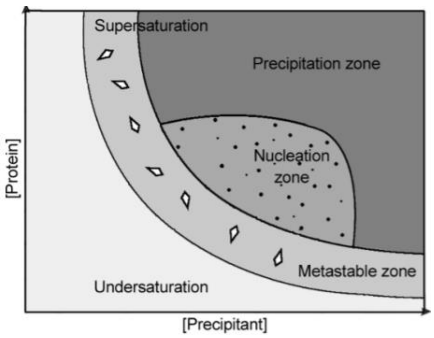


Figure 1.5 – Protein solubility phase diagram. Protein concentration as a function of precipitant concentration. Crystals are formed in the nucleation zone and grow under the metastable zone. In the unsaturated region, protein remains soluble, contrarily to the precipitation region, where the protein aggregates and precipitates. Figure from Pichlo, C. et al. (2016) [47].

Conditions are adjusted in each crystallization trial to slowly approach the supersaturated state in the phase diagram (Figure 1.5), specifically the nucleation zone, inducing crystal formation and subsequent growth in the metastable zone [47].

Keeping the complexity and concentration of the protein buffer to a minimum allows the precipitant solution to "dominate" the crystallization condition and assess what conditions are ideal to crystallize the macromolecule. A 10 $mg \cdot ml^{-1}$ protein concentration is usually used as a rule-of-thumb for initial crystallization trials, although this concentration can be adjusted to each

case. All protein samples must have high purity, homogeneity and folding for any crystallization experiment attempt to be successful [44,46].

Current crystallization methods may include dialysis, batch, counter-diffusion or, more commonly, vapour diffusion. In the latter, sitting-drops (Figure 1.6.A), where the crystallization drops are set on a support, or hanging-drops experiments, where the drops are suspended in a siliconized coverslip (Figure 1.6.B), are in contact with a reservoir of a larger volume of the crystallization solution [44,46,48].

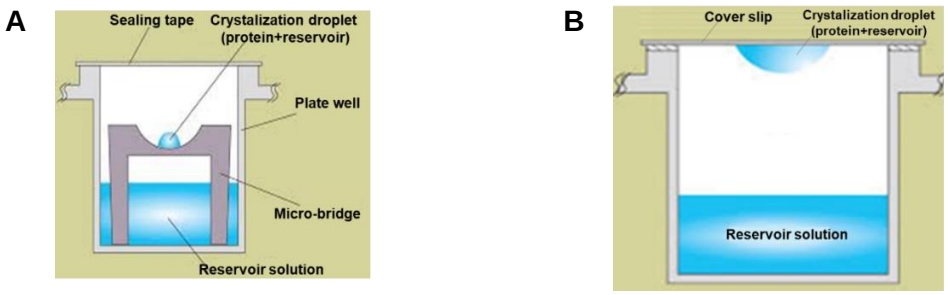


Figure 1.6 – Vapour-diffusion techniques: Sitting-drop (A) and hanging-drop (B). Adapted from McPherson, A. & Gavira, J. A (2013) [48].

As it is impossible to predict the rate at which crystals might form, therefore these should be monitored regularly over time. Once crystalline material is observed, the crystallization condition is often optimized to a larger scale, testing a range of precipitant concentrations and/or pHs to obtain suitable crystals of larger size and higher quality for data collection. If the protein does not crystallize after successive crystallization attempts, it is advised to rethink the expression and purification processes, the tag sequence and length, or test different buffer compositions, namely through Thermal Shift Analysis, Dynamic Light Scattering, or, Circular Dichroism [49], to help increase protein solubility, stability and sample homogeneity [32,46].

1.6.2.2. Data collection

Once a 3-dimensional crystal of the macromolecule is obtained, the crystal is mounted in a loop with a cryoprotectant solution, (e.g., crystallization condition supplemented with low-weight polyethylene glycols (PEGs), glycerol, sucrose, or other cryoprotectants), and placed in the goniometer of the diffractometer. Although the diffraction experiment can be performed at room temperature, usually the data acquisition is made at cryogenic temperatures with a steam of liquid nitrogen at 100 K (-173.15 °C) pointed to the crystal. The objective of the addition of the above-mentioned cryoprotectant solutions is to induce phase-transition from liquid to the solid-

state by vitrifying (rather than freezing) the sample. This avoids the formation of ice crystals that interfere with the diffraction pattern (presence of "ice-rings") and shadow diffraction data from the macromolecule. Most importantly, low-temperature diffraction measurements reduce crystal decay from radiation damage and formation of radicals, and so enable a longer time for data acquisition. [46]

In-house diffractometers, such as the one at ITQB NOVA (*X8 Proteum Diffractometer*, Bruker), generate an X-ray beam from the collision of electrons with an anode, focusing it on the crystal and recording the diffraction pattern in a detector. However, state-of-the-art synchrotrons are a better alternative to generate X-ray radiation [46], e.g., the European Synchrotron Radiation Facility (ESRF, Grenoble – France) [50]. A synchrotron is a cyclic particle accelerator that produces radiation in the electromagnetic spectrum from the acceleration of electrons [51]. Synchrotron X-rays are more intense, with a tunable wavelength, and enable faster data collections at higher resolution compared to in-house X-ray diffractometers [46].

1.6.2.3. Data processing

After collection, the data needs to be processed to yield interpretable data. As previously mentioned in *section 1.6.1*, the phases for all reflections need to be calculated to construct electron density maps.

Structure factors are calculated as the sum of all scattered waves by all the electrons in the unit cell with a given position, direction and phase (Equation 2) [44]. The reciprocal space lattice together with the structure factors is the Fourier transform of the electron density distribution in the crystal [44].

$$F(h,k,l) = \sum_j^N f_j \, e^{2\pi \cdot i\left(hx_{(j)}+ky_{(j)}+lz_{(j)}\right)} \qquad \text{(Equation 2: Structure factors)}$$

However, one has no information about the atom positions (*x*, *y*, *z*). Electron density is a periodic function that represents the atomic distribution of the atoms within the crystal structure and is defined in Equation 3, where *V* is the unit cell volume [44]. This equation is the inverse Fourier transform of the structure factors.

$$\rho(x,y,z) = \frac{1}{V}\sum_{hkl} |F_{hkl}| e^{-2\pi i(hx+ky+lz)+i\alpha(hkl)} \qquad \text{(Equation 3: Electron density function)}$$

However, the electron density function is dependent on the determination of the phase information for each reflection (α(*hkl*)). As previously stated, the phase information cannot be directly extracted from the diffraction experiment and only the structure factor amplitudes ($|F_{hkl}|$), derived from the square root of the reflections' intensities are known. Thus, computing the electron density map is only possible when the phases for all reflections are determined [44].

Currently, data is often processed with automatic pipelines, where the intensities for each reflection (spot) in the acquired images (diffraction patterns) are indexed, integrated, the crystal's space group is assigned based on statistical assumptions, scaled, converted to amplitudes, and the data is merged and converted to an MTZ file format. The overall statistics for the output data file reflect upon the quality of the selected processing strategy [46].

Data quality is accessed by a number of parameters, the most commonly used are **completeness**, **resolution**, $\boldsymbol{R_{meas}}$, $\mathbf{CC_{1/2}}$ and **signal-to-noise ratio [** I/σ(I)**]**. **Completeness** is one of the most important parameters to evaluate the quality of data collection. Completeness is defined in percentage as the total number of reflections measured over all theoretically possible reflections for that crystal [44]. A good quality data collection set has 100 % completeness, although smaller completeness (93 % overall completeness, and 70 % in the last resolution shell), are also accepted by specialized journals in the field (e.g., *Acta Cryst.* Journals [52]). **Resolution** by definition is the smallest distance between crystal lattice planes that is resolved in the diffraction pattern. The resolution of the diffraction pattern will directly define the degree of detail on the electron density maps, affect their interpretation and, consequently, limit the resolution of the final model [53]. This parameter directly evaluates the quality of the data and is used to decide if the quality is good enough to proceed with further data processing [34]. $\boldsymbol{R_{meas}}$ is an improvement of the previously used R_{merge}, as it is independent of data multiplicity, and indicates data consistency [54]. Nowadays, $\mathbf{CC_{1/2}}$ is the most used parameter to determine the high-resolution cut-off limit surpassing I/σ(I). I/σ(I) is the average reflection intensity against the estimated error of the measurements. As a rule-of-thumb, this parameter should have a value higher than 2 to differentiate real data from noise introduced from external sources (namely the detector's background) [34], however, since the introduction of $CC_{1/2}$ one finds datasets in the literature with I/σ(I) values as low as 0.8 [55–57].

Depending on the targeted macromolecule, the crystallographer needs to decide upon the **phasing method**. For example, if there is a protein with high homology to the targeted protein and its structure is already deposited in PDB, it might be useful to start by trying the molecular replacement method. In other cases, if the targeted protein already harbours a heavy-atom (e.g., an iron in a cytochrome's heme moiety), or if the crystal was previously soaked or co-crystallized before data collection with a heavy atom (e.g., Hg, Pt, Ag, Au, Pb, among others), or even if the

protein was produced as a selenomethionine variant, one might choose isomorphic replacement or anomalous diffraction methods to solve the "phase-problem". The latest method can only be employed if the X-ray beam wavelength is tunable. The direct methods are the least used for phasing and are only employed for high-resolution diffraction data of small(er) macromolecules [46]. A more detailed description about each method can be found below.

In the **molecular replacement method**, phases from a homologous structure deposited in the PDB are used as initial phases to calculate an initial electron density map for the molecule in study. As a rule-of-thumb, the amino acid sequences between the previously determined structure and the target structure should share sequence identity of at least 25 %, although it is not a guarantee that both structures have enough homology [44,46]. In short, and as previously stated, every molecule is defined in space by six parameters, three rotation angles (orientation) and three translations (position). This method, based on the Patterson function, employs a rotation function and a translation function to calculate these parameters and attain the coordinates for each atom. Once atomic positions (coordinates) are solved, the electron density maps are calculated (following Equation 3) [58]. Recently this year, AlphaFold emerged as a novel machine learning approach to predict protein structures with high accuracy based on their amino acid sequence. Therefore, these predicted models are alternatives to take into consideration when choosing a structure for molecular replacement [59].

In single and multiple wavelength anomalous diffraction, or isomorphic replacement, experiments, crystals with heavy atoms need to be produced through *in vitro* protein production, namely by selenomethionine derivatization, or supplemented by crystal soaking or co-crystallization. Mercury compounds that tend to covalently bind to cysteine residues and halides, e.g. bromide and iodide, which have propensity to nest in hydrophobic cavities in the macromolecule's surface, are also widely used [44,46,60]. **Isomorphic replacement methods** (**SIR/MIR**) compare intensity differences between equivalent reflections of native crystals and crystals with the heavy atom. These crystals need to have the same lattice parameters and the macromolecule structures should only differ in the presence or absence of the ordered heavy atom [44,46,61]. In **anomalous diffraction methods** (**SAD/MAD**) the X-ray beam wavelength is tuned to a value near the absorption edge of the resonant heavy atom, producing an anomalous dispersion signal. Crystals with the heavy atom will present a shift in the phase value for all atoms, which is more accentuated for those near the heavy atom. In the MAD method, different wavelengths are used to measure the crystal data [44,46,62].

In the case of **direct methods**, a very high-resolution dataset (1.2 Å or better) for small macromolecules (typically within 200 to 1,000 non-hydrogen atoms) is needed to calculate the

phases of all reflections. If the phases of some reflections are known or estimated, then the phases of the remaining reflections can be deduced [44,46].

After the "phase problem" is solved and electron density maps are calculated, the next step is building an initial model that fits the electron density maps. This model could be the one used for molecular replacement, or in case there is no homologous structure available, a new model needs to be built from scratch, based on the interpretation of the electron density information. Nowadays, automatic tools are available to facilitate this step by applying multi-step procedures that combine several possible solutions into a single (best) model. This best solution is then submitted to several cycles of model building, refinement and rebuilding, that alternate with model-based density modifications, to produce a complete atomic model [63]. Once, an initial model is attained, the phases are improved along with **iterative cycles** of crystallographic **refinement** and **model building.**

1.6.3. Model quality

The quality of fitting the atomic model into the observed diffraction data can be assessed through analysis of several overall statistical parameters, along with individual atomic information such as stereochemistry. These parameters need to be evaluated carefully to avoid misfitting, or overfitting, to the diffraction data, and obtain an incorrect structural model. The most commonly used parameters to evaluate the quality of the model are R_{work}, R_{free}, RMSD's (root-mean-square deviations), and overall geometric parameters of the model (e.g., Ramachandran plot, rotamer outliers, and clashscore). To some extent, overall B-factors (or ADP's – atomic displacement parameters), can also be used as evaluation parameters.

Traditionally, the parameter to evaluate the overall accuracy of a crystallographic model is the $\boldsymbol{R_{work}}$, also depicted as R, R-factor or R_{cryst} (Equation 4) [64]. F_{obs} and F_{calc} refer to the amplitude values measured and calculated from the model, respectively, for each reflection on the diffraction pattern. Most often, well-refined structures have a R_{work} lower than 0.2 (20 %), but this value is resolution dependent, as lower resolution structures can show higher R_{work} values [32].

$$R_{work} = \sum \frac{||F_{obs}| - |F_{calc}||}{|F_{obs}|} \qquad \text{(Equation 4: } R_{work}\text{)}$$

In 1992, Brünger [64] introduced the $\boldsymbol{R_{free}}$ as a bias-free parameter. With a role similar to R_{work}, i.e., to evaluate the overall quality of a crystallographic model, this reliable and unbiased parameter is calculated in the same way as R_{work}, but using a subset of the diffraction data (typically a random selection of 5 % of the total reflections), that is never used for model refinement. R_{work} and R_{free} values are expected to decrease proportionally in each refinement cycle for a well-interpreted density map. Noteworthy, a difference between R_{free} and R_{work} above 7 % can indicate overfitting of the calculated atomic model to the experimental data. Moreover, when this difference is below 2 %, the model might be biased and the reflections marked with "Free-R flags" were likely used within the model refinement steps. Hence, an analysis of both parameters is useful and indicative of the overall quality of the model [64].

The **B-factor**, or Atomic Displacement Parameter (measured in $Å^2$), describes the degree of vibration (motion) of an atom around its mean position in the model. This parameter reflects the degree of disorder of an atom compared to its neighbors. A high B-factor value translates to a high degree of uncertainty of an atom fit into the electron density maps. Disordered loops, or poorly occupied ligands/cofactors, usually have high B-factors [46,65,66].

The **RMSD** represents the difference between standardized (theoretical), values, and the model's overall geometrical parameters for bond lengths, angles, and chiral planes. Values between 0.015 and 0.025 Å, and below 2^o (preferably below 1.5^o), are generally accepted as good indication of model correctness for bond lengths and angles, respectively [46,65,66].

The present project aims to study proteins of unknown structure, whose functions are critical for the survival of human pathogenic bacteria. These proteins are the **CstB** from ***Staphylococcus aureus*** (***Sa*CstB**), and **Coenzyme A persulfide reductase** from ***Enterococcus faecalis*** (***Ef*CoAPR**), and will be discussed in *Chapters II* and *III*, respectively.

The mechanisms behind the function of these proteins inside the bacterial cells are, so far, not fully understood. A complete X-ray structure could help unravel this knowledge, as the protein structure is intimately related to its function. This research would culminate into developing new alternatives to fight antibiotic resistant bacteria by proposing putative novel drugs that, by targeting these enzymes, might hinder pathogenic bacteria propagation and reduce human morbidity and mortality rates worldwide.

The present Master Thesis aims to:

- Determine the complete crystallographic Apo structure of the wild type (WT) variant of CstB from *Staphylococcus aureus* (*Sa*CstB), along with structures in different redox (intermediate) states;

- Determine the 3D structures of *Sa*CstB single variants C201S and C408S, and double variant C201S/C408S, and compare these structures to the WT form of the enzyme;

- Determine the first complete crystallographic Holo structure of CoAPR from *Enterococcus faecalis* (*Ef*CoAPR) and structures in complex with substrate-like ligands.

Moreover, we propose to:

- Compare the complete structure of WT of *Sa*CstB to a previous incomplete model of this protein;

- Compare the WT of *Sa*CstB structure with similar proteins deposited in the PDB;

- Compare the WT of *Ef*CoAPR structure with similar proteins deposited in the PDB.

Chapter II

SaCstB

2. CstB from the human pathogenic bacterium *Staphylococcus aureus* (*Sa*CstB)

2.1. *Staphylococcus aureus*

Staphylococcus aureus is a facultative anaerobe gram-positive bacterium and the major cause of human and animal infections. In the early 1960s, the methicillin-resistant *Staphylococcus aureus* (MRSA) strain emerged with acquired resistance to a wide group of antibiotics and, since then, several worldwide MRSA outbreaks have been reported [67]. MRSA bacteria are sub-classified into two different groups, each with a specific set of antibiotic resistance characteristics, the healthcare-associated MRSA (HA-MRSA) and the community-associated MRSA (CA-MRSA). According to Premier Healthcare and Cerner Health Facts databases, in 2017, nearly 120,000 cases of MRSA bloodstream infections were reported, leading to nearly 20,000 deaths in the United States [68]. Common clinical infections associated with MRSA include bacteremia (presence of bacteria in the bloodstream), infective endocarditis (inflammation of the endocardium), osteoarticular, skin and soft tissue infections, and pleuropulmonary infections [69].

S. aureus clinical infections remain to the present date of high concern and the development of new methods to fight this major human pathogen are of paramount importance. One of the paths to hinder the propagation of this organism is the disruption of its metabolic homeostasis. As previously referred, H_2S is an essential molecule to most organisms and in *S. aureus* dysregulation of its metabolic pathways could cause serious deleterious effects. In 2011, Shatalin et al. [70] correlated for the first time endogenous H_2S concentration with antibiotic resistance in *S. aureus* and other bacteria.

2.2. *Sa*CstB

The first step in the H_2S catabolism involves a two-electron oxidation of H_2S to low molecular weight (LMW) thiol persulfides (R-SSH), through the activity of sulfide:quinone oxidoreductases (SQRs). These persulfide compounds are the substrate for enzymes with a persulfide dioxygenase (PDO) activity. Sulfurtransferases, also known as rhodaneses can work in parallel with PDO enzymes to transport these persulfide compounds and subject them to further oxidation to sulfite (SO_3^{2-}) [27,71] (Figure 2.1).

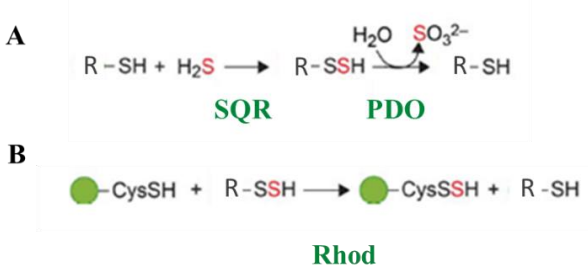


Figure 2.1 – First steps of hydrogen sulfide metabolism, evidencing SQR, PDO and Rhodanese (Rhod) activities. (Reaction A) H_2S catabolic reaction to SO_3^{2-} with SQR and PDO enzymatic activities. (Reaction B) Rhodanese (**green circle**) and thiol persulfides (R-SSH) in an autocatalyzed persulfidation reaction of the rhodanese enzyme. Adapted from Giedroc, D. P. (2017) [27].

In 2015, Shen et al. [71] reported the discovery of the operon *cst* (*copper-sensing operon repressor-like sulfur transferase*) in *S. aureus*, which encodes genes for multidomain sulfurtransferase CstA, persulfide dioxygenase CstB (EC 1.13.11.18 [72]) and SQR enzyme (Figure 2.2). Hence, this operon encodes a nearly complete H_2S oxidation system, activated in the presence of polysulfides and repressed upon the association with the reduced CstR repressor enzyme. This organization has been proposed to indicate a cooperative function of these enzymes in H_2S metabolism [27,71].

Figure 2.2 – *Cst* operon from *Staphylococcus aureus* encoding for CstA, CstB and SQR enzymes and inhibited by the presence of reduced CstR ($CstR_{red}$). Adapted from Giedroc, D. P. (2017) [27].

In the same study, Shen et al.[71] proposed an organization for CstB into 3 domains from amino acid sequence alignment studies of CstB and homologs of known crystal structures, namely *S. aureus* metallo-β-lactamase [PDB ID: 3R2U, Minasov, G. et al. (2011), *unpublished work*], *Alicyclobacillus acidocaldarius* β-lactamase-rhodanese fusion protein (PDB ID: 3TP9, Michalsha, K. et al. (2011), *unpublished work*), and human ethylmalonic encephalopathy protein 1 (hETHE1, PDB ID: 4CHL [73]). Based on this analysis, the three-domain organization proposed for CstB comprises an N-terminal nonheme iron (II) **persulfide dioxygenase (PDO) domain**, a **pseudorhodanese homology linker domain (RHD)** and a conserved C-terminal **rhodanese (Rhod) domain** (Figure 2.3).

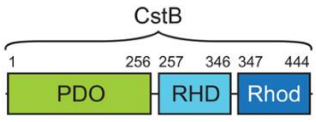


Figure 2.3 – *Sa*CstB protein sequence evidencing the three catalytic domains. The N-terminal persulfide dioxygenase (PDO) domain (**green**), a rhodanese homology domain (RHD) (**light-blue**), and a C-terminal rhodanese domain (Rhod) (**dark-blue**). Adapted from Shen J. et al. (2015) [71].

The protein structural arrangement, together with the biochemical and functional characterization of this enzyme, allowed three catalytic activities to be proposed for CstB: firstly, a **coupled persulfide dioxygenase-persulfide activity** (cPDO-PT); secondly, **a persulfide transferase activity** (PT); and thirdly, a **thiosulfate transferase (or rhodanese) activity** (TST). The first activity (cPDO-PT) is catalyzed by both PDO and Rhod domains and the latter two activities (PT and TST) are catalyzed by the Rhod domain, as schemed in Figure 2.4 [71].

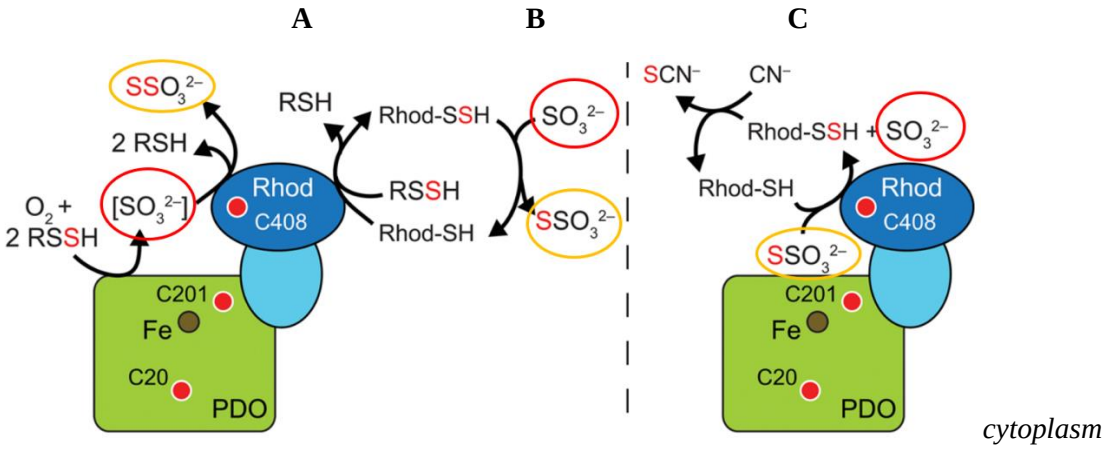


Figure 2.4 – Cartoon model of *S. aureus* CstB elucidating the enzymatic activities and the three-domain proposed structural organization. **Green**: N-terminal nonheme iron (II) persulfide dioxygenase (PDO); **Light-blue**: pseudorhodanese homology domain (RHD); **Dark-blue**: C-terminal rhodanese (Rhod) domain. (A) Coupled persulfide dioxygenase-persulfide activity (cPDO-PT). (B) Persulfide transferase activity (PT). (C) Thiosulfate transferase activity (TST). Adapted from Shen et al. 2015 [71].

CstB enzymatic activities are connected through a complex series of mechanisms that are activated accordingly to the cell's physiological necessities. According to Shen et al. [71], the PDO domain has one catalytic cysteine that might be important for structural organization, Cys-20, and one cysteine that is essential for catalysis, Cys-201, while the Rhod domain has one catalytic cysteine, Cys-408. A proposed model for the *Sa*CstB structure highlights the location of Cys-201 near the active site of the enzyme, namely within close range of the protein's nonheme iron center (Figure 2.4).

Illustrated in **Figure 2.4.A**, the persulfide compounds released from SQR activity (R-SSH) are used as substrates for the persulfide dioxygenase activity and converted into SO_3^{2-}, which, in turn, can be persulfidated to thiosulfate ($S_2O_3^{2-}$), in a **coupled persulfide dioxygenase-persulfide transferase activity** (cPDO-PT) mechanism. This mechanism requires the coordination of PDO and Rhod domains.

Interestingly, SO_3^{2-} can act both as substrate and final product of the rhodanese domain. Moreover, as depicted in **Figure 2.4.B**, R-SSH compounds can also be substrates for the persulfidation of the rhodanese catalytic cysteine (Cys-408). The persulfidated rhodanese cysteine and SO_3^2 can then be used as substrates for the **persulfide transferase activity** (PT), releasing $S_2O_3^{2-}$ as the final product. Alternatively, when Rhod domain functions with thiosulfate transferase activity, $S_2O_3^{2-}$ acts as a substrate to this domain to ultimately convert cyanide (CN^-) into thiocyanate (SCN^-), releasing SO_3^{2-} as a side product.

Recently, we have determined the X-ray structure of *Sa*CstB to 1.93 Å resolution [*unpublished data*]. The crystallographic model was "incomplete" as we could not interpret electron density for the rhodanese domain and the Cys-201 loop. (Figure 2.5). This work was carried out within my Final Project in Biochemistry at FCT NOVA, entitled "*Structural studies on enzymes involved in hydrogen sulfide detoxification*". Hence, for the present Master thesis, we have continued the studies to elucidate the 3D structure of full-length *Sa*CstB. The work was performed at Archer Lab, Membrane Protein Crystallography Group, ITQB NOVA. This project runs in collaboration with Professor Dr. David P. Giedroc, Indiana University – Bloomington, USA. We aim at the biochemical and structural characterization of enzymes involved in hydrogen sulfide homeostasis in pathogenic bacteria. This chapter depicts the work performed at Archer Lab to determine the crystal structures of non-heme Fe (II) persulfide dioxygenase-rhodanese fusion protein from *Staphylococcus aureus* (*Sa*CstB) [71].

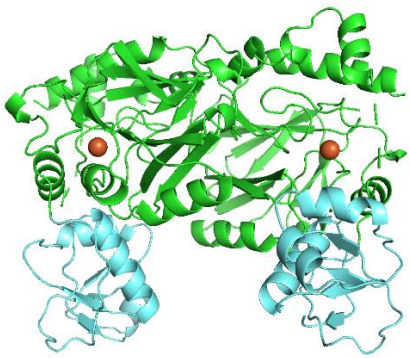


Figure 2.5 – Crystallographic structure of *Sa*CstB without the Rhod domain to 1.93 Å resolution ("CstB incomplete model"). Dimeric structure with persulfide dioxygenase (PDO) and pseudorhodanese homology domain (RHD). Phasing in *PHENIX* [74,75] by molecular replacement using metallo-β-lactamase from *S. aureus* (PDB ID: 3R2U, Minasov, G. et al. (2011), no follow-up publication) model as a template. Image created in *PyMOL* [76].

2.3. Materials and Methods

Gene expression and protein production were performed in the Giedroc Lab as previously described in Shen et al. (2015) [71], with slight modifications. In brief, to produce the *Sa*CstB recombinant protein, the *cstb* gene was fused with a 6-polyhistidine tag, overexpressed on a *pET15b* plasmid and transformed in *E. coli Rosetta*. *E. coli* cells were lysed, centrifugated and precipitated with polyethylenimine and ammonium sulfate. The proteins were purified by Ni-NTA affinity chromatography (IMAC), followed by a size exclusion chromatography (SEC column G200), using 25 mM MES pH 6.0, 50 mM NaCl, 5 % (V/V) glycerol and 2 mM TCEP buffer. *Sa*CstB amino acid sequence can be consulted in Appendix 1. *Sa*CstB single variants *Sa*CstB_C201S, *Sa*CstB_C408S and the double variant *Sa*CstB_C201S/C408S were produced from a mutated *cstb* gene following the same protocol as described above.

2.3.1. Sample preparation

A total of 2 ml of CstB at 9.5 $mg \cdot ml^{-1}$ in 25 mM MES at pH 6.0, 50 mM NaCl, 5 % (V/V) glycerol and 2 mM TCEP buffer, referred in this work as "*Sa*CstB with TCEP", were ultracentrifuged for 1 hour at 40,000 rpm (217,200 x *g*, *Ultracentrifuge optima TL-100, TLA-100.3 Fixed-Angle Rotor, Beckman Coulter*), in order to remove any precipitate or particules in solution. At this stage, three procedures were tested to remove TCEP from the protein sample.

The **first procedure** tested the removal of TCEP "in-column", consisting in the injection of 450 µl of protein at 9.5 $mg \cdot ml^{-1}$ in a *Superdex 200 (S200) Increase 10/300 GL* (*GE Healthcare Life Sciences*) column. Elution was performed at 0.4 $ml \cdot min^{-1}$ flow rate in the original buffer without TCEP [25 mM MES at pH 6.0, 50 mM NaCl and 5 % (V/V) glycerol].

The **second procedure** used sequential steps of concentration and dilution using the TCEP-free buffer in an *Amicon ultra-MILLIPORE 30 kDa* concentrator at 3500 x *g* (Centrifuge 5804 R, *Eppendorf*). Concentration/dilution steps were performed at against a total of 30 ml of protein buffer without TCEP. Along this step, an initial 300 µl of protein at 9.5 $mg \cdot ml^{-1}$ were concentrated to a final volume of 250 µl at ~9.7 $mg \cdot ml^{-1}$ (~1.02 concentration factor). Subsequently, the 250 µl of protein were injected in a *Superdex 200 (S200) Increase 10/300 GL* (*GE Healthcare Life Sciences*) column, following the same protocol as described above.

The **third procedure** consisted of an overnight dialysis using a *Micro Float-A-Lyser, Spectra/Por* (MWCO 3.5 – 5 kDa). In this step, 500 µl of protein at 9.5 $mg \cdot ml^{-1}$ were dialyzed against 1 L of TCEP-free protein buffer, followed by concentration using an *Amicon ultra-MILLIPORE 30 kDa* concentrator at 3500 x *g* (Centrifuge 5804 R, *Eppendorf*). The concentrated 450 µl of protein at ~9.5 $mg \cdot ml^{-1}$ were subsequently injected in a *Superdex 200 (S200) Increase*

10/300 GL (*GE Healthcare Life Sciences*) column, following the same protocol as described above.

*Sa*CstB protein variants

A total of 3 ml at ~2 $mg \cdot ml^{-1}$ of each variant sample in 25 mM Tris at 8.0 pH, 150 mM NaCl, 5 mM EDTA and 2 mM TCEP buffer were ultracentrifuged for 1 hour at 40,000 x rpm (217,200 x *g*, *Ultracentrifuge optima TL-100, TLA-100.3 Fixed-Angle Rotor, Beckman Coulter*). The protein buffer was exchange to 25 mM MES at pH 6.0, 50 mM NaCl and 5 % (V/V) glycerol through an overnight dialysis using a *Micro Float-A-Lyser, Spectra/Por* (MWCO 3.5 – 5 kDa). Prior to the polishing step in the S200 column, all samples were incubated with 0.7 equiv of $FeSO_4 \cdot DTT$ for 30 minutes. Protein samples were then concentrated to ~15 $mg \cdot ml^{-1}$ and 450 µl were injected in the *S200* column, using the same protocol as described above.

Protein fractions from the *S200* columns were pooled together, and labeled as "**Pool ++**" and "**Pool +**", accordingly to their distance to the center of the protein peak in the SEC chromatogram and concentrated as described in Table 2.1.

Table 2.1 – *Sa*CstB protein samples (volume and concentration) previous to crystallization.

***Sa*CstB**	**Pool ++**	**Pool +**
WT "with TCEP"	1.7 ml at 9.5 $mg \cdot ml^{-1}$	1.4 ml at 10.7 $mg \cdot ml^{-1}$
WT "without TCEP"	0.1 ml at 10.7 $mg \cdot ml^{-1}$	0.03 ml at 7.9 $mg \cdot ml^{-1}$
Variant **C201S** "without TCEP"	0.1 ml at 10.3 $mg \cdot ml^{-1}$	1.5 ml at 0.5 $mg \cdot ml^{-1}$
Variant **C408S** "without TCEP"	0.2 ml at 9.4 $mg \cdot ml^{-1}$	0.9 ml at 0.5 $mg \cdot ml^{-1}$
Variant **C201S/C408S** "without TCEP"	0.2 ml at 11.5 $mg \cdot ml^{-1}$	0.75 ml at 0.5 $mg \cdot ml^{-1}$

The purity of all samples was assessed with a precast SDS-PAGE gel (Bolt 4-12 % Bis-Tris Plus, *Invitrogen, Thermo Fisher Scientific*). UV-Vis spectra and protein concentration at 280 nm were determined using a *NanoDrop ONE Microvolume UV-Vis Spectrophotometer* (*Thermo Fisher Scientific*). A 48.82 $M^{-1} \cdot cm^{-1}$ molar absorptivity, 51.96 kDa molecular weight were implemented on the NanoDrop settings and the respective protein buffer was used as a blank solution for all protein quantifications.

2.3.2. Crystallization experiments

Accordingly to previous crystallization result hits for *Sa*CstB, optimized crystallization experiments were performed around condition F12 from the PACT Premier screen [77] [20 % (V/V) PEG 3350, 0.1 M Bis-Tris propane and 0.2 M sodium malonate dibasic monohydrate] using vapour diffusion sitting-drop method [48].

Protein samples from the **WT of "*Sa*CstB with TCEP"** and **WT of "*Sa*CstB without TCEP"** ("**Pool ++**") were tested against a gradient of 17 to 23 % (V/V) of PEG 3350, 0.1 M Bis-Tris propane buffer with a pH range from 6.1 to 7.0, and 0.2 M of sodium malonate, following vapour diffusion sitting-drop method [48]. Reservoirs from a *48-well sitting drop MRC Maxi* crystallization plate (*Swissci*) were filled with 184 µl of reservoir solution in *Dragonfly* (*SPT Labtech*), followed by mixing in a *MXone* (*SPT Labtech*) mixer. Drops of (500+500) nl of *Sa*CstB and reservoir solution were dispensed using *Mosquito LCP* (*TTP Labtech*) liquid dispenser robot, at 20 ºC and 70 % humidity. Handmade drops of (1+1) µl of *Sa*CstB and reservoir solution in gradient range as described above were prepared using vapour diffusion hanging-drop method [48]. Reservoirs from a *24-well Linbro* crystallization plate (*Molecular dimensions*) were filled with 500 µl of the reservoir solution in a *Dragonfly* (*SPT Labtech*), followed by manual mixing. In both methods, different protein concentrations ranging from 8 to 10 $mg{\cdot}ml^{-1}$ were tested.

_Sa_CstB protein variants

For the ***Sa*CstB variants,** initial crystallization screening experiments were carried out using "**Pool ++**" of *Sa*CstB C408S and C201S/C408S variants, both concentrated to 9.5 $mg{\cdot}ml^{-1}$ and the C201S variant to 8 $mg{\cdot}ml^{-1}$. Crystallization screenings were performed using PACT Premier commercial kit screen from *Molecular Dimensions* [77]. Reservoirs from a 96-well sitting drop iQ plate (*TTP Labtech*) were filled with 40 µl of each screening condition. Three drops per well with 100 nl of reservoir solution and 100 nl of each variant were dispensed with a Mosquito LCP (*TTP Labtech*) liquid dispenser robot, at 20 ºC and 70 % humidity, following the vapour diffusion sitting-drop method [48].

Selected crystallization conditions were optimized for the ***Sa*CstB variants** "**Pool ++**" samples. A gradient of 17 to 23 % (V/V) of PEG 3350, 0.1 M Bis-Tris propane buffer with a pH range from 6.1 to 7.0, and 0.2 M of sodium citrate, following vapour diffusion sitting-drop method [48]. Reservoirs from a *48-well sitting drop MRC Maxi* crystallization plate (*Swissci*) were filled with 184 µl of reservoir solution in *Dragonfly* (*SPT Labtech*), followed by mixing in a *MXone* (*SPT Labtech*) mixer. Drops of (500+500) nl of *Sa*CstB and reservoir solution were dispensed using *Mosquito LCP* (*TTP Labtech*) liquid dispenser robot, at 20 ºC and 70 % humidity. For the

***Sa*CstB single variant C201S**, **co-crystallizations** with 1 mM of $Na_2S_2O_3$, SO_3^{2-} and GSH (substrate-like ligands) were prepared in *48-well sitting drop MRC Maxi* crystallization plates (*Swissci*) filled with 184 µl of reservoir solution in *Dragonfly* (*SPT Labtech*), followed by mixing in a *MXone* (*SPT Labtech*) mixer. Drops of (480+20+500) nl of *Sa*CstB, substrate-like ligand and reservoir solutions, respectively, were dispensed using *Mosquito LCP* (*TTP Labtech*) liquid dispenser robot, respectively, at 20 ºC and 70 % humidity.

2.3.3. Data collection and processing

Apo *Sa*CstB wildtype and variant crystals were transferred to 1 µl drops containing 40 % (V/V) PEG 3350 cryo-protectant solution. Other *Sa*CstB wildtype and variant crystals were soaked in 1 µl drops of the same cryo-protectant solution, supplemented with 1 mM of substrate-like ligands (H_2S, SO_3^{2-}, Cys, CysSS, homo-Cys, S-methyl Cys, S-benzyl Cys, S-allyl Cys, GSH, GSSH, alliin, spermidine, spermine). Incubations were performed immediately before freezing (**quick soaking**), or for 5-minute period (**long soaking**). All crystals were mounted in *Dual Thickness MicroLoops LD™* long neck loops from *MiTeGen* (35 mm or 50 mm), flash-cooled in liquid nitrogen and sent to a synchrotron facility for data collection.

X-ray diffraction data collection on **WT of SaCtB** crystals was performed on microfocus ID23-2 beamline [78] at European Synchrotron Radiation Facility (ESRF, Grenoble – France) [50], with a fixed wavelength of 0.873 Å (14.2 keV), and recorded in *DECTRIS PILATUS3 X 2M* detector. Data for ***Sa*CstB C201S variant** X-ray diffraction data collection was performed on the microfocus ID30A-3 beamline [79] at European Synchrotron Radiation Facility (ESRF, Grenoble – France) [50], with a fixed wavelength of 0.968 Å (12.812 keV), and recorded in a *DECTRIS Eiger X 4M* detector. X-ray diffraction data collection of ***Sa*CstB C408S and C201S/C408S variants** crystals was performed on microfocus MX 14.1 beamline at Berlin Electron Storage Ring Society for Synchrotron Radiation (BESSY II, Berlin – Germany) [80], with a fixed wavelength of 0.918 Å (12.159 keV), and recorded with *DECTRIS PILATUS3 S 6M* detector.

Diffraction data were indexed and integrated using *XDS* [81], space-group assignment with *POINTLESS* [82], and scaling with *AIMLESS* [83] and *STARANISO* [84]. All programs were used within the *autoPROC* data-processing pipeline [85,75]. Final diffraction data were converted to MTZ format with *CTRUNCATE* [86–89], and at this stage, a set corresponding to 5 % of the total measured reflections was created and identified with Free-R flags [90,91].

2.3.4. Phasing, model building and refinement

Data quality was assessed with *phenix.xtriage* tool [92] within *PHENIX* suite of programs [74,75]. Phasing of experimental data for **WT of *Sa*CstB** was done by molecular replacement using *PHASER* [93] as implemented in *PHENIX* [74,75]. Two different search models were used: firstly, the previously determined "CstB incomplete model" (Figure 2.5), for phasing the persulfide dioxygenase (PDO) and pseudorhodanese homology (RHD) domains; secondly, a rhodanese (Rhod) domain model generated in the I-TASSER server [94,95,96,97], building a consensus model out of 20 rhodanese structures deposited in PDB. Following Matthews Coefficient analysis within *phenix.xtriage* [92], two molecules of each search model were searched for phasing the experimental data (asymmetric unit defined as a dimer of the full-length *Sa*CstB).

Iterative model building and refinement were carried out in a cyclic manner with *phenix.refine* [98] within the *PHENIX* [74,75] suite of programs, *BUSTER-TNT* [99,100], and *COOT* [101], until a complete model was built and refinement convergence achieved. *Sa*CstB models were validated with *MolProbity* [102] as implemented in *PHENIX* [74,75]. After initial molecular replacement phasing, phases for all subsequent *Sa*CstB isomorphic datasets were determined by an initial rigid-body refinement with *BUSTER-TNT* [99,100] using the corresponding Apo dataset of each *SaCstB* variant. The macro "Missing Atoms", together with the "-L" flag ("presence of an unknown ligand"), were used in *BUSTER-TNT* [99,100] to search for "unmodelled" electron density on the maps.

Structural illustrations were rendered with *PyMOL* [76] and *COOT* [101] programs.

2.4. Results and Discussion

Protein purification

Prior to all experiments performed in our Lab, protein purity and homogeneity of the samples received from the Giedroc Lab were assessed with SDS-PAGE (Figure 2.6).

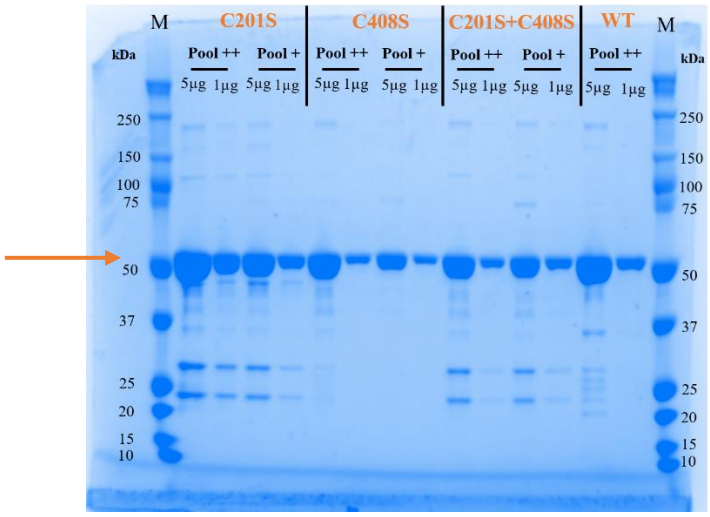


Figure 2.6 – 4-12 % precast Bis-Tris SDS-PAGE gel from *Novex NuPAGE*^{TM} of purified *Sa*CstB sample fractions. Gel ran at 80 V (with *Electrophoresis Power Supply - EPS 201* from *Amersham Biosciences*). (Pool ++) purest protein fractions and (Pool +) less pure protein samples. *Roti-Load* (1x) loading buffer from *Roth*. (M) Molecular weight marker *Precision Plus Dual Color Standards* from *Bio-RAD*. Protein buffer 25 mM MES at pH 6.0, 50 mM NaCl, 5 % (V/V) glycerol and 2 mM TCEP. *Sa*CstB protein bands (**orange arrow**).

In Figure 2.6 contaminants of lower and higher molecular weights than *Sa*CstB are observed in the samples; however, these are present at much lower concentrations than the target protein. Broad protein bands near 50 kDa (**orange arrow**) are indicative of the presence of CstB, which has a theoretical molecular mass of 51.96 kDa (ProtParam tool analysis from ExPASy server [103]). The majority of these contaminants will be removed in the following purification steps.

We have previously determined the crystallographic structure of *Sa*CstB to 1.93 Å [Figure 2.5 - *unpublished data*]. However, this model was incomplete rendering no diffraction electron density to the protein's third domain, the rhodanese (Rhod) domain. This was attributed either to the crystal packing of the molecules in the crystal, or, more plausible, to the flexibility between the linker loop between the domains 2 and 3 (RHD and Rhod domains, respectively). Hence, we proceeded to test and optimize various crystallization conditions, although unsuccessful to solve the problem in question. Herein, we address the challenge to find crystallization conditions that could stabilize or even restrain the flexibility of this third domain.

Our initial line of thought was that the catalytic cysteines Cys-201 and Cys-408 might need to establish a disulfide bond in order to perform their role in the reaction mechanisms. A putative bond between these two cysteines could restrict the rhodanese domain mobility and enable the acquisition of new diffraction data so that all domains could be resolved in the electron density maps. Therefore, we decided to test this new hypothesis and remove the reducing agent TCEP from the protein buffer to allow disulfide bond formation.

We tested **three different strategies to remove TCEP** from the protein buffer:

I – Removal of TCEP "in column" using a SEC-column running against the protein buffer devoid of TCEP (Figure 2.7.I and Appendix 2).

II – Removal of TCEP after several steps of protein concentration/dilutions followed by SEC; again, all proceedings against the protein buffer devoid of TCEP (Figure 2.7.II and Appendix 3).

III – Removal of TCEP with overnight dialysis followed by SEC; likewise, all proceedings against the protein buffer devoid of TCEP (Figure 2.7.III and Appendix 4).

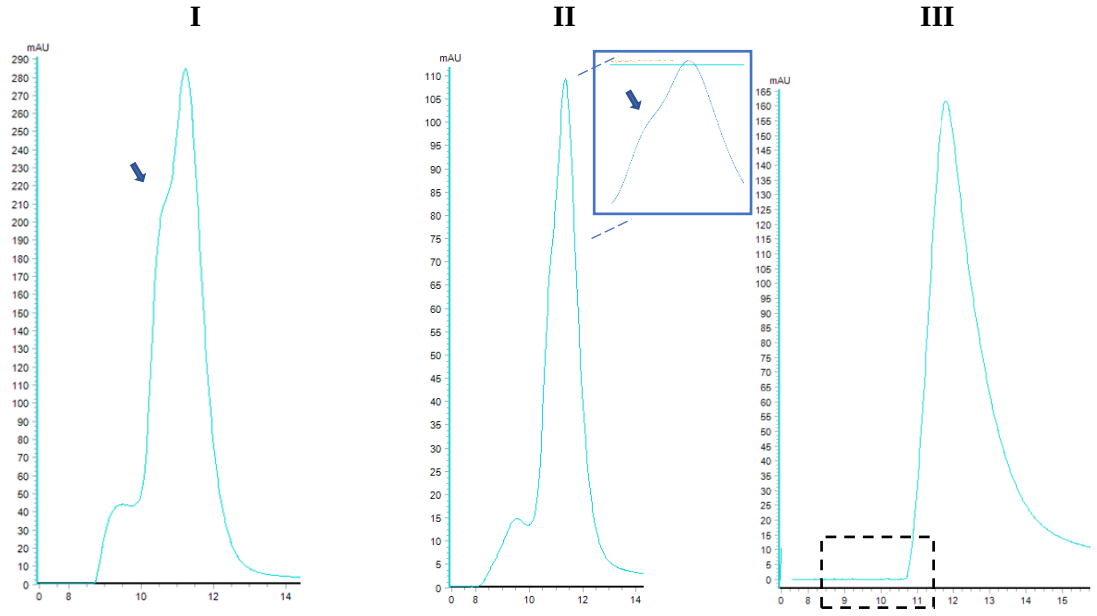


Figure 2.7 – Size-exclusion chromatogram obtained for *Sa*CstB purification after removal of TCEP "in column" purification (I); after successive steps of protein concentration and dilutions followed by SEC (II); and after overnight dialysis followed by SEC (III). The **light-blue** line indicates the protein absorbance registered on AKTA Pure connected to a *Superdex 200 Increase 10/300 GL* column (*Cytiva*). 450 µl of *Sa*CstB were injected and eluted with 25 mM MES pH 6.0, 50 mM NaCl and 5 % (V/V) glycerol buffer with a 0.4 $ml \cdot min^{-1}$ flow rate. 200 µl fractions were collected. The **blue arrow** highlights the presence of an additional peak in the chromatogram.

In Figure 2.7 it is possible to observe a clear shift between populations in solution. In Figures 2.7.I and 2.7.II, three peaks are present in the chromatograms, representing populations of distinct molecular weights, and, therefore being eluted from the column at different elution volumes (or elution time). Following SDS-PAGE analysis (Figure 2.6), we propose that the short broad peak of the first two chromatograms (Figures 2.7.I and 2.7.II) corresponds to contaminants in solution of higher molecular weight than *Sa*CstB, and the other two peaks correspond to *Sa*CstB in two different oligomeric states. Moreover, is possible to observe that overnight dialysis is the most effective method to remove TCEP and obtain homogenous *Sa*CstB samples. This is demonstrated by the presence of a single sharp protein peak in Figure 2.7.III, corresponding to a single protein oligomeric state. The size-exclusion chromatogram profile is identical for all *Sa*CstB variants and to Figure 2.7.III. Although the tetrameric state of the enzyme, as characterized by Shen et al. (2015) [71], is the proposed physiological state, we observe a shift from higher molecular weight species to lower molecular weights (elution volumes ~10 and 12 ml, respectively). According to a calibration curve with standard proteins for the used size-exclusion column (*data not shown*), this shift occurs from a tetrameric to a dimeric assembly in solution, as observed in the crystal asymmetric unit (*see below*).

SaCstB crystallization and structure determination

*Sa*CstB and its variants crystallized under optimized conditions from a range of 17–23 % (V/V) PEG 3350, 0.1 M Bis-Tris propane buffer at pH 6.1–7.0 and 0.2 M sodium malonate/ sodium citrate. WT and variant protein crystals appeared within 5 days with a regular, eight-sided, uncolored bipyramidal morphology and grew up to approximate maximum dimensions of ~0.2 x 0.2 x 0.4 mm^3 (Figure 2.8.A and 2.8.B). *Sa*CstB crystals belong to the tetragonal primitive space group $P\ 4_3\ 2_1\ 2$ with unit cell parameters a = b = ~150 Å, c =~125 Å and α = β = γ = 90 º (see Table 2.3 for details). Although *phenix.xtriage* tool [92-75] suggests the presence of three molecules in the asymmetric unit, searches for two molecules of each search model yielded high TFZ and LLG scores, with no further molecules being found after *PHASER* [93] search for an additional molecule of each search model. Therefore, the crystal contains two monomers in the asymmetric unit, corresponding to a Matthew's coefficient [104] (V_m) of 3.32 $Å^3 \cdot Da^{-1}$ and a solvent content of ~63 %.

Figures 2.8.A and 2.8.B are examples of typical *Sa*CstB crystals, and Figure 2.8.C displays an illustrative X-ray diffraction pattern for the WT of *Sa*CstB.

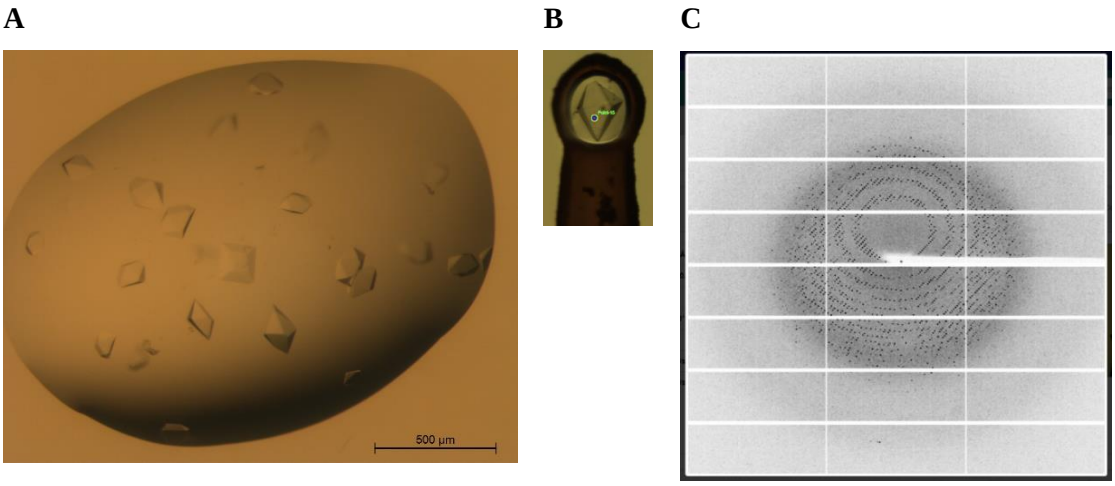


Figure 2.8 – (A) Drop with *Sa*CstB C201S/C408S crystals grown after four days in 23 % (V/V) PEG 3350, 0.1 M Bis-Tris propane buffer at pH 7.0 and 0.2 M sodium citrate. Photo acquired in *Leica MZ16 Stereomicroscope* (*Leica Microsystems*) with 5x ampliation. (B) WT of *Sa*CstB crystal grown in 18 % (V/V) PEG 3350, 0.1 M Bis-Tris propane buffer at pH 6.7 and 0.2 M sodium malonate, mounted in a 50 µm aperture *DT MicroLoop LD* (*Jena Bioscience*). Photo acquired in the *Basler GigE* camera at ESRF. (C) X-ray diffraction pattern of (B) crystal collected with *DECTRIS Eiger X 4M* detector of ID30A-3 beamline at ESRF, Grenoble, France.

Table 2.2 summarizes the WT Apo *Sa*CstB and variant crystals sent to synchrotron and are further discussed in the present Thesis.

Table 2.2 – WT of *Sa*CstB and variant *Sa*CstB crystals summary.

Crystal ID	Crystallization condition	Cryoprotectant condition
WT of *Sa*CstB Crystal IDs: **Apo *Sa*CstB**	**Drop volume**: 500 nl of protein + 500 nl of reservoir solution **Reservoir volume**: 184 µl **Method**: Vapour diffusion sitting-drop **Temperature**: 20 ºC **Crystallization solution**: **18 %** (V/V) PEG 3350, 0.1 M Bis-Tris propane buffer at pH 6.7 and 0.2 M sodium malonate	**40 %** (V/V) PEG 3350 0.1 M Bis-Tris propane buffer pH 6.7 0.2 M sodium malonate
C201S single variant *Sa*CstB **co-crystallized with $Na_2S_2O_3$** Crystal IDs: ***Sa*CstB C201S**	**Drop volume**: 480 nl of protein + 20 nl of $Na_2S_2O_3$ + 500 nl of reservoir solution **Reservoir volume**: 184 µl **Method**: Vapour diffusion sitting-drop **Temperature**: 20 ºC **Crystallization solution**: **23 %** (V/V) PEG 3350, 0.1 M Bis-Tris propane buffer at pH 6.7 and 0.2 M sodium citrate	**40 %** (V/V) PEG3350 0.1 M Bis.Tris propane buffer pH 6.7 0.2 M sodium citrate
C408S single variant *Sa*CstB **soaking with GSH** Crystal IDs: ***Sa*CstB C408S**	**Drop volume**: 500 nl of protein + 500 nl of reservoir solution **Reservoir volume**: 184 µl **Method**: Vapour diffusion sitting-drop **Temperature**: 20 ºC **Crystallization solution**: **21.8 %** (V/V) PEG 3350, 0.1 M Bis-Tris propane buffer at pH 7.0 and 0.2 M sodium citrate	**40 %** (V/V) PEG3350 0.1 M Bis.Tris propane buffer pH 7.0 0.2 M sodium citrate 1 mM **GSH**

C201S/C408S double variant *Sa*CstB **soaking with SO_3^{2-} and $Na_2S_2O_3$** Crystal IDs: ***Sa*CstB C201S/C408S**	**Drop volume**: 500 nl of protein + 500 nl of reservoir solution **Reservoir volume**: 184 µl **Method**: Vapour diffusion sitting-drop **Temperature**: 20 ºC **Crystallization solution**: **23 %** (V/V) PEG 3350, 0.1 M Bis-Tris propane buffer at pH 7.0 and 0.2 M sodium citrate	**40 %** (V/V) PEG3350 0.1 M Bis.Tris propane buffer pH 7.0 0.2 M sodium citrate 1 mM SO_3^{2-} 1 mM $Na_2S_2O_3$

Table 2.3 summarizes the results of X-ray data collection, processing, refinement statistics and model quality parameters for WT of *Sa*CstB (Apo) and *Sa*CstB variants. This table will suffer alterations before publication of these results.

Table 2.3 – Data collection processing, refinement statistics and model quality parameters for *S. aureus* CstB WT and variants C201S, C408S and C201S/C408S.

	***Sa*CstB WT**		***Sa*CstB C201S**	
Data Collection				
Synchrotron Facility	ESRF (Grenoble – France)		ESRF (Grenoble – France)	
Beamline Detector	ID23-2 DECTRIS PILATUS3 X 2M		ID30A-3 Eiger X 4M	
Wavelength (Å)	0.873		0.968	
Data Processing				
	autoPROC/ STARANISO	***autoPROC/ AIMLESS***	***autoPROC/ STARANISO***	***autoPROC/ AIMLESS***
Resolution range (Å)[a]	58.82-2.69 (2.83-2.69)	47.95-3.02 (3.07-3.02)	94.49-3.19 (3.40-3.19)	79.28 – 3.55 (3.61-3.55)
Crystal system	Tetragonal		Tetragonal	
Space group	*P* 4_3 2_1 2		*P* 4_3 2_1 2	
Unit cell parameters **a, b, c (Å)** **α, β, ϒ (º)**	148.93, 148.93, 125.36 90, 90, 90		145.70, 147.70, 124.14 90, 90, 90	

Total number of reflections	222748 (13134)	175930 (9102)	164918 (11843)	134903 (6850)
Total number of unique reflections	33300 (1666)	27177 (1354)	19171 (959)	16531 (795)
Multiplicity	6.7 (7.9)	6.5 (6.7)	8.6 (12.3)	8.2 (8.6)
Completeness (%)	-	95.7 (97.0)	-	98.8 (99.7)
Spherical	83.9 (30.1)	-	84.0 (25.1)	-
Ellipsoidal	91.3 (49.6)	-	93.5 (61.5)	-
Mean I/σ(I)	8.2 (1.3)	9.7 (2.2)	7.5 (1.4)	8.4 (2.2)
R_{merge} (%) [b]	17.1 (161.7)	14.4 (83.2)	23.8 (183.8)	19.6 (94.9)
R_{meas} (%) [c]	18.6 (173.1)	15.6 (90.0)	25.2 (191.4)	20.9 (100.5)
$R_{p.i.m}$ (%) [d]	7.0 (61.4)	5.9 (33.5)	8.0 (52.7)	6.8 (31.5)
$CC_{1/2}$ (%) [e]	99.5 (51.4)	99.5 (74.6)	99.4 (58.9)	99.4 (73.4)
Model Refinement				
	autoPROC/ STARANISO	***autoPROC/ AIMLESS***	***autoPROC/ STARANISO***	***autoPROC/ AIMLESS***
Total number of reflections used in refinement	27175 (2702)	-	19169 (193)	-
Total number of reflections used for R_{free}	1404 (146)	-	971 (11)	-
R_{work} (%) [f]	19.16 (28.32)	-	19.13 (30.07)	-
R_{free} (%) [g]	21.67 (28.87)	-	23.13 (29.25)	-
RMSD Bonds (Å)[h]	0.008	-	0.002	-
RMSD Angles (°) [h]	1.94	-	0.5	-
Number of atoms	-	-	-	-
Protein residues	885	-	889	-
Non-hydrogen atoms	6958	-	6985	-
Macromolecules	6932	-	6974	-
Ligands	9	-	10	-
Waters	17	-	1	-
Ramachandran plot	-	-	-	-
Most favoured (%)	95.50	-	95.23	-
Outliers (%)	0.30	-	0.00	-
Rotamer	0.40	-	1.31	-

outliers (%)				
Clashscore [i]	1.68	-	3.56	-
Molprobity score [j]	1.23	-	1.44	-
Average *B*-factors (Å²)	66.17	-	90.07	-
Protein	66.19	-	90.01	-
Ligands	77.50	-	130.54	-
Solvent	48.33	-	52.04	-

	***Sa*CstB C408S**		***Sa*CstB C201S/C408S**	
Data Collection				
Synchrotron Facility	BESSY II (Berlin – Germany)		BESSY II (Berlin – Germany)	
Beamline & Detector	MX 14.1 & *DECTRIS PILATUS3 S 6M*		MX 14.1 & *DECTRIS PILATUS3 S 6M*	
Wavelength (Å)	0.918		0.918	
Data Processing				
	autoPROC/ STARANISO	***autoPROC/ AIMLESS***	***autoPROC/ STARANISO***	***autoPROC/ AIMLESS***
Resolution range (Å)[a]	104-38-2.44 (2.55-2.44)	104.38-2.70 (2.75-2.70)	52.48 – 2.20 (2.28 – 2.20)	95.78 – 2.20 (2.23-2.20)
Crystal system	Tetragonal		Tetragonal	
Space group	$P\ 4_3\ 2_1\ 2$		$P\ 4_3\ 2_1\ 2$	
Unit cell parameters **a, b, c (Å)** **α, β, ϒ (º)**	147.62, 147.62, 124.82 90, 90, 90		148.42, 148.42, 125.38 90, 90, 90	
Total number of reflections	559614 (27794)	463966 (25788)	820748 (41228)	948723 (48666)
Total number of unique reflections	43915 (2196)	36530 (1886)	62327 (3117)	71690 (3538)
Multiplicity	12.7 (12.7)	12.7 (13.7)	13.2 (13.2)	13.2 (13.8)
Completeness (%)	-	95.0 (100.0)	-	100 (100)
Spherical	85.2 (36.5)	-	91.1 (36.3)	-
Ellipsoidal	88.5 (51.4)	-	96.2 (60.8)	-
Mean I/σ(I)	13.9 (1.5)	16.4 (2.5)	17.6 (1.4)	15.4 (0.8)
R_{merge} (%) [b]	15.1 (181.6)	13.0 (113.5)	12.2 (198.4)	14.2 (338.8)

R_{meas} (%) [c]	15.8 (189.2)	13.6 (117.9)	12.7 (206.4)	14.8 (351.8)
$R_{p.i.m}$ (%) [d]	4.4 (52.8)	3.8 (31.7)	35.5 (56.5)	4.0 (94.4)
$CC_{1/2}$ (%) [e]	99.9 (63.4)	99.9 (80.6)	99.9 (54.3)	99.9 (34.6)
Model Refinement				
	autoPROC/ STARANISO	***autoPROC/ AIMLESS***	***autoPROC/ STARANISO***	***autoPROC/ AIMLESS***
Total number of reflections used in refinement	43911 (1525)	-	71262 (6964)	-
Total number of reflections used for R_{free}	2267 (71)	-	3587 (336)	-
R_{work} (%) [f]	18.04 (27.23)	-	17.46 (30.8)	-
R_{free} (%) [g]	22.46 (34.24)	-	20.44 (35.99)	-
RMSD Bonds (Å)[h]	0.013	-	0.012	-
RMSD Angles (°) [h]	1.158	-	1.092	-
Number of atoms	-	-	-	-
Protein residues	898	-	897	-
Non-hydrogen atoms	7133	-	7281	-
Macromolecules	7033	-	7022	-
Ligands	10	-	10	-
Waters	90	-	249	-
Ramachandran plot	-	-	-	-
Most favoured (%)	95.53	-	96.00	-
Outliers (%)	0.00	-	0.00	-
Rotamer outliers (%)	0.52	-	0.00	-
Clashscore [i]	3.74	-	1.23	-
Molprobity score [j]	1.34	-	1.12	-
Average *B*-factors ($Å^2$)	52.97	-	57.69	-
Protein	53.09	-	57.96	-
Ligands	72.85	-	68.99	-
Solvent	41.36	-	49.59	-

[a] Information in parenthesis refers to the last resolution shell.

[b] $R_{merge} = \sum_{hkl} \sum_i |I_i(hkl) - \overline{I(hkl)}| \,/\, \sum_{hkl} \sum_i I_i(hkl)$.

[c] $R_{meas} = \sum_{hkl} [N/(N-1)]^{\frac{1}{2}} \sum_i |I_i(hkl) - \overline{I(hkl)}| \,/\, \sum_{hkl} \sum_i I_i(hkl)$.

[d] $R_{p.i.m} = \sum_{hkl} [1/(N-1)]^{\frac{1}{2}} \sum_i |I_i(hkl) - \overline{I(hkl)}| \,/\, \sum_{hkl} \sum_i I_i(hkl)$.

[e] $CC_{1/2}$ as described in Karplus & Diederichs (2012). Science, 336(6084): 1030–1033.

[f] $R_{work} = \sum_h \sum_k \sum_l \frac{\{||F_o(h,k,l)| - |F_c(h,k,l)||\}}{\sum_h \sum_k \sum_l |F_o(h,k,l)|}$, where F_o and F_c are the observed and calculated structure factors for reflection *h*, respectively.

[g] R_{free} was calculated the same way as R_{work} but using only 5% of the reflections which were selected randomly and omitted from refinement.

[h] RMSD, root mean square deviation.

[i] Clashscore is the number of unfavourable all-atom steric overlaps ≥ 0.4Å per 1000 atoms. Word *et al*. (1999). Mol Biol, 285(4):1711-33.

[j] *MolProbity* score provides a single number that represents the central *MolProbity* protein quality statistics; it is a log-weighted combination of clashscore, Ramachandran not favoured and bad side-chain rotamers, giving one number that reflects the crystallographic resolution at which those values would be expected.

For the sake of clarity, the following discussion is focused on WT of *Sa*CstB Apo model unless otherwise stated. *Sa*CstB variant structures will be discussed whenever there are differences from the WT structure worth mentioning.

The full-length WT *Sa*CstB X-ray model was refined to a final resolution of 2.69 Å with anisotropic diffraction limits of 2.69 Å, 2.69 Å and 2.97 Å and final R_{work} of 19.16 % and R_{free} of 21.17 % (*STARANISO* [84]). The final *Sa*CstB model comprises 885 amino acid residues from Gly-1 to Leu-442 in both chains (out of 444), 1 amino acid residue from the His-tag in chain A, 2 iron ions, 17 waters and 2 malonate molecules.

The electron density maps are clearly defined, apart from the disordered loop Asp-156 to Gly-166, which shows higher thermal motion parameters (B-factors) compared to neighbor residues. Interestingly, Ala-197 to Gly-206 loop is disordered only for the *Sa*CstB single variant C201S and double variant C201S/C408S. Hence, we propose that this disorder may affect enzymatic activity, as, in accordance to Shen et al. [71], a site-specific mutation of Cys-201 decreases the activity of these variants of *Sa*CstB, compared to the wild-type (WT) variant. No interpretable density was observed for the remaining amino acid residues of the His-tag, suggesting, as usual and expected, high flexibility around the N-terminal region.

The Ramachandran plot, as assessed by *MolProbity* [102] within *PHENIX* [74,75] shows that all nonglycine amino acid residues lie within allowed regions, except for Lys-203 (chain A). Furthermore, only 0.4 % of residues are rotamer outliers. Together with a clashscore of 1.68, this model displays a *MolProbity* score of 1.23 well below the resolution observed for this structure.

SaCstB overall fold and similar structures

According to PISA (Protein Interfaces Surfaces and Assemblies) program [105,106], the dimer interface has a calculated dissociation energy (ΔG^{diss}) of -17.4 kcal · mol^{-1} and an approximate area of 1100 $Å^2$, corresponding to ~5.8 % of the total solvent-accessible area of each monomer. The dimer interface is mainly hydrophilic and stabilized by 11 H-bonds and 2 salt bridges, with most of the residues contributing to this stabilization located in loop regions. Overall, no disulfide bonds were established between Cys-20, Cys-201 and Cys-408, nor these residues were found in the persulfidated state.

Both *Sa*CstB chains are very similar showing an RMSD of 0.32 Å for 445 aligned C_α atoms (chains A and B superposition performed with "Secondary Structure Matching" tool, within *COOT* [101,107]). As proposed by Shen et al. [71] each monomer has three domains (Figure 2.9A): a core N-terminal nonheme Iron (II) persulfide dioxygenase (PDO) domain, a pseudorhodanese homology domain (RHD) linking domains 1 and 3, and a conserved C-terminal rhodanese (Rhod) domain. Secondary structure analysis [108] revealed that each *Sa*CstB monomer consists of eleven α-helices, four-stranded parallel β-sheets and five-stranded anti-parallel β-sheets. Figure 2.9.A displays the complete crystallographic dimeric structure for *Sa*CstB, while Figure 2.9.B recreates the physiological tetramer assembly, generated by symmetry operations within the crystal unit cell.

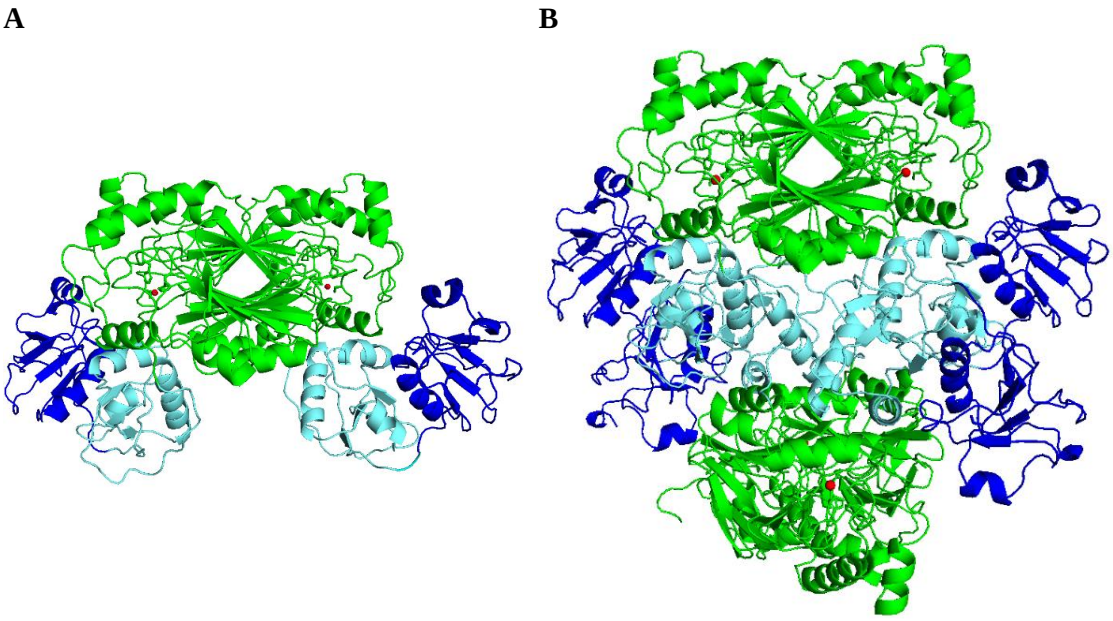


Figure 2.9 – (A) Cartoon representation of *Sa*CstB full-length crystallographic model to 2.69 Å resolution (CstB "as-isolated"). (B) Cartoon representation *Sa*CstB full-length tetrameric model generated by symmetry operations. Persulfide dioxygenase (PDO) domain (Green). Pseudorhodanese homology domain (RHD) (Cyan). Rhodanese (Rhod) domain (Dark-blue). Iron ion (red sphere) in the catalytic center of the PDO domain. Image created in *PyMOL* [76].

Although the closest protein analogue to *Sa*CstB, the metallo-β-lactamase from *S. aureus* (PDB ID: 3R2U) forms a homotetramer in the crystal asymmetric unit and *Sa*CstB biochemical characterization also indicates a tetrameric organization [71], we herein observed that removing TCEP from solution drives *Sa*CstB towards a dimeric assembly both in solution and crystal asymmetric unit. Nevertheless, it is easily observed how the tetrameric assembly can occur in solution, since the crystal packing allows this assembly, and it is generated by simple symmetry operations (Figure 2.9.B). Interestingly, both dimeric and tetrameric symmetry-generated assembly structures show the rhodanese domains standing out from the core of the protein. This grants an additional degree of flexibility to this domain, which we propose to be crucial for the interplay between the PDO and Rhod domain activities.

A search for the most similar structures to *Sa*CstB in the PDB highlighted the **metallo-β-lactamase from *S. aureus*** (***Sa*MβL**, PDB ID: **3R2U**, Minasov, G. et al. (2011), *no follow-up publication*), the **β-lactamase-rhodanese fusion protein from *Alicyclobacillus acidocaldarius*** [***Aa*βL-Rhod**, PDB ID: **3TP9**, Michalsha, K. *et al.* (2011), *no follow-up publication*] and the **human ethylmalonic encephalopathy protein 1** (**hETHE1**, PDB ID: 4CHL [73]. Secondary-structure matching (SSM) analysis between *Sa*CstB and these enzymes showed RMSDs of ~0.76 Å, 1.86 Å and 1.54 Å for 337, 408 and 211 aligned C_α atoms, and 74.2, 43.4 and 26.1 % of shared sequence identity, respectively. According to the Dali online server [109], most proteins within the first 50 hits belong to the metallo-β-lactamase superfamily, being the ***Sa*MβL**, ***Aa*βL-Rhod**, and

the persulfide dioxygenase-rhodanese fusion protein from *Burkholderia phytofirmans* [Motl, N. et al. (2017), PDB ID: 5VE3 [110]], representatives of this superfamily with Z-scores of 50.3, 48.2 and 30.1, respectively.

Figures 2.10.A and 2.10.B represent the structural superposition between the *Sa*CstB full-length structure and the previous *Sa*CstB incomplete model (Figure 2.5) for the dimer and monomeric structures, respectively. Figures 2.10.C, D and E illustrate the alignment between *Sa*CstB and *Sa*MβL, hETHE1 [73] and *Aa*βL-Rhod proteins, respectively.

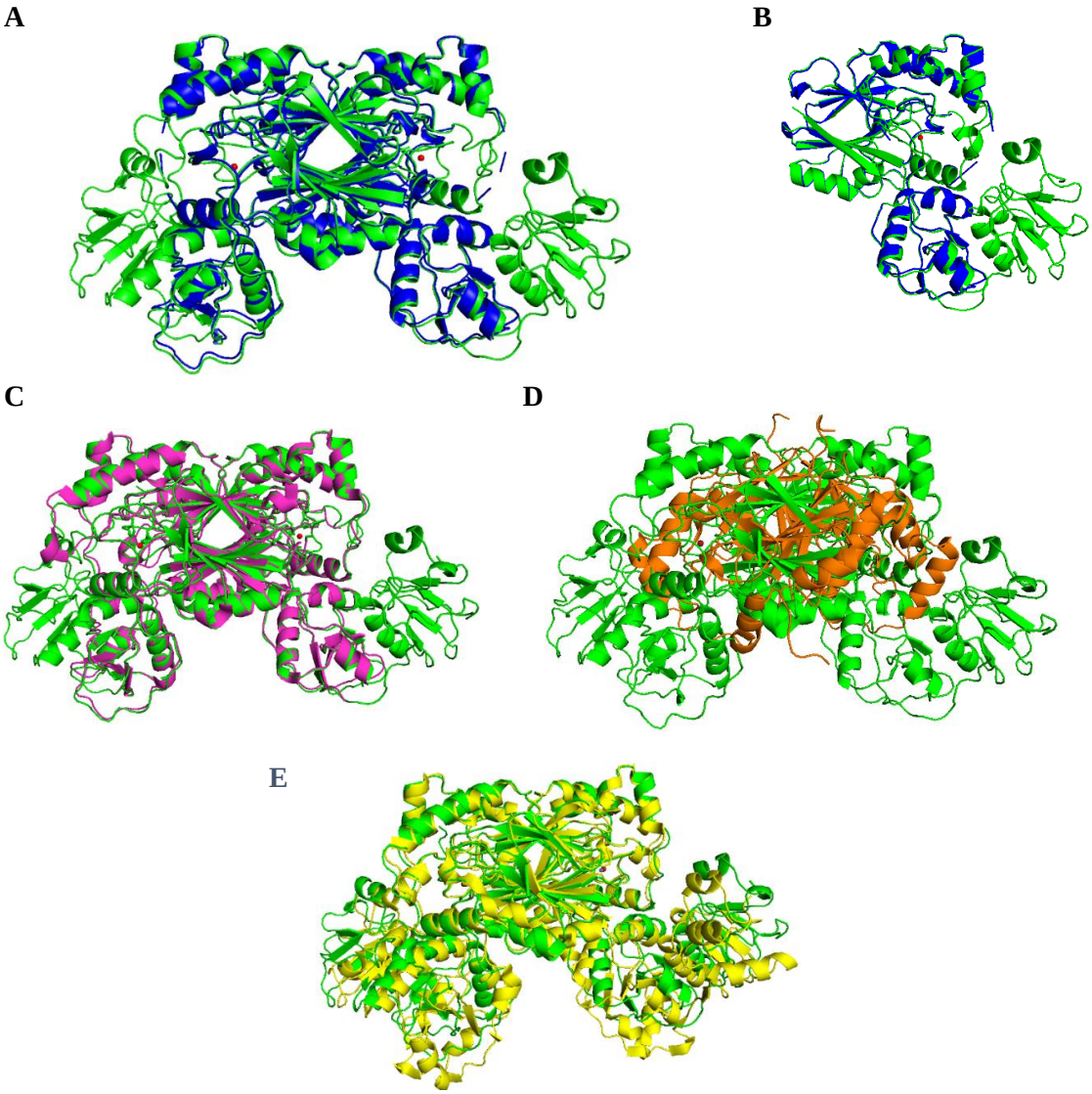


Figure 2.10 – Cartoon representations of the structural alignment between WT *Sa*CstB (**green**) model to 2.69 Å resolution with "*Sa*CstB incomplete model" (**blue**) to 1.93 Å resolution (RMSD value of 0.60 Å) as a dimer (A) and monomer (B), and with (C) *Sa*MβL model (pink) to 2.10 Å resolution, (D) hETHE1 [73] model (**orange**) to 2.61 Å resolution and (E) *Aa*βL-Rhod model (**yellow**) to 2.70 Å resolution with (**grey spheres**) zinc ion in the catalytic site. Iron ions (**red spheres**). Images generated in *PyMOL* [76].

In Figures 2.10.A and 2.10.B we can clearly see the rhodanese domain present in the ***Sa*CstB** model, in opposition to the previous incomplete model solved at 1.93 Å (Figure 2.5). We propose that the TCEP removal was the key factor to stabilize the rhodanese domain, Cys-201 loop region and elucidate the full-length *Sa*CstB structure.

Figure 2.10.C clearly demonstrates high structural similarity between ***Sa*CstB** and ***Sa*MβL** proteins (RMSD of ~0.76 Å for the aligned 337 C_α atoms), both present in *S. aureus*. The most evident similarity between these models is the presence of two domains, PDO and pseudorhodanese, with approximate 74.2 % of sequence identity, which could explain the overall structure similarity [111]. The most preeminent structural differences between the two models are the unmodelled rhodanese domain in *Sa*MβL, different conformations of loop regions, and, most strikingly, the unmodelled loop containing the Cys-201 (from residues Pro-200 to Lys-203) in *Sa*MβL model. This last observation also supports the previous remark regarding the high mobility of this loop and its absence in the "*Sa*CstB incomplete model". Moreover, the absence of rhodanese domain in *Sa*MβL makes this model alone ineffective to solve the full-length *Sa*CstB through molecular replacement. Hence, this further evidence the need to use online tools (e.g., I-TASSER) to build a homology model for the rhodanese domain or solve the "phase-problem" with another method. On this note, *Sa*MβL phasing was done using anomalous data acquired from a single-wavelength anomalous diffraction (SAD) experiment at ~0.9786 Å wavelength. Therefore, a SAD experiment could have been planned making use of the non-heme iron site in *Sa*CstB, in case of the impossibility to solve its structure by molecular replacement.

Interestingly, *Sa*MβL protein was solved from a crystal with a unit cell with the same space group as *Sa*CstB ($P\,4_3\,2_1\,2$), however with very different unit cell dimensions (a = b = 92.5 Å, c = 385.4 Å, α = β = γ = 90º). This observation is expected from crystals with different asymmetric units (homotetramer for *Sa*MβL versus homodimer for *Sa*CstB), even though the crystallization conditions were similar. In short, *Sa*MβL crystals were produced with metallo-β-lactamase at 7.6 $mg{\cdot}ml^{-1}$ in 0.01 M Tris buffer at pH 8.3 and 0.25 M NaCl, and crystallized in 25 % (V/V) PEG 3350, 0.1 M Tris at pH 8.5 and 0.2 M MgCl at room temperature with vapour-diffusion sitting-drop method.

Figure 2.10.D demonstrates the striking difference between ***Sa*CstB** and **hETHE1** [73], showing that even though these proteins have similar persulfide dioxygenase activity, hETHE1 overall folding is very different from the core PDO domain in *Sa*CstB. These structural differences reflect their low sequence identity (26.1 %), as correctly predicted in 2015 by Shen et al. [71].

Figure 2.10.E shows the overall structural similarity between ***Sa*CstB** and ***Aa*βL-Rhod** proteins (RMSD of ~1.86 Å for 408 aligned C_α atoms), with the three domains, PDO, RHD and

Rhod, properly aligned and superimposed. The loop with Cys-201 is also present in the *Aa*βL-Rhod model with an almost identical conformation to *Sa*CstB. The most pronounced differences lay on the loop regions more exposed to the solvent and the presence of a zinc iron coordinating the active site in the *Aa*βL-Rhod model instead of the typical iron coordination complex in *Sa*CstB.

Figure 2.11 highlights part of the amino acid sequence conservation between *Sa*CstB, *Sa*MβL, *Aa*βL-Rhod and hETHE1 proteins.

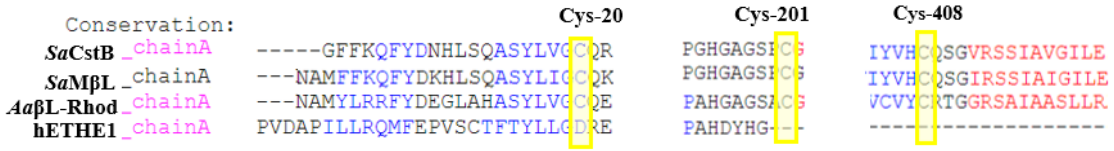


Figure 2.11 – Amino acid sequence alignment of CstB from *S. aureus* and the proteins deposited in the Protein Data Bank *Sa*MβL, *Aa*βL-Rhod and hETHE1 [73]. Predicted secondary structure features are colored **red** for α-helices and **blue** for β-strands. Cysteine numeration follows *Sa*CstB amino acid sequence (**yellow highlights**) [71]. Image adapted from PROMALS3D multiple sequence and structure alignment online server [112].

In Figure 2.11 is possible to observe the *Sa*CstB cysteine conservation in the discussed protein models. In agreement with what was previously stated, hETHE1 model [73] has the least conserved primary sequence, amongst the aligned sequences, while, the three cysteines are strictly conserved in *Sa*MβL and *Aa*βL-Rhod, being the closest related protein sequences to *Sa*CstB.

SaCstB active site

The full-length *Sa*CstB model comprises all three cysteines, Cys-20, Cys-201 and Cys-408. These are distanced 29.2 Å, 3.7 Å and 27.1 Å away from the iron ion of the same protomer, respectively. Notably, Cys-408 is only ~4 Å away from the iron ion of the neighboring protomer. We can speculate that this might be an important structural feature for the intercommunication of the three catalytic activities of *Sa*CstB. Interestingly, in the non-heme iron site (Figure 2.12 panel A), it is possible to observe a tetragonal pyramidal coordination of the iron ion covalently bounded to the amino acids His-56, His-119 and Asp-145. The iron ion was modelled with 100 % occupancy, and is distanced ca. 2.2 Å, 2.2 Å and 2.0 Å from the aforementioned amino acid residues, respectively. Interestingly, the amino acids surrounding the active site's cavity are mainly hydrophobic (namely, Ile-57, His-58, Ala-59 and Leu-157), but the iron ion is further coordinated by two ordered waters that complete the tetragonal pyramidal arrangement. Moreover, one of these water molecules establishes an H-bond with Asp-60 (~2.7 Å).

Another important feature, as highlighted in Figure 2.12 panel B, is the presence of Cys-201 in close proximity (~3.7 Å) to the iron ion; however, too far distanced to form a covalent bound. This structural arrangement is most likely one of the key structural elements important to the catalytic activity of this enzyme.

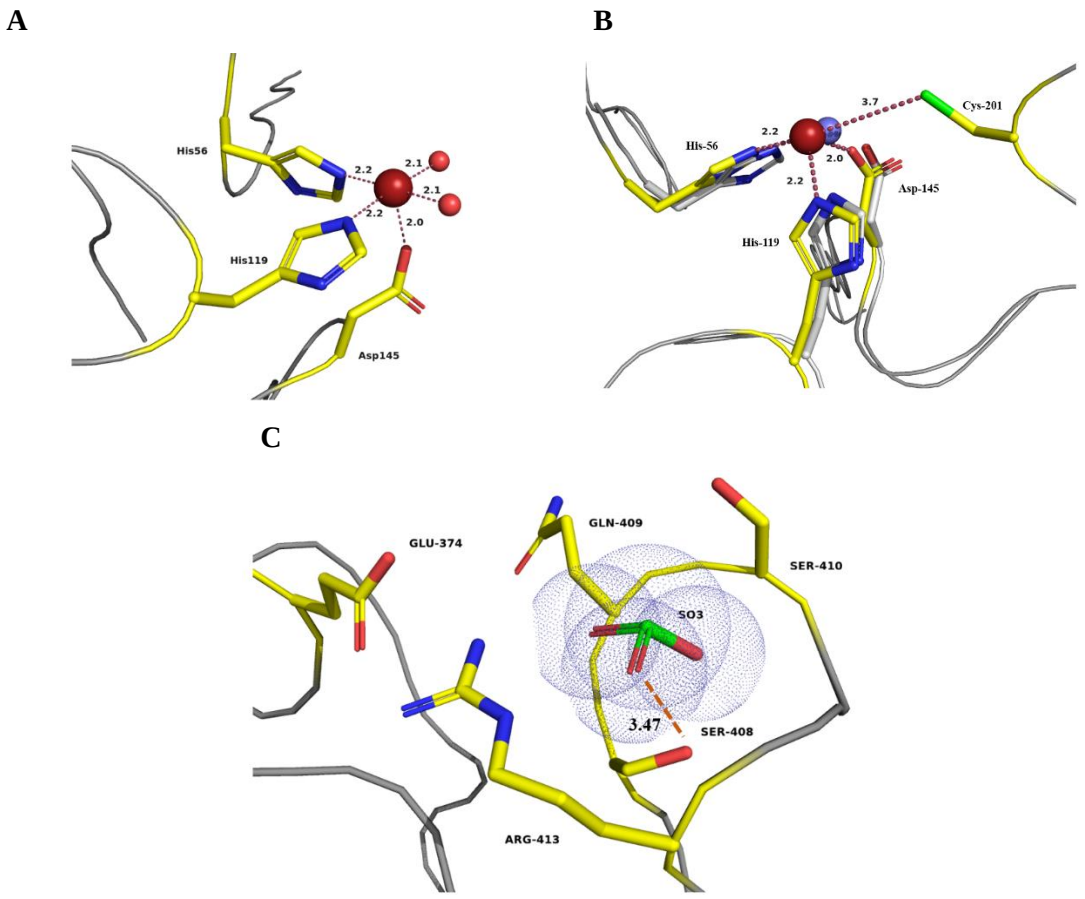


Figure 2.12 – (A) Wires and sticks representation highlighting the three catalytic site residues (carbon atoms in **yellow**) in coordination with the iron ion (**red sphere**) and two waters (**orange spheres**), with their respective distances (in Å) to the iron ion, for the WT *Sa*CstB dataset processed to ~2.7 Å resolution. (B) Wires and sticks representation evidencing the iron site residues, with their respective distances (in Å) to the iron ion, and structural alignment of chain A from *Sa*CstB (carbon atoms in **yellow**) and chain A from *Aa*βL-Rhod (carbon atoms in **grey**). Zinc ion (**purple sphere**). (C) Wires and sticks representation of the *Sa*CstB C201S/C408S variant model to ~2.2 Å resolution with a refined SO_3^{2-} ion near Ser-408. Images were created in *PyMOL*[76].

In ***Aa*βL-Rhod**, the iron site is very similar to ***Sa*CstB**, apart from the zinc atom (Figure 2.12 panel B **purple sphere**) in the center of the catalytic site. In Figure 2.12 panel B it is possible to see that His-56, His-119 and Asp-145 are conserved, covalently bond to the zinc atom, and in very similar spatial positions as in *Sa*CstB.

SaCstB variant structures

Crystal structures of *Sa*CstB variants presented an identical overall fold to the WT of *Sa*CstB. Surprisingly, one SO_3^{2-} ion was refined in all variant models, even for *Sa*CstB C201S variants. Sulfite was always trapped in the same position and orientation at the catalytic site of rhodanese domain (Figure 2.12 panel C), and is coordinated through hydrogen bonds to Cys/Ser-408, Ser-410 and Arg-413. The environment surrounding SO_3^{2-} is mainly composed by polar amino acid residues, e.g., Glu-374, Gln-409 and Arg-413. Furthermore, and taking these observations into consideration, a new sequence alignment search revealed that Glu-374, Cys-408, Gln-409, Ser-410 and Arg-413 are strictly conserved in *Sa*MβL and *Aa*βL-Rhod sequences, suggesting that the environment surrounding the Cys-408 residue is important for catalysis. Moreover, continuous electron density was observed between Ser-408 and SO_3^{2-} ion, distanced at 3.47 Å (O^{γ}-S). Although more data needs to be analyzed before we could propose any mechanism of action, we might speculate that the presence of these residues, namely Arg-413, might be important for the proper orientation of SO_3^{2-} ion. Noteworthy, electron density for the SO_3^{2-} ion was always observed in the maps, even without being a component of the crystallization condition or added to the crystal in soaking experiments. This further reinforces the importance of sulfite for the enzyme, namely as a reaction product from the first catalytic activity and a substrate for the second, which was likely trapped during protein production. In contrast, no other ligands (from expression, purification or crystallization), were observed in *Sa*CstB variant models.

Several follow-up studies are currently ongoing, namely model building and crystallographic refinement of 23 additional datasets (with resolutions between 2.2 and 4 Å), not only for the WT, but also for the *Sa*CstB variants in complex with different substrates, substrate-like moieties and products of each catalytic activity. However, it feels safe to say at this point that the C201S substitution in all variants causes a disordering of the loop where this residue lies. This important catalytic residue seems to be less ordered in the *Sa*CstB variants than in the presence of Cys-201. Moreover, differences are also observed in the coordination sphere of the non-heme iron ion, namely the absence of the covalent bond with Asp-145. This observation emphasizes the importance of the cysteine residue for catalysis, and points to the possibility of other interactions at the catalytic site.

2.5. Concluding Remarks

- We have previously crystallized the WT of *Sa*CstB and determined its 3D structure to 1.93 Å resolution, although the third domain (Rhod) of the enzyme was not observed in the electron density maps, this result was crucial for the present work;
- Herein we described an optimized purification/crystallization procedure that allowed the structure characterization of the full-length WT of *Sa*CstB to 2.69 Å resolution;
- Crystallographic structures were also determined for single variant C201S, C408S, and double variant C201S/C408S variants to 3.19 Å, 2.44 Å and 2.20 Å, respectively;
- *Sa*CstB and its variants crystallize as dimers, with each monomer consisting of three domains, PDO, RHD and Rhod, folded to eleven α-helices, four-stranded parallel β-sheet and five-stranded anti-parallel β-sheets;
- The iron site consists of an iron ion coordinated by residues His-56, His-119 and Asp-145, distanced at ~2.2 Å, 2.2 Å and 2.0 Å, respectively, and two water molecules;
- Structural analysis was performed to all *Sa*CstB variants, and its structural features compared to metallo-β-lactamase from *S. aureus* (*Sa*MβL), β-lactamase and rhodanese fusion protein from *A. acidocaldarius* and hETHE1 from *H. sapiens*, (PDB IDs 3R2U, 3TP9 and 4CHL, respectively);
- Several sulfur-based ligands (e.g., SO_3^{2-}, $Na_2S_2O_3$, GSH, CysSS, ...) were tested in co-crystallization and soaking experiments;
- No enzymatic side products could be resolved in any WT *Sa*CstB structures;
- One SO_3^{2-} ion was present in all *Sa*CstB variant structures; it is distanced at ~3.47 Å away from Cys/Ser-408 (S^{γ} or O^{γ}), and is coordinated by residues Cys/Ser-408, Ser-410 and Arg-413 through hydrogen bonds;
- The determination of *Sa*CstB X-ray structure and its variants C201S, C408S, and C201S/C408S, marks an important milestone towards a better understanding of this enzyme catalytic mechanism.

Chapter III

EfCoAPR

3. Coenzyme A persulfide reductase from the human pathogenic bacterium *Enterococcus faecalis* (*Ef*CoAPR)

3.1. *Enterococcus faecalis*

Alongside *S. aureus*, *Enterococci* are one of the leading causes of healthcare-associated infections (also known as nosocomial infections). Within this genus, *Enterococcus faecalis* (*E. faecalis*) is the most predominant species present in the human gastrointestinal tract, and, developing, over the years, a high resistance phenotype to antibiotics. *Enterococci* are gram-positive facultative anaerobes, with an optimal growth temperature around 35 ºC. These pathogens are leading causes of nosocomial bacteremia with an overall mortality rate near 30 %, often associated with endocarditis [113].

3.2. Coenzyme A persulfide reductase

In *E. faecalis*, a conserved transcribed *cst*-like operon, similar to *S. aureus*' *cst* operon, is hypothesized to encode for two rhodanese enzymes (RhodA and RhodB) and an NADH/FAD-dependent coenzyme A (CoA) persulfide reductase (*Ef*CoAPR) [114].

*Ef*CoAPR is a cytosolic multidomain enzyme with a theoretical molecular mass of 59.33 kDa. *Ef*CoAPR structure comprises an N-terminal CoA disulfide reductase (CDR) domain and a C-terminal rhodanese homology domain (RHD), the latest in resemblance with *Sa*CstB. However, while CstB-like enzymes rely on LMW persulfide oxidative chemistry to maintain cellular homeostasis at the mitochondrial level, in anaerobic conditions, *Ef*CoAPR uses reductive chemistry to maintain the sulfide levels in the organism [71,114]. The two catalytic cysteines present in *Ef*CoAPR domains, Cys-42 in CDR domain and Cys-508 in the RHD domain, are required for *Ef*CoAPR activity. Site mutations of these residues lead to a complete and irreversible loss in protein function. Although the persulfide activity is evident for the CDR domain, with high substrate specificity to CoASSH, the role of the RHD domain and Cys-508 in the catalysis remains unclear [114,115]. *Ef*CoAPR reduces coenzyme A persulfide (CoASSH), producing free thiol CoA (CoASH) and H_2S, in the presence of FAD and NADH (Figure 3.1). This activity is essential to regenerate H_2S, reduced coenzyme A intracellular levels and maintain metabolic homeostasis [114].

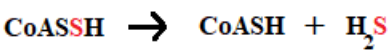


Figure 3.1 – Coenzyme A persulfide reductase (*Ef*CoAPR) activity. Reduction of coenzyme A persulfide (CoASSH) to free thiol coenzyme A (CoASH) while producing hydrogen sulfide in a FAD-dependent and NADH-requiring mechanism. Adapted from Shen et al. (2018) [114].

To date, no structure characterization has been published for *Ef*CoAPR [114], with the most similar structure being that of CoA disulfide reductase-rhodanese fusion protein (CDR-RHD) from *Bacillus anthracis* (PDB ID: 3ICS) [116]. Therefore, our study aims to further analyze this structural arrangement. This work is interconnected with the previously reported *Sa*CstB structures, *Ef*CoAPR and *Sa*CstB have been proposed to employ similar strategies to limit the accumulation of LMW persulfides, RSS, CoASSH and other products formed under sulfide stress and cellular signaling [114]. Both structures can provide valuable clues for understanding sulfide homeostasis in two important pathogens.

In resemblance to the work presented for the *Sa*CstB enzyme, this chapter describes the X-ray structure characterization of CoA-persulfide reductase from Enterococcus *faecalis* (*Ef*CoAPR) performed at Archer Lab. It is a joint project with Professor Dr. David P. Giedroc, Indiana University – Bloomington, USA.

3.3. Materials and Methods

Gene expression and protein production were performed in the Giedroc Lab as previously described in Shen et al. (2018) [114], with slight modifications. In brief, *coaP* gene from *E. faecalis* strain OG1RF was fused with a 6-polyhistidine tag and cloned into pET15b expression plasmid. The expression plasmid was transformed into *E. coli* Rosetta and cultured in LB medium at 37º C. *Ef*CoAPR expression was induced with 1 mM IPTG for 16 hours at 37 ºC. After cell lysis by sonication, DNA removal, protein salting-out and pellet resuspension, His-tagged *Ef*CoAPR was purified by Ni-NTA affinity chromatography (IMAC), followed by size exclusion chromatography (*G200 HiLoad 16/600 Superdex*), using 20 mM MES pH 6.0, 30 mM NaCl and 5 % (V/V) glycerol buffer. Pure *Ef*CoAPR fractions were collected, concentrated and stored at -80 ºC [114]. The protein amino acid sequence can be consulted in Appendix 5.

3.3.1. Sample preparation

*Ef*CoAPR was concentrated using an *Amicon ultra-MILLIPORE 30 kDa* by repeated concentration steps at a centrifugal force of 3500 x *g* (Centrifuge 5804 R, *Eppendorf*). A final volume of ~1.2 ml at 23 $mg \cdot ml^{-1}$ could be obtained. The protein sample was then ultracentrifuged for 1 hour at 40,000 x rpm (217,200 x *g*, *Ultracentrifuge optima TL-100, TLA-100.3 Fixed-Angle Rotor*, *Beckman Coulter*), aliquoted and stored at -80 ºC.

UV-Vis spectra were obtained with a NanoDrop ONE Microvolume UV-Vis Spectrophotometer (*Thermo Fisher Scientific*). Protein concentration at 280 nm was also calculated in a NanoDrop ONE Microvolume UV-Vis Spectrophotometer (*Thermo Fisher Scientific*), using a 29.47 $M^{-1} \cdot cm^{-1}$ molar absorptivity, and a theoretical molecular weight of 61.88 kDa. The protein buffer, 20 mM MES pH 6.0, 30 mM NaCl and 5 % (V/V) glycerol, was used as a blank solution for all protein quantifications.

3.3.2. Crystallization experiments

Several methodologies were attempted to crystallize *Ef*CoAPR. We have initially used different commercially available crystallization screens, namely JCSG+, BCS Screen, PACT Premier and Shot Gun, all from *Molecular Dimensions*. Moreover, different protein concentrations ranging from 8-23 $mg \cdot ml^{-1}$, and co-crystallization with different ligands e.g., FAD, NADH, NAD+, GSSH, GSH, acetyl-CoA, alliin, S-methyl and S-allyl cysteine, were also tested. All crystallization experiments herein described were conducted following vapour diffusion sitting-drop method [48]. Reservoirs from a *96-well 3-drop iQ plate* (*TTP Labtech*) were filled with

40 µl of reservoir solutions from each crystallization screen (*Molecular Dimensions*). Drops containing 100 nl of protein at various concentrations and 100 nl of reservoir solution were dispensed using Mosquito LCP (TTP Labtech) liquid dispenser robot, at 20 ºC and 70 % humidity.

*Ef*CoAPR crystal seeds were produced after a single protein crystal developed under G3 condition from PACT Premier [20 % (V/V) PEG 3350, 0.1 M Bis-Tris propane pH 7.5, 0.2 M NaI]. 2 µl of reservoir solution were added on top of the crystallization drop and the crystal was crushed using a glass crystal crusher (*Hampton Research*). The seeds were aspirated to an 1.5 ml eppendorf tube with a stainless-steel bead (*Hampton Research*) and diluted to a final volume of 35 µl along with several cycles of aspiration/washing. The seeds were then vortexed for 30 seconds and placed in ice for 10 seconds. This step was repeated 2 more times. A 20 µl seed stock with (1:5) dilution was prepared from the initial seed stock for cross-seeding crystallization experiments.

Reservoirs from a *96-well 3-drop iQ plate* (*TTP Labtech*) were filled with 40 µl of each PACT Premier screening solutions. Two drops per well with 100 nl of *Ef*CoAPR at 20 $mg \cdot ml^{-1}$, 70 nl of reservoir solution and 30 nl of seeds were dispensed using Mosquito LCP (TTP Labtech) liquid dispenser robot, at 20 ºC and 70 % humidity. One drop contained seeds from the initial seed stock and the second drop seeds at a (1:5) dilution.

A scale-up test based on the best hits from PACT Premier screen [77] and seeding technique was performed for a gradient of 18 to 22 % (V/V) of PEG 3350, 0.1 M Bis-Tris propane pH 6.4-6.6, and 0.2 M NaI or NaF. Reservoirs from a *48-well 2-drop MRC Maxi* crystallization plate (*Swissci*) were filled with 184 µl of reservoir solution in *Dragonfly* (*SPT Labtech*), followed by mixing in a *MXone* (*SPT Labtech*) mixer. Drops of (500+150+350) nl of *Ef*CoAPR at 20 $mg \cdot ml^{-1}$, seeds and reservoir solution, respectively, were dispensed using *Mosquito LCP* (*TTP Labtech*) liquid dispenser robot, at 20 ºC and 70 % humidity. One drop contained seeds from the seed stock and the second drop seeds with 1:5 dilution.

3.3.3. Data collection and processing

Native Apo crystals grown under 20 % (V/V) PEG 3350, 0.1 M Bis-Tris propane buffer at pH 6.50 and 0.2 M sodium iodide were transferred to 1 µl drops containing **40 %** (V/V) PEG 3350, 0.1 M Bis-Tris propane buffer at pH 6.50 and 0.2 M sodium iodide cryo-protectant solution. The crystals were mounted in *Dual Thickness MicroLoops LD*^{TM} long neck loops from *MiTeGen* (35 mm or 50 mm), flash-cooled in liquid nitrogen and sent to European Synchrotron Radiation Facility (ESRF, Grenoble – France) [50] for data collection. The X-ray diffraction data collection was performed on microfocus beamline ID23-2, with a fixed wavelength of 0.873 Å (14.2 keV), and recorded with *DECTRIS PILATUS3 X 2M* detector.

The diffraction data were indexed and integrated using *XDS* [81], space-group assignment with *POINTLESS* [82], and scaling with *AIMLESS* [83] and *STARANISO* [84]. All programs were used within the *autoPROC* data processing pipeline [85,75]. Final diffraction data were converted to MTZ format with *CTRUNCATE* [86–89], and a set corresponding to 5 % of the total measured reflections was created and identified with Free-R flags [90,91].

3.3.4. Phasing, model building and refinement

Data quality was assessed with *phenix.xtriage* tool [92] within *PHENIX* suite of programs [74,75]. The structure was solved by molecular replacement (MR) with *PHASER* [93] as implemented in *PHENIX* [74,75], using the X-ray structure of CoADR-RHD from *Bacillus anthracis* (PDB ID: 3ICS) [116], as the search model. The template structure was devoid of any cofactors, solvent molecules and other ligands, and only one chain (monomer), was used for MR-search. The initial *Ef*CoAPR model was corrected with automated model building using the *AutoBuild* tool as implemented in *PHENIX* [74,75] and *BUSTER-TNT* "-L" macro [99,100] was executed in the refinement command to search for unmodelled density on the electron density maps. Iterative model building and refinement were carried out in a cyclic manner with *phenix.refine* [98] within *PHENIX* [74,75] suite of programs, *BUSTER-TNT* [99,100] and *COOT* [101], until a complete model was built and refinement convergence achieved. *Ef*CoAPR model was validated with the *MolProbity* program [102] implemented within *PHENIX* suite of programs [74,75].

Structural illustrations were rendered using *PyMOL* [76] and *COOT* [101] programs.

3.4. Results and Discussion

Protein purification

Figure 3.2.A shows the chromatogram profile after size exclusion chromatography performed at Dr. Giedroc's Lab, before the protein was sent to Archer Lab. Upon arrival, protein purity and homogeneity were assessed with SDS-PAGE (Figure 3.2.B).

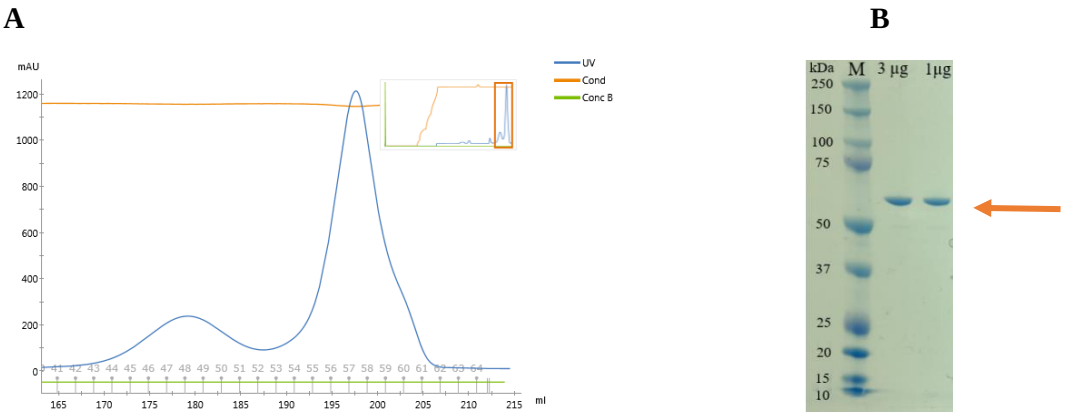


Figure 3.2 – (A) Size-exclusion chromatogram from a *G200 HiLoad 16/600 Superdex* column with 20 mM MES pH 6.0, 30 mM NaCl, 5 % (V/V) glycerol elution buffer. 4 ml of sample was injected on the column with a flow rate of 0.7 ml·min^{-1}, 2 ml fractions were collected and the protein absorbance was registered on AKTA Pure. (B) 4-12 % precast MOPS SDS-PAGE gel from *Novex NuPAGE™*. Gel of purified *Ef*CoAPR samples ran at 80 V with an *Electrophoresis Power Supply - EPS 201* from *Amersham Biosciences*. *Roti-Load* (1x) loading buffer from *Roth*. (M) Molecular weight marker *Precision Plus Dual Color Standards* from *Bio-RAD*. Protein buffer 25 mM MES pH 6.0, 100 mM NaCl, 5 % (V/V) glycerol and 2 mM TCEP. *Ef*CoAPR protein bands (**orange arrow**).

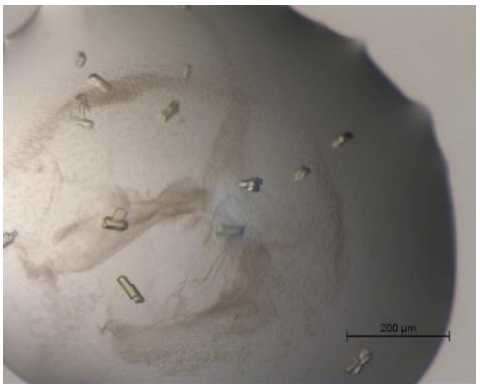
The sharp high peak in chromatogram of Figure 3.2.A corresponds to *Ef*CoAPR sample. These fractions were collected and sent to our Lab to proceed with crystallization experiments. The small broad peak shows the contaminants that were discarded in the gel filtration purification step. In Figure 3.2.B a single band around 60 kDa (**orange arrow**) is observed, which indicates a highly pure protein sample. This molecular weight is consistent with the protein theoretical molecular mass of 61.88 kDa (ProtParam tool analysis from ExPASy server [103]) and the quality of the samples is suitable to proceed with crystallization experiments.

EfCoAPR crystallization and structure determination

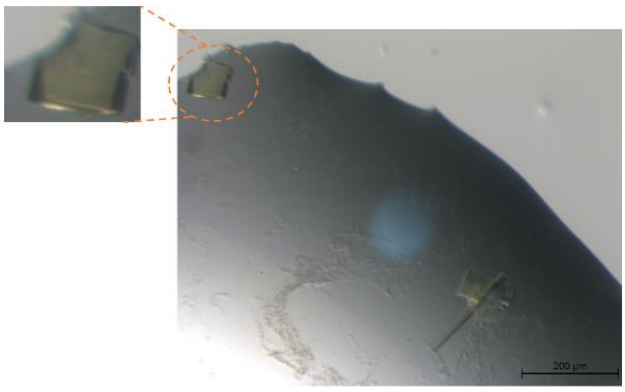
The first native Holo *Ef*CoAPR crystals developed after 9 days in drops from the PACT Premier screen [77] and grew to approximate maximum dimensions of 0.17 x 0.07 x 0.05 mm^3. Interestingly, these conditions were very similar to those where CstB crystals were grown, with 20 % (V/V) PEG 3350, 0.1 M Bis-Tris propane at pH 6.5 and 0.2 M of different halide salts (NaF, NaBr and NaI). Native *Ef*CoAPR crystals are yellow-greenish rectangular prisms, characteristic of FAD-dependent enzymes such as *Ef*CoAPR. However, these crystals showed poor diffraction ~8 to 10 Å, at synchrotron radiation source. Scale-up tests were carried out to optimize and produce better quality crystals, although with no success.

After three months, *Ef*CoAPR crystals were grown in G3 condition from the initial PACT Premier screen (similar to the previous condition, although with NaI salt and buffer pH at 7.5). These crystals (Figure 3.3.A) were used for seeding experiments using the PACT Premier screen [77] and after three days new crystals were formed (Figure 3.3.B). Figure 3.3.C displays an illustrative diffraction pattern for native Holo *Ef*CoAPR crystal mounted in a 50 mm loop (Figure 3.3.B) produced with micro-seeding experiment with 1:5 seeds dilution.

A



B



C

D

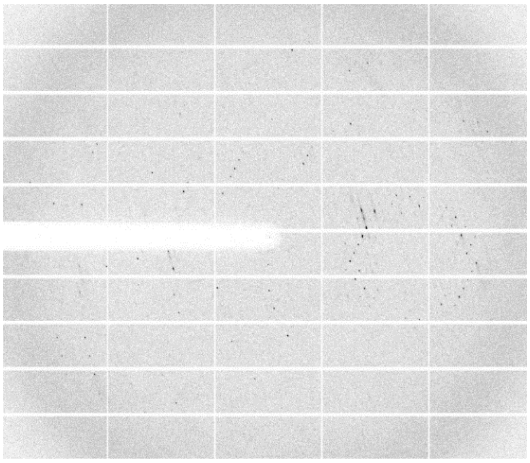


Figure 3.3 – (A) Drop with *Ef*CoAPR crystals grown after three months in 20 % (V/V) PEG3350, 0.1 M Bis-Tris Propane at pH 7.5 and 0.2 M sodium iodide. (B) Drop with *Ef*CoAPR crystals from (A) 1:5 seeds dilution, grown after three days in 20 % (V/V) PEG 3350, 0.1 M Bis-Tris propane buffer at pH 6.50 and 0.2 M sodium iodide. Photos acquired in Leica MZ16 Stereomicroscope (Leica Microsystems) with 11.5x ampliation. (C) *Ef*CoAPR crystal mounted in a 50 µm aperture DT MicroLoop LD (Jena Bioscience). (D) X-ray diffraction pattern of (B) crystal collected at 1.95 Å with DECTRIS PILATUS3 X 2M of ID23-2 beamline at ESRF, Grenoble, France.

The *Ef*CoAPR Holo crystal (Figure 3.3.B) belongs to the orthorhombic space group *C* 2 2 2_1 with approximate unit cell parameters of a = 139.8 Å, b = 194.8 Å, c = 91.5 Å, and α = β = ϒ = 90º. The crystal asymmetric unit contains two monomers of *Ef*CoAPR, corresponding to a Matthew's coefficient [104] (V_m) of 2.26 $Å^3 \cdot Da^{-1}$ and a solvent content of around 46 %.

Table 3.1 summarizes the crystallization condition for WT Holo *Ef*CoAPR crystal discussed in this Thesis.

Table 3.1 - WT of Holo *Ef*CoAPR crystallization summary.

Crystal ID	Crystallization condition	Cryoprotectant condition
WT of *Ef*CoAPR Crystal ID: **Holo *Ef*CoAPR**	**Drop volume**: 100 nl of protein + 30 nl of seeds + 70 nl of reservoir solution **Reservoir volume**: 40 ul **Method**: Vapour diffusion sitting-drop **Temperature**: 20 ºC **Crystallization solution**: 20 % (V/V) PEG 3350, 0.1 M Bis-Tris propane buffer at pH 6.50 and 0.2 M sodium iodide	**40 %** (V/V) PEG 3350 0.1 M Bis-Tris propane buffer pH 6.50 0.2 M sodium iodide

Table 3.2 summarizes the results of X-ray data collection, processing, refinement statistics and model quality parameters for WT Holo *Ef*CoAPR. This table will suffer alterations before publication of these results.

Table 3.2 – Data collection, processing, refinement statistics and model quality parameters for *E. faecalis* Holo CoAPR structure.

	Holo *Ef*CoAPR	
Data Collection		
Synchrotron Facility	ESRF (Grenoble – France)	
Beamline **Detector**	ID23-2 *DECTRIS PILATUS3 X 2M*	
Wavelength (Å)	0.873	
Data Processing		
	autoPROC/STARANISO	***autoPROC/AIMLESS***
Resolution range (Å)[a]	48.25 – 2.05 (2.26 – 2.05)	45.73 – 2.52 (2.57-2.52)
Crystal system	Orthorhombic	
Space group	*C* 2 2 2_1	
Unit cell parameters **a, b, c (Å)** **α, β, ϒ (º)**	139.80, 194.81, 91.46 90, 90, 90	
Total number of reflections	229727 (9588)	186084 (9493)
Total number of unique reflections	54207 (2710)	42108 (2113)
Multiplicity	4.2 (3.5)	4.4 (4.5)
Completeness (%)	-	99.3 (99.7)
Spherical	68.7 (13.5)	-
Ellipsoidal	92.5 (65.3)	-
Mean I/σ(I)	7.2 (1.6)	8.8 (2.2)
R_{merge} (%) [b]	13.6 (81.8)	12.0 (65.1)
R_{meas} (%) [c]	15.5 (94.2)	13.7 (74.0)
$R_{p.i.m}$ (%) [d]	7.4 (45.7)	6.4 (34.4)
$CC_{1/2}$ (%) [e]	99.6 (64.9)	99.6 (82.7)
Model Refinement		
	autoPROC/STARANISO	***autoPROC/AIMLESS***
Total number of reflections used in refinement	42075 (4209)	-
Total number of reflections used for R_{free}	2147 (206)	-
R_{work} (%) [f]	20.12 (30.76)	-
R_{free} (%) [g]	23.26 (33.43)	-
RMSD Bonds (Å) [h]	0.004	-
RMSD Angles (º) [h]	0.66	-

Number of atoms	-	-
Protein residues	1083	-
Non-hydrogen atoms	8529	-
Macromolecules	8119	
Ligands	168	-
Waters	242	-
Ramachandran plot	-	-
Most favoured (%)	96.7	
Outliers (%)	0.3	-
Rotamer outliers (%)	1.2	-
Clashscore [i]	1.27	-
Molprobity score [j]	1.11	-
Average *B*-factors (Å^2)	38.14	-
Protein	38.41	-
Ligands	31.52	-
Solvent	33.42	-

[a] Information in parenthesis refers to the last resolution shell.

[b] $R_{merge} = \sum_{hkl} \sum_i |I_i(hkl) - \overline{I(hkl)}| \,/\, \sum_{hkl} \sum_i I_i(hkl)$.

[c] $R_{meas} = \sum_{hkl} [N/(N-1)]^{\frac{1}{2}} \sum_i |I_i(hkl) - \overline{I(hkl)}| \,/\, \sum_{hkl} \sum_i I_i(hkl)$.

[d] $R_{p.i.m} = \sum_{hkl} [1/(N-1)]^{\frac{1}{2}} \sum_i |I_i(hkl) - \overline{I(hkl)}| \,/\, \sum_{hkl} \sum_i I_i(hkl)$.

[e] $CC_{1/2}$ as described in Karplus & Diederichs (2012). Science, 336(6084): 1030–1033.

[f] $R_{work} = \sum_h \sum_k \sum_l \frac{\{||F_o(h,k,l)| - |F_c(h,k,l)||\}}{\sum_h \sum_k \sum_l |F_o(h,k,l)|}$, where F_o and F_c are the observed and calculated structure factors for reflection *h*, respectively.

[g] R_{free} was calculated the same way as R_{work} but using only 5% of the reflections which were selected randomly and omitted from refinement.

[h] RMSD, root mean square deviation.

[i] Clashscore is the number of unfavourable all-atom steric overlaps ≥ 0.4Å per 1000 atoms. Word *et al*. (1999). Mol Biol, 285(4):1711-33.

[j] *MolProbity* score provides a single number that represents the central *MolProbity* protein quality statistics; it is a log-weighted combination of clashscore, Ramachandran not favoured and bad side-chain rotamers, giving one number that reflects the crystallographic resolution at which those values would be expected.

The *Ef*CoAPR X-ray model was refined to a final resolution of 2.05 Å with anisotropic diffraction limits of 2.01 Å, 2.36 Å and 2.46 Å and a final R_{work} of 20.12 % and an R_{free} of 23.26 % (*STARANISO* [84]). Although *phenix.xtriage* tool [92-75] detected translational NCS that could affect data interpretation, phasing by molecular replacement using *Ba*CoADR-RHD [116] as search model was successful. After several cycles of model refinement and model building, the final *Ef*CoAPR

model comprises 1,083 amino acid residues from Gly-1 to Pro-544 (chain A) and to Glu-546 (chain B), with two molecules of FAD, two 3'-phosphate-adenosine-5'-diphosphate (PAP) and 242 waters.

The electron density maps are clearly defined, leading to unambiguous building of *Ef*CoAPR model, apart from the disordered N-terminal His-tag, the loop Arg-476 to Gln-480 in chain B and the last two residues of chain A (Glu-545 and Glu-546). All these regions are solvent-exposed, implying higher flexibility. Moreover, no electron density was observed for NADH, suggesting that it may not be ordered in the crystal. Interestingly, the electron density map revealed a blurry blob in the region where CoA is expected to be, based on the CoADR-RHD from *Bacillus anthracis* (PDB ID: 3ICS) [116]. This electron density should correspond to CoA, with clear density for the PAP moiety, although the pantothenate and cysteamine moieties (Figure 3.4.A) could not be resolved. PAP is ~13.3 Å apart from FAD riboflavin (Figure 3.4.B), which was fully modelled and refined (*see active site section for a more detailed explanation*).

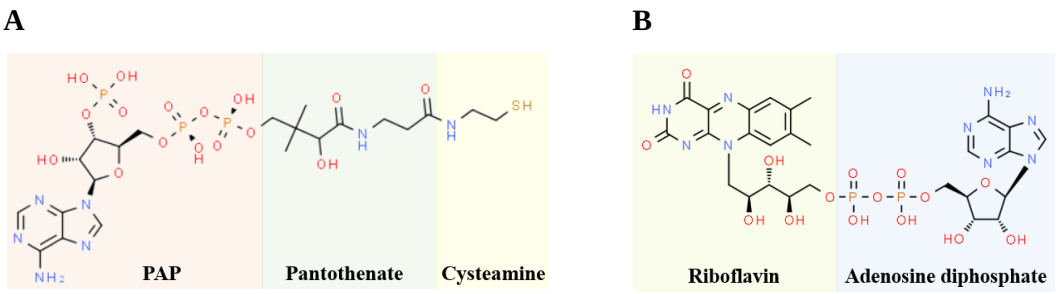


Figure 3.4 – (A) Coenzyme A (CoA) molecular structure, evidencing the PAP, pantothenate and cysteamine moieties. 2D representation adapted from ChemSpider (ID 6557) [117]. (B) Flavin adenine dinucleotide (FAD) molecular structure, evidencing the riboflavin and adenosine diphosphate (ADP) moieties. 2D representation adapted from ChemSpider (ID 559059) [117].

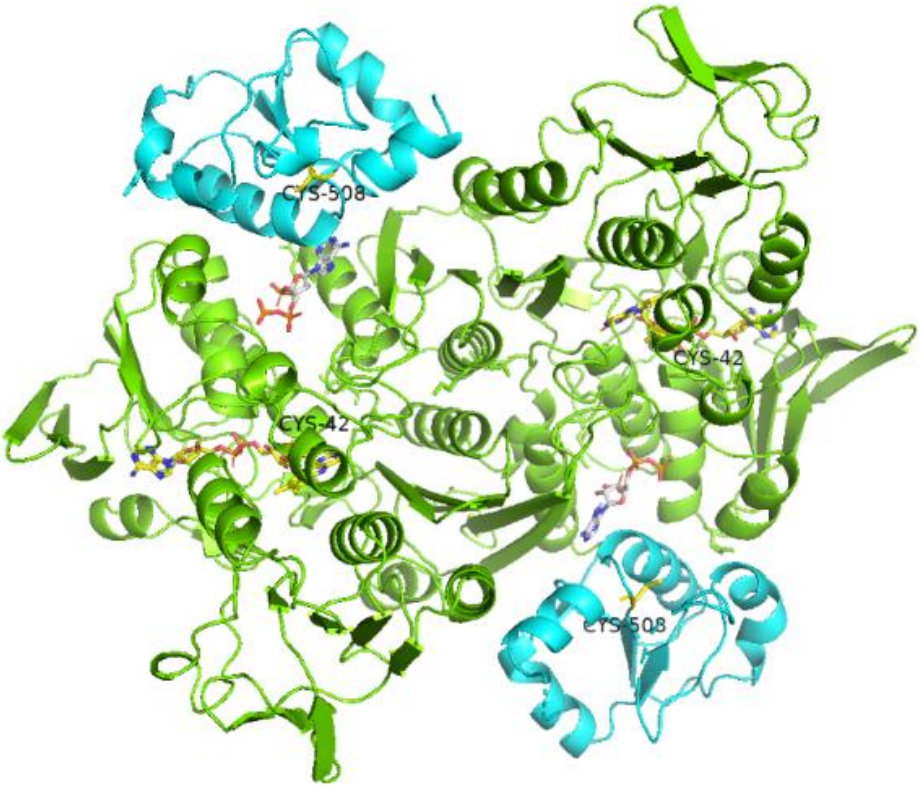
The Ramachandran plot, assessed with *MolProbity* [102] within *PHENIX* suite of programs [74,75] revealed that all nonglycine amino acid residues are located within allowed regions, except for Val-10 (chain A and B), Phe-481 (chain A). Overall, the final CoAPR model displays good geometry and stereochemistry, statistics with 1.2 % rotamer outliers, and a *MolProbity* score of 1.11.

EfCoAPR overall fold and similar structures

The physiological oligomeric assembly of *Ef*CoAPR is proposed to be homodimeric [114], and indeed a dimer is found in the crystal asymmetric unit. Each monomer is composed by two functional domains: an N-terminal CoA disulfide reductase (CDR) and a C-terminal rhodanese homology domain (RHD). According to PISA program [105,106], the dimer interface is predicted to have a dissociation energy (ΔG^{diss}) of -24.4 kcal · mol^{-1} and an approximate area of 3600 $Å^2$,

corresponding to ~14.7 % of the total solvent-accessible area of each monomer. The dimer interface is mainly stabilized by hydrophilic residues, with 59 H-bonds and 26 salt bridges. No disulfide bonds were established between the catalytic residues Cys-42 and Cys-508, which are ca. 26.6 Å apart. *Ef*CoAPR chains A and B are nearly identical, showing an RMSD of ~0.40 Å for 539 aligned C_α atoms (chains superposition performed with "Secondary Structure Matching" tool, within *COOT* [101,107]). Secondary structure analysis [108] revealed that each monomer consists of fifteen α-helices, six-stranded parallel β-sheets and seven-stranded anti-parallel β-sheets arranged in a globular shape and packed as dimer. Figure 3.5.A shows the crystallographic dimeric structure for *Ef*CoAPR solved to 2.05 Å resolution. Figures 3.5.B and 3.5.C depict the structural superposition of *Ef*CoAPR with the two most similar structures deposited in PDB, *Ba*CoADR-RHD [116] and SlPV-4 Npsr [115] (PDB ID's: 3ICS and 3NTA, respectively).

A



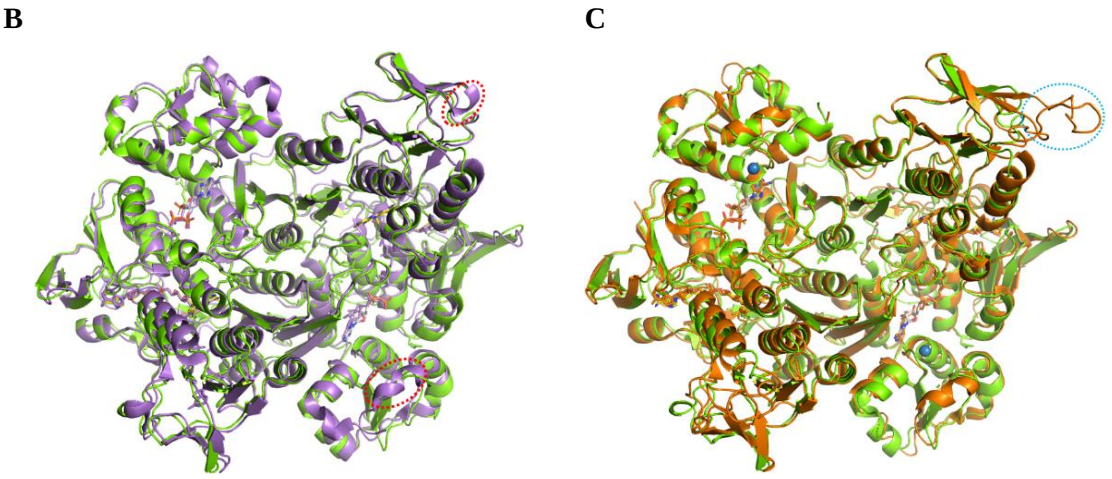


Figure 3.5 – (A) Cartoon representation of *Ef*CoAPR (Holo) full-length crystallographic model to 2.05 Å resolution colored by domain. CoA disulfide reductase (CDR) domain (**Chartreuse**). Pseudorhodanese homology domain (RHD) (**Cyan**). Carbon atoms colored yellow for FAD and white for PAP (following Wallen et al. 2009 [116] color code). (B) Cartoon representation of *Ef*CoAPR (**Chartreuse**) structural alignment with *Ba*CoADR-RHD (**purple**) model to 1.94 Å [116] and (C) *Sl*PV-4 Npsr to 2.01 Å resolution (**orange**) [115]. Chloride atoms present in *Sl*PV-4 Npsr model (**blue spheres**). Images generated in *PyMOL* [76].

According to Dali online server [109], amongst the first 100 hits with the highest structural similarity to *Ef*CoAPR model, we find proteins classified as coenzyme-A disulfide reductases (CoADR), FAD-dependent pyridine nucleotide-disulfide oxidoreductases and NADH (per)oxidases and NAD-dependent persulfide reductases (Npsr). **CoADR-RHD from *Bacillus anthracis*** (***Ba*CoADR-RHD**, PDB ID: 3ICS [116]) has the highest sequence identity, ~51.7 % for the 542 aligned C_α atoms, an RMSD of ~0.98 Å and a Z-score of 49.2. SSM analysis between *Ef*CoAPR and the wild-type **PV-4 Npsr, an FAD-dependent pyridine nucleotide-disulfide oxidoreductase from *Shewanella loihica*** (***Sl*PV-4 Npsr**, PDB ID: 3NTA [115]) and the **NAD-dependent persulfide reductase from *Archaeoglobus fulgidus*** (***Af*Npsr**, PDB ID: 6PFZ [118]), showed RMSDs of ~1.01 Å and 1.66 Å for the 535 and 503 aligned C_α atoms, and ~46.4 % and ~38.0 % of shared sequence identity, with Z-scores of 47.1 and 45.2, respectively.

The 3D structure of ***Ef*CoAPR** reveals the presence of CDR larger domain (residues 1-446) comprising Cys-42 and RHD smaller domain (residues 447-544) with Cys-508, in a dimeric arrangement (Figure 3.5.A). Previous studies on *Ef*CoAPR identified these cysteines as essential to CoA persulfide reductase activity [114], however, their long distance from each other (~26.6 Å) suggests major structural adjustments to accommodate a mechanism involving the cooperation of these residues.

Figure 3.5.B demonstrates the high structural similarity between ***Ef*CoAPR** and ***Ba*CoADR-RHD** [116] models (RMSD of ~0.98 Å). The most pronounced similarities are the regions nearby the active site, with an almost perfect superposition of CoA (PAP moiety) and

FAD cofactors. These similarities can be attributed to the high structural conservation of both CDR and RHD domains in this family of enzymes [114], even with ~51.7 % of sequence identity, since folding is more conserved than the amino acid sequence [111]. The main structural differences between these models are spotted on the loop regions, especially the Phe-215 to Gly-119 loop and the Asn-471 to Phe-480 disordered loop (red-dotted circles) in *Ef*CoAPR, which are represented as α-helices in *Ba*CDR-RHD. This observation is quite expected as amino acid residues in these loops are not conserved (Figure 3.6), although it has been demonstrated that α-helices can undergo several site mutations while maintaining structural stability [119,120].

Figure 3.5.C shows the structural superposition of ***Ef*CoAPR** and ***Sl*PV-4 Npsr** [115], evidencing a high structural similarity over the two models. The most pronounced similarities are observed in the core CDR domain, especially the two active sites and the dimer interface region. The most prominent differences are the quaternary structure of the RHD domains, the conformation of the loop regions, especially the ones more solvent-exposed, the larger loop consisting of residues His-225 to Ser-245 in *Sl*PV-4 Npsr (blue-dotted circle), and the presence of a chloride ion at 7.0 Å and 2.9 Å away from the adenine region (CoA) and RHD catalytic cysteine (Cys-531) [115], respectively, in both chains of *Sl*PV-4 Npsr.

Figure 3.6 highlights part of the amino acid sequence conservation among *Ef*CoAPR, *Ba*CDR-RHD [116], *Sl*PV-4 Npsr [115] and *Af*Npsr [118] proteins.

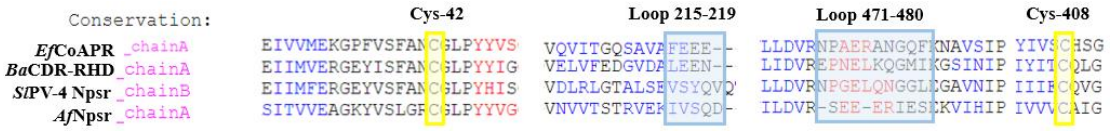


Figure 3.6 – Amino acid sequence alignment of CoAPR from *E. faecalis* and *Ba*CDR-RHD [116], *Sl*PV-4 Npsr [115] and *Af*Npsr [118]. Predicted secondary structure features are colored **red** for α-helices and **blue** for β-strands. *Ef*CoAPR catalytic cysteines (**yellow highlights**). Loop regions (blue highlights). Residue numeration follows *Ef*CoAPR amino acid sequence [114]. Image adapted from PROMALS3D multiple sequence and structure alignment online server [112].

In Figure 3.6 it is possible to observe that the catalytic cysteines Cys-42 and Cys-508 are highly conserved among the aforementioned protein models. Moreover, Cys-42 is conserved in all known CDRs and CDR-RHDs [114]. In agreement with the previous discussion, *Ba*CDR-RHD [116] has the most conserved primary structure with *Ef*CoAPR, in contrast with *Af*Npsr [118].

EfCoAPR active site

Regarding substrate accessibility to the active site, is possible to see in Figure 3.7.A a large opening cavity in the CoA binding pocket with solvent access and approximate dimensions of 24.5 Å x 13.0 Å x 16.0 Å (width x height x depth).

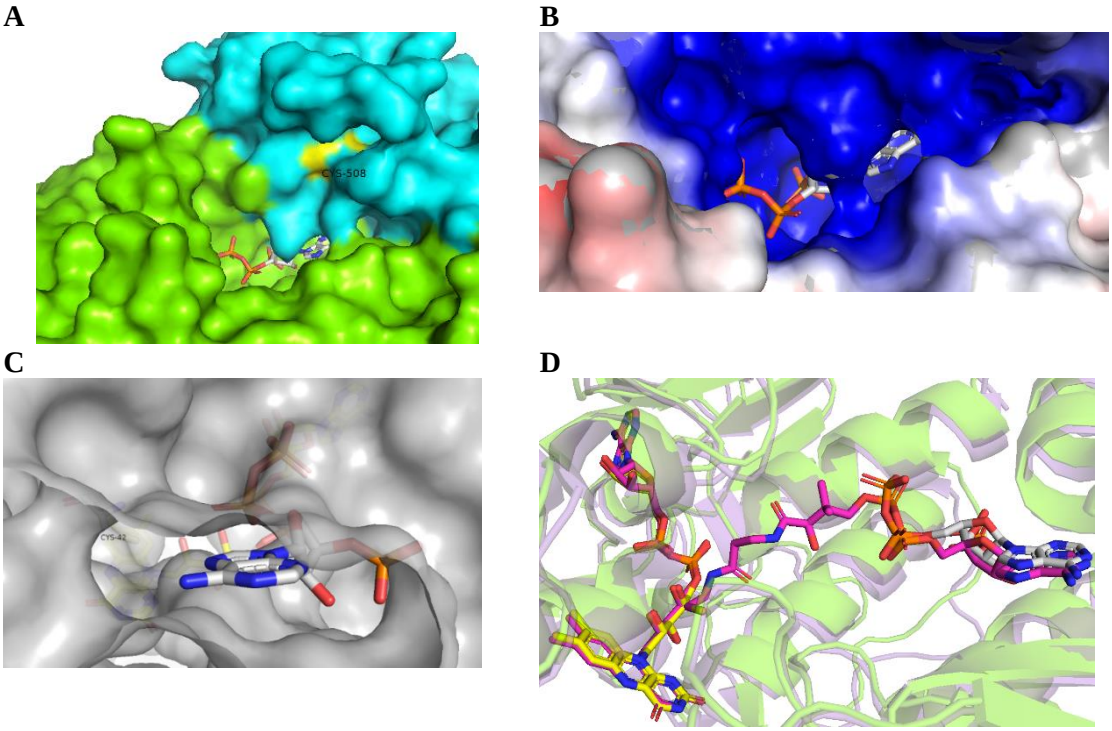


Figure 3.7 – (A) Surface representation of *Ef*CoAPR coenzyme A cavity pocket colored by monomer (**green** and **blue**). Cys-508 labeled and colored in **yellow**. (B) Surface view of CoA pocket showing the surrounding electrostatic potential. (C) Inside view of the CoA tunnel to FAD and Cys-42 with 20 % transparency for the surface representation of *Ef*CoAPR residues. (D) Active site view with the alignment of *Ef*CoAPR and *Ba*CoADR-RHD models [116]. Cartoon representation of *Ef*CoAPR (**Chartreuse**) and *Ba*CoADR-RHD (**purple**) [116] with 60 % transparency. *Ef*CoAPR carbon atoms colored yellow for FAD and white for PAP (following Wallen et al. 2009 [116] color code). *Ba*CoADR-RHD carbon atoms colored **purple**. Images generated in *PyMOL* [76]

This CoA binding cleft is located at the intersection of CDR and RHD domains of opposite subunits (Figure 3.7.A). The adenine end of the PAP is deeply buried in the RDH domain at ~10.7 Å from Cys-508. The PAP is well-ordered, suggesting the presence of non-covalent interactions with both domains, e.g., Arg-19 and/or Arg-306 NH_2 and Lys431 Nζ. Moreover, in Figure 3.7.B, the electropositive surface near Cys-508 could explain an electrostatic attraction of the negatively charged PAP to this pocket, as described for *S. loihica* CDR-RHD [115]. This interaction may help to hold CoA the binding pocket, even after catalysis. The pantothenate and cysteamine CoA moieties (Figure 3.7.A) are disordered, with no electron density implying high flexibility. This previous observation is also reported for *Ba*CoADR-RHD, Wallen et al. (2009),

when detected a precedent for the pantothenate to act as a swinging arm in a wide variety of enzymes and its flexibility might be essential for the catalytic mechanism [116].

Figure 3.7.C shows a shallow tunnel with an approximate depth of 24.8 Å, connecting the aperture of the solvent-exposed catalytic Cys-508 and CoA cavity located in the RHD domain, to catalytic Cys-42 and FAD groove, located in the CDR domain of the opposite monomer (Figure 3.7.A). This tunnel is likely an essential feature for catalysis as it connects the FAD and CoA binging pockets to the solvent. The FAD is also in contact with the solvent, although through three different channels of much smaller dimensions (~10x smaller) compared to the CoA binding cleft. Cys-42 is oriented in the *re* face of the FAD isoalloxazine ring. FAD shows clear density in the electron density map and is well-ordered establishing H-bonds and ionic interactions with the surrounding amino acid residues.

Figure 3.7.D demonstrates an almost perfect superposition between *Ef*CoAPR and *Ba*CoADR-RHD cofactors, although CoA could not be fully modelled in *Ef*CoAPR, likely due to a high degree of flexibility enabled by the broad and shallow pocket presented in Figure 3.7.C.

At present, two additional data sets are under analysis, one of them collected from a crystal soaked with CoA. Moreover, a non-hydrolyzable CoA-analogue is being synthesized, and *Ef*CoAPR C42S, C508S, and C42S/C508S variants are being produced at Giedroc Lab. We hope that these follow-up studies can shed some light into the mechanism of action for the catalytic activity of *Ef*CoAPR.

3.5. Concluding Remarks

- We have successfully crystallized the full-length CoAPR enzyme from *E. faecalis*; moreover, we could develop a seeding crystallization strategy that yielded good quality diffraction crystals marking a significant improvement from the initial crystals (~8 Å to 2.05 Å);
- We have determined the first full-length crystallographic structure of *Ef*CoAPR to 2.05 Å;
- The 3D structure revealed a similar overall fold to CoADR-RHD from *B. anthracis* (PDB ID: 3ICS), with RMSD of ca. 0.98 Å. These proteins crystallize as dimers and each monomer has two domains, CDR and RHD, in close contact with the neighboring protomer;
- Each *Ef*CoAPR monomer is folded to fifteen α-helices, six-stranded parallel β-sheet and seven-stranded anti-parallel β-sheets;
- *Ef*CoAPR rendered clear electron density for FAD cofactor, and partially for CoA cofactor;
- The catalytic Cys-42 is on the *re* side of FAD cofactor, and, CoA is in the vicinity of Cys-508 of the neighboring protomer;
- A shallow tunnel connects the two catalytic cysteines, CoA and FAD cofactors; moreover, Cys-508 is solvent-accessible, which most likely is an important feature for catalysis;
- The proximity between domains of opposite protomers might also be essential for biological activity;
- The crystallographic structure of *Ef*CoAPR marks an important step towards a better understanding of this enzyme catalytic mechanism.

Chapter IV

Overall Concluding Remarks

4. Concluding remarks

4.1. *Sa*CstB

The work herein presented aimed at determining several crystallographic structures of CstB from *Staphylococcus aureus*, a joint project with Professor Dr. David P. Giedroc, Indiana University – Bloomington, USA.

Several methodologies were employed that allowed us to surpass the main problem: acquire electron density corresponding to the enzyme rhodanese domain. With that in mind, we set up a strategy encompassing the change of protein buffer components, optimization of crystallization screening conditions, and crystallization methodology (vapour diffusion) with co-crystallization and crystal-soaking with several ligands.

These efforts were most successful and we could determine the first full-length Apo structure for the WT of *Sa*CstB to 2.69 Å resolution. Moreover, we have determined three more X-ray structures, corresponding to the single variants C201S and C408S, and the double variant C201S/C408S to 3.19 Å, 2.4 Å and 2.2 Å resolution, respectively.

The non-heme iron present at the active site is coordinated by, His-56, His-119 and Asp-145 and two additional waters. Interestingly, the catalytic Cys-201 is in close proximity to the iron ion (3.7 Å), creating an architecture that is likely important for catalysis.

Several sulfur-based ligands were tested (e.g., SO_3^{2-}, $Na_2S_2O_3$, GSH, CysSS, GSSH, ...) in co-crystallization and soaking experiments for both WT and *Sa*CtB variants. As a result of these experiments, it was possible to visualize one SO_3^{2-} ion locked in the rhodanese domain. Sulfite acts both as reaction substrate and product of this domain, depending on the catalytic half-reaction being performed. Noteworthy, sulfite is only present in the variant *Sa*CstB structures, and it is too far distanced from the Cys-408 S^{γ} (~3.47 Å) to establish a SS-bond.

Moreover, amino acid sequence and structural alignments of *Sa*CstB with *S. aureus* metallo-β-lactamase (*Sa*MβL, PDB ID: 3R2U), *A. acidocaldarius* β-lactamase and rhodanese fusion protein (*Aa*βL-Rhod, PDB ID: 3TP9) and *H. sapiens* hETHE1 (PDB ID: 4CHL), provided new insights about their structural similarities and differences.

Overall, we believe that the crystallization structures determined within the timeline of the present work mark a crucial step towards a better understanding of this enzyme's catalytic mechanism, as well as for understanding H_2S homeostasis in this very important pathogen.

4.2. *Ef*CoAPR

The studies herein conducted, intended to characterize the X-ray structure of WT of CoAPR from *Enterococcus faecalis*, a collaboration with Professor Dr. David P. Giedroc, Indiana University – Bloomington, USA.

Within the timeframe of this project, we have tested a vast array of possible crystallization conditions and procedures. However, the first *Ef*CoAPR crystals were of poor quality and diffracted to only ~8 – 10 Å. Remarkably, we used these crystals for seeding experiments and this procedure was as a silver bullet for the project: not only we could measure good diffraction data, we were also able to determine the full-length *Ef*CoAPR crystallographic structure by molecular replacement, which was refined to 2.05 Å resolution.

As expected, *Ef*CoAPR overall folding was very similar to CoADR-RHD from *B. anthracis* (PDB ID 3ICS). *Ef*CoAPR has a dimeric arrangement with two domains for each monomer, CDR and RHD, which are in close contact with the domains of the opposite monomer. The electron density maps rendered clear features for both cofactors, FAD and CoA (the latest only partially resolved). Furthermore, *Ef*CoAPR structure depicts a tunnel that connects the two catalytic cysteines, Cys-42 and Cys-508, with CoA and FAD cofactors of the opposite protomers. Moreover, this tunnel is also solvent-accessible which, together with the above features, most likely is relevant for the catalysis to occur.

In this project, we have also compared the amino acid sequences and structures of *Ef*CoAPR with those of *S. loihica* FAD-dependent pyridine nucleotide-disulfide oxidoreductase (PV-4 Npsr) and the amino acid sequence of NAD-dependent persulfide reductase (Nspr) from *A. fulgidus* (PDB IDs: 3NTA and 6PFZ, respectively).

Overall, we believe that the *Ef*CoAPR crystallization structure determined within the timeframe of the present work marks a crucial step towards a better understanding of this enzyme's catalytic mechanism. Together with the crystallographic structures determined for *Sa*CstB, we believe our work is an important contribution towards the understating of H_2S homeostasis in two very important human pathogens.

We believe our structural studies on *Sa*CstB and *Ef*CoAPR represent a landmark to better understand the catalytic mechanisms of these enzymes and their involvement in H_2S homeostasis. A long-term goal would be to contribute for the rational design of novel drugs against these pathogenic bacteria, *Staphylococcus aureus* and *Enterococcus faecalis* (*see future perspectives*).

5. Future Perspectives

The work on these enzymes continues to obtain more functional and structural information on the enzymes of interest. Several ***Ef*CoAPR variants,** namely **C42S**, **C508S** and **C42S/C508S,** are currently being produced at Giedroc Lab and we expect to determine their crystal structures. Chemical compounds are being synthesized at Giedroc Lab, namely **non-hydrolysable substrate analogues**, and we intend to use them in crystallization experiments, expecting to "lock" *Sa*CstB and *Ef*CoAPR proteins in intermediate steps of catalysis. This will give us insights into the catalytic mechanisms of these enzymes.

Methyl-TROSY NMR spectroscopy will be explored at Giedroc Lab to assess interdomain conformational dynamics in various ligand-bound states.

Although not included in this thesis, **EPR studies** are underway for *Sa*CstB. We hope to assess structural changes on the iron-site environment upon incubation with sulfide-based compounds, such as SO_3^{2-}, $Na_2S_2O_3$ and GSH. In addition, preliminary **cryo-EM** data have been collected on the target enzymes. Single particle analysis cryo-EM will be used as an alternative plan if crystallization problems arise upon incubation with the non-hydrolysable analogues. 2D-particle classification may disclose different conformations for these enzymes upon ligand binding, providing a "movie-like" grasp for the mode of action of *Sa*CstB and *Ef*CoAPR.

Fluorescence-based thermal denaturation assays, namely **nanoDSF**, will be employed to identify potential sulfide-based compounds that interact and stabilize the targeted proteins. These assays use intrinsic tryptophan fluorescence to detect and characterize protein-ligand interactions.

In silico methods, such as **molecular docking**, will be implemented to predict the fit of selected compounds into the active site of the target enzymes and study their atomic interactions. The goal is to describe the association mode and predict the association energy for these interactions in order to select strong binding compounds.

The structural data will be combined with kinetic data and activity studies to optimize the **identification of potential drug candidates**. Our ambition is to take this project further and use the structural knowledge to synthesize selective inhibitors by structure-based drug design (SBDD) for these H_2S-regulating enzymes. These efforts will consolidate on finding alternative routes to tackle antibiotic-resistance pathways in human pathogenic bacteria, *Staphylococcus aureus* and *Enterococcus faecalis*. Herein, we plan to access **Diamond's XChem fragment screening facility** with the goal of selecting promising fragments by **fragment-based screenings**, that could develop into potential anti-bacterial drug candidates. The obtained (protein:fragment) crystallographic structures will allow the design of molecules of increased complexity and

inhibitory potential. Finally, the **inhibitory effect** of the selected compounds will be evaluated by **functional assays.** In ***Sa*CstB** with monitorization of O_2 consumption by high-resolution respirometry and production of fluorescent derivates by HPLC with fluorescence detection. The latest making use of a fluorescent agent, such as, ammonium 7-fluoro-2,1,3-benzoxadiazole-4-sulfonate (SBD-F), which reacts selectively with thiols; In ***Ef*CoAPR** with monitorization of NAD(P)H oxidation by UV-Vis absorption spectroscopy.

6. References

(1) U.S. National Library of Medicine. PubChem. https://pubchem.ncbi.nlm.nih.gov/ (accessed Oct 18, 2020).

(2) Walsh, B. J. C.; Brito, J. A.; Giedroc, D. P. Hydrogen Sulfide Signaling and Enzymology. In *Reference Module in Chemistry, Molecular Sciences and Chemical Engineering Elsevier*; Elsevier Inc., 2020. https://doi.org/10.1016/B978-0-12-409547-2.14699-2.

(3) Beauchamp, R. O. et al. A Critical Review Of The Literature On Hydrogen Sulfide Toxicity. *CRC Crit. Rev. Toxicol.* **1984**, *13* (1), 25–29. https://doi.org/10.3109/10408448409029321.

(4) Scheele, C. *Chemical Observations and Experiments on Air and Fire*; J. Johnson, 1780.

(5) Meyer, B. Chapter 12 - Industrial Uses of Sulfur and Its Compounds. In *Sulfur, Energy, and Environment*; Elsevier, 1977; pp 279–290. https://doi.org/10.1016/b978-0-444-41595-0.50015-2.

(6) Wang, R. Two's Company, Three's a Crowd: Can H2S Be the Third Endogenous Gaseous Transmitter? *FASEB Journal. Res. Commun.* **2002**, *16* (13), 1792–1798. https://doi.org/10.1007/978-1-59259-806-9_1.

(7) Roth, S. H.; Skrajny, B.; Reiffenstein, R. J. Alteration of the Morphology and Neurochemistry of the Developing Mammalian Nervous System By Hydrogen Sulphide. *Clin. Exp. Pharmacol. Physiol.* **1995**, *22* (5), 379–380. https://doi.org/10.1111/j.1440-1681.1995.tb02024.x.

(8) Zanardo, R. C. O. et al. Hydrogen Sulfide Is an Endogenous Modulator of Leukocyte-mediated Inflammation. *FASEB J.* **2006**, *20* (12), 2118–2120. https://doi.org/10.1096/fj.06-6270fje.

(9) Teague, B.; Asiedu, S.; Moore, P. K. The Smooth Muscle Relaxant Effect of Hydrogen Sulphide in Vitro: Evidence for a Physiological Role to Control Intestinal Contractility. *Br. J. Pharmacol.* **2002**, *137* (2), 139–145. https://doi.org/10.1038/sj.bjp.0704858.

(10) Zhao, W. et al. The Vasorelaxant Effect of H2S as a Novel Endogenous Gaseous KATP Channel Opener. *EMBO J.* **2001**, *20* (21), 6008–6016. https://doi.org/10.1093/emboj/20.21.6008.

(11) Kimura, H.; Shibuya, N.; Kimura, Y. Hydrogen Sulfide Is a Signaling Molecule and a Cytoprotectant. *Antioxidants Redox Signal.* **2012**, *17* (1), 45–57. https://doi.org/10.1089/ars.2011.4345.

(12) Malone Rubright, S. L.; Pearce, L. L.; Peterson, J. Environmental Toxicology of Hydrogen Sulfide. *Nitric Oxide. Elsevier* **2017**, *71*, 1–13. https://doi.org/10.1016/j.niox.2017.09.011.

(13) Kilburn, K. H. Case Report: Profound Neurobehavioral Deficits in an Oil Field Worker Overcome by Hydrogen Sulfide. *Am. J. Med. Sci.* **1993**, *306* (5), 301–305. https://doi.org/10.1097/00000441-199311000-00005.

(14) Snyder, J. W. et al. Occupational Fatality and Persistent Neurological Sequelae after Mass Exposure to Hydrogen Sulfide. *Am. J. Emerg. Med.* **1995**, *13* (2), 199–203. https://doi.org/10.1016/0735-6757(95)90094-2.

(15) Guidotti, T. L. Hydrogen Sulphide. *Occup. Med.* **1996**, *46* (5), 367–371. https://doi.org/10.1093/occmed/46.5.367.

(16) Guidotti, T. L. Hydrogen Sulfide: Advances in Understanding Human Toxicity. *Int. J. Toxicol.* **2010**, *29* (6), 569–581. https://doi.org/10.1177/1091581810384882.

(17) Kilburn, K. H. Exposure to Reduced Sulfur Gases Impairs Neurobehavioral. *South. Med. J.* **1997**, *90* (10), 997–1006. https://doi.org/10.1097/00007611-199710000-00006.

(18) Occupational Safety and Health Administration. United States department of labor. https://www.osha.gov/ (accessed Jan 20, 2021).

(19) Dorman, D. C. et al. Cytochrome Oxidase Inhibition Induced by Acute Hydrogen Sulfide Inhalation: Correlation with Tissue Sulfide Concentrations in the Rat Brain, Liver, Lung, and Nasal Epithelium. *Toxicol. Sci. Soc. Toxicol.* **2002**, *65* (1), 18–25. https://doi.org/10.1093/toxsci/65.1.18.

(20) Khan, S.; O'Brien, P. J. Role of the Cellular Redox State in Modulating Acute Ethanol Toxicity in Isolated Hepatocytes. *Clin. Biochem.* **1999**, *32* (7), 585–589. https://doi.org/10.1016/S0009-9120(99)00059-4.

(21) Smith, L.; Kruszyna, H.; Smith, R. P. The Effect of Methemoglogin on the Inhibition of Cytochrome c Oxidase by Cyanide, Sulfide or Azide. *Biochem. Pharmacol.* **1977**, *26* (23), 2247–2250. https://doi.org/10.1016/0006-2952(77)90287-8.

(22) Truong, D. H. et al. Molecular Mechanisms of Hydrogen Sulfide Toxicity. *Drug Metab. Rev.* **2006**, *38* (4), 733–744. https://doi.org/10.1080/03602530600959607.

(23) Nelp, M. T. et al. Potent Activation of Indoleamine 2,3-Dioxygenase by Polysulfides. *J. Am. Chem. Soc.* **2019**, *141* (38), 15288–15300. https://doi.org/10.1021/jacs.9b07338.

(24) Hosoki, R.; Matsuki, N.; Kimura, H. The Possible Role of Hydrogen Sulfide as an Endogenous Smooth Muscle Relaxant in Synergy with Nitric Oxide. *Biochem. Biophys. Res. Commun.* **1997**, *237* (3), 527–531. https://doi.org/10.1006/bbrc.1997.6878.

(25) Wang, R. Hydrogen Sulfide: The Third Gasotransmitter in Biology and Medicine. *Antioxidants Redox Signal.* **2010**, *12* (9), 1061–1064. https://doi.org/10.1089/ars.2009.2938.

(26) Panthi, S. et al. Review Article. Physiological Importance of Hydrogen Sulfide : Emerging Potent Neuroprotector and Neuromodulator. Review Article. *Oxid. Med. Cell. Longev.* **2016**, *2016*, 1–11. https://doi.org/https://doi.org/10.1155/2016/9049782.

(27) Giedroc, D. P. A New Player in Bacterial Sulfide-Inducible Transcriptional Regulation. *Mol. Microbiol. John Wiley Sons Ltd.* **2017**, *105* (3), 347–352. https://doi.org/10.1111/mmi.13726.

(28) Pfennig, N. The Phototrophic Bacteria and Their Role in the Sulfur Cycle. *Plant Soil* **1975**, *43*, 1–16. https://doi.org/https://doi.org/10.1007/BF01928472.

(29) Brune, D. C. Sulfur Oxidation by Phototrophic Bacteria. Review. *Biochim. Biophys. Acta - Bioenerg.* **1989**, *975* (2), 189–221. https://doi.org/10.1016/S0005-2728(89)80251-8.

(30) Toliver-Kinsky, T. H2S, a Bacterial Defense Mechanism against the Host Immune Response. *Infect. Immun.* **2019**, *87* (1). https://doi.org/10.1128/IAI.00272-18.

(31) Shatalin, K. et al. H2S: A Universal Defense Against Antibiotics in Bacteria. *Science (80-.).* **2011**, *334* (6058), 986–990. https://doi.org/10.1126/science.1209855.

(32) Hickman, A. B.; Davies, D. R. Principles of Macromolecular X-Ray. *Curr. Protoc. Protein Sci.* **1997**, *10:*, 17.3.1-17.3.15. https://doi.org/10.1002/0471140864.ps1703s10.

(33) Kendrew, J. C. et al. A Three-Dimensional Model of the Myoglobin Molecule Obtained by x-Ray Analysis. *Nature* **1958**, *181* (4610), 662–666. https://doi.org/10.1038/181662a0.

(34) Wlodawer, A. et al. Protein Crystallography for Non-Crystallographers, or How to Get the Best (but Not More) from Published Macromolecular Structures. *FEBS J.* **2008**, *275* (1),

1–21. https://doi.org/10.1111/j.1742-4658.2007.06178.x.

(35) Creative Biostructure. Comparison of Crystallography, NMR and EM https://www.creative-biostructure.com/comparison-of-crystallography-nmr-and-em_6.htm (accessed Oct 13, 2021).

(36) Ziegler, S. J.; Mallinson, S. J. B.; St. John, P. C.; Bomble, Y. J. Advances in Integrative Structural Biology: Towards Understanding Protein Complexes in Their Cellular Context. *Comput. Struct. Biotechnol. J.* **2021**, *19*, 214–225. https://doi.org/10.1016/j.csbj.2020.11.052.

(37) Marion, D. An Introduction to Biological NMR Spectroscopy. *Mol. Cell. Proteomics* **2013**, *12* (11), 3006–3025. https://doi.org/10.1074/mcp.O113.030239.

(38) Jacques, D. A.; Trewhella, J. Small-Angle Scattering for Structural Biology - Expanding the Frontier While Avoiding the Pitfalls. *Protein Sci.* **2010**, *19* (4), 642–657. https://doi.org/10.1002/pro.351.

(39) Röntgen, W. C. On a New Kind of Rays. *Science*, **1896**, *3* (56), 227–231.

(40) Friedrich, W.; Knipping, P.; M., L. *Interferenz-Erscheinungen Bei Röntgenstrahlen.*; Sitz Bayer Akad Wiss, 1912.

(41) Jaskolski, M.; Dauter, Z.; Wlodawer, A. A Brief History of Macromolecular Crystallography, Illustrated by a Family Tree and Its Nobel Fruits. *FEBS J.* **2014**, *281* (18), 3985–4009. https://doi.org/10.1111/febs.12796.

(42) Bragg, W. L.; A, P. R. S. L. The Structure of Some Crystals as Indicated by Their Diffraction of X-Rays. *Proc. R. Soc. London. Ser. A, Contain. Pap. a Math. Phys. Character* **1913**, *89* (610), 248–277. https://doi.org/10.1098/rspa.1913.0083.

(43) Romão, M. J. Cristalografia de Proteinas: Metodologias e Aplicações Em Bioquímica. *Bol. Biotecnol.* **1996**, *53*, 18–36.

(44) Llamas-Saiz, A. L.; van Raaij, M. J. X-Ray Crystallography of Biological Macromolecules: Fundamentals and Applications. In *Proteins in Solution and at Interfaces*; Ruso, J. M., Piñeiro, Á., Eds.; Wiley. John Wiley & Sons, Inc., 2013; pp 1–22. https://doi.org/10.1002/9781118523063.ch1.

(45) Drenth, J. X-Ray Diffraction : Principles. *Encycl. LIFE Sci.* **2003**, 1–8. https://doi.org/10.1038/npg.els.0002721.

(46) Brito, J. A.; Archer, M. Practical Approaches to Biological Inorganic Chemistry. In *X-ray Crystallography*; Elsevier Inc., 2013; pp 217–255. https://doi.org/10.1016/B978-0-444-56351-4.00009-9.

(47) Pichlo, C. et al. Production, Crystallization and Structure Determination of C. Difficile PPEP-1 via Microseeding and Zinc-SAD. *J. Vis. Exp.* **2016**, No. 118, 1–12. https://doi.org/10.3791/55022.

(48) McPherson, A.; Gavira, J. A. Introduction to Protein Crystallization. *Acta Crystallogr. Sect. FStructural Biol. Commun.* **2014**, *70* (1), 2–20. https://doi.org/10.1107/S2053230X13033141.

(49) Kwan, T. O. C. et al. Selection of Biophysical Methods for Characterisation of Membrane Proteins. *Int. J. Mol. Sci.* **2019**, *20* (10:2605). https://doi.org/10.3390/ijms20102605.

(50) Theveneau, P. et al. The Upgrade Programme for the Structural Biology Beamlines at the European Synchrotron Radiation Facility-High Throughput Sample Evaluation and Automation. *J. Phys. Conf. Ser.* **2013**, *425* (012001). https://doi.org/10.1088/1742-6596/425/1/012001.

(51) ESRF. Synchrotron science. https://www.esrf.fr/about/synchrotron-science (accessed Oct 11, 2020).

(52) Lamzin, V. S.; Wilson, K. S. Automated Refinement of Protein Models. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **1993**, *D49*, 129–147. https://doi.org/10.1107/s0907444992008886.

(53) Martz, E.; Theis, K. et al. Resolution. Proteopedia https://proteopedia.org/wiki/index.php/Resolution#Electron_Density_Map_vs._Resolution (accessed Oct 17, 2020).

(54) Karplus, P. A.; Diederichs, K. Assessing and Maximizing Data Quality in Macromolecular Crystallography. *Curr. Opin. Struct. Biol.* **2015**, *34*, 60–68. https://doi.org/10.1016/j.sbi.2015.07.003.

(55) Ochoa, J. M. et al. Structural Characterization of Hexameric Shell Proteins from Two Types of Choline-Utilization Bacterial Microcompartments. *Acta Crystallogr. Sect. F Struct. Biol. Commun.* **2021**, *F77*, 275–285. https://doi.org/10.1107/S2053230X21007470.

(56) Murakawa, T. et al. Microcrystal Preparation for Serial Femtosecond X-Ray Crystallography of Bacterial Copper Amine Oxidase. *Acta Crystallogr. Sect. F Struct. Biol. Commun.* **2021**, *F77*, 356–363. https://doi.org/10.1107/S2053230X21008967.

(57) Rechiche, O.; Lee, T. V.; Lott, J. S. Structural Characterization of Human Peptidyl-Arginine Deiminase Type III by X-Ray Crystallography. *Acta Crystallogr. Sect. F Struct. Biol. Commun.* **2021**, *D77*, 334–340. https://doi.org/10.1107/S2053230X21009195.

(58) Evans, P.; McCoy, A. An Introduction to Molecular Replacement. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2008**, *64* (Pt 1), 1–10. https://doi.org/10.1107/S0907444907051554.

(59) Jumper, J.; Evans, R.; Pritzel, A. et al. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**, *596*, 583–589. https://doi.org/10.1038/s41586-021-03819-2.

(60) Hendrickson, W. A.; Horton, J. R.; LeMaster, D. M. Selenomethionyl Proteins Produced for Analysis by Multiwavelength Anomalous Diffraction (MAD): A Vehicle for Direct Determination of Three Dimensional Structure. *EMBO J.* **1990**, *9* (5), 1665–1672. https://doi.org/10.1002/j.1460-2075.1990.tb08287.x.

(61) Green, D. W.; Ingram, V. M.; Perutz, M. F. The Structure of Haemoglobin - IV. Sign Determination by the Isomorphous Replacement Method. *Proc. R. Soc. Lond.* **1954**, *225* (1162), 287–307. https://doi.org/https://doi.org/10.1098/rspa.1954.0203.

(62) Hendrickson, W. A. Determination of Macromolecular Structures from Anomalous Diffraction of Synchrotron Radiation. *Science (80-.).* **1991**, *254* (5028), 51–58. https://doi.org/10.1126/science.1925561.

(63) Terwilliger, T. C. et al. Iterative Model Building, Structure Refinement and Density Modification with the PHENIX AutoBuild Wizard. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2008**, *D64*, 61–69. https://doi.org/10.1107/S090744490705024X.

(64) Brünger, A. Free R Value: A Novel Statistical Quantity for Assessing the Accuracy of Crystal Structures. *Nature* **1992**, *355*, 472–475. https://doi.org/https://doi.org/10.1038/355472a0.

(65) Wlodawer, A. et al. Protein Crystallography for Non-Crystallographers, or How to Get the Best (but Not More) from Published Macromolecular Structures. *FEBS J.* **2007**, *275* (1), 1–21.

(66) Rhodes, G. *Crystallography Made Crystals Clear. A Guide For Users Of Macromolecular*

Models, third.; Hayhurst, J., Ed.; Elsevier Inc., 2006.

(67) Mulligan, M. E. et al. Methicillin-Resistant Staphylococcus Aureus: A Consensus Review of the Microbiology, Pathogenesis, and Epidemiology with Implications for Prevention and Management. *Am. J. Med.* **1993**, *94* (3), 313–328. https://doi.org/10.1016/0002-9343(93)90063-u.

(68) Kourtis, A. P. et al. Vital Signs: Epidemiology and Recent Trends in Methicillin-Resistant and in Methicillin-Susceptible Staphylococcus Aureus Bloodstream Infections — United States. *MMWR. Morb. Mortal. Wkly. Rep.* **2019**, *68* (9), 214–219. https://doi.org/10.15585/mmwr.mm6809e1.

(69) Healthline. Healthline Media a Red Ventures Company. https://www.healthline.com/ (accessed Feb 11, 2020).

(70) Shatalin, K. et al. H2S: A Universal Defense Against Antibiotics in Bacteria. *Science.* **2011**, *334* (6058), 986–990. https://doi.org/10.1126/science.1209855.

(71) Shen, J. et al. Staphylococcus Aureus CstB Is a Novel Multidomain Persulfide Dioxygenase-Sulfurtransferase Involved in Hydrogen Sulfide Detoxification. *Biochem. Am. Chem. Soc.* **2015**, *54* (29), 4542–4554.

(72) Technische Universität Braunschweig. BRENDA. The Comprehensive Enzyme Information System. https://www.brenda-enzymes.org (accessed Nov 17, 2020).

(73) Pettinati, I. et al. Crystal Structure of Human Persulfide Dioxygenase: Structural Basis of Ethylmalonic Encephalopathy. *Hum. Mol. Genet.* **2015**, *24* (9), 2458–2469. https://doi.org/10.1093/hmg/ddv007.

(74) Adams, P. D. et al. PHENIX: A Comprehensive Python-Based System for Macromolecular Structure Solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2010**, *66* (2), 213–221. https://doi.org/10.1107/S0907444909052925.

(75) Winn, M. D. et al. Overview of the CCP4 Suite and Current Developments. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2011**, *D67*, 235–242. https://doi.org/10.1107/S0907444910045749.

(76) The PyMol Molecular Graphics System. Version 2.4.1. Schrödinger, LLC.

(77) Molecular Dimensions. PACT premier. https://www.moleculardimensions.com/products/2384-PACT-empremierem/ (accessed Feb 10, 2021).

(78) Flot, D. et al. The ID23-2 Structural Biology Microfocus Beamline at the ESRF. *J. Synchrotron Radiat.* **2010**, *17*, 107–118. https://doi.org/10.1107/S0909049509041168.

(79) Von Stetten, D. et al. ID30A-3 (MASSIF-3) - A Beamline for Macromolecular Crystallography at the ESRF with a Small Intense Beam. *J. Synchrotron Radiat.* **2020**, *27*, 844–851. https://doi.org/10.1107/S1600577520004002.

(80) Mueller, U.; Förster, R.; Hellmig, M. et al. The Macromolecular Crystallography Beamlines at BESSY II of the Helmholtz-Zentrum Berlin: Current Status and Perspectives. *Eur. Phys. J. Plus* **2015**, *130* (7), 141–151. https://doi.org/10.1140/epjp/i2015-15141-2.

(81) Kabsch, W. XDS. *Acta Crystallogr. D. Biol. Crystallogr.* **2010**, *66* (Pt 2), 125–132. https://doi.org/10.1107/S0907444909047337.

(82) Evans, P. R. An Introduction to Data Reduction: Space-Group Determination, Scaling and Intensity Statistics. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2011**, *67* (4), 282–292. https://doi.org/10.1107/S090744491003982X.

(83) Evans, P. Scaling and Assessment of Data Quality. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2006**, *62* (Pt 1), 72–82. https://doi.org/10.1107/S0907444905036693.

(84) Tickle, I. J. et al. Staraniso. *Glob. Phasing Ltd, Cambridge, UK*. **2016**.

(85) Vonrhein, C. et al. Data Processing and Analysis with the AutoPROC Toolbox. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2011**, *67* (4), 293–302. https://doi.org/10.1107/S0907444911007773.

(86) French, S.; Wilson, K. On the Treatment of Negative Intensity Observations. *Acta Crystallogr. Sect. A Found. Adv.* **1978**, *A34*, 517–525. https://doi.org/https://doi.org/10.1107/S0567739478001114.

(87) Yeates, T. O. Simple Statistics for Intensity Data from Twinned Specimens. *Acta Crystallogr. Sect. A Found. Adv.* **1988**, *A44*, 142–144. https://doi.org/10.1107/S0108767387009632.

(88) Padilla, J. E.; Yeates, T. O. A Statistic for Local Intensity Differences: Robustness to Anisotropy and Pseudo-Centering and Utility for Detecting Twinning. *Acta Crystallogr. - Sect. D Biol. Crystallogr.* **2003**, *D59*, 1124–1130. https://doi.org/10.1107/S0907444903007947.

(89) Dauter, Z. Estimation of Anomalous Signal in Diffraction Data. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2006**, *D62*, 867–876. https://doi.org/10.1107/S0907444906023481.

(90) Brünger, A. T. Free R Value: Cross-Validation in Crystallography. *Methods Enzymol.* **1997**, *277*, 366–396. https://doi.org/10.1016/S0076-6879(97)77021-6.

(91) Brünger, A. Free R Value: A Novel Statistical Quantity for Assessing the Accuracy of Crystal Structures. *Nature* **1992**, *355*, 472–475. https://doi.org/https://doi.org/10.1038/355472a0.

(92) Zwart, P. H.; Grosse-Kunstleve, R. W.; Adams, P. D. Xtriage and Fest: Automatic Assessment of X-Ray Data and Substructure Structure Factor Estimation. *CCP4 Newsl. Winter. Contrib. 7.* **2005**.

(93) McCoy, A. J. et al. Phaser Crystallographic Software. *J. Appl. Crystallogr.* **2007**, *40* (4), 658–674. https://doi.org/10.1107/S0021889807021206.

(94) Zhang Lab. I-TASSER. Protein Structure & Function Predictions https://zhanglab.ccmb.med.umich.edu/I-TASSER/.

(95) Yang, J. et al. The I-TASSER Suite: Protein Structure and Function Prediction. *Nat. Methods* **2015**, *12*, 7–8. https://doi.org/10.1038/nmeth.3213.

(96) Roy, A.; Kucukural, A.; Zhang, Y. I-TASSER: A Unified Platform for Automated Protein Structure and Function Prediction. *Nat. Protoc.* **2010**, *5* (4), 725–738. https://doi.org/10.1038/nprot.2010.5.

(97) Zhang, Y. I-TASSER Server for Protein 3D Structure Prediction. *BMC Bioinformatics* **2008**, *9* (60), 1–8. https://doi.org/10.1186/1471-2105-9-40.

(98) Afonine, P. V. et al. Towards Automated Crystallographic Structure Refinement with Phenix.Refine. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2012**, *68*, 352–367. https://doi.org/10.1107/S0907444912001308.

(99) Bricogne, G. et al. BUSTER. Version 2.10.4. Cambridge, United Kingdom: Global Phasing Ltd. 2017.

(100) Blanc, E. et al. Refinement of Severely Incomplete Structures with Maximum Likelihood in BUSTER-TNT. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2004**, *D60*, 2210–2221.

https://doi.org/10.1107/S0907444904016427.

(101) Emsley, P.; Cowtan, K. Coot: Model-Building Tools for Molecular Graphics. *Acta Crystallogr.* **2004**, *D60*, 2126–2132. https://doi.org/10.1107/S0907444904019158.

(102) Chen, V. B. et al. MolProbity: All-Atom Structure Validation for Macromolecular Crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2010**, *D66*, 12–21. https://doi.org/10.1107/S0907444909042073.

(103) Gasteiger, E. et al. Protein Identification and Analysis Tools on the ExPASy Server. *Proteomics Protoc. Handbook, Humana Press* **2005**, 571–607. https://doi.org/10.1385/1592598900.

(104) Matthews, B. W. Solvent Content of Protein Crystals. *J. Mol. Biol.* **1968**, *33* (2), 491–497. https://doi.org/10.1016/0022-2836(68)90205-2.

(105) Krissinel, E.; Henrick, K. Inference of Macromolecular Assemblies from Crystalline State. *J. Mol. Biol.* **2007**, *372* (3), 774–797. https://doi.org/10.1016/j.jmb.2007.05.022.

(106) EMBL-EBI. PDBePISA (Proteins, Interfaces, Structures and Assemblies) https://www.ebi.ac.uk/pdbe/pisa/ (accessed Jun 18, 2021).

(107) Krissinel, E.; Henrick, K. Secondary-Structure Matching (SSM), a New Tool for Fast Protein Structure Alignment in Three Dimensions. *Acta Crystallogr.* **2004**, *D60*, 2256–2268. https://doi.org/10.1107/S0907444904026460.

(108) Kabsch, W.; Sander, C. Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen-Bonded and Geometrical Features. *Biopolymers* **1983**, *22* (12), 2577–2637. https://doi.org/https://doi.org/10.1002/bip.360221211.

(109) Holm, L. DALI and the Persistence of Protein Shape. *Protein Sci.* **2020**, *29*, 128–140. https://doi.org/10.1002/pro.3749.

(110) Motl, N. et al. Structural and Biochemical Analyses Indicate That a Bacterial Persulfide Dioxygenase-Rhodanese Fusion Protein Functions in Sulfur Assimilation. *J. Biol. Chem.* **2017**, *292* (34), 14026–14038. https://doi.org/10.1074/jbc.M117.790170.

(111) Chothia, C.; Lesk, A. M. The Relation between the Divergence of Sequence and Structure in Proteins. *EMBO J.* **1986**, *5* (4), 823–826. https://doi.org/https://doi.org/10.1002/j.1460-2075.1986.tb04288.x.

(112) Pei, J.; Kim, B. H.; Grishin, N. V. PROMALS3D: A Tool for Multiple Protein Sequence and Structure Alignments. *Nucleic Acids Res.* **2008**, *36* (7), 2295–2300. https://doi.org/10.1093/nar/gkn072.

(113) Dubin, K.; Pamer, E. G. Enterococci and Their Interactions with the Intestinal Microbiome. *Microbiol. Spectr.* **2014**, *5* (6). https://doi.org/10.1128/9781555819705.ch13.

(114) Shen, J. et al. Hydrogen Sulfide Sensing through Reactive Sulfur Species (RSS) and Nitroxyl (HNO) in Enterococcus Faecalis. *ACS Chem. Biol.* **2018**, *13* (6), 1610–1620. https://doi.org/10.1021/acschembio.8b00230.

(115) Warner, M. D. et al. Characterization of an NADH-Dependent Persulfide Reductase from Shewanella Loihica PV-4: Implications for the Mechanism of Sulfur Respiration via FAD-Dependent Enzymes. *Biochemistry* **2011**, *50* (2), 194–206. https://doi.org/10.1021/bi101232y.

(116) Wallen, J. R.; Mallett, T. C.; Boles, W.; Parsonage, D.; Furdui, C. M.; Karplus, P. A.; Claiborne, A. Crystal Structure and Catalytic Properties of Bacillus Anthracis CoADR-RHD: Implications for Flavin-Linked Sulfur Trafficking. *Biochemistry* **2009**, *48* (40),

9650–9667. https://doi.org/10.1021/bi900887k.

(117) Royal Society of Chemistry. ChemSpider. Search and share chemistry. Version 2021.0.16.0. http://www.chemspider.com/ (accessed Jun 30, 2021).

(118) Shabdar, S. et al. Structural and Kinetic Characterization of Hyperthermophilic NADH-Dependent Persulfide Reductase from Archaeoglobus Fulgidus. *Archaea* **2021**, *2021*, 1–9. https://doi.org/10.1155/2021/8817136.

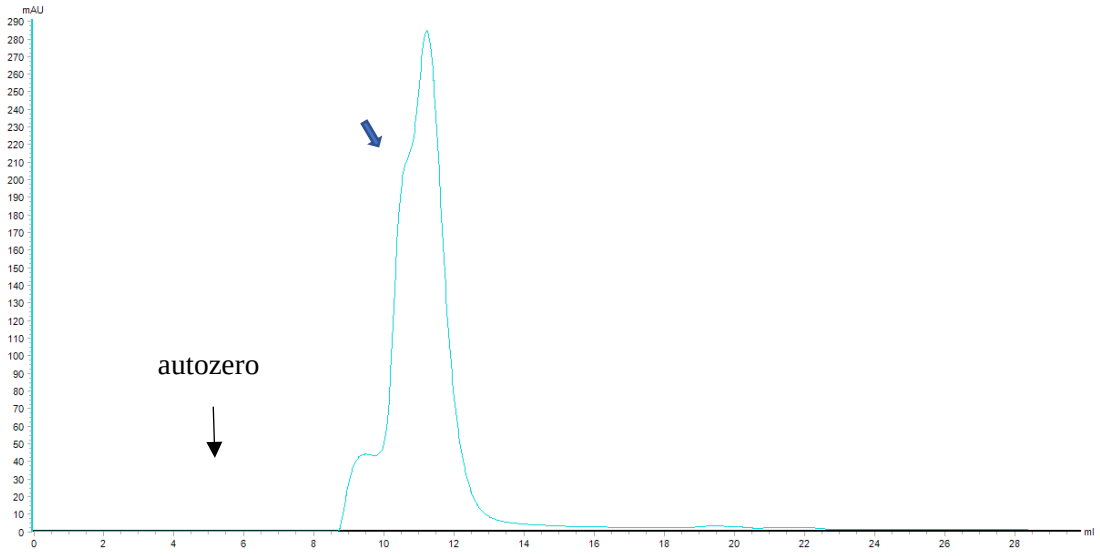
(119) Blaber, M.; Baase, W. A.; Gassner, N.; Matthews, B. W. Alanine Scanning Mutagenesis of the α-Helix 115-123 of Phage T4 Lysozyme: Effects on Structure, Stability and the Binding of Solvent. *J. Mol. Biol.* **1995**, *246* (2), 317–330. https://doi.org/10.1006/jmbi.1994.0087.

(120) Sitbon, E.; Pietrokovski, S. Occurrence of Protein Structure Elements in Conserved Sequence Regions. *BMC Struct. Biol.* **2007**, *7* (3). https://doi.org/10.1186/1472-6807-7-3.

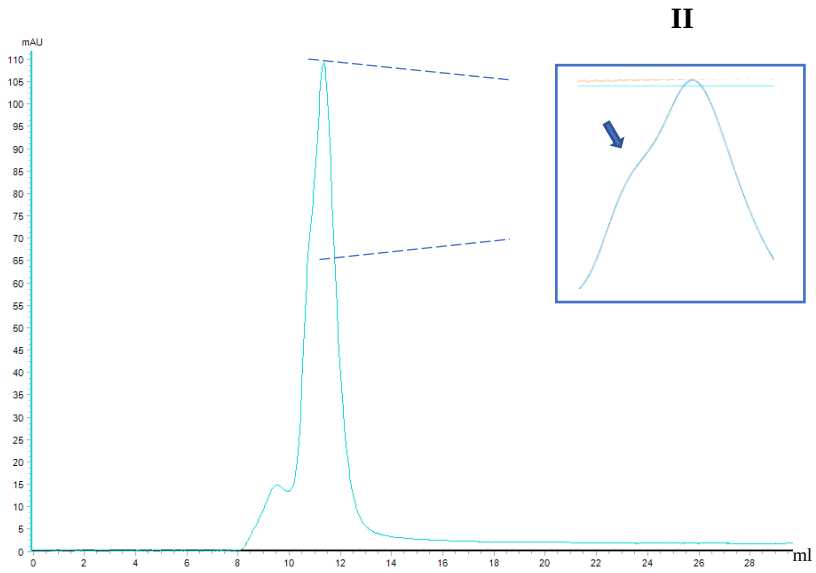
Appendix

HHHHHHDYDI PTTENLYFQG AMGFFKQFYD NHLSQASYLV G**C**QRTGEAII IDPVRDLSKY IEVADSEGLT ITQATETHIH ADFASGIRDV AKRLNANIYV SGEGEDALGY KNMPSKTQFV KHGDIIQVGN VKLEVLHTPG HTPESISFLL TDLGGGSSVP MGLFSGDFIF VGDIGRPDLL EKSVQIKGST EISAKQMYES VQNIKNLPDY VQIWPGHGAG SP**C**GKALGAI PISTIGYEKI NNWAFNEIDE TKFIESLTSN QPAPPHHFAQ MKQVNQFGMN LYQSYDVYPS LDNKRVAFDL RSKEAFHGGH TKGTINIPYN KNFINQIGWY LDFEKDIDLI GDKSTVEKAK HTLQLIGFDK VAGYRLPKSG ISTQSVHSAD MTGKEEHVLD VRNDEEWNNG HLDQAVNIPH GKLLNENIPF NKEDKIYVH**C** QSGVRSSIAV GILESKGFEN VVNIREGYQD FPESLK

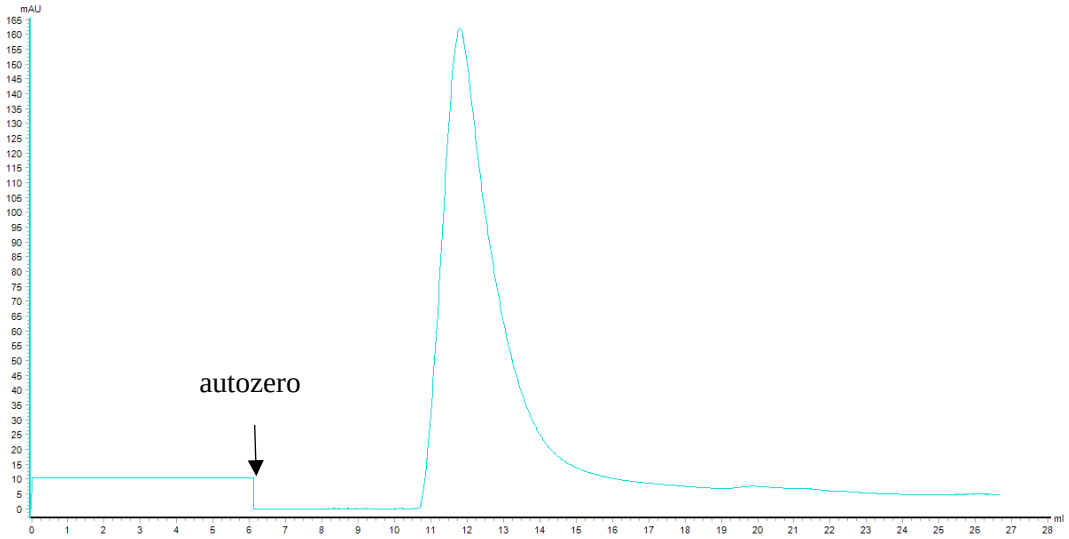
Appendix 1 – *Sa*CstB amino acid sequence. His-tag (orange), PDO domain (green), RHD domain (light-blue) and Rhod domain (dark-blue).



Appendix 2 – Size-exclusion chromatogram obtained for SaCstB purification. The blue line indicates the protein absorbance registered on AKTA Pure connected to a Superdex 200 Increase 10/300 GL column (Cytiva). 450 µl of SaCstB were injected and eluted with 25 mM MES pH 6.0, 50 mM NaCl and 5 % (V/V) glycerol buffer with a 0.4 $ml \cdot min^{-1}$ flow rate. 200 µl fractions were collected. Blue arrow highlights the presence of a "shoulder" in the chromatogram.



Appendix 3 – (I) Size-exclusion chromatogram obtained for *Sa*CstB purification after successive steps of protein concentration and dilution. The blue line indicates the protein absorbance registered on *AKTA Pure* connected to a *Superdex 200 Increase 10/300 GL* column (*Cytiva*). 450 µl of *Sa*CstB were injected and eluted with 25 mM MES pH 6.0, 50 mM NaCl and 5 % (V/V) glycerol buffer with a 0.4 ml·min^{-1} flow rate. 200 µl fractions were collected. Blue arrow (II) highlights the presence of a "shoulder" in the chromatogram.



Appendix 4 –Size-exclusion chromatogram obtained for *Sa*CstB purification after overnight dialysis. The blue line indicates the protein absorbance registered on AKTA Pure connected to a *Superdex 200 Increase 10/300 GL column (Cytiva)*. 450 µl of *Sa*CstB were injected and eluted with 25 mM MES pH 6.0, 50 mM NaCl and 5 % (V/V) glycerol buffer with a 0.4 ml·min^{-1} flow rate. 200 µl fractions were collected.

HHHHHHDYDI PTTENLYFQG AMGKIVIIGG VAGGMSAATR LRRLMEDAEI VVMEKGPFVS FANCGLPYYV SGEIAEREQL LVQTPEALKA RFNLDVRPHH EVVAIDPIEK VITVKHETEI LTEHYDKLIL SPGAKPFVPP ITGLAEAKNV FSLRNVPDLD QIMTALTPET KRAVVIGAGF IGLEMAENLQ KRGLEVTLVE KAPHVLPPLD EEMAAFVKAE LSKNNVQVIT GQSAVAFEEE GQVIRLEDGQ TLASDLTILS VGVQPENTLA VEAGVATGLR GGIVVDEHYQ TNQPDIYAVG DAIVVKQQIT QEDALISLAS PANRQGRQVA DVIAGLERKN QGSIGTAIVR VFDLTAASTG LSERAAKAAG LTTAVVHISG KDHAGYYPGA TDLQLKLVFH PTTGEIYGAQ GIGAKGVDKR IDILATAIKG QLTIFDLPEL EFTYAPPFGS AKDPVNMLGY AAMNLVEGLS ENVQWYELSN ELAKGAVLLD VRNPAERANG QFKNAVSIPL NELRERLEEL DKSTEYIVSC HSGLRSYIAE RMLKQAGISA KNLDGAFALY RMVKPEELEN V

Appendix 5 – *Ef*CoAPR amino acid sequence. His-tag (orange), CDR domain (purple) and RHD domain (light-blue).

Well #	Conc.	Salt	Conc.	Buffer	pH	Conc.	Precipitant
A1			0.1 M	SPG	4.0	25 % w/v	PEG 1500
A2			0.1 M	SPG	5.0	25 % w/v	PEG 1500
A3			0.1 M	SPG	6.0	25 % w/v	PEG 1500
A4			0.1 M	SPG	7.0	25 % w/v	PEG 1500
A5			0.1 M	SPG	8.0	25 % w/v	PEG 1500
A6			0.1 M	SPG	9.0	25 % w/v	PEG 1500
A7	0.2 M	Sodium chloride	0.1 M	Sodium acetate	5.0	20 % w/v	PEG 6000
A8	0.2 M	Ammonium chloride	0.1 M	Sodium acetate	5.0	20 % w/v	PEG 6000
A9	0.2 M	Lithium chloride	0.1 M	Sodium acetate	5.0	20 % w/v	PEG 6000
A10	0.2 M	Magnesium chloride hexahydrate	0.1 M	Sodium acetate	5.0	20 % w/v	PEG 6000
A11	0.2 M	Calcium chloride dihydrate	0.1 M	Sodium acetate	5.0	20 % w/v	PEG 6000
A12	0.01 M	Zinc chloride	0.1 M	Sodium acetate	5.0	20 % w/v	PEG 6000
B1			0.1 M	MIB	4.0	25 % w/v	PEG 1500
B2			0.1 M	MIB	5.0	25 % w/v	PEG 1500
B3			0.1 M	MIB	6.0	25 % w/v	PEG 1500
B4			0.1 M	MIB	7.0	25 % w/v	PEG 1500
B5			0.1 M	MIB	8.0	25 % w/v	PEG 1500
B6			0.1 M	MIB	9.0	25 % w/v	PEG 1500
B7	0.2 M	Sodium chloride	0.1 M	MES	6.0	20 % w/v	PEG 6000
B8	0.2 M	Ammonium chloride	0.1 M	MES	6.0	20 % w/v	PEG 6000
B9	0.2 M	Lithium chloride	0.1 M	MES	6.0	20 % w/v	PEG 6000
B10	0.2 M	Magnesium chloride hexahydrate	0.1 M	MES	6.0	20 % w/v	PEG 6000
B11	0.2 M	Calcium chloride dihydrate	0.1 M	MES	6.0	20 % w/v	PEG 6000
B12	0.01 M	Zinc chloride	0.1 M	MES	6.0	20 % w/v	PEG 6000
C1			0.1 M	PCTP	4.0	25 % w/v	PEG 1500
C2			0.1 M	PCTP	5.0	25 % w/v	PEG 1500
C3			0.1 M	PCTP	6.0	25 % w/v	PEG 1500
C4			0.1 M	PCTP	7.0	25 % w/v	PEG 1500
C5			0.1 M	PCTP	8.0	25 % w/v	PEG 1500
C6			0.1 M	PCTP	9.0	25 % w/v	PEG 1500
C7	0.2 M	Sodium chloride	0.1 M	HEPES	7.0	20 % w/v	PEG 6000
C8	0.2 M	Ammonium chloride	0.1 M	HEPES	7.0	20 % w/v	PEG 6000
C9	0.2 M	Lithium chloride	0.1 M	HEPES	7.0	20 % w/v	PEG 6000
C10	0.2 M	Magnesium chloride hexahydrate	0.1 M	HEPES	7.0	20 % w/v	PEG 6000
C11	0.2 M	Calcium chloride hexahydrate	0.1 M	HEPES	7.0	20 % w/v	PEG 6000
C12	0.01 M	Zinc chloride	0.1 M	HEPES	7.0	20 % w/v	PEG 6000
D1			0.1 M	MMT	4.0	25 % w/v	PEG 1500
D2			0.1 M	MMT	5.0	25 % w/v	PEG 1500
D3			0.1 M	MMT	6.0	25 % w/v	PEG 1500
D4			0.1 M	MMT	7.0	25 % w/v	PEG 1500
D5			0.1 M	MMT	8.0	25 % w/v	PEG 1500
D6			0.1 M	MMT	9.0	25 % w/v	PEG 1500
D7	0.2 M	Sodium chloride	0.1 M	Tris	8.0	20 % w/v	PEG 6000
D8	0.2 M	Ammonium chloride	0.1 M	Tris	8.0	20 % w/v	PEG 6000
D9	0.2 M	Lithium chloride	0.1 M	Tris	8.0	20 % w/v	PEG 6000
D10	0.2 M	Magnesium chloride hexahydrate	0.1 M	Tris	8.0	20 % w/v	PEG 6000
D11	0.2 M	Calcium chloride dihydrate	0.1 M	Tris	8.0	20 % w/v	PEG 6000
D12	0.002 M	Zinc chloride	0.1 M	Tris	8.0	20 % w/v	PEG 6000

Appendix 6 – PACT Premier screen from Molecular Dimensions (1/2) [77].

Well #	Conc.	Salt	Conc.	Buffer	pH	Conc.	Precipitant
E1	0.2 M	Sodium fluoride				20 % w/v	PEG 3350
E2	0.2 M	Sodium bromide				20 % w/v	PEG 3350
E3	0.2 M	Sodium iodide				20 % w/v	PEG 3350
E4	0.2 M	Potassium thiocyanate				20 % w/v	PEG 3350
E5	0.2 M	Sodium nitrate				20 % w/v	PEG 3350
E6	0.2 M	Sodium formate				20 % w/v	PEG 3350
E7	0.2 M	Sodium acetate trihydrate				20 % w/v	PEG 3350
E8	0.2 M	Sodium sulfate				20 % w/v	PEG 3350
E9	0.2 M	Potassium sodium tartrate tetrahydrate				20 % w/v	PEG 3350
E10	0.02 M	Sodium/potassium phosphate				20 % w/v	PEG 3350
E11	0.2 M	Sodium citrate tribasic dihydrate				20 % w/v	PEG 3350
E12	0.2 M	Sodium malonate dibasic monohydrate				20 % w/v	PEG 3350
F1	0.2 M	Sodium fluoride	0.1 M	Bis-Tris propane	6.5	20 % w/v	PEG 3350
F2	0.2 M	Sodium bromide	0.1 M	Bis-Tris propane	6.5	20 % w/v	PEG 3350
F3	0.2 M	Sodium iodide	0.1 M	Bis-Tris propane	6.5	20 % w/v	PEG 3350
F4	0.2 M	Potassium thiocyanate	0.1 M	Bis-Tris propane	6.5	20 % w/v	PEG 3350
F5	0.2 M	Sodium nitrate	0.1 M	Bis-Tris propane	6.5	20 % w/v	PEG 3350
F6	0.2 M	Sodium formate	0.1 M	Bis-Tris propane	6.5	20 % w/v	PEG 3350
F7	0.2 M	Sodium acetate trihydrate	0.1 M	Bis-Tris propane	6.5	20 % w/v	PEG 3350
F8	0.2 M	Sodium sulfate	0.1 M	Bis-Tris propane	6.5	20 % w/v	PEG 3350
F9	0.2 M	Potassium sodium tartrate tetrahydrate	0.1 M	Bis-Tris propane	6.5	20 % w/v	PEG 3350
F10	0.02 M	Sodium/potassium phosphate	0.1 M	Bis-Tris propane	6.5	20 % w/v	PEG 3350
F11	0.2 M	Sodium citrate tribasic dihydrate	0.1 M	Bis-Tris propane	6.5	20 % w/v	PEG 3350
F12	0.2 M	Sodium malonate dibasic monohydrate	0.1 M	Bis-Tris propane	6.5	20 % w/v	PEG 3350
G1	0.2 M	Sodium fluoride	0.1 M	Bis-Tris propane	7.5	20 % w/v	PEG 3350
G2	0.2 M	Sodium bromide	0.1 M	Bis-Tris propane	7.5	20 % w/v	PEG 3350
G3	0.2 M	Sodium iodide	0.1 M	Bis-Tris propane	7.5	20 % w/v	PEG 3350
G4	0.2 M	Potassium thiocyanate	0.1 M	Bis-Tris propane	7.5	20 % w/v	PEG 3350
G5	0.2 M	Sodium nitrate	0.1 M	Bis-Tris propane	7.5	20 % w/v	PEG 3350
G6	0.2 M	Sodium formate	0.1 M	Bis-Tris propane	7.5	20 % w/v	PEG 3350
G7	0.2 M	Sodium acetate trihydrate	0.1 M	Bis-Tris propane	7.5	20 % w/v	PEG 3350
G8	0.2 M	Sodium sulfate	0.1 M	Bis-Tris propane	7.5	20 % w/v	PEG 3350
G9	0.2 M	Potassium sodium tartrate tetrahydrate	0.1 M	Bis-Tris propane	7.5	20 % w/v	PEG 3350
G10	0.02 M	Sodium/potassium phosphate	0.1 M	Bis-Tris propane	7.5	20 % w/v	PEG 3350
G11	0.2 M	Sodium citrate tribasic dihydrate	0.1 M	Bis-Tris propane	7.5	20 % w/v	PEG 3350
G12	0.2 M	Sodium malonate dibasic monohydrate	0.1 M	Bis-Tris propane	7.5	20 % w/v	PEG 3350
H1	0.2 M	Sodium fluoride	0.1 M	Bis-Tris propane	8.5	20 % w/v	PEG 3350
H2	0.2 M	Sodium bromide	0.1 M	Bis-Tris propane	8.5	20 % w/v	PEG 3350
H3	0.2 M	Sodium iodide	0.1 M	Bis-Tris propane	8.5	20 % w/v	PEG 3350
H4	0.2 M	Potassium thiocyanate	0.1 M	Bis-Tris propane	8.5	20 % w/v	PEG 3350
H5	0.2 M	Sodium nitrate	0.1 M	Bis-Tris propane	8.5	20 % w/v	PEG 3350
H6	0.2 M	Sodium formate	0.1 M	Bis-Tris propane	8.5	20 % w/v	PEG 3350
H7	0.2 M	Sodium acetate trihydrate	0.1 M	Bis-Tris propane	8.5	20 % w/v	PEG 3350
H8	0.2 M	Sodium sulfate	0.1 M	Bis-Tris propane	8.5	20 % w/v	PEG 3350
H9	0.2 M	Potassium sodium tartrate tetrahydrate	0.1 M	Bis-Tris propane	8.5	20 % w/v	PEG 3350
H10	0.02 M	Sodium/potassium phosphate	0.1 M	Bis-Tris propane	8.5	20 % w/v	PEG 3350
H11	0.2 M	Sodium citrate tribasic dihydrate	0.1 M	Bis-Tris propane	8.5	20 % w/v	PEG 3350
H12	0.2 M	Sodium malonate dibasic monohydrate	0.1 M	Bis-Tris propane	8.5	20 % w/v	PEG 3350

4

Appendix 7 – PACT Premier screen from Molecular Dimensions (2/2) [77].