Characterisation of the Transcriptional Regulator CysB in *Neisseria gonorrhoeae*

A thesis

submitted in partial fulfilment

of the requirements for the degree

of

Master of Science (Research) in Cellular and Molecular Biology

at

The University of Waikato

by

Hannah Rose Klaus

2022

Abstract

Neisseria gonorrhoeae is an obligate human pathogen and the causative agent of the sexually transmitted infection gonorrhoea. Rapid emergence of antibiotic resistant strains of *N. gonorrhoeae* has given rise to multi-drug resistant gonorrhoea. Combined with increasing incidence and the most recent resistance event in 2018, there is an urgent need for new antimicrobial treatments for gonorrhoeal infection. *N. gonorrhoeae* has a limited number of transcriptional regulators, one of which, CysB, is essential and regulates genes required for *de novo* cysteine biosynthesis. The cysteine biosynthetic pathway is a promising target for the development of novel antimicrobials. Cysteine plays an important role in protein folding and function, and is the precursor for many thiol-containing molecules within *N. gonorrhoeae* including effective oxidative stress defences such as glutathione. *N. gonorrhoeae* displays unique differences in assimilation of sulfur and cysteine biosynthesis, yet little is known about how this altered pathway is regulated in *N. gonorrhoeae*.

The cysteine regulon consists of the genes required for sulfur transport, assimilation and cysteine biosynthesis. The transcriptional regulator CysB regulates this regulon in other organisms and is essential in *N. gonorrhoeae*. Understanding how *N. gonorrhoeae* regulates gene expression will improve our understanding of the pathogenicity of *N. gonorrhoeae*. In order to understand how CysB mediates gene expression, we need to characterise DNA-binding activity and elucidate the three-dimensional structure. In this thesis, we determine the DNA-binding activity of CysB, its three-dimensional structure and investigate conformational changes of CysB. Our results indicate that CysB regulates genes in the altered cysteine regulon, but that CysB does not respond to the inducer *N*-acetylserine. The three-dimensional structure of CysB is the first reported full-length CysB structure and was solved to 2.73 Å. Analysis of this structure, coupled with small angle scattering data suggest that CysB from *N. gonorrhoeae* lacks the ability to undergo the hypothesised large scale conformational changes upon inducer binding. This indicates that CysB may have an alternate mechanism in *N. gonorrhoeae*.

Collectively, the data presented in this thesis represents key advances in our understanding of how *N. gonorrhoeae* regulates gene expression by characterisation of the transcriptional regulator CysB. Data presented here is the basis for future work to identify other targets within *N. gonorrhoeae* that may be investigated for the development of new antimicrobials for treatment of extensively drug resistant gonorrhoea.

Acknowledgements

I would like to acknowledge first and foremost, my amazing supervisor Dr Joanna Hicks. Jo, you have been nothing but supportive, encouraging and inspiring throughout all the highs and lows of the past few years. Your passion for research has been infectious and your enthusiasm has inspired me to look for the silver linings in every result. I have no doubt my passion for research stems from yours. Thank you for all your help guiding and mentoring me. I am especially grateful for all the amazing opportunities, it is greatly appreciated!

Thank you to Judith for being caring and supportive and always being a friendly and encouraging ear. You have done an amazing job keeping the lab running smoothly. Thank you for bringing your endless baking talents and plentiful shared morning teas into my life!

A special thank you to Professor Vis Arcus and to Keely Oldham for helping me to process my structural data and analyse my structure. Thank you Keely for answering all my many, many, many structure questions and encouraging me when Coot kept crashing…

To all members of the C2 lab family, past and present, thank you! You have all been a super team to be a part of and helped make this experience a wonderful one! Thank you for all your tea and lunch breaks, space to vent when things weren't going so well, ideas and suggestions and celebration when things were going well. I am especially appreciative to have shared so many celebrations with you all, many, many, many shared morning teas and plenty of laughs A special thank you to the Apartyment members for such an exciting experience at QMB and those who travelled to the Synchrotron for your quality humour and such fun experiences!

I would like to thank the University of Waikato and the Waikato Graduate Women Educational Trust for funding me through my research.

To Craig, thank you for your continued encouragement and unwavering support. Thank you for your pep talks and helping me remember why this was important to me. I appreciate you listening to my many rambles and providing perspective when needed. You have been there through the most trying times, and for that, I am forever grateful.

To my family, I could not have come this far in life, and in research without your unwavering support, encouragement and unconditional love. Thank you for always sending me back to uni with renewed purpose, a car plied full of food and a full heart. I appreciate your continued interest in my passions! Thank you to my Nan for always having a calm, relaxing space when I needed time to recharge, and for always sending me back with a full tummy and a full heart!

Table of Contents

List of Figures

List of Tables

List of Equations

Abbreviations

Å	angstrom
°	degrees
°C	degrees celsius
%	percent
ABC	ATP-binding cassette
Abs	absorbance
A/Ala	alanine
AMR	antimicrobial resistance
APS	adenosine 5'-phosphosulfate
AR	antibiotic resistance
ATP	adenosine 5'-triphosphate
bp	base pair
BSA	bovine serum albumin
Cα	α-carbon
CBS	CysB binding site
CDC	Centre for Disease Control
cm	centimetre
CoA	co-enzyme A
CSC	cysteine synthase complex
C-terminal	carboxyl end of protein

C/Cys	cysteine
Da	Dalton
D/Asp	aspartic acid/aspartate
DBD	DNA-binding domain
D/Glu	glutamic acid/glutamate
DNA	deoxyribonucleic acid
DTT	dithiothreitol
EDTA	ethylenediamine tetra acetic acid
EMSA(s)	electrophoretic mobility shift assay(s)
F/Phe	phenylalanine
FPLC	Fast Protein Liquid Chromatography
g	G-force
GCB	gonococcal base medium
G/Gly	glycine
H_2O	water
H/His	histidine
HIV	human immunodeficiency virus
HTH	helix-turn-helix
IMAC	immobilised metal affinity chromatography
I/Ile	isoleucine
IPTG	isopropylthio-β-D-galactosidase

IR	inverted repeat
K/Lys	lysine
Kav	gel phase distribution co-efficient
kb	kilobase
kcal	kilocalorie
KCl	potassium chloride
kDa	kilo Dalton
L	litre
LB	luria bertani
L/Leu	leucine
LLG	log-likelihood gain
LTTR	LysR-type transcriptional regulator
M	molar
m	metre
mM	millimole
mm	millimetre
mAU	milli-absorbance units
MDR	multi-drug resistant
M/Met	methionine
min	minutes
mIR	mutated inverted repeat

mg	milligram
MgCl$_2$	magnesium chloride
mL	millilitre
mol	mole
M$_r$	molecular weight
MQ	milliQ
MX	macromolecular crystallography
MWCO	molecular weight cut off
NaCl	sodium chloride
NADPH	nicotinamide adenine dinucleotide phosphate
NAS	N-acetylserine
ng	nanogram
nl	nanolitre
nm	nanomole
NTD	N-terminal domain
N-terminal	amino terminus of protein
OAS	O-acetylserine
OASS	O-acetylserine sulfhydrylase
Opa	opacity-associated
PAGE	polyacrylamide gel electrophoresis
PAP	phosposadenosylphosphate

PAPS	phosposadenosylphosphosulfate
PCR	polymerase chain reaction
PDB	Protein Data Bank
P/Pro	proline
Q/Gln	glutamine
r.m.s.d.	root mean square deviation
R/Arg	arginine
RD	regulatory domain
Rg	radius of gyration
RMS	root mean square
Rpm	revolutions per minute
s	second(s)
SAXS	small angle X-ray scattering
SAT	serine acetyltransferase
sbp	sulfate binding protein
SDS	sodium dodecyl sulfate
SEC	size exclusion chromatography
S/Ser	serine
STI	sexually transmitted disease
TAE	Tris/Acetate/EDTA
TBE	Tris/Borate/EDTA

TE	Tris/EDTA
TG	Tris/Glycine
TEMED	tetramethylethylenediamine
TFZ	translation function Z score
T/Thr	threonine
tNCS	translational non-crystallographic symmetry
Tris	tris(hydroxymethyl)aminomethane
μg	microgram
μL	microlitre
μM	micromolar
μm	micromole
UV	ultra violet
V	volts
V/Val	valine
v/v	volume per volume
W/Trp	tryptophan
w/v	weight per volume
WAXS	wide angle X-ray scattering
Y/Tyr	tyrosine
WHO	World Health Organisation
wHTH	winged helix-turn-helix

Chapter 1: Introduction

1.1 Introduction

Neisseria gonorrhoeae as an obligate human pathogen, is the causative agent of the sexually transmitted infection (STI) gonorrhoea. Gonorrhoea remains a major public health, social and economic burden as the second most prevalent STI with an annual incidence of 87 million cases worldwide where marginalised populations disproportionately bear the burden (WHO, 2017). Alarmingly the number of cases in Aotearoa New Zealand have doubled from 2014 to 2019 and continue to be on the rise where Māori and Pacific populations are disproportionately affected (ESR, 2015, 2022). Since antibiotic treatment began in the 1930's, *N. gonorrhoeae* has demonstrated a remarkable ability to develop resistance to antimicrobials, having developed resistance to every available class of antibiotic previously used for its treatment, with the latest report of multi-drug resistance in 2018 to the current dual therapy of azithromycin and ceftriaxone (Eyre et al., 2018). This highlights the rapid evolution of resistant gonorrhoea and the need for novel therapeutics to prevent and treat extensively drug resistant strains of *N. gonorrhoeae*.

Targeting amino acid synthesis is a novel and promising route for the development of new therapeutics and antimicrobials. Cysteine biosynthesis and sulfate assimilation pathways are absent in mammals and have garnered increasing interest as potential therapeutics for pathogenic bacteria such as *Salmonella typhimurium* and *Mycobacterium tuberculosis* (Benoni et al., 2016; Campanini et al., 2015; Schnell, Sriram, & Schneider, 2015; Spyrakis et al., 2013). However, the sulfate assimilation pathway is absent in *N. gonorrhoeae* (Hicks & Mullholland, 2018) and the cysteine biosynthetic pathway of *N. gonorrhoeae* is only partially characterised (Oldham, Prentice, Summers, & Hicks, 2022). The transcriptional regulator of such pathways, CysB, is one of the few transcriptional regulators in *N. gonorrhoeae*, is differentially expressed upon gonorrhoeal infection and has been shown to be an essential gene (McClure et al., 2015; Remmele et al., 2014). Therefore, this research investigates the transcriptional regulator CysB and its interactions within the *N. gonorrhoeae* genome to further understand the cysteine biosynthetic pathway in *N. gonorrhoeae* and its potential as a novel therapeutic target.

1.2 Gonorrhoea pathogenesis

N. gonorrhoeae (also called gonococcus) is a Gram-negative, diplococci bacterium spread primarily through sexual contact. Gonorrhoeal infection occurs at human mucosal surfaces with the primary site of infection being the mucosal surfaces of the lower urogenital tract of both men and women. Other sites of gonococci infection include exposed mucosal surfaces of the rectum, pharynx and conjunctivae. Symptoms characteristic of lower-urogenital gonorrhoea infection are more readily identifiable in men compared to women. Problematically, in a large number of cases, gonorrhoeal infection is predominantly asymptomatic but remains transmissible, with women disproportionately affected compared to men (50% vs 10%) (Workowski, 2010). Asymptomatic carriage can have ramifications for long-term health outcomes and can facilitate the emergence of antibiotic resistant strains. Ascending gonococcal infection can occur in up to 45% of infected women and can result in pelvic inflammatory disease which may cause infertility and ectopic pregnancies (Edwards & Apicella, 2004). Gonococcal infections can have severe long-term implications to reproductive, maternal and new-born health including in women; chronic pelvic pain, spontaneous abortion and neonatal blindness, and in men; reduced fertility. It has also been reported that gonorrhoeal infection can increase susceptibility to and transmission of the human immunodeficiency virus (HIV) (Edwards & Apicella, 2004; Kirkcaldy, Weston, Segurado, & Hughes, 2019). Disseminated gonococcal infection although uncommon, has been reported to lead to septic arthritis, endocarditis and meningitis (Kirkcaldy et al., 2019; Masi & Eisenstein, 1981). Diagnosis of asymptomatic gonococcal infection is challenging and relies predominantly on regular screening efforts among populations who have access to such healthcare. This has far reaching implications on our understanding of the incidence, prevalence and transmission of gonorrhoea infection. Furthermore, clinical isolates are infrequently subjected to screening for resistance markers, further inhibiting our understanding of the evolution, spread and transmission of antimicrobial resistant gonorrhoea.

While infection of the nasopharynx has a significantly lower incidence and is asymptomatic, it is an important site of infection as such infections are considered to play a major role in development of antimicrobial resistance (AMR) in *N. gonorrhoeae* (Adamson & Klausner, 2021). *Neisseria* species are naturally transformable and readily obtain DNA from the environment which are integrated into the chromosome (Ochman, Lawrence, & Groisman, 2000). The nasopharynx is home to other commensal *Neisseria* species, which due to prior insufficient exposure to antibiotics, can harbour antibiotic resistance (AR) elements. The

natural transformability of *Neisseria* species creates an ideal environment for horizontal gene transfer to occur amongst *Neisseria* species within the nasopharynx (Ochman et al., 2000) which is postulated as the leading mechanism for AMR in *N. gonorrhoeae* (Adamson & Klausner, 2021; Dong & Klausner, 2019; Unemo & Shafer, 2014). Indeed, multiple ceftriaxone treatment failures all involved pharyngeal infections (Adamson & Klausner, 2021).

N. gonorrhoeae's ability to mitigate host immune defences is one of the key mechanisms that drive gonococcal pathogenicity. The natural transformability of *Neisseria* species contributes to antigenic variation, preventing recognition of the bacterium by the human immune system and allowing multiple gonorrhoea infections over one's lifetime (Criss, Kline, & Seifert, 2005; Stern, Brown, Nickel, & Meyer, 1986). A range of virulence factors have been implicated in the success of the bacterium to adapt to the host microenvironment. Many of these virulence factors including the pilus, opacity-associated (Opa) and outer membrane proteins undergo phase and/or antigenic variation which has thus far hampered gonococcal vaccine development and prevents adaptive immunity to the gonococcus (Edwards & Apicella, 2004).

1.3 Emergence of antimicrobial strains of *N. gonorrhoeae*

Of all STIs, *N. gonorrhoeae* has the highest levels of antimicrobial resistance. Due to the alarmingly rapid emergence of antimicrobial resistant strains of *N. gonorrhoeae* since treatment began in the 1930's, *N. gonorrhoeae* has earned "superbug" status by the World Health Organisation (WHO). Subsequently, *N. gonorrhoeae* has been classed as a "Priority 2" pathogen by the WHO and listed among the top five AMR threats by the Centers for Disease Control and Prevention (CDC). The CDC currently estimates that approximately half of all new gonococcal infections in the United States of America are drug resistant. Over decades of antibiotic use, *N. gonorrhoeae* has developed resistance to all available frontline antibiotics used for its treatment, including, in chronological order, sulfonamides, penicillin, tetracyclines, aminoglycosides, fluoroquinolones, macrolides (azithromycin) and cephalosporins (ceftriaxone and cefixime) (Figure 1.1). Resistance to these treatments occurs rapidly, with resistant strains replacing the sensitive bacterial populations within two decades (Costa-Lourenço, Santos, Moreira, Fracalanzza, & Bonelli, 2017).
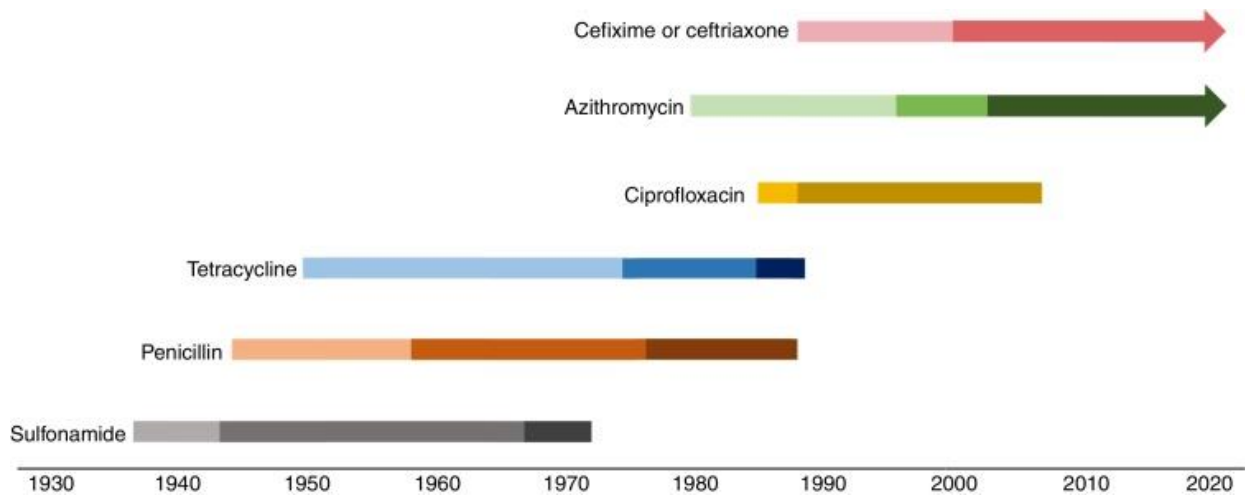
Figure 1.1 Timeline of acquired resistance to antibiotics used for N. gonorrhoeae treatment. Darkened bands show where resistance events occurred, and the end of the bands represents when antibiotics were no longer used for treatment. Figure taken from (Costa-Lourenço et al., 2017).

Following widespread failure of monotherapy treatments, the WHO recommends dual-therapy approaches with the oral antibiotic azithromycin and the injectable extended spectrum cephalosporin ceftriaxone as the current prescribed treatment suitable for most settings (WHO, 2016). However, even with dual therapy, the first instances of frontline AMR strains have emerged both globally and locally with completely resistant strains identified in England and Australia in 2018 giving rise to multi-drug resistant (MDR) strains. In response to MDR strains of *N. gonorrhoeae*, increased doses of antibiotics are prescribed where sensitivity is known with multiple molecular detection assays under development to assist successful treatment (Rochelle P. Walensky, 2021). While there has been no detection of MDR *N. gonorrhoeae* strains in New Zealand yet, strains with increasing resistance to antimicrobials have been detected across New Zealand (ESR, 2021). A recent study tested the resistance of 400 clinical *N. gonorrhoeae* isolates from New Zealand and reported seven 'pre-MDR' clinical isolates showing either decreased susceptibility to extended spectrum cephalosporins or spectinomycin (class I drugs) and resistance to two or more category II drugs. This study also reported 36 and 10 isolates which displayed decreased susceptibility to azithromycin and ceftriaxone respectively, while seven isolates displayed complete resistance to azithromycin (R. S. Lee et al., 2018). The presence of decreased susceptibility to current frontline antibiotics already detected in New Zealand among *N. gonorrhoeae* isolates, combined with increasing global incidence rates of MDR-gonorrhoea creates a real possibility of encountering such MDR-gonorrhoea in New Zealand in the future. Undisputedly, *N. gonorrhoeae* infection is a global crisis with the threat of antimicrobial resistant gonorrhoea providing urgency to develop novel therapeutic interventions to overcome AMR. Understanding key pathways involved in

4

infection, pathogenicity, survival and persistence of *N. gonorrhoeae,* as well as the regulatory mechanisms that govern these is crucial for novel therapeutic development.

1.4   Host immune response, L-cysteine and defence against oxidative stress

One of the main challenges for pathogenic bacteria upon colonisation is evasion of the host immune response and mitigation of the hostile host environment. Upon invasion of the female genital tract, *N. gonorrhoeae* encounters a hostile host environment colonised by other commensal bacteria that induce a host immune response which includes the release of various sources of oxidative stress. Alongside oxidative stress, this host response includes the release of antimicrobial peptides that interfere with the bacterial cell wall, proteins that sequester nutrients important to pathogens, fluctuating antibiotic levels, and initiation of the adaptive immune system directed at *N. gonorrhoeae* surface antigens (Hill, Masters, & Wachter, 2016).

N. gonorrhoeae displays an impressive ability to defend against oxidative stress employing many mechanisms to aid survival and resistance (Quillin & Seifert, 2018). Oxidative stress describes the situation where the production or presence of reactive oxygen species is imbalanced to their rate of detoxification. Oxidative stress can cause damage to DNA, proteins, cell membranes and often results in cell death highlighting the importance for effective oxidative stress defences upon colonisation. *N. gonorrhoeae* encounters reactive oxygen species such as the superoxide anion ($O_2^-\cdot$), hydrogen peroxide (H_2O_2) and the hydroxyl radical ($OH\cdot$) from the commensal *Lactobacillus spp*. populations of the female genital tract and through the innate immune response to infection. Commensal *Lactobacilli* generate reactive oxygen species through secreting hydrogen peroxide which inhibits gonococcal growth and by secretion of lactate which lowers the pH to promote aerobic respiration leading to endogenous production of reactive oxygen species during *N. gonorrhoeae's* metabolism (Seib et al., 2006). However, the main oxidative stress challenge for *N. gonorrhoeae* is from the host immune system. Detection of gonococcal infection recruits polymorphonuclear neutrophils to the infection site which mediate a rapid release of reactive oxygen species termed an 'oxidative burst' (El-Benna et al., 2016). Where neutrophils would ordinarily phagocytose pathogens and eliminate them by oxygen-dependent and oxygen-independent mechanisms, *N. gonorrhoeae* has displayed a remarkable ability to survive and replicate within polymorphonuclear neutrophils (Simons, Nauseef, & Apicella, 2005). In order to mitigate such oxidative stress, *N. gonorrhoeae* has evolved a novel combination of effectors and regulators to sense and respond to oxidative stress.

Cysteine is an important molecule in *N. gonorrhoeae* as it is a branch point for incorporation of sulfur into many compounds including iron-sulfur clusters, thiol-containing amino acids, biotin, co-enzyme A and is the rate-limiting amino acid in the synthesis of the reducing system glutathione (Kredich, 2008). Glutathione is a low-molecular weight tripeptide (L-γ-glutamyl-L-cysteinyl-glycine) and is considered the first line of defence against oxidative stress. Glutathione is a radical scavenger and restores oxidised macromolecules by hydrogen donation forming the oxidised form glutathione disulfide. The glutathione oxidoreductase maintains the reduced pool of glutathione and is essential, highly expressed (Remmele et al., 2014) and part of the OxyR regulon (Seib et al., 2007). *N. gonorrhoeae* strains lacking glutathione reductase and its regulator OxyR have been shown to be more susceptible to reactive oxygen species killing *in vitro* and demonstrate decreased survival within human cervical epithelial cells (Seib et al., 2007). Glutathione has been reported to be three-fold greater in concentration in *N. gonorrhoeae* compared to *Escherichia coli* (>15mM vs. 5mM (Seib et al., 2006)) which highlights an important role of glutathione as a powerful antioxidant system in *N. gonorrhoeae*. However, *N. gonorrhoeae* lacks any glutathione transporters to satisfy this increased internal redox system. As the incorporation of cysteine is the rate-limiting step in glutathione synthesis, this leaves glutathione synthesis contingent on cysteine made available within the cell. Additionally, any uptake by the cell of cysteine or its oxidised form cystine is likely insufficient due to the minimal exogenous sulfur pool in host serum (cysteine, methionine and cystine collectively are reportedly 136.2 µM (Lensmire & Hammer, 2019)). This leaves glutathione synthesis contingent on *de novo* cysteine synthesis via the cysteine biosynthetic pathways (Hicks & Mullholland, 2018).

1.5 Sulfate assimilation and cysteine biosynthesis in *Neisseria* species

The *de novo* synthesis of L-cysteine from inorganic sulfur is the main mechanism utilised by plant and bacterial species to acquire environmental inorganic sulfur and incorporate reduced sulfur into organic compounds (Kredich, 2008). Cysteine biosynthesis is highly conserved across bacteria, protozoa and higher plant species (Jeelani, Sato, Soga, & Nozaki, 2017; Mino & Ishikawa, 2003) but the biosynthetic pathway and all associated enzymes are not present in mammals as cysteine is synthesised from methionine. While there is 90% homology in the nucleotide sequence between the genomes of the pathogenic *Neisseria. meningitidis* and *N. gonorrhoeae*, there are significant differences in the cysteine biosynthetic pathway in *N. gonorrhoeae* (Hicks & Mullholland, 2018).

1.5.1 Active transport of sulfate

Cysteine biosynthesis begins with the import of sulfur sources. There are two main classes of sulfate transporter proteins, the major facilitator superfamily and the ATP-binding cassette (ABC) superfamily (Kertesz, 2001). These transporters use active transport to import inorganic sulfate into microbial cells. Genomic analysis showed that *N. gonorrhoeae* has just one of these sulfate uptake systems, that of the ABC transporter superfamily (Hicks & Mullholland, 2018). This transporter (Figure 1.2) is composed of a periplasmic sulfate binding protein (*sbp*), two permeases (*cysU* and *cysW*) and the ATP-binding subunit (*cysA*). Genomic analysis showed that the sulfate permease CysU is present in both *N. meningitidis* and *N. gonorrhoeae* but was absent in the *N. gonorrhoeae* strain FA1090 (Hicks & Mullholland, 2018). While the sbp transporter is able to transport sulfate (Kredich, 2008), given that *N. gonorrhoeae* can grow on thiosulfate as a sole source of sulfur and this sbp transporter is the only sulfate transporter present, it is also postulated that this transporter may also transport thiosulfate (Hicks & Mullholland, 2018).
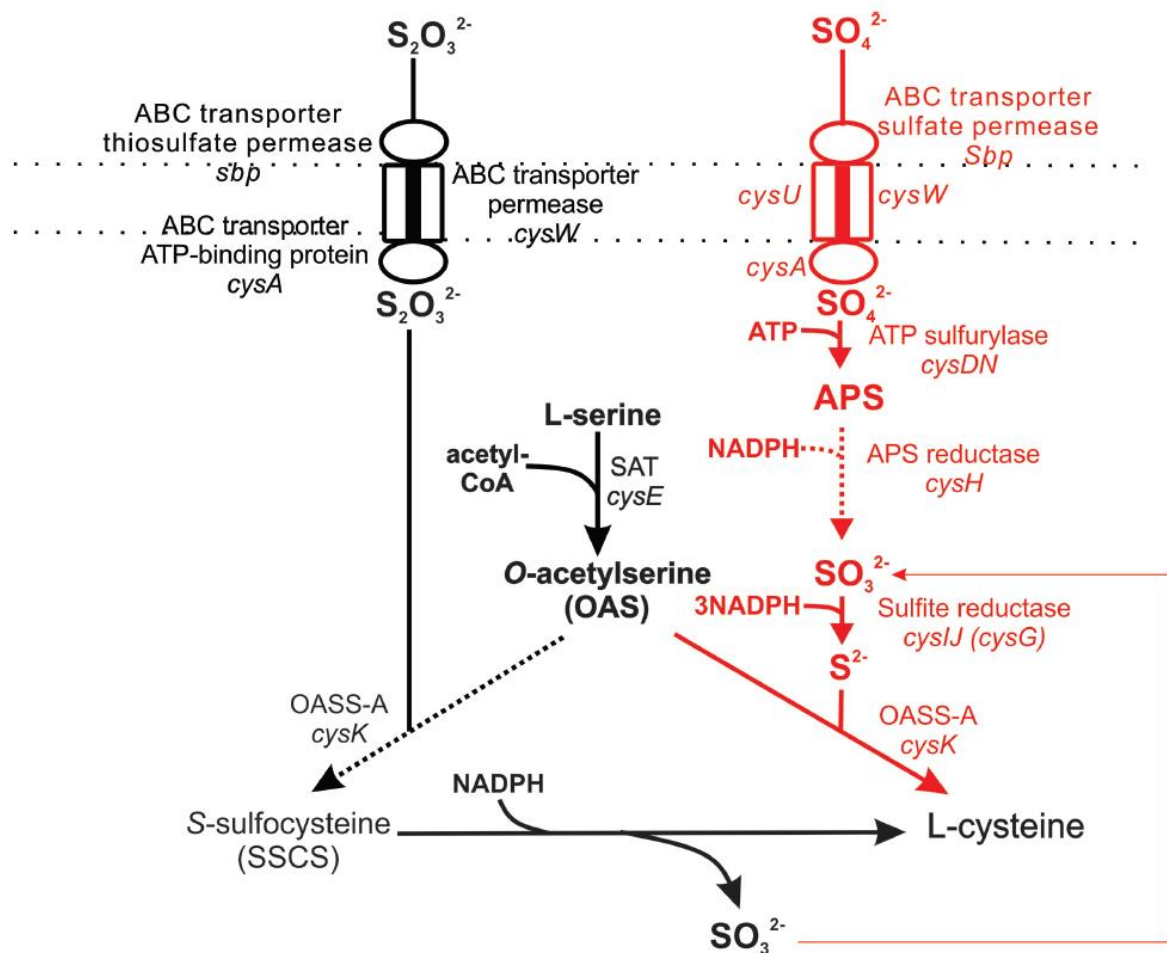
Figure 1.2 Sulfate transporter and the sulfate reduction pathway in Neisseria species. The sulfate reduction pathway shown in red is absent in N. gonorrhoeae. Differences between Neisseria species and other bacteria are represented by dotted lines – APS reductase directly reduces APS to sulfite, and OASS-A (CysK) is the only OASS isoform present in Neisseria. Figure taken from (Hicks & Mullholland, 2018).

1.5.2 The sulfate reduction pathway

Following import of sulfate into the cell, sulfate must be reduced to bisulfide for incorporation into L-cysteine. In *E. coli* reduction of sulfate to bisulfide occurs in successive steps via the sulfate reduction system. This first involves the activation of sulfate in an ATP-dependent reaction catalysed by ATP sulfhydrylase (CysDN) to form adenosine 5'-phosphosulfate (APS). Second, APS is phosphorylated by APS kinase (CysC) to form phosposadenosylphosphosulfate (PAPS) in an ATP-dependent reaction and subsequently reduced to sulfite and phosposadenosylphosphate (PAP) by PAPS reductase (CysH). In *Neisseria* species a single enzyme APS reductase (also termed CysH) reduced APS to sulfite directly. The final step is the reduction of sulfite to sulfide by the multi-subunit enzyme sulfite reductase (CysJI) in an NADPH-dependent reaction (Kredich, 2008).

The sulfate reduction pathway is arranged differently across different bacterial species. *E. coli* contains three transcriptional units; the operons *cysDNC* and *cysJIH* and the single gene *cysG*. *Neisseria* species have all genes required for sulfate reduction under one operon with the *N. meningitidis* containing the *cysGHDNJI* operon, while *N. gonorrhoeae* has an unusual *cysGNJI* operon (Figure 1.3). Analysis of this operon in *N. gonorrhoeae* revealed a 3500 bp deletion from the *N. gonorrhoeae* genome (~2,219 kb genome) which removed the *cysH* and *cysD* genes entirely and results in truncated *cysG* and *cysN* genes (Hicks & Mullholland, 2018). Additionally, the *cysJ* and *cysI* genes have premature stop codons denoting them as pseudogenes. Deletion of *cysD* would remove the ability of *N. gonorrhoeae* to reduce sulfate to APS, with *cysH* deletion preventing subsequent reduction of APS to sulfide. The in frame stop codons of *cysJ* and *cysI* render *cysJ* and *cysI* as pseudogenes which would prevent reduction of sulfite to sulfide. The combination of these genetic deletions, truncations and pseudogenes renders the sulfate reduction pathway non-functional in *N. gonorrhoeae*. This supports *N. gonorrhoeae's* inability to grow on sulfate as a sole source of sulfur (Le Faou, 1984). A functional sulfate reduction operon or individual genes are not located elsewhere in the *N. gonorrhoeae* genome (Hicks & Mullholland, 2018) indicating *N. gonorrhoeae* has a markedly different cysteine biosynthesis pathway to other pathogenic bacteria including other *Neisseria* species. As the sulfate reduction pathway is non-functional in *N. gonorrhoeae*, yet *N. gonorrhoeae* contains just one sulfate transporter system but is unable to grow on sulfate as a sole source of sulfur, other, as yet, unknown mechanisms must be at play to ensure reduction of sulfate for synthesis of cysteine.
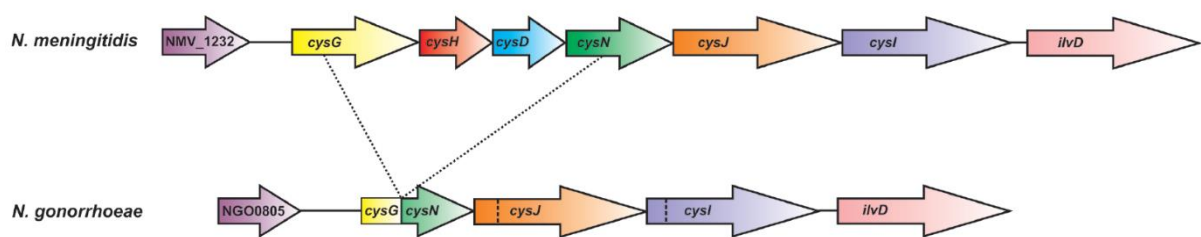
Figure 1.3 The sulfate reduction operon of N. meningitidis and N. gonorrhoeae. Dotted lines represent the 3500b bp genomic deletion in N. gonorrhoeae truncating cysG and cysN genes and deleting cysH and cysD. Dashed lines in N. gonorrhoeae cysJ and cysI genes represent stop codons generating pseudogenes. Figure taken from (Hicks & Mullholland, 2018).

1.5.3 Cysteine biosynthesis

The cysteine biosynthesis pathway is a two-step reaction catalysed by two enzymes, serine acetyltransferase (SAT, *cysE*) and *O*-acetylserine-sulfhydrylase (OASS, *cysK/cysM*). First the side-chain carboxyl of L-serine is acetylated by SAT in a CoA-dependent reaction to form *O*-

acetyl-L-serine. Second, *O*-acetylserine is replaced with sulfide by OASS in a condensation reaction to form L-cysteine.

cysE has been denoted as an essential gene in *N. gonorrhoeae* (Remmele et al., 2014) and while it is not essential in *N. meningitidis*, a *cysE* deletion strain shows impaired growth in media and increased sensitivity to antibiotics (Capel et al., 2016). Serine acetyltransferase has been explored as a target for inhibition (Agarwal, Jain, Bhattacharya, & Azam, 2008; C. Chen et al., 2019; Magalhães et al., 2020) and structural characterisation of SAT from *N. gonorrhoeae* (Oldham et al., 2022) allows the investigation of SAT as a target for therapeutic development in *N. gonorrhoeae*. *O*-acetylserine-sulfhydrylase has two isoforms, OASS-A (*cysK*) which utilises sulfide and OASS-B (*cysM*) which utilises thiosulfate (Kredich, 2008). While most bacteria have both isoforms to allow cysteine biosynthesis using sulfide and thiosulfate, *N. gonorrhoeae* has just OASS-A (*cysK*). This also presents an interesting opportunity for inhibition of OASS-A in *N. gonorrhoeae*.

1.6 Regulation of cysteine biosynthesis

Regulation of the cysteine biosynthetic pathway is two-fold. Regulation occurs at both the protein and DNA level. Feedback inhibition and a bienzymatic complex termed the cysteine synthase complex regulate the pathway at the protein level. DNA-level regulation occurs via transcriptional regulation of the many genes involved in the pathway by the transcriptional regulator CysB.

1.6.1 The cysteine synthase complex and feedback inhibition

The cysteine synthase complex (CSC) is a bienzyme complex formed by interaction of the two enzymes involved in cysteine synthesis – SAT and OASS. The complex is comprised of one SAT hexamer and two OASS dimers (Benoni et al., 2017; Kredich, Becker, & Tomkins, 1969). While most enzyme complexes form to channel substrates between them, this complex is unusual in that the C-terminal tail of SAT inserts into the active site of OASS competing with the OASS substrate *O*-acetylserine. The formation of this complex modulates cysteine production by inhibiting the activity of OASS and increasing the activity of SAT. Where SAT is regulated by feedback inhibition of L-cysteine, sensitivity of SAT to such inhibition is reduced when SAT forms the CSC. Formation of the complex is mediated by sulfur availability. With high sulfur availability the complex is stabilised by bisulfide which allows increased SAT activity due to reduced sensitivity of substrate inhibition by L-serine (Benoni et al., 2017). When sulfur availability is low, accumulation of millimolar amounts of *O*-

acetylserine signals sulfur starvation and dissociation of the complex which allows activity of OASS to resume (Kredich et al., 1969; Wang & Leyh, 2012).

1.6.2 The cysteine regulon

The cysteine regulon is composed of genes involved in the sulfate/thiosulfate transporter, the sulfate reduction pathway and the cysteine synthase complex. Most genes of the cysteine regulon are regulated by the transcription factor CysB, except *cysE* which is constitutively expressed (Jones-Mortimer, Wheldrake, & Pasternak, 1968; Kredich, 1971). The cysteine regulon has been shown to be upregulated among other pathogenic bacteria *in vitro* under oxidative stress (Turnbull & Surette, 2010) in the presence of nitric oxide (Santi-Rocca et al., 2012) and *in vivo* during infection and long-term survival (Fontán, Aris, Ghanny, Soteropoulos, & Smith, 2008; Schnappinger et al., 2003). The sulfate reduction operon is essential for *N. meningitidis* growth and upon host cell colonisation genes of the cysteine regulon have been shown to be upregulated, yet no changes to the expression of the transcriptional regulator *cysB* were observed in *N. meningitidis* (Capel et al., 2016; Mendum, Newcombe, Mannan, Kierzek, & McFadden, 2011). CysB activates transcription of the cysteine regulon genes by binding to promoter regions to promote transcription. However, CysB exhibits negative autoregulation of its own promoter, only allowing transcription of *cysB* when CysB is no longer bound to the *cysB* promoter region. Transcriptome analysis of three clinical isolates of *N. gonorrhoeae* during natural infection compared to growth in chemically defined media, showed downregulation of *cysB* in two of the three subjects (McClure et al., 2015). Current transcriptome data has not investigated the relationship between expression of the cysteine regulon and *cysB,* nor investigated this relationship during stages of infection.

1.7 Transcriptional regulation in *N. gonorrhoeae*

Neisseria gonorrhoeae's success as an obligate human pathogen can be attributed to three key factors; 1) its large asymptomatic carrier population, 2) its ability to quickly acquire antibiotic resistance and 3) its ability to evade the host immune system. Gene expression and regulation are key to the success of these attributes in gonorrhoeal infection.

One mechanism for gene regulation is through DNA-binding transcriptional regulators. *N. gonorrhoeae* has just ~34 putative transcriptional regulators within its genome based on homology searches. This is far fewer than other organisms such as *E. coli* that have greater than 200 transcriptional regulators (Schielke, Frosch, & Kurzai, 2010) and points to the importance of these transcriptional regulators for *N. gonorrhoeae* infection, survival and

pathogenesis. Of the ~34 transcriptional regulators in *N. gonorrhoeae*, 14 have been partially or fully characterised and have various gene targets involved in nutrient uptake/availability (Jackson et al., 2010; Overton et al., 2006; Sunkavalli, McClure, & Genco, 2022), stress responses (Kidd, Potter, Apicella, Jennings, & McEwan, 2005; Kim & Holmes, 2012; Mongkolsuk & Helmann, 2002; Schook, Stohl, Criss, & Seifert, 2011; Teramoto, Inui, & Yukawa, 2013; Wu et al., 2006; Zaim & Kierzek, 2003), efflux pumps (Ayala & Shafer, 2019; Folster, Dhulipala, Nicholas, & Shafer, 2007; Hagman & Shafer, 1995; E. H. Lee & Shafer, 1999), surface antigen variation and *pilE* regulation (De Reuse & Taha, 1997; Holley, Ayala, & Shafer, 2020; Matthias & Rest, 2014; Sainsbury et al., 2009). Regulators of particular interest for therapeutic development which have been well-characterised are MtrR and its regulator MpeR which regulate the Mtr efflux pump, and OxyR which regulates response to oxidative stress.

Regulatory mechanisms which have been described that underpin *N. gonorrhoeae's* ability to sense and respond to a hostile environment include the superoxide-dependent OxyR, Mn-dependent PerR, Zn-dependent thiol-based NmLR, iron-dependent Fur and oxygen-dependent FNR. These mechanisms serve to detoxify the host environment under different conditions where transcriptome analysis during infection reports the upregulation of OxyR and Fur responses (McClure et al., 2015). The roles of regulators OxyR, PerR and Fur alongside MtrR and MpeR are shown in Figure 1.4 to highlight the many ways in which *N. gonorrhoeae* responds to various stress stimuli upon host infection. This shows the importance of gene regulation for *N. gonorrhoeae* infection and survival in the female genital tract.
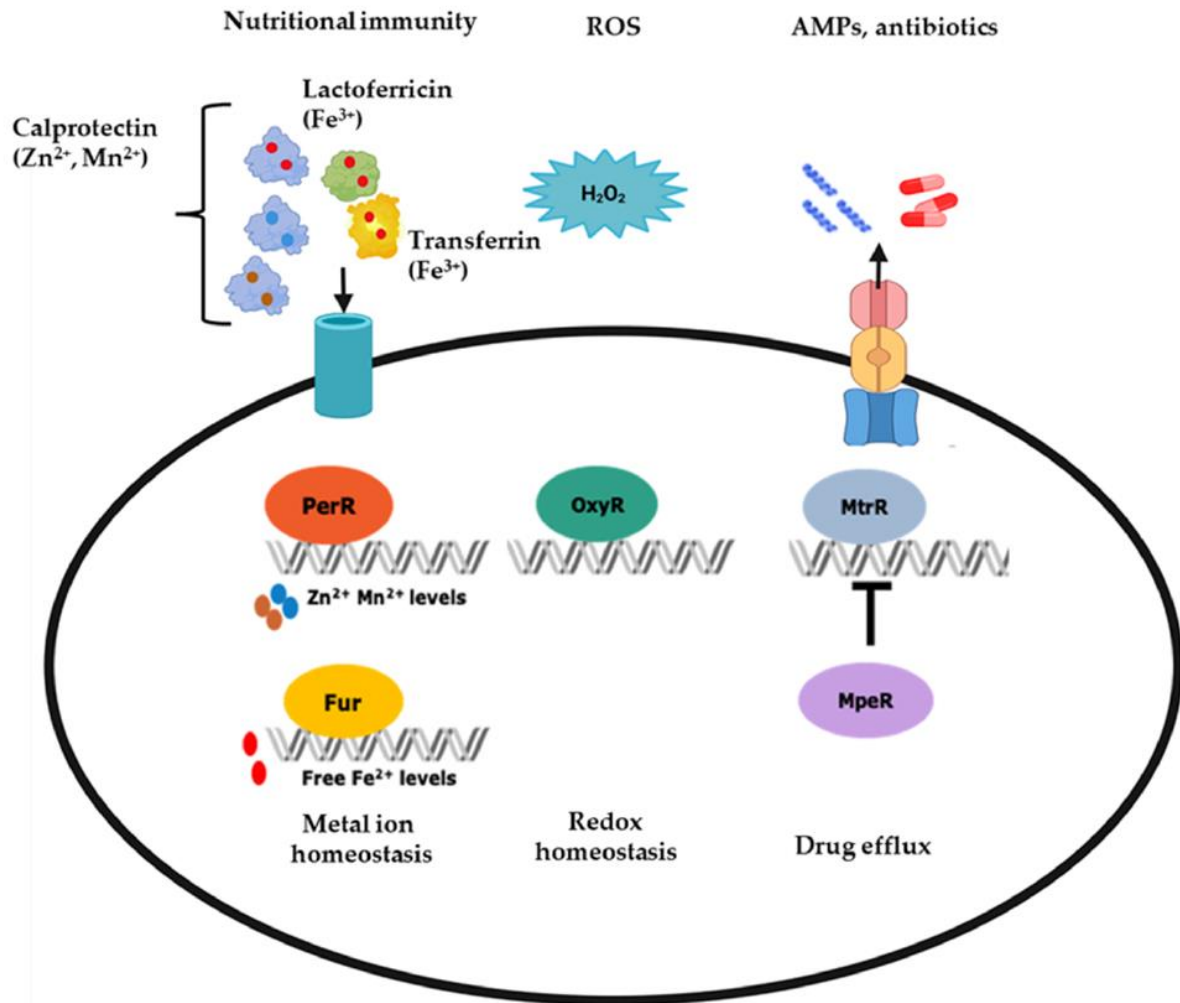
Figure 1.4 Transcriptional regulators that respond to stress stimuli in N. gonorrhoeae. The roles of well-characterised transcriptional regulators PerR, Fur, OxyR, MtrR and MpeR in regulating pathways involved in response to stress stimuli during N. gonorrhoeae infection. PerR and Fur are involved in nutrient availability through nutritional immunity and metal ion homeostasis with mechanisms such as the uptake of metal chelating proteins including lactoferrin for use of sequestered metal ions. OxyR mediates peroxide stress maintaining redox homeostasis. MtrR and its regulator MpeR regulate the expression of the Mtr efflux pump to remove antibiotics and hydrophobic agents from the cell. Figure taken from (Sunkavalli et al., 2022).

Multiple *N. gonorrhoeae* transcriptomic studies have been undertaken across a variety of conditions including natural infection and *in vitro* studies. These studies have generated useful data which allows identification of important or essential genes for target by novel therapeutics and can facilitate the identification of novel gene targets of transcriptional regulators (Sunkavalli et al., 2022). Transcriptomic studies have investigated expression levels of genes and their essentiality (Remmele et al., 2014) as well as differential expression of genes from clinical isolates during natural infection (McClure et al., 2015). Perhaps unsurprisingly, of the four clinical isolates in this study, genes relating to the pilin, transcription and stress response showed the highest ratio of enrichment during natural infection.

From these studies we have identified CysB as a transcriptional regulator that is highly expressed and essential during *in vitro* growth, yet downregulated during natural infection. CysB is one of the few transcriptional regulators in *N. gonorrhoeae* and is postulated to regulate *N. gonorrhoeae's* altered cysteine regulon. This essentiality, differential expression during infection, the limited number of *N. gonorrhoeae* transcriptional regulators as well as CysB's role in regulating an altered cysteine regulon warrant investigation of this transcriptional regulator and the role it may play in *N. gonorrhoeae* infection, survival and/or pathogenesis.

1.8 CysB

CysB as a transcriptional regulator is a member of the LysR-type transcriptional regulator (LTTR) family. The LTTR family was first described by Henikoff (Henikoff, Haughn, Calvo, & Wallace, 1988) and, as LysR was the most well-characterised transcriptional regulator of the family at the time, became the family namesake. Since their discovery, LTTRs have been reported ubiquitously amongst bacteria with functional orthologues reported in archaea and eukaryotic organisms (Pérez-Rueda & Collado-Vides, 2001; Stec et al., 2006; Sun & Klein, 2004). Currently the LTTR family is the largest known family of prokaryotic DNA-binding proteins with members identified based on their amino acid sequence (Schell, 1993). LTTRs act as global transcriptional regulators and may act as either activators or repressors of either a single gene or operonic genes. LTTRs are found across a wide range of bacteria and regulate a similarly diverse range of functions ranging from virulence and motility, to oxidative stress responses and metabolism. Despite this vast diversity, LTTRs do share some common characteristics; LTTRs are of similar size (~330 amino acids) functioning as homodimers or homotetramers, all LTTRs contain a structurally well-conserved helix-turn-helix (HTH) DNA binding motif at the N-terminus while the C-terminus is arranged to contain an inducer-binding cavity where co-inducers play an important role in LTTR activation (Maddocks & Oyston, 2008).

CysB has been partially characterised in other organisms but to date has not been characterised in *N. gonorrhoeae*. Functional studies report CysB's binding activity to cysteine regulon promoters and inducer activity in *Salmonella typhimurium* and *E. coli*. Structural reports describe truncated forms of CysB monomers and dimers from other organisms where no full-length CysB structures have yet been reported, nor any truncated structures from *N. gonorrhoeae*.

1.8.1 Biochemical characterisation of CysB

Expression of the cysteine regulon, and hence biosynthesis of cysteine, is dependent on the transcriptional regulator CysB. Gel filtration studies report CysB is a homotetramer with identical $M_r = $ ~36,000 subunits (Jagura-Burdzy & Hulanicka, 1981; Miller & Kredich, 1987). As a member of the LTTR family, CysB has dual regulatory function which is dependent on the binding of co-factors termed inducers and anti-inducers. Upon inducer binding, CysB activates transcription of the cysteine regulon while exerting negative autoregulation of its own *cysB* promoter. *N*-acetylserine, and to a lesser extent *O*-acetylserine, act as inducers while sulfate and thiosulfate act as anti-inducers (Figure 1.5). These co-factors bind to the inducer-binding cavity or secondary site to modify the interactions with DNA. While no CysB structure has yet been solved with thiosulfate to confirm its site of interaction, the similar structures of sulfate and thiosulfate suggest that thiosulfate likely binds in the inducer-binding cavity where sulfate binds. *O*-acetylserine is the product of the constitutively expressed serine acetyltransferase (*cysE*) and spontaneously isomerises to form *N*-acetylserine which accumulates in the cell.
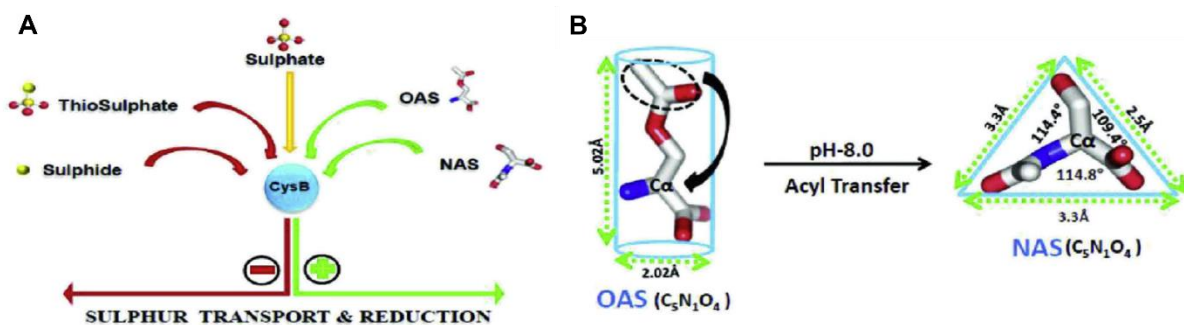
Figure 1.5 Inducers and anti-inducers of CysB regulating the cysteine regulon genes. ***A*** *The anti-inducers thiosulfate and sulfate alongside the inducer O*-acetylserine *and N*-acetylserine*.* ***B*** *Stereochemical features of inducers O*-acetylserine *and N*-acetylserine*. The arrow indicates the spontaneous isomerisation of O*-acetylserine *to N*-acetylserine *by acyl transfer to form a triangular N*-acetylserine *from a cylindrical O*-acetylserine*. Figure taken from (Mittal, Singh, & Kumaran, 2017).*

CysB binding activity to promoters of the cysteine regulon has been well characterised in *S. typhimurium* and *E. coli*. These studies show that a single CysB tetramer binds to the promoters of *cysB, cysJ* (of the *cysJIH* operon), *cysK* and *cysP* (of the *cysPUWAM* operon) in the absence of inducer. However, activation of transcription of the *cysJ*, *cysK* and *cysP* genes by CysB is dependent on inducer binding (Ostrowski & Kredich, 1989, 1990). Hydroxyl-radical foot printing studies and nucleotide sequence alignments indicate that within the promoter regions of *cysJ*, *cysK* and *cysP* genes, CysB binds to 19 bp half sites convergently

orientated and separated by 1-3 bp termed a CysB-binding site (CBS) which encompass regions of ~40 bp (Hryniewicz & Kredich, 1995). The spacing between the binding half sites is important for activation (Colyer & Kredich, 1996). Within each of the *cysJ, cysK* and *cysP* promoters there are multiple binding sites including an activation site located just upstream of the -35 region. These activation sites are termed CBS-J1, CBS-K1 and CBS-P1 within the *cysJ, cysK* and *cysP* promoters respectively and are displayed for clarification in Figure 1.6. Accessory sights are found upstream and downstream of these activation sites termed CBS-J2 and CBS-J3 in the *cysJ* promoter, CBS-K2 in the *cysK* promoter and CBS-P2 and CBS-P3 in the *cysP* promoter (Figure 1.6). *N*-acetylserine stimulates binding to CBS-J2 and CBS-P2 but inhibits binding to CBS-K2 and CBS-P3, yet these accessory sites have unknown function (Tyrrell et al., 1997).

Within its own promoter, CysB binds a single repressor site designated CBS-B1 (Figure 1.6) which lies within the RNA polymerase-binding region of the promoter (Ostrowski & Kredich, 1991). It is suggested that when CysB is bound to its own promoter it occludes the RNA-polymerase binding region thereby preventing transcription (Ostrowski & Kredich, 1991). *N*-acetylserine lowers the affinity of CysB for CBS-B1 and *in vitro* studies show CysB completely releases its promoter at high concentrations of *N*-acetylserine ergo exerting negative autoregulation of its promoter (Hryniewicz & Kredich, 1995; Lynch, Tyrrell, Smerdon, Briggs, & Wilkinson, 1994; Mittal et al., 2017; Ostrowski & Kredich, 1991). Comparison of the *cysB, cysJ, cysK* and *cysP* promoter sequences from *S. typhimurium* and *E. coli* reveals elements of identity but no strong consensus sequence to define what constitutes a CysB-binding site (Kredich, 1992).
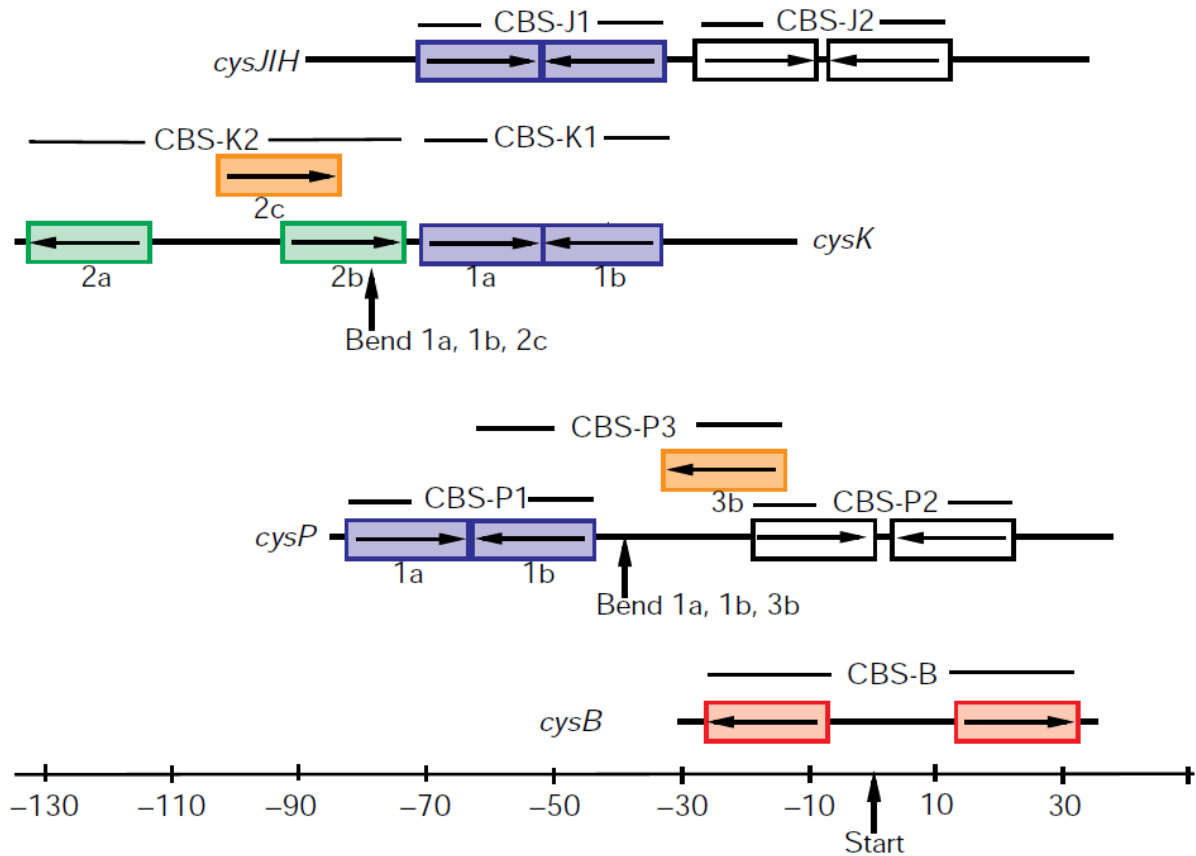
Figure 1.6 CysB-binding sites within the cysJ, cysK, cysP and cysB promoters of S. typhimurium. The boxes represent 19 bp half sites where each box is bound by a single CysB N-terminal domain. Activation sites (blue) are shown just upstream of the -35 region with accessory sites (white) downstream. The repressor sites (red) within the cysB promoter overlap the RNA polymerase site near the transcriptional start site. N-acetylserine stimulates binding to the activation sites and accessory sites CBS-J2 and CBS-P2, while inhibiting binding to the cysB repressor sites and accessory site CBS-K2. Additional half sites (orange) are thought to be involved in DNA bending which occurs at points indicated by arrows within the cysK and cysP promoters. Figure taken from (Tyrrell et al., 1997).

While CysB binds to the *cysK* and *cysP* promoters without inducer, the absence of inducer is reported to cause a 100° bend in the DNA just upstream of the activation sites (Hryniewicz & Kredich, 1991, 1994). This is caused through CysB binding to the activation site CBS-K1 or CBS-P1 and the additional overlapping half site CBS-K2c or CBS-P3b (Figure 1.6). Upon inducer binding, the DNA bend is reduced to 50° which allows CysB to bind the activation site exclusively. For both promoters, the additional half site is located one helical turn away from the activation site. The mechanism by which this occurs is not yet understood.

The lack of a CysB binding site consensus sequence, the existence of multiple overlapping binding sites and sites of unknown function, alongside the involvement of DNA bending, result in unusually complex interactions between CysB and cysteine regulon promoters. This makes identification of novel CysB binding sites within other genes difficult.

1.8.2 Structural characterisation of CysB

CysB structures solved by X-ray crystallography have been reported from just three organisms; *Klebsiella aerogenes*, *Salmonella typhimurium* and *Pseudomonas aeruginosa* plus an LTTR involved in sulfur starvation (Cbl) from *E. coli*. Due to difficultly generating crystals for the full-length CysB protein, these structures report truncated forms of CysB. The CysB monomer can be divided into two domains – an N-terminal DNA-binding domain and a C-terminal regulatory domain arranged to contain an inducer-binding cavity. The structure of the regulatory domain of CysB has been reported in each of these organisms with the first N-terminal DNA-binding domain reported in 2019 (Song et al., 2019). The first reported CysB regulatory domain structure was reported in *K. aerogenes* (PDB code 1AL3 (Tyrrell et al., 1997) which modelled the regulatory domain spanning residues 88-324.

The regulatory domain of CysB has an ellipsoid shape with an axial ratio of 2:1 consisting of two similar α/β domains, domain I and II, connected by two cross-over regions (Figure 1.7). Domain I (residues 88-162 and 270-292) has five β sheets (β_A-β_D and β_J) arranged to form a β sheet topology $\beta_B\beta_A\beta_C\beta_J\beta_D$ where β_J is antiparallel to the others. This forms a 'core' sheet within the domain. Four α helices (α_I-α_{III} and α_{VII}) also make up domain I with three of the helices inserted in loop regions between the core β sheets and the fourth (α_{VII}) is after β_J. Directly after β_J the C-terminal arm (residues 293-318) is formed by helices α_{VIII} and α_{IX} and sheet β_K which loops away from the core sheets and terminates at the interface between domain I and II. Cross over regions occur between β_D and β_E as well as β_I and β_J. Domain II consists of a single sequence of polypeptides (residues 166-265) with a similar arrangement to domain I consisting of five β sheets and three α helices. Five β sheets form a core with the strand order $\beta_I\beta_E\beta_H\beta_F\beta_G$ where β_E is antiparallel, which is again interspersed with three helices (α_{IV}-α_{VI}). Each of these domains have 'Rossmann fold' like topology, and leaving aside the C-terminal arm, are related by pseudosymmetry. Between these two domains is a cavity approximately 10 Å deep by 6 Å in diameter lined with polar residues from the loop regions and cross-over polypeptide segments. This cavity forms the inducer-binding cavity which in this structure was occupied by a sulfate anion and surrounding water molecules. The inducer-binding cavity is narrowed at one end by the C-terminal end of helix α_I and the authors postulated that this may sterically restrict access to the binding cleft.
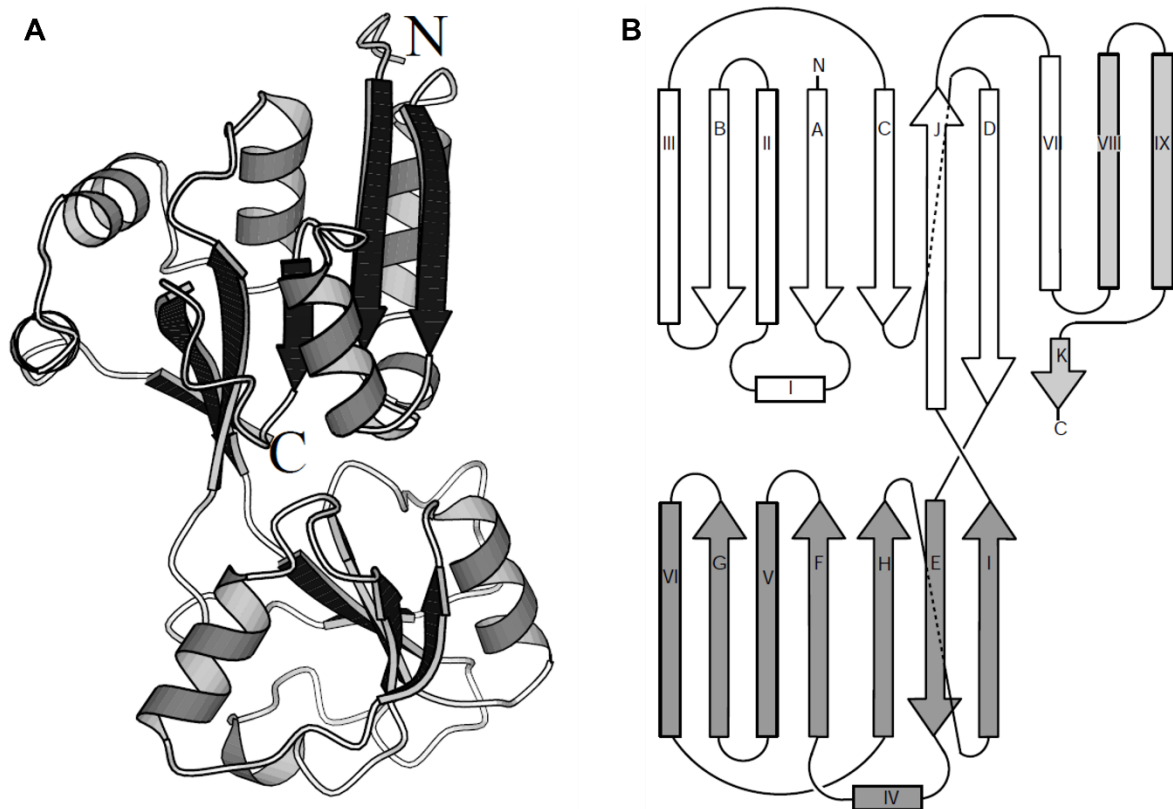
*Figure 1.7 Regulatory domain (residues 88-324) of CysB from K. aerogenes (PDB code 1AL3). **A** Stereoview of the regulatory domain of CysB highlighting secondary structure elements. α helices are shown as ribbons and β strands as arrows indicating their direction. The N and C termini are labelled. The inducer-binding cavity is located in the cleft between the two domains roughly where the C terminus label is. **B** Topological arrangement of secondary structures within the CysB RD. α helices are represented as rectangles and β strand as arrows. The two domains (I above, II below) and the C-terminal are shaded differently. Figures taken from (Tyrrell et al., 1997)*

Structural data for the CysB regulatory domain demonstrate that the CysB monomer is related to a neighbouring CysB molecule by a twofold axis of symmetry. The interface between these two monomers buries a surface of ~2,400 Å^2 on each molecule which supports dimer formation of the CysB regulatory domain. This dimeric regulatory domain crystal supported gel filtration studies of the purified regulatory domain where a ~50,000 Da product was formed (Tyrrell et al., 1997). This is in contrast to the full-length CysB which forms a tetramer. Dimer formation further restricts access to the inducer-binding cavity where the openings to the two cavities form a channel.

Other subsequently reported regulatory domain (RD) structures from *E. coli* (*Ec*Cbl-RD, PDB code 2FYI; (Stec et al., 2006)), *S. typhimurium* (*St*CysB-RD, PDB codes 4GXA, 4GWO, 4M4G, 4LQ5, 4LQ2; (Mittal et al., 2017)) and *P. aeruginosa* (*Pa*CysB-RD, PDB codes 5Z50; (Song et al., 2019)) show high structural and amino acid sequence similarity to that of the

regulatory domain of CysB from *K. aerogenes* (*Ka*CysB-RD). These structures differ in which ligands are bound in the inducer-binding cavity. The *Ka*CysB-RD (1AL3) and *St*CysB-RD (4GXA) structures report a single sulfate anion in the inducer-binding cavity, while the *Pa*CysB-RD (5Z50) structure reports one sulfate anion in the inducer-binding cavity with several sulfate anions surrounding the cavity. The structures of *Ec*Cbl-RD (2FYI) and *St*CysB-RD (4GWO) report *apo* forms with no density in the inducer-binding cavity. One *St*CysB-RD structure reports *N*-acetylserine (4M4G) in the inducer-binding cavity while two other structures of *St*CysB-RD report *O*-acetylserine adjacent to the inducer-binding cavity (4LQ2 and 4LQ5). *N*-acetylserine, and to a lesser extent *O*-acetylserine, are inducers of CysB activity while sulfate is an anti-inducer. Each of these ligands remodel the inducer-binding cavity when bound and small differences can be seen in the surrounding secondary structures (Figure 1.8).
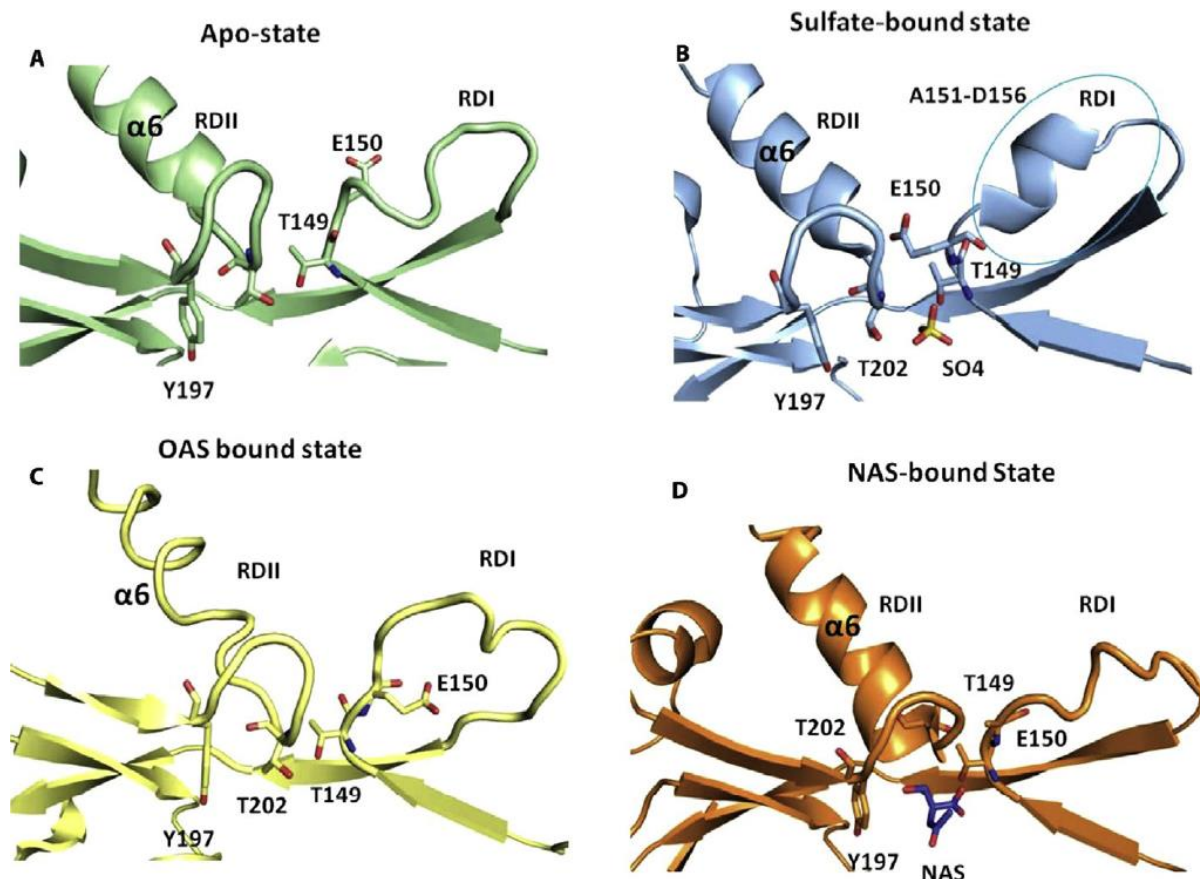
*Figure 1.8 Inducer-binding cavity of St-CysB-RD in different states with differing ligands. **A** The apo state with no ligand bound. **B** The anti-inducer bound state with sulfate bound. **C** Remodelling of the inducer-binding cavity when inducer O-acetylserine is bound to a secondary binding site (secondary site not shown here). **D** The inducer-binding cavity in an inducer-bound state with inducer N-acetylserine bound. Labelled residues undergo ligand specific conformational changes which results in the gain or loss of secondary structures, including α_6 and a loop region in domain I. Figure taken from (Mittal et al., 2017).*

The *St*CysB-RD structures reported by Mittal are the first to describe the inducer-bound regulatory domain of CysB (Mittal et al., 2017). The *N*-acetylserine-bound structure shows *N*-acetylserine binding into the previously described inducer-binding cavity while *O*-acetylserine-bound structures display *O*-acetylserine binding at a second site (Figure 1.8). This second site, hereby termed site-2, is located at the C-terminal arm on the exterior of the protein. Interestingly, *O*-acetylserine only occupied site-2 and not the inducer-binding cavity in the two crystals generated, even when crystals were soaked at higher concentrations of *O*-acetylserine indicating a more complex *St*CysB activation mechanism. Comparison of the inducer-binding cavity between *apo* and *O*-acetylserine-induced states displayed remodelling of the inducer-binding cavity when *O*-acetylserine is bound to site-2 shown in Figure 1.8. This suggests that the inducer-binding cavity and site-2 are allosterically coupled, whereby binding to site-2

remodels the inducer-binding cavity. Further investigation of the two binding sites revealed that the inducer-binding cavity excludes binding of *O*-acetylserine while allowing binding of *N*-acetylserine and the sulfate anion, while site-2 preferentially recognises *O*-acetylserine (Mittal et al., 2017). *N*-acetylserine and sulfate are the primary inducer and anti-inducer respectively, while *O*-acetylserine appears to have a more secondary inducer role. It is proposed that this selective binding of *O*-acetylserine and sulfate has physiological consequences. The inducer-binding cavity may be used to sense sulfate availability where decreased sulfate levels indicate increasing levels of *O*-acetylserine, while site-2 may sense the accumulation of *O*-acetylserine before generation of significant *N*-acetylserine and may remodel the inducer-binding cavity in preparation for *N*-acetylserine binding. This presents a dual sensor ability which would allow CysB to monitor the first and last step of cysteine biosynthesis.

The first structure of a CysB N-terminal domain (NTD) was reported from *P. aeruginosa* (*Pa*CysB-NTD, PDB codes 5Z4Y and 5Z4Z; (Song et al., 2019)) which modelled residues 1-90. These structures consist of four α-helices (α_{1-4}) and two β-sheets (β_{1-2}) which are arranged into a winged-helix-turn-helix (wHTH) motif where the fourth helix, termed the linker helix, relates the N-terminal domain to the regulatory domain (Figure 1.9). The wHTH is comprised of helices α_2 and α_3 with the two anti-parallel β sheets forming a 'wing'. Comparison with the full-length structure of the LTTR, BenM, from *Acinetobacter baylyi* in complex with DNA showed that helix α_3 was inserted into the major DNA groove with the wing positioned to bind into the minor groove (Song et al., 2019). This supports the expected mechanism for wHTH-DNA interactions (Pabo & Sauer, 1984). Of the two reported *pa*CysB-NTD structures, one shows an ordered wing motif while the other structure shows a disordered motif. This indicates flexibility within the wing that may provide certain conformational changes required for targeted DNA binding. Within this winged HTH motif are two positively charged lysine residues (Lys35 and Lys54) which are important for DNA binding activity. The N-terminal domain in both structures formed a dimer where the linker helices dimerised antiparallel to one another to position the N-terminal domain facing toward one another to form a DNA-binding domain (DBD) Figure 1.9.
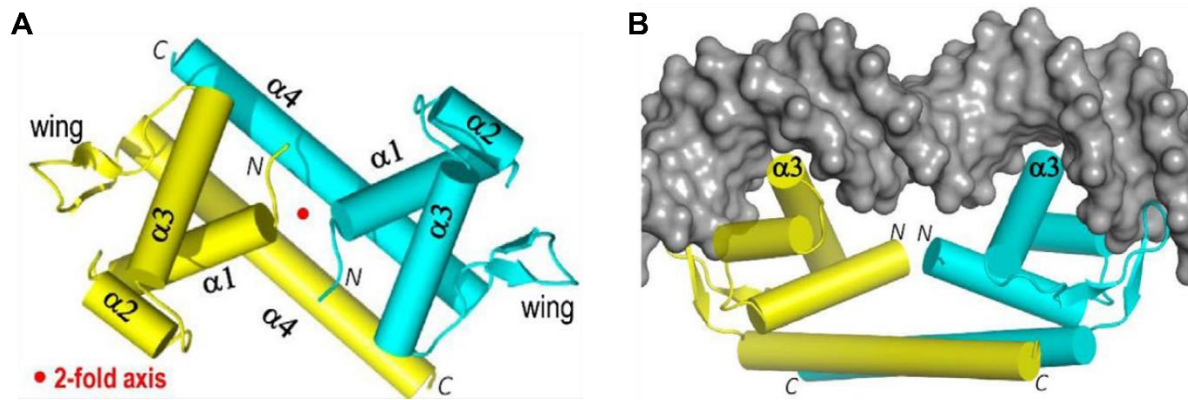
*Figure 1.9 Dimer of N-terminal domains (residues 1-90) of CysB from P. aeruginosa (PDB code 5Z4Z, 5Z4Y). **A** Cartoon representation of the dimerised N-terminal domain with each monomer coloured blue or yellow. α helices are labelled and the wing shows two small antiparallel β sheets. **B** Surface model of DNA in grey where the α$_3$ helix of each N-terminal domain binds into a major groove of DNA and the wings are positioned into the minor groove of DNA. Figures taken from (Song et al., 2019)*

These reported truncated CysB structures give insight into the individual regulatory domains and N-terminal domains of CysB. The regulatory domains indicate ligand-dependent structural differences within the inducer-binding cavity and surrounding secondary structures. It is hypothesised that these ligands can cause conformational changes that are relayed to the DNA-binding domains to effect DNA binding through altering the affinity for specific DNA sequences or large-scale conformational changes that may alter the entire protein arrangement.

The absence of a characterised full-length CysB structure prevents conclusions around the arrangement of the functional CysB tetramer and how any conformational changes might be relayed through the protein.

1.9   Research Objectives

The aim of the research conducted herein was to characterise the structure and function of CysB from *N. gonorrhoeae*. Characterisation of CysB activity would provide understanding of how CysB regulates an altered cysteine regulon and whether previously reported activity from other organisms is true of the superbug *N. gonorrhoeae*. Structural characterisation would provide further evidence to understand the mechanisms by which CysB activity is governed. Understanding how *N. gonorrhoeae* regulates gene expression will improve our understanding of the pathogenicity of *N. gonorrhoeae*. The following objectives were set to achieve this:

Objective 1:

Characterise the binding activity of CysB to promoters within the cysteine regulon. This will involve gel-based electrophoretic mobility shift assays (EMSAs) to cysteine regulon promoters and attempts to determine the CysB binding site within the *cysB* promoter.

Objective 2:

Determine the three-dimensional structure of CysB using crystallisation and X-ray diffraction. We will attempt to collect a full-length *apo*-CysB structure, and an inducer bound *N*-acetylserine-CysB structure.

Objective 3:

Investigate large-scale conformational changes in CysB in different inducer states in solution. This will be done by small angle X-ray scattering (SAXS) to visualise any large-scale conformational changes not detectable in crystal structures.

Objective 4:

Partially characterise the role of CysB *in vivo* to investigate its essentiality and potential transcriptional role during stages of infection. This will involve *in vivo* work such as the generation of a knockout and/or knockdown strains.

Chapter 2: Materials and Methods

2.1 Cloning of *cysB* for expression in *Escherichia coli*

The *cysB* gene NGO_1578 was codon optimised for *E. coli* and ordered from Geneart (Thermo Fisher) (Appendix A: Cloning information for CysB). The synthetic *cysB* construct was cloned into expression vector pET28b-PstI with a C-terminal hexahistidine-tag, between PstI and Xhol restriction sites by Dr Joanna Hicks (The University of Waikato). The NGO_1578 pET28b plasmid was checked by DNA sequencing before transformation into *E. coli* BL21 (DE3) for protein expression. Positive transformants were selected for by growing overnight at 37°C in Luria-Bertani (LB)-agar (Appendix A.2) supplemented with 50 µg.mL^{-1} kanamycin.

2.2 Long-term storage of BL21 CysB expression strains

Seeder cultures were prepared from positive transformants, by inoculating single colonies or a loop of glycerol stock into LB broth, (10 mL) supplemented with kanamycin (50 µg.mL^{-1}) and incubated at 37°C, shaking at 180 rpm overnight. Glycerol stocks were prepared from seeder cultures, by adding 0.5 mL of seeder culture to 0.5 mL sterile 50 % glycerol (*v/v*) and stored at -80°C.

2.3 CysB expression cultures

Large scale expression cultures were prepared by inoculating the 10 mL seeder culture (prepared as per Section 2.2) into 1 L LB broth containing 50 µg.mL^{-1} kanamycin. Cultures were incubated at 37°C, 180 rpm until the optical density at 600 nm (OD$_{600}$) was between 0.5 - 0.7 indicating mid-log growth optimal for protein expression. Cultures were induced with 1 mL 0.75 M isopropylthio-β D-galactosidase (IPTG, final concentration 0.75 mM) and incubated at 22°C, 180 rpm overnight.

Expression cultures were centrifuged at 4 600 *g*, 4°C for 20 minutes, after which the supernatant was discarded. The remaining pellet was resuspended in 20 mL of lysis buffer (50 mM Tris pH 8.0, 200 mM NaCl, 20 mM imidazole) and transferred to a 50 mL falcon tube. Resuspended pellets were centrifuged for a further 20 minutes at 4 600 *g*, 4°C, supernatant discarded, and pellets were stored at -80°C.

2.4   Purification of CysB

CysB was isolated and purified using Immobilised Metal Affinity Chromatography (IMAC) and Size Exclusion Chromatography (SEC). SDS-PAGE gel electrophoresis was used to visualise the purity of purified protein and determine solubility. Oligomeric state was measured using gel filtration chromatography. Buffers found in Appendix AB.1.

2.4.1   Immobilised Metal Affinity Chromatography (IMAC)

Cell pellets (Section 2.3) were thawed at room temperature and resuspended in 25 mL lysis buffer (50 mM Tris pH 8.0, 200 mM NaCl, 20 mM imidazole), with the addition of one cOmplete™, Mini, EDTA-free Protease Inhibitor Cocktail tablet (Roche, Switzerland) to inhibit cellular proteases. Cells (25 mL volume) were lysed on ice via sonication using a ¼ " microtip probe (QSONICA Q700) with the following pulse settings: alternating 1 second bursts at 12 % power with 1 second intervals of no sonication (1.5 min of sonication in total).

The sonicated lysate was centrifuged at 9 000 g, 4°C for 20 min to separate soluble and insoluble fractions. The supernatant was filtered through 1.2, 0.45 and 0.2 μm Minisart filters. IMAC purification was carried out using an NGC FPLC system (BioRad). The filtered supernatant was loaded onto a 5 mL Histrap column (Cytiva) pre equilibrated with lysis buffer via the NGC sample pump at a flow rate of 2 mL.min^{-1}. The column was washed with 15 mL lysis buffer at 2 mL.min^{-1} and held at isocratic flow for 15 mL with 3 % elution buffer (50 mM, Tris pH 8.0, 200 mM NaCl, 1 M imidazole) to remove any unbound proteins. Bound proteins were eluted into collection tubes with a gradient of 3-50 % elution buffer over 50 mL at a flow rate of 2 mL.min^{-1} to elute the hexaHis-tagged CysB. Fractions (2 mL) were collected across the gradient and those corresponding to the elution peak on the 280 nm chromatogram were collected and stored at 4°C. Eluted protein (peak fractions) were assessed by SDS-PAGE gel electrophoresis (Section 2.4.3) alongside samples of the pellet (insoluble), supernatant (column load) and other peaks in the 280 nm and 260 nm chromatograms. Fractions were stored for up to 1 day at 4°C prior to concentration and further purification.

2.4.2   Size Exclusion Chromatography (SEC)

For purification of CysB via size exclusion chromatography, IMAC peak fractions were pooled and concentrated in a 10,000 Da MWCO spin concentrator (Pall Corporation) by centrifugation at 3 600 g for 20 min at 8°C or until the volume of protein concentrate was reduced to 5 mL.

The concentrate was manually loaded into a 5 mL injection loop on the Bio-Rad NGC FPLC through a 0.2 μm filter for injection onto the column. The concentrate was run through an S200 16/60 SEC column (Cytiva), preequilibrated with size exclusion buffer (50 mM Tris pH 8.0, 200 mM NaCl) at a flow rate of 0.5 mL.min^{-1}. After sample injection, one column volume (130 mL) of SEC buffer was run at 0.5 mL.min^{-1} with 2 mL fractions collected the entire run. Fractions of eluted protein, as determined by the corresponding peak at 280 nm in the chromatogram, were collected, alongside other peaks in the 280 nm chromatogram which were assessed by SDS-PAGE gel electrophoresis. Protein fractions were quantified by Denovix® and stored at room temperature protected from light (up to 28 days).

2.4.3 SDS-PAGE gel electrophoresis

SDS-PAGE gels were composed of both a resolving and stacking layer, at 12 % and 5 % polyacrylamide, respectively, components are listed in Table 2.1. All gels were made in house in a multi-gel caster (Hoefer) capable of making either 5 or 10 mini gels. The resolving layer was prepared by mixing the components listed in Table 2.1 under "resolving layer" and filling the caster, approximately 3 cm from the top. One millilitre of isopropanol was gently pipetted on top of each of the resolving gels which were left to polymerise (~45 min) at room temperature. Once set, the isopropanol was decanted, and the tops of the resolving gels washed with water to ensure no residual isopropanol. The stacking layer was prepared by combining the components listed under "stacking layer" in Table 2.1, which was added to the top edge of the gel caster. A ten/fifteen well comb was inserted into the top of each gel, before leaving to polymerise at room temperature (~30 min). Gels were stored wrapped in wet tissue at 4°C for up to four weeks.

Table 2.1 Components to make 5x 12 % SDS-PAGE mini gels (Appendix AB.2).

Reagents	Resolving gel (15 %)	Stacking gel (5 %)
MQ H$_2$O	10.05 mL	8.5 mL
30 % acrylamide	12 mL	2.125 mL
Resolving Buffer (1.5 M Tris, pH 8.8)	7.5 mL	-
Stacking Buffer (1.0 M Tris, pH 6.8)	-	1.6 mL
10 % SDS	300 μL	125 μL
10 % APS*	150 μL	63 μL
TEMED	15 μL	6.3 μL

* Fresh ammonium persulfate (APS) stock solution was prepared for each set of gels.

Protein samples were prepared for gel-loading by addition of 5 μL 4x SDS loading dye to 15 μL of sample. Pellet and supernatant samples from IMAC purification were prepared by addition of 5 μL 4x SDS loading dye to 3 μL of sample with 12 μL lysis buffer. All samples were heated to 95°C for 5 min prior to gel loading. Precision Plus Protein™ ladder (Bio-Rad Laboratories, USA) (10 μL) was loaded in the first well of each gel with 20 μL of sample loaded per well. Gels were run in 1 x Tris-Glycine SDS-PAGE running buffer (Appendix AB.3) at 100 V, until the loading dye migrated to the end of the stacking gel where it was increased to 150 V. The gel was stopped when the dye migrated to the end of the resolving gel.

Protein bands were visualised by staining in Coomassie Fairbanks stain (0.05 % Coomassie Blue, 25 % isopropanol, 10 % acetic acid) (Appendix AB.3). Gels were covered with Coomassie Fairbanks stain and heated by microwaving (30 seconds), followed by incubation at room temperature with agitation (5 minutes, 120 rpm). Stained gels were rinsed with distilled water, covered with destaining solution (10 % acetic acid) and again heated by microwaving (30 seconds), followed by incubation at room temperature with agitation (5 minutes, 120 rpm). Gels were left in destaining solution until residual background dye had been removed. Destained gel bands were visualised using an iBright FL1000 (Thermo Fisher).

2.4.4 Measuring the oligomeric state of CysB

An ENrich™ SEC 650 10 x 300 gel filtration column (BioRad Laboratories, USA) was calibrated in size exclusion buffer (50 mM Tris, pH 8.0, 200 mM NaCl) using Gel Filtration Standards (BioRad Laboratories, USA) according to manufacturer's instructions. The

calibration curve can be found in Appendix AB.1. 200 μL of the CysB IMAC concentrate was manually loaded onto an ENrich650 column pre-equlibrated in SEC buffer. The molecular weight of CysB was calculated taking the peak elution volume (V_e) and calculating the K_{av} gel phase distribution co-efficient using Equation 2.1, where V_o is the column void volume and V_c the total column volume. The calibration standard curve relating K_{av} and molecular weight was used to determine the molecular weight of CysB from the calculated weight of K_{av}.

Equation 2.1 Formula for calculating gel phase distribution co-efficient (K_{av}), where V_e is the peak elution volume, V_o is the column void volume and V_c is the total column volume.

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

2.4.5 Measuring protein concentration using the Denovix®

Protein concentration was determined by Denovix® DS-11 FX measuring absorbance at 280 nm. This protein quantification method works by measuring the absorbance of tryptophan residues (280 nm). Protein concentration was calculated using Beer-Lambert law (Equation 2.2), where absorbance readings were corrected by dividing the Denovix® reading (1 Abs = 1 mg.mL^{-1}) by the molar absorption co-efficient of the CysB protein, ($\varepsilon = 0.824$ L mol^{-1}.cm^{-1}) as determined by ProtParam (Gasteiger et al., 2005).

Equation 2.2: Beer-Lambert law, where A = absorbance, ε = molar absorption co-efficient (L mol^{-1}.cm^{-1}), c = concentration (mol.L^{-1}) and l = pathlength (cm).

$$A = \varepsilon c l$$

2.5 Electrophoretic Mobility Shift Assays (EMSAs) of CysB binding activity

Electrophoretic mobility shift assays (EMSAs) were used to investigate CysB's DNA binding activity to DNA promoter fragments hypothesised to contain CysB binding sites. CysB was purified as per Section 2.4.

2.5.1 EMSA assay parameters

CysB protein for use in EMSAs was concentrated from dilute size exclusion fractions as needed (8°C, 5 000 *g*). All DNA fragments and oligos (IDT) were resuspended following manufacturer's instructions. Varying molar ratios of DNA:protein were incubated with 1x DNA binding buffer (40 mM Tris pH 8.0, 10 mM MgCl$_2$, 100 mM KCl, 1 mM DTT) at 37 °C

for 5 minutes. Ratios of DNA:CysB were determined based on one CysB tetramer binding to one DNA strand where the molecular weight of CysB was 140.8 kDa as a tetramer (35.2 kDa per monomer as predicted by Protparam (Gasteiger et al., 2005)). 10 μL samples were combined with 1 μL 10x DNA loading dye (Appendix AB.4) and run at 100 V on either a TBE (Tris-Boric acid-EDTA) gel in 1x TBE buffer or an agarose gel with 1x TAE (Tris-acetate-EDTA) buffer dependent on DNA fragment size. Gels were stained in either 1x SYBR Green EMSA nucleic acid stain (Molecular Probes™, kit E33075) or 1x SYBR Gold Nucleic Acid stain (ThermoFisher) diluted with 1x TBE buffer, protected from light with gentle agitation (details in Sections 2.5.2 & 2.5.3).

2.5.2 TBE gels for EMSAs

TBE gels were prepared similarly to preparing SDS-PAGE gels. TBE gels were prepared with two layers in a "stacking" and "resolving" type format where each layer was a different acrylamide percentage. This was to encourage the DNA:protein complex to enter the gel in the lower percent acrylamide layer, while slowing down the free DNA in the higher percent acrylamide layer to ensure visualisation of the free DNA. Components can be found in Table 2.2. All gels were made in house in a multi-gel caster (Hoefer) capable of making either 5 or 10 mini gels. The higher acrylamide percentage layer was prepared by mixing the components listed in Table 2.1 under "bottom layer" and filling the caster, approximately 3 cm from the top. One millilitre of isopropanol was gently pipetted on top of each of the resolving gels which were left to polymerise (~45 min) at room temperature. Once set, the isopropanol was decanted, and the tops of the resolving gels washed with water to ensure no residual isopropanol. The lower acrylamide percentage layer was prepared by combining the components listed under "top layer" in Table 2.2, which was added to the top edge of the gel caster. A ten/fifteen well comb was inserted into the top of each gel, before leaving to polymerise at room temperature (~30 min). Gels were stored wrapped in tissue soaked in 1x TBE buffer at 4°C for up to one week.

Table 2.2 Components of 8%/10% TBE gel for EMSAs (Appendix AB.2).

Reagent	Bottom layer (10%)	Top layer (8%)
Acrylamide:bis-acrylamide (29:1) (40%)	7.5 mL	4.0 mL
TBE Buffer (10X)	3.0 mL	2.0 mL
MQ H_2O	19.5 mL	14.0 mL
10 % APS*	300 μL	200 μL
TEMED	9 μL	6 μL

* Fresh ammonium persulfate (APS) stock solution was prepared for each set of gels.

Incubated samples were loaded on the TBE gel and run at 100 V. Time trials were completed for each fragment to attempt visualisation of both the free DNA and DNA:protein complex bands on the gel. Gels were stained with either 1x SYBR Green EMSA nucleic acid stain protected from light with gentle agitation (20 minutes) or 1x SYBR Gold nucleic acid stain protected from light (5-10 minutes). Gels were imaged using an iBright FL1000 with default settings for SYBR stain (excitation 495 nm, emission 520 nm).

2.5.3   Agarose gels for EMSAs

Agarose gels (HydraGene LE Agarose) were made with 1x TAE Buffer (Appendix AB.5) at either 1 % or 2 % agarose (*w/v*) (dependent on DNA fragment size). Agarose and 1x TAE buffer were combined and microwaved until the agarose had dissolved. The agarose was cooled briefly and poured into a levelled Mini-Sub Cell GT gel caster (BioRads) or OWL™ D2 Wide-Gel caster (ThermoFisher Scientific). A well comb was inserted into the top (and occasionally middle) of the gel, which was left to set at room temperature (approximately 30-45 minutes). Samples were loaded onto the agarose gel and run at 100 V for 40-100 minutes. Gels were post-stained in 1x SYBR Gold nucleic acid stain protected from light with gentle agitation (120 rpm) until the loading dye had been removed from the gel. Gels were imaged using an iBright FL1000 with default settings for nucleic acid gels (excitation 495 nm, emission 520 nm).

2.5.4   Oligo preparation

Oligos for EMSAs were designed in Geneious Prime (Biomatters) with some sequences containing manually mutated nucleotides. Oligos (IDT) were resuspended according to manufacturer's instructions. Briefly, the DNA was centrifuged at 3 000 *g* for 5 seconds, reconstituted with 1 x TE Buffer (10 mM Tris pH 8.0, 1 mM EDTA) to a final concentration of 100 μM. Equal volumes of forward and reverse oligos were combined and incubated at 95°C

(5 minutes) and cooled slowly to room temperature (>60 minutes). Oligo concentration was determined by Denovix®, oligos were diluted with ddH$_2$O for use in EMSAs and the quality of annealed oligos was checked by TBE gel.

2.5.5 Polymerase Chain Reaction (PCR)

PCR reactions were used to amplify DNA for EMSAs and genomic deletion of *cysB* by spot-transformations (Section 2.8). All PCR reactions were set up using 5x HOT FIREPol® Blend Master mix (Solis BioDyne) with appropriate DNA templates and primers. All sequences for DNA templates, primers are detailed in Appendix AB.5. Primers were resuspended in TE buffer to 100 µM and working stocks made by diluting in MQ water to 10 µM. Reactions were set up as in Table 2.3.

Table 2.3 PCR reaction conditions for DNA amplification.

Component	Volume (µL)
5 × HOT FIREPol® Blend	4
10 µM forward primer	0.5
10 µM reverse primer	0.5
DNA template (5ng/ul final concentration)	
MQ H2O	14.8

PCR reactions were set up in replicates with one negative control, in which the DNA template was omitted and replaced with equal volume of MQ H$_2$O. PCR cycling conditions were as per Table 2.4.

Table 2.4 PCR cycling conditions.

Step	Temperature (^{0}C)	Time (min)	
Polymerase activation	95	15:00	
Denaturation	95	0:20	
Annealing	57*	0:30	Repeat 29x
Extension	72	0:30/1:00**	
Final Extension	72	10:00	
Hold	12	∞	

* Annealing temperature was dependant on specific primers used.

** Extension time was dependent on the DNA product and was based on 1 min extension time per 1 000 DNA bases.

2.5.6 PCR purification and agarose gel electrophoresis

Following PCR amplification, all 20 μL amplified reactions were combined and purified using the High Pure PCR Clean-Up Micro Kit (Roche) or QIAquick® PCR purification kit (Qiagen), dependant on DNA fragment size, as per manufacturer's instructions. Purified DNA concentration was determined by absorbance at 260 nm by Denovix®. Products were diluted using ddH$_2$O if required. Purified PCR product was stored at -20°C.

Agarose gels (HydraGene LE Agarose) made with 1x TAE Buffer and Thiazole Orange DNA stain were made as per Section 2.5.3. The negative control, pre, and post purification DNA samples were prepared for agarose gel electrophoresis by addition of 1 μL 10 × DNA loading dye to 5 μL of DNA sample (concentration dependant). Samples were loaded onto the agarose gel with 7 μL of Invitrogen 1kb+ ladder (Appendix AB.5) and run in a dedicated OWL electrophoresis tank (Thermo Fisher Scientific) containing 1x TAE with a LighteningVolt OSP250-L electrophoresis machine (Biolab scientific) at 100 V for 40 minutes at room temperature. Gels were visualised on an iBright FL1000 with default settings for nucleic acid gels (excitation 495 nm, emission 520 nm).

2.5.7 Thermostability assay with Sypro Orange

Protein stability was assessed via thermostability assays using SYPRO Orange dye across a temperature gradient of 25°C – 99°C. This measures increases in fluoresence as the SYPRO dye binds to hydrophobic regions of the protein exposed during temperature denaturation. Thermostability assays were prepared in PCR tubes by addition of SYPRO Orange dye (final concentration 1x) with CysB protein (5 μL at 0.8 mg.mL^{-1}) and different buffer components to a final volume of 25 μL. PCR tubes were run in a Rotor Gene PCR machine (Corbett Research) over a temperature gradient of 25°C – 99°C. Raw data was visualised using GraphPad Prism version 9.4.1 for windows.

2.6 Structural determination of CysB by X-Ray Crystallography

2.6.1 Protein Preparation

CysB protein was expressed and purified and detailed in Sections 2.3 & 2.4 respectively. Protein concentration was determined by Denovix® as per Section 2.4.5.

2.6.2 High throughput screens

High throughput crystallisation screens were set up to identify promising crystallisation conditions that could be further optimised to produce crystals suitable for X-ray diffraction. These included the following screens: Index™-HR2-144, PEGRx HT™-HR2-086, SaltRx™ 1-HR2107 and the SaltRx 2–HR2109 (Hampton Research, U.S.A.). Screens were set up in low profile 96-2 well INTELLI-PLATEs using the Mosquito® crystallisation robot (TTP LabTech Ltd., USA). Each crystallisation condition had a reservoir solution (crystallisation solution) of 200 μL. Sitting drops (200 nl) consisted of a 1:1 mixture of reservoir solution and concentrated protein (3.4 mg.mL^{-1}). Screens were incubated at 18°C and checked for crystal formation after seven weeks using a dissection microscope.

2.6.3 Hanging drop fine screens

Promising crystallisation conditions were further optimised through scaling up to hanging drop screens. 24 well VDX trays were set up with crystallisation solution (well reservoir) volumes of 500 µl or 1000 µl from concentrated stock solutions. Hanging drops were prepared by mixing a 1:1, 1:2 or 2:1 ratio (1:1 µl, 2:2 µl, 1:2µL or 2:1µL) of crystallisation solution and concentrated protein solution on a siliconised cover slip, which was inverted drop-side down, over a well lined with silicone grease. For each condition, crystallisation parameters such as pH, drop size and the concentration of protein and precipitant were optimised. Crystallisation drops were checked for crystal growth under a dissection microscope, daily for one week and weekly thereafter.

2.6.4 Final structure crystallisation condition

Purified protein was concentrated to 3.4 mg.mL^{-1} and stored on ice for a short period of time. This optimised crystallisation screen was based on a condition identified from high throughput crystallisation tray HR2-144 IndexTM : 35% Tacsimate™ (*v/v*) pH 7.0 (Hampton Research, USA). Tacsimate™ is a mixture of seven organic acids and was used both as a precipitant and a buffer in CysB crystallisation (refer to Appendix AC.1 for Tacsimate™ composition). Trays were set up as per Section 2.6.3. Drops were comprised of 2:2 µl drops (crystallisation solution : protein concentrate). All crystals were grown at 18°C. The crystal from which the X-ray data was collected was grown in 33% (*v/v*) Tacsimate™, pH 6.6.

2.6.5 Crystal preparation for X-ray diffraction

Each crystal was transferred using a cryo-loop, to a cryo-protectant solution, consisting of a modified crystallisation solution with the addition of 20 % glycerol (*v/v*) as a cryo-protectant.

Crystals were then flash cooled in liquid nitrogen. Crystals were stored short-term in liquid nitrogen (two days) before data collection.

2.6.6 Data collection

X-ray diffraction data was collected remotely at the Australian Synchrotron (Melbourne, Victoria) on the MX2 Beamline (McPhillips et al., 2002), equipped with an EIGER x 16M detector (Dectris, Switzerland).

2.6.7 Indexing, integration and scaling

The CysB dataset diffraction images were indexed, integrated and scaled, using XDS (Kabsch, 2010). Merging of reflections was carried out using AIMLESS (Evans & Murshudov, 2013) from the CCP4 suite (Winn et al., 2011). Data quality was assessed through AIMLESS (Evans & Murshudov, 2013). FreeR flag dataset (5 % of dataset reflections not to be used in refinement and used to compute R_{free}) was generated in AIMLESS.

2.6.8 Detection of twinning and non-crystallographic symmetry

Data was analysed for evidence of twinning and translational non-crystallographic symmetry (tNCS) in AIMLESS from the CCP4 suite (Evans & Murshudov, 2013; Winn et al., 2011) and phenix.xtriage from PHENIX suite (Adams et al., 2010; Liebschner et al., 2019). No evidence of twinning or translational non-crystallographic symmetry was detected.

2.6.9 Matthew's coefficient

The total number of monomers in the asymmetric unit was determined by calculating the solvent content using the Matthew's coefficient (Matthews, 1968) program, as a part of the CCP4 program suite (Winn et al., 2011).

2.6.10 Molecular replacement

As no complete structures of CysB have been solved prior to this research, the partial structures of CysB from *Pseudomonas aeruginosa* (5Z4Y & 5Z50) were retrieved from the protein data bank (PDB). A single monomer of each was extracted from the files and 5Z50 was separated into two parts in PyMOL (The PyMOL Molecular Graphics System, Version 2.3.4 Schrödinger, LLC). Molecular replacement was carried out using *phenix.phaser* (McCoy et al., 2007), from the PHENIX suite (Adams et al., 2010; Liebschner et al., 2019).

2.6.11 Model building and refinement

The model was initially built and refined using *phenix.autobuild* (Terwilliger et al., 2008) from the PHENIX suite (Adams et al., 2010). The resulting structure was further built manually using the program COOT (Emsley & Cowtan, 2004; Emsley, Lohkamp, Scott, & Cowtan, 2010). For manual building and refinement in COOT, the 2Fo-Fc and Fo-Fc electron density maps, were set to 1σ and 3σ, respectively. After each round of structure manipulation, *phenix.refine* was used to run rounds of real-space refinement (Afonine et al., 2012).

2.6.12 Ligand fitting and refinement

The structure was prepared for ligand fitting by building and refining the protein structure and adding in waters using *phenix.refine* (Afonine et al., 2012). CheckmyBlob (Kowiel et al., 2019) was used to identify unmodelled density. Visual inspection and manual real space refinement were carried out in COOT (Emsley & Cowtan, 2004; Emsley et al., 2010).

2.6.13 Structural analysis

Assessment of Ramachandran outliers was carried out using *MolProbity* (V. B. Chen et al., 2010), from the PHENIX suite (Adams et al., 2010; Liebschner et al., 2019). Images of the model and maps were generated using programs PyMOL (The PyMOL Molecular Graphics System, Version 2.3.4 Schrödinger, LLC) and COOT (Emsley & Cowtan, 2004; Emsley et al., 2010). Hydrogen bonds were classified as being moderate for distances between 2.5-3.2 Å and weak for 3.3-4.0 Å (Jeffrey, 1997). Bond distances were measured in PyMOL and COOT. Comparing structures was conducted through calculating root mean square distances (r.m.s.d.) using alignment tool in PyMOL. The surface area and buried residue calculations for the tetramer interfaces were carried out using PDBePISA (Krissinel & Henrick, 2007). ENDscript 3.0 was used to identify secondary structural features (Robert & Gouet, 2014), alongside visual inspection in PyMOL.

2.7 Structural determination of CysB by Small Angle X-ray Scattering (SAXS)

2.7.1 SAXS Sample preparation

Samples were prepared for SEC-SAXS analysis by purification methods detailed in Section 2.4. The third and fourth sets of samples had an altered purification method which saw the addition of 10 % *v/v* glycerol in all purification buffers. A high salt wash step during IMAC was also included with 4 M NaCl in lysis buffer (50 mM Tris pH 8.0, 4 m NaCl, 20 mM

imidazole, 10 % *v/v* glycerol) alongside two incubations with Benzonase® Nuclease (Sigma-Aldrich), to the resuspended pellet and pooled IMAC fractions, to attempt to remove any bound nucleic acids from the sample. Where samples were sent for analysis, final volume of 200 μL of each concentrated sample was sent on ice with corresponding SEC buffers to the Australian SAXS/WAXS BeamLine (Ryan et al., 2018). Upon arrival at the Australian SAXS/WAXS BeamLine, samples were stored at 4°C until analysis. Samples were centrifuged and filtered to remove any aggregation that may have occurred during transport and storage of samples. Where samples were analysed in person, ample CysB sample was concentrated to 2 mg.mL^{-1} and transported on ice to the Australian Synchrotron. Upon arrival samples were stored at room temperature, protected from light until analysis. Further concentration (5 000 *g*, 8°C), centrifugation and filtering ensured CysB was free of visible aggregation with Denovix® measurements giving precise information about the concentration of CysB loaded onto the column.

2.7.2   SAXS Data Collection

Measurements were performed at the Australian Synchrotron SAXS/WAXS Beamline (Ryan et al., 2018) equipped with a Dectris-Pilatus3-2M detector. Samples were run via SEC-SAXS analysis, where 50 μL of sample was loaded onto a Superdex 200 Increase 5/150 GL (Cytiva) column at a flowrate of 0.4 mL.min^{-1} with co-flow in place (Kirby et al., 2016). The column was pre-equilibrated with the corresponding buffer, SEC buffer (50 mM Tris pH 8.0, 200 mM NaCl, 10% *v/v* glycerol) or SEC buffer with 10 mM final concentration *N*-acetylserine inducer. Measurements were taken with a sample detector distance of 2.00 m at 12°C with an X-ray wavelength of 1.0322 Å. Data was collected as samples flowed through a 1.5 mm thin-walled glass capillary at one second intervals with an inline UV trace recorded.

2.7.3   SAXS Data Analysis

Several programs from the ATSAS Suite (Manalastas-Cantos et al., 2021) were used in conjunction with the Australian Synchrotron *Scatterbrain* software to analyse the SEC-SAXS data. UV traces from each sample run were inspected in GraphPad Prism version 9.4.1 for windows to confirm uniform species and give an initial indication on data quality. Buffer subtraction to deconvolute peaks was performed in CHROMIXS from the ATSAS Suite (Panjkovich & Svergun, 2018) with manual inspection of image files undertaken in *Scatterbrain* to inform data quality. *Primus* from the ATSAS Suite was used to assess data quality by visual analysis of the Guinier plot, Kratky plot and Porod distribution, while plots

were made using GraphPad Prism version 9.4.1 for windows. Manipulation of the Porod distribution by an artificially large Rmax allowed any subtle aggregation to be detected in samples.

2.7.4 SAXS Model Building

The Porod distribution with an appropriate Rmax was used to generate an output file for each sample for theoretical modelling in DAMMIN (Svergun, 1999), DAMMIF (Franke & Svergun, 2009) and GASBOR (Svergun, Petoukhov, & Koch, 2001) from the ATSAS Suite. Averaging of DAMMIF models and generation of a surface model from scattering data was performed using DAMAVER (Volkov & Svergun, 2003). A theoretical scattering pattern was obtained from CRYSOL from the ATSAS Suite (Franke et al., 2017) using the final X-ray crystallography model.

2.8 *cysB* genomic deletion in *Neisseria gonorrhoeae*

Genomic deletion of *cysB* was attempted to probe the hypothesised essentiality of the *cysB* gene in *N. gonorrhoeae*.

2.8.1 Construct for cysB genomic deletion in *N. gonorrhoeae*

The DNA template for homologous recombination was designed by Dr Joanna Hicks using Geneious Prime (Biomatters) to contain 150 bp of flanking sequence (corresponding to 150 bp upstream and downstream of the *cysB* gene respectively) either side of a kanamycin resistance gene (*kanR*). The DNA construct was ordered as a Geneblock from Twist Bioscience (USA). Lyophilised DNA was resuspended in TE buffer to a final concentration of 20 ng.μL^{-1} for use as a DNA template in subsequent PCR reactions. Primers were designed using Geneious Prime (Biomatters) and ordered from IDT (USA) to amplify the entire DNA construct (CysB KO FWD: 5'- ATAGTCTATCATGCCGAAA -3' and CysB KO REV: 5'- CCTGTACGAACATTTCAGAC -3') which can be found in Appendix AB.5. The construct was purified using a QIAquick® PCR purification kit (Qiagen) following manufacturer's instructions.

2.8.2 Spot transformations of *N. gonorrhoeae*

Purified PCR DNA products (prepared as per Section 2.5.6) were pooled and combined with ultra-distilled water to a final concentration of 20 ng.μL^{-1}. Two 10 μL amounts of PCR product were spotted onto GCB agar (Appendix AB.5) and left to soak into the agar and dry. This was repeated onto New York agar (Fort Richards, New Zealand). A viable *N. gonorrhoeae* colony

was taken from fresh GCB agar and streaked across the plates through each spot of DNA and incubated at 37°C for 48 hours. Single colonies that grew within the spotted DNA circles from both GCB and New York agar were streaked onto GCB agar supplemented with 50 µg.mL^{-1} kanamycin and incubated at 37°C for 48 hours to select for integration of the DNA construct into the *N. gonorrhoeae* chromosome. Primers were designed by Dr Joanna Hicks to confirm integration of the *cysB* knockout construct via colony PCR of any positive transformants (primers can be found in Appendix AB.5).

Chapter 3: Results and Discussion

3.1 Expression and Purification of CysB

In order to successfully characterise CysB from *N. gonorrhoeae,* purified soluble and active protein is required. Prior to the start of this MSc research, Dr Joanna Hicks had cloned CysB into *E. coli* and optimised protein expression and purification (Sections 2.1, 2.2, 2.3 & 2.4). During assay development protein stability issues were encountered which were mitigated with protein purification and storage optimisation.

CysB was successful purified by IMAC and size exclusion chromatography. The IMAC purification chromatogram in Figure 3.1A shows that CysB elutes at ~25 % elution buffer (265 mM of imidazole). Non-specific unbound proteins elute in the flow through and early in the elution gradient. SDS-PAGE gel analysis (Figure 3.1B) indicates that CysB was adequately removed from the cell pellet during sonication as the majority of the protein was present in the fractions corresponding to the chromatogram peak, with the majority of *E. coli* proteins present in the supernatant successfully removed during IMAC purification. As a result, the CysB protein recovered from IMAC purification was relatively pure and shown to be ~35 kDa via SDS gel assessment (indicated by a green arrow in Figure 3.1B), consistent with the predicted molecular weight of the CysB monomer (with His-tag, 35.2 kDa) as determined by Protparam (Gasteiger et al., 2005).
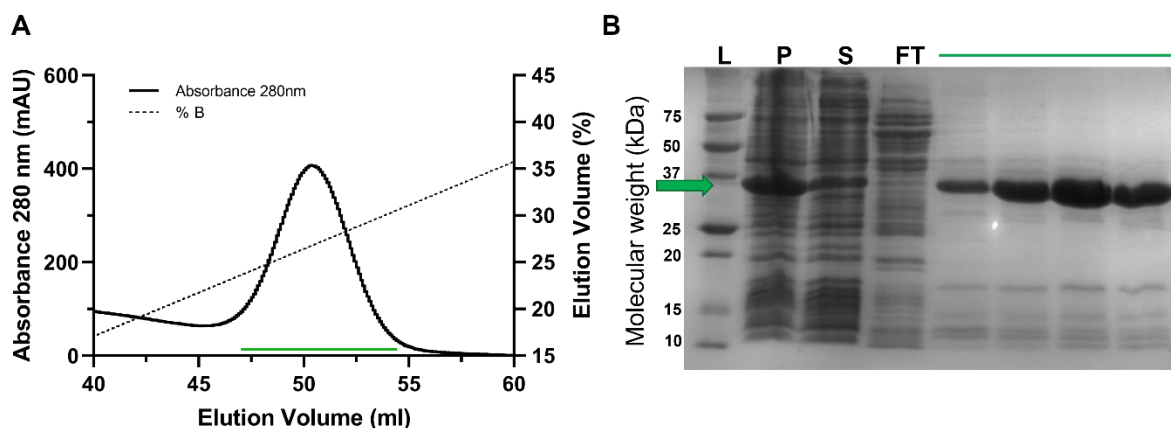
*Figure 3.1 IMAC purification of CysB. **A** IMAC chromatogram of A_{280} nm shows volume (mL) (X-axis), absorbance 280_{nm} (mAU) (left Y-axis) and concentration of elution buffer (%B) (right Y-axis). Green bar indicates collected protein that was assessed by SDS-PAGE gel. Data from protein loading and flow through between 0-40 mL has been excluded for scaling purposes. **B** 12 % SDS-PAGE gel of CysB IMAC purification. CysB protein (35.2 kDa) is labelled with a green arrow and is present in samples insoluble pellet (P), the soluble supernatant (S) and IMAC fractions (green bar). Molecular weights of Precision Plus Protein Standards (L) in kDa are labelled. Purification was carried out at room temperature. Figure generated using Prism version 9.4.1.*

IMAC purified CysB (Figure 3.1, green bar) underwent a second purification step via size exclusion purification (Section 2.4.2). CysB elutes as a single peak from an S200 16 60 SEC column with an elution volume of ~60-75 mL (Figure 3.2A). The corresponding SDS-PAGE gel demonstrates contaminating protein present in the IMAC purified CysB have been successfully removed yielding highly pure CysB protein (Figure 3.2B).
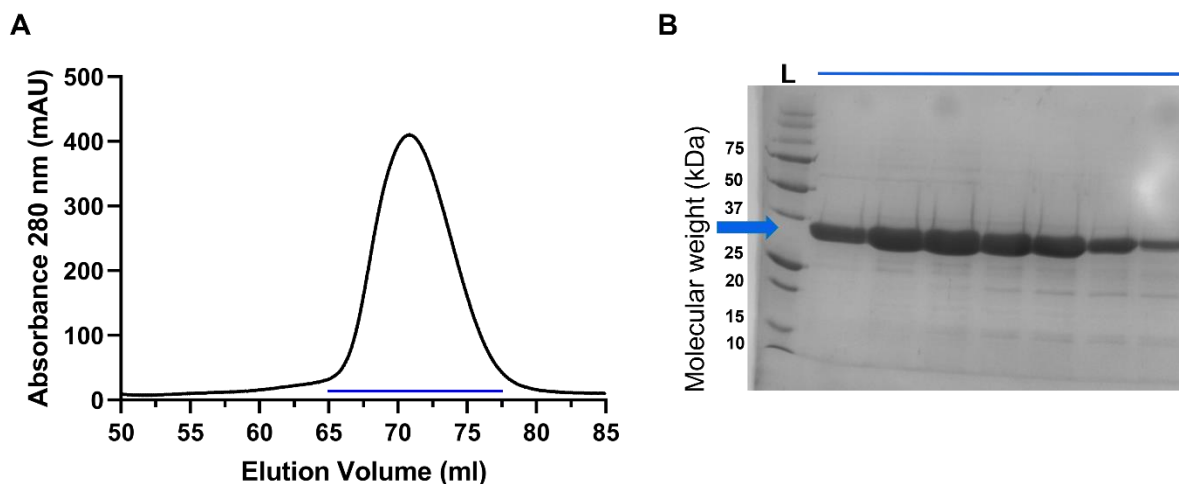
*Figure 3.2 SEC purification of CysB. **A** S200 16 60 size exclusion chromatogram shows elution of a large single peak. Blue bar indicates collected protein that was assessed by SDS-PAGE gel. **B** SDS-PAGE gel of purified CysB from corresponding fractions (blue bar). The blue arrow indicates CysB is present at high purity in all fractions across the peak. Molecular weights of Precision Plus Protein Standards (L) in kDa are labelled. Purification was carried out at room temperature. Figure generated using Prism version 9.4.1.*

IMAC purified CysB was also analysed using an analytical ENrich650 gel filtration column to determine the corresponding molecular weight. CysB eluted as a single peak from the analytical ENrich650 gel filtration column with an elution volume of 12.66 mL (Figure 3.3A). Using the ENrich650 calibration curve (Appendix AB.1), the apparent molecular weight of CysB from the elution volume is 162.22 kDa. The predicted molecular weight of the CysB monomer is 35.2 kDa which suggests CysB elutes as a tetramer (4 x 35.2 kDa = 140.8 kDa). The residual molecular weight of the eluted CysB species may be due to bound nucleic acid present during purification (Figure 3.3B) as CysB is a DNA-binding protein. No protein was present in the void volume of the column (8.57 mL) indicating that CysB did not aggregate and remained soluble during purification.
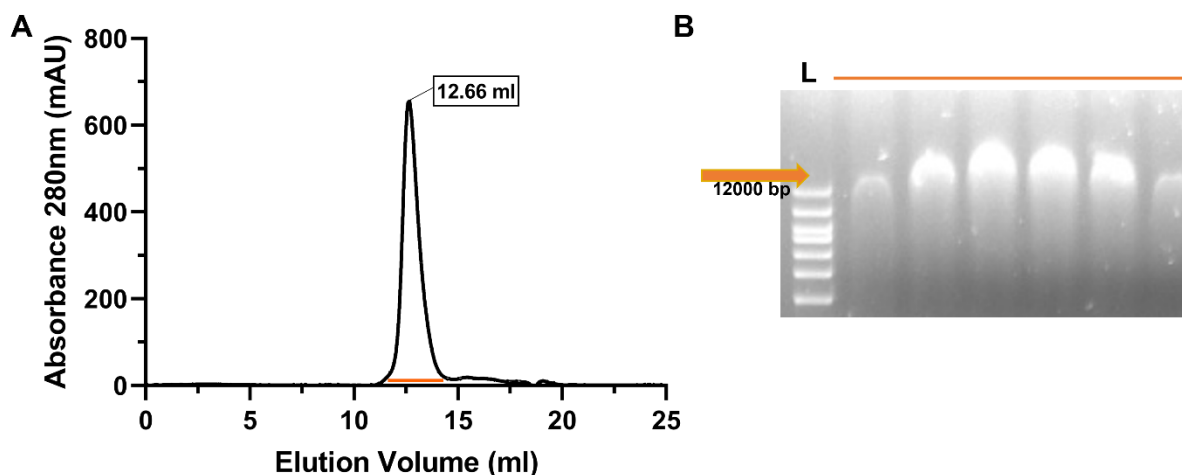
*Figure 3.3 ENrich650 purification of CysB. **A** ENrich650 gel filtration chromatogram shows elution of a single large peak at 12.66 mL. Orange bar indicates collected protein that was analysed by agarose gel. **B** 1 % Agarose gel of CysB fractions (orange arrow) showing bands above the largest nucleic acid standard in the ladder (L) at 12,000 bp indicating the presence of a large species with nucleic acid present. This profile across the fractions corresponds to the same pattern across fractions seen in SDS-PAGE gels indicating it is indeed CysB protein bound to nucleic acids. Figure generated using Prism version 9.4.1.*

CysB was purified for Small Angle X-ray Scattering experiments as per Section 2.7.1. This purification method saw the addition of 10% *v/v* glycerol to all purification buffers and also added two incubations of CysB protein with Benzonase® Nuclease and a high salt wash step with 4 M NaCl salt during the IMAC purification which helped to remove any bound nucleic acids. SDS-PAGE gel analysis saw high purity of CysB across all corresponding fractions with no evidence of residual Benzonase® in the ENrich650 fractions (M_r of 26 kDa monomer) and no evidence of CysB in the high salt wash fraction (Figure 3.4A). These CysB IMAC fractions were also run on an ENrich650 column equilibrated with the appropriate buffer. In this case the elution volume was 12.40 mL and the molecular weight was determined to be 132.08 kDa (Calibration curve available in Appendix AB.1). An agarose gel of the high salt wash fraction shows the presence of DNA at ~150 bp (Figure 3.4B). Again, no protein was present in the void volume of the column (9.25 mL) indicating that CysB did not aggregate and remained soluble during purification upon removal of any bound nucleic acids.
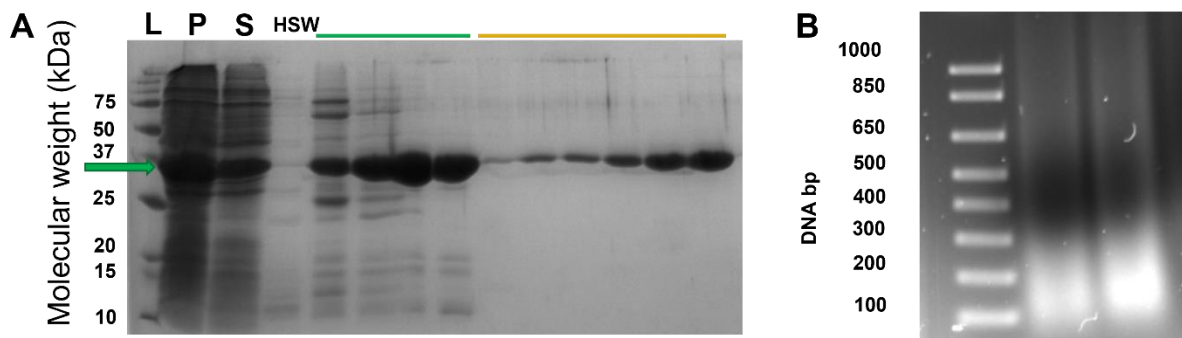
Figure 3.4 SDS-PAGE gel of CysB fractions to remove bound nucleic acids. Fractions from IMAC and ENrich650 purification method altered to remove bound nucleic acids were run on a gel to check purity and removal of nucleic acids. **A** *12 % SDS-PAGE gel showing the insoluble fraction (P), soluble supernatant (S), high salt wash fraction (HSW), IMAC fractions (green bar) and ENrich650 fractions (orange bar). The green arrow indicates CysB protein. No CysB protein or any strong bands indicating proteins were observed in the high salt wash fraction. This indicates that the observed peak in the chromatogram was due to nucleic acids and not protein.* **B** *1 % agarose gel of the high salt wash fractions showing the presence of DNA at ~150 bp.*

These results demonstrate that CysB from *N. gonorrhoeae* is a soluble protein that remains folded during IMAC and SEC purification.

3.2   DNA Binding Activity by Electrophoretic Mobility Shift Assays

The DNA binding activity of CysB has been characterised in homologues from *S. typhimurium* and *E. coli* (Colyer & Kredich, 1996; Lochowska, Iwanicka-Nowicka, Plochocka, & Hryniewicz, 2001) but to date has not been characterised for CysB from *N. gonorrhoeae*. In both *S. typhimurium* and *E. coli*, CysB regulates the cysteine regulon by binding to the promoter region of its own gene, *cysB*, the regions upstream of the *cysPWUAM* and *cysJIH* operons and the single gene *cysK*. Within the *cysJ, cysK* and *cysP* promoters in *S. typhimurium* and *E. coli,* there are multiple convergently arranged half sites (Figure 1.6). The cysteine regulon of *N. gonorrhoeae* has a 3,500 bp genomic deletion resulting in truncated genes, deleted genes and pseudo-genes, deeming the *cysGHDNJI* operon non-functional in an altered cysteine regulon (Hicks & Mullholland, 2018) (Figure 1.3).

O-acetylserine is a pathway intermediate for cysteine biosynthesis and spontaneously isomerises to form *N*-acetylserine. Both *O*-acetylserine and *N*-acetylserine act as inducers for CysB activity, where *N*-acetylserine is reported as a more potent inducer (Mittal et al., 2017; Ostrowski & Kredich, 1989). CysB reportedly binds to half sites within regions upstream of the *cysK* and *cysP* promoters causing a bend in the DNA to 100° (Hryniewicz & Kredich, 1991). Upon binding of *N*-acetylserine, CysB reportedly binds a different combination of half

sites which relieves this bend to 50° (Hryniewicz & Kredich, 1991), while inducer binding releases CysB entirely from the *cysB* promoter (Mittal et al., 2017; Ostrowski & Kredich, 1991). As CysB has been reported to regulate the genes within the cysteine regulon, we set out to investigate the binding activity of CysB to the remaining known gene targets present in *N. gonorrhoeae*. We also set out to investigate any response of CysB to the potential inducer, *N*-acetylserine.

Electrophoretic mobility shift assays (EMSAs) allow a shift in the size of a species to be visualised on a gel. For DNA binding assays, the smaller species (free DNA) runs faster and can be used as a measure of how much DNA has been bound by the protein while the larger species (DNA:protein complex) runs slower. Ideally, both of these species would be visualised on the gel as bands to confirm the DNA of interest becomes bound to the protein.

In preliminary EMSAs both species were able to be visualised on the gel (Figure 3.6). However, these assays had to be re-optimised and a replicable DNA:protein band, with the absence of the free DNA band indicating formation of the DNA:CysB complex, was unable to be visualised. As such, we have reported the presence or absence of the free DNA band as indicative of binding activity, where the presence of the free DNA band indicates no binding activity, while the absence of the free DNA band indicates CysB binding activity to the DNA of interest.

3.2.1   Preliminary Electrophoretic Mobility Shift Assays (EMSAs)

Initial binding activity of CysB was confirmed by assaying binding of CysB to its *cysB* promoter using a SYBR Green nucleic acid stain (Molecular Probes™ kit E33075) which stained for DNA. This assayed binding to a 164 bp substrate that encompassed the region 160 bp upstream from the translational start site (ATG) and the first four nucleotides of the *cysB* gene (Figure 3.5). As the concentration of CysB increased, the free DNA band gradually disappeared while a larger species band corresponding to the DNA:CysB complex appeared (Figure 3.6). This confirmed that CysB from *N. gonorrhoeae* is a functional DNA-binding protein and does indeed bind to its own *cysB* promoter.
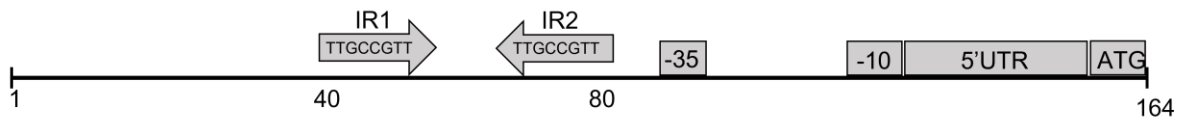
Figure 3.5 cysB promoter in N. gonorrhoeae. 164 bp substrate assayed for CysB binding activity including 160 bp upstream of the ATG translational start codon. The 5' untranslated region (5'UTR), -10 and -35 sites are shown. Two inverted repeats are found upstream of the -35 sequence with the sequence 5'-TTGCCGTT-3'.
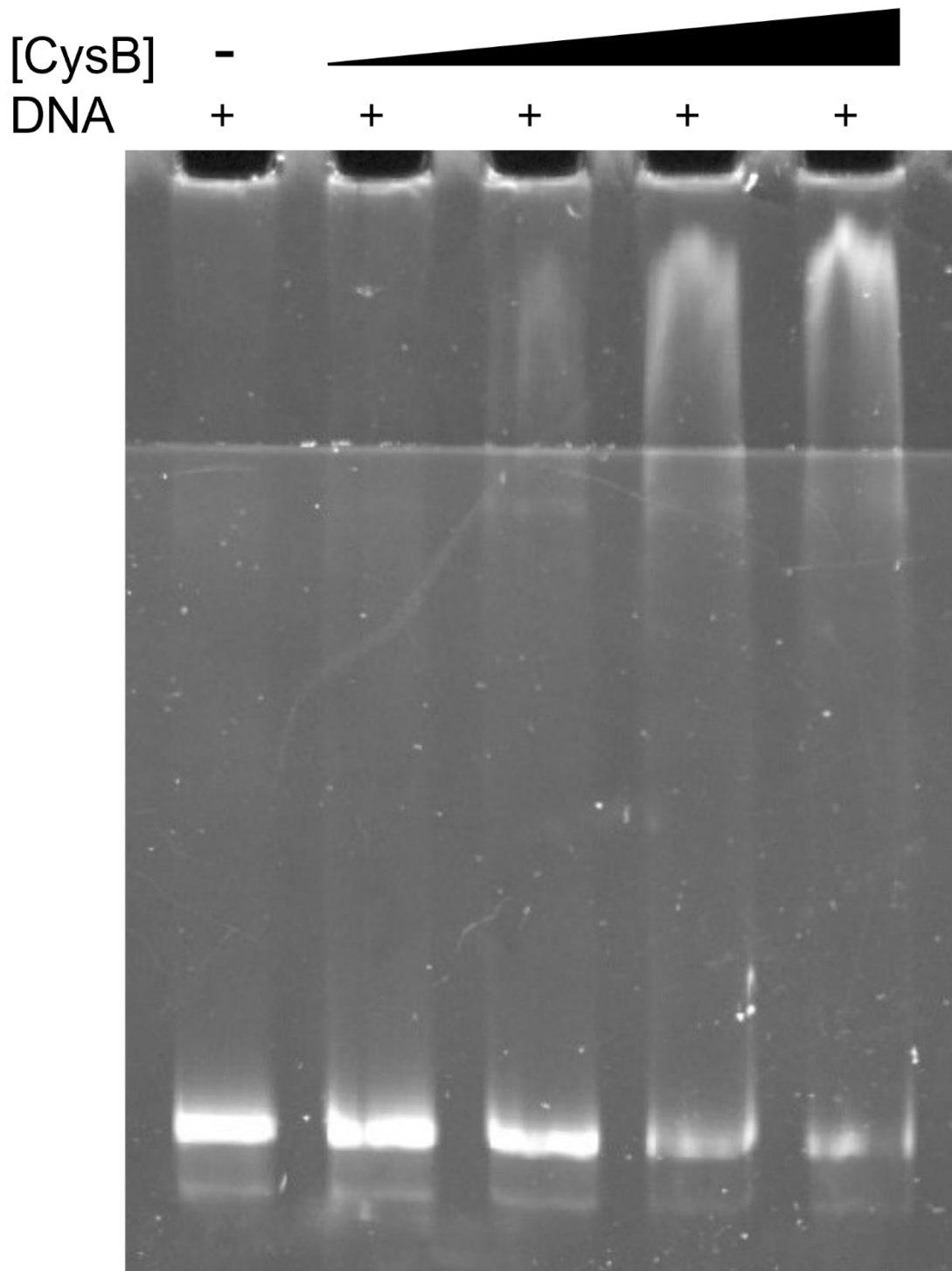
Figure 3.6 Preliminary CysB binding activity. CysB binding to the cysB promoter region 160 bp upstream of the translation start codon ATG (Figure 3.5). The free DNA band is the lower band which can be seen disappearing as increasing CysB is added. The DNA:CysB complex band is the higher band which appears when increasing CysB is added. 20ng DNA was used in each reaction with increasing DNA:CysB molar ratios of 1:5, 1:10, 1:20, 1:50.

3.2.2 Optimisation of EMSAs

Following these initial binding experiments, we went into COVID19 lockdown. Upon return to the lab and repeating these experiments all the DNA was getting stuck in the wells of the gels. Despite extensive optimisation efforts including varying the percent of acrylamide within gels, including multiple different acrylamide percentage layers within the gel to slow down the free DNA and encourage the DNA:CysB complex to run into the gel, varying acrylamide cross linker percentages, different loading dyes, time trials, running the gels at lower speeds and using both SYBR Green and SYBR Gold stains, these results shown in Figure 3.6 were unable to be replicated.

It was noticed during this optimisation that there was DNA present in the well of the protein only samples when stained with SYBR Gold for nucleic acids (Figure 3.7A). We attempted to remove any nucleic acid bound to CysB by the addition of Benzonase® Nuclease during CysB purification and including a high salt wash step during IMAC purification (Section 2.4.1). These attempts appeared successful in removing the majority of bound nucleic acid, as seen in an agarose gel (Figure 3.4) and reduced the amount of DNA seen in the protein only samples on the gel (Figure 3.7B).
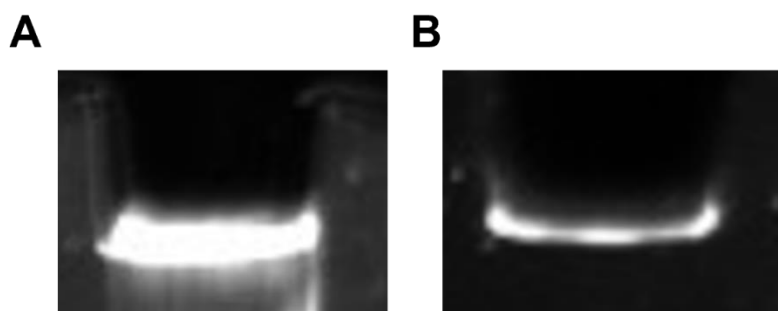
*Figure 3.7 Residual DNA bound to CysB from protein expression. Wells containing CysB protein only were stained for nucleic acids with SYBR Gold nucleic acid stain. Excess nucleic acids can be seen in the wells indicating residual DNA bound to CysB after protein expression and purification. **A** Protein only well with purification method outlined in Section 2.4. The well shows the presence of extensive bound nucleic acids. **B** Protein only well after optimisation of protein purification (Section 2.7.1) to remove any bound nucleic acids. This well shows some residual bound nucleic acid but has a marked reduction from the initial purification method as shown by the well in **A**.*

We hypothesised that the removal of any bound nucleic acids from purification may improve the binding activity seen in the EMSAs, however this did not appear to be the case indicating CysB had higher affinity for the DNA substrates assayed rather than any nucleic acids that remained bound during purification. We also hypothesised that this bound DNA from protein expression may give an indication on the minimum size of DNA required for CysB binding.

CysB, with this bound DNA from protein expression, was incubated with Proteinase K to degrade the protein and run the remaining DNA on an agarose gel. Staining with either thiazole orange or SYBR Gold for increased sensitivity showed no bands indicating the amount of DNA may have been too small to visualise on an agarose gel or the protein may not have completely protected the DNA from Benzonase® degradation.

During these optimisation attempts a pattern among a series of gels was noticed that saw all samples remain in the wells of the gel and not run into the gel when the CysB protein was present or for the DNA only samples without CysB protein present (Figure 3.8A). Following extensive optimisation efforts with acrylamide gels and protein purification, it was investigated whether CysB formed soluble aggregates as an explanation for the consistent lack of successful EMSA results. Protein stability can be investigated via thermostability assays to determine whether different conditions, may affect the stability of the protein. Thermostability assays employ a fluorescent dye, such as SYPRO Orange, which binds to hydrophobic regions of proteins and emit a fluorescent signal. These thermostability assays use a temperature gradient across 25°C – 99°C to stimulate temperature denaturation of the protein. As proteins unfold at higher temperatures the hydrophobic regions become exposed and the SYPRO Orange dye binds resulting in an increase in fluorescent signal. If a protein is aggregated at low temperatures this fluorescent signal will begin high and slowly decrease.

The storage temperature of CysB was optimised to store dilute size exclusion fractions at room temperature protected from light. When concentrating CysB at high g forces, precipitate formed on the centrifugal filters. Lower g forces during concentration minimised visible precipitation and ensured CysB did not become overconcentrated as CysB was prone to aggregation at higher concentrations. Thermostability assays using SYPRO orange showed that CysB was stable under these conditions for up to 28 days post purification (Figure 3.9 – red line).

We therefore hypothesised that there may be a destabilising effect in the DNA binding buffer or running buffer used in the EMSAs. Preparation of fresh binding buffer and fresh TBE buffer resulted in the same effect where all samples remained in the wells of the gels (Figure 3.8A). As such, thermostability assays of CysB using SYPRO Orange, in the presence of different components of the DNA binding buffer, were undertaken to investigate whether any individual components had a destabilising effect on CysB.

The original binding buffer (40 mM Tris pH 8.0, 10 mM $MgCl_2$, 100 mM KCl, 1 mM DTT, 100 $\mu g.\mu L^{-1}$ BSA) was used to characterise the DNA binding of CysB from *S. typhimurium* (Ostrowski & Kredich, 1990) and *E. coli* (Lochowska et al., 2001). Results from the thermostability assays showed that when BSA was included in the DNA binding buffer that all protein was aggregated, including BSA (Figure 3.9 – dark blue line), as shown by the high fluorescent signal at low temperatures of 25°C indicating exposed hydrophobic residues within the BSA protein. It was confirmed that BSA was aggregated under these conditions by a BSA alone thermostability assay which was also aggregated at 25°C (Figure 3.9 – black line). BSA was subsequently removed from the binding buffer. Thermostability assays of CysB with the modified DNA binding buffer (without BSA) and TBE buffer did see a slight destabilising effect on CysB (Figure 3.9 – orange line), but this occurred above temperatures of 40°C. As such, it was ensured that all EMSAs were run with cool flowing water to prevent increased temperatures of the gels and prevent CysB aggregation.

Successful EMSA results were seen following the removal of BSA from the DNA binding buffer. This resulted in visible bands for different sized DNA substrates on the gel and the absence of free DNA bands or differing free DNA band intensities when CysB was present indicating binding activity and hence successful optimisation of EMSAs. While we were still unable to show the DNA:CysB complex band on the gel, the different intensities of the free DNA band with different assayed DNA substrates provided confidence that this was a difference in CysB binding strength. We have therefore reported CysB binding activity based on the presence, absence or varying intensity of the free DNA band when incubated with CysB. Here, the presence of the free DNA band when incubated with CysB indicates no binding activity, lower intensity of the free DNA band indicates some CysB binding activity and absence or near absence indicates strong CysB binding activity.
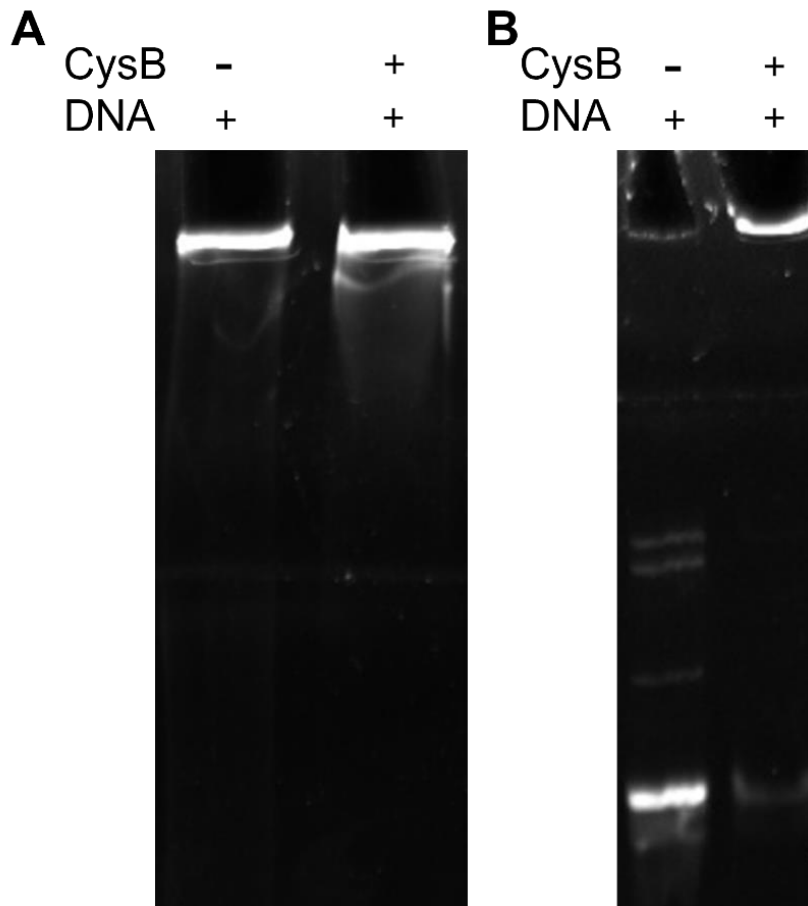
*Figure 3.8 Investigation of DNA binding buffer in EMSAs. All DNA remained stuck in the wells of the gel, even as a DNA only sample, as seen when stained with SYBR Gold for nucleic acids. All reactions contained 20 ng DNA and DNA binding buffer. **A** DNA only and DNA:CysB 1:100 molar ratio samples incubated with original DNA binding buffer containing 100 µg.µL^{-1} BSA. The samples remained in the wells including the DNA only sample due to aggregated BSA under these conditions. **B** The same DNA only and DNA:CysB 1:100 molar ratio sample prep incubated with modified DNA binding buffer without BSA. The DNA only sample was now free to run into the gel as seen by the free DNA band while the DNA and CysB sample remained in the well indicating CysB binding activity to the DNA substrate assayed. This shows a successful EMSA result upon removal of BSA from the DNA binding buffer.*
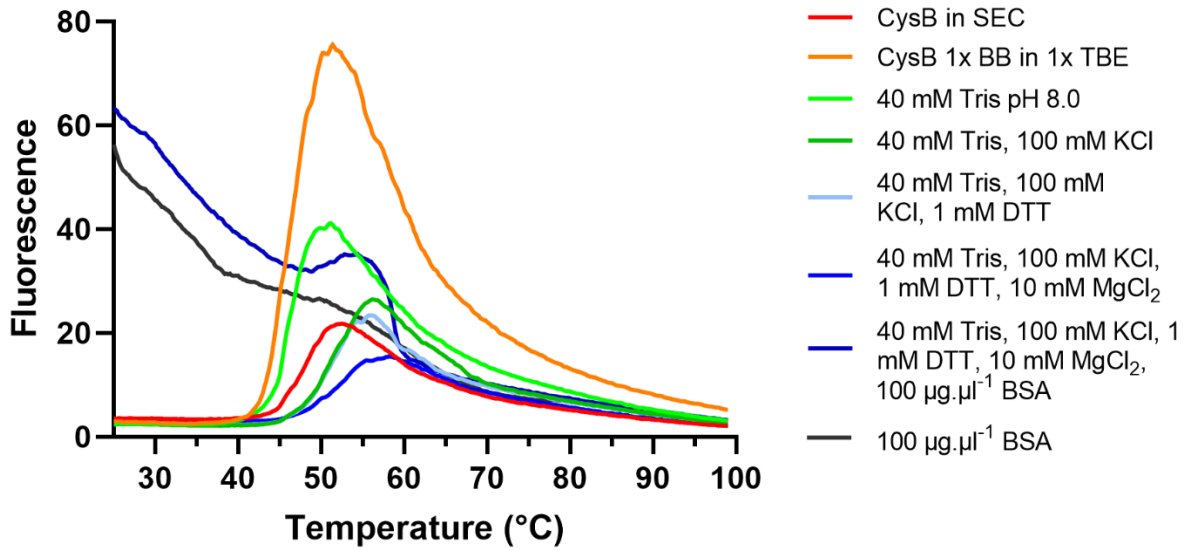
Figure 3.9 Thermostability assay of CysB with DNA binding buffer components. CysB remained stable when stored dilute at room temperature protected from light (red line). CysB with the modified DNA binding buffer (BB) without BSA and in the presence of the running buffer 1x TBE (orange line) showed no aggregation at temperatures below 40°C but did show a slight destabilisation, shown by a shift to the left where the increase in fluorescence begins, compared with when no running buffer was added (medium blue line (40 mM Tris, 100 mM KCl, 1 mM DTT, 10 mM MgCl$_2$)). The different components of the DNA binding buffer without BSA (light green, dark green, light blue and medium blue) show CysB is folded and stable. The addition of KCl further stabilises CysB as shown by a shift to the right in the dark green line compared with the light green line. The addition of MgCl$_2$ further stabilises CysB (medium blue line). Both samples containing BSA (dark blue and black lines) show high fluorescence at 25°C indicating that BSA has aggregated under these conditions at temperatures below 25°C. In the sample with both BSA and CysB (dark blue line) the fluorescence signal starts high and has a second peak where the CysB peak appears in the other samples containing CysB. This indicates that even when BSA was present and aggregated that CysB remained folded. The removal of BSA from the DNA binding buffer (dark blue line, and with running buffer - orange line) show that CysB is stable and folded in solution and does not indicate the presence of any soluble CysB aggregates. Figure generated using Prism version 9.4.1.

3.2.3 CysB binding to the *cysB* promoter

CysB has been reported to negatively autoregulate the expression of its own promoter. It has been hypothesised that when CysB binds to the CysB binding site within the *cysB* promoter, it occludes the RNA polymerase binding site and therefore, when CysB is bound to its own promoter, prevents transcription of *cysB* (Ostrowski & Kredich, 1991). The results presented here (Figure 3.6) show that CysB does in fact bind to the promoter region of the *cysB* gene in *N. gonorrhoeae* and we wanted to further explore specifically where upstream of the translational start codon, ATG, CysB binds. Two avenues were pursued to do this. The first was assaying binding to different substrates within the *cysB* promoter, and the second was

through mutating the inverted repeats upstream of the translation start codon to investigate whether they play an important role in CysB binding.

As there is no consensus sequence that constitutes a CysB binding site, we hypothesised that identifying the exact binding regions within the *cysB* promoter would provide greater understanding of CysB binding sites. Understanding this binding site could allow identification of such binding sites with the promoter regions of other known binding targets, and potentially allow the identification of novel CysB binding targets within the *N. gonorrhoeae* genome. We identified two inverted repeats within the *cysB* promoter. These sequences were 5'-TTGCCGTT-3' and were separated by 19 bp. The upstream inverted repeat (IR1) was 109 bp upstream of the translational start site while the downstream inverted repeat (IR2) was 82 bp upstream of the translational start site (Figure 3.10A). CysB reportedly binds to convergently arranged half sites and we hypothesised that these inverted repeats may be such binding sites. To investigate this, primers were designed to amplify sections of the 164 bp substrate to generate five substrates that encompassed different portions of the *cysB* promoter (Figure 3.10B).
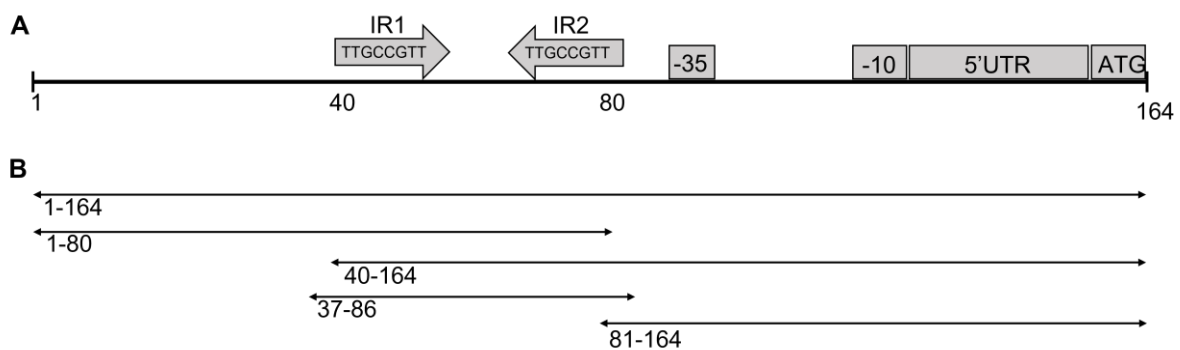
*Figure 3.10 Arrangement of the cysB promoter and substrates of interest for assays. **A** Arrangement of the cysB promoter in N. gonorrhoeae. The 164bp region shown was used for assaying binding which included the translation start site (ATG), 5' untranslated region (5'UTR), -10 and -35 sequences. Two inverted repeats were identified with the sequence 5'-TTGCCGTT-3'. **B** Different substrates within the cysB promoter to investigate the location of important binding regions within the cysB promoter. These substrates are named based on the regions encompassed wihtin the substrates.*

CysB binding to these five substrates within the *cysB* promoter (Figure 3.10B) was assayed. This included the 164 bp substrate that showed initial binding success (substrate 1-164) alongside substrates that included both of the inverted repeats with the sequences either side and a short 50 bp substrate that included only both repeats (substrate 37-86). The results showed that CysB bound to all the substrates assayed but had reduced binding efficiency to the

short 50 bp substrate and also to the substrate that included both inverted repeats and the region upstream, (substrate 1-80). All substrates that contained the 81-164 region (which encompasses the -35 sequence, -10 sequence, 5'UTR and translational start codon ATG) had a complete shift. The 50 bp substrate (substrate 37-86) contained both inverted repeats and did not show a complete shift. These results indicate that there is a sequence downstream of the inverted repeats that is important in CysB binding and that 50 bp may be too short for complete binding or may not contain the correct binding sites.
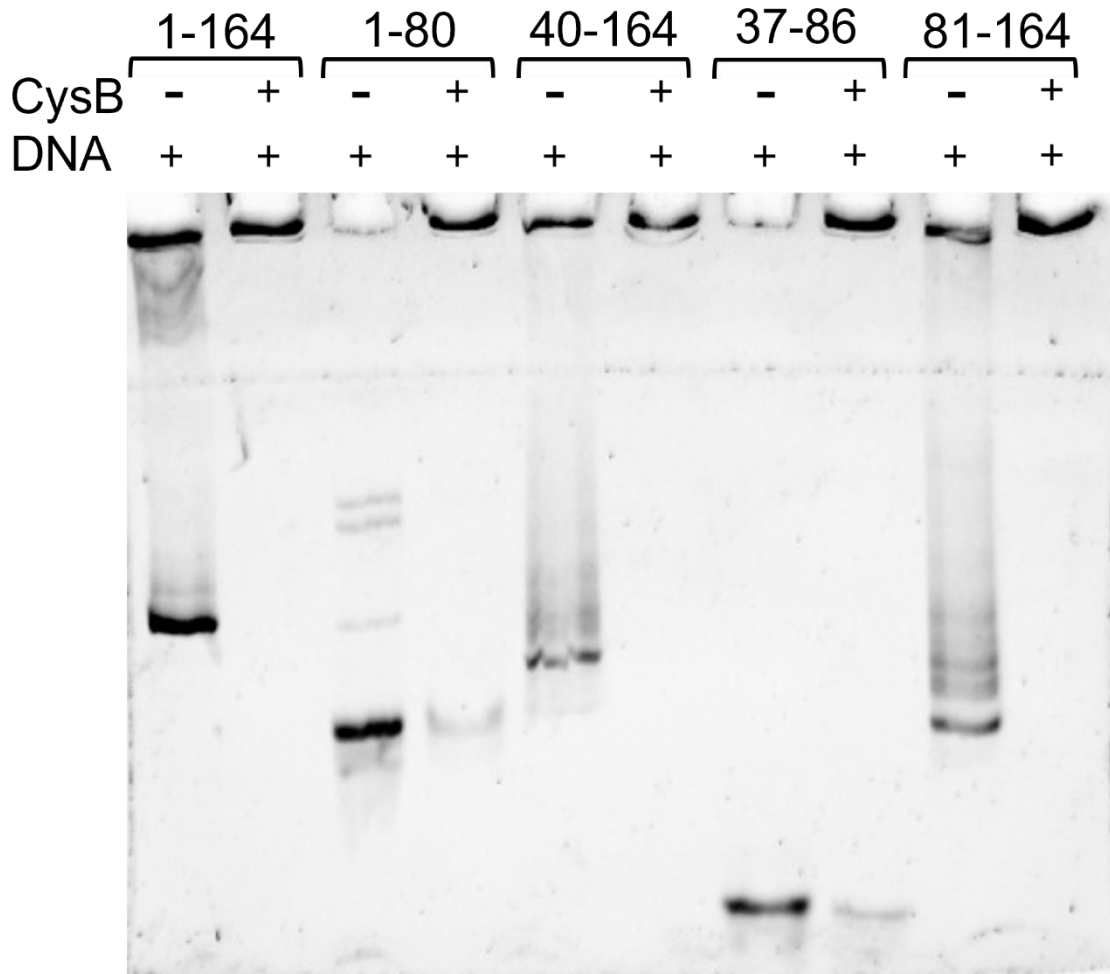
Figure 3.11 CysB preferentially binding regions within the cysB promoter. Five substrates from within the cysB promoter were assayed for CysB binding (Figure 3.10). Each substrate was assayed at 20ng of DNA with a DNA only sample and then with a molar ratio of 1:100 DNA:CysB. The DNA substrates are named based on the region they encompassed upstream of the translational start site (Figure 3.10) where 164 is four nucleotides into the cysB coding region and 1 is 160 bp upstream of the ATG start codon. The absence of free DNA bands in the samples that contained CysB indicate CysB binding to these substrates. The faint free DNA band in the samples that contained CysB with the 1-80 substrate and 37-86 substrate indicate CysB binding but to a lower intensity than samples with a complete shift.

To further investigate the role of these inverted repeats and attempt to elucidate their potential as CysB half sites, these repeats were mutated, as well as the gap between them to alter the position of the repeats relative to one another and binding activity was assayed. To do this, five versions of the 50 bp piece of DNA that encompassed the inverted repeats were generated. In two versions, each inverted repeat was individually manually scrambled from 5'-TTGCCGTT-3' to 5'-AGTCACGT-3' while the other repeat remained 5'-TTGCCGTT-3'. These substrates were termed mIR1 where the upstream repeat (IR1) was scrambled and mIR2 where the downstream repeat (IR2) was scrambled. In another version both inverted repeats were replaced by this scrambled sequence 5'-AGTCACGT-3' which was termed mIR1+2. The two

inverted repeats are separated by 19 bp in the *cys*B promoter, therefore, the final two versions kept the 5'-TTGCCGTT-3' sequences but changed the distance between these repeats to shorten it by 3 bp (short) or extend it by addition of 3 bp (long) (Figure 3.12).
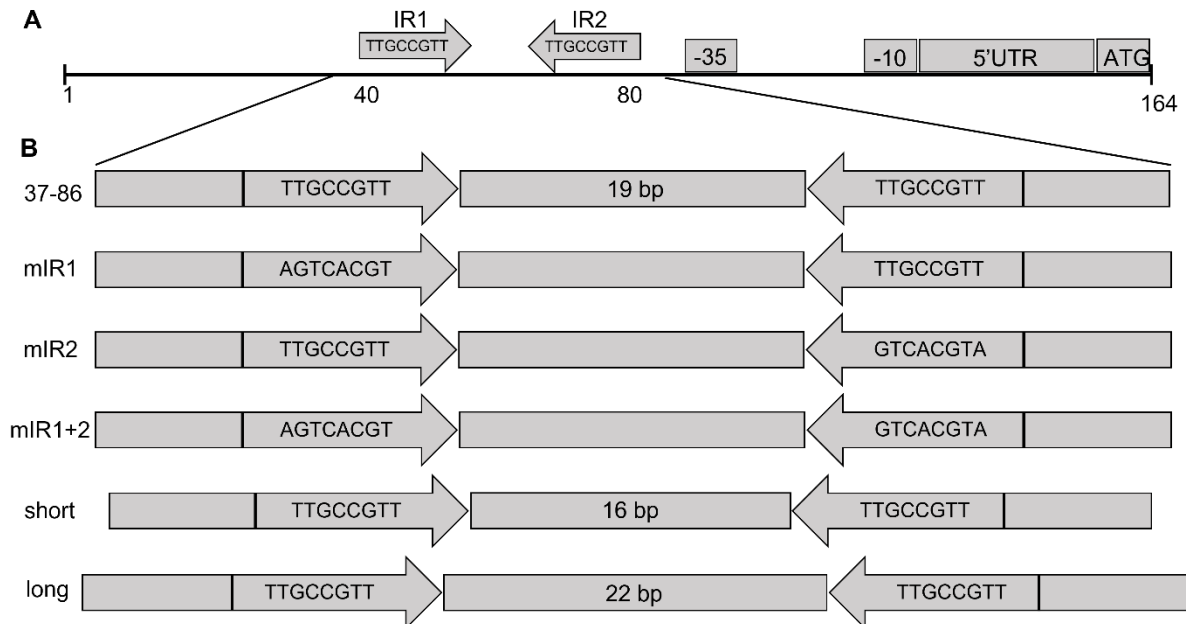
*Figure 3.12 Mutated inverted repeat sequences within the cysB promoter. The inverted repeat sequences were mutated to investigate their involvement in CysB binding. **A** The sequence upstream of the translational start site (ATG) of cysB showing the position of the inverted repeat sequences. **B** The different substrates assayed for binding. 37-86 is a 50bp extract from the cysB promoter sequence encompassing both inverted repeats with the sequence 5'-TTGCCGTT-3'. mIR1 and mIR2 have each mutated either the inverted repeat 1 (IR1) or inverted repeat 2 (IR2) to 5'-AGTCACGT-3' keeping the other repeat unmutated. mIR1+2 has both inverted repeats mutated. 'short' has a reduced number of nucleotides between the inverted repeats as a 47 bp substrate, while 'long' has increased the number of nucleotides by addition of CAA in the middle of the 19 bp distance as a 53 bp substrate. Sequences can be found in Appendix AB.5.*

When both repeats were present unmutated within a 50 bp DNA substrate, there was almost complete binding (Figure 3.13). Mutating the IR1 saw a decrease in binding efficiency shown by the presence of more remaining free DNA, where mutation of the IR2 saw a further decrease in binding efficiency (Figure 3.13). When both of these repeats were mutated (mIR1+2), there was similar binding efficiency as when the IR2 was mutated (Figure 3.13). This may indicate that a 50 bp substrate is too short to see efficient DNA binding or that the mutated sequence retained enough similarity within important nucleotide positions to avoid detrimental effects to DNA binding. Alternatively, this may indicate that the region critical for CysB binding may be downstream of the IR2 as supported by the substrates assayed above (Figure 3.10, Figure 3.11) and any sequences important for binding contained within these inverted repeats may contribute toward binding activity. Interestingly, when the distance between these repeats was

decreased by 3 bp to 14 bp or increased by 3 bp to 20 bp there was a complete absence of any binding activity (Figure 3.13). This indicates that while the inverted repeats may not be the exact CysB binding half sites, their position relative to one another or the position of specific nucleotides within this region may be critical to CysB binding activity.

An unexpected result was the binding of CysB to the promoter region of *cysE* (Figure 3.13). We had initially hypothesised the *cysE* promoter would be a good negative control for CysB binding activity as *cysE* is reportedly constitutively expressed (Jones-Mortimer et al., 1968; Kredich, 1971). However, EMSA results showed that CysB did bind this DNA substrate, and with high efficiency which was unexpected. Additionally, CysB appeared to have a higher affinity for this *cysE* substrate over the other substrates assayed including the 50 bp substrate that contained both unmutated inverted repeats. This may indicate that CysB may exert some regulatory effect on the expression of *cysE* although further investigation is required to confirm this.
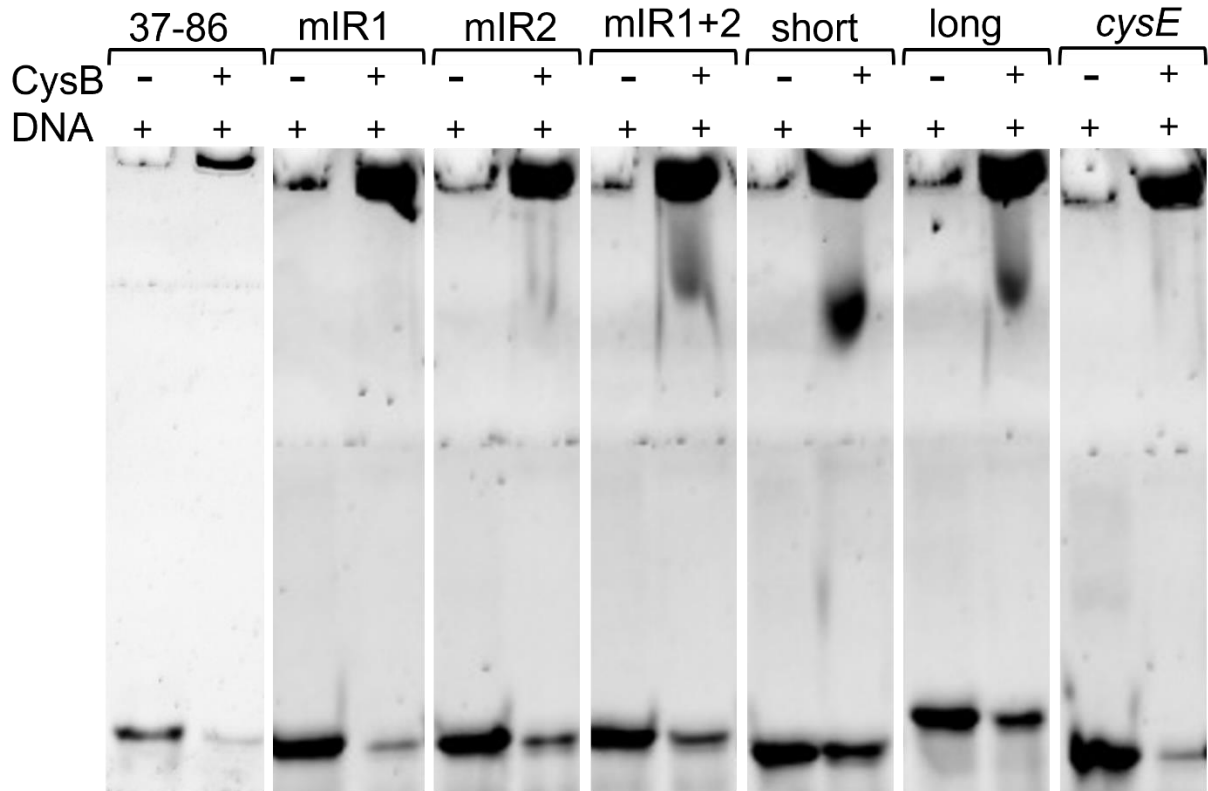
Figure 3.13 CysB binding of mutated inverted repeat substrates within the cysB promoter. Six substrates were assayed for CysB binding (Figure 3.12) alongside a 50 bp extract from the cysE promoter region from N. gonorrhoeae. Each substrate was assayed at 20ng of DNA with a DNA only sample and then with a molar ratio of 1:100 DNA:CysB. The DNA substrates are named based on the mutation (Figure 3.12). The free DNA band at the bottom of the gel is indicative of CysB binding activity. Here the faint band in the DNA:CysB sample with the 37-86 substrate indicates high CysB binding activity where the faded band in the DNA:CysB sample with the mIR1 substrate indicates CysB binding but not to the same extent as seen with the 37-86 substrate. The band in the DNA:CysB samples with the mIR2 and mIR1+2 substrates shows that mutation of the IR2 further decreases Cysb binding. The dark bands in the DNA:CysB samples with the short and long substrates indicate there is little to no CysB binding when the distance between the inverted repeats is altered. The faint band in the DNA:CysB sample with the cysE substrate indicates CysB binding to this substrate also. There is a band starting to appear higher in the gel in the substrates mIR1+2, short and long however, this was unreliable and did not reproduce and so may be ignored.

3.2.4 CysB response to the inducer *N*-acetylserine with the *cysB* promoter

In *S. typhimurium*, when CysB is bound to the *cysB* promoter, the addition of the inducer *N*-acetylserine (NAS) relieves CysB from its own promoter (Mittal et al., 2017). This is seen in EMSAs as the reappearance of the free DNA band, which is thought to free CysB to allow binding to other promoters of genes within the cysteine regulon (Mittal et al., 2017). We attempted to replicate this result by assaying binding of CysB to the 164 bp substrate of its *cysB* promoter in the presence of N-acetylserine. Interestingly, even at concentrations of 30 mM *N*-acetylserine, concentrations where unfolding of the protein was observed in Small

Angle X-ray Scattering data (Figure 3.38C), CysB would not release from the *cysB* promoter (Figure 3.14). Repeated attempts with various *N*-acetylserine stocks, different protein and DNA preparations saw the same result, that the addition of the inducer *N*-acetylserine did not allow release of the *cysB* promoter.
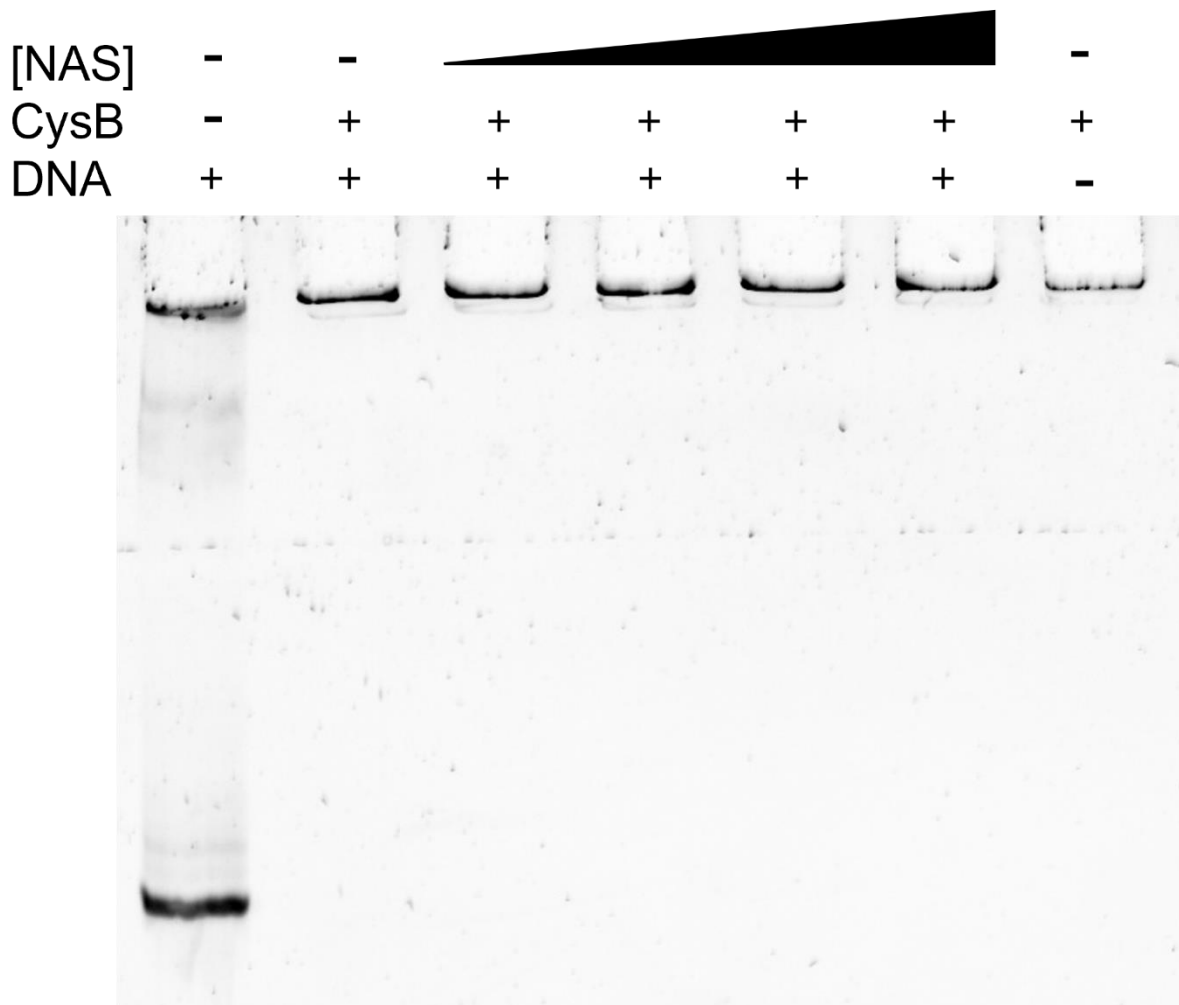
Figure 3.14 Effect of addition of inducer N-acetylserine on CysB binding. The full 1-164 bp substrate from the cysB promoter was assayed at 20 ng DNA, 1:100 molar ratios of DNA:CysB and increasing amounts of the inducer N-acetylserine (NAS) (1 mM, 5 mM, 10 mM, 20 mM final concentrations). The absence of the free DNA band in the DNA:CysB sample without inducer present indicates CysB binding to the 164 bp substrate (previously shown in Figure 3.6 and Figure 3.11). The addition of increasing inducer N-acetylserine is expected to release CysB from the cysB promoter and would show the free DNA band reappearing as the concentration of N-acetylserine increased. The absence of free DNA bands at all N-acetylserine concentrations indicates that CysB remains bound to the cysB promoter in the presence of the inducer N-acetylserine.

This is a particularly interesting result as it contrasts the literature and brings into question the hypothesised mechanism by which CysB may exert its regulatory effects. CysB has been hypothesised to undergo large scale conformation changes upon inducer binding (Mittal et al.,

2017) in line with large scale conformation changes seen amongst the LTTR family member, DntR, upon inducer binding (Lerche et al., 2016). These hypothesised conformational changes within CysB supposedly change the arrangement of the DNA-binding domains which allows them to slide between the multiple half sites reported within the *cysP* and *cysK* promoters to relieve DNA bending of these promoters. These changes in the positioning of the DNA-binding domains may allow CysB to release the *cysB* promoter due to lack of any additional half sites within the *cysB* promoter. CysB from *N. gonorrhoeae* appears to lack the ability to release the *cysB* promoter when the inducer *N*-acetylserine is present. It may be possible that a combination of the primary and secondary inducers *N*-acetylserine and *O*-acetylserine, respectively, may be required to induce any changes. Structural characterisation may provide insight into the potential mechanism by which CysB from *N. gonorrhoeae* may respond to inducers.

3.2.5 Optimisation of DNA substrates and fluorescent DNA substrates

While the SYBR Green or SYBR Gold nucleic acid stain stains for all nucleic acids and allowed the visualisation of the free DNA band, in order to accurately visualise specific and non-specific controls, fluorescently labelled DNA substrates were required. Specific controls ensure that the protein specifically binds to the DNA of interest rather than the fluorophore within the fluorescently labelled DNA substrate. To assay this, the protein is saturated with a non-labelled DNA substrate which is added in excess to the fluorescently labelled DNA substrate of the same sequence. In this control, the non-labelled DNA substrate should outcompete the labelled DNA substrate, and when visualising by florescence the shift should be lost, with the free DNA band becoming visible. Non-specific controls ensure that the protein is binding to sequences within the assayed DNA substrate rather than non-specifically binding, confirming that the protein is folded and functional and that is not a non-specific binder. To assay this, the protein is saturated with a non-specific DNA substrate that the protein is not specific for in the presence of a fluorescently labelled DNA substrate that the protein is known to bind. In this control, the labelled specific DNA substrate should bind, and the shift should remain.

We attempted to make the five substrates within the *cysB* promoter fluorescent by employing a two-round PCR method. Here the first round amplified the desired substrates using forward primers that had an additional 12 bp sequence as part of the primer that would not bind the DNA template. The role of this additional 12 bp sequence was to bind to a second primer ordered that was the complement of this sequence with a 5' 6-FAM™ fluorophore (IDT)

attached. The second round of PCR used this FAM-labelled primer in an attempt to add a FAM tag to each substrate without needing to order a new FAM-labelled primer for each new substrate that may be investigated. This approach showed some success but was unable to generate sufficient FAM labelled material for use in EMSAs due to the complex nature of the approach. We ordered a FAM-labelled primer for the 164 bp DNA substrate that showed success in binding assays (Figure 3.5, Figure 3.6) and attempted a single round of PCR. This was more successful but was unable to be optimised within the timeframe for this work. We also ordered a FAM-labelled oligo and assayed binding to this with specific and non-specific controls.

This substrate was 50 bp and contained both the inverted repeats within the *cysB* promoter (Figure 3.5) where the upstream inverted repeat was mutated from the sequence 5'-TTGCCGTT-3' to 5'-AGTCACGT-3'. This mutated inverted repeat 1 (mIR1) substrate was ordered with a 5' 6-FAM™ fluorophore (IDT) attached to perform specific and non-specific controls. Assaying CysB binding to this FAM-labelled mIR1 substrate showed almost complete binding with only a faint free DNA band remaining (Figure 3.15). This band did not decrease in intensity with increasing concentrations of CysB (Figure 3.15). The FAM-labelled DNA that became bound to CysB is clearly visible in the wells of the gel, while where no DNA was added, in the protein only sample, there is no visible DNA in the well. The specific control (sp) sample shows that the faint free DNA band does not change intensity compared with that of the DNA:CysB samples (Figure 3.15). Here we would expect to lose the shift as CysB would be saturated by and therefore bind the unlabelled mIR1 substrate. This would allow the FAM-labelled mIR1 substrate to be released and be visualised as a bright free DNA band comparable to that of the DNA only free DNA band. As this did not occur, it is possible that the 50x ratio of labelled:non-labelled DNA was too low to see the expected result, where perhaps a 100x ratio of labelled:non-labelled DNA may be sufficient. The non-specific control (non-sp) sample shows the faint free DNA does not change intensity compared with that of the DNA:CysB samples (Figure 3.15). Here we expect the shift remains as CysB preferentially binds to the mIR1 substrate over the 'short' substrate, which has three less nucleotides between two unmutated inverted repeats as a 47 bp substrate, which it does not show binding affinity for. The DNA remaining in the well of the non-specific control sample indicates CysB binds the FAM-labelled mIR1 substrate, which is further shown by the faint free DNA band which remains comparable to that of the DNA:CysB samples.
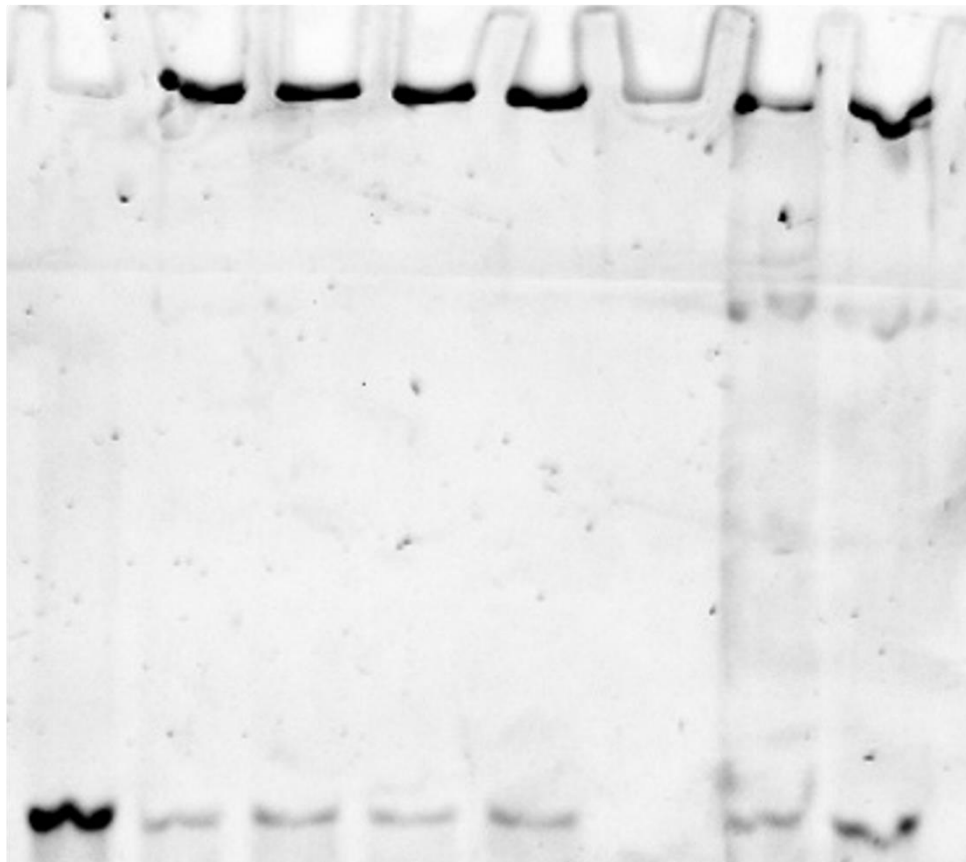
Figure 3.15 Specific and non-specific controls for EMSAs. The mutated inverted repeat 1 (mIR1) substrate with a 5' 6-FAM™ fluorophore attached was assayed for CysB binding and specific and non-specific controls for EMSAs. DNA was assayed at 20 ng in each sample with increasing molar ratios of DNA:CysB of 1:10, 1:25, 1:50 and 1:100. The specific control (sp) consisted of 20 ng FAM-labelled mIR1, DNA:CysB molar ratio of 1:100 and the non-labelled mIR1 at 50x the concentration of the FAM-labelled mIR1. This showed that the shift remained where it was expected to be lost. The non-specific control (non-sp) consisted of 20 ng FAM-labelled mIR1, DNA:CysB molar ratio of 1:100 and the non-labelled, non-specific substrate that CysB doesn't bind, termed 'short', at 50x the concentration of the FAM-labelled mIR1. This showed that the shift remained which indicated CysB bound to the FAM-labelled mIR1 exhibiting specific binding.

While this assay requires further optimisation, it does show that the FAM-labelled substrates allow visualisation of CysB binding. The specific control did not show the expected outcome, where the free DNA band reappears, which may be a result of an insufficient ratio of labelled:non-labelled DNA substrate. Optimisation of this assay should investigate higher ratios of labelled:non-labelled DNA substrate. The inability to optimise generation of fluorescently labelled DNA substrates lead to extensive characterisation of CysB binding to various DNA substrates using SYBR Gold nucleic acid stain.

3.2.6   CysB binding to cysteine regulon genes

Upon investigation of the *N. gonorrhoeae* genome, it became clear there were a few differences amongst the cysteine regulon genes between *N. gonorrhoeae* and *E. coli* and *S. typhimurium.* While other bacteria have two isoforms of the *O*-acetylserine-sulfhydrylase (OASS) enzyme responsible for the second step in cysteine biosynthesis, *N. gonorrhoeae* has only one isoform. This OASS isoform in *N. gonorrhoeae* is denoted as *cysK* and hence no *cysM* is present (in the *cysPUWAM* operon). The *cysP* gene from the *cysPUWAM* operon found in *E. coli* and *S. typhimurium* is replaced by a periplasmic sulfate binding protein (*sbp*) elsewhere in the *N. gonorrhoeae* genome which is hypothesised to bind both sulfate and thiosulfate (Hicks & Mullholland, 2018). While the remaining genes from the *cysPUWAM* operon (*cysU, cysW, cysA*) which are permeases and an ATP-binding subunit, which make up the rest of the transporter, are found as single genes in *N. gonorrhoeae*. Hence the remaining gene targets of the cysteine regulon in *N. gonorrhoeae* are *cysK, sbp, cysU, cysW, cysA* and CysB's own gene *cysB*. We therefore investigated whether CysB binds to the regions upstream of each of these genes *cysK, sbp, cysU, cysW* and *cysA*.

The binding of CysB to the regions upstream of the translational start site within the *cysP* and *cysK* promoters has been reported to cause a bend in the DNA with an approximate angle of 100° (Hryniewicz & Kredich, 1994). This bend is relieved to approximately 50° upon the binding of the inducer *N*-acetylserine which in turn allows transcription of these genes (Hryniewicz & Kredich, 1994). It is hypothesised that this bending and change in bend angle is due to the presence of multiple DNA half sites within the promoter regions and conformational changes that occur within the structure of CysB to promote binding to different half sites. The potential arrangement of which half sites CysB occupies with and without inducer have been partially characterised which make predictions about the exact sequences required for CysB binding difficult.

The binding ability of CysB to the promoter regions of each of the *cysK, sbp, cysU, cysW* and *cysA* genes from *N. gonorrhoeae* was investigated. To ensure that all potential half sites were encompassed, regions ranging from 300 - 502 bp were amplified from the promoters of the cysteine regulon genes in *N. gonorrhoeae* (Figure 3.16). The assayed promoter regions from *cysK* and *sbp* lacked the first 29 bp or 11 bp immediately upstream of the translational start site of the *cysK* or *sbp* genes respectively (Figure 3.16A, B). The *cysK* transcriptional start site is 106 bp upstream of the translational start site and hence excluding the first 29 bp immediately

upstream of the translational start site would likely have still included all CysB binding sites. The *sbp* gene has a truncation in the 5' region of the gene rendering it a pseudogene. The assayed region therefore included most of this truncation to include the historical translational start site and hence likely encompassed any remaining CysB binding sites (Figure 3.16B). The transcriptional start sites for *cysK,* and *cysB* in *N. gonorrhoeae* have been described (Remmele et al., 2014), while the transcriptional start sites for *sbp, cysU, cysW* and *cysA* in *N. gonorrhoeae* have not. The regions assayed from the *cysU, cysW* and *cysA* promoters therefore, encompassed the entire region between the translational start site of the *cysU, cysW* or *cysA* gene, including a portion of the *cysU, cysW* or *cysA* gene coding sequence, and the upstream coding sequence of the upstream gene (Figure 3.16C, D, E). This would likely have ensured all CysB binding sites would have been encompassed in the assayed region.
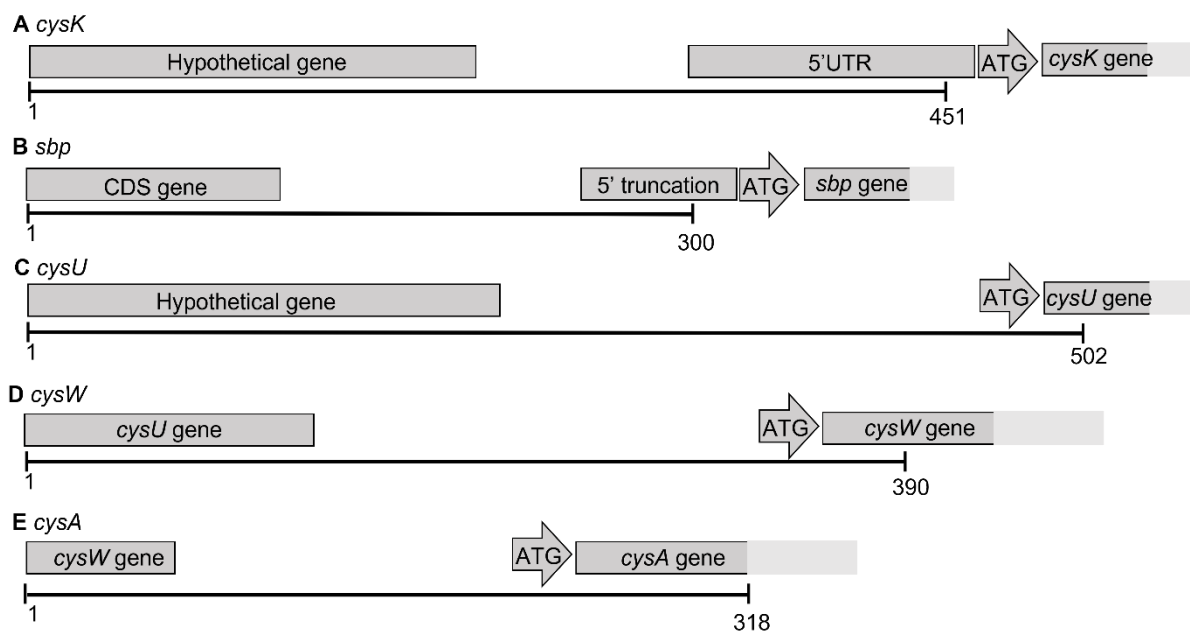
Figure 3.16 Genomic context of assayed cysteine regulon substrates. The 300-502 bp substrates from the cysteine regulon promoters assayed for CysB binding within the N. gonorrhoeae genome. Each substrate encompassed the region upstream of the translational ATG start codon (ATG) and a portion of the upstream gene. Faded grey boxes indicate the downstream coding sequence of the genes. The cysK substrate (A) did not included the ATG start site but did include that majority of the 5'untranslated region (5'UTR). The sbp substrate (B) also did not include the ATG start site but included most of the 5' truncation which would therefore include the historic ATG start site and likely encompass any CysB binding sites. The cysU (C), cysW (D) and cysA (E) genes are located in tandem within the N. gonorrhoeae genome but are separate genes. As the 5' untranslated regions of each of these genes has not yet been characterised (Remmele et al., 2014), each substrate includes the beginning of the relative gene coding sequence (CDS) to ensure all CysB binding sites were likely encompassed. Primer sequences can be found in Appendix AB.5

Results demonstrate that CysB does bind each of these promoter regions of the *cysK, sbp, cysU, cysW* and *cysA* genes (Figure 3.17). This suggests that the regions upstream of each gene contain CysB binding sites and that while each gene is present as a single gene within *N. gonorrhoeae*, these genes remain under CysB regulation when no longer arranged in the *cysPUWAM* operon reported in *E. coli* and *S. typhimurium* (where *cysP* is replaced by *sbp* and *cysM* is absent from *N. gonorrhoeae*).

CysB showed greater affinity for the *cysK* and *cysA* promoters with poorer affinity for the *sbp* promoter where a complete shift is not observed while a complete shift is observed for all other promoter regions. Recently, another member of our lab identified that *sbp* from *N. gonorrhoeae* is a pseudogene and is non-functional *in vitro*. This lack of function of the sulfate binding protein may explain the incomplete binding seen with CysB to the *sbp* promoter. This also poses interesting questions surrounding the role of the other genes involved in the sulfate transporter *cysU, cysW* and *cysA*. As each of these genes are single genes and appear to be under the control of CysB, there may be other functions for these transporter proteins within *N. gonorrhoeae*.
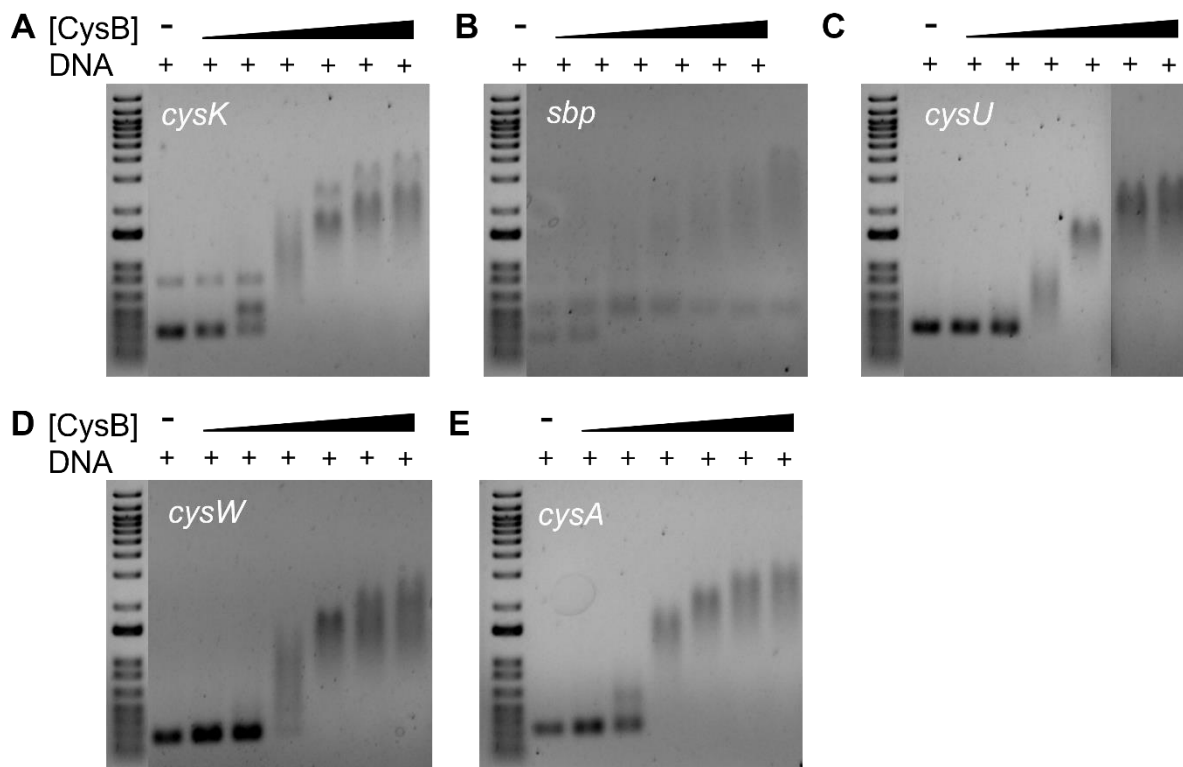
*Figure 3.17 Cysteine regulon promoters with a CysB gradient. DNA was assayed at 20 ng in each sample with an increasing concentration of CysB across the DNA:CysB molar ratios of 1:1, 1:10, 1:25, 1:50, 1:75, 1:100. The cysteine regulon promoters of cysK, sbp, cysU, cysW and cysA were assayed (**A-E**). A complete shift is observed in all promoters except the sbp promoter. Both the free DNA (lower band) and DNA:CysB complex (upper band) can be visualised. The secondary band within the cysK assay was not present when this assay was optimised but appeared after PCR to generate more of the substrate. The same gel shift was observed during optimisation and as such we are confident that the lower band is the cysK substrate of interest which undergoes a shift when it becomes bound to CysB.*

The addition of 10 mM of the inducer *N*-acetylserine saw that CysB remained bound to each of the *cysK, sbp, cysU, cysW* and *cysA* promoters (Figure 3.18). This appears to support the hypothesised mechanism of the multiple half site arrangement of *cysP* and *cysK* promoters that relieves DNA bending keeping CysB bound to the DNA upon inducer binding to these promoters. The arrangement of *cysU, cysW* and *cysA* as single genes requires investigation to identify whether there are multiple half sites within these promoters as well. However, the lack of CysB releasing the *cysB* promoter at concentration of *N*-acetylserine up to 20 mM raises questions around whether any changes in DNA bending angles is occurring within these promoters when 10 mM *N*-acetylserine is present. These substrates had to be run on an agarose gel due to their size. Agarose gels are not sensitive enough to show any changes in the way the DNA:CysB band has run through the gel when *N*-acetylserine was present preventing any conclusions into DNA bending being made. Further experiments to investigate the DNA

bending angle are required to assess whether CysB relieves any bend in the DNA upon inducer binding which may require smaller DNA substrates or lower acrylamide percentage TBE gels. Further experiments also need to be conducted to confirm how many half sites are contained within the promoter regions of the *cysK, sbp, cysU, cysW* and *cysA* genes of *N. gonorrhoeae*.
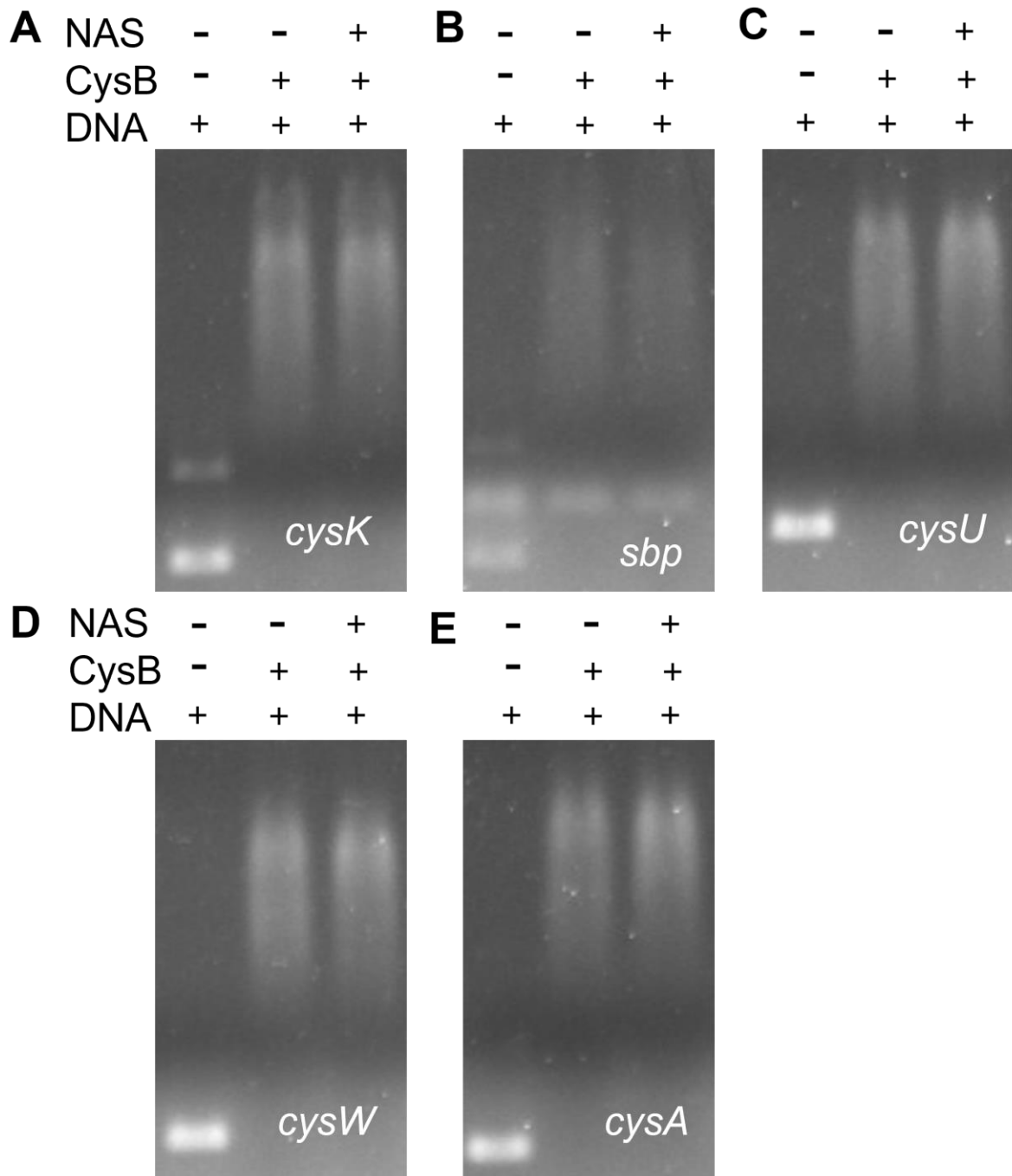
*Figure 3.18 Cysteine regulon promoters with inducer N-acetylserine. Each of the cysteine regulon promoters from cysK, sbp, cysU, cysW and cysA were assayed at 20 ng DNA. Each assay consisted of a DNA only sample, DNA:CysB at a molar ratio of 1:100 and the same DNA:CysB ratio with 10 mM N-acetylserine (final concentration). The addition of 10 mM N-acetylserine retained the shift observed when CysB bound to the cysteine regulon promoters of cysK, sbp, cysU, cysW and cysA (**A-E**).*

67

The results presented here demonstrate that CysB from *N. gonorrhoeae* is a functional DNA-binding protein. We have shown that CysB does bind its own *cysB* promoter and preferentially binds to sequences within this indicating that the CysB binding site may be located within the region 82bp upstream of the *cysB* translational start site. While the inverted repeat sequences within the *cysB* promoter do not appear to be the exact binding sequences, they appear to impact the strength of CysB binding where mutation of these sequences reduces the strength of CysB binding. By altering the distance between these inverted repeats, both decreasing and increasing the distance between the repeats, CysB binding is almost lost completely. This indicates that while the repeat sequences are important, the distance relative to one another appears more critical for binding activity. We hypothesise from these results that there may be just a few nucleotides that are required to promote CysB binding where the distance between these important nucleotides is critical for binding. As the wHTH DNA-binding motif within CysB binds into the major and corresponding minor grooves of DNA, perhaps the nucleotides required for binding must be positioned at corresponding major and minor grooves to promoter CysB binding. Structural studies would help to understand how CysB is positioned to bind the DNA and may explain these findings. Further mutational studies should attempt to elucidate the exact binding region and in particular any nucleotides critical for CysB binding activity.

The EMSA results reported here were replicable and reliable and as such give us good confidence in the CysB binding activity assayed here. These results also satisfy the requirements of the specific and non-specific controls, albeit in a more convoluted way. These assays were conducted using DNA substrates without any modifications or fluorescent tags. The successful binding results (Figure 3.6) indicate that CysB binds to the DNA substrates in absence of any fluorescent tag satisfying the need for specific controls. By assaying binding of CysB to multiple DNA substrates upon the same gels, the binding results showed that CysB preferentially bound to certain substrates, as shown when comparing binding between mIR1 and the short or long substrates within the *cysB* promoter (Figure 3.13). These results, alongside the thermostability assays show that CysB is a folded, functional, specific binder, satisfying the need for non-specific controls. As such we have high confidence that the results reported here are true and reflect CysB binding activity.

CysB also binds to the promoter regions of each of the separate genes *cysK, sbp, cysU, cysW* and *cysA*. This result suggests that CysB binding sites must be present within the regions of each gene assayed. *sbp* from *N. gonorrhoeae* has a truncation in the 5' portion of the gene

rendering it as a pseudogene. This 5' truncation may explain the partial shift observed when CysB was present as the promoter region of *sbp* may have also been altered. If there are just a few nucleotides important for CysB binding which have strict distances between them to promote CysB binding, as our results imply, these nucleotides or the position of them relative to one another may have been affected when the 5' portion of the *sbp* gene was truncated. The partial shift may also be a result of the region of DNA assayed as it did not include the first 29 bp upstream of the functional translational start site. *sbp* as a pseudogene may explain why the *cysU, cysW* and *cysA* genes are separate genes in *N. gonorrhoeae*. The periplasmic sulfate binding protein (*sbp*), two permeases (*cysU* and *cysW*) and ATP-binding subunit (*cysA*) form the ABC transporter which is responsible for sulfate uptake. Genomic analysis of *N. gonorrhoeae* showed *sbp* was the only sulfate transporter present and may be responsible for both sulfate and thiosulfate uptake (Hicks & Mullholland, 2018). The discovery of *sbp* as a pseudogene, and CysB binding sites upstream of each of the *cysU, cysW* and *cysA* genes may indicate that these genes remain important for sulfur transport for use in cysteine biosynthesis and hint there may be other transporters or mechanisms of sulfate transport/reduction/generation to satisfy the sulfur requirement for *de novo* cysteine biosynthesis.

When the inducer *N*-acetylserine is present, CysB remains bound to the promoter regions of the *cysK, sbp, cysU, cysW* and *cysA* genes, as well as the *cysB* promoter region. While this supports the hypothesised mechanism of relieving DNA bending within the *cysK* and *cysP* promoters, CysB's inability to release the *cysB* promoter is interesting.

Further work is required to understand if any DNA bending is occurring within the promoter regions of the *cysK, cysU, cysW* and *cysA* genes and whether the inducer *N*-acetylserine relieves this bending. Further work is also required to understand any changes that may or may not be occurring within the CysB protein when bound to the *cysB* promoter in the presence and absence of inducer. X-ray crystallography and solution-based small angle X-ray scattering of CysB without and with the inducer *N*-acetylserine may help understand any changes that may, or may not, be occurring under these conditions. The presence of any half sites within the *cysU, cysW* and *cysA* genes should also be investigated.

3.3   Structural characterisation of CysB by X-ray crystallography

CysB is a transcriptional regulator of the cysteine regulon and member of the LysR-type transcriptional regulator (LTTR) family. Structural characterisation of the CysB protein can

give insight into the arrangement of the protein, and of particular importance, arrangement of the DNA-binding domains. To date, there are limited crystal structures of CysB in the PDB from just three organisms, with all deposited structures being either the individual ligand binding domain or the DNA-binding domain, respectively. These structures describe truncated forms of the protein and demonstrate a dimeric arrangement of the truncated domains, where gel filtration studies describe a tetramer of the full-length CysB protein. There are also limited deposited structures of LTTR family members, highlighting the difficulty in working with these proteins and indeed successfully crystallising them. As there are no full-length structures of CysB, the arrangement of the tetramer is not well understood. As such, we aimed to purify, crystallise and determine the structure of CysB from *N. gonorrhoeae* using X-ray crystallography in the hopes of solving the first tetrameric CysB structure.

3.3.1 Crystallisation of CysB

As part of an undergraduate research project, I set up high throughput crystallisation screens of CysB (Section 2.6.2). This identified multiple conditions that supported crystal growth seven weeks after preparing the screens. As part of a summer research project, I optimised these crystallisation conditions through hanging drop fine screens (Section 2.6.3). The majority of crystals grown in these conditions diffracted above 3.0 Å, which was to low resolution for structure solving and required further optimisation. At the beginning of this master's research, I collected a single dataset diffracting to 2.73 Å from a fine screen condition. The crystals from this condition were grown in a 4 μL drop with the condition consisting of 33 % *v/v* Tacsimate ™, pH 6.6 where 2 μL of CysB protein at 3.4 mg.mL^{-1} was added to 2 μL of the crystallisation solution. The condition showed crystals grew within 24 hours. The crystals were looped seven weeks after initially appearing, were coated in a cryoprotectant (20 % *v/v* glycerol in fresh precipitant solution) for ~30 seconds and flash frozen in liquid nitrogen for data collection at the Australian Synchrotron MX2 Beamline (McPhillips et al., 2002). From one crystal a single dataset was collected from a 360° rotation at 2.73 Å which was used for downstream data processing.

3.3.2 CysB data processing

Data collected was initially indexed in space group P6$_1$ 2 2 but was determined to have been incorrectly assigned using *zanuda* (Lebedev & Isupov, 2014) from the CCP4 suite (Winn et al., 2011). The data was subsequently reindexed in space group P6$_1$ (based on *zanuda*_output), integrated and scaled using XDS (Kabsch, 2010). A FreeR flag dataset was generated using

AIMLESS from the CCP4 suite (Winn et al., 2011) for refinement and dataset quality statistics (Table 3.1). Data quality statistics indicated the dataset was of good quality with parameters meeting the minimum data quality thresholds. AIMLESS and *phenix.xtriage* analysis of twinning and translational non-crystallographic symmetry (tNCS) showed the dataset had low percentage of twinning fraction (0.021), therefore no attempt to detwin the dataset was made. The dataset collected showed good quality data to a resolution of 2.73 Å, as such, resolution was not cut back, and the data were used for downstream processing.

Table 3.1 Data collection statistics for CysB. Statistics for the highest resolution shell are in brackets. Minimum thresholds are shown beside the parameters in brackets. Data statistics generated by AIMLESS and phenix.tableone.

Data Statistic	CysB structure	
Space group	P 61	
Wavelength (Å)	0.953739	
Cell dimensions		
a/b/c (Å)	94.826 / 94.826 / 341.704	
α/β/γ (°)	90 / 90 / 120	
Monomers in asymmetric unit	4	
Resolution range (Å)	41.96 – 2.733	(2.831 – 2.733)
Number of observed reflections	91251	(8887)
Number of unique reflections	45667	(4484)
R_{merge} (≤ 0.8)	0.01787	(0.2979)
R_{pim}	0.01787	(0.2979)
$CC_{(1/2)}$ (≥ 0.3)	1	(0.841)
Mean I/σI (≥ 2.0)	24.94	(2.45)
Completeness ($> 95\%$)	99.75	(98.46)
Multiplicity	2.0	(2.0)
Wilson B factor (Ang^2)	75.53	

3.3.3 Solving the CysB structure

There are no full-length homologous CysB structures currently deposited in the PDB. There are, however, truncated structures from three organisms that report either the C-terminal regulatory domain or the N-terminal DNA-binding domain separately. The regulatory domain has been solved from *K. aerogenes* (1AL3, 1.80 Å), from *S. typhimurium* in complex with

different inducers and anti-inducers (4GXA, 4GWO, 4M4G, 4LQ5, 4LQ2, 2.4-2.8 Å) and from *P. aeruginosa* (5Z50, 2.21 Å). The DNA-binding domain has only been solved from *P. aeruginosa* (5Z4Y, 2.40 Å). These truncated forms showed good sequence similarity (≥ 30%) and conservation of secondary structures (Figure 3.19) identifying them as good candidates for molecular replacement. The models from *P. aeruginosa* had good sequence similarity to that of CysB from *N. gonorrhoeae*, 54.5 % similarity to the DNA-binding domain (5Z4Y) and 49.7 % similarity to the regulatory domain (5Z50). As both the regulatory domain and DNA-binding domain had been solved from *P. aeruginosa*, and these models had good sequence similarity to CysB from *N. gonorrhoeae*, these models (5Z4Y & 5Z50) were used as search models for molecular replacement.

Solving this structure was a difficult process with iterative rounds of manual building in COOT (Emsley & Cowtan, 2004; Emsley et al., 2010) from the CCP4 suite (Winn et al., 2011) and refinement using *phenix.refine* (Afonine et al., 2012) to generate a logical model from the individual domains used as search models. Ambiguous density due to the 2.73 Å resolution also made positioning of sidechains difficult.
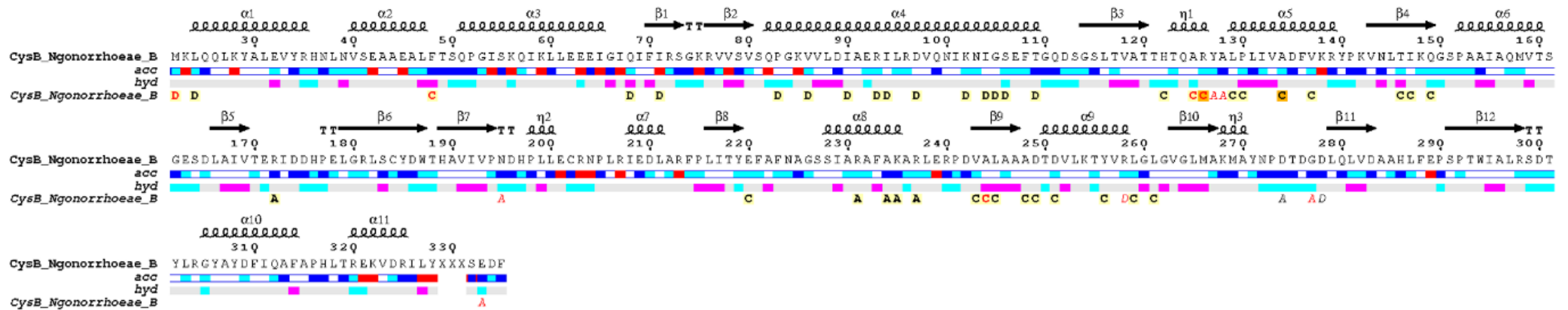
Figure 3.19 ENDscript analysis of the secondary structure of CysB from N. gonorrhoeae. For each residue solvent accessibility and hydrophobicity is labelled below the sequence. Solvent accessibility is labelled "acc" and is rated in terms of blue being accessible, cyan less accessible and white being completely buried from solvent (red being not calculated). Hydrophobicity is labelled as "hyd" and is ranked as cyan being hydrophilic, purple being hydrophobic and white being neutral. Secondary structure elements are labelled in grey above the peptide sequence, with α-helices, β-strands, and β-turns, represented by linked loops, arrows and TT, respectively. Letters indicate noncrystallographic interaction with chain denoted by letter. Refer to Appendix AC.2 for chain labelling of the monomers in the ASU. Chain B was used for analysis. Figure generated using ENDscript 2.0 (Robert & Gouet, 2014).

The Mathew's co-efficient was used to predict the number of monomers in the asymmetric unit (Matthews, 1968). The most likely number of monomer copies was determined to be four, calculated with a predicted molecular weight of 35.2 kDa for the CysB monomer (molecular weight calculated using Protparam (Gasteiger et al., 2005). Thus the *P. aeruginosa* CysB models (5Z4Y & 5Z50) were used as search models for molecular replacement with four monomers in the asymmetric unit.

Initial attempts using the *P. aeruginosa* DNA-binding domain (5Z4Y) and the regulatory domain (5Z50) as models for molecular replacement generated multiple solutions with poor TFZ and LLG scores. We hypothesised that this was due to the cross-over regions between domain I and domain II of the regulatory domain which allows flexibility within the structure. We therefore split the regulatory domain into two models to allow for structural flexibility (Figure 3.20). When the DNA-binding domain and the split regulatory domains were used as a search model, Phaser found one solution with acceptable TFZ and LLG scores of 13.6 and 410.135 respectively (McCoy et al., 2007).
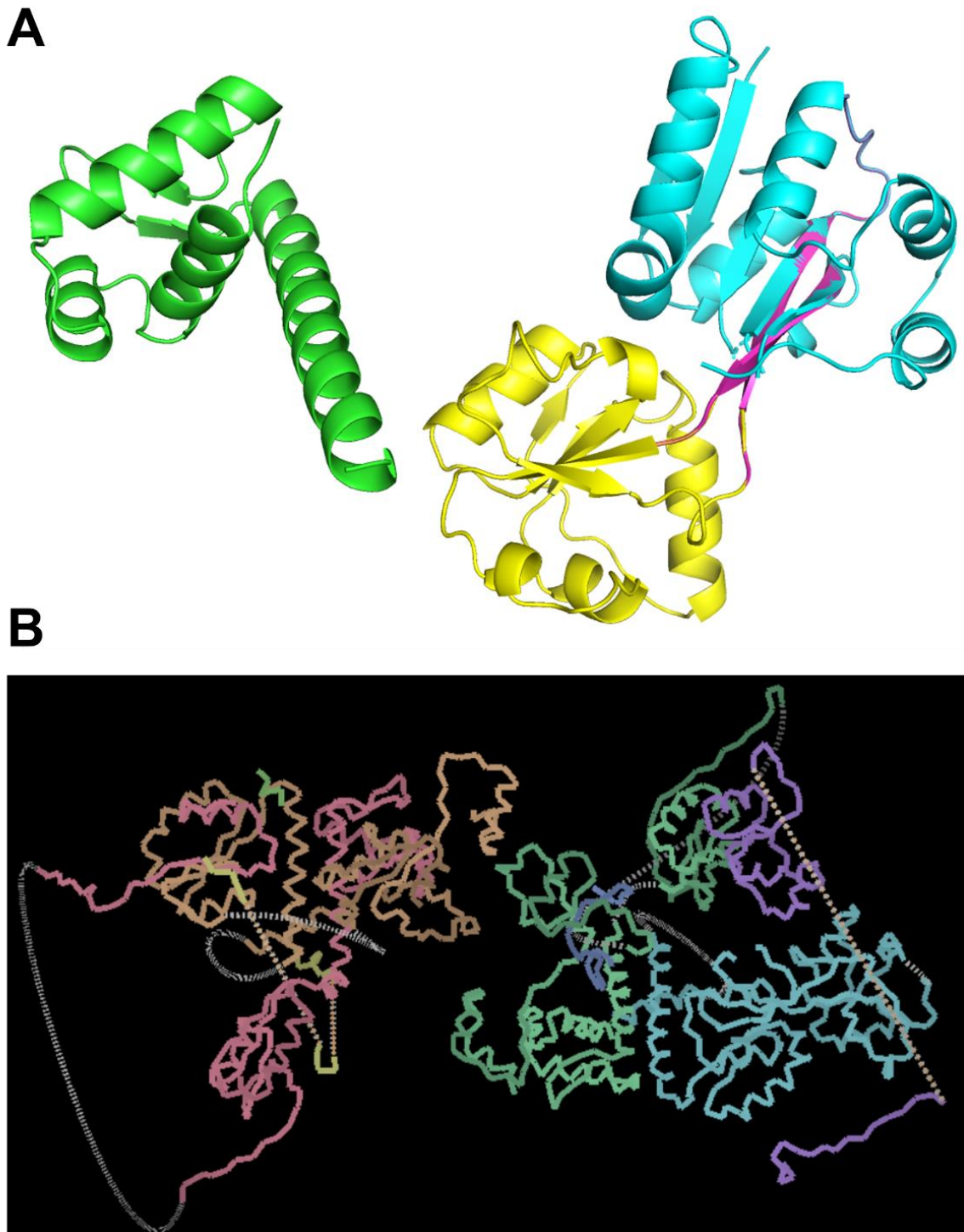
*Figure 3.20 Initial input models for molecular replacement and autobuild output. **A** Input models used for molecular replacement modified from 5Z4Y and 5Z50. The 5Z4Y N-terminal domain model (green) was not modified, while the 5Z50 C-terminal regulatory domain was modified. The regulatory domain (5Z50) was split into two smaller domains (shown in yellow and blue) with the crossover regions deleted (shown in pink). **B** Output from phenix.autobuild visualised in Coot showing multiple chains in a random orientation where the two portions of the regulatory domain are not linked. Figure generated using PyMOL version 2.3.4 and Coot.*

The CysB model was built using the *autobuild.phenix* programme (Terwilliger et al., 2008) from the PHENIX suite (Adams et al., 2010). The programme was supplied with the *phenix.phaser* (McCoy et al., 2007) output model and map, reflection file (with FreeR flag dataset) and the CysB amino acid sequence. Default settings were used (Appendix AC.3), with the exception of not placing waters in refinement. Rebuild was set to auto to allow the programme to build outside of the model where there may be unmodelled density, and to edit the structure file to match the CysB sequence file provided. The structure went through six iterative building rounds and three cycles of refinement. The resulting structure had an acceptable R_{work}/R_{free} score for the starting model of 0.2507/0.3069. The resulting structure had multiple fragments (18 in total) and had not built the two portions of the regulatory domain relative to one another such that they would form one full C-terminal regulatory domain (Figure 3.20).

Through investigation of the model generated by this initial round of *autobuild.phenix,* we were able to identify density that supported symmetry-related chains and were able to manually build two chains (from the initial 18 chain fragments). We then supplied these fully built chains as individual monomers for search models (Figure 3.21) and *phenix.phaser* found one solution with good TFZ and LLG scores of 76.2 and 6788.248 respectively. The *autobuild.phenix* programme was run using the same settings as described above and generated a good R_{work}/R_{free} score for the starting model of 0.2280/0.2821. In the asymmetric unit, four monomers were built and arranged in a homotetramer. The majority of each monomer was built by automated building (95%), however, there were sections of the protein's amino acid sequence missing (Figure 3.21). The remaining unbuilt peptide sequences were manually built in COOT. During manual building it was found that due to ambiguous density at 2.73 Å resolution, the autobuilding process had placed residues incorrectly within the density. The chains within the regulatory domains had to be rebuilt to allow the protein backbone to sit correctly within the density as there was continuous density with no gaps in density indicating no gaps in the Cα backbone, and as such, the regulatory domain chains had to be rebuilt. This was an iterative process of manual building in Coot and refinement using *phenix.refine* to build the remaining residues and place the protein chains correctly within the density.
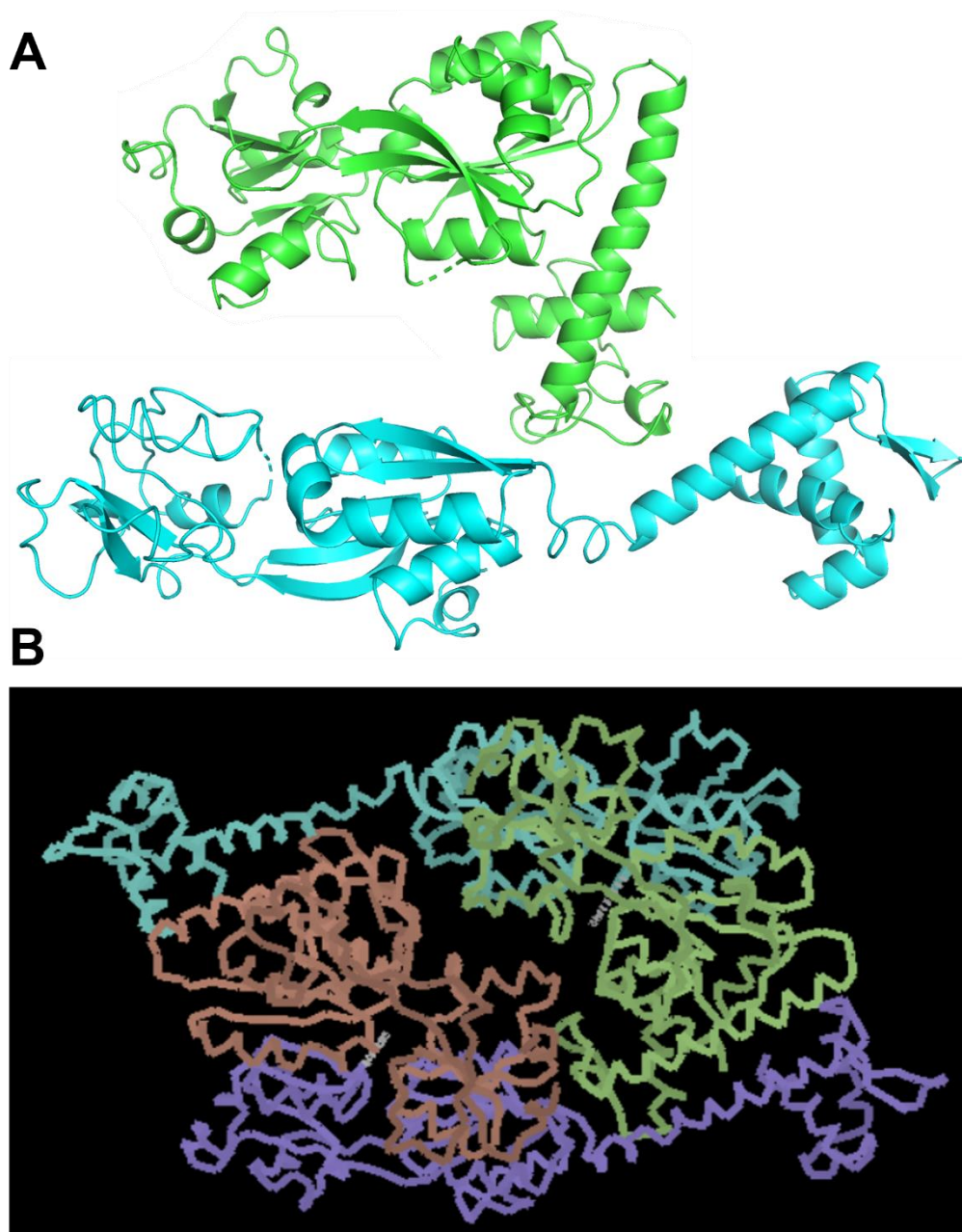
*Figure 3.21 Final input models for molecular replacement and autobuild output. **A** Input models (shown in green and blue) used for the second round of molecular replacement which were built from the first round of autobuild (Figure 3.20). **B** Output from the second round of autobuild visualised in Coot showing four chains arranged in a tetramer. Gaps are present in the regulatory domains of two of the monomers (blue and purple) shown as dotted lines. Figure generated using PyMOL version 2.3.4 and Coot.*

3.3.4 Analysis of the CysB structure

The final structural model for CysB was solved at 2.73 Å with a final R_{work}/R_{free} of 0.1973/0.2474 (Table 3.2). Structure and maps files for the CysB model can be found in Appendix AC.4. Four monomers were built within the asymmetric unit, and all assemble to form a homotetramer, as expected by our gel filtration results (Section 3.1). In all four

monomers within the asymmetric unit there was no density for the final two residues at the C-terminus including the hexahistidine-tag. One monomer had built residues Met1-Leu306, two of the monomers had built residues Met1-Ile310, and the other monomer had residues Met1-Phe314 with a gap between residues Tyr307-Ser311 (Appendix AC.5). The overall structure was well built with an average monomer length of 309.5 residues (309.5/316, 97.9 % built). The preliminary PDB validation report can be found in Appendix AC.6. The CysB monomer consists of two domains. An N-terminal DNA-binding domain spanning residues Met1-Gln90 including a linker helix which connects it to the C-terminal regulatory domain, and a C-terminal regulatory domain spanning residues Asp91-Ile316 which contains the inducer-binding cavity (Figure 3.22).

Table 3.2 Final CysB model quality statistics. Statistics generated by phenix.tableone.

Data statistic	CysB
R-work	0.1973
R-free	0.2474
No. of protein residues	1237
Total number of non-hydrogen atoms	9441
Macromolecules	9419
Solvent	22
RMS	
Bonds (Å)	0.011
Angles (°)	1.48
Average B-factor (Å)	69.72
Macromolecules	69.74
Solvent	62.26
Ramachandran analysis	
No. of residues in favoured regions (%)	95.27
No. of residues in allowed regions (%)	3.67
No. of residues in outlier regions (%)	1.06
Rotamer outliers (%)	0.11
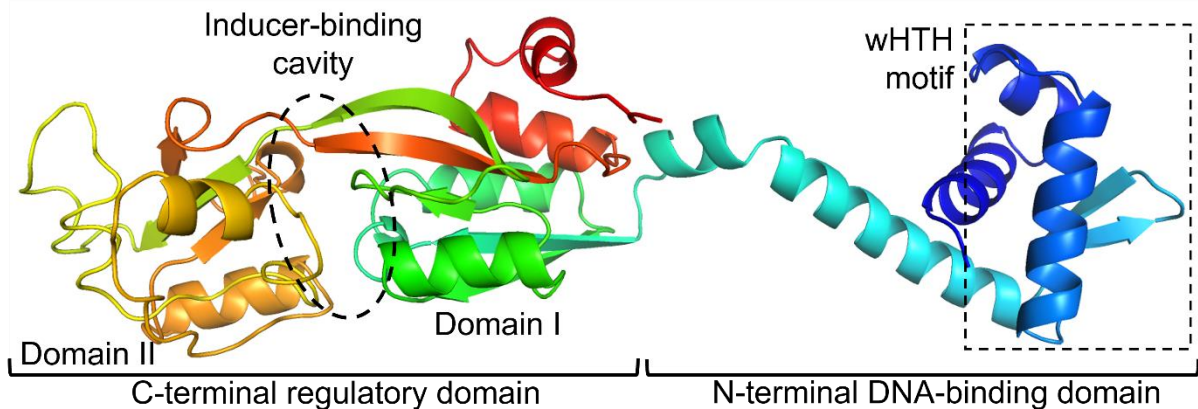
Figure 3.22 Structure of the CysB monomer from N. gonorrhoeae. The extended monomer is coloured from the N-terminus (blue) to the C-terminus (red). The N-terminal DNA-binding domain and C-terminal regulatory domains are labelled. The inducer-binding cavity and the winged helix-turn-helix motif are highlighted. Within the regulatory domain, domains I and II are indicated on either side of the inducer-binding cavity. Figure generated in PyMOL version 2.3.4.

3.3.4.1 *Analysis of the CysB tetramer*

This is the first solved full-length CysB structure and reveals unique insight into the formation of the homotetramer. The overall dimensions of the tetramer are approximately 130 Å x 94 Å x 66 Å with a central cavity (Figure 3.23). Analysis of the tetramer shows it is a dimer of dimers related by a two-fold axis. This axis within the X-ray crystallography structure may be an axis of pseudosymmetry which could be due to crystal packing restraints and may explain the inability to solve the structure in the higher symmetry space group $P6_1 2 2$. The surface model of the tetramer (Figure 3.24A) shows a circular arrangement with the regulatory domains positioned on the top of the protein and the DNA-binding domains underneath. The arrangement of the tetramer positions both DNA-binding domains facing one another on the same plane (Figure 3.24B) which supports the reported interaction of one CysB tetramer with convergently arranged DNA half sites (Hryniewicz & Kredich, 1994, 1995).
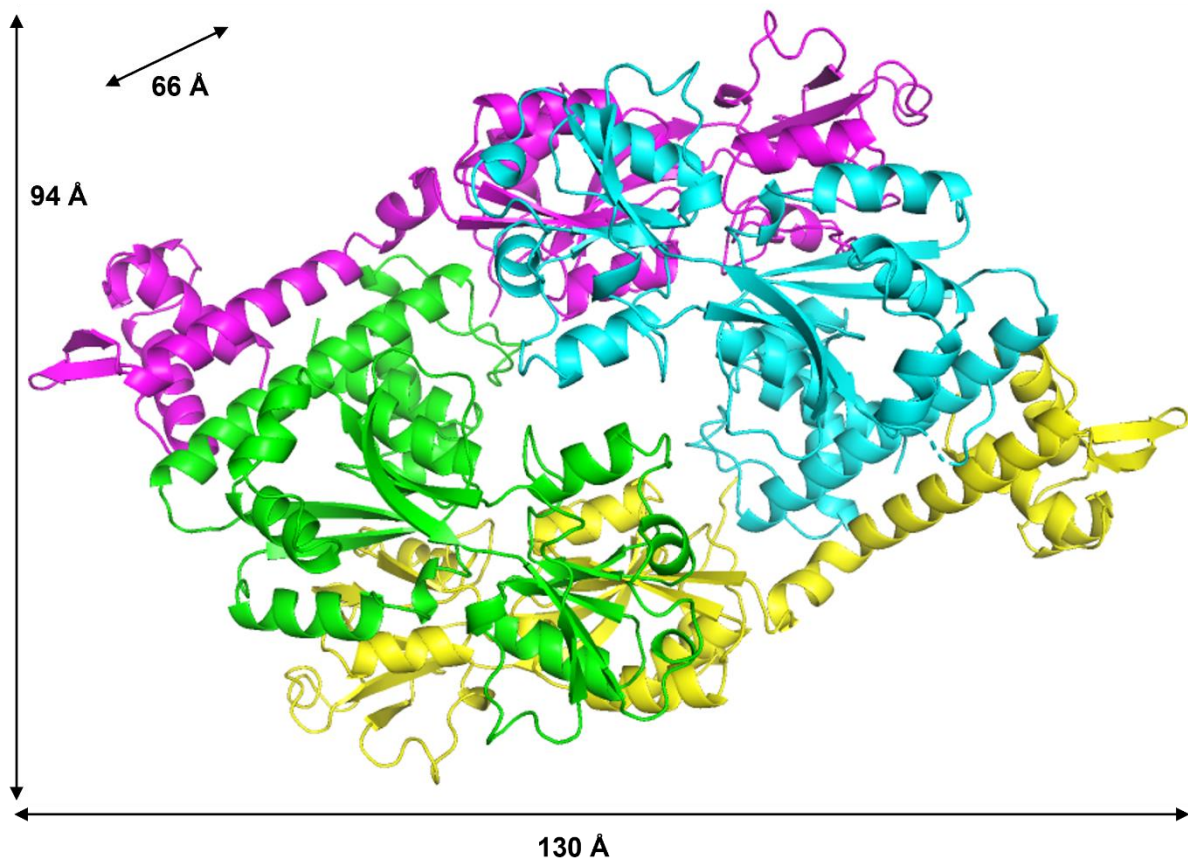
Figure 3.23 X-ray Crystallography structure of the CysB tetramer. The CysB tetramer with extended monomers in yellow and pink and compact monomers in cyan and green. This colour scheme is used throughout unless otherwise stated. Dimensions of the tetramer are displayed showing the tetramer is 130 Å x 94 Å x 66 Å. Figure generated using PyMOL version 2.3.4.
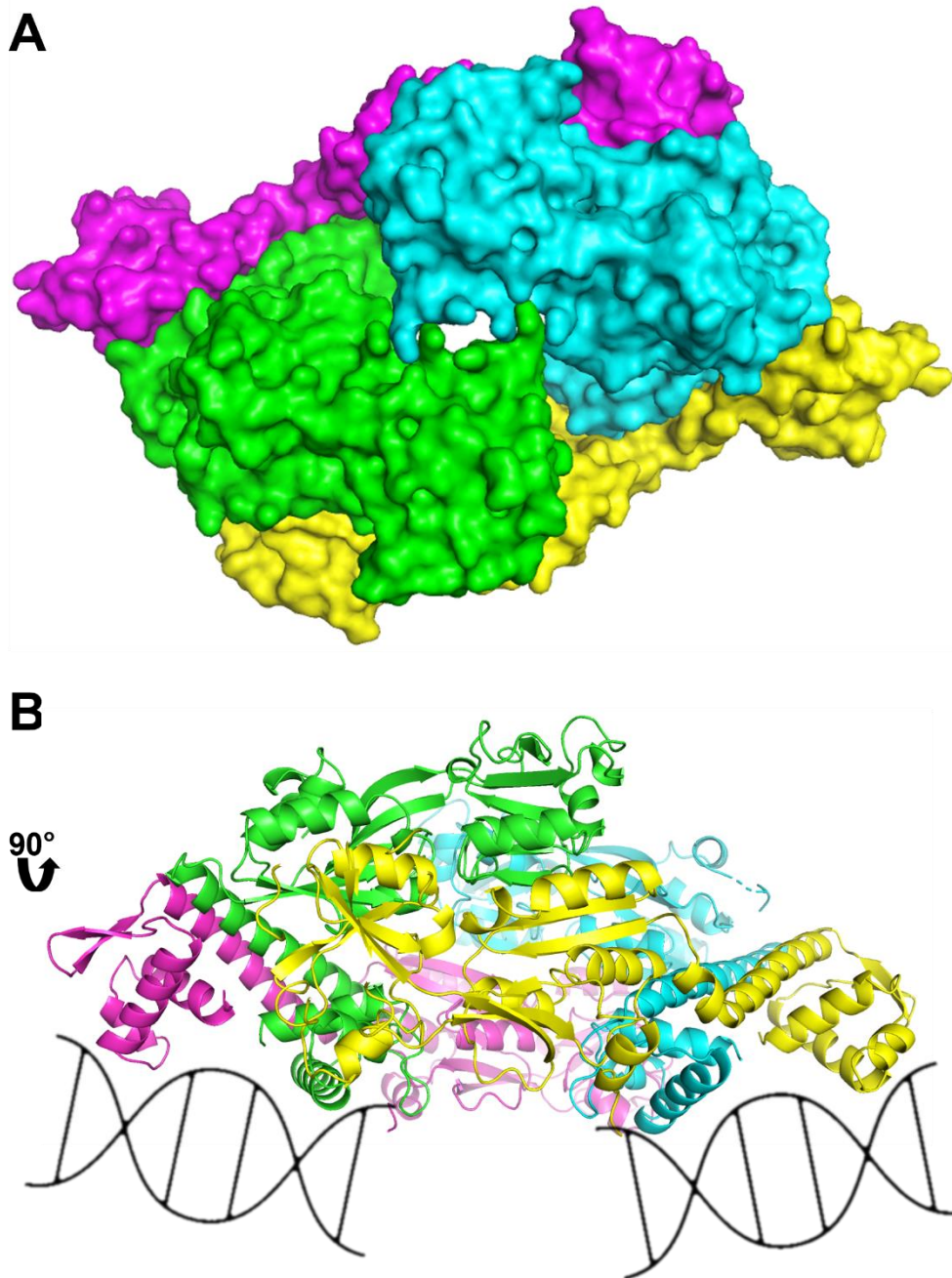
*Figure 3.24 The surface model of CysB and arrangement of the DNA-binding domains. **A** Surface model of CysB showing a circular arrangement of the tetramer. The regulatory domains (cyan and green) are positioned 'on top' of the protein. **B** Arrangement of the DNA-binding domains on the same plane when the tetramer is rotated 90°. Cartoon DNA placeholders demonstrate how each of the DNA-binding helices (described in 3.3.4.2) are positioned to bind into the major grooves of DNA. Figure generated using PyMOL version 2.3.4.*

Interestingly, two forms of the monomer were built, an extended and compact form. The extended or compact conformation is determined by a hinge region at the C-terminal end of the linker helix (residues Gln61-Gly89) which positions the N-terminal domain either extended away from the C-terminal regulatory domain or tucked up adjacent to it (Figure 3.25A,B). In

the extended form, the linker helix is one residue longer with a kink formed between residues Lys81-Ile83 which is where the remaining residues of the helix tend away from those that are straight (Figure 3.25C). Comparison of these two forms reveals no major differences between the extended and compact monomer form other than this kink in the linker helix (Figure 3.25C). There is strong alignment between the Cα of the two compact and of the two extended monomers giving r.m.s.d. values of 0.086 for the two compact monomers and 0.097 for the two extended monomers. Amongst the N-terminal DNA-binding domains there is strong fit between the two compact and between the two extended monomers with r.m.s.d. values of 0.149 and 0.096 for the compact monomers and extended monomers respectively. Amongst the C-terminal regulatory domains, there is also strong fit between the two compact and between the two extended monomers with r.m.s.d. values of 0.080 and 0.096 for the compact monomers and extended monomers respectively. There is more variation within the Cα between one extended and one compact monomer where alignment of the DNA-binding domain, residues Met1-Ile80 excluding the hinge of the linker helix, generates an r.m.s.d. value of 0.346, while alignment of the regulatory domain, residues 90-316, generates an r.m.s.d. value of 0.228.
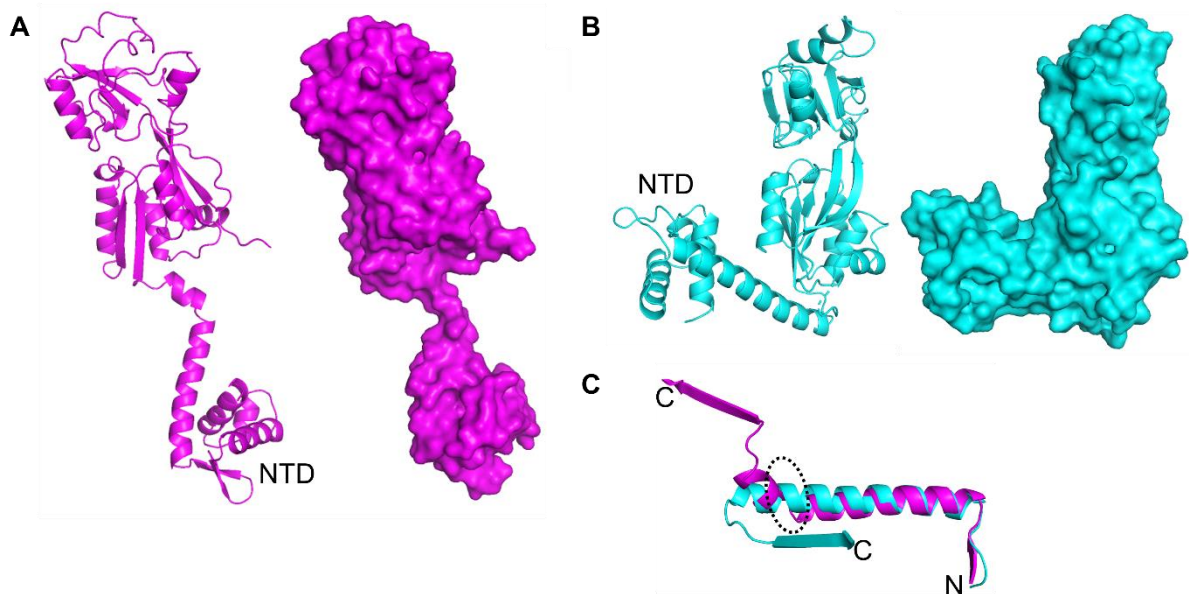
*Figure 3.25 Extended and compact monomers of CysB. Cartoon and surface representation of the extended monomer in pink (**A**) and compact monomer in cyan (**B**). The N-terminal domain is labelled NTD. **C** The two forms of the linker helix overlaid with the hinge highlighted and the N- and C-termini labelled N and C respectively. The kink in the linker helix positions the C-terminal regulatory domain either extend away or compact against the helix. This is shown by the first β sheet of the regulatory domain at the C-terminal end of the helix where the extended form is shown in pink and the compact form in cyan. Figure generated using PyMOL version 2.3.4.*

Within the tetramer there are two dimerisation interfaces, one through an antiparallel arrangement of the linker helix from an extended and compact monomer, and the other through association between alternate C-terminal subdomains (domain I and II) of an extended and compact monomer (Figure 3.26). Association between the two antiparallel linker helices (Figure 3.26A) has a large interface of 1 316 Å^2 (PDBePISA, (Krissinel & Henrick, 2007)) consisting of predominantly hydrophobic residues. Association between alternate C-terminal subdomains (Figure 3.26B) has a large interface of 1 156 Å^2 also consisting of predominantly hydrophobic residues. The calculated free Gibbs energy for dissociation of the tetramer is 19.9 kcal.mol^{-1} while both interfaces have a complex formation significance score of 1.000 indicating the tetramer is thermodynamically stable and both interfaces are necessary for tetramer formation. The total surface area of the tetramer is 53 040 Å^2 where the buried surface area is 10 690 Å^2.
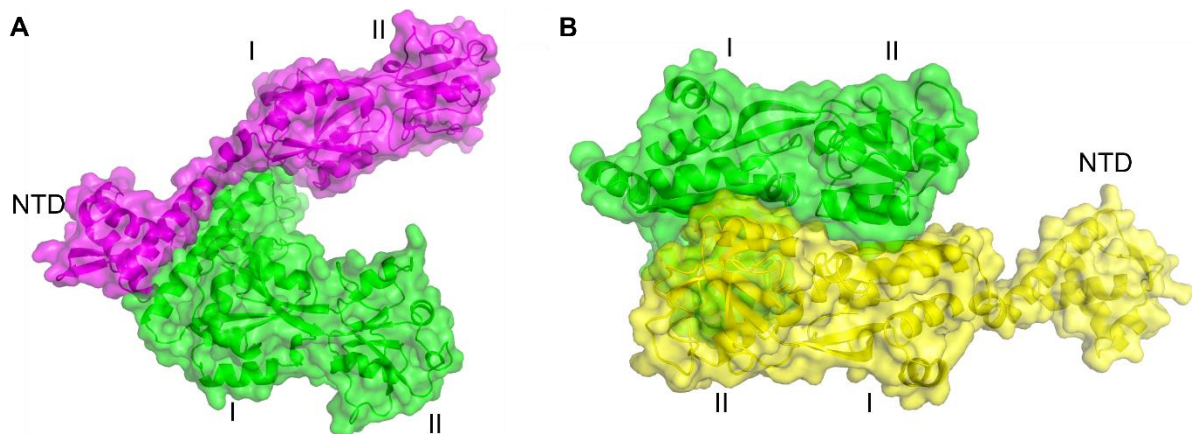
Figure 3.26 Dimerisation interfaces within the CysB tetramer. The N-terminal domain (NTD), domain I (I) and domain II (II) are labelled for orientation. Extended monomers are shown in pink and yellow while the compact monomer that interacts with both extended monomers is shown in green (this compact monomer represents the same interactions for both compact monomers). A Dimerisation of linker helices to position the DNA-binding domains relative to one another. The DNA-binding domain of the compact monomer is occluded from view by the compact monomer but sits near the N-terminal domain of the extended monomer. B Dimerisation interface between two regulatory domains where the subdomains domain I and II are positioned opposite to one another. Figure generated using PyMOL version 2.3.4.

3.3.4.2 Analysis of the N-terminal DNA-binding domain

The N-terminal DNA-binding domain of CysB from *N. gonorrhoeae* spans residues Met1-Gln91. This domain consists of four α helices (α$_{1-4}$) and two β sheets (β$_{1,2}$) (Figure 3.27A). The secondary structure elements from N-terminal to C-terminal are α$_1$ (Lys2-Arg14), α$_2$ (Val19-Leu26), α$_3$ (Gln30-Ile44), β$_1$ (Phe49-Ser52), β$_2$ (Arg55-Val59) and α$_4$ (Gln61-Gly89). Helices

α_2 and α_3 form a helix-turn-helix motif while sheets β_1 and β_2 form the 'wing' which come together to form the winged helix-turn-helix (wHTH) DNA binding motif (Figure 3.27A).
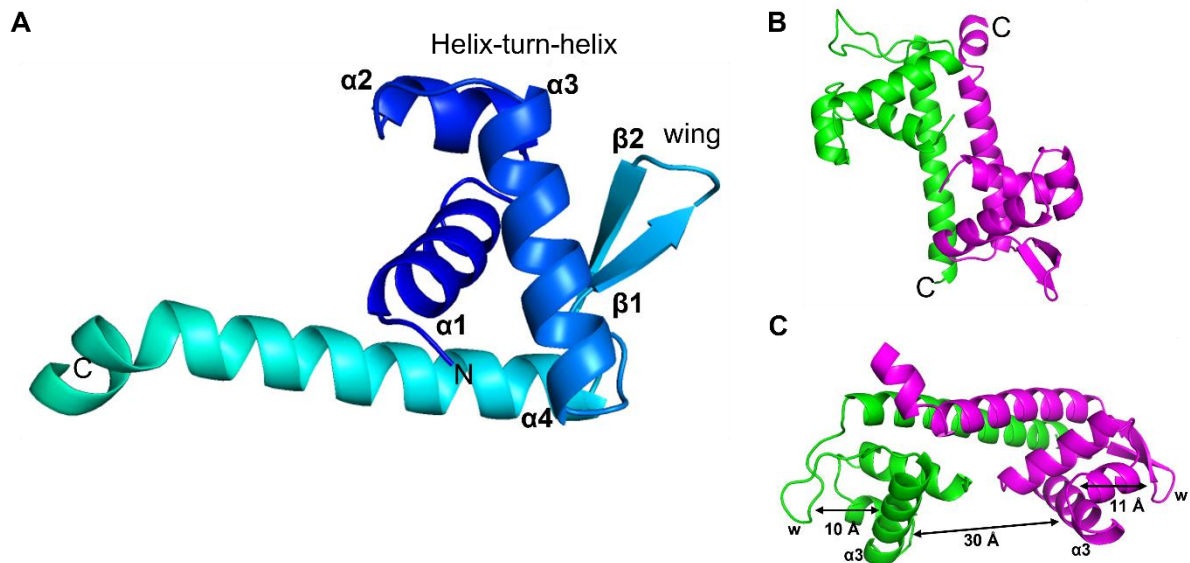
*Figure 3.27 N-terminal DNA-binding domain of CysB. **A** Secondary structure elements within the monomer which are labelled from the N- to the C-terminus in ascending order. The monomer is coloured from the N-terminus (blue) to the C-terminus (red) with N and C labels denoting the N- and C-termini. **B** Dimerisation of the N-terminal domains along the linker helix. **C** A rotated view of the dimerised DNA-binding domain to position both α_3 helices to bind into the major grooves of the DNA. Measurements are shown for the distances between the wing and α_3 helix (10/11 Å) and between the two α_3 helices (30 Å). Figure generated using PyMOL version 2.3.4.*

The structure of CysB from *N. gonorrhoeae* shows the N-terminal domain dimerises via the linker helix (α_4) (Figure 3.27B). Here the linker helices from one compact and one extended monomer dimerise antiparallel to one another to position the DNA domains opposite one another. This supports the observation of convergently arranged DNA half sites within the promoter regions of *S. typhimurium* and *E. coli* where CysB binds (Hryniewicz & Kredich, 1995; Lochowska et al., 2004). Within a wHTH motif, one helix of the HTH (α_3) binds into the major groove of DNA (Brennan & Matthews, 1989) while the wing is positioned to interact with the phosphate backbone in the corresponding minor groove (Zaim & Kierzek, 2003). In the dimerised N-terminal domain of the CysB structure, the α_3 helices are positioned 30 Å apart (Figure 3.27C) which is consistent with the arrangement of other LTTR structures (Monferrer et al., 2010) and consistent with the distance between major grooves within DNA. Each wing is positioned 10-11 Å from the α_3 helix which supports interaction of these with the phosphate backbone of the minor groove of DNA. It is interesting to note that the wing of the compact monomers does not strictly form a β sheet but rather a loop. This may be due to crystal packing

restraints or flexibility within the wing where perhaps bound DNA is required to visualise the exact conformation of this wing.

A peptide sequence alignment of CysB from *N. gonorrhoeae* to other organisms (Figure 3.28) shows high conservation within the DNA-binding domain. The sequence within the helix α_3 and the turn between the β sheets that make up the wing are very highly conserved. This may suggest that CysB binding sites within the cysteine regulon promoters may have some conservation between organisms.

The peptide sequence alignment of CysB proteins shows strong sequence and structural conservation (Figure 3.28). CysB from *N. gonorrhoeae* and *N. meningitidis* have almost identical sequences with only two amino acid substitutions seen between these organisms. Both of these substitutions occur within the regulatory domain where Ala213 within α_{10} of *N. gonorrhoeae* is replaced by Ser213 in *N. meningitidis*, and Arg219 of *N. gonorrhoeae* is replaced by Gln219 in *N. meningitidis* within the loop region between α_{10} and β_9. These amino acid positions are not highly conserved among CysB proteins (Figure 3.28). The CysB peptide sequences of *N. gonorrhoeae* and *N. meningitidis* align well to the peptide sequences from *P. aeruginosa, E. coli, S. typhimurium* and *K. aerogenes* with the exception of two additional amino acids in both *N. gonorrhoeae* and *N. meningitidis*. The first insertion is one glutamic acid at position 218 within the same loop region as the substitution at position 219, while the second insertion is a cysteine at position Cys181. The overall peptide alignment of CysB sequences has high conservation throughout the protein with some regions having multiple consecutive conserved residues (Figure 3.28). Within the N-terminal domain there is a very strong consensus sequence with only a few positions prior to the linker helix showing less or non-conserved residues. Within the C-terminal regulatory domain, however, there is less conservation with smaller regions of consecutive conserved residues. At the C-terminus, *N. gonorrhoeae* and *N. meningitidis* have a truncated C-terminal arm void of the final ten residues present in the other sequences.

N.gonorrhoeae

α1   α2   α3   β1   β2

```
N.gonorrhoeae   1 MKLQQLKYALEVYRHNLNVSEAAEALFTSQPGISKQIKLLEEEIGIQIFIRSGKRVVSVS
N.meningitidis  1 MKLQQLKYALEVYRHNLNVSEAAEALFTSQPGISKQIKLLEEEIGIQIFIRSGKRVVSVS
P.aeruginosa    1 MKLQQLRYIWEVAHHDLNVSATAQSLYTSQPGISKQIRLLEDELGVEVFARSGKHLTRVT
E.coli          1 MKLQQLRYIVEVVNHNLNVSSTAEGLYTSQPGISKQVRMLEDELGIQIFSRSGKHLTQVT
S.typhimurium   1 MKLQQLRYIVEVVNHNLNVSSTAEGLYTSQPGISKQVRMLEDELGIQIFARSGKHLTQVT
K.aerogenes     1 MKLQQLRYIVEVVNHNLNVSSTAEGLYTSQPGISKQVRMLEDELGIQIFARSGKHLTQVT
consensus>70      MKLQQL.Y..EV..H#LNVS..A#.L%TSQPGISKQ!.$LE#E.G!#!F.RSGK....V.
```

N.gonorrhoeae

α4   β3   η1   α5

```
N.gonorrhoeae  61 QPGKVVLDIAERILRDVQNIKNIGSEFTGQDSGSLTVATTHTQARYALPLIVADFVKRYP
N.meningitidis 61 QPGKVVLDIAERILRDVQNIKNIGSEFTGQDSGSLTVATTHTQARYALPLIVADFVKRYP
P.aeruginosa   61 PAGERIIHTAGEILRKVESIKQIAQEFSNEKKGTLSIATTHTQARYALPNVISGFIKQYP
E.coli         61 PAGQEIIRIAREVLSKVDAIKSVAGEHTWPDKGSLYIATTHTQARYALPNVIKGFIERYP
S.typhimurium  61 PAGQEIIRIAREVLSKVDAIKSVAGEHTWPDKGSLYIATTHTQARYALPGVIKGFIERYP
K.aerogenes    61 PAGQEIIRIAREVLSKVEAIKSVAGEHTWPDKGSLYVATTHTQARYALPGVIKGFIERYP
consensus>70      ..G..!..iA..!L..V#.IK.!..E.t..d.GsL.!ATTHTQARYALP.!!..F!.rYP
```

N.gonorrhoeae

β4   α6   β5   β6   β7   η2

```
N.gonorrhoeae  121 KVNLTIKQGSPAAIAQMVTSGESDLAIVTERIDDHPELGRLSCYDWTHAVIVPNDHPLLE
N.meningitidis 121 KVNLTIKQGSPAAIAQMVTSGESDLAIVTERIDDHPELGRLSCYDWTHAVIVPNDHPLLE
P.aeruginosa   121 DVSLHMHQGTPMQIAEMAADGTVDFAIATEALELFGDLIMMPCYRWNRAIVPHGHPLTK
E.coli         121 RVSLHMHQGSPTQIADAVSKGNADFAIATEALHLYEDLVMLPCYHWNRAIVVTPDHPLAG
S.typhimurium  121 RVSLHMHQGSPTQIAEAVSKGNADFAIATEALHLYDDLVMLPCYHWNRSIVVTPDHPLAA
K.aerogenes    121 RVSLHMHQGSPTQIAEAVSKGNADFAIATEALHLYDDLVMLPCYHWNRSIVVTPDHPLAS
consensus>70       .V.L...QGsP..IA#.v..Gn.D.AI.TE......#L..$.CY.W....!!V..dHPL..
```

N.gonorrhoeae

α7   β8   α8   β9   α9

```
N.gonorrhoeae  181 CRNPLRIEDLARFPLITYEFAENAGSSIARAFAKARLERPDVALAAADTDVLKTYVRLGL
N.meningitidis 181 CRNPLRIEDLARFPLITYEFAENAGSSIARAFSKARLEQPDVALAAADTDVLKTYVRLGL
P.aeruginosa   181 L.PKLTLEALAEQPIVTYVFGFTGRSKLDEAFSQRGL.VPKVVFTAADADVIKTYVRLGL
E.coli         181 K.KAITIEELAQYPLVTYTFGFTGRSELDTAFNRAGL.TPRIVFTATDADVIKTYVRLGL
S.typhimurium  181 T.SSVTIEALAQYPLVTYTFGFTGRSELDTAFNRAGL.TPRIVFTATDADVIKTYVRLGL
K.aerogenes    181 K.GSVTIEELAQYPLVTYTFGFTGRSELDTAFNRAGL.TPRIVFTATDADVIKTYVRLGL
consensus>70       ......iE.LA.yPl!TY.F.F...S.....AF..a.L..P.!...A.D.DV.KTYVRLGL
```

N.gonorrhoeae

β10   η3   β11   β12   α10

```
N.gonorrhoeae  241 GVGLMAKMAYNPDTDGDLQLVDAAHLFEPSPTWIALRSDTYLRGYAYDFIQAFAPHLTRE
N.meningitidis 241 GVGLMAKMAYNPDTDGDLQLVDAAHLFEPSPTWIALRSDTYLRGYAYDFIQAFAPHLTRE
P.aeruginosa   239 GVGIVAHMAVDPKLDNDLVILDASHLFESSVTKIGFRRGTFLRGFMCDFIEKFAPHLTRE
E.coli         239 GVEVIASMAVDPVADPDLVRVDAHDIFSHSTTKIGFRRSTFLRSYMYDFIQRFAPHLTRD
S.typhimurium  239 GVGVIASMAVDPLADPDLVRIDAHDIFSHSTTKIGFRRSTFLRSYMYDFIQRFAPHLTRD
K.aerogenes    239 GVGVIASMAVDPVSDPDLVKLDANGIFSHSTTKIGFRRSTFLRSYMYDFIQRFAPHLTRD
consensus>70       GVg..A.MA.#P..D.DL...DA...F..S.T.I..R..T%LR.%.yDFI#.FAPHLTR#
```

N.gonorrhoeae

α11

```
N.gonorrhoeae  301 KVDRILYTPISEDFSI..........
N.meningitidis 301 KVDRILYTPISEDFSI..........
P.aeruginosa   299 LLAKAVQCHNKAELDELFDGIELPVY
E.coli         299 VVDAAVALRSNEEIEVMFKDIKLPEK
S.typhimurium  299 VVDTAVALRSNEEIEAMFQDIKLPEK
K.aerogenes    299 VVDTAVALRSNEDIEAMFKDIKLPEK
consensus>70       .vd.......e#..............
```

Figure 3.28 CysB peptide sequence alignment. The N. gonorrhoeae FA1090 CysB peptide sequence was aligned to the following CysB homologues; N. meningitidis, P. aeruginosa, E. coli, S. typhimurium and K. aerogenes. Highly conserved residues are highlighted in red, less conserved residues in yellow and non-conserved residues in white. Secondary structures are shown above the sequence where conserved, with the consensus sequence shown below. Sequence alignment created using Clustal Omega (Sievers et al., 2011). Figure created using ESPript 3.0 (Robert & Gouet, 2014).

When the solved CysB reported here is overlaid with the solved N-terminal domain from *P. aeruginosa* (PDB code; 5Z4Y) used for molecular replacement, there is good homology between these two structures with an r.m.s.d. value of 0.538. The main difference between these structures is between the forms of the linker helix (α_4 residues Gln61-Thr88/Gly89). The tetrameric CysB structure from *N. gonorrhoeae* shows both a straight and kinked linker helix. This kink in the linker helix does not appear in the N-terminal domain of CysB from *P. aeruginosa* (5Z4Y) which may be due to the lack of the C-terminal regulatory domain. This kink is observed in other full-length LTTRs but only in the extended forms of the LTTR monomers.

Alignment of the CysB N-terminal domain to the N-terminal domain of the LTTR family member CbnR from *Cupriavidus necator* solved with DNA bound (PDB code; 7D98 (Giannopoulou et al., 2021)) shows good structural homology with an r.m.s.d. score of 0.992 (Figure 3.29A). In the CbnR structure, both α_3 helices bind into the major DNA groove while the wing sits adjacent toward what would be the minor groove (Figure 3.29A). Once DNA binding is modelled in our structure, it becomes clear the importance of the highly conserved positively charged residues Lys35, Lys38, Arg51, Lys54 and Arg55 among CysB proteins (Figure 3.28). Within helix α_3, Lys35 and Lys38 appear to interact with the major groove of the DNA, although there was no density for the sidechains of these residues, they appear to be facing toward where the helix would interact with DNA (Figure 3.29B). Within the wing (β_1 and β_2) there are three highly conserved positively charged residues, Arg51 which interacts with Glu41 to anchor the wing relative to helix α_3 (discussed below) and Lys54 and Arg55 which have density in some chains (Figure 3.29C-F) and appear to be placed to interact with the phosphate backbone of the minor groove of the DNA. The residues Lys35 within the DNA-binding helix α_3 and Lys54 within the wing, have been shown to be critical for DNA binding where mutation to alanine abolishes binding activity (Song et al., 2019). Other residues where mutations have shown diminished DNA binding include Glu11, Ser20, Thr22, Glu41, Leu44 and Ile48, where Thr22 and Leu44 are not conserved in *N. gonorrhoeae* and are instead mutated to Ala22 and Ile44 (Lochowska et al., 2001). While the role of Glu41 may be in stabilising the wHTH motif, the role of these other residues is as yet unclear, including any potential effects of the mutation of the residues Thr22 and Leu44 in CysB of *N. gonorrhoeae*.
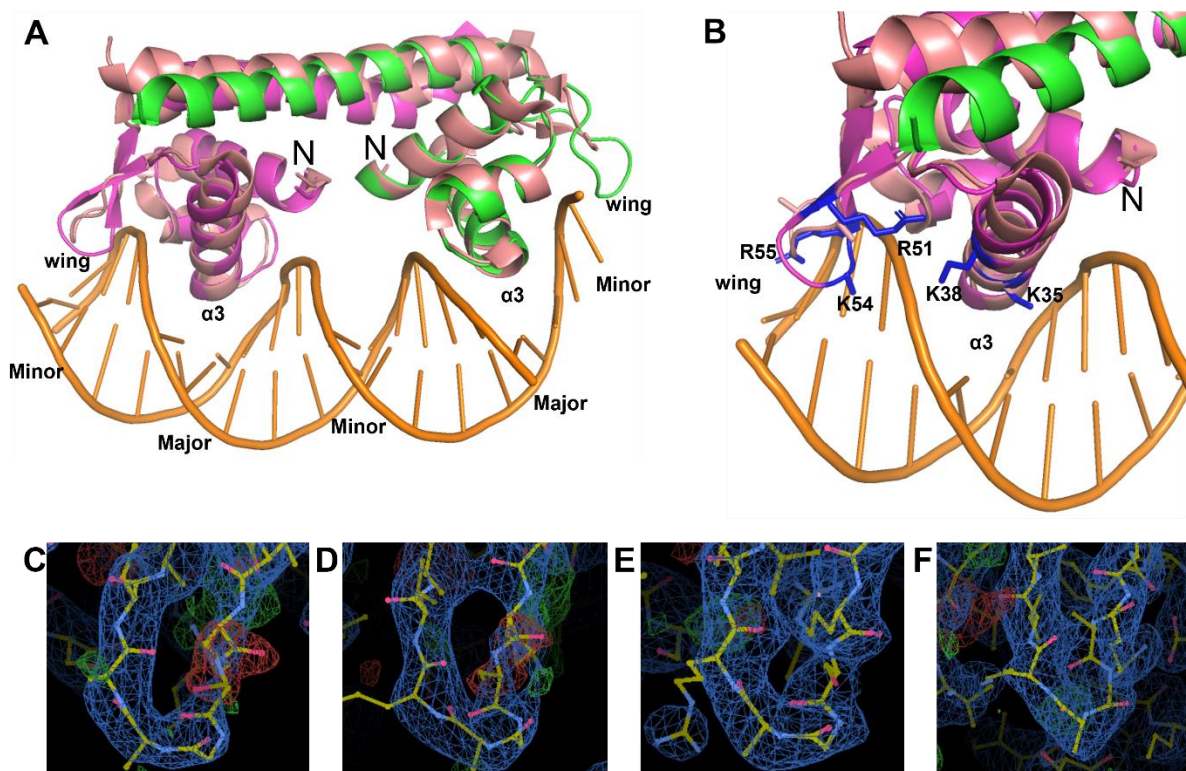
*Figure 3.29 DNA modelled onto the CysB DNA-binding domain and residues important for binding. The CysB DNA-binding domain showed good alignment to the DNA-binding domain of CbnR with DNA bound. This alignment shows the DNA-binding domains within CysB are positioned to bind into the major and minor grooves of DNA. **A** CysB DNA-binding domain aligned to the CbnR model that was solved with DNA bound (7D98). **B** Side chains of residues within the α_3 helix and wing. Critical binding residues are labelled. **C-F** 2Fo-Fc map contoured to 1 σ showing the density built for each of the four wings within the chains A-D respectively in the N. gonorrhoeae CysB structure. 2Fo-Fc map (blue mesh) is contoured to 1 σ and Fo-Fc map is contoured to 3 σ (green and red mesh). Green mesh = positive density, red mesh = negative density. Figure generated using PyMOL version 2.3.4 and Coot.*

Elucidation of the full-length tetrameric form of CysB allows comparison between the CysB structure and other LTTRs. Peptide sequence alignments of CysB with other LTTRs show no strong sequence similarity among LTTRs (Figure 3.30). Attempts to align the structures of LTTRs with CysB resulted in poor r.m.s.d. values or failed all together. With such poor conservation of both structure and sequence it is perhaps surprising that there are four highly conserved residues among all solved structures of members of the LTTR family (Figure 3.30). These residues include two leucines (CysB numbering; Leu6 and Leu40) and one arginine (Arg51) conserved across all LTTRs, and one glutamic acid (Glu41) which is conserved in all structures except DntR. All of these conserved residues are found within the DNA-binding winged helix-turn-helix domain which is present among LTTRs. The glutamic acid that is not conserved in DntR, may be due to the extended wing loop present in DntR (Smirnova et al., 2004). The relevance of these conserved residues only becomes apparent when mapped onto

the structure (Figure 3.31). The structure of CysB shows the arginine residue (Arg51) positioned within the wing and glutamic acid residue (Glu41) within the DNA-binding helix (α_3) interact which anchors the wing to the α_3 (Figure 3.31). This would ensure the wing is positioned relative to helix α_3 to promote binding of the helix α_3 into the major groove of DNA and the wing towards the corresponding minor DNA groove which may prevent steric clashes between helix α_3 and the wing within the protein when interacting with DNA. The role of the two leucine residues is less clear. The leucines could be involved in hydrophobic packing and/or van der waals forces to stabilise the position of the DNA-binding domains relative to the linker helices (Figure 3.31).
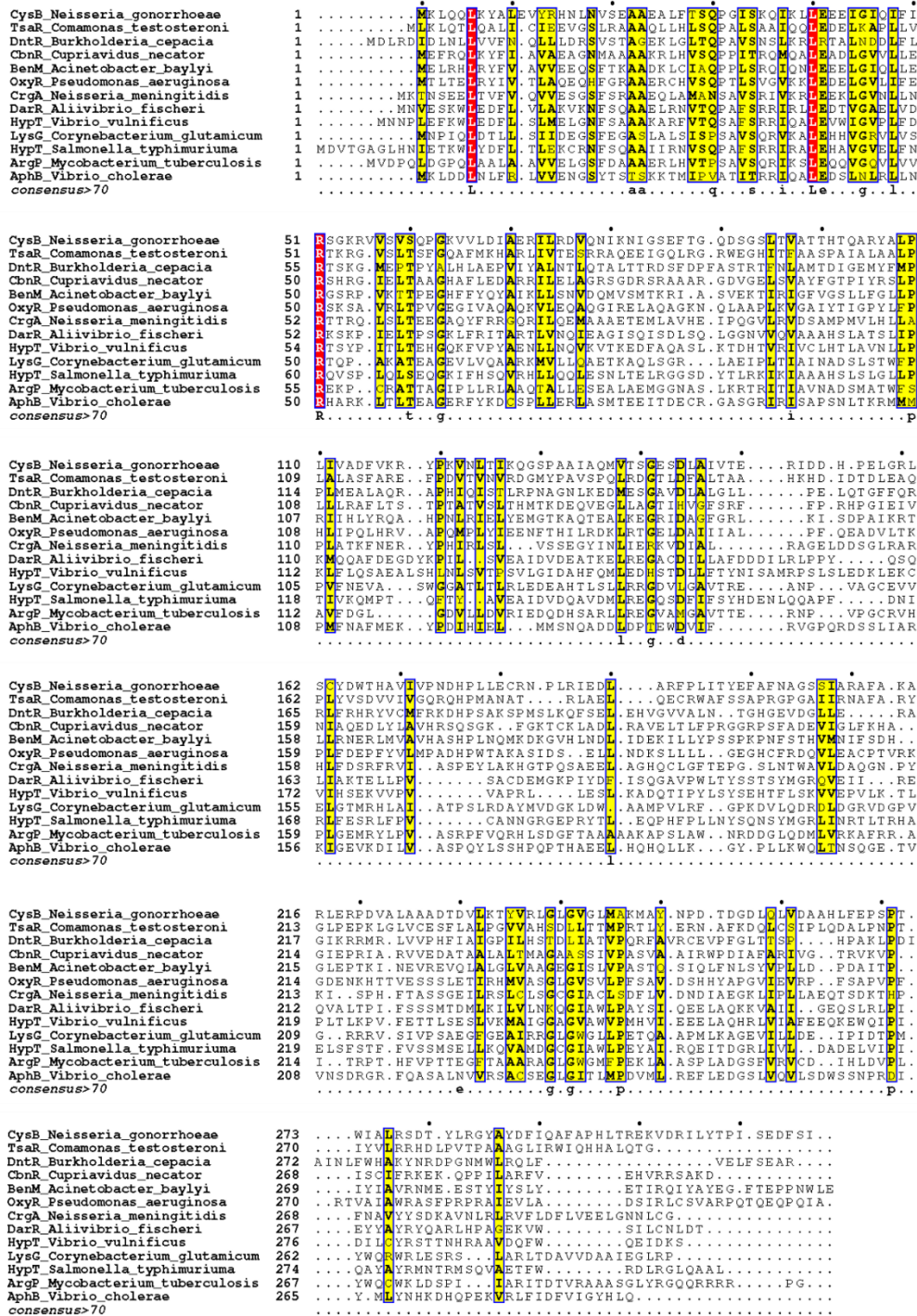
Figure 3.30 CysB peptide sequence alignment with other LTTRS. The N. gonorrhoeae FA1090 CysB peptide sequence was aligned to the following LTTRs whose structures have been solved; TsaR of Comamonas testosteroni (3FXU), DntR of Burkholderia cepacian (5AE5), CbnR of Cupriavidus necator (7D98), BenM of Acinetobacter baylyi (3K1N), OxyR of Pseudomonas aeruginosa (4Y0M), CrgA of Neisseria meningitidis (3HHG), DarR of Aliivibrio fischeri (7DWN), Hypt of Vibrio vulnificus (5Y9S), LysG of Corynebacterium glutamicum (6XTU), HypT of Salmonella typhimuriuma (5YDW), ArgP of Mycobacterium tuberculosis (3ISP) and AphB of Vibrio cholerae (3SZP). Highly conserved residues are highlighted in red, less conserved residues in yellow and non-conserved residues in white. The N-terminal DNA-binding domain is made up of the first ~90 residues. Sequence alignment created using Clustal Omega (Sievers et al., 2011). Figure created using ESPript 3.0 (Robert & Gouet, 2014).
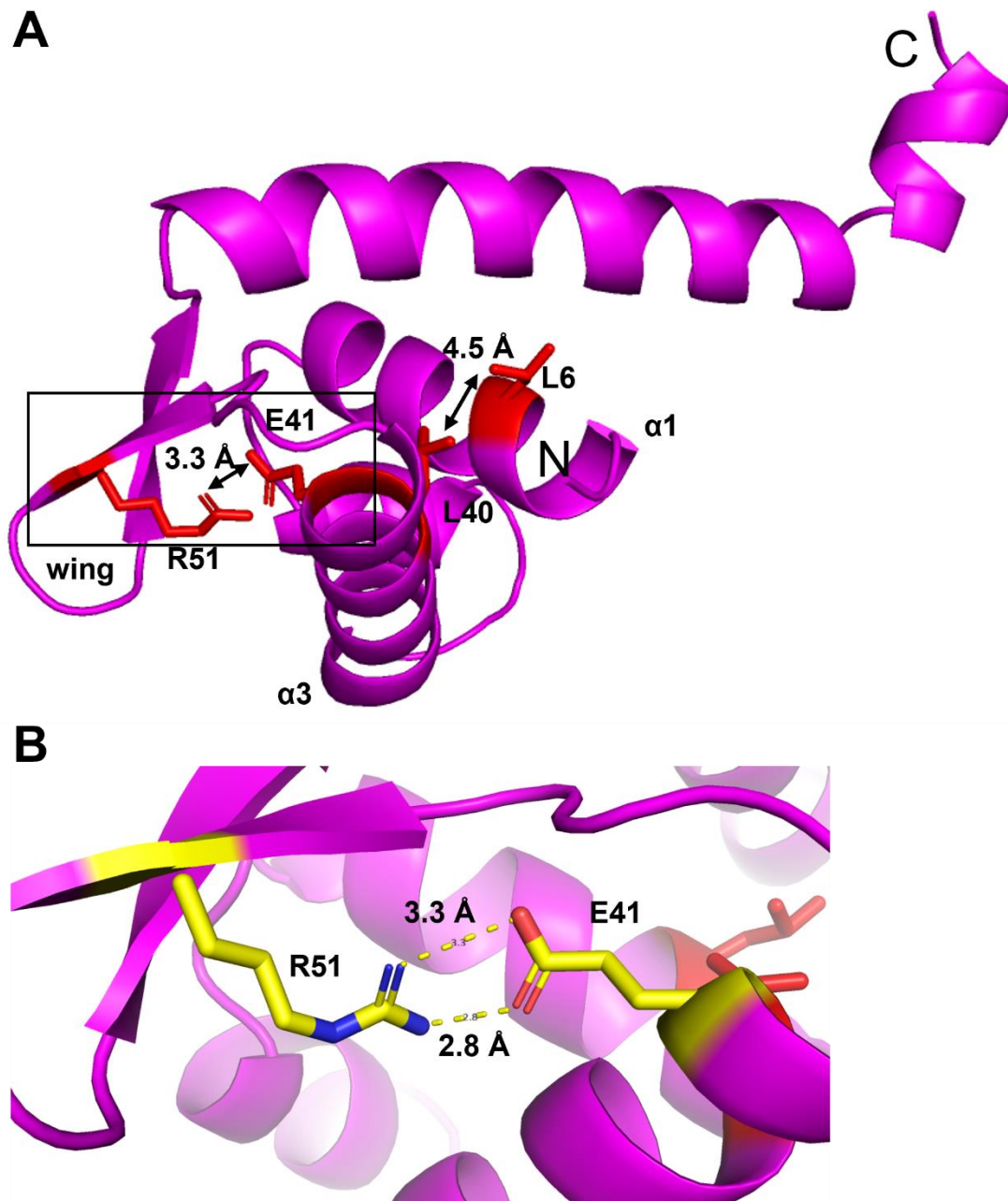
*Figure 3.31 The highly conserved residues among LTTRs mapped on the CysB structure. **A** These residues (shown in red) are located in the N-terminal DNA-binding domain and appear to stabilise the positioning of the helices within the wHTH motif. The N- and C-termini are labelled N and C respectively. Helices α1 and α3 are labelled, as is the wing. Distances are shown between the Leu6 and Leu40 of 4.5 Å and between Arg51 and Glu41 of 3.3 Å. The interaction between Arg51 and Glu41 within the box is expanded. **B** An expanded view of the Arg51 – Glu41 interaction. Here the sidechains are coloured by element and show the interactions between the sidechains with bond distances of 2.8 Å and 3.3 Å. Figure generated using PyMOL version 2.3.4.*

3.3.4.3   Analysis of the C-terminal regulatory domain

The C-terminal regulatory domain of CysB from *N. gonorrhoeae* spans residues Asp91-Ile316. It forms an ellipsoid shape with an axial ratio of 2:1 consisting of two similar α/β domains,

domain I and II (Figure 3.32). The α/β domains display Rossmann-like fold topology and are connected by two cross-over regions (Figure 3.32). The arrangement within the CysB regulatory domain of *N. gonorrhoeae* is similar to the arrangement of the previously solved regulatory domain of CysB from *K. aerogenes* (PDB code 1AL3 (Tyrrell et al., 1997)) described in Section 1.8.2 with the exception of a truncated C-terminal arm (r.m.s.d. value of 2.211). Briefly, the CysB regulatory domain consists of ten α helices and eight β sheets arranged into two subdomains, domain I and II (Figure 3.32). Domain I (Figure 3.32) is the larger of the two subdomains and consists of five α helices ($\alpha_{5\text{-}7}$ and $\alpha_{14\text{-}15}$) and five β sheets ($\beta_{3\text{-}6}$ and β_{12}) spanning residues Ser94-Trp166 and Ser270-Ile305. The secondary structure elements from the N-terminus to the C-terminus are β_3 (Ser94-Thr99), α_5 (His101-Arg105), α_6 (Leu108-Arg118), β_4 (Val122-Gln128), α_7 (Pro131-Ser140), β_5 (Leu145-Val148), β_6 (Leu158-Trp166), β_{12} (Ser270-Arg277), α_{14} (Gly284-Phe293) and α_{15} (Arg299-Ile305). The first crossover region occurs between β_6 and β_7, while the second crossover region occurs between α_{13} and β_{10}. These helices and sheets are arranged to form a core β sheet with the topology $\beta_4\beta_3\beta_5\beta_{12}\beta_6$ where β_{12} is antiparallel to the others which is surrounded by α helices α_5, α_6, α_7 and α_{14}, and α_{15} of the C-terminus (Figure 3.32A)

Domain II (Figure 3.32) also consists of five α helices ($\alpha_{8\text{-}13}$) and five β sheets ($\beta_{7\text{-}11}$) spanning residues His168-Leu266. The secondary structure elements from the N-terminus to the C-terminus are β_7 (His168-Pro173), α_8 (Pro177-Glu180), α_9 (Ile187-Arg192), β_8 (Leu195-Tyr198), α_{10} (Ile208-Ala215), β_9 (Val222-Ala226), α_{11} (Thr229-Leu238), β_{10} (Gly243-Ala246), α_{12} (Lys247-Ala249), β_{11} (Leu258-Asp262) and α_{13} (Ala264-Leu266). These helices and sheets also form a core β sheet with the topology $\beta_{11}\beta_7\beta_{10}\beta_8\beta_9$ where β_7 is antiparallel to the others and is surrounded by α helices $\alpha_{8\text{-}13}$. Between the two subdomains is a cavity approximately 16 Å by 10 Å which is lined by polar residues from the crossover regions and loop regions. This cavity forms the inducer-binding cavity. In the CysB structure, the regulatory domains dimerise between one compact and one extended monomer where the subdomains are positioned opposite to one another (Figure 3.32B) . In the CysB dimer this forms a channel where the two inducer-binding cavities are positioned so that the crossover regions are on opposite sides of the channel.

92
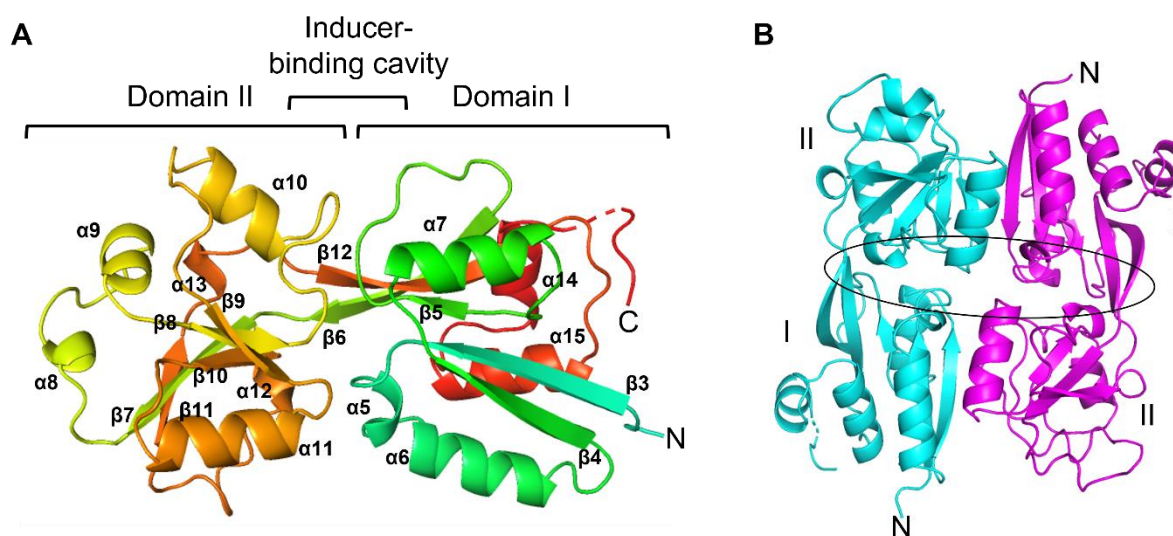
*Figure 3.32 C-terminal regulatory domain of the CysB monomer. **A** Secondary structure elements are labelled from the N- to the C-terminus in ascending order. The monomer is coloured from the N-terminus (blue) to the C-terminus (red) with the letters N and C denoting the N- and C-termini. **B** Dimeric arrangement of the regulatory domain in the CysB tetramer. The inducer-binding cavity (highlighted) from two monomers forms a channel in the dimeric arrangement. Figure generated using PyMOL version 2.3.4.*

Despite conservation of secondary structure elements amongst CysB proteins, when the compact regulatory domain of CysB from *N. gonorrhoeae* was overlaid with the regulatory domain of all solved CysB structures, r.m.s.d. values ranged from 1.996 to 2.272. Secondary structures appeared at similar positions within the polypeptide chain but there was poor alignment particularly amongst the loop regions (Figure 3.33A). Alignment to the regulatory domain of the extended CysB monomer gave worse r.m.s.d. scores, as did alignment to the dimeric form of the regulatory domains. There are two potential explanations for this structural mismatch. Either this misalignment is due to crystal packing restraints across various crystallisation conditions or may be a result of only the regulatory domain being crystallised. Both of these scenarios appear unlikely as each structure was crystallised in different conditions yet align well to one another, and the crystal structures from *S. typhimurium* were the result of attempts to crystallise the full-length protein but saw ordered density for just the regulatory domains. So perhaps, there are enough residue substitutions amongst the peptide sequence of CysB from *N. gonorrhoeae*, which along with the two inserted residues, may afford some small changes in the conformation of the CysB regulatory domain.

Inspection of the inducer-binding cavity within each monomer during model building revealed no density was present indicating that the CysB structure presented here is in the *apo*-form. Comparison of the inducer-binding site to that from the *apo*-structure from *S. typhimurium*

(4GXA) shows similar positions for conserved residues Thr100, Thr102, Thr149, Glu150 and Tyr219 (Figure 3.33B). However, there was a substitution in the *N. gonorrhoeae* sequence for the conserved residue Thr202 to glutamine, which due to the insertion of Cys181 is now position 203 (Figure 3.33B and C). Interestingly the loop containing this substitution is positioned away from the binding cavity (Figure 3.33B) where it is positioned toward the cavity in all other structures including the *apo*-regulatory domain solved from *S. typhimurium* (4GXA). Investigation of the *N*-acetylserine-bound regulatory domain from *S. typhimurium* (4M4G) shows that Thr202 does not appear to interact with *N*-acetylserine in the inducer-binding cavity which may explain its absence in *N. gonorrhoeae*.
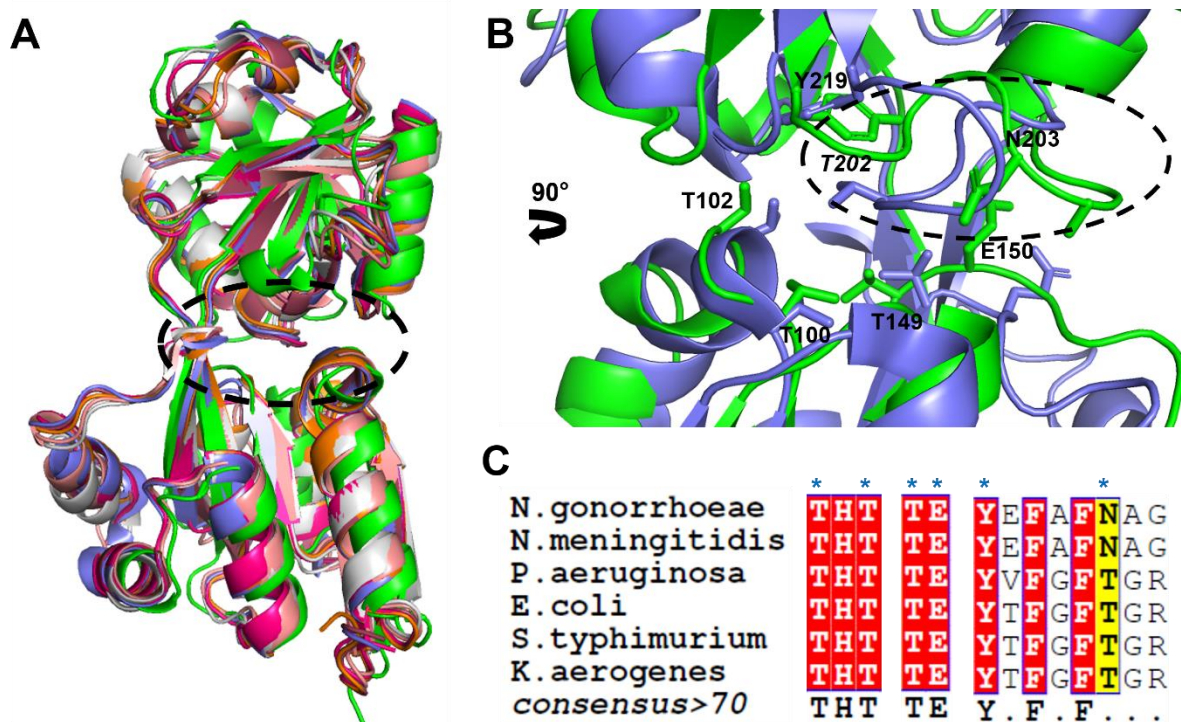
*Figure 3.33 Structural overlay of CysB regulatory domain structures and the inducer-binding cavity. **A** Structural overlay of CysB from N. gonorrhoeae (green) with the regulatory domain from K. aerogenes (1AL3 in peach) where sulfate is bound, P. aeruginosa (5Z50 in grey) where sulfate is bound and S. typhimurium in the apo-state (4GXA in purple), with sulfate bound (4GXO in orange), with N-acetylserine bound (4M4G in burgundy red) and with O-acetylserine bound in the secondary site (4LQ2 in hot pink). The inducer-binding cavity is highlighted. **B** The inducer-binding cavity of the apo-state CysB from N. gonorrhoeae (green) and S. typhimurium (purple). Residues important for inducer/anti-inducer binding are labelled. Thr202 is in italics as this residue is present in the S. typhimurium structure but is replaced by Gln203 in the N. gonorrhoeae structure. The loop that is positioned toward the inducer-binding cavity in S. typhimurium (purple) is positioned away from the inducer-binding cavity in N. gonorrhoeae (green) which is highlighted. **C** Extract from the CysB peptide sequence alignment showing conservation of most of the residues important for inducer/anti-inducer binding. Residues labelled with * are important for binding and are labelled in **B**. Figure generated using PyMOL version 2.3.4.*

A secondary binding site was reported by Mittal et al. (2017) which described the binding of the secondary inducer *O*-acetylserine to allow remodelling of the inducer-binding cavity (Mittal et al., 2017). This site is described as a concave, shallow cavity which is surface exposed and allows the binding of *O*-acetylserine exclusively (Mittal et al., 2017). Analysis of this secondary binding site (site-2) in the structure of CysB from *N. gonorrhoeae* (Figure 3.34A) shows the absence of the C-terminal arm that forms the top of the binding site (Figure 3.34B) including the residues involved in anchoring *O*-acetylserine (Mittal et al., 2017). Mittal describes a hypothesised mechanism by which *O*-acetylserine binding to site-2 may confer changes to the inducer-binding cavity, which begins through the anchoring of *O*-acetylserine by the final residues of the C-terminal arm. These residues are not present in CysB from *N. gonorrhoeae* or from *N. meningitidis*. Additionally, in the regulatory domain structures from *S. typhimurium* there are 19 residues built at the C-terminus where there was limited to no density to build the remaining eight terminal residues in this reported CysB structure from *N. gonorrhoeae*. Even where there was density at the C-terminus of some monomers to build residues Tyr307-Phe314, the final residue that aligned to previously solved structures was Leu306, with any further built residues not aligning to the structures from *S. typhimurium* or forming secondary structures (Figure 3.34B).

Despite lacking the C-terminal residues, the majority of residues involved in forming this binding cleft are present in the *N. gonorrhoeae* CysB structure where any substitutions are among resides that do not appear to interact directly with the cleft (Figure 3.34C). Residues involved in forming this cleft consist of residues between helices α_5 and α_6 Arg105-Leu108, within helix α_{14} Ala292-Ala294, the crossover region Cys163-His168, and if present, the final four residues of the C-terminal arm. The hydrophobic cleft involved in binding *O*-acetylserine is lined with residues Tyr106, Tyr164, Met248, and Phe293 (Figure 3.34C). The 2Fo-Fc map contoured to 1 σ shows strong density for these hydrophobic residues while the cleft itself is void of any density (structural files in Appendix AC.4). Mutational analysis in CysB from *E. coli* demonstrated that residues within the C-terminal arm may be important in DNA-binding activity as loss of these residues reduced the ability of CysB to repress transcription (Lochowska et al., 2001).
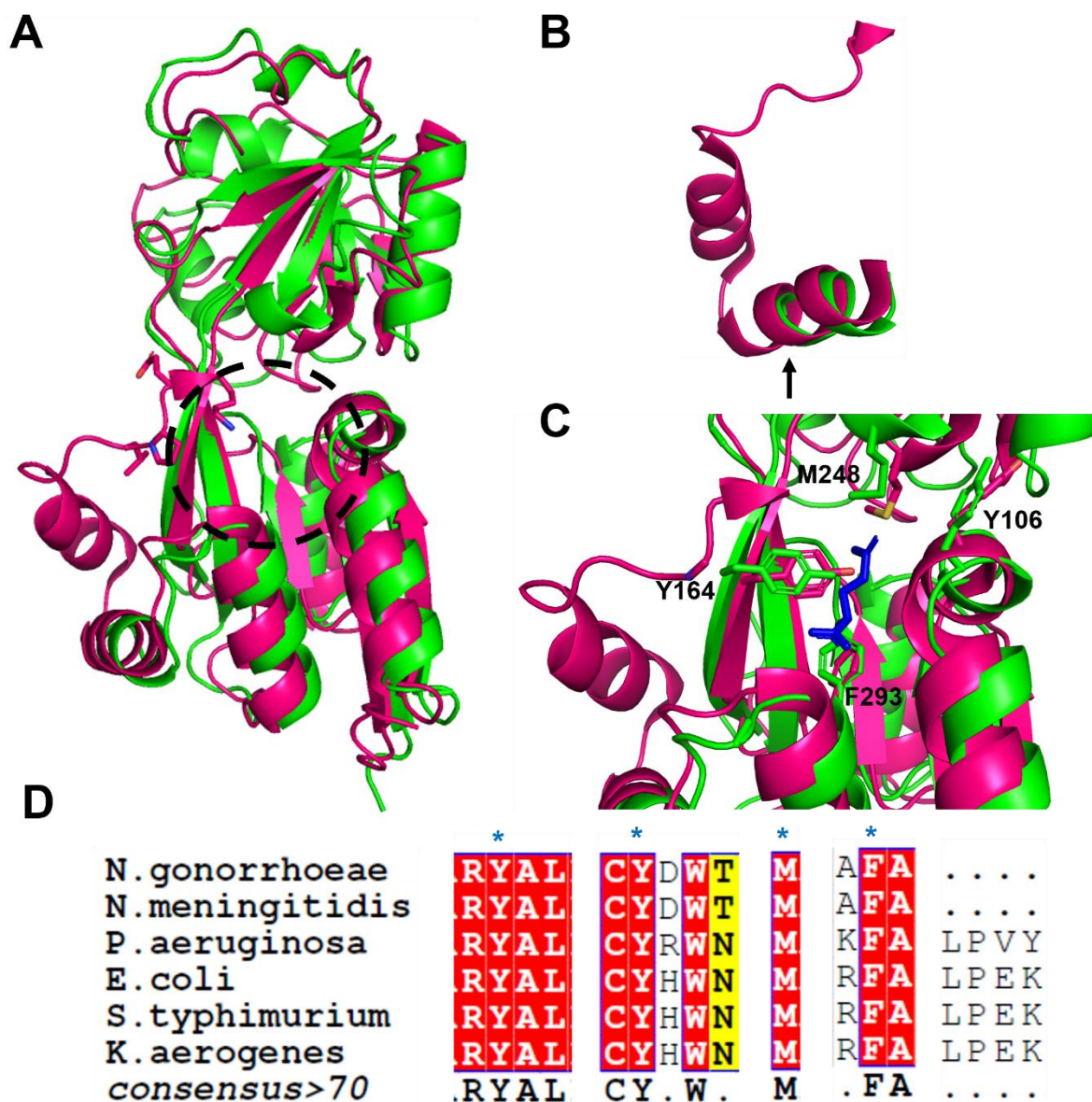
Figure 3.34 The secondary inducer-binding *site. **A** Structural overlay of the regulatory domain of N. gonorrhoeae (green) and the O-acetylserine-bound regulatory domain from S. typhimurium (hot pink) PDB code 4GXA. The location of the secondary binding site is highlighted. **B** The C-terminal arm missing from the N. gonorrhoeae structure but present in other solved CysB structures. An arrow indicates the final residue Leu306 which aligns to the previously solved CysB regulatory domain structures built in the compact monomer of CysB from N. gonorrhoeae. **C** Structural overlay of the secondary binding site within CysB from N. gonorrhoeae (green) and S. typhimurium (hot pink). O-acetylserine is shown in dark blue bound into the cleft in the 4LQ2 structure. The residues Tyr106, Tyr164, Met248 and Phe293 involved in forming the hydrophobic cleft are shown and labelled. **D** Extract from peptide sequence alignment of CysBs showing the conservation of residues involved in the secondary binding site. * above residues indicate those involved in forming the hydrophobic cleft. Figure generated using PyMOL version 2.3.4.*

Amongst the regulatory domains of the CysB protein there is a small interaction between two helices of the two compact monomers (α_{10}) which buries a surface of 398.7 Å^2 (Figure 3.35). This interface is not thermodynamically stable and is not significant for complex formation but has an interesting arrangement. Here two sets of three arginine residues align and appear to interact (Figure 3.35A). Arginines at position 151 and 210 from one chain are placed on either side of Arg216 from the opposing chain to form one interaction (Figure 3.35B,C). All arginine residues in this interaction are only present in the CysB proteins from *N. gonorrhoeae* and *N. meningitidis* as substitutions. In fact, upon further investigation of the CysB peptide sequence alignment (Figure 3.28), there are many substitutions where the sequence of *N. gonorrhoeae* and *N. meningitidis* contain an arginine residue in place of various amino acids including glutamine (at position 14), histidine (55), glutamic acid (72), serine (75), alanine (151), methionine (160), threonine (186), glutamine (192), threonine (210) and glycine (216). Further inspection of this helix interaction in COOT showed there was strong density for these residues (Figure 3.35D) with bond distances between atoms in the arginine side chains between 3.6-4.4 Å indicating the presence of weak hydrogen bonding. Due to the positive nature of arginine residues this interaction is likely due to crystal packing and may not be a true interaction. Further investigation into the role of these arginine residues may elucidate their potential function.
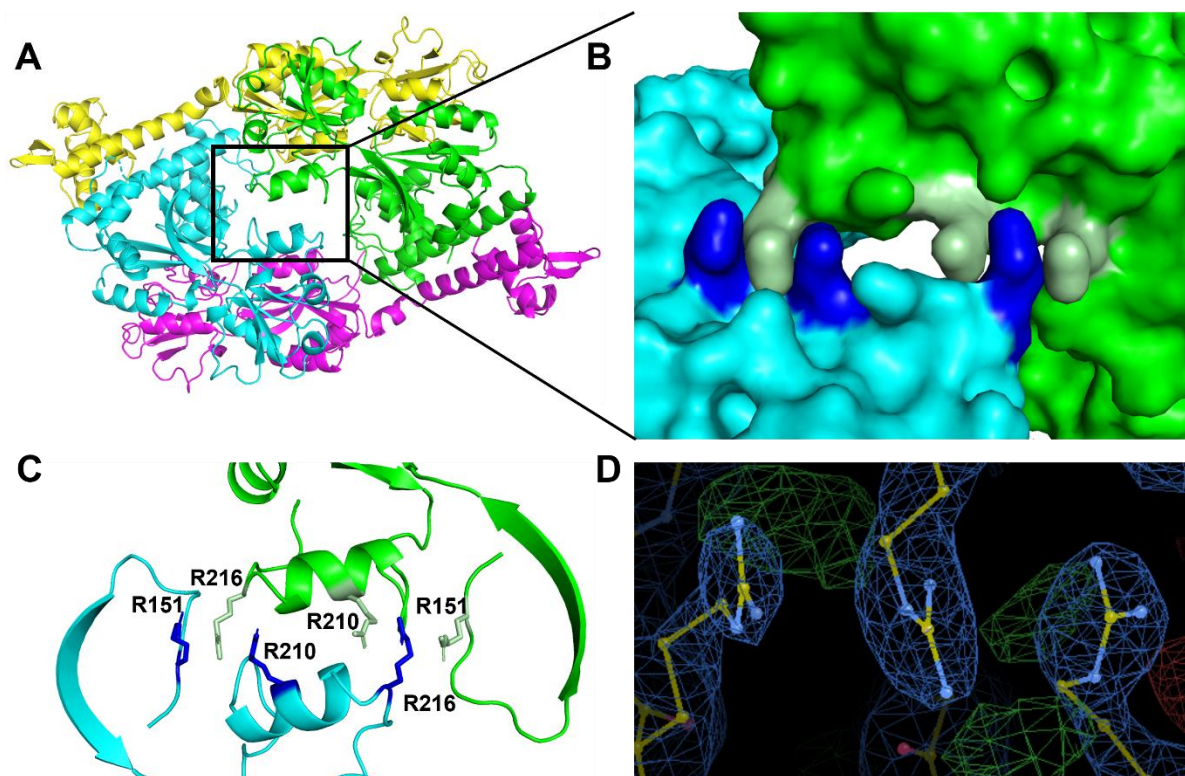
*Figure 3.35 Helix interaction within the regulatory domains of CysB. **A** Location of the two α_{10} helices within the CysB tetramer. This interaction only occurs between these two helices of the compact monomer in the structure. **B** Surface map of this interaction where the arginine residues are shown in dark blue and pale green to show their distinct interaction. **C** Labelled side chains and their positions relative to the helix. Arg151 is on a loop region preceding the helix, Arg210 is part of the helix and Arg216 is in the loop after the helix. **D** 2Fo-Fc map contoured to 1σ showing density for each of the arginine residues during model building indicating strong density and confidence in their placement. Figure generated using Coot and PyMOL version 2.3.4.*

An interaction between these two α_{10} helices within the two compact monomers has been described in other LTTR structures where these helices were termed contact helices. In the solved structures of TsaR (Monferrer et al., 2010), CbnR (Muraoka et al., 2003), tetragonal crystal forms of BenM (Clark, Haddad, Neidle, & Momany, 2004), the reduced form of OxyR (Choi et al., 2001), Pa01-PR (unpublished, PDB code 3FZV) and DntR (Smirnova et al., 2004) an interaction between these two contact helices was observed and described as a transient interaction. For the contact helices of TsaR, CbnR and Pa01-PR the non-buried side of the helices showed a polar surface with positively and negatively charged residues. For this surface in TsaR it was postulated that a 180° rotation would allow such residues to form complementary interactions and as such these transient interactions may be involved in structure stabilisation or a conformational change between the *apo*-state and inducer-bound state. Analysis by Monferrer of all other solved LTTR structures shows a β-turn containing a pair of either proline or glycine residues on one side of the inducer-binding cavity immediately

preceding the contact helix and a conserved glycine residue at the other end of the helix (Figure 3.36A). It is hypothesised that flexibility of the contact helix is due to the limited mainchain torsion angles of glycine and the ability of proline residues to adopt reversible *cis* or *trans* conformations allowing reorientation of the polypeptide chain (Monferrer et al., 2010). The only exception is OxyR where two cysteine residues are positioned either side of the contact helices and form a disulfide bridge in the oxidised form disrupting the helix and presenting a loop instead (Choi et al., 2001; Zheng, Åslund, & Storz, 1998).

These conserved glycine and proline residues surrounding the contact helices among LTTRs provide one explanation for potential labile interactions that may be involved in inducing a conformational change within the structure between an *apo*-state and an inducer-bound state. However, these glycine and proline residues conserved among LTTRs are not conserved in the sequence of CysB from *N. gonorrhoeae*. A sequence alignment of LTTRs by Monferrer et al. (2010) (Figure 3.36A) shows the presence of glycine residues in the CysB sequence from *K. aerogenes* (Monferrer et al., 2010). Alignment of CysB sequences from *N. gonorrhoeae* and *N. meningitidis* with other CysB sequences including that from *K. aerogenes* reveals that CysB from both *N. gonorrhoeae* and *N. meningitidis* have lost these conserved residues (Figure 3.36B). The CysB sequence from *N. gonorrhoeae* has alanine residues in place of the conserved glycines (or proline) at positions 201 and 204 and an arginine in place of the conserved glycine following the helix at position 216, which in the *N. gonorrhoeae* CysB structure participates in the contact helices interactions (Figure 3.35C). In the CysB structure from *N. gonorrhoeae* there is one glycine present preceding the helix upstream of the conserved residue positions (Gly205) and one proline further downstream of the helix (Pro220) which may confer some re-orientation of the helix (Figure 3.36C).
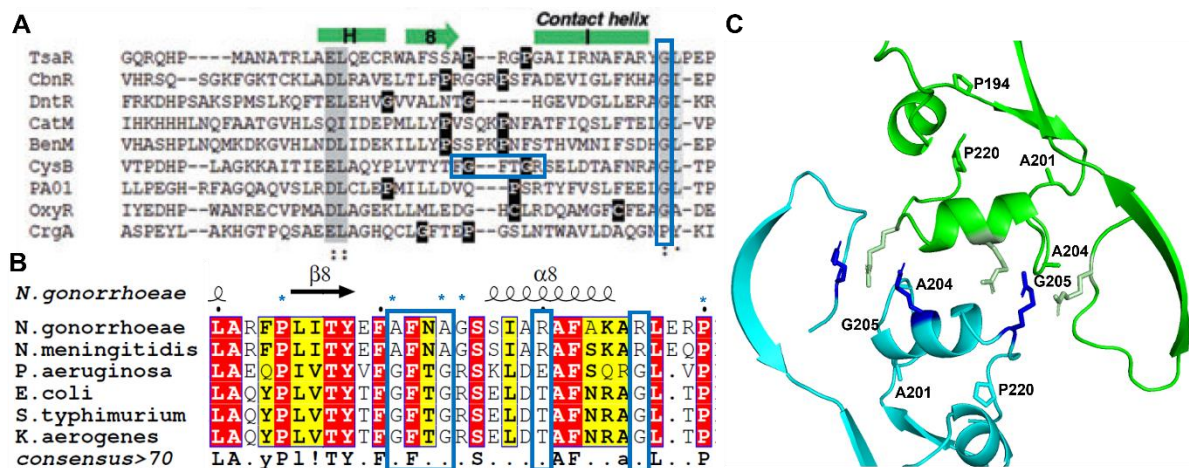
*Figure 3.36 Conserved residues amongst LTTRs and CysBs. **A** Extract of a peptide sequence alignment of structurally solved LTTRs adapted from (Monferrer et al., 2010). This shows the peptide region surrounding the contact helix with conserved residues in grey boxes and structurally conserved glycine and proline residues in black boxes for each sequence. The blue boxes indicate the residues of interest. The CysB peptide sequence was from K. aerogenes. **B** Extract of the peptide alignment of CysBs including that from K. aerogenes for comparison with CysB from N. gonorrhoeae. The first blue box indicates the mutations of glycines to alanines in N. gonorrhoeae and N. meningitidis. The second and final boxes show mutations to arginines in the N. gonorrhoeae and N. meningitidis sequences. The final box includes the mutation to arginine in place of the structurally conserved glycine shown in **A**. * above residues indicate which residues are labelled on the corresponding sequence figure in **C**. **C** Labelled residues (denoted by * above the residues in the sequence alignment in **B**) mapped on and around the contact helices of CysB from N. gonorrhoeae. The two alanine residues (Ala201 and Ala204) replace the conserved glycines preceding the helix, while the only proline near this helix is Pro194. Figures generated using PyMOL version 2.3.4 and taken from the previously reported Figure 3.28.*

The Gly>Ala substitution at position 204 is of particular interest because this sits in the loop that is involved in inducer binding within the inducer-binding cavity. The absence of a conserved threonine residue within this loop region (substituted from Thr202 to Gln203) coupled with the Gly>Ala substitution may help to explain the orientation of this loop away from the inducer-binding cavity in the CysB structure from *N. gonorrhoeae* (Figure 3.33B). It is also possible that these substitutions may alter the interactions CysB has with inducers and anti-inducers. It is of particular interest that the CysB sequence from both *N. gonorrhoeae* and *N. meningitidis* has the same changes to alanines resulting in the loss of these conserved glycine and/or proline residues and both lack the conserved threonine in the inducer-binding cavity. This warrants further investigation into these protein structures and their functional interactions with inducers and anti-inducers. If it is indeed the conserved residues relative to the contact helices that is involved in a conformational change upon inducer binding, CysB from *N. gonorrhoeae* and *N. meningitidis* may have limited or different conformational changes

100

upon inducer binding which may impact the ability of CysB from *N. gonorrhoeae* and *N. meningitidis* to bind DNA.

Interestingly, within the CysB peptide sequence, there is one cysteine residue present at position 163 conserved across all organisms (Figure 3.28). The insertion of an additional cysteine residue, Cys181, in the CysB sequences of *N. gonorrhoeae* and *N. meningitidis* is particularly interesting. When mapped onto the CysB structure, both of these cysteine residues are on the solvent-accessible surface of each monomer (Figure 3.37). These residues are not involved in any dimerisation interface and are located within a β sheet (β6) near the inducer-binding cavity (Cys163) and in a loop region at one end of the protein wihtin domain II (Cys181). The presence of an additional cysteine in each monomer of CysB from *N. gonorhoeae* requires eight cysteine residues to form one CysB tetramer, which is in turn required to activate cysteine biosynthesis. The presence of eight cysteine residues in the transcriptional regulator which activates the genes required for cysteine synthesis, the *de novo* synthesis pathway required for many sulfur containing molecules within the cell including glutathione, is surely not a random error and indicates a role for these cysteine residues. The hypothesised role of these cysteine residues has not been described nor investigated experimentally. We hypothesise that these cysteine residues may interact with other thiol-containing molecules within the cell, such as free cysteine, which may alter the regulation of CysB. It may also be possible that the four cysteine residues within the two compact monomers may be placed to form an octamer with another CysB tetramer under oxidising conditions, although this would need to be investigated experimentally to confirm any such interaction.
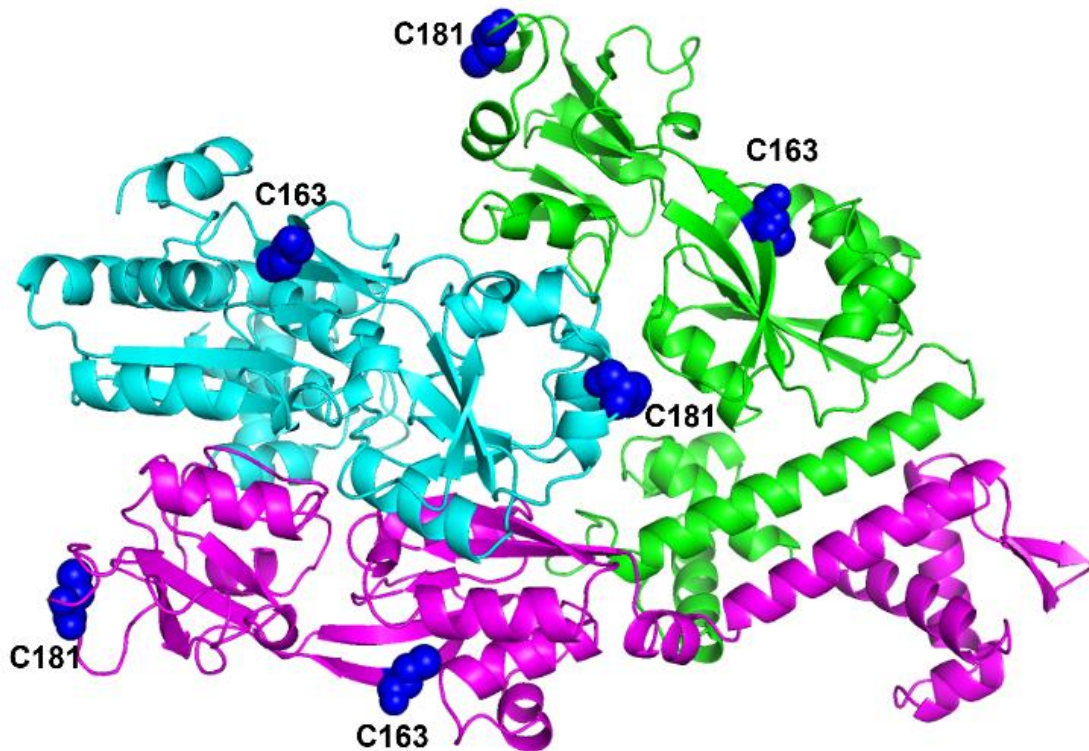
Figure 3.37 The cysteine residues within the CysB monomers. Cysteines are shown as dark blue spheres with one monomer omitted for visual clarity. Two cysteine residues are present within each monomer, one within a loop region at the end of the monomer (Cys181) while the other is in the sheet β6 (Cys163) which contributes to the crossover region and hence the inducer-binding cavity within the regulatory domain. Figure generated using PyMOL version 2.3.4.

Analysis of the solved full-length structure of CysB from *N. gonorrhoeae,* alongside investigation into conservation of residues within the CysB family has revealed some interesting aspects to CysB from *N. gonorrhoeae*. The peptide sequence of CysB from *N. gonorrhoeae* and *N. meningitidis* shows a truncated C-terminus, many substitutions and two insertions. Some of these substitutions occur at positions important in the inducer-binding cavity (Thr>Gln202/203), within regions hypothesised to allow conformation changes (contact helix 194-220) and regions reported to effect DNA binding activity (Thr>Ala22, Leu>Ile44). Furthermore, there is an additional cysteine residue in the peptide sequence which results in eight cysteines required to synthesise one CysB tetramer, an interesting aspect in the transcriptional regulator required to activate cysteine synthesis through the cysteine regulon. This range of changes in the peptide sequence of CysB from *N. gonorrhoeae*, coupled with structural investigation of the relative positioning of these residues and the overall tetrameric form of CysB, alludes to an interesting mechanism of induction. Currently, no full-length forms of CysB in an inducer-bound state have been elucidated and limited conformations changes

were reported in the *N*-acetylserine-bound regulatory domain, which was solved by soaking *apo*-CysB regulatory domain crystals in the inducer *N*-acetylserine which in turn may have prevented any large-scale conformational changes occurring within the crystal. Our attempts to grow crystals of CysB with the inducer *N*-acetylserine were unsuccessful, and as such, we set out to investigate any conformational changes occurring in CysB from *N. gonorrhoeae* between *apo-* and inducer-bound states using small angle X-ray scattering (SAXS).

3.4 Structural characterisation of CysB by Small Angle X-ray Scattering

In order to investigate conformational changes in CysB upon inducer binding we employed Small Angle X-ray Scattering (SAXS). Attempts to crystallise CysB with the inducer *N*-acetylserine bound using methods described in Sections 2.6.1, 2.6.2 & 2.6.3 were unsuccessful despite continued efforts. As such, solution-based X-ray scattering was an appealing solution as it would allow any conformation changes to occur in solution, free from various crystallisation components and crystal packing restraints. We investigated the scattering of CysB without inducer (CysB-*apo*) and with the addition of inducer *N*-acetylserine (NAS) (CysB-*NAS*).

3.4.1 Sample Optimisation and Optimisation of Data Collection

Analysis of protein scattering by SAXS requires optimisation of protein preparation, stability and concentration. We sent three sets of purified CysB protein samples (purified via methods described in Section 2.4) to the Australian Synchrotron SAXS/WAXS Beamline for SAXS analysis using the SEC-SAXS set up (Ryan et al., 2018) with co-flow in place (Kirby et al., 2016). Due to border closures and an inability to travel to Australia as a result of the COVID-19 pandemic, the Beamline scientists collected scattering data for these protein samples.

The first set of SAXS samples were analysed in June 2021 (experiment code 17284a). These samples consisted of one CysB only sample at 3.6 mg.mL^{-1} and one sample with CysB at 3.5 mg.mL^{-1} with 10 mM *N*-acetylserine. This protein concentration was used for crystallisation of CysB. This first set of SAXS samples showed that CysB eluted as a single peak from the size exclusion column (Appendix AD.1) and scattered well. However, there was a high background to noise ratio in the data, resulting in large error in the scattering plot as well as in key SAXS parameters such as the radius of gyration (Rg) which is the average distance of atoms from the centre of the particle and gives an average radius of the protein. Based on these data we deemed the protein concentration was too low and future analysis should focus on optimising protein concentration.

The second set of SAXS samples were analysed in October 2021 (experiment code 17630a). We wanted to analyse the scattering of CysB at two different concentrations, higher than previously tested (above). We therefore sent two samples of purified CysB at ~4.7 mg.mL^{-1} and ~7.2 mg.mL^{-1} and two further samples with *N*-acetylserine added to each (final concentration 10 mM). This second set of SAXS samples encountered many Beamline issues including a three-week delay where the protein was kept at 4°C and detector issues rendering the data obtained unsuitable for analysis, as seen in multiple peaks in the UV traces (Appendix AD.1) and hence no conclusions could be made.

The third set of SAXS samples were analysed in March 2022 (experiment code 18012a). Again, we wanted to analyse the scattering profiles of purified CysB to determine the optimal concentration. Therefore, in total we sent six CysB samples at three concentrations (~5 mg.mL^{-1}, ~6 mg.mL^{-1} and ~7 mg.mL^{-1}) with and without the addition of 10 mM *N*-acetylserine. We also sent an additional sample at 6.3 mg.mL^{-1} with 30 mM *N*-acetylserine. For optimal stability of the protein, CysB was purified and concentrated three days prior to SAXS analysis to reduce degradation. Purification of these samples was via the altered method described in Section 2.7.1 to remove any bound nucleic acids that may have been interfering with CysB scattering. All samples eluted as a single peak from the size exclusion column indicating sample stability and homogeneity (Appendix AD.1).

The scattering profiles for each sample were of high quality with some small differences detected between CysB in the *apo* form and with the inducer *N*-acetylserine bound. During background subtraction however, there was an issue with the CysB-*apo* data which prevented the Porod region from being visualised in the scattering plot using common background subtraction methods (Figure 3.38A). The data within the scattering plot can be separated into three regions; the Guinier region at low q values which allows analysis of data quality indicating aggregation or interparticle interference in the sample, the Kratky region which gives information on the flexibility of the protein and the Porod region of the scattering plot at high q values which gives information on the shape of the protein. The Porod region is used to generate the P(r) distribution which is consequently used to generate the input files for model building programmes.

While the Guinier and Kratky plots could be visualised from the original scattering plot (Figure 3.38B, C), the multiplication factor had to be changed from the default 1.00 to 0.99 to visualise the data required to allow a complete scattering pattern (Figure 3.38D). The Guinier plots

(Figure 3.38B) revealed that CysB with the addition of 30 mM *N*-acetylserine aggregated as shown by a 'smiling' plot where there is an upward trend at both low and high q values. The Kratky plot (Figure 3.38C) of this CysB-*NAS* data with 30 mM *N*-acetylserine confirmed that CysB was aggregated as the data did not return to the baseline and instead was tracking upwards. Comparison of the Kratky plots between CysB-*apo* where the multiplication factor was 1.00 and CysB-*NAS* at 10 mM *N*-acetylserine (also with a multiplication factor of 1.00) showed a delay in the CysB-*NAS* data returning to baseline which indicated some flexibility differences in CysB when inducer was bound (Figure 3.38C).

Altering the multiplication factor altered all regions of the scattering plot (Figure 3.38D) and hence altered downstream processing using this data including the P(r) distribution plots and model building. This reduced our confidence in any models produced from the altered scattering data which prevented any conclusive differences between the samples being drawn as model building is required to inform the overall shape of the protein in solution where Kratky plots are used as an indicator of differences but are not themselves conclusive.
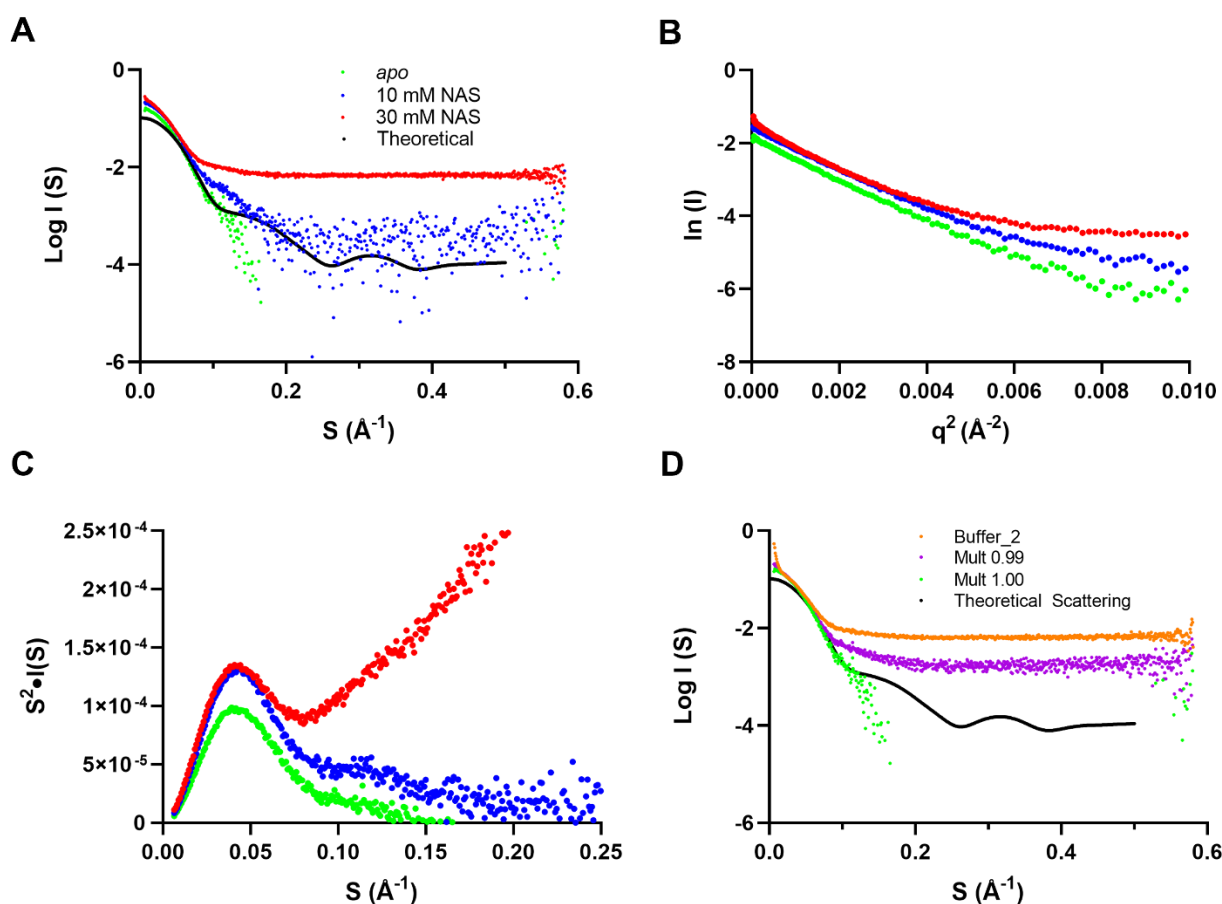
*Figure 3.38 SAXS data quality plots during sample optimisation. The key SAXS data quality plots - Scattering, Guinier and Kratky plots of CysB samples with inducer N-acetylserine showing further optimisation was required. CysB-apo data is shown in green, CysB-NAS at 10 mM final concentration data is shown in blue and CysB-NAS at 30 mM final concentration data is shown in red. **A** The scattering data shows that the Porod region of the CysB-apo data (green) is not visible. There are some differences in the scattering data between the visible CysB-apo data and the CysB-NAS at 10 mM (blue) data. The theoretical scattering generated from the X-ray crystallography model (in black) shows a good fit to the visible CysB-apo data where differences are seen between the theoretical scattering and the CysB-NAS at 10 mM. The CysB-NAS at 30 mM (red) scattering data is markedly different to the other scattering data. **B** The Guinier plots of the scattering data. The upwards curve at both low and high q values of the CysB-NAS at 30 mM (red) indicates aggregation in the sample. Both the CysB-apo (green) and CysB-NAS at 10 mM (blue) data indicates no aggregation. **C** The Kratky plots from the scattering data. The CysB-apo (green) and CysB-NAS at 10 mM (blue) data returns to baseline indicating the protein is folded under these conditions. The delay in the CysB-NAS at 10 mM data returning to baseline indicates more flexibility in the protein than CysB in the CysB-apo state. The CysB-NAS at 30 mM (red) data trends upwards and does not return to the baseline further confirming that the protein is aggregated under these conditions. **D** Scattering plots of the CysB-apo data using different background subtraction methods. When the multiplication factor was set to the default value of 1.00 (green) the Porod region of the scattering data is not visible. When the multiplication factor was changed to 0.99 (purple) the scattering data begins to trend away from the visible data at multiplication factor of 1.00 and from the theoretical scattering (black) where the points at high q no longer overlay with the visible multiplication factor 1.00 points in green. When a blank buffer run was used as the buffer subtraction (orange), the whole scattering profile changed and no longer resembled either the visible multiplication factor 1.00 data or the theoretical scattering data. Figures generated using Prism version 9.4.1.*

The removal of border restrictions allowed us to travel to collect a fourth set of SEC-SAXS data in person at the Australian Synchrotron SAXS/WAXS Beamline during August 2022 (experiment codes 18012d & 18574). CysB was analysed at 6 mg.mL^{-1} without and with 10mM of inducer *N*-acetylserine. CysB was purified (Section 2.7.1) four days prior to analysis and was concentrated to 2mg.mL^{-1} prior to travel and then to 6mg.mL^{-1} at the Synchrotron before analysis. The addition of 0.1% sodium azide to reduce radiation damage had no effect on data quality indicating the 10% *v/v* glycerol as a buffer additive was adequate to protect from radiation damage. The UV traces showed most samples eluted as a single peak however, most peaks had a preceding shoulder not seen in previous SAXS data collected as part of this research (Figure 3.39A). Background subtraction using CHROMIXS (Panjkovich & Svergun, 2018) showed there were multiple species within the peak (Figure 3.39B) and allowed some deconvolution of peaks by analysing the Radius of Gyration (Rg) values across the peak to only include data points across a narrow range of Rg values (Rg = 42 ± 2). Scattering patterns were of high quality and all data could be visualised using common background subtraction methods (where the multiplication factor is set to the default value of 1.00) (Figure 3.39C). Guinier analysis confirmed samples were free from aggregation and interparticle interference (Figure 3.39D). Raw data files can be found in Appendix AD.2.
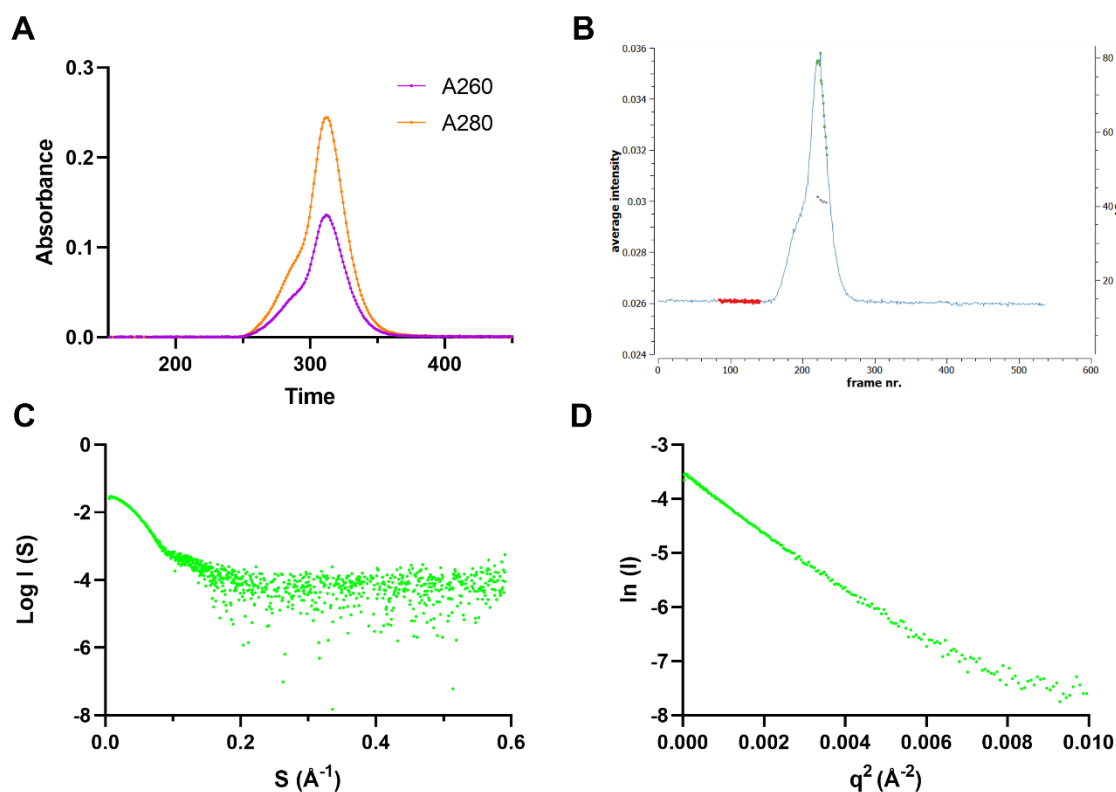
*Figure 3.39 SAXS data quality plots of final CysB-apo sample. CysB eluted as a single peak from the SEC column as seen in the UV trace (**A**). The shoulder in the UV trace (**A**) could be deconvoluted using CHROMIXS (**B**). The CHROMIXS plot shows the scattering intensity over time. The points in green represent sample data points, the red data points are buffer for buffer subtraction. Analysis of the Rg values across the peak allowed Rg values larger than 42 ± 2 to be excluded. The remaining grey points show the narrow range of Rg values of the remaining data points. These points were used in downstream data processing including the scattering plot (**C**) and Guinier plot (**D**). The scattering plot and Guinier plot indicate no aggregation or interparticle interference was present in this data set. Figures generated using Prism version 9.4.1 and CHROMIXS from the ATSAS Suite 3.1.3.*

3.4.2 Analysis of SAXS data

The data collected from samples during the fourth SAXS trip showed a good scattering profile where all the data was visible using default background subtraction methods. The Guinier plots showed no evidence of aggregation (uptick at low q) or interparticle interference (downtick at low q) indicating good quality data. The Kratky plots showed the protein was folded as the plot returned to baseline. Extrapolation of the Rmax to an artificially large value showed no additional peaks indicating no aggregation was present in the samples. The quality of this data therefore allowed comparison between CysB without inducer (CysB-*apo*) and with the inducer *N*-acetylserine (CysB-*NAS*). While some samples showed evidence of aggregation, there were enough datasets collected without and with inducer present to allow comparison between two data sets seemingly free of aggregation.

The scattering plots of CysB-*apo* and CysB-*NAS* show very similar profiles (Figure 3.40) and are consistent with a globular protein. The scattering profiles reported here showed the same scattering profile for all collected CysB datasets across all trips indicating stability of CysB during SAXS. The theoretical scattering calculated from the final X-ray crystallography structure is also shown on the graph (Figure 3.40). Alignment of the theoretical CysB scattering data to the CysB-*apo* and CysB-*NAS* data gave χ^2 values of 0.8157 for the fit of the theoretical scattering to the CysB-*apo* data, and 1.515 for the fit of the theoretical scattering data to the experimental data from CysB-*NAS*. This mismatch is not unexpected as proteins experience dynamic movement in solution while the X-ray data is from a rigid crystal structure. It may also be possible that in solution without DNA bound, that CysB is a flexible structure where bound DNA may lock it into a less flexible conformation. Any flexibility in solution would be captured in the scattering data. The calculated molecular mass based on the Bayesian Inference from the Porod volume was 185.8 kDa, which combined with the globular pair wise distribution (P(r)) supports a CysB tetramer in the SAXS data.

Figure 3.40 Scattering plot of CysB-apo vs CysB-NAS. CysB-apo data is shown in green while CysB-NAS data is shown in blue. The scattering profiles of the CysB-apo and CysB-NAS data are highly similar to one another. Both the CysB-apo and CysB-NAS scattering profiles are similar to the theoretical scattering profile generated from the X-ray crystallography structure (black). Figure generated in Prism version 9.4.1.

The Guinier plots of the CysB-*apo* and CysB-*NAS* datasets indicate good quality data was attained free from aggregation and interparticle interference (Figure 3.41). The radius of gyration (Rg) is an approximate measure of the average radius of the protein (Å) by describing the volume distribution of the sample. The Rg values of CysB-*apo* and CysB-*NAS* calculated from the Guinier analysis were 42.09 ± 0.20 and 40.98 ± 0.24 respectively while the Rg values based on the P(r) distribution were 42.90 ± 0.03 and 41.84 ± 0.02 respectively. The small difference between the Rg values calculated from the Guinier analysis and via the P(r) distribution is expected as the Guinier analysis uses a small range of data within the Guinier region, while the P(r) distribution uses all the data to calculate the Rg value. The P(r) distribution is, therefore, a more reliable calculation of Rg. The magnitude of difference between the Rg values of the CysB-*apo* and CysB-*NAS* data is the same (~1 Å) which gives confidence in the small differences between these values. There is a small difference between the Rg values of the CysB-*apo* and CysB-*NAS* data sets, which may indicate small changes between the *apo* and inducer-bound states, where the CysB-*NAS* state is more compact.
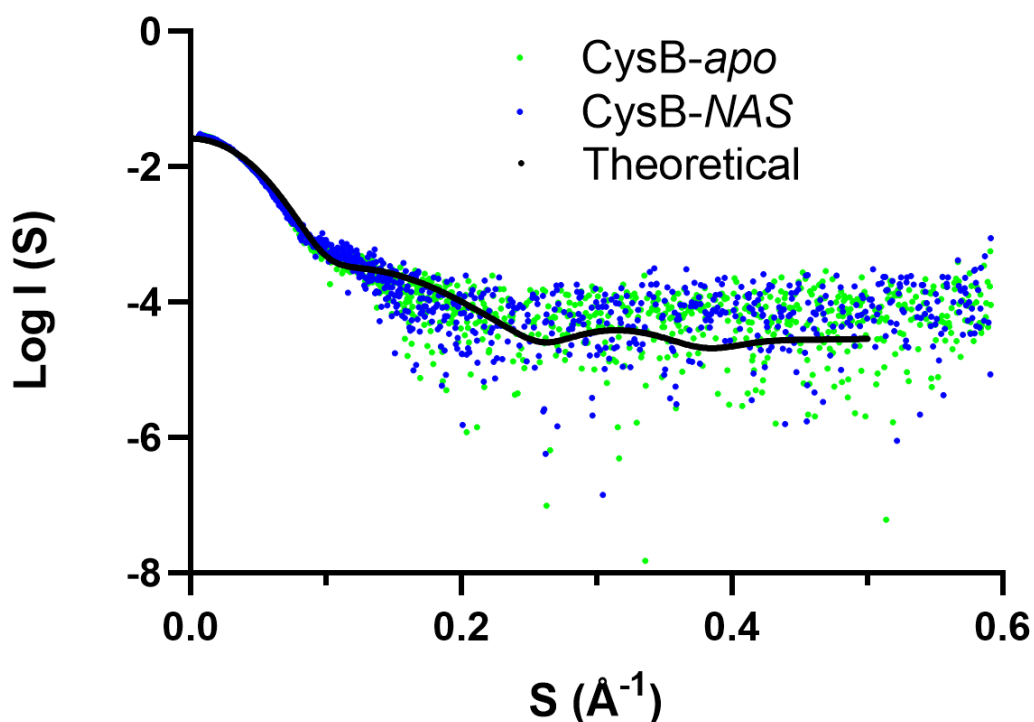
Figure 3.41 Guinier plot of CysB-apo vs CysB-NAS. CysB-apo data is shown in green while CysB-NAS data is shown in blue. The Guinier plots of the CysB-apo and CysB-NAS indicate that there is no aggregation of interparticle interference present in these samples as shown by a straight line with no curvature at low q. Figure generated in Prism version 9.4.1.

The P(r) distribution plot shows a bell-shaped curve which indicates the protein has a globular shape in both the CysB-*apo* and CysB-*NAS* states (Figure 3.42). This indicates the overall shape of the protein does not drastically change upon inducer binding as seen in the LTTR family member DntR from *Burkholderia cepacia* using SAXS (Lerche et al., 2016). DntR scattering data exhibited differences in the Pr distribution upon ligand binding, leading to large scale changes in the protein from a globular shape to a hollow sphere upon ligand binding. These data from DntR are supported by dummy atom models showing a large open cavity in the middle of the tetramer upon ligand binding (Lerche et al., 2016). These large changes are not present in our scattering data, with the Pr distributions of *apo* and ligand-bound scattering data highly similar (Figure 3.42). The Rmax value is the distance across the widest part of the protein. This value is automatically calculated when generating the P(r) distribution plot, however, this is an approximate value. The P(r) distribution generates where the distribution re-approaches zero which is deemed the Rmax value. Where the distribution drops too quickly

to zero, the Rmax value is likely calculated too small. Manually altering the Rmax value allows visualisation of what the true Rmax value may be, which can have a profound effect on the model building, as the saved output of this distribution is used as the input for model building programs such as DAMMIN (Svergun, 1999), DAMMIF (Franke & Svergun, 2009) and GASBOR (Svergun et al., 2001). Upon investigation an Rmax of 160 Å showed a smooth approach to zero for the CysB-*apo* data, while an Rmax of 150 Å showed a smooth approach to zero for the CysB-*NAS* data.
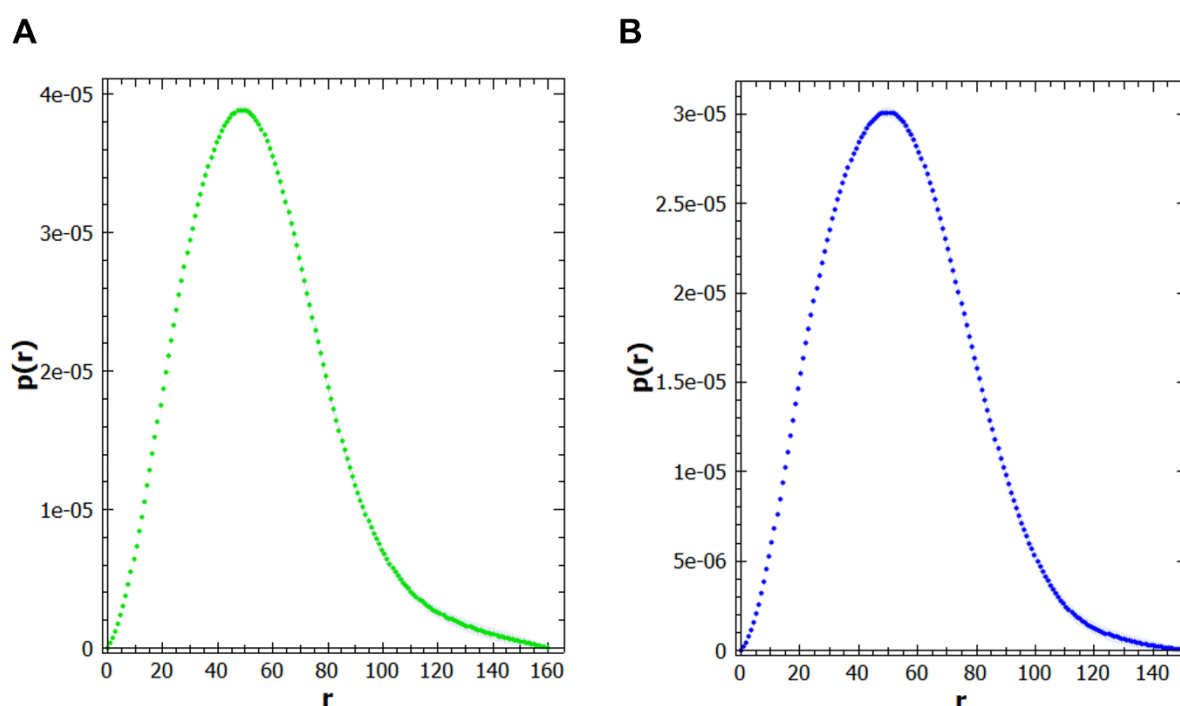
*Figure 3.42 P(r) distributions of CysB-apo and CysB-NAS. CysB-apo data is shown in green (**A**) while CysB-NAS data is shown in blue (**B**). The P(r) distribution shows CysB is a globular protein in both the apo and inducer-bound states due to the bell-shaped curve. A gradual regression to zero at an Rmax of 160 for the CysB-apo dataset indicates that the maximum distance (Rmax) across CysB in an apo state is 160 Å. A gradual regression to zero at an Rmax of 150 for the CysB-NAS dataset indicates that maximum distance across CysB when N-acetylserine is bound is 150 Å. Figures generated in primus from the ATSAS suite version 3.1.3.*

The Kratky plots can give information on the flexibility of the protein and can determine if the protein is unfolded. The Kratky plots of both the CysB-*apo* and CysB-*NAS* samples return to the baseline which indicates that CysB in both samples is folded (Figure 3.43). The delay in return indicates flexibility within the structures although there are no obvious differences in flexibility between these samples (Figure 3.43). If there were large-scale conformational changes between CysB-*apo* and CysB-*NAS* we would expect to see a difference in flexibility

112

between the two forms which would be visible as a horizontal shift in the Kratky plot and a delay in the return of the plot to baseline.
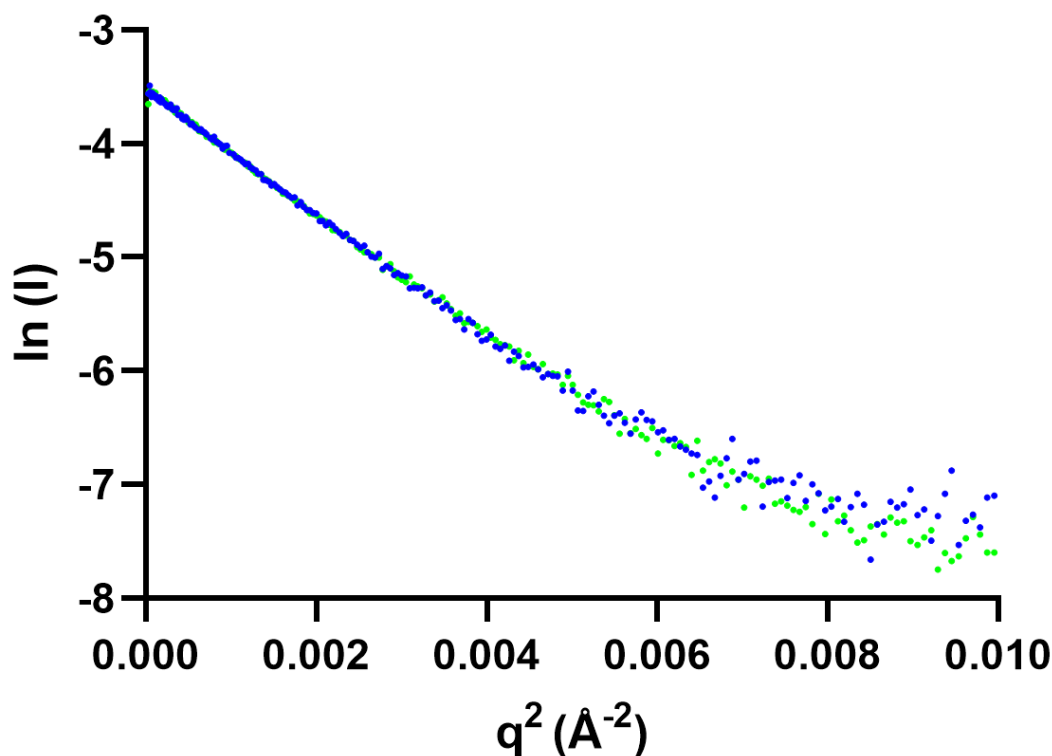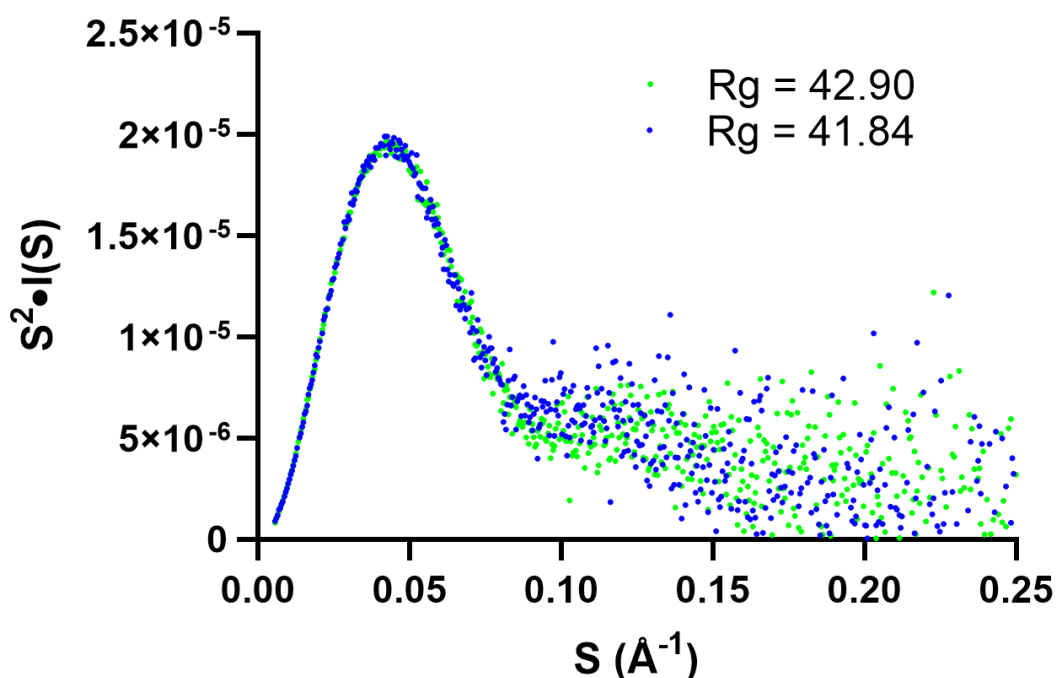

Figure 3.43 Kratky plot of CysB-apo vs CysB-NAS. CysB-apo data is shown in green while CysB-NAS data is shown in blue. The Kratky plots of both the CysB-apo and CysB-NAS data shows a similar profile with both data sets returning to baseline indicated CysB is folded in both datasets. The absence of any notable differences indicates there is no difference in flexibility of CysB without and with inducer N-acetylserine bound. Radius of gyration (Rg) values generated from the Porod volume are displayed. Figure generated in Prism version 9.4.1.

3.4.3 Model building from SAXS data

The model-building programs DAMMIN (Svergun, 1999), DAMMIF (Franke & Svergun, 2009) and GASBOR (Svergun et al., 2001) work by bead-modelling to create dummy atom models where one 'bead' is placed within the measured experimental 'density' and calculations are run to determine whether the 'bead' is more likely to be protein or solvent. This model-building approach does not require the 'protein beads' to be in a biologically relevant orientation. i.e. connected as a polypeptide chain, which allows some error in the generated models. These models produce a low-resolution output which is more representative of the surface model of the protein and cannot give any atomic detail. SAXS is therefore a complimentary low resolution technique and should be used alongside high resolution structural techniques such as X-ray crystallography. Structural models generated from the CysB-*apo* and CysB-*NAS* data using DAMMIN, DAMMIF and GASBOR were all comparable

to one another and had similar dimensions to the overall dimensions of the X-ray crystallography structure of CysB. The model outputs from DAMMIF were averaged using DAMAVER (Volkov & Svergun, 2003) to generate an averaged surface of the CysB protein from the experimental scattering data.

Alignment of these built surface structures from the CysB-*apo* and CysB-*NAS* scattering data to the X-ray crystallography-solved structure was conducted using the PyMOL plugin SASpy SUPALM (Konarev, Petoukhov, & Svergun, 2016) which generates NSD values (normalised spatial discrepancy). NSD values allow comparison between high resolution data, i.e. from an X-ray crystallography structure, to low resolution data, from the SAXS envelope (Kozin & Svergun, 2001). For ideally superimposed objects the NSD value tends toward zero where NSD values exceeding one indicate the objects substantially differ from one another. The magnitude of the NSD value, however, also depends on the level of detail represented between the objects, for example where all atoms in the X-ray crystallography structure are used while the surface models include 'beads' representing an area of data (Franke et al., 2017). Hence, NSD values should be interpreted with caution.

The P(r) distribution from the CysB-*apo* scattering data showed a suitable return of the distribution to zero at an Rmax of 160 Å (Figure 3.42A) which was used for model building. The CysB-*apo* SAXS surface model (Figure 3.44) shows 'density' surrounding the X-ray crystallography structure. There appears to be two concave areas of the SAXS model on the same plane which align somewhat to the DNA-binding domains of the X-ray structure (Figure 3.44). Alignment of the X-ray crystallography structure to the SAXS surface model generated from the CysB-*apo* data gave an NSD value of 8.0548. This indicates differences between the X-ray crystallography structure and the SAXS surface model. This large difference may be the result of the high-resolution atomic detail within the X-ray crystallography structure, while the SAXS model contains dummy beads. Visual inspection of this overlay shows the SAXS model is much larger than the X-ray structure which may indicate flexibility within CysB when in the CysB-*apo* state which may have been captured and averaged across the scattering data (Figure 3.44).

Figure 3.44 CysB-apo SAXS surface model. The DAMAVER generated surface SAXS model from the CysB-apo scattering data overlaid with the X-ray crystallography structure. The X-ray structure is coloured by chain in line with previous X-ray figures. The SAXS model 'density' surrounds the X-ray structure. Two concave areas of the surface SAXS model can be seen on the same plane which align somewhat to the DNA-binding domains of the X-ray structure. Figure generated in PyMOL version 2.3.4.

While the P(r) distribution from the CysB-*apo* scattering data showed a suitable return of the distribution to zero at an Rmax of 160 Å, the CysB-*NAS* scattering data showed a more suitable return of the distribution to zero at 150 Å (Figure 3.42B). This further supported the differences in Rg values indicating that CysB with the inducer *N*-acetylserine bound is slightly more compact than CysB in the CysB-*apo* state. The CysB-*NAS* SAXS model (Figure 3.45) shows density packed around the X-ray crystallography structure. Here there appears to be two concave regions on the same plane that align somewhat with the DNA-binding domains of the X-ray structure (Figure 3.45). Alignment of the X-ray crystallography structure to the SAXS surface model generated from the CysB-*NAS* data gave an NSD value of 4.8700. While this indicates differences between the X-ray structure and the CysB-*NAS* SAXS model, this is a much smaller NSD value than the overlay of the X-ray structure with the CysB-*apo* SAXS model indicating this CysB-*NAS* SAXS model more closely resembles the CysB X-ray structure.
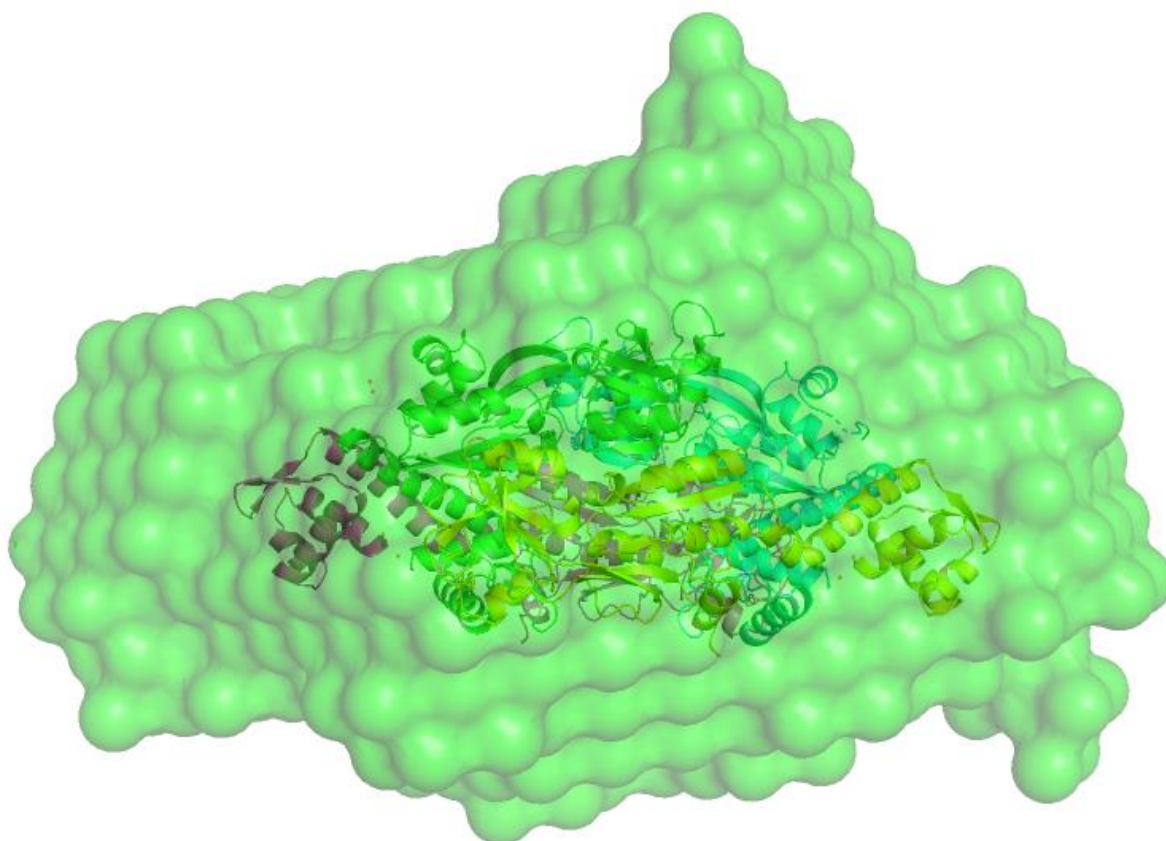
Figure 3.45 CysB-NAS SAXS surface model. The generated surface SAXS model from the CysB-NAS scattering data overlaid with the X-ray crystallography structure. The X-ray structure is coloured by chain in line with previous X-ray figures. The SAXS model 'density' surrounds the X-ray structure. Two concave areas of the surface SAXS model can be seen on the same plane which align somewhat to the DNA-binding domains of the X-ray structure. Figure generated in PyMOL version 2.3.4.

As the presented DAMAVER model averages many DAMMIF models, flexibility of CysB in solution would be captured and hence averaged in this model, generating a larger SAXS model as shown in the CysB-*apo* SAXS model (Figure 3.46). This can be further seen when comparing the CysB-*apo* SAXS model with the CysB-*NAS* SAXS model (Figure 3.46). Here the differences in Rg values indicating CysB becomes more compact when inducer *N*-acetylserine is bound become more profound. The CysB-*apo* SAXS model is larger than the CysB-*NAS* model and has additional density (Figure 3.46).
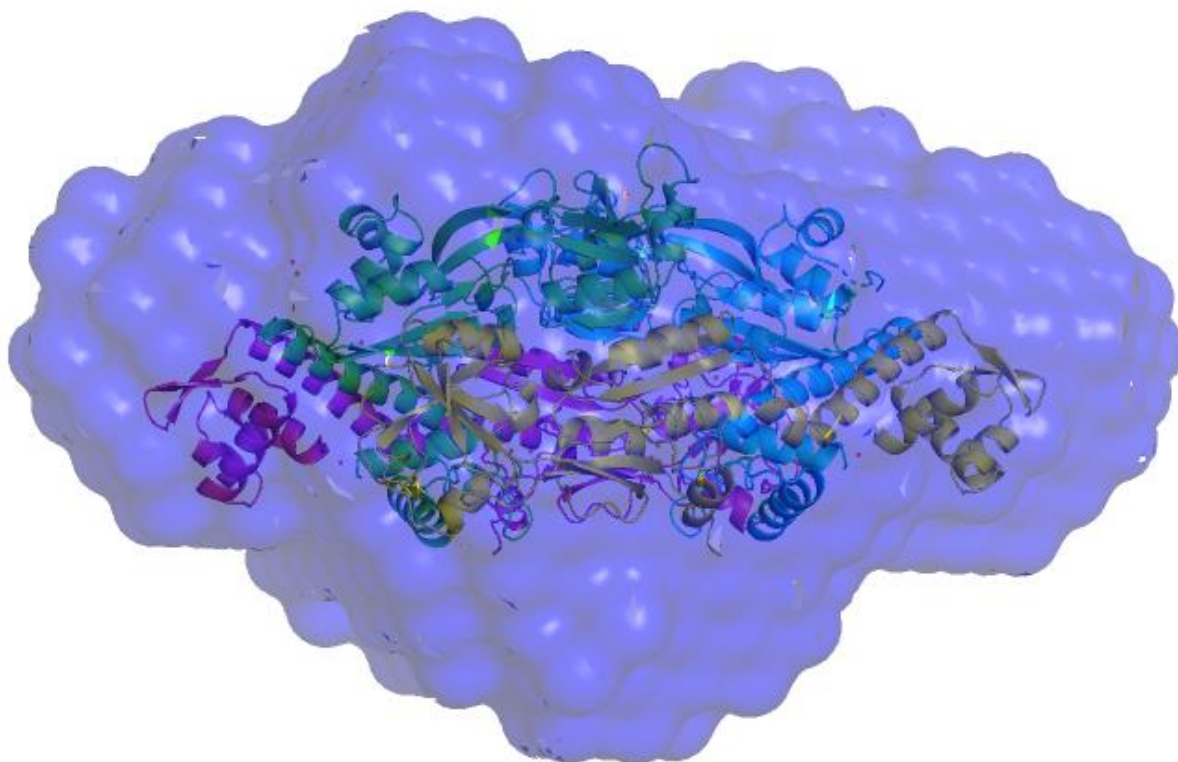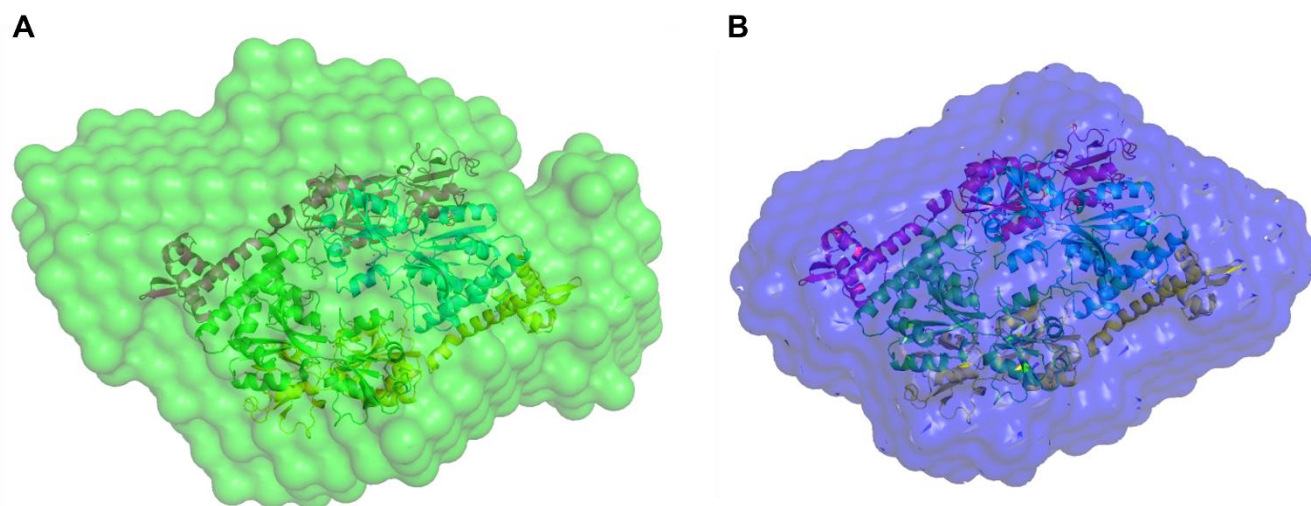
Figure 3.46 CysB-apo vs CysB-NAS SAXS surface models. The generated surface SAXS models from the CysB-apo and CysB-NAS scattering data, each overlaid with the X-ray crystallography structure. The X-ray structure is coloured by chain in line with previous X-ray figures. This shows the differences in 'density' of each model surrounding the X-ray structure. Both SAXS models are oriented the same for comparison, viewing the contact helix interaction op 'top' of the structure. Figures generated in PyMOL version 2.3.4.

When the SAXS models generated from the CysB-*apo* and CysB-*NAS* data were overlaid, an NSD value of 0.7138 indicated these were similar models. The model profiles show minimal differences within the core of the model (Figure 3.47) with some differences around the outer surface of the models and additional density in the CysB-*apo* model (Figure 3.47). Both of these models fit with the P(r) distributions of the CysB data demonstrating a globular protein with no large open cavity like that seen in the scattering data of the LTTR family member DntR (Lerche et al., 2016) i.e. no hollow shape in the inducer-bound state (Figure 3.45). Interestingly, the CysB-*NAS* model appears to more closely resemble the X-ray structure based on the NSD value and visual inspection of the alignment to the surface model (Figure 3.46), even though the X-ray structure is an *apo*-structure.
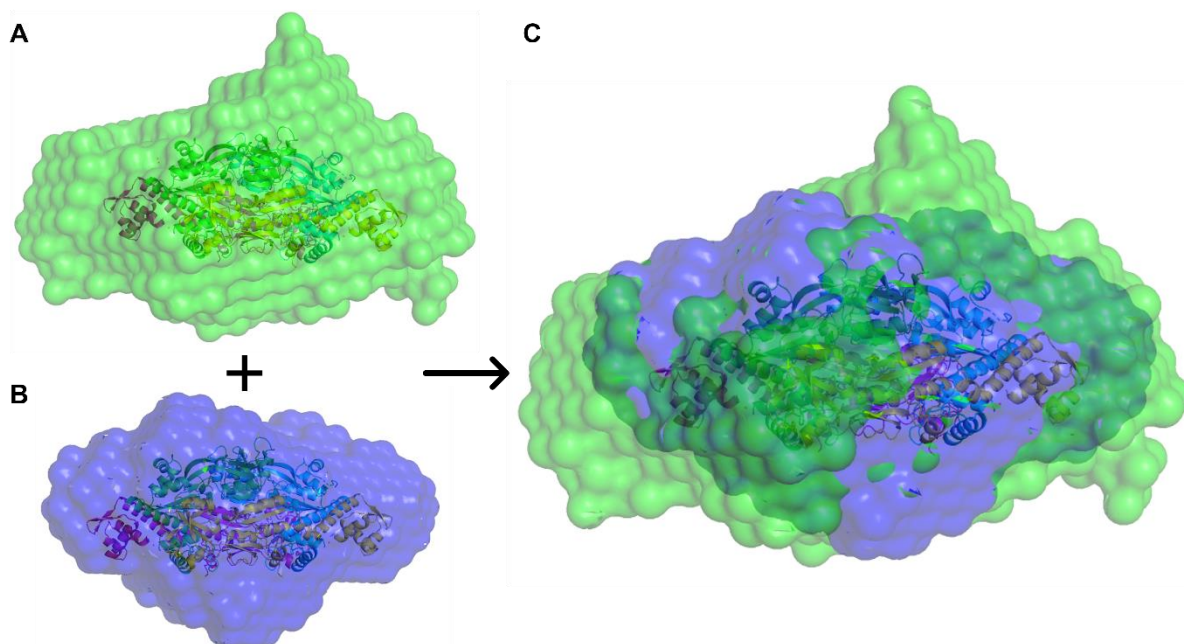
*Figure 3.47 Overlay of the CysB-apo and CysB-NAS SAXS surface models. The individual SAXS surface models from the CysB-apo (**A**) and CysB-NAS (**B**) scattering data when overlaid (**C**). All models are oriented with the DNA-binding domains on the same plane for comparison. In the overlaid model (**C**) the core 'density' overlays well while the CysB-apo model shows additional density surrounding the CysB-NAS SAXS model. Figures generated in PyMOL version 2.3.4.*

Between the SAXS models generated from the CysB-*apo* and CysB-*NAS* scattering data, there appears to be more density within the SAXS model from the CysB-*apo* data above the regulatory domains than that seen in the SAXS model from the CysB-*NAS* data (Figure 3.47C). If inducer-binding leads to small conformational changes resulting in a slightly more compact state, as shown by a smaller Rg value, perhaps the regulatory domain may experience small conformational changes leading to this flexible density, only shown in the CysB-*apo* state, becoming more compact and hence no longer seen in the inducer-bound CysB-*NAS* model. This would explain the additional density in the CysB-*apo* model as CysB may be more flexible without inducer bound, as indicated by the higher Rg value.

The low-resolution surface model makes it difficult to differentiate exactly how the DNA-binding domains are arranged between the CysB-*apo* and CysB-*NAS* SAXS models (Figure 3.47). There appears to be a small difference in the positioning of the DNA-binding domains where the inducer-bound structure has the DNA-binding domains positioned away from the same plane in a 'V'-like shaped position. In the CysB-*apo* SAXS model the DNA-binding domains appear almost as an 'm'-like shape. This SAXS data was collected from CysB which was purified to remove any bound nucleic acids (Section 2.7.1). As CysB is a DNA-binding protein, perhaps bound DNA is required to lock CysB into a more rigid conformation in order

to differentiate the DNA-binding domains with confidence between the two models. DAMAVER surface models can be found in Appendix AD.3.

Our SAXS data indicate there may be small differences between the CysB-*apo* and CysB-*NAS* states, as indicated by small changes in the Rg values. It is important to remember that proteins are dynamic in solution and SAXS is a low-resolution structural technique that can give valuable information about the behaviour of a protein in solution but is not able to inform atomic detail. Due to the lack of obvious differences between the CysB-*apo* and CysB-*NAS* SAXS plots and theoretical models, there is no evidence to support the occurrence of large-scale conformational changes upon inducer binding under these conditions. However, we hypothesise small changes within the protein upon inducer binding, but further scattering data should be collected to confirm this. At the time of submitting this thesis, we are preparing CysB samples for SAXS beamtime that will validate data presented here of CysB in an *apo-* state and with the inducer *N*-acetylserine bound. The data collected during this beamtime will include CysB bound to DNA to attempt to visualise any conformational changes within CysB upon DNA binding.

The EMSA results reported in Section 3.2.4 indicate that when the inducer *N*-acetylserine is bound, CysB does not release the *cysB* promoter as is observed in CysB from *E. coli*. Analysis of the X-ray crystallography structure (Section 3.3.4) revealed a host of amino acid substitutions including in positions important within the inducer-binding cavity and residues among the contact helices conserved across the LTTR family. This SAXS data suggests that CysB from *N. gonorrhoeae* does not undergo large scale conformational changes and supports the EMSA and X-ray crystallography findings that CysB may not respond to the inducer *N*-acetylserine. It may be entirely possible that bound DNA is required to visualise any large-scale conformational changes via SAXS or that CysB from *N. gonorrhoeae* simply does not undergo these large-scale changes reported amongst other members of the LTTR family and as such may have a different mechanism.

3.5   *in vivo* characterisation of CysB

3.5.1   Essentiality of CysB and proposed role *in vivo*

CysB is a transcriptional regulator that activates the cysteine regulon and in turn activates cysteine biosynthesis. The sulfate reduction pathway is essential and upregulated in *N. meningitidis* (Capel et al., 2016; Mendum et al., 2011) and the cysteine regulon has been shown to be upregulated among other pathogenic bacteria across a range of states. The absence

of the sulfate reduction pathway in *N. gonorrhoeae* results in an altered cysteine regulon. CysB has been identified as an essential gene in *N. gonorrhoeae* (Remmele et al., 2014) and is differentially expressed during infection (McClure et al., 2015). This raises interesting questions regarding the role of CysB in *N. gonorrhoeae*. To determine the role of CysB *in vivo* and whether CysB is indeed essential in *N. gonorrhoeae*, *cysB* gene deletion by homologous recombination in *N. gonorrhoeae* using a plate transformation was trialled.

3.5.2 *N. gonorrhoeae cysB* deletion construct for homologous recombination

A DNA construct was designed with Dr Joanna Hicks and ordered as from Twist Bioscience. The construct contained the *N. gonorrhoeae* kanamycin resistance gene (*kanR*) from (Ramsey, Hackett, Kotha, & Dillard, 2012) flanked by 150 base pairs of DNA sequence homologous to the DNA sequence upstream and downstream of *cysB* in the *N. gonorrhoeae* genome. The construct also included the *Neisseria* DNA uptake sequence (DUS) necessary for uptake of DNA (Figure 3.48).
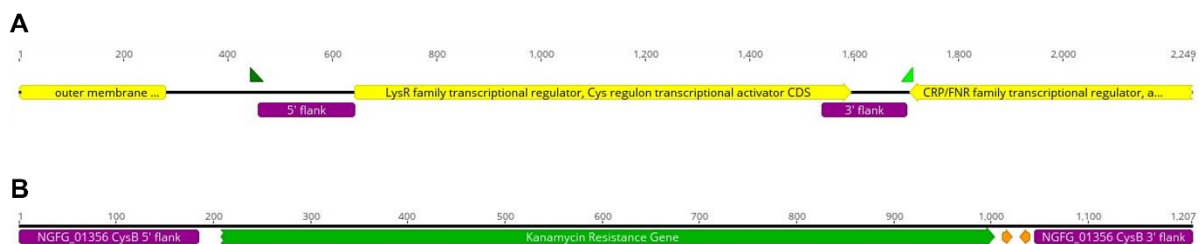
*Figure 3.48 DNA constructure of the N. gonorrhoeae cysB knockout. **A** Genomic context of the cysB gene (LysR family transcriptional regulator). The 150 bp homologous regions within the N. gonorrhoeae genome are shown in purple which were included in the knockout construct (**B**). Location of forward (dark green arrow) and reverse (light green arrow) primers are shown outside the homologous regions which were used to check for integration of the construct. **B** The cysB knockout construct with a kanamycin resistance gene in place of the cysB gene. The Neisseria DNA uptake sequences are shown as orange arrows. Figure created using Geneious Prime (Biomatter).*

Forward and reverse primers for the amplification of the entire DNA construct were also designed with Dr Joanna Hicks. Primer sequences can be found in Appendix AB.5. Amplification was performed via PCR (Section 2.5.5) and product purity was assessed via an agarose gel (Figure 3.49) as per Section 2.5.6. The gel shows a single distinct band at the expected size indicating successful amplification and purification of the *cysB* knockout construct.
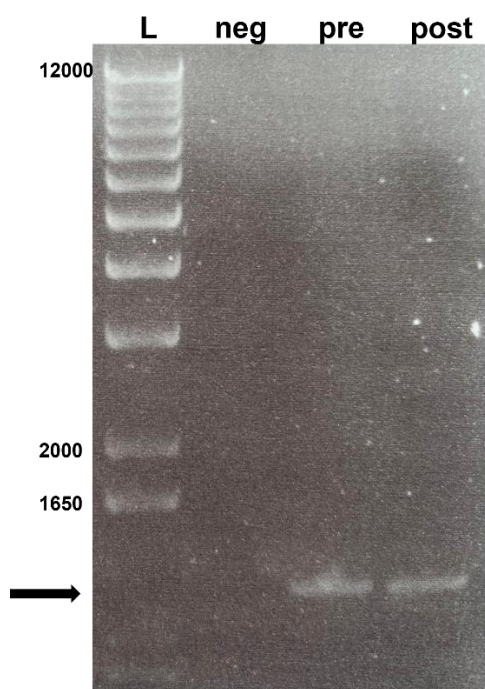
Figure 3.49 PCR amplification and purity results of the cysB knockout construct. The amplified and purified cysB knockout construct was run on a 1 % agarose gel. There are no bands in the negative control and a distinct single band (arrow) in the pre-purified and post-purified samples indicating successful amplification and purity of the cysB knockout PCR product.

3.5.3 Transformation of *N. gonorrhoeae*

Following preparation of the *cysB* knockout construct by PCR, two 10 μL amounts of PCR product (20 ng.μL^{-1}) were spotted onto GCB agar plates as per Section 2.8.2. Piliated colonies from a fresh GCB agar plate were selected and streaked through the DNA circles for transformation and incubated overnight. Single colonies that grew within the spotted DNA circles were streaked across GCB plates supplemented with 50 μg.mL^{-1} kanamycin and incubated as per Section 2.8.2 to select for positive transformants. The plates were checked after 48 hours of incubation and no growth was visible. This protocol was attempted on two occasions, both with no growth observed upon kanamycin selection. At the same time in our lab, other non-essential genes were successfully knocked out of *N. gonorrhoeae* using the same protocol which served as positive controls for our protocol, reagents and *N. gonorrhoeae* strains, confirming our negative results as true results.

The lack of growth on the kanamycin selection plates presents two possible conclusions. 1) that the *N. gonorrhoeae* was not successfully transformed and therefore lacked the kanamycin resistance gene to confer survival on the kanamycin plate. Or 2) that the transformation was successful and the *cysB* knockout construct conferring kanamycin resistance was integrated into the *N. gonorrhoeae* genome, however *cysB* is essential and therefore *N. gonorrhoeae* was not able to

121

grow or survive without the *cysB* gene. Both of these conclusions seemingly confirm the essentiality of *cysB* in *N. gonorrhoeae*.

To further understand the role of CysB in *N. gonorrhoeae* and confirm essentiality, we designed a knockdown *cysB* construct with a tetracycline inducible promoter (Figure 3.50). This would allow *cysB* expression to be controlled to understand the behaviour of *N. gonorrhoeae* under different conditions i.e. oxidative stress, and also investigate the genes regulated by CysB via transcriptomics. Due to issues with Twist Biosciences synthesising this construct, we have been unable to investigate the role of CysB via a knockdown strain in this work due to time constraints.

Figure 3.50 DNA constructure of the N. gonorrhoeae cysB knockdown. This construct was designed to integrate directly upstream of the cysB gene to replace the cysB promoter region with an inducible tetracycline promoter to control cysB expression. The 150bp homologous regions are shown in purple where the 3' homologous flank includes the cysB translational start codon ATG. In place of the cysB gene (replaced by the kanamycin resistance gene in the knockout construct (Figure 3.48)) is a kanamycin resistance gene (KanR) with the KanR promoter (light green arrow) and KanR terminator (red arrow). Tetracycline was used as an inducible promoter where the promoter is shown as a small light green arrow just upstream of the 3' 150 bp homologous flank for the cysB gene. The tetracycline repressor gene (tetR) was also included. Neisseria DNA uptake sequences are shown as two orange arrows upstream of the tetR gene. Figure created using Geneious Prime (Biomatter).

These results support the hypothesised essentiality of *cysB* in the *N. gonorrhoeae* genome (Remmele et al., 2014). The generation of a knockdown *cysB* strain was not possible during this work but will provide information on the exact role of CysB during infection as it has been identified to be downregulated during infection (McClure et al., 2015). A knockdown strain would allow investigation of the transcriptional activation of the known cysteine regulon targets assayed in this work (Section 3.2) during infection. It is also possible that transcriptomics with a *cysB* knockdown strain may allow for identification of other gene targets in *N. gonorrhoeae* which is not entirely far-fetched due to such a limited number of transcriptional regulators present in the *N. gonorrhoeae* genome.

Chapter 4: Conclusions and Future Work

Neisseria gonorrhoeae, as an obligate human pathogen, is the causative agent of the sexually transmitted infection gonorrhoea. Since treatment for gonococcal infection began in the 1930's, *N. gonorrhoeae* has rapidly acquired resistance to every class of antibiotic used for its treatment (Costa-Lourenço et al., 2017). The recent emergence of multi-drug resistance *N. gonorrhoeae* strains resistant to the current dual-therapy treatment highlights the rapid evolution of *N. gonorrhoeae* and the need for novel therapeutics to prevent extensively drug resistance strains of *N. gonorrhoeae* (Wi et al., 2017). Targeting amino acid synthesis is a novel and promising route for the development of new therapeutics and antimicrobials. Cysteine biosynthesis and sulfur assimilation pathways are absent in mammals and have garnered increasing interest as potential therapeutics for pathogenic bacteria.

N. gonorrhoeae encounters constant oxidative stress during pathogenesis. Colonisation of the urogenital tract, and hence the spread of *N. gonorrhoeae* to new hosts hinges on *N. gonorrhoeae*'s ability to survive oxidative stress from commensal *Lactobacilli spp* and the phagosome of neutrophils encountered in host immune defence (Hill et al., 2016; Seib et al., 2006). Mitigation of such stressors relies on reducing systems within *N. gonorrhoeae*. Given the necessity of L-cysteine-derived compounds including glutathione, antioxidants, iron-sulfur cluster proteins and the role of cysteine in protein folding and function, the cysteine biosynthetic pathway for *de novo* synthesis of L-cysteine is a promising target for developing new therapeutics for *N. gonorrhoeae*. However, *N. gonorrhoeae* has an altered cysteine regulon with a non-functional sulfate reduction pathway and a pseudogene sulfate binding protein which seemingly prevents the import and subsequent reduction of sulfate to sulfide (Hicks & Mullholland, 2018). The function of the remaining cysteine regulon genes and regulation of this altered cysteine regulon is therefore, not well understood. In this thesis we have biochemically and structurally characterised the cysteine regulon transcriptional regulator, CysB.

CysB is a LysR-type transcriptional regulator (LTTR) that regulates the cysteine regulon. CysB activates genes of the cysteine regulon and exhibits negative autoregulation of its own *cysB* promoter. This dual regulatory function of CysB is dependent on the binding of co-factors, termed inducers and anti-inducers. CysB reportedly bends the DNA of cysteine regulon

promoters to an angle of 100° which is relieved to 50° upon inducer binding (Hryniewicz & Kredich, 1991, 1994). Upon inducer binding, CysB releases entirely from the *cysB* promoter (Mittal et al., 2017). Large scale conformational changes are hypothesised in CysB upon inducer binding to relieve DNA bending and release CysB from the *cysB* promoter, although no such structural evidence has yet been reported to support this.

In this thesis, we have characterised the DNA-binding activity of CysB to the promoter regions of the *cysB, cysK, sbp, cysU, cysW* and *cysA* genes. The promoter region upstream of the *cysB* gene contains two inverted repeat sequences that appear to be involved in CysB binding. Our results suggest that while these inverted repeat sequences may not be directly involved in CysB binding, the distance of these sequences relative to one another is critical for CysB binding. Our results also indicate the presence of a sequence important in CysB binding downstream of these inverted repeats, within the 80 bp upstream of the translation start codon ATG. Interestingly, the addition of inducer *N*-acetylserine to CysB binding assays did not release CysB from the *cysB* promoter as expected (Mittal et al., 2017). This suggests that CysB may not respond to the inducer *N*-acetylserine.

We have characterised the DNA-binding activity of CysB to the promoter regions of other cysteine regulon genes *cysK, sbp, cysU, cysW* and *cysA*. In *N. gonorrhoeae* the *cysU, cysW* and *cysA* genes are separate genes where in other bacteria, such as *E. coli*, they are part of the *cysPUWAM* operon. Assayed binding to the *cysK, sbp, cysU, cysW* and *cysA* promoters showed that CysB bound to each of these promoter regions. This supports the single gene arrangement of the *cysU, cysW* and *cysA* genes in *N. gonorrhoeae* and suggests CysB remains the transcriptional regulator of these genes. The *sbp* gene in *N. gonorrhoeae* is a pseudogene due to a 5' truncation. As the ABC transporter consisting of the *sbp, cysU, cysW* and *cysA* proteins is the only sulfur transport system present in *N. gonorrhoeae* (Hicks & Mullholland, 2018), the identification of *sbp* as a pseudogene leads to interesting questions surrounding the function and CysB-mediated regulation of the remaining *cysU, cysW* and *cysA* genes. Due to a non-functional sulfate reduction pathway and *sbp* as a pseudogene, coupled with *N. gonorrhoeae*'s ability to grow on thiosulfate as a sole source of sulfur, there may be other sulfur import and/or reduction mechanisms within *N. gonorrhoeae*. This may explain the CysB-mediated regulation of the separate *cysU, cysW* and *cysA* genes. CysB remained bound to these promoter regions when the inducer *N*-acetylserine is present at 10 mM concentrations. This suggests CysB

remains bound to these promoters and may exhibit DNA bending, although future work is required to inform this.

We have solved the structure of CysB from *N. gonorrhoeae* to 2.73 Å resolution which is the first full-length tetrameric CysB structure solved. This structure provides the first understanding of the tetrameric arrangement of CysB and provides insights into the arrangement of the DNA-binding domains relative to one another and appear to support the observed convergently arranged half sites in the *cysP, cysK* and *cysJ* promoters from *S. typhimurium* (Hryniewicz & Kredich, 1995). Interestingly, the peptide sequences of CysB from *N. gonorrhoeae* and *Neisseria meningitidis* show a host of amino acid substitutions including among highly conserved residues in the inducer-binding cavity and residues reported to effect DNA-binding activity. Within the inducer-binding cavity the highly conserved Thr202 which binds the anti-inducer sulfate, has been lost due to an insertion and mutation to Gln203. The loop containing this residue is also positioned away from the inducer-binding cavity. If CysB from *N. gonorrhoeae* does not respond to the inducer *N*-acetylserine, this may explain the loss of this conserved residue and alternate conformation of the loop. CysB from *N. gonorrhoeae* and *N. meningitidis* both have truncated C-terminal arms which are hypothesised to be involved in anchoring the secondary inducer *O*-acetylserine to the secondary inducer binding site.

A particularly interesting aspect to the CysB structure is the contact helix interaction and loss of conserved residues in this region. The contact helix has been implicated in the large scale conformation changes that LTTRs reportedly experience upon inducer binding (Lerche et al., 2016; Monferrer et al., 2010). Analysis of the CysB structure reported here and the CysB peptide sequence show that the conserved glycine and prolines residues surrounding the contact helix have been lost in CysB from *N. gonorrhoeae*. These glycine and proline residues are hypothesised to afford conformational changes in the regulatory domain upon inducer binding (Monferrer et al., 2010). The absence of these conserved residues in CysB from *N. gonorrhoeae* supports the EMSA results where CysB does not respond to *N*-acetylserine when bound to the *cysB* promoter. This is further supported by the lack of observed large scale conformational changes in our SAXS data with the inducer *N*-acetylserine bound.

Based on our findings reported here, we hypothesise that due to amino acid substitutions among highly conserved residues in the contact helices, CysB from *N. gonorrhoeae* does not experience large scale conformational changes upon inducer binding. Our EMSA assays

suggest that CysB may also not respond to the inducer *N*-acetylserine. Future work to further investigate this hypothesis will include SAXS with DNA bound to investigate whether DNA is required to observe any conformational changes and DNA bending experiments to investigate whether the presence of *N*-acetylserine relieves bending of the cysteine regulon promoters (Hryniewicz & Kredich, 1991, 1994). Future work will also investigate the presence of CysB binding sites and half sites within the promoter regions of *cysK, cysU, cysW* and *cysA* from *N. gonorrhoeae* which will indicate whether there are additional half sites that CysB may or may not bind between to cause DNA bending. Upon these findings, I investigated the large scale conformational changes reported among other LTTRs. While the literature constantly alludes to these proteins undergoing large scale conformational changes upon inducer binding and the involvement of LTTRs in DNA bending, structures to date show only one LTTR member that has been shown to adopt both the compact *apo* form and the holo-spherical inducer-bound form, determined via SAXS (Lerche et al., 2016). Other LTTR X-ray structures solved without and with inducers do not display any large scale conformational changes. This seemingly supports the lack of observed large scale conformational changes in CysB from *N. gonorrhoeae*.

To confirm that CysB cannot undergo these large scale conformation changes upon inducer binding, crystallisation attempts are ongoing with the inducer *N*-acetylserine. Future work will also investigate crystallisation of CysB with the secondary inducer *O*-acetylserine, with free cysteine to probe any conformational changes involving the solvent accessible cysteine residues, under both reducing and oxidising conditions, as well as crystallisation attempts with DNA containing the known CysB binding sites. This will therefore require elucidation of the nucleotides critical for CysB binding. Identification of these critical nucleotides may allow identification of novel CysB targets within the *N. gonorrhoeae* genome. As *cysB* is essential, the genes it regulates may have critical functions within *N. gonorrhoeae* and may be targets for future development of new antimicrobials. The generation of a *cysB* knockdown strain will allow *in vivo* investigation, such as transcriptomics of *N. gonorrhoeae* during stage of infection and under oxidative stress, to determine the relationship between CysB and the cysteine regulon genes. Such *in vivo* work also has the potential to uncover previously unknown targets within *N. gonorrhoeae* that may undergo CysB-mediated regulation.

N. meningitidis is the leading cause of bacterial meningitis, with Aotearoa, New Zealand having some of the highest rates globally (Lopez & Sexton, 2013). There is high sequence

similarity (99.4 %) between CysB from *N. gonorrhoeae* and *N. meningitidis* with only two amino acid substitutions between them. As *N. meningitidis* has a functional sulfate reduction pathway, but contains the same amino acid substitutions resulting in the loss of highly conserved residues within the contact helices, investigation of CysB from *N. meningitidis* would be particularly interesting and may confirm our hypothesised lack of CysB conformational changes due to the lack of movement around the contact helices. It would also be particularly interesting to investigate whether CysB from *N. meningitidis* responds to the inducer *N*-acetylserine.

The results presented in this thesis represent a major leap forward in our understanding of transcriptional regulation of cysteine biosynthesis and the cysteine regulon in *N. gonorrhoeae*. These results are also the basis for future work to understand the mechanism of CysB from *N. gonorrhoeae* and further our understanding of regulation within *N. gonorrhoeae*. Overall, this research has contributed to a greater understanding of a previously uncharacterised essential transcriptional regulator within *N. gonorrhoeae* and may pave the way for the identification of novel targets and hence development of novel therapeutics to combat antimicrobial resistance in *N. gonorrhoeae*.

References

Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., . . . Grosse-Kunstleve, R. W. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D: Biological Crystallography, 66*(2), 213-221.

Adamson, P. C., & Klausner, J. D. (2021). The Staying Power of Pharyngeal Gonorrhea: Implications for Public Health and Antimicrobial Resistance. *Clinical Infectious Diseases, 73*(4), 583-585. doi:10.1093/cid/ciab074

Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., . . . Adams, P. D. (2012). Towards automated crystallographic structure refinement with phenix. refine. *Acta Crystallographica Section D: Biological Crystallography, 68*(4), 352-367.

Agarwal, S. M., Jain, R., Bhattacharya, A., & Azam, A. (2008). Inhibitors of Escherichia coli serine acetyltransferase block proliferation of Entamoeba histolytica trophozoites. *International journal for parasitology, 38*(2), 137-141.

Ayala, J. C., & Shafer, W. M. (2019). Transcriptional regulation of a gonococcal gene encoding a virulence factor (L-lactate permease). *PLoS pathogens, 15*(12), e1008233.

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

Benoni, R., Pertinhez, T. A., Spyrakis, F., Davalli, S., Pellegrino, S., Paredi, G., . . . Mozzarelli, A. (2016). Structural insight into the interaction of O-acetylserine sulfhydrylase with competitive, peptidic inhibitors by saturation transfer difference-NMR. *FEBS letters, 590*(7), 943-953.

Brennan, R. G., & Matthews, B. W. (1989). The helix-turn-helix DNA binding motif. *Journal of Biological Chemistry, 264*(4), 1903-1906. doi:https://doi.org/10.1016/S0021-9258(18)94115-3

Campanini, B., Pieroni, M., Raboni, S., Bettati, S., Benoni, R., Pecchini, C., . . . Mozzarelli, A. (2015). Inhibitors of the sulfur assimilation pathway in bacterial pathogens as enhancers of antibiotic therapy. *Current medicinal chemistry, 22*(2), 187-213.

Capel, E., Zomer, A. L., Nussbaumer, T., Bole, C., Izac, B., Frapy, E., . . . Jamet, A. (2016). Comprehensive identification of meningococcal genes and small noncoding RNAs required for host cell colonization. *MBio, 7*(4), e01173-01116.

Chen, C., Yan, Q., Tao, M., Shi, H., Han, X., Jia, L., . . . Ma, X. (2019). Characterization of serine acetyltransferase (CysE) from methicillin-resistant Staphylococcus aureus and inhibitory effect of two natural products on CysE. *Microbial pathogenesis, 131*, 218-226.

Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., . . . Richardson, D. C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D: Biological Crystallography, 66*(1), 12-21.

Choi, H.-J., Kim, S.-J., Mukhopadhyay, P., Cho, S., Woo, J.-R., Storz, G., & Ryu, S.-E. (2001). Structural basis of the redox switch in the OxyR transcription factor. *Cell, 105*(1), 103-113.

Clark, T., Haddad, S., Neidle, E., & Momany, C. (2004). Crystallization of the effector-binding domains of BenM and CatM, LysR-type transcriptional regulators from Acinetobacter sp. ADP1. *Acta Crystallographica Section D: Biological Crystallography, 60*(1), 105-108.

Colyer, T. E., & Kredich, N. M. (1996). In vitro characterization of constitutive CysB proteins from Salmonella typhimurium. *Molecular microbiology, 21*(2), 247-256.

Costa-Lourenço, A. P. R. d., Santos, K. T. B. d., Moreira, B. M., Fracalanzza, S. E. L., & Bonelli, R. R. (2017). Antimicrobial resistance in Neisseria gonorrhoeae: history, molecular mechanisms and epidemiological aspects of an emerging global threat. *brazilian journal of microbiology, 48*, 617-628.

Criss, A. K., Kline, K. A., & Seifert, H. S. (2005). The frequency and rate of pilin antigenic variation in Neisseria gonorrhoeae. *Molecular microbiology, 58*(2), 510-519.

De Reuse, H., & Taha, M.-K. (1997). RegF, an SspA homologue, regulates the expression of the Neisseria gonorrhoeae pilE gene. *Research in Microbiology, 148*(4), 289-303.

Dong, H. V., & Klausner, J. D. (2019). Neisseria gonorrhoeae resistance driven by antibiotic use. *Nature Reviews Urology, 16*(9), 509-510.

Edwards, J. L., & Apicella, M. A. (2004). The molecular mechanisms used by Neisseria gonorrhoeae to initiate infection differ between men and women. *Clinical microbiology reviews, 17*(4), 965-981.

El-Benna, J., Hurtado-Nedelec, M., Marzaioli, V., Marie, J. C., Gougerot-Pocidalo, M. A., & Dang, P. M. C. (2016). Priming of the neutrophil respiratory burst: role in host defense and inflammation. *Immunological reviews, 273*(1), 180-193.

Emsley, P., & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D: Biological Crystallography, 60*(12), 2126-2132.

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

ESR. (2015). *Sexually Transmitted Infections in New Zealand: Annual Surveillance Report 2014*. Retrieved from Porirua, New Zealand:

https://surv.esr.cri.nz/PDF_surveillance/STISurvRpt/2014/FINAL2014AnnualSTIRep ort.pdf

ESR. (2021). *Antimicrobial resistance and molecular epidemiology of Neisseria gonorrhoeae in New Zealand, 2018-2019*. Retrieved from Porirua: New Zealand: https://surv.esr.cri.nz/PDF_surveillance/Antimicrobial/Gono/NgonoSurvey2019_FIN AL.pdf

ESR. (2022). *Sexually Transmitted Infections in New Zealand: Annual Surveillance Report 2017/2018/2019*. Retrieved from Porirua, New Zealand: https://surv.esr.cri.nz/PDF_surveillance/STISurvRpt/2017/FINALSTIANNUALREP ORT17_18_1930032022.pdf

Evans, P. R., & Murshudov, G. N. (2013). How good are my data and what is the resolution? *Acta Crystallographica Section D: Biological Crystallography, 69*(7), 1204-1214.

Eyre, D. W., Sanderson, N. D., Lord, E., Regisford-Reimmer, N., Chau, K., Barker, L., . . . Andersson, M. I. (2018). Gonorrhoea treatment failure caused by a Neisseria gonorrhoeae strain with combined ceftriaxone and high-level azithromycin resistance, England, February 2018. *Euro Surveill, 23*(27). doi:10.2807/1560-7917.Es.2018.23.27.1800323

Folster, J. P., Dhulipala, V., Nicholas, R. A., & Shafer, W. M. (2007). Differential regulation of ponA and pilMNOPQ expression by the MtrR transcriptional regulatory protein in Neisseria gonorrhoeae. *Journal of Bacteriology, 189*(13), 4569-4577.

Fontán, P., Aris, V., Ghanny, S., Soteropoulos, P., & Smith, I. (2008). Global transcriptional profile of Mycobacterium tuberculosis during THP-1 human macrophage infection. *Infection and immunity, 76*(2), 717-725.

Franke, D., Petoukhov, M., Konarev, P., Panjkovich, A., Tuukkanen, A., Mertens, H., . . . Jeffries, C. (2017). ATSAS 2.8: a comprehensive data analysis suite for small-angle

scattering from macromolecular solutions. *Journal of applied crystallography, 50*(4), 1212-1225.

Franke, D., & Svergun, D. I. (2009). DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *Journal of applied crystallography, 42*(2), 342-346.

Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). Protein identification and analysis tools on the ExPASy server. In *The proteomics protocols handbook* (pp. 571-607): Springer.

Giannopoulou, E. A., Senda, M., Koentjoro, M. P., Adachi, N., Ogawa, N., & Senda, T. (2021). Crystal structure of the full-length LysR-type transcription regulator CbnR in complex with promoter DNA. *The FEBS Journal, 288*(15), 4560-4575.

Hagman, K. E., & Shafer, W. M. (1995). Transcriptional control of the mtr efflux system of Neisseria gonorrhoeae. *Journal of Bacteriology, 177*(14), 4162-4165.

Henikoff, S., Haughn, G. W., Calvo, J. M., & Wallace, J. C. (1988). A large family of bacterial activator proteins. *Proceedings of the National Academy of Sciences, 85*(18), 6602-6606.

Hicks, J. L., & Mullholland, C. V. (2018). Cysteine biosynthesis in Neisseria species. *Microbiology, 164*(12), 1471-1480.

Hill, S. A., Masters, T. L., & Wachter, J. (2016). Gonorrhea-an evolving disease of the new millennium. *Microbial cell, 3*(9), 371.

Holley, C. L., Ayala, J. C., & Shafer, W. M. (2020). Transcriptional control of the gonococcal ompA gene by the MisR/MisS two-component regulatory system. *Scientific reports, 10*(1), 1-10.

Hryniewicz, M. M., & Kredich, N. M. (1991). The cysP promoter of Salmonella typhimurium: characterization of two binding sites for CysB protein, studies of in vivo transcription

initiation, and demonstration of the anti-inducer effects of thiosulfate. *Journal of Bacteriology, 173*(18), 5876-5886.

Hryniewicz, M. M., & Kredich, N. M. (1994). Stoichiometry of binding of CysB to the cysJIH, cysK, and cysP promoter regions of Salmonella typhimurium. *Journal of Bacteriology, 176*(12), 3673-3682.

Hryniewicz, M. M., & Kredich, N. M. (1995). Hydroxyl radical footprints and half-site arrangements of binding sites for the CysB transcriptional activator of Salmonella typhimurium. *Journal of Bacteriology, 177*(9), 2343-2353.

Jackson, L. A., Ducey, T. F., Day, M. W., Zaitshik, J. B., Orvis, J., & Dyer, D. W. (2010). Transcriptional and functional analysis of the Neisseria gonorrhoeae Fur regulon. *Journal of Bacteriology, 192*(1), 77-85.

Jagura-Burdzy, G., & Hulanicka, D. (1981). Use of gene fusions to study expression of cysB, the regulatory gene of the cysteine regulon. *Journal of Bacteriology, 147*(3), 744-751.

Jeelani, G., Sato, D., Soga, T., & Nozaki, T. (2017). Genetic, metabolomic and transcriptomic analyses of the de novo L-cysteine biosynthetic pathway in the enteric protozoan parasite Entamoeba histolytica. *Scientific reports, 7*(1), 1-15.

Jeffrey, G. A. (1997). *An introduction to hydrogen bonding* (Vol. 12): Oxford university press New York.

Jones-Mortimer, M., Wheldrake, J., & Pasternak, C. (1968). The control of sulphate reduction in Escherichia coli by O-acetyl-L-serine. *Biochemical journal, 107*(1), 51-53.

Kabsch, W. (2010). Integration, scaling, space-group assignment and post-refinement. *Acta Crystallographica Section D: Biological Crystallography, 66*(2), 133-144.

Kertesz, M. A. (2001). Bacterial transporters for sulfate and organosulfur compounds. *Research in Microbiology, 152*(3-4), 279-290.

Kidd, S. P., Potter, A. J., Apicella, M. A., Jennings, M. P., & McEwan, A. G. (2005). NmLR of Neisseria gonorrhoeae: a novel redox responsive transcription factor from the MerR family. *Molecular microbiology, 57*(6), 1676-1689.

Kim, J.-S., & Holmes, R. K. (2012). Characterization of OxyR as a Negative Transcriptional Regulator That Represses Catalase Production in Corynebacterium diphtheriae. *PLOS ONE, 7*(3), e31709. doi:10.1371/journal.pone.0031709

Kirby, N., Cowieson, N., Hawley, A. M., Mudie, S. T., McGillivray, D. J., Kusel, M., . . . Ryan, T. M. (2016). Improved radiation dose efficiency in solution SAXS using a sheath flow sample environment. *Acta Crystallographica Section D: Structural Biology, 72*(12), 1254-1266.

Kirkcaldy, R. D., Weston, E., Segurado, A. C., & Hughes, G. (2019). Epidemiology of gonorrhoea: a global perspective. *Sexual health, 16*(5), 401-411.

Konarev, P. V., Petoukhov, M. V., & Svergun, D. I. (2016). Rapid automated superposition of shapes and macromolecular models using spherical harmonics. *Journal of applied crystallography, 49*(3), 953-960.

Kowiel, M., Brzezinski, D., Porebski, P. J., Shabalin, I. G., Jaskolski, M., & Minor, W. (2019). Automatic recognition of ligands in electron density by machine learning. *Bioinformatics, 35*(3), 452-461.

Kozin, M. B., & Svergun, D. I. (2001). Automated matching of high-and low-resolution structural models. *Journal of applied crystallography, 34*(1), 33-41.

Kredich, N. M. (1971). Regulation of L-cysteine biosynthesis in Salmonella typhimurium: I. Effects of growth on varying sulfur sources and O-acetyl-L-serine on gene expression. *Journal of Biological Chemistry, 246*(11), 3474-3484.

Kredich, N. M. (1992). The molecular basis for positive regulation of cys promoters in Salmonella typhimurium and Escherichia coli. *Molecular microbiology, 6*(19), 2747-2753.

Kredich, N. M. (2008). Biosynthesis of cysteine. *EcoSal Plus, 3*(1).

Kredich, N. M., Becker, M. A., & Tomkins, G. M. (1969). Purification and characterization of cysteine synthetase, a bifunctional protein complex, from Salmonella typhimurium. *Journal of Biological Chemistry, 244*(9), 2428-2439.

Krissinel, E., & Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. *Journal of Molecular Biology, 372*(3), 774-797.

Le Faou, A. (1984). *Sulphur nutrition and metabolism in various species of Neisseria.* Paper presented at the Annales de l'Institut Pasteur/Microbiologie.

Lebedev, A. A., & Isupov, M. N. (2014). Space-group and origin ambiguity in macromolecular structures with pseudo-symmetry and its treatment with the program Zanuda. *Acta Crystallographica Section D: Biological Crystallography, 70*(9), 2430-2443.

Lee, E. H., & Shafer, W. M. (1999). The farAB-encoded efflux pump mediates resistance of gonococci to long-chained antibacterial fatty acids. *Molecular microbiology, 33*(4), 839-845.

Lee, R. S., Seemann, T., Heffernan, H., Kwong, J. C., Gonçalves da Silva, A., Carter, G. P., . . . Stinear, T. P. (2018). Genomic epidemiology and antimicrobial resistance of Neisseria gonorrhoeae in New Zealand. *Journal of Antimicrobial Chemotherapy, 73*(2), 353-364.

Lensmire, J. M., & Hammer, N. D. (2019). Nutrient sulfur acquisition strategies employed by bacterial pathogens. *Current opinion in microbiology, 47*, 52-58.

Lerche, M., Dian, C., Round, A., Lönneborg, R., Brzezinski, P., & Leonard, G. A. (2016). The solution configurations of inactive and activated DntR have implications for the sliding dimer mechanism of LysR transcription factors. *Scientific reports, 6*(1), 1-14.

Liebschner, D., Afonine, P. V., Baker, M. L., Bunkóczi, G., Chen, V. B., Croll, T. I., . . . McCoy, A. J. (2019). Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallographica Section D: Structural Biology, 75*(10), 861-877.

Lochowska, A., Iwanicka-Nowicka, R., Plochocka, D., & Hryniewicz, M. M. (2001). Functional Dissection of the LysR-type CysB Transcriptional Regulator: Regions important for DNA binding, inducer response, oligomerization, and positive control. *Journal of Biological Chemistry, 276*(3), 2098-2107. doi:https://doi.org/10.1074/jbc.M007192200

Lochowska, A., Iwanicka-Nowicka, R., Zaim, J., Witkowska-Zimny, M., Bolewska, K., & Hryniewicz, M. M. (2004). Identification of activating region (AR) of Escherichia coli LysR-type transcription factor CysB and CysB contact site on RNA polymerase alpha subunit at the cysP promoter. *Molecular microbiology, 53*(3), 791-806.

Lopez, L., & Sexton, K. (2013). *The Epidemiology of Meningococcal Disease in New Zealand in 2012, 2013*. Retrieved from Wellington, New Zealand:

Lynch, A., Tyrrell, R., Smerdon, S., Briggs, G., & Wilkinson, A. (1994). Characterization of the CysB protein of Klebsiella aerogenes: direct evidence that N-acetylserine rather than O-acetylserine serves as the inducer of the cysteine regulon. *Biochemical journal, 299*(1), 129-136.

Maddocks, S. E., & Oyston, P. C. F. (2008). Structure and function of the LysR-type transcriptional regulator (LTTR) family proteins. *Microbiology, 154*(12), 3609-3623. doi:https://doi.org/10.1099/mic.0.2008/022772-0

Magalhães, J., Franko, N., Raboni, S., Annunziato, G., Tammela, P. i., Bruno, A., . . . Campanini, B. (2020). Inhibition of nonessential bacterial targets: discovery of a novel serine O-acetyltransferase inhibitor. *ACS medicinal chemistry letters, 11*(5), 790-797.

136

Manalastas-Cantos, K., Konarev, P. V., Hajizadeh, N. R., Kikhney, A. G., Petoukhov, M. V., Molodenskiy, D. S., . . . Borges, C. (2021). ATSAS 3.0: expanded functionality and new tools for small-angle scattering data analysis. *Journal of applied crystallography, 54*(1), 343-355.

Masi, A. T., & Eisenstein, B. I. (1981). Disseminated gonococcal infection (DGI) and gonococcal arthritis (GCA): II. Clinical manifestations, diagnosis, complications, treatment, and prevention. *Seminars in Arthritis and Rheumatism, 10*(3), 173-197. doi:https://doi.org/10.1016/S0049-0172(81)80002-9

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

Matthias, K. A., & Rest, R. F. (2014). Control of pili and sialyltransferase expression in N eisseria gonorrhoeae is mediated by the transcriptional regulator CrgA. *Molecular microbiology, 91*(6), 1120-1135.

McClure, R., Nudel, K., Massari, P., Tjaden, B., Su, X., Rice, P. A., & Genco, C. A. (2015). The Gonococcal Transcriptome during Infection of the Lower Genital Tract in Women. *PLOS ONE, 10*(8), e0133982. doi:10.1371/journal.pone.0133982

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., & Read, R. J. (2007). Phaser crystallographic software. *Journal of applied crystallography, 40*(4), 658-674.

McPhillips, T. M., McPhillips, S. E., Chiu, H.-J., Cohen, A. E., Deacon, A. M., Ellis, P. J., . . . Phizackerley, R. P. (2002). Blu-Ice and the Distributed Control System: software for data acquisition and instrument control at macromolecular crystallography Beamlines. *Journal of synchrotron radiation, 9*(6), 401-406.

Mendum, T. A., Newcombe, J., Mannan, A. A., Kierzek, A. M., & McFadden, J. (2011). Interrogation of global mutagenesis data with a genome scale model of Neisseria meningitidis to assess gene fitness in vitro and in sera. *Genome biology, 12*(12), 1-15.

Miller, B. E., & Kredich, N. M. (1987). Purification of the cysB protein from Salmonella typhimurium. *Journal of Biological Chemistry, 262*(13), 6006-6009.

Mino, K., & Ishikawa, K. (2003). A novel O-phospho-L-serine sulfhydrylation reaction catalyzed by O-acetylserine sulfhydrylase from Aeropyrum pernix K1. *FEBS letters, 551*(1-3), 133-138.

Mittal, M., Singh, A. K., & Kumaran, S. (2017). Structural and biochemical characterization of ligand recognition by CysB, the master regulator of sulfate metabolism. *Biochimie, 142*, 112-124. doi:10.1016/j.biochi.2017.08.011

Monferrer, D., Tralau, T., Kertesz, M. A., Dix, I., Solà, M., & Usón, I. (2010). Structural studies on the full-length LysR-type regulator TsaR from Comamonas testosteroni T-2 reveal a novel open conformation of the tetrameric LTTR fold. *Molecular microbiology, 75*(5), 1199-1214.

Mongkolsuk, S., & Helmann, J. D. (2002). Regulation of inducible peroxide stress responses. *Molecular microbiology, 45*(1), 9-15.

Muraoka, S., Okumura, R., Ogawa, N., Nonaka, T., Miyashita, K., & Senda, T. (2003). Crystal structure of a full-length LysR-type transcriptional regulator, CbnR: unusual combination of two subunit forms and molecular bases for causing and changing DNA bend. *Journal of Molecular Biology, 328*(3), 555-566.

Ochman, H., Lawrence, J. G., & Groisman, E. A. (2000). Lateral gene transfer and the nature of bacterial innovation. *Nature, 405*(6784), 299.

Oldham, K. E., Prentice, E. J., Summers, E. L., & Hicks, J. L. (2022). Serine acetyltransferase from Neisseria gonorrhoeae; structural and biochemical basis of inhibition. *Biochemical journal, 479*(1), 57-74.

Ostrowski, J., & Kredich, N. M. (1989). Molecular characterization of the cysJIH promoters of Salmonella typhimurium and Escherichia coli: regulation by cysB protein and N-acetyl-L-serine. *Journal of Bacteriology, 171*(1), 130-140.

Ostrowski, J., & Kredich, N. M. (1990). In vitro interactions of CysB protein with the cysJIH promoter of Salmonella typhimurium: inhibitory effects of sulfide. *Journal of Bacteriology, 172*(2), 779-785.

Ostrowski, J., & Kredich, N. M. (1991). Negative autoregulation of cysB in Salmonella typhimurium: in vitro interactions of CysB protein with the cysB promoter. *Journal of Bacteriology, 173*(7), 2212-2218.

Overton, T. W., Whitehead, R., Li, Y., Snyder, L. A., Saunders, N. J., Smith, H., & Cole, J. A. (2006). Coordinated regulation of the Neisseria gonorrhoeae-truncated denitrification pathway by the nitric oxide-sensitive repressor, NsrR, and nitrite-insensitive NarQ-NarP. *Journal of Biological Chemistry, 281*(44), 33115-33126.

Pabo, C. O., & Sauer, R. T. (1984). Protein-DNA recognition. *Annual review of biochemistry, 53*(1), 293-321.

Panjkovich, A., & Svergun, D. I. (2018). CHROMIXS: automatic and interactive analysis of chromatography-coupled small-angle X-ray scattering data. *Bioinformatics, 34*(11), 1944-1946.

Pérez-Rueda, E., & Collado-Vides, J. (2001). Common history at the origin of the position–function correlation in transcriptional regulators in archaea and bacteria. *Journal of molecular evolution, 53*(3), 172-179.

Quillin, S. J., & Seifert, H. S. (2018). Neisseria gonorrhoeae host adaptation and pathogenesis. *Nature Reviews Microbiology, 16*(4), 226.

Ramsey, M. E., Hackett, K. T., Kotha, C., & Dillard, J. P. (2012). New complementation constructs for inducible and constitutive gene expression in Neisseria gonorrhoeae and Neisseria meningitidis. *Applied and environmental microbiology, 78*(9), 3068-3078.

Remmele, C. W., Xian, Y., Albrecht, M., Faulstich, M., Fraunholz, M., Heinrichs, E., . . . Rudel, T. (2014). Transcriptional landscape and essential genes of Neisseria gonorrhoeae. *Nucleic acids research, 42*(16), 10579-10595. doi:10.1093/nar/gku762

Robert, X., & Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. *Nucleic acids research, 42*(W1), W320-W324.

Rochelle P. Walensky, D. H., Daniel B. Jernigan, Rebecca Bunnell, Jennifer Layden, Michael F. Iademarco. (2021). Sexually Transmitted Infections Treatment Guidelines, 2021. *CDC MMWR Recomm Rep, 70*(4).

Ryan, T. M., Trewhella, J., Murphy, J. M., Keown, J. R., Casey, L., Pearce, F. G., . . . Kobe, B. (2018). An optimized SEC-SAXS system enabling high X-ray dose for rapid SAXS assessment with correlated UV measurements for biomolecular structure analysis. *Journal of applied crystallography, 51*(1), 97-111.

Sainsbury, S., Lane, L. A., Ren, J., Gilbert, R. J., Saunders, N. J., Robinson, C. V., . . . Owens, R. J. (2009). The structure of CrgA from Neisseria meningitidis reveals a new octameric assembly state for LysR transcriptional regulators. *Nucleic acids research, 37*(14), 4545-4558. doi:10.1093/nar/gkp445

Santi-Rocca, J., Smith, S., Weber, C., Pineda, E., Hon, C.-C., Saavedra, E., . . . Coppée, J.-Y. (2012). Endoplasmic reticulum stress-sensing mechanism is activated in Entamoeba histolytica upon treatment with nitric oxide. *PLOS ONE, 7*(2), e31777.

Schell, M. A. (1993). Molecular biology of the LysR family of transcriptional regulators. *Annual review of microbiology, 47*(1), 597-626.

Schielke, S., Frosch, M., & Kurzai, O. (2010). Virulence determinants involved in differential host niche adaptation of Neisseria meningitidis and Neisseria gonorrhoeae. *Med Microbiol Immunol, 199*(3), 185-196. doi:10.1007/s00430-010-0150-5

Schnappinger, D., Ehrt, S., Voskuil, M. I., Liu, Y., Mangan, J. A., Monahan, I. M., . . . Nathan, C. (2003). Transcriptional adaptation of Mycobacterium tuberculosis within macrophages: insights into the phagosomal environment. *The Journal of experimental medicine, 198*(5), 693-704.

Schnell, R., Sriram, D., & Schneider, G. (2015). Pyridoxal-phosphate dependent mycobacterial cysteine synthases: Structure, mechanism and potential as drug targets. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics, 1854*(9), 1175-1183.

Schook, P. O. P., Stohl, E. A., Criss, A. K., & Seifert, H. S. (2011). The DNA-binding activity of the Neisseria gonorrhoeae LexA orthologue NG1427 is modulated by oxidation. *Molecular microbiology, 79*(4), 846-860. doi:https://doi.org/10.1111/j.1365-2958.2010.07491.x

Seib, K. L., Wu, H.-J., Kidd, S. P., Apicella, M. A., Jennings, M. P., & McEwan, A. G. (2006). Defenses against oxidative stress in Neisseria gonorrhoeae: a system tailored for a challenging environment. *Microbiol. Mol. Biol. Rev., 70*(2), 344-361.

Seib, K. L., Wu, H. J., Srikhanta, Y. N., Edwards, J. L., Falsetta, M. L., Hamilton, A. J., . . . McEwan, A. G. (2007). Characterization of the OxyR regulon of Neisseria gonorrhoeae. *Molecular microbiology, 63*(1), 54-68.

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., . . . Söding, J. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology, 7*(1), 539.

Simons, M. P., Nauseef, W. M., & Apicella, M. A. (2005). Interactions of Neisseria gonorrhoeae with adherent polymorphonuclear leukocytes. *Infection and immunity, 73*(4), 1971-1977.

Smirnova, I. A., Dian, C., Leonard, G. A., McSweeney, S., Birse, D., & Brzezinski, P. (2004). Development of a bacterial biosensor for nitrotoluenes: the crystal structure of the transcriptional regulator DntR. *Journal of Molecular Biology, 340*(3), 405-418.

Song, Y., Yang, C., Chen, G., Zhang, Y., Seng, Z., Cai, Z., . . . Liang, H. (2019). Molecular insights into the master regulator CysB-mediated bacterial virulence in Pseudomonas aeruginosa. *Molecular microbiology, 111*(5), 1195-1210. doi:10.1111/mmi.14200

Spyrakis, F., Singh, R., Cozzini, P., Campanini, B., Salsi, E., Felici, P., . . . Kellogg, G. E. (2013). Isozyme-specific ligands for O-acetylserine sulfhydrylase, a novel antibiotic target. *PLOS ONE, 8*(10), e77558.

Stec, E., Witkowska-Zimny, M., Hryniewicz, M. M., Neumann, P., Wilkinson, A. J., Brzozowski, A. M., . . . Bujacz, G. D. (2006). Structural basis of the sulphate starvation response in E. coli: crystal structure and mutational analysis of the cofactor-binding domain of the Cbl transcriptional regulator. *Journal of Molecular Biology, 364*(3), 309-322.

Stern, A., Brown, M., Nickel, P., & Meyer, T. F. (1986). Opacity genes in Neisseria gonorrhoeae: control of phase and antigenic variation. *Cell, 47*(1), 61-71.

Sun, J., & Klein, A. (2004). A lysR-type regulator is involved in the negative regulation of genes encoding selenium-free hydrogenases in the archaeon Methanococcus voltae. *Molecular microbiology, 52*(2), 563-571.

Sunkavalli, A., McClure, R., & Genco, C. (2022). Molecular Regulatory Mechanisms Drive Emergent Pathogenetic Properties of Neisseria gonorrhoeae. *Microorganisms, 10*(5), 922. Retrieved from https://www.mdpi.com/2076-2607/10/5/922

Svergun, D. I. (1999). Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophysical journal, 76*(6), 2879-2886.

Svergun, D. I., Petoukhov, M. V., & Koch, M. H. (2001). Determination of domain structure of proteins from X-ray solution scattering. *Biophysical journal, 80*(6), 2946-2953.

Teramoto, H., Inui, M., & Yukawa, H. (2013). OxyR acts as a transcriptional repressor of hydrogen peroxide-inducible antioxidant genes in Corynebacterium glutamicum R. *Febs j, 280*(14), 3298-3312. doi:10.1111/febs.12312

Terwilliger, T. C., Grosse-Kunstleve, R. W., Afonine, P. V., Moriarty, N. W., Zwart, P. H., Hung, L.-W., . . . Adams, P. D. (2008). Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallographica Section D: Biological Crystallography, 64*(1), 61-69.

Turnbull, A. L., & Surette, M. G. (2010). Cysteine biosynthesis, oxidative stress and antibiotic resistance in Salmonella typhimurium. *Research in Microbiology, 161*(8), 643-650.

Tyrrell, R., Verschueren, K. H. G., Dodson, E. J., Murshudov, G. N., Addy, C., & Wilkinson, A. J. (1997). The structure of the cofactor-binding fragment of the LysR family member, CysB: a familiar fold with a surprising subunit arrangement. *Structure, 5*(8), 1017-1032. doi:https://doi.org/10.1016/S0969-2126(97)00254-2

Unemo, M., & Shafer, W. M. (2014). Antimicrobial resistance in Neisseria gonorrhoeae in the 21st century: past, evolution, and future. *Clinical microbiology reviews, 27*(3), 587-613.

Volkov, V. V., & Svergun, D. I. (2003). Uniqueness of ab initio shape determination in small-angle scattering. *Journal of applied crystallography, 36*(3), 860-864.

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

WHO. (2016). *WHO guidelines for the treatment of Neisseria gonorrhoeae*. Retrieved from Geneva: World Health Organisation: https://www.who.int/publications/i/item/9789241549691

WHO. (2017). *Prioritization of pathogens to guide discovery, research and development of new antibiotics for drug-resistant bacterial infections, including tuberculosis*. Retrieved from Geneva: World Health Organisation: https://apps.who.int/iris/handle/10665/311820

Wi, T., Lahra, M. M., Ndowa, F., Bala, M., Dillon, J.-A. R., Ramon-Pardo, P., . . . Unemo, M. (2017). Antimicrobial resistance in Neisseria gonorrhoeae: global surveillance and a call for international collaborative action. *PLoS medicine, 14*(7), e1002344.

Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., . . . McCoy, A. (2011). Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D: Biological Crystallography, 67*(4), 235-242.

Workowski, K. B., S. (2010). Sexually Transmitted Diseases Treatment Guideline. *CDC*.

Wu, H. J., Seib, K. L., Srikhanta, Y. N., Kidd, S. P., Edwards, J. L., Maguire, T. L., . . . Jennings, M. P. (2006). PerR controls Mn-dependent resistance to oxidative stress in Neisseria gonorrhoeae. *Molecular microbiology, 60*(2), 401-416.

Zaim, J., & Kierzek, A. M. (2003). The structure of full-length LysR-type transcriptional regulators. Modeling of the full-length OxyR transcription factor dimer. *Nucleic acids research, 31*(5), 1444-1454.

Zheng, M., Åslund, F., & Storz, G. (1998). Activation of the OxyR transcription factor by reversible disulfide bond formation. *Science, 279*(5357), 1718-1722.

Appendix A: Cloning information for CysB

A.1 Sequences for CysB (NGO_1578)

<u>CysB FA 1090 nucleotide sequence NC_022240.1 (1552596...1553546)</u>

ATGAAATTACAACAATTGAAATACGCCTTAGAAGTTTACCGGCACAACCTGAAC
GTTTCCGAAGCGGCCGAAGCCTTATTCACATCACAACCCGGCATCTCCAAACAAA
TCAAATTGCTGGAAGAAGAAATCGGCATTCAGATTTTTATCCGCAGCGGCAAGC
GCGTGGTTTCGGTCTCGCAGCCGGGCAAGGTGGTTTTGGATATTGCGGAACGTAT
TTTGCGCGATGTACAGAACATTAAAAATATCGGCAGCGAGTTTACCGGACAGGA
CAGCGGTTCGCTGACGGTTGCCACGACGCATACGCAGGCGCGCTATGCCCTACC
GTTGATTGTTGCCGATTTTGTGAAACGCTATCCGAAAGTCAATCTGACCATCAAA
CAGGGAAGCCCTGCCGCCATCGCCCAAATGGTTACTTCGGGAGAATCTGATTTGG
CGATTGTTACGGAACGGATAGACGACCACCCCGAGTTGGGAAGGCTTTCCTGCT
ACGACTGGACACACGCGGTGATTGTGCCGAACGACCATCCGCTGCTCGAATGCA
GGAATCCCCTCCGTATTGAAGATTTGGCGAGGTTTCCGCTGATTACTTATGAATTT
GCATTCAATGCGGGCAGCAGCATCGCGCGGGCATTCGCCAAAGCCCGTTTGGAA
CGGCCGGATGTCGCATTGGCGGCGGCGGATACGGATGTATTGAAGACTTATGTG
CGCTTGGGTTTGGGCGTGGGACTGATGGCGAAAATGGCATACAACCCGGATACG
GACGGCGATTTGCAGCTTGTGGATGCGGCACACCTGTTCGAGCCGTCGCCGACGT
GGATAGCTTTGCGCAGCGATACTTATTTGCGCGGATATGCCTACGACTTTATCCA
AGCGTTTGCGCCGCACCTGACACGCGAGAAGGTGGATAGGATTCTGTACACGCC
CATCAGCGAGGATTTTTCGATTTAG

<u>CysB peptide sequence. C-terminal HexaHis-tag (underlined) 316 residues. 35.2/140.8 kDa (monomer/tetramer) | Theoretical pI 5.85</u>

<u>MGSSHHHHHHSSGLVPRGSHM</u>MKLQQLKYALEVYRHNLNVSEAAEALFTSQPGISK
QIKLLEEEIGIQIFIRSGKRVVSVSQPGKVVLDIAERILRDVQNIKNIGSEFTGQDSGSLT
VATTHTQARYALPLIVADFVKRYPKVNLTIKQGSPAAIAQMVTSGESDLAIVTERIDD
HPELGRLSCYDWTHAVIVPNDHPLLECRNPLRIEDLARFPLITYEFAFNAGSSIARAFA
KARLERPDVALAAADTDVLKTYVRLGLGVGLMAKMAYNPDTDGDLQLVDAAHLF
EPSPTWIALRSDTYLRGYAYDFIQAFAPHLTREKVDRILYTPISEDFSI

pET28b description

Table A.1 Description of the pET28b Vector

Vector	Description
pET28b	*E. coli* expression vector, 5368 bps, N-terminal thrombin cleavage site, Kanamycin resistance, C and N-terminal His-tags, T7-promoter.

GMO approval numbers:

E. coli: GMD101146

N. gonorrhoeae: GMD102338

A.2 CysB expression media

Luria Bertani (LB) Agar (1 L)

Combine the following:

10 g Peptone

10 g NaCl

5 g Yeast extract

5 g Agar

1 L of distilled water

Autoclave at 121°C before use.

LB Broth (1 L)

Same method as LB agar, but no addition of agar.

Appendix B: Protein purification and assay information

B.1 Purification buffers for CysB crystallography and assays

Table B.1 Purification buffers for CysB EMSA assays

	Chemical composition
Lysis buffer	50 mM Tris (pH 8.0)
	200 mM NaCl
	20 mM Imidazole
High salt wash buffer	50 mM Tris (pH 8.0)
	4 M NaCl
	20 mM Imidazole
Elution buffer	50 mM Tris (pH 8.0)
	200 mM NaCl
	1 M Imidazole
Size Exclusion buffer	50 mM Tris (pH 8.0)
	200 mM NaCl

Table B.2 Purification buffers for CysB SAXS experiments

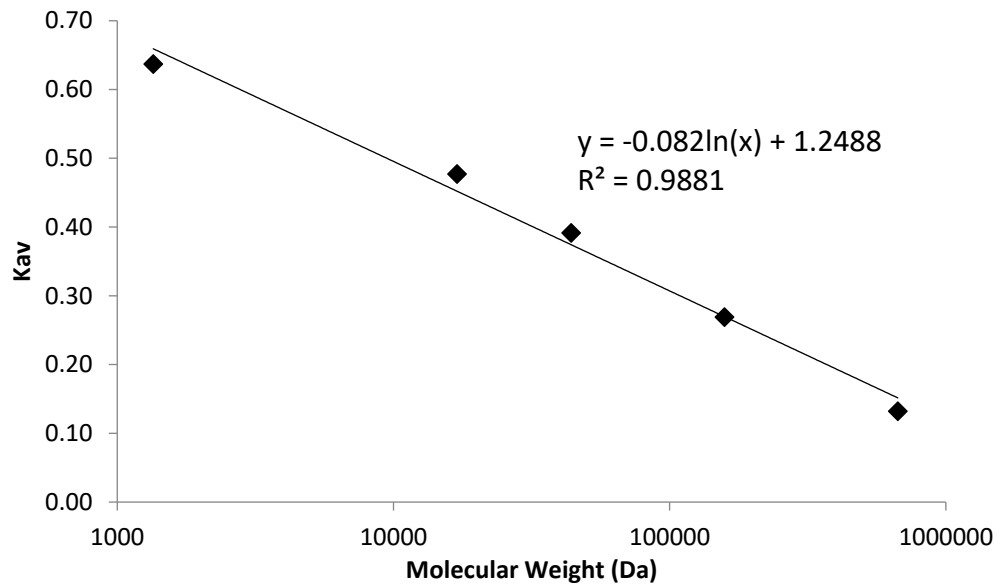|	**Chemical composition**
Lysis buffer	50 mM Tris (pH 8.0)
	200 mM NaCl
	20 mM Imidazole
	10 % *v/v* glycerol
High salt wash buffer	50 mM Tris (pH 8.0)
	4 M NaCl
	20 mM Imidazole
	10 % *v/v* glycerol
Elution buffer	50 mM Tris (pH 8.0)
	200 mM NaCl
	1 M Imidazole
	10 % *v/v* glycerol
Size Exclusion buffer	50 mM Tris (pH 8.0)
	200 mM NaCl
	10 % *v/v* glycerol

BioENrich SEC 650 Calibration Curves

Figure B.1 Calibration curve for gel filtration column ENrich650. Calibration curved was used to calculate the approximate molecular weight and oligomeric state of CysB. Column was calibrated in 50 mM Tris (pH 8.0), 200 mM NaCl.
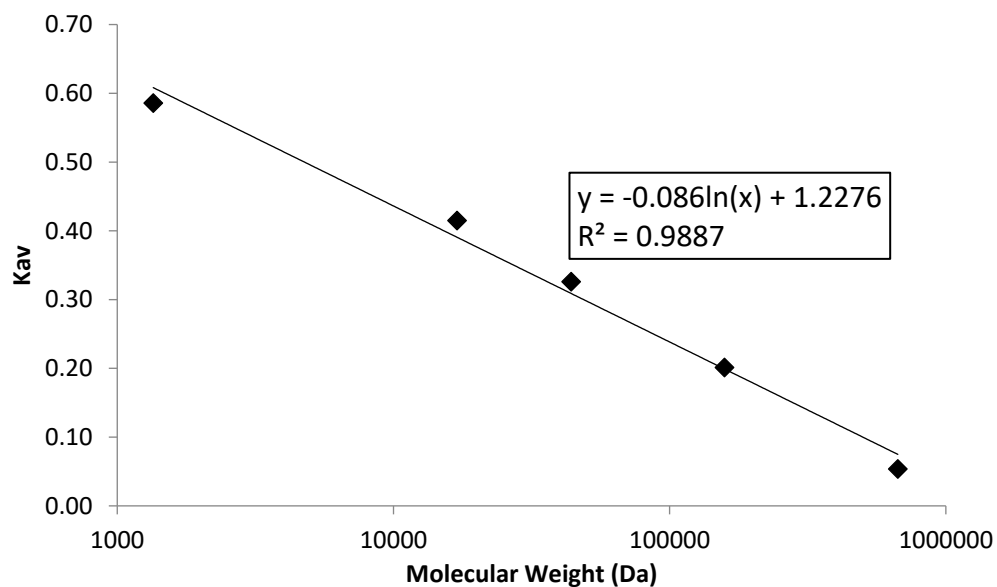

Figure B.2 Calibration curve for gel filtration column ENrich650. Calibration curved was used to calculate the approximate molecular weight and oligomeric state of CysB. Column was calibrated in 50 mM Tris (pH 8.0), 200 mM NaCl, 10% v/v glycerol.

B.2 SDS- and TBE-PAGE gel composition

Table B.3 12% SDS-PAGE gel composition (5 mini gels)

Reagents	Resolving gel (15 %)	Stacking gel (5 %)
Water	10.05 mL	8.5 mL
30 % acrylamide	12 mL	2.125 mL
Resolving Buffer (1.5 M Tris, pH 8.8)	7.5 mL	-
Stacking Buffer (1.0 M Tris, pH 6.8)	-	1.6 mL
10 % SDS	300 μL	125 μL
10 % APS	150 μL	63 μL
TEMED	15 μL	6.3 μL

Table B.4 8%/10% TBE gel composition (5 mini gels)

Reagent	Bottom layer (10%)	Top layer (8%)
Acrylamide:bis-acrylamide (29:1) (40%)	7.5 mL	4.0 mL
TBE Buffer (10X)	3.0 mL	2.0 mL
dH$_2$O	19.5 mL	14.0 mL
Ammonium persulfate (10%)	300 μL	200 μL
TEMED	9 μL	6 μL

B.3 SDS- and TBE-PAGE gel buffers compositions

SDS loading dye

250 mM Tris (pH 6.8)

20% glycerol (w/v)

4% SDS (w/v)

10% beta-2-mercaptoethanol (w/v)

0.025% bromophenol blue (w/v)

Tris-glycine SDS buffer

25 mM Tris (pH 8.5)

250 mM glycine

0.1% SDS (w/v)

Fairbanks staining solution

0.05% coomasie blue

25% isopropanol

10% acetic acid

Destaining solution

10% acetic acid

Tris-Boric acid-electrophoresis (TBE) buffer (10x)

1 M Tris

1 M Boric acid

0.02 M EDTA (disodium salt)

TBE running buffer (1x)

0.1 M Tris

0.1 M Boric acid

0.002 M EDTA (disodium salt)

B.4 DNA binding assay buffers

DNA Binding buffer (5x)

200 mM Tris (pH 8.0)

50 mM $MgCl_2$

500 mM KCl

5 mM DTT

DNA Loading dye (10x)

3mL 30% (v/v) Glycerol

25mg 0.25% (w/v) Bromophenol Blue

25mg 0.25% (w/v) Xylene Cyanol FF

Make up to 10mL with MQ water

B.5 PCR and gene knockout

TAE running buffer (1x)

40 mM Tris

1.42 mL glacial acetic acid

1 mM EDTA (disodium salt)

Table B.5 Primers for cysB knockout

Primer description	Primer sequence 5' ➜ 3'
To amplify the *cysB* knockout construct	
NGFG_01356 CysB KanR FWD	ATA GTC TAT CAT GCC GAA A
NGFG_01356 CysB KanR REV	CCT GTA CGA ACA TTT CAG AC
To confirm *cysB* knockout construct integration	
NGFG_01356 CysB Int PCR FWD	CAG CCT GTT TCC TGA TAG TCT AT
NGFG_01356 CysB Int PCR REV	CAT TTG ATT AAC CTG TAC GAA

Table B.6 Primers to amplify sequences within the cysB promoter. The 5'FAM sequence is underlined.

Primer description	Primer sequence 5' ➔ 3'
cysB promoter – 1-164 bp region (no FAM tag)	
CysB pro FWD	ATA CTG CAG ATG AAA CTG CAA CA
CysB pro REV	TCA TGG CGG GTC GGT
cysB promoter – 1-164 bp region	
CysB pro 1 + FAM tag FWD	<u>CGT CGA GAC GTG</u> TTC CGT GTC ATT
CysB pro REV	TCA TGG CGG GTC GGT
cysB promoter – 1-80 bp region	
CysB pro 1 + FAM tag FWD	<u>CGT CGA GAC GTG</u> TTC CGT GTC ATT
CysB pro 80 REV	CTT TGC CGT TGT GTT TGT TG
cysB promoter – 40-164 bp region	
CysB pro 40 + FAM tag FWD	<u>CGT CGA GAC GTG</u> ACG ATT GCC GTT G
CysB pro REV	TCA TGG CGG GTC GGT
cysB promoter – 81-164 bp region	
CysB pro 81 + FAM tag FWD	<u>CGT CGA GAC GTG</u> CCC CAT ACC GTA
CysB pro REV	TCA TGG CGG GTC GGT

Table B.7 Oligos sequences for DNA-binding assays

Oligo description	Oligo sequence 5' ➔ 3'
cysE promoter – 30 bp upstream of TSS	
CysE pro FWD	CCGCCGCCGA ATACCCTACC GCAAAAACCA TATAAACGGT AAGATACCGC
CysE pro REV	GCGGTATCTT ACCGTTTATA TGGTTTTTGC GGTAGGGTAT TCGGCGGCGG

	cysB promoter 37-86 bp region
Ng_CysB 40-80 50bp FWD	CGCACGATTG CCGTTGCACA AAAACAACAA ACACAACGGC AAAGCCCCAT
Ng_CysB 40-80 50bp REV	ATGGGGCTTT GCCGTTGTGT TTGTTGTTTT TGTGCAACGG CAATCGTGCG
	cysB promoter inverted repeat 1 mutated (mIR1)
CysB BS1 mut FWD	CGCACGAAGT CACGTACACA AAAACAACAA ACACAACGGC AAAGCCCCAT
CysB BS1 mut REV	ATGGGGCTTT GCCGTTGTGT TTGTTGTTTT TGTGTACGTG ACTTCGTGCG
	cysB promoter inverted repeat 2 mutated (mIR2)
CysB BS2 mut FWD	CGCACGATTG CCGTTGCACA AAAACAACAA ACATCAGTGC ATAGCCCCAT
CysB BS2 mut REV	ATGGGGCTAT GCACTGATGT TTGTTGTTTT TGTGCAACGG CAATCGTGCG
	cysB promoter inverted repeats 1 and 2 mutated (mIR1+2)
CysB BS1+BS2 mut FWD	CGCACGAAGT CACGTACACA AAAACAACAA ACATCAGTGC ATAGCCCCAT
CysB BS1+BS2 mut REV	ATGGGGCTAT GCACTGATGT TTGTTGTTTT TGTGTACGTG ACTTCGTGCG
	cysB promoter shorter insert between inverted repeats 1 and 2 (short)
CysB shorter insert FWD	CGCACGATTG CCGTTGCACA AACAACAACA CAACGGCAAA GCCCCAT
CysB shorter insert REV	ATGGGGCTTT GCCGTTGTGT TGTTGTTTGT GCAACGGCAA TCGTGCG
	cysB promoter longer insert between inverted repeats 1 and 2 (long)
CysB longer insert FWD	CGCACGATTG CCGTTGCACA AAAACAACAA CAAACACAAC GGCAAAGCCC CAT
CysB longer insert REV	ATGGGGCTTT GCCGTTGTGT TTGTTGTTGT TTTTGTGCAA CGGCAATCGT GCG

Table B.8 Primers to amplify promoters of the cysteine regulon

| Primer description	Primer sequence 5' → 3'
sbp promoter – region encompassing 11 bp upstream TSS – 312bp upstream TSS, 119 bp into neighbouring coding sequence.	
sbp Pro FWD	CAT CTC TCA ACT CCG TCC CG
sbp Pro REV	GGA TGG TTT GAA TCG GCT GC
cysA promoter – region encompassing 81 bp downstream TSS – 237 bp upstream TSS, 96 bp into neighbouring coding sequence.	
CysA Pro FWD	GGG GAC GTT GAG GTT GAT GT
CysA Pro REV	GAA TAC AAC TTC ACC GGC GC
cysU promoter – region encompassing 23bp downstream TSS – 479 bp upstream TSS, 299 bp into neighbouring coding sequence.	
CysU Pro FWD	CCG GGT GTT TTG AGC GGT AA
CysU Pro REV	CCG CAA TCT TTA GCC GCT TC
cysW promoter – region encompassing 36 bp downstream of TSS – 354 bp upstream TSS, 166 bp into neighbouring coding sequence.	
CysW Pro FWD	TTC GGT CAG GTT GGG ATT GG
CysW Pro REV	GAT TTC CGA AAT CCT GCC GC
cysK promoter – region encompassing 29 bp upstream TSS – 480 bp upstream of TSS, 285 bp into neighbouring coding sequence.	
CysK Pro FWD	ATG ACG CTT GCT TTT GTC GG
CysK Pro REV	AGT AAG GCG CAA TCC CCA TC

Invitrogen 1kb+ DNA ladder
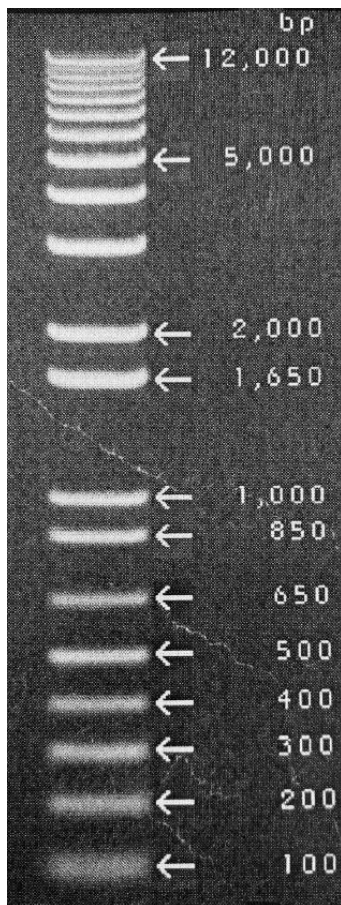
Figure B.3 Invitrogen 1kb+ DNA ladder used in agarose gels.

GCB Agar

36.25 g GCB base (Difco)

1.25 g Agar

1 L MQ H2O

Split into two 500 mL Schott bottles

Autoclaved

500 µL of 50 mg.mL-1 kanamycin added to one Schott bottle

Poured into petri dishes and set for ~20 mins

Set plates stored at 4° C

Appendix C: Crystallisation and Structure solving information

C.1 Tacsimate™ composition (100% (*v/v*), pH 7.0) – HR2-755

1.8305 M Malonic acid

0.25 M Ammonium citrate tribasic

0.12 M Succinic acid

0.3 M DL-Malic acid

0.4 M Sodium acetate trihydrate

0.5 M Sodium formate

0.16 M Ammonium tartrate dibasic
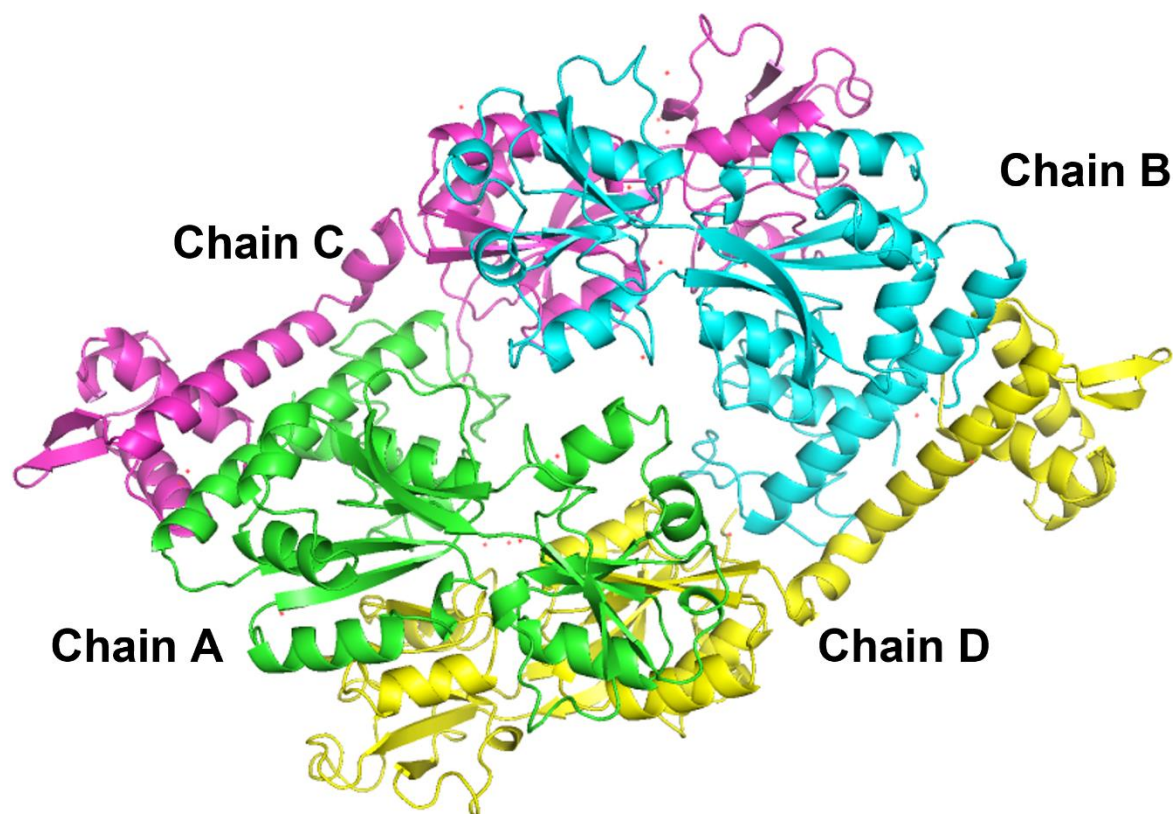
C.2 Contents of asymmetric unit for CysB.

Figure C.1 Contents of the asymmetric unit for the CysB model. Waters are represented as red crosses. Figure generated using PyMOL version 2.3.4.

C.3 Input settings for *phenix.autobuild.*

Initial and final *phenix.autobuild* settings

Figure C.2 Input settings for initial and final phenix.autobuild for CysB structure.

C.4 CysB structural files

The CysB final X-ray Crystallography structure (.pdb) and map (.mtz) files can be found in the Google Drive link below.

https://drive.google.com/drive/folders/19P_lG8t78TUaxIDB6w6u43sPI3JIGiSU?usp=share_link

C.5 CysB chain statistics

Table C.1 CysB structure statistics for each monomer in the asymmetric unit.

Chain	Residue chain built	Residues missing	Total built
A	Met1-Leu306	10	306/316
B	Met1-Tyr307 Ser11-Phe314	5	311/316
C	Met1-Ile310	6	310/316
D	Met1-Ile310	6	310/316

C.6 PDB Validation statistics of X-ray crystallography structure

Full PDB Validation report in the Google Drive link below.

https://drive.google.com/drive/folders/19P_lG8t78TUaxIDB6w6u43sPI3JIGiSU?usp=share_link
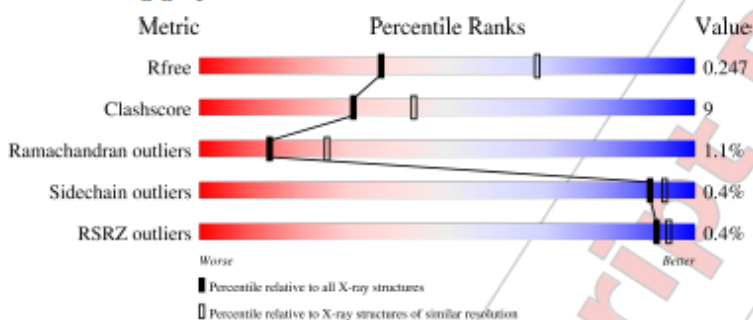
Preliminary Full wwPDB X-ray Structure Validation Report

1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.73 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.

Metric	Percentile Ranks	Value
Rfree		0.247
Clashscore		9
Ramachandran outliers		1.1%
Sidechain outliers		0.4%
RSRZ outliers		0.4%

Worse ... *Better*

▌ Percentile relative to all X-ray structures
▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1271 (2.76-2.72)
Clashscore	141614	1322 (2.76-2.72)
Ramachandran outliers	138981	1297 (2.76-2.72)
Sidechain outliers	138945	1298 (2.76-2.72)
RSRZ outliers	127900	1243 (2.76-2.72)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments of the lower bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria respectively. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions <=5% The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	B	311	83% 16%
2	A	306	80% 19%
3	C	310	82% 16%
3	D	310	80% 19%

WORLDWIDE
wwPDB
PROTEIN DATA BANK

Figure C.3 PDB Validation statistics of CysB X-ray crystallography structure.

Appendix D: SAXS model building

D.1   CysB UV traces during SEC-SAXS and CHROMIXS profiles

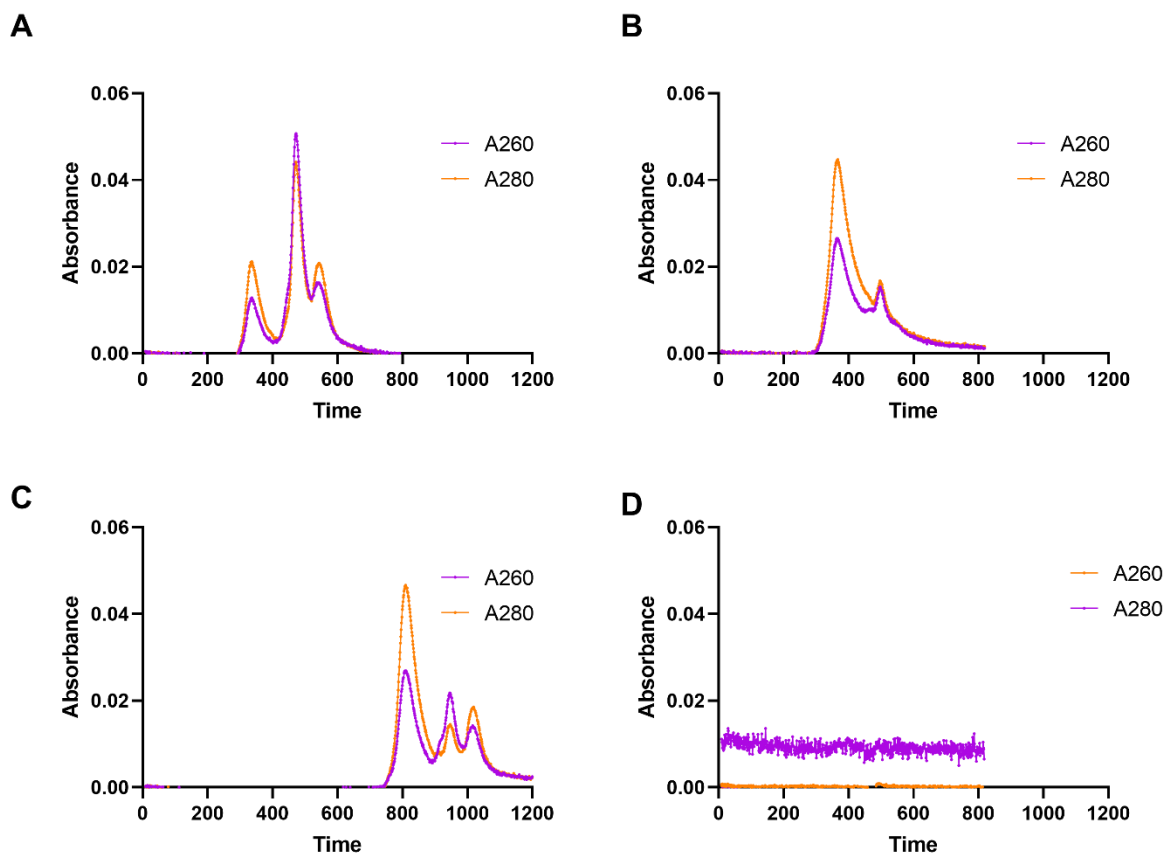**Poor UV traces from unusable data**

*Figure D.1 Poor UV traces from SAXS data during SEC-SAXS. These UV traces are from CysB eluting from the size exclusion column from samples sent during trip two to the Australian Synchrotron SAXS/WAXS Beamline for collection. This trip experienced many delays and detector issues. The UV traces from each of the samples (**A-D**) show multiple peaks indicating the data was of unusable quality. The 260 nm chromatogram is shown in purple, the 280 nm chromatogram is shown in orange. Figures generated using Prism version 9.4.1.*

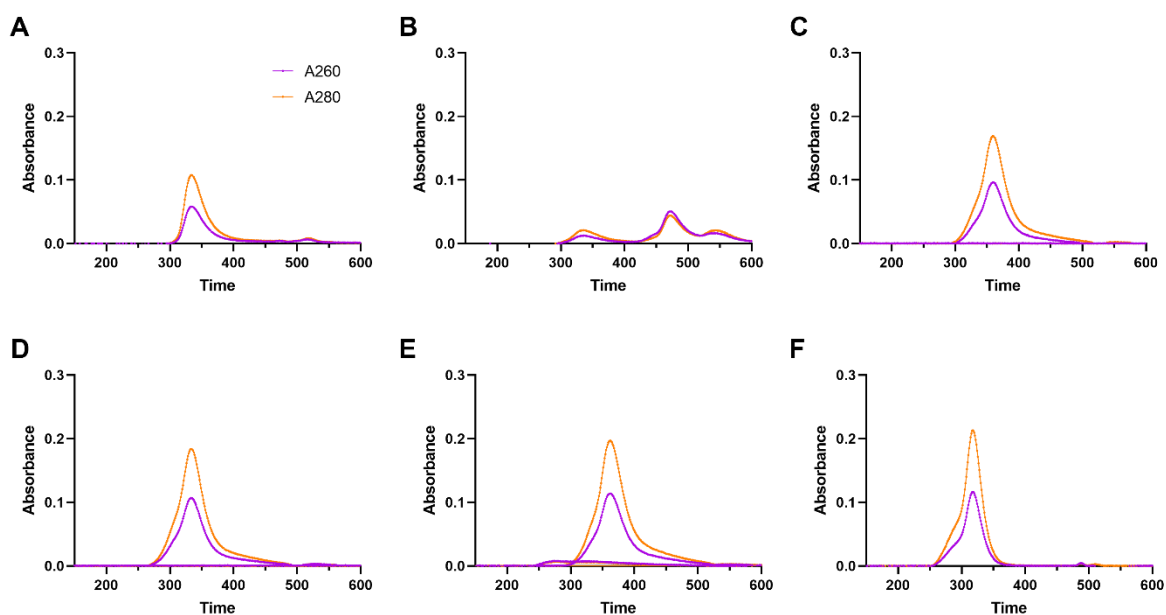# Optimisation of CysB concentration

Figure D.2 Optimisation of CysB concentration for SAXS. UV traces of CysB samples with increasing concentration of CysB. The 260 nm chromatogram is shown in purple, the 280 nm chromatogram is shown in orange. Similar elution tomes and UV profiles show stability of CysB during SEC-SAXS and give confidence in results obtained across samples. ***A*** *UV trace of CysB at 3.6mg.mL^{-1} from the first SAXS trip.* ***B*** *UV trace of CysB at 4.7 mg.mL^{-1} from the second SAXS trip.* ***C-E*** *UV traces of CysB at 4 mg.mL^{-1} (**C**), 5 mg.mL^{-1} (**D**) and 6 mg.mL^{-1} (**E**) from the third SAXS trip. These samples show a correlation between increasing concentration of CysB and increasing absorbance.* ***F*** *UV trace of CysB at 6 mg.mL^{-1} from the fourth SAXS trip. Figures generated using Prism version 9.4.1.*
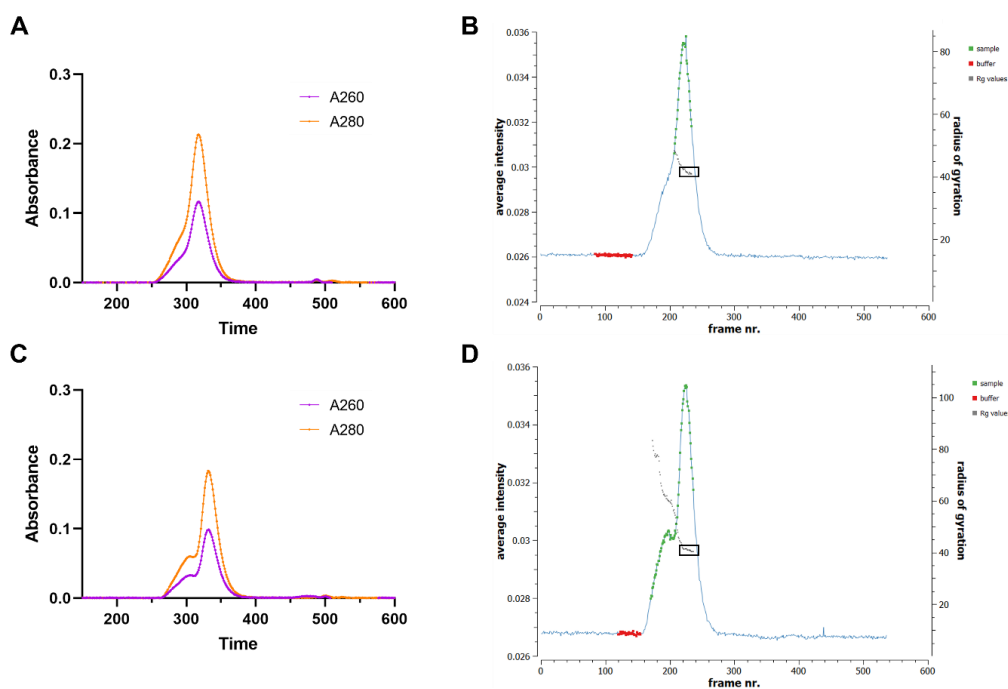
Final CysB-*apo* and CysB-*NAS* data sets

*Figure D.3 UV traces and CHROMIXS plots of final CysB-apo and CysB-NAS samples. The UV traces show a single peak with a preceding shoulder. The CHROMIXS plots allow for analysis of Rg values across the peak and hence deconvolution of the peak. The increased Rg values at the start of the peak indicate a larger species due to the larger Rg values. Data points used from the peak for downstream data processing including the Guinier, Kratky and P(r) distribution plots is highlighted within the box. All other data points were excluded. The UV trace (**A**) and CHROMIXS plot (**B**) for the CysB-apo data show a small preceding shoulder. The UV trace (**C**) and CHROMIXS plot (**D**) for the CysB-NAS data show a larger preceding shoulder. The data points from this shoulder were excluded for downstream processing. Figures generated using Prism version 9.4.1 and CHROMIXS from the ATSAS Suite 3.1.3.*

D.2 SAXS scattering data files

The raw scattering data for the CysB-*apo* and CysB-*NAS* data presented in this thesis can be found in the Google Drive link below.

https://drive.google.com/drive/folders/19P_lG8t78TUaxIDB6w6u43sPI3JIGiSU?usp=share_link

D.3 SAXS DAMAVER models

The SAXS surface models of CysB-*apo* and CysB-*NAS* presented in this thesis can be found in the Google Drive link below.

https://drive.google.com/drive/folders/19P_lG8t78TUaxIDB6w6u43sPI3JIGiSU?usp=share_link