Biochemical and Structural Studies of Metallo-β-Lactamase Superfamily Enzymes



A thesis submitted to the board of the Faculty of Physical Sciences of the University of Oxford in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Aistė Skorupskaitė

St Anne's College, University of Oxford

Michaelmas Term 2019

Abstract

Aistė Skorupskaitė DPhil Thesis

St Anne's College Michaelmas Term 2019

Biochemical and Structural Studies of Metallo-β-Lactamase Superfamily Enzymes

The metallo-β-lactamase (MBL) superfamily is characterised by a unique αββα structural fold and is widespread in nature. Whilst best known for its name-lending members that hydrolyse β-lactam antibiotics and contribute to antimicrobial resistance, other members of the MBL superfamily also utilise metal ions to catalyse various hydrolysis and redox reactions. The work described in this thesis focuses on characterisation of two MBL proteins – cytidine monophosphate *N*-acetylneuraminic acid hydroxylase (CMAH) and glyoxalase II (GloB).

CMAH catalyses hydroxylation of a sialic acid cytidine monophosphate *N*-acetylneuraminic acid (CMP-Neu5Ac) to CMP-*N*-glycolylneuraminic acid (CMP-Neu5Gc) in the presence of O_2, cytochrome (cyt) b_5, cyt b_5 reductase and NADH and is found in animals of the Deuterostomia lineage. Human (h) CMAH is catalytically inactive due to a deletion which occurred approximately 3 million years ago. Understanding the differences between hCMAH and its catalytically active homologues could provide insight into the CMAH mechanism and the function of hCMAH in humans. In this thesis, development of expression systems for hCMAH and its active mouse homologue mCMAH, their initial characterisation and efforts to obtain the first CMAH structure are described.

Despite the lack of catalytic activity of hCMAH, the human *CMAHP* gene is transcribed in cells suggesting an alternative function. Studies on the cellular localisation of hCMAH revealed cytoplasmic localisation with some compartmentalisation being observed. Additionally, attempts were made to detect proteins interacting with hCMAH and mCMAH in mammalian cells.

GloB is involved in the methylglyoxal detoxification pathway. GloB catalyses hydrolysis of *S*-D-lactoylglutathione to glutathione and D-lactate. The thesis described work on the characterisation of GloB from *Escherichia coli*. Kinetic studies using various transition metals are described as well as the search for potential inhibitors. Crystal structures of GloB at 1.1 Å and 1.8 Å resolution are reported and initial attempts to obtain the structure of an enzyme-substrate complex are described.

Overall, the work described in this thesis explores the biochemical, functional and structural features of two very different MBL superfamily proteins and provides insights into the scope of differences present within the superfamily.

Acknowledgements

Firstly I would like to thank my supervisor Prof. Chris Schofield for an opportunity to work in his lab, for support and guidance, and being an endless source of new ideas.

The biggest thanks goes to Dr Ilaria Pettinati for her guidance, advice and supervision. Thanks for teaching me most of the techniques, especially when it came to mammalian cell work and not giving up on me when the I came to her with the same question for the third time. Without her support, many of the things described in this thesis would not have been possible.

For help with different aspects of my research, I would like to thank past and current members of the CJS group: Dr Tony Tumber (Rapid Fire MS), Shuang Liu (Native MS), Dr Karina Calvopiña (kinetic assays), Dr Mike McDonough and Dr Ian Cliffton (crystallography and data processing), Dr Jürgen Brem (various aspects of the projects), Dr Joan Gannon (immunuofluorescence microscopy), Dr Madeline Kavanagh (GloB plasmid and purification protocol). Thank you to Dr David Owen and Dr Martin Walsh (Research Complex at Harwell and Diamond Light Source) for help with crystallography, and to Dr William Myers (CAESR) for his help with EPR measurements. I would also like to thank Dr Adam Hardy and Manso Jubier Herminio for their lab management and making sure everything runs smoothly.

Thank you goes to past and present members of the CJS group Tharindi Panduwawala, Pauline Lang, Jos Kamps, Patrick Rabe, Karina Calvopiña, Jürgen Brem, Ilaria Pettinati, Sam Cahill, Thomas Leißing, and many others for their friendship and creating an enjoyable atmosphere in the lab. I would like to thank Dr Ilaria Pettinati, Dr Thomas Leißing, Jiao-Yu Wang, Dr Sam Cahill, Dr Chiara Maniaci, Dr Tony Tumber, Reda Stankūnaitė, Dr Mike McDonough, Dr Ian Cliffton, Dr Karina Calvopiña, and Dr Alistair Farley for taking time to proofread my thesis and providing comments and advice.

I am very grateful to my friends outside the lab Giedrė Čepukaitytė, Reda Stankūnaitė, Karolina Subko, and Emma Cawood for their friendship throughout the years and always providing a listening ear.

Thanks go to Evgeny Romanenkov for his support and encouregement and for trying to keep me sane during the most stressful moments. Your support means a lot to me!

Finally, a massive thank you goes to my family for their unconditional love, support and encourement. Be Jūsų rūpesčio ir palaikymo visa šitai nebūtų buvę įmanoma. Ačiū!

Declaration

Contributions from collaborators are described at the beginning of each chapter and highlighted in text were appropriate. The rest of the work is my own.

Contents

Abstract ..... **i**

Acknowledgements ..... **ii**

Declaration ..... **iii**

Abbreviations ..... **x**

Chapter 1: General Introduction **1**

1.1. History of the metallo-β-lactamase superfamily 2

1.2 Functional and structural range 3

1.2.1 Metallo-β-lactamases 7

1.2.2 Nucleic acid processing MBL fold enzymes 9

1.2.3 Other MBL fold hydrolases 12

1.2.4 MBL fold oxidoreductases 15

1.2.5 Enzymes with multiple and unknown functions 16

1.3 Metal binding within the MBL superfamily 19

1.4 MBL fold proteins and disease 20

1.5 Summary 22

1.6 Objectives 23

1.7 References 24

Chapter 2: Biochemical and Biophysical Characterisation of CMAH **40**

2.1 Introduction 41

2.1.1 Sialic acids 41

2.1.2 Cytidine monophosphate *N*-acetylneuraminic acid hydroxylase 42

2.1.3 Loss of CMAH activity in humans 45

2.1.4 Other inactivation events 46

2.2 Objectives 46

2.3 Production of recombinant hCMAH and mCMAH proteins 47

2.3.1 Cloning 47

2.3.2 Bacterial expression trials 48

2.3.3 Large scale bacterial expression 50

2.3.4 Mammalian expression systems 52

2.4 Characterisation 56

2.4.1 Metal content analysis 56

2.4.2 Deglycosylation experiments 57

2.4.3 Mass spectrometry and Circular Dichroism spectrosopy 58
2.4.4 Oligomerisation state analysis 61
2.5 Characterisation of FeS cluster 64
2.5.1 UV-Vis spectroscopy 64
2.5.2 Electron paramagnetic resonance 64
2.6 Hydroxylase activity assay 66
2.6.1 Assay optimisation and NADH decomposition 68
2.6.2 Reaction timecourse 69
2.7 Crystallographic studies 71
2.7.1 Thermal shift assays 71
2.7.2 hCMAH crystallisation trials 74
2.7.3 mCMAH crystallisation trials 76
2.9 Summary and discussion 79
2.10 References 80

Chapter 3: Investigating the Role of CMAH in Living Cells 87

3.1 Introduction 88
3.1.1 Expression of human CMAH gene 88
3.1.2 CMAH localisation and mechanism of action 89
3.1.3 Cytochrome b_5 and cytochrome b_5 reductase 89
3.2 Objectives 90
3.3 Immunofluorescence microscopy 90
3.3.1 Subcellular localisation of CMAH proteins 90
3.3.2 hCMAH does not localise to the endoplasmic reticulum or the mitochondria 91
3.3.3 Clustering of hCMAH in HeLa cells 93
3.4 Search for interacting proteins 93
3.4.1 Generating a stably mCMAH expressing TCMK-1 cell line 94
3.4.2 Pulldown experiments 95
3.4.3 Co-immunoprecipitation 96
3.5 Summary and discussion 97
3.6 References 99

Chapter 4: Biochemical and Structural Studies of *Escherichia coli* Glyoxalase II 103

4.1 Introduction 104
4.1.1. Formation and toxicity of methylglyoxal 104
4.1.2 Glyoxalase system 106

4.1.3 Glyoxalases as potential drug targets 108
4.2 Objectives 109
4.3 Protein production and purification 109
4.4 Characterisation of GloB 111
4.4.1 Metal content analysis 111
4.4.2 Non-denaturing MS 111
4.4.3 Circular dichroism 113
4.5 Kinetic and inhibition studies 115
4.5.1 Kinetics and metal substitution 115
4.5.2 Search for GloB inhibitors 117
4.6 Crystallography 121
4.6.1 Buffer optimisation 121
4.6.2 Crystallisation trials 122
4.6.3 Search for an alternative crystal form 127
4.6.4 GloB structure within the Glx2 family 134
4.7 Summary and conclusions 135
4.8 References 138

Chapter 5: Summary and Conclusions **146**
5.1 Biochemical and Biophysical Characterisation of CMAH 147
5.2 Investigating the Role of CMAH in Living Cells 148
5.3 Biochemical and Structural Studies of *Escherichia coli* Glyoxalase II 149
5.4 References 151

Chapter 6: Materials and Methods **153**
6.1 Materials 156
6.1.1 Cloning reagents 156
6.1.2 Competent cells and cell growth media 156
6.1.2 Mammalian cell lines and cell growth media 156
6.1.3 Protein purification columns 156
6.2 General procedures 157
6.2.1 pH measurements 157
6.2.2 Centrifugation 157
6.2.3 DNA and protein concentration measurements 157
6.2.4 Filtering solutions 157
6.3 Bacterial growth media, stock solution and buffer preparation 157

6.3.1 LB-agar medium preparation 157
6.3.2 Bacterial media preparation 158
6.3.3 Antibiotic stock preparation 158
6.3.4 IPTG stock solution preparation 158
6.3.5 Buffer preparation 158
6.4 Electrophoretic techniques 159
6.4.1 Agarose gel electrophoresis 159
6.4.2 SDS-PAGE 160
6.4.3 Western blotting 162
6.5 Molecular Cloning 164
6.5.1 Primers and vectors 164
6.5.2 Codon-optimised gene sequences 165
6.5.3 Polymerase chain reaction 165
6.5.4 Purification of PCR products 166
6.5.5 DNA digestion and ligation 166
6.5.6 Bacterial transformations 167
6.5.7 Plasmid extraction 167
6.6 Bacterial growth and protein production 167
6.6.1 Starter culture preparation 167
6.6.2 Expression trials 167
6.6.3 Large scale culture 168
6.7 Mammalian cell culture 168
6.7.1 General techniques 168
6.7.2 Maintenance of HeLa and TCMK-1 cells 168
6.7.3. Maintenance of adherent HEK293 168
6.7.4 Suspension adaptation for HEK293 169
6.7.5 Maintenance of HEK293 in suspension 169
6.7.6 Freezing and thawing cells 169
6.7.7 Cell transfection and stable cell line generation 170
6.7.8 Transient expression using Expi293 system 170
6.7.9 Transient expression using ExpiCHO system 170
6.7.10 Sample preparation for Western blotting 170
6.8 Protein purification 171
6.8.1 Cell lysis 171
Bacterial cells *171*

Expression trials *171*
Mammalian cells *171*
6.8.2 General procedures 172
Tag cleavage by 3C protease *172*
Protein concentrating *172*
Buffer exchange and protein storage *172*
6.8.3 Chromatographic techniques 172
Ni-affinity chromatography *172*
Size exclusion chromatography *172*
Elution of untagged protein and removal of His_6-tag and 3C protease *174*
Ion exchange chromatography *174*
Further ion exchange using spin columns *174*
6.9 X-ray crystallography 174
6.9.1 Broad screen and optimisation conditions 174
6.9.2 Crystallisation experiments 174
6.9.3 Soaking experiments 175
6.9.4 Crystal harvesting 175
6.9.5 Diffraction data collection and processing 175
6.9.6 Phasing and structure refinements 176
6.9.7 Data deposition 176
6.10 Thermal shift assays 176
6.11 Metal content analysis 176
6.11.1 ICP-MS 176
6.11.2 MicroPIXE 176
6.12 Generation of Apo enzymes 177
6.13 UV-visible spectroscopy 177
6.14 Circular Dichroism spectroscopy 177
6.15 Mass spectrometry 177
6.15.1 Non-denaturing MS 177
6.15.2 RapidFire High-throughput MS 178
6.16 mCMAH assays 178
6.17 GloB assays 179
6.17.1 Assay conditions 179
6.17.2 Measurements 179
6.17.3 Data analysis 180

6.18 Protein-protein interactions 180
6.18.1 Crosslinking and decrosslinking 180
6.18.2 Pulldowns 180
6.18.3 Co-immunoprecipitation 181
6.19 Immunofluorescence 181
6.19.1 Coverslip preparation 181
6.19.2 Image collection and processing 182
6.20 References 183

Appendix 1: Protein Sequences 185
Appendix 2: Compounds Tested for GloB Inhibition 187

Abbreviations

(v/v)	Volume to volume ratio
(w/v)	Weight to volume ratio
2TY	2x Yeast tryptone medium
aCPSF	Archaeal cleavage and polyadenylation specificity factor
AGE	Advance glycation endproduct
BcII	*Bacillus cereus* β-lactamase II
bp	Base pair
CAPS	*N*-Cyclohexyl-3-aminopropanesulfonic acid buffer
CD	Circular dichroism spectroscopy
CHO	Chinese hamster ovary cells
CMAH	Cytidine monophosphate *N*-acetylneuraminic acid hydroxylase
CMAHP	Human *CMAH* pseudogene
CMP	Cytidine monophosphate
co-IP	Co-immunoprecipitation
CPSF	Cleavage and polyadenylation specificity factor
CPSF100	CPSF 100 kDa subunit
CPSF73	CPSF 73 kDa subunit
cryo-EM	Cryogenic electron microscopy
cyt b_5	Cytochrome b_5
DAPI	4′,6-Diamidino-2-phenylindole
DHAP	Dihydroxyacetone phosphate
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleoside triphosphate
DPA	Dipicolinic acid, 2,6-pyridine-dicarboxylic acid
DTNB	5,5'-Dithiobis-2-nitrobenzoic acid
EDTA	Ethylenediaminetetraacetic acid
EPR	Electron paramagnetic resonance
ER	Endoplasmic reticulum
ESI	Electrospray ionization
ETHE1	Ethylmalonic encephalopathy 1 protein
FAD	Flavin adenine dinucleotide

FDP	Flavodiiron protein
FPLC	Fast protein liquid chromatography
G-418	Geneticin 418
GAP	Glyceraldehyde-3-phosphate
GFP	Green fluorescent protein
GloB	Glyoxalase II
Glx	Glyoxalase
Glx1	Glyoxalase I
Glx2	Glyoxalase II
GSH	Glutathione
HAGH	Hydroxyacylglutathione hydrolase (glyoxalase II)
hCMAH	Human CMAH
HEK293	Human embryonic kidney 293 cells
HeLa	Henrietta Lacks cells
HEPES	4-(2-Hydroxyethyl)-1-piperazineethanesulfonic acid
HH	Hoyeraal-Hreidarsson syndrome
HPLC	High-performance liquid chromatography
IC_{50}	Half maximal inhibitory concentration
ICP-MS	Inductively coupled plasma mass spectrometry
IF	Immunofluorescence
IMAC	Immobilized metal affinity chromatography
INTS	Integrator complex subunits
IPTG	Isopropyl β-D-1-thiogalactopyranoside
LC	Liquid chromatography
m/z	Mass-to-charge ratio
MBL	Metallo-β-lactamase
MBP	Maltose binding protein
mCMAH	Mouse CMAH
MEM	Minimum Essential Medium
MG	Methylglyoxal
MPD	2-Methyl-2,4-pentanediol
MPH	Methyl parathion hydrolase
mRNA	Messenger RNA
MS	Mass spectrometry
NADH	Nicotinamide adenine dinucleotide

NAPE-PLD	*N*-Acylphosphatidylethanolamine-hydrolysing phospholipase D
Neu5Ac	*N*-Acetylneuraminic acid
Neu5Gc	*N*-Glycolylneuraminic acid
NHEJ	Non-homologous end joining
NMR	Nuclear magnetic resonance
OAA	Octylammonium acetate
OD_{600}	Optical density at 600 nm
OGDC-E2	E2 component of the 2-oxoglutarate dehydrogenase
OPH2	Organophosphate hydrolase 2
PBS	Phosphate-buffered saline
PBS-T	Phosphate bufferred saline supplemented with 0.1% Triton X-100
PCR	Polymerase chain reaction
PDB	Protein Data Bank
PEG	Polyethylene glycol
PNKD	Paroxysmal non-kinesigenic dyskinesia
ppm	Parts per million
PqsE	Pseudomonal quinolone signal E protein
Pso2	Psoralen-sensitive 2
Q-TOF	Quadrupole time-of-flight mass spectrometry
RIPA	Radioimmunoprecipitation assay buffer
RMSD	Root-mean-square deviation
RNA	Ribonucleic acid
ROO	Rubredoxin oxygen:oxidoreductase
SBL	Serine β-lactamase
SDS-PAGE	Sodium dodecyl sulfate–polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
SEC-MALS	Size exclusion chromatography - multiangle light scattering
Sia	Sialic acid
SLG	*S*-D-Lactoylglutahione
SNM1	Sensitive to nitrogen mustard 1
snRNA	Small nuclear RNA
SPE	Solid phase extraction
TAE	Tris-acetate buffer containing EDTA
TB	Terrific Broth medium
TCEP	Tris(2-carboxyethyl)phosphine

TCMK-1	Transformed C3H Mouse Kidney-1 cells
TEMED	Tetramethylethylenediamine
TIM	Triose phosphate isomerase
TLA	L-Tartaric acid
TLC	Thin layer chromatography
T_m	Melting temperature
Tris	Tris(hydroxymethyl)aminomethane buffer
tRNA	Transfer RNA
UV	Ultraviolet
UV-Vis	Ultraviolet–visible spectroscopy
VIM-2	Verona integron-encoded metallo-β-lactamase
WB	Western blot
WT	Wild type
β-CASP	Metallo-β-lactamase-associated, CPSF, Artemis, SNM1/PSO2 domain

Standard abbreviations for nucleotides, amino acids, SI-units and prefixes and chemical elements are used throughout.

Chapter 1: General Introduction

Contents

1.1. History of the metallo-β-lactamase superfamily 2
1.2 Functional and structural range 3
1.2.1 Metallo-β-lactamases 7
1.2.2 Nucleic acid processing MBL fold enzymes 9
1.2.3 Other MBL fold hydrolases 12
1.2.4 MBL fold oxidoreductases 15
1.2.5 Enzymes with multiple and unknown functions 16
1.3 Metal binding within the MBL superfamily 19
1.4 MBL fold proteins and disease 20
1.5 Summary 22
1.6 Objectives 23
1.7 References 24

1.1. History of the metallo-β-lactamase superfamily

The β-lactam antibiotics were first introduced to the market in the 1940s[1] and have since become one of the most widely used class of antimicrobial agents, constituting around 60% of total antibiotics taken worldwide[2]. β-Lactam antibiotics contain a reactive four-membered β-lactam ring (**Figure 1.1**) and act by inhibiting bacterial cell wall biosynthesis.[3] One of the major modes of antimicrobial resistance to β-lactams is the production of enzymes – β-lactamases – hydrolysing the β-lactam ring and thus inactivating the antibiotic. Serine β-lactamases (SBLs) were discovered in 1940 by Abraham and Chain[4], while the first metallo-β-lactamase (MBL) was discovered in 1965 by Sabath and Abraham[5].

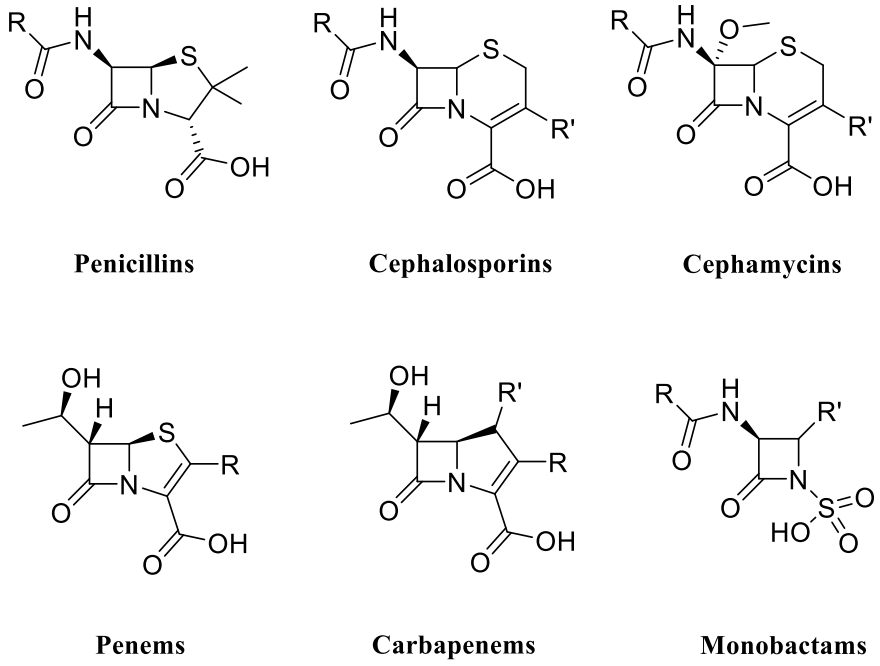


Figure 1.1: Core structures of the major classes of β-lactam antibiotics.

The first MBL was identified in *Bacillus cereus* as a cephalosporinase that was inhibited by treatment with ethylenediaminetetraacetic acid (EDTA) and other metal chelating agents with its activity being restored upon addition of Zn^{2+} salt.[6] It was the second β-lactamase discovered in *B. cereus* [5] and is now known as BcII (*B. cereus* β-lactamase II). BcII was the only known MBL for 16 years until the identification of Zn-dependent β-lactamase L1 in *Stenotrophomonas maltophilia* in 1982.[7, 8] This was followed by the discovery of numerous other MBLs in the 1990s and 2000s[9]. While initially not considered to be of clinical significance, MBLs have since become a substantial concern due the discovery of MBL-encoding mobile genetic elements and their ability to hydrolyse penems and carbapenems used as last resort antibiotics.[10, 11]

A crystal structure of BcII reported in 1995 revealed a novel αββα three-dimensional structural fold distinct from other metallopeptidases.[12] Increase in available gene and protein sequences and improvement of automated bioinformatic methods have led to the expansion of MBL structural superfamily to include a wide range of members, such as glyoxalase II (Glx2), arylsulfatases, phosphodiesterases, and others.[13, 14] Five conserved sequence motives have been defined – (i) Asp84, (ii) His116-X-His118-X-Asp120-His121, (iii) His196, (iv) Asp221, (v) His263 (residue numbers are given using BBL (class B β-lactamase) numbering[15, 16]), with (iv) Cys221 replacing Asp221 in most classes of bacterial β-lactamases.[17] A crystal structure of human glyoxalase II (Glx2) was determined in 1999 making Glx2 the first non β-lactamase and the first eukaryotic protein to be experimentally demonstrated to possess the MBL fold.[18] A structure of the first MBL fold redox protein rubredoxin oxygen:oxidoreductase (ROO), from bacterium *Desulfovibrio gigas*, followed shortly after.[19] To date, there are over 600 structures of more than 80 unique proteins belonging to the MBL superfamily in the Protein Data Bank (PDB) with over 70,000 sequences of confirmed and predicted members listed on the European Bioinformatics Institute Pfam database.[20]

1.2 Functional and structural range

Reactions attributed to members of the MBL superfamily include, in addition to β-lactamase activity, the hydrolysis of thioesters, phosphoesters (including RNA and DNA), sulfates, as well as redox reactions (**Figure 1.2**). While the overuse of β-lactam antibiotics likely contributed to the spread of β-lactamase MBLs in bacteria, the wide distribution of MBL proteins within living organisms and their large variety of function suggest that the 'original' function of the ancestor MBL proteins was likely unrelated to the β-lactamase activity.[9, 17] The identity of the ancestral proteins is not known, however, due to multiple MBL superfamily members with known nucleic acid processing activity, nucleic acid hydrolase activity is a likely contender.[9, 21]

Functional diversity of the MBL superfamily is aided by presence of other domains. Additional domains in MBL fold carrying proteins include, among others, β-CASP (metallo-β-lactamase-associated CPSF Artemis SNM1/PSO2), RNA-specificity domains found in nucleic acid processing proteins, rhodanese-like domain found in cysteine desulfurase, flavodoxin and FeS cluster binding domains in redox proteins, and several domains with unknown functions (**Figure 1.3**).[20] List of proteins discussed in this chapter and their details are given in **Table 1.1**.

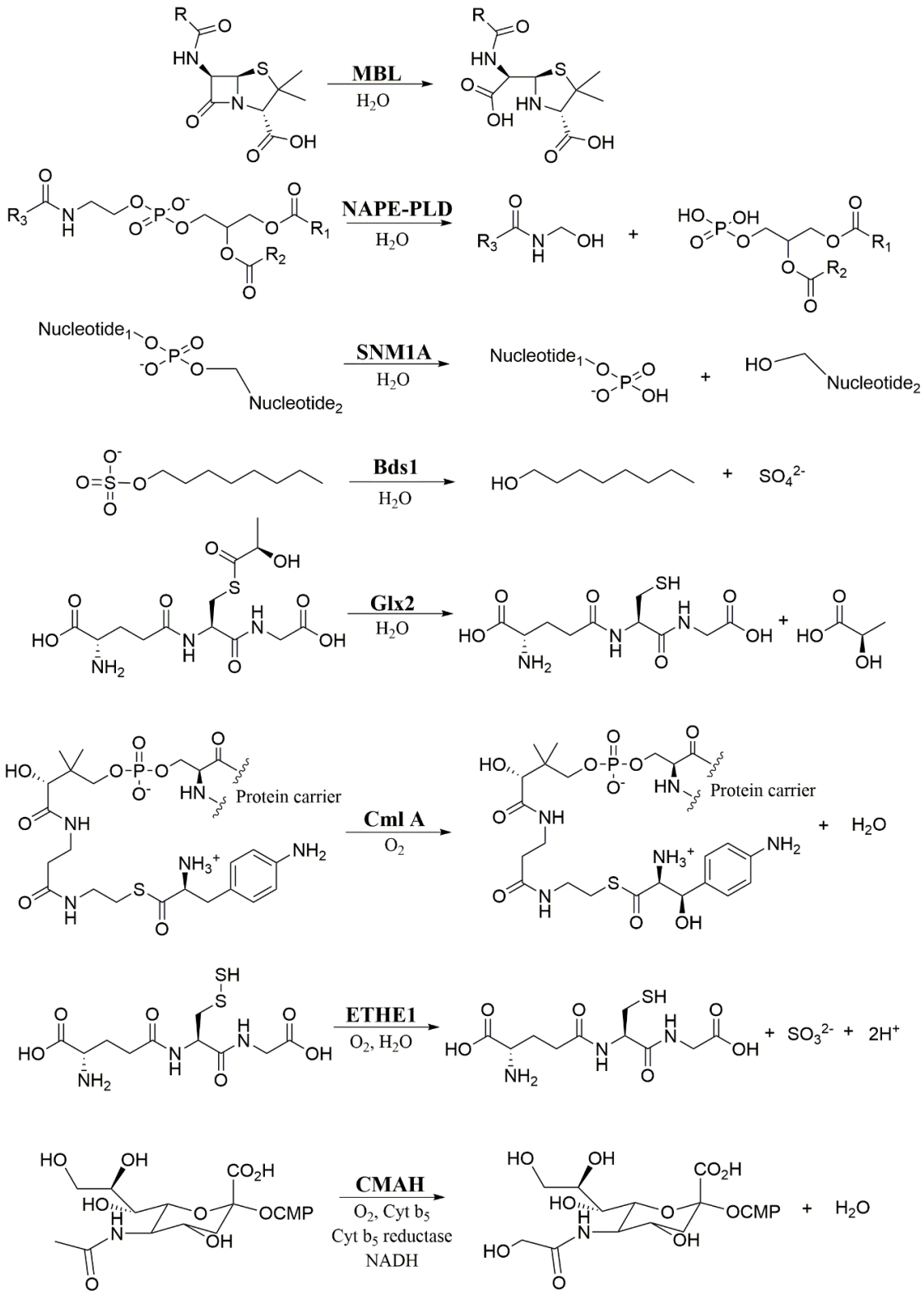


Figure 1.2: The range of reactions catalysed by members of the MBL superfamily. MBL refers to metallo-β-lactamases. NAPE-PLD - *N*-acyl phosphatidylethanolamine phospholipase D; SNM1A – sensitive to nitrogen mustard 1; Bds1 – bacterially-derived sulfatase 1; Glx2 – glyoxalase II; CmlA - 4-amino-L-phenylalanyl-[CmlP-peptidyl-carrier-protein] 3-hydroxylase; ETHE1 - Ethylmalonic encephalopathy 1 protein; CMAH – cytidine monophosphate *N*-acetylneuraminic acid hydroxylase. Cyt b_5 – cytochrome b_5, CMP – cytidine monophosphate.

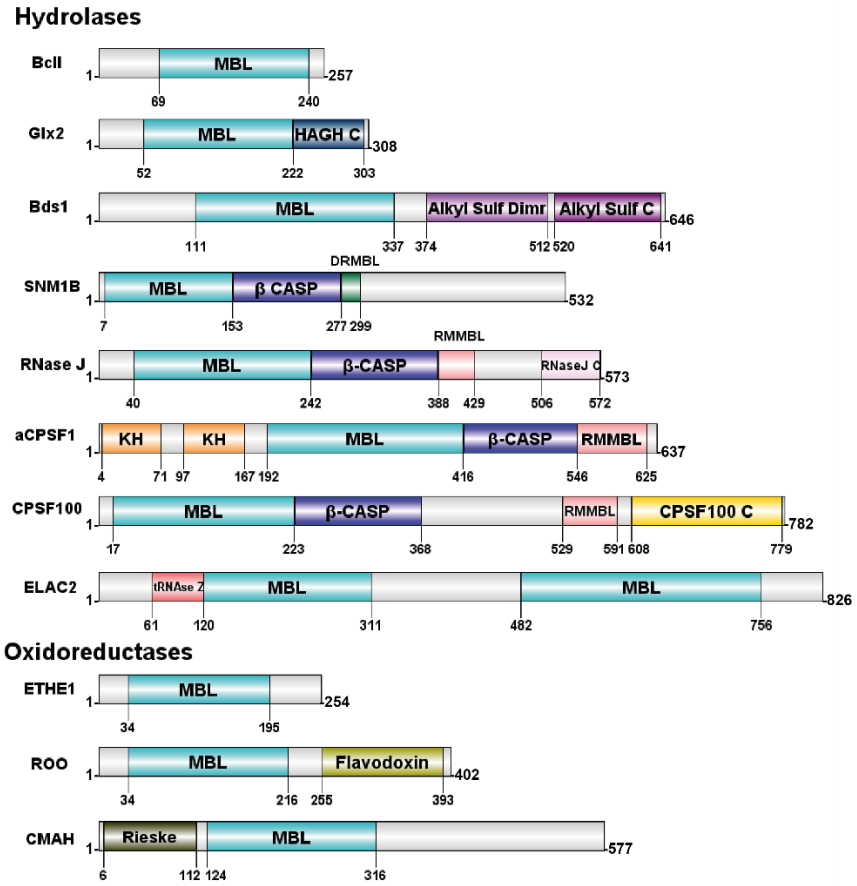


Figure 1.3: Domain architectures of selected MBL superfamily enzymes. Numbers correspond to amino acid residues in proteins of organisms specified as predicted by InterProScan[22]. BcII – *Bacillus cereus* β-lactamase II, Glx2 – human glyoxalase II, Bds1 – *Saccharomyces cerevisiae* bacterially derived sulfatase 1, aCPSF1 – archaeal (*Methanosarcina mazei*) CPSF100 protein, CPSF100 – human CPSF100; ELAC2 – human long form tRNase Z, ETHE1 -human Ethylmalonic encephalopathy 1 protein, ROO – *Desulfovibrio gigas* rubredoxin oxygen:oxidoreductase, CMAH – mouse cytidine monophosphate *N*-acetylneuraminic acid hydroxylase. HAGH C – *C*-terminal domain of hydroxyacylglutathione hydrolase (glyoxalase II). Alkyl Sulf Dimr – alkyl sulfatase dimerization domain. Alkyl Sulf C – alkyl sulfatase *C*-terminal domain. DRMBL – DNA repair MBL domain. RMMBL – RNA specificity domain. RNase J C – RNase J *C*-terminal domain. KH – K Homology domain. CPSF100 C – Cpsf 100 *C*-terminal domain. Figure was created using DOG 2.0 software[23].

Daiyasu et al. have previously proposed assignment of the MBL superfamily members into 16 groups based on their biological function.[17] However, as numerous MBL proteins currently have no assigned functions, more groups will likely be added in the future.[17, 20]

Table 1.1: Summary of the proteins with known three-dimensional structures described in this chapter. PDB and UniProtKB[24] accession codes are listed. References are given for the structures shown. N – no reference available.

Protein name	Organism	Function	PDB ID	Uniprot ID	Ref.
BcII	*Bacillus cereus*	β-lactamase (B1)	4C09	P04190	[25]
Sfh-1	*Serratia fonticola*	β-lactamase (B2)	3SD9	Q9RMI1	[26]
L1	*Stenotrophomonas maltophilia*	β-lactamase (B3)	1SML	P52700	[27]
SNM1A	*Homo sapiens*	DNA repair nuclease	4B87	Q6PJP8	[28]
SNM1B	*Homo sapiens*	DNA repair nuclease	5AHO	Q9H816	[28]
aCPSF1	*Methanosarcina mazei*	pre-mRNA endonuclease	2XR1	Q8PZ03	[29]
CPSF73	*Homo sapiens*	pre-mRNA endonuclease	2I7T	Q9UKF6	[30]
RNase J	*Thermus thermophilus*	RNA processing nuclease	3BK1	Q72JJ7	[31]
tRNase Z	*Bacillus subtilis*	tRNA maturation nuclease	4GCW	P54548	[32]
LACTB2	*Homo sapiens*	mitochondrial RNA nuclease	4AD9	Q53H82	[33]
MBLAC1	*Homo sapiens*	pre-mRNA nuclease	4V0H	A4D2B0	[34]
NAPE PLD	*Homo sapiens*	phospholipase	4QN9	Q6IQ20	[35]
Pce	*Streptococcus pneumoniae*	teichoic acid phosphorylcholine esterase	2BIB	Q8DQ62	[36]
MPH	*Pseudomonas* sp. (strain WBC-3)	Methyl parathion hydrolase	1P9E	Q841S6	[37]
OPHC2	*Pseudomonas oleovorans*	organophosphorus hydrolase	4LE6	Q5W503	[38]
SdsA1	*Pseudomonas aeruginosa*	alkylsulfatase	2CFU	Q9I5I9	[39]
Glx2	*Homo sapiens*	Thioesterase (methylglyoxal detoxification)	1QH5	Q16775	[18]
PqsE	*Pseudomonas aeruginosa*	2-aminobenzoylacetyl-CoA thioesterase	2Q0I	P20581	[40]
UlaG	*Escherichia coli*	lactonase	2WYM	P39300	[41]
ROO	*Desulfovibrio gigas*	Oxygen reductase	1E5D	Q9F0J6	[19]
FlRb	*Escherichia coli*	NO reductase	4D02	Q46877	[42]
CmlA	*Streptomyces venezuelae*	β-hydroxylase in chloramphenicol biosynthesis	5KIK	F2RB80	[43]
ETHE1	*Homo sapiens*	Persulfide dioxygenase	4CHL	O95571	[44]
PNGM-1	Deep sea sediment uncultured bacteria	β-lactamase and tRNase	6J4N	A0A2U8UYM6	[45]
Tm1162	*Thermotoga maritima*	β lactamase and phosphodiesterase	3X30	Q9X0P5	[46]
PRF	*Burkholderia phytofirmans*	sulfur assimilation	5VE3	B2TEQ2	[47]
Tm0894	*Thermotoga maritima*	unknown	1ZTC	Q9WZZ6	N
Bphy_0182	*Burkholderia phytofirmans*	unknown	6AO1	B2JKI6	N
SACOL0046	*Staphylococcus aureus*	unknown	3R2U	A0A0H2X0Y8	N
Tflp	*Caldanaerobacter subterraneus* (subsp. *tengcongensis*)	unknown	2P4Z	Q8R8V2	[48]

1.2.1 Metallo-β-lactamases

Lending their name to the superfamily, the 'true' MBLs are, arguably, its best studied members. They constitute class B of the β-lactamases (classes A, C and D being comprised of serine-β-lactamases) and are further divided into B1, B2 and B3 subclasses based on their sequence alignment.[16, 49, 50] While the sequences between the subclasses, as well as within, vary, the overall MBL structural fold is highly conserved (**Figure 1.4A**).[9]

The B1 subclass is the largest of the three MBL subclasses. B1 MBLs possess a broad catalytic activity targeting penicillins, cephalosporins and carbapenems.[51] B1 MBLs share around 30% sequence identity and are mostly encoded on mobile genetic elements.[9] Their active site contains two Zn binding sites – Zn1 bound by three histidines, His116, His118, and His196, and Zn2 coordinated by Asp120, Cys221, and His263. The Zn ions are considered to act as Lewis acid. In addition, there are two water molecules bound, one bridging between the two Zn ions and one coordinating to Zn2, both of which are involved in the hydrolysis of the substrate (**Figure 1.4B**).[52]

The B2 subclass is much smaller than the B1 and possesses a narrow substrate range with a selectivity for carbapenems.[51, 53] The most significant difference between the B1 and B2 classes results from a mutation of His116 to Asn116, which reduces the affinity of one of the metal binding sites leading to B2 MBLs normally binding a single Zn ion in their active state.[54] Binding of the second Zn ion has been reported to lead to inhibition of the catalytic activity.[54, 55] Similarly to the B1 subclass, there are two water molecules bound at the active site of B2 enzymes with one coordinating the bound Zn ion and the other is positioned at the second metal binding site, coordinated to the His118 (**Figure 1.4A**). These water molecules are involved in substrate hydrolysis (**Figure 1.4C**) thus potentially explaining inhibition of B2 MBLs by the binding of a second Zn ion as this would make residues required for water coordination unavailable.[55]

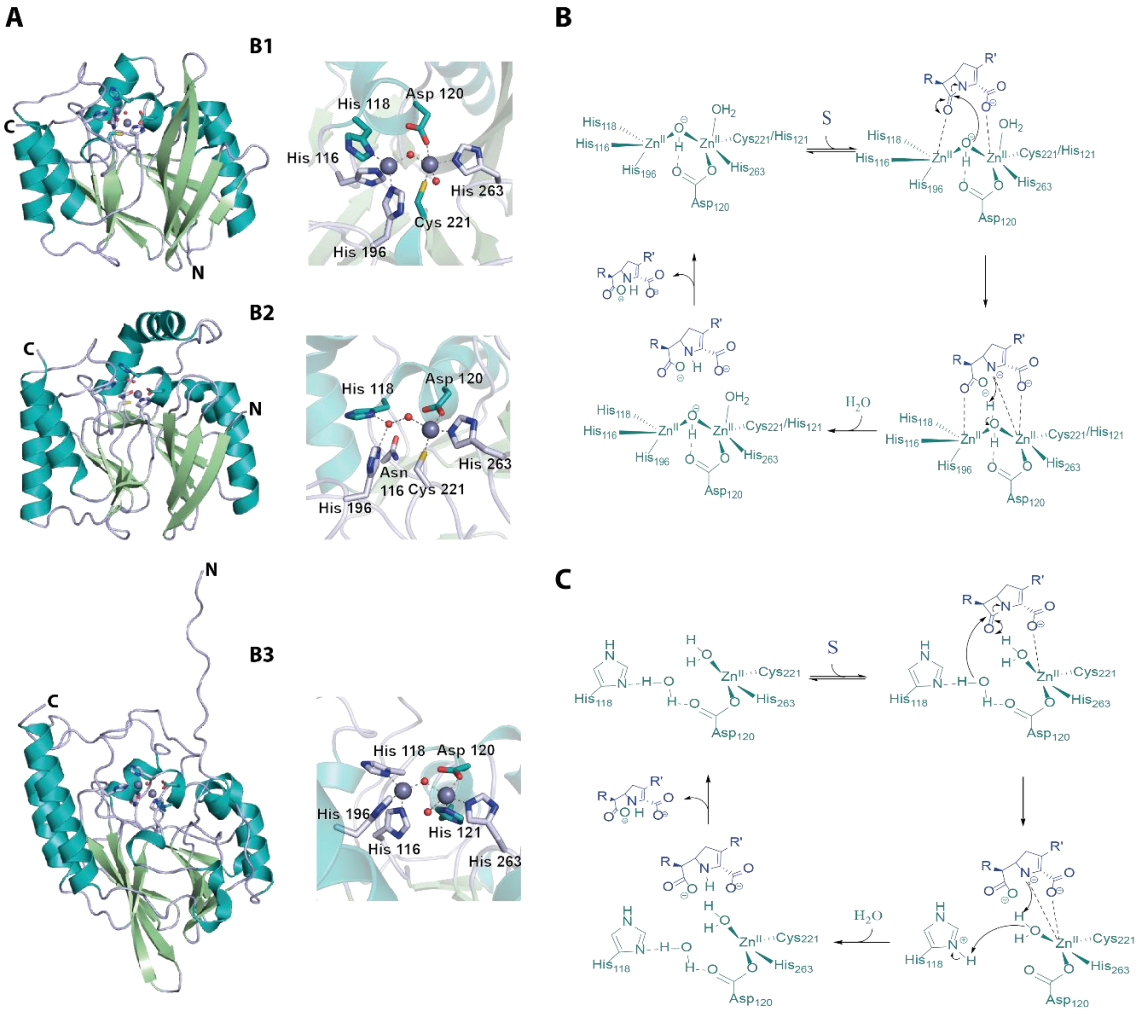


Figure 1.4: (A) Structures and active site views of B1, B2 and B3 subclasses of MBLs, as exemplified with BcII (PDB: 4C09)[25], Sfh-1(PDB: 3SD9)[26] and L1 (PDB: 1SML)[27], showing the conserved active site residues and the αββα fold. Active site Zn ions are shown as grey spheres, red spheres denote water molecules. Residues are numbered using BBL numbering. (B) Proposed reaction mechanism for carbapenem hydrolysis by B1 and B3 MBLs. (C) Proposed reaction mechanism for carbapenem hydrolysis by B2 MBLs.

The MBL subclass B3 is diverged from the other two classes and is proposed to have evolved β-lactamase activity separately from the B1 and B2 enzymes.[56] B3 MBLs typically have two Zn ions bound at the active site.[57, 58] However, unlike B1 MBLs, B3 MBLs do not have Cys221, which is substituted for Ser221, and is no longer involved in metal binding. Instead, in the B3 MBLs His121 is involved in ligand binding (**Figure 1.4A**).[27, 57, 59] Examples of the B3 subclass include GOB-1 from *Elizabethkingia meningoseptica*, which has been reported to be active with both one and two Zn ions bound, as well as containing an additional His116 to Gln116 mutation which further modifies the metal binding site[57, 60]. L1 from *S. maltophilia* which is a tetramer, while the rest of the 'true' MBLs are monomeric.[8, 58] Mechanistically, the B3 MBL enzymes do not appear to differ from the B1 subclass (**Figure 1.4B**), and, like B1, also have a broad substrate range.[51]

1.2.2 Nucleic acid processing MBL fold enzymes

Many of the MBL superfamily members have nucleic acid processing activities including DNA crosslink repair proteins and ribonucleases working on different types of RNA (3-dimensional structures of select MBL nucleases are shown in **Figure 1.5**).[17, 61] A subfamily of MBL nucleases contains a β-CASP domain (shown in blue in **Figure 1.5**) inserted between conserved motifs (iv) Asp221 and (v) His263 near their *C*-terminus.[62] Members of the β-CASP family include both DNA and RNA nucleases.

β-CASP DNA nucleases appear to only be present in eukaryotes and include the DNA crosslink repair proteins PSO2 (psoralen-sensitive) and SNM1 (sensitive to nitrogen mustard).[62, 63] Three proteins - SNM1A, B and C – have been identified in vertebrates while only a single protein appears to be present in other eukaryotes[63] (structures of SNM1A and SNM1B are shown in **Figure 1.5A**). Human SNM1A and its functional homolog in yeast, PSO2, are reported to be involved in interstrand crosslink repair.[64, 65] Human SNM1A (DCLRE1A) has been reported to possess 5'-3' exonuclease activity *in vitro*[66], as well as structure-specific endonuclease activity *in vitro* and in yeast cells[67]. Human SNM1B (DCLRE1B, Apollo) possesses 5'-3' exonuclease activity and is reported to be involved in telomere maintenance.[68, 69] Depletion of both SNM1A and SNM1B results in accumulation of interstrand crosslinks and increased sensitivity to mutagens, such as ionising radiation or cisplatin[66, 70, 71]. Human SNM1C (DCLRE1C or Artemis) is both an exonuclease acting on single and double strand DNA, and an endonuclease which acts on DNA hairpin structures.[72, 73] SNM1C plays a crucial role in the NHEJ (non-homologous end joining) pathway involved in double-strand break repair and is also involved in T-cell diversification.[72, 74, 75]

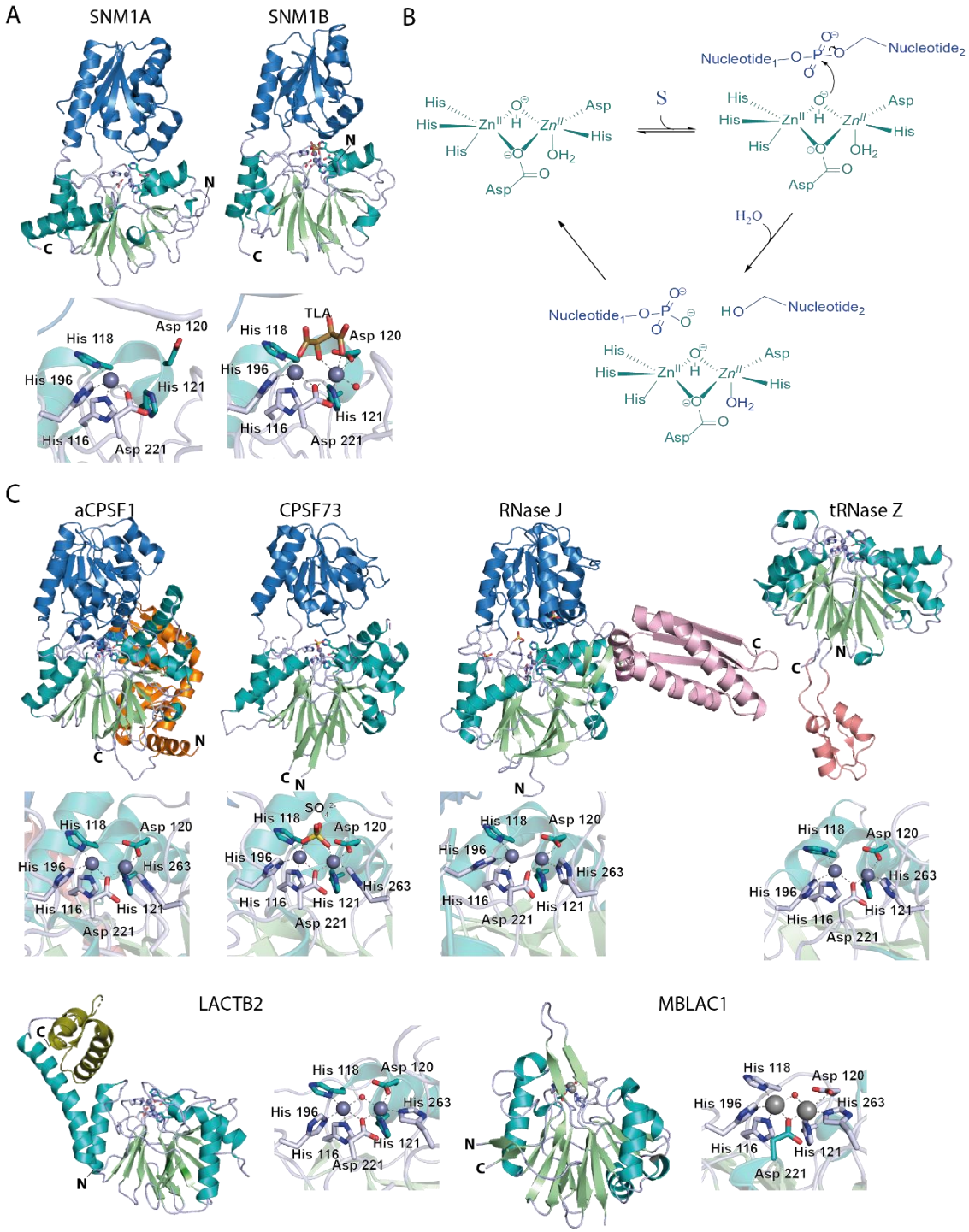


Figure 1.5: Nucleic acid processing enzymes of the MBL superfamily. (A) Views from crystal structures of human DNA repair proteins SNM1A[28] (PDB: 4B87) and SNM1B[28] (PDB: 5AHO) with magnified views of the active sites. (B) Proposed outline reaction mechanism of SNM1 nucleases. (C) Structures of RNA nucleases: Cleavage and polyadenylation specificity factor 1 from archaeon *Metanosarcina mazei*[29] (aCPSF1, PDB: 2XR1), human CPSF73[30] (PDB: 2I7T), *Thermus thermophilus* RNase J[31] (PDB: 3BK1), *Bacillus subtilis* tRNase Z[32] (PDB: 4GCW), human mitochondrial endoribonuclease LACTB2[33] (PDB: 4AD9), and human histone pre-mRNA endonuclease MBLAC1[34] (PDB: 4V0H). Zn ions are shown as grey spheres, water is show as red spheres. β-CASP (blue), KH (orange), RNase J *C*-terminal (pink), tRNase Z (salmon), and LactB2 winged helix (olive) domains are shown. Note that proteins are shown as monomers for simplicity. Residues are numbered using standard BBL numbering. TLA – L-tartrate.

Known RNA nucleases of the β-CASP family include the Cleavage and Polyadenylation Specificity Factor (CPSF) 73 kDa and 100 kDa subunits found in eukaryotes and archaea, animal Integrator Complex Subunits (INTS) 9 and 11, and prokaryotic RNase J.[61, 62, 76]

CPSF is a multiprotein complex formed in eukaryotic cells, which is involved in pre-mRNA processing.[30, 77, 78] CPSF complex is responsible for polyadenylation site definition, pre-mRNA cleavage and recruitment of a poly(A) polymerase.[79] Eukaryotic CPSF73 possesses both endo- and exonuclease activities and is reported to be involved in histone pre-mRNA processing[80]. CPSF100 lacks ribonuclease activity, which is likely due to the absence of catalytic residues within the HXHXDH motif and its inability to bind metal ions.[30] CPSF73 and CPSF100 are reported to form a heterodimer which is reported to enable substrate binding, suggesting that CPSF100 likely plays a regulatory role.[81] Archaeal CPSF proteins can be separated into 3 groups – aCPSF1, 1b and 2.[82] The largest of the groups, aCPSF1 is reported to be orthologous to CPSF73 and appears to be present in all archaea.[82] Several aCPSF proteins of all groups have been reported to possess endo- and/or exonuclease activities[82-85], while sequence analysis and structural studies have shown that aCPSF1 proteins contain two *N*-terminal K homology (KH) domains (shown in orange in **Figure 1.5C**).[29, 83, 86]

INTS9 and INTS11 are part of an integrator complex which is involved in regulating gene expression by hydrolysing small nuclear (sn) RNAs.[87] Similarly to CPSF100, INTS9 is reported to be catalytically inactive, while, ITS11, like CPSF73, possesses endonuclease activity and forms a complex with INTS9.[88, 89]

RNase J is found in prokaryotes and contains a characteristic *C*-terminal domain (highlighted in pink in **Figure 1.5C**).[31] Some bacteria possess a single RNase J, while others produce two distinct paralogs – RNase J1 and RNase J2.[90] Bacterial RNase J paralogs possess endo- or both exo- and endonuclease activities and are involved in RNA maturation and degradation, while archaeal RNase J proteins have been reported to possess distinct endo- and exonuclease activities.[84]

Ribonuclease activity in the MBL superfamily is not limited to the β-CASP family. Other proteins involved in RNA processing are tRNase Z and its eukaryotic homologs commonly known as ELAC1 and 2, mitochondrial ribonuclease LACTB2, and histone mRNA endonuclease MBLAC1.[17] tRNase Z generates the mature 3'-end of tRNA and can be found in short dimeric form (prokaryotes and ELAC1 in eukaryotes) or long form (ELAC2 in eukaryotes) containing two MBL domains, likely resulting from a gene fusion event.[91] In eukaryotes, ELAC1 is found in the cytoplasm, while ELAC2 is present in the nucleus and the mitochondria.[92] LACTB2 is

another mitochondrial ribonuclease cleaving single stranded RNA.[33] MBLAC1 is a metazoan protein involved in 3' end processing of histone mRNA.[34]

1.2.3 Other MBL fold hydrolases

Most MBL superfamily enzymes are hydrolases. In addition to β-lactams and nucleic acids, numerous other substrates are used, which contain hydrolysable groups such as, among others, thioesters, sulfates, phosphoesters, and esters. Several distinct hydrolases have been characterised structurally (**Figure 1.6**).

In addition to the phosphodiesters of nucleic acids, other phosphate containing substrates are hydrolysable by the MBL superfamily. *N*-Acylphosphatidylethanolamine-hydrolysing phospholipase D (NAPE-PLD), a membrane-associated phospholipase, hydrolyses *N*-acylphosphatidylethanolamines and is reported to be involved in regulation of lipid metabolism in animal tissues.[35, 93] Teichoic acid phosphorylcholine esterase Pce is a bacterial cell wall bound MBL esterase found in *Streptococcus pneumoniae* involved in host-pathogen interactions and pathogenesis (cell wall binding domain shown in red in **Figure 1.6**).[36, 94] Pce releases phosphorylcholine from cell wall teichoic or lipoteichoic acids and is involved in colonisation of the host nasopharynx.[94, 95] Smaller organic phosphates, which are often neurotoxic and are commonly used as agricultural pesticides, can also be hydrolysed by MBL family proteins.[96] Some soil-dwelling microorganisms are capable of utilising organophosphates as carbon and phosphorus sources. MBL enzymes methyl parathion hydrolase (MPH) and organophosphate hydrolase 2 (OPH2) are involved in the breakdown of organophosphates and have been characterised structurally (**Figure 1.6**).[37, 38]

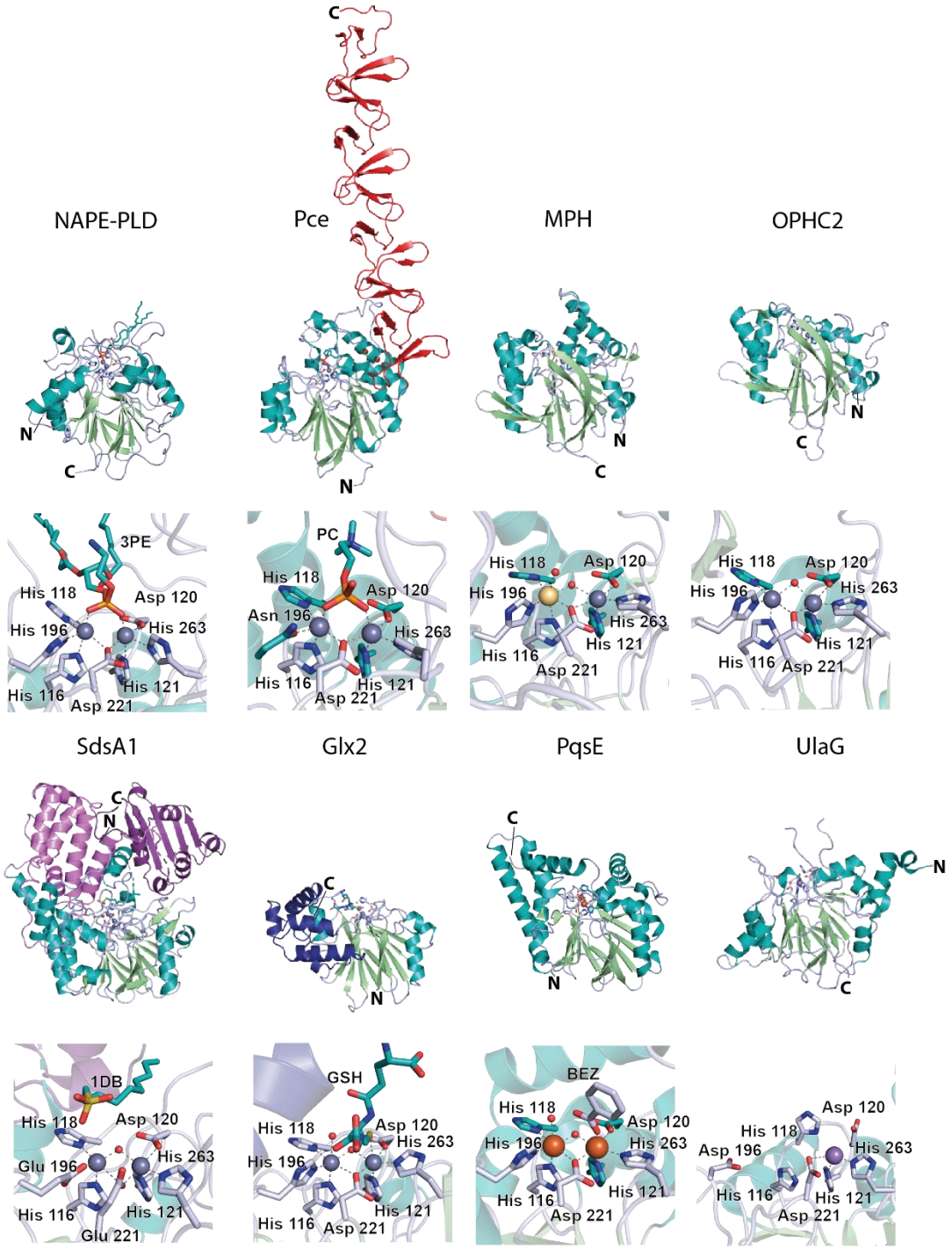


Figure 1.6: Structures of MBL superfamily hydrolases: human *N*-acyl phosphatidylethanolamine phospholipase D (NAPE-PLD, PDB: 4QN9)[35], *S. pneumoniae* teichoic acid phosphorylcholine esterase Pce (PDB: 2BIB)[36], *Pseudomonas* sp. (strain WBC-3) methyl parathion hydrolase (MPH, PDB: 1P9E)[37], *Pseudomonas oleovorans* organophosphorus hydrolase OPHC2 (PDB: 4LE6)[38], *Pseudomonas aeruginosa* alkylsulfatase SdsA1 (PDB: 2CFU)[39], human Glx2 (PDB: 1QH5)[18], *P. aeruginosa* quinolone signal response protein PqsE (PDB: 2Q0I)[40], and *E. coli* lactonase UlaG (PDB: 2WYM)[41]. 3PE – phosphatidylethanolamine, PC – phosphorylcholine, 1DB – decane-1-sulfonic acid, GSH – glutathione, BEZ – benzoic acid. Zn (grey), Fe (orange), Mn (purple), Cd (yellow), and Ca (brown) ions are shown as spheres. Water molecules are shown as red spheres. Cell wall binding repeat (red), alkylsulfatase dimerisation (lilac), alkylsulfatase *C*-terminal (dark purple), and glyoxalase II *C*-terminal (dark blue) domains are highlighted.

Sulfatases or sulfohydrolases hydrolyse sulfate esters and are found in all domains of life[97]. Sulfatases are grouped into three established classes and existence and differing definition of the fourth group has been proposed several times to date[98, 99]. Sulfatase class 3 is comprised of alkylsulfatase members of the MBL superfamily and contains homologues of *Pseudomonas aeruginosa* SdsA1 hydrolysing sodium dodecyl sulfate (SDS).[39] SdsA1 (**Figure 1.6**) and three homologues - pseudomonal Pisa1 (PDB: 2YHE)[100], SdsAP (PDB: 4NUR, unpublished) and *E. coli* YjcS (PDB: 4PDX, unpublished) have been characterised structurally and found to be comprised of two other domains in addition to the MBL domain – alkylsulfatase dimerisation (shown in lilac) and alkylsulfatase *C*-terminal (shown in dark purple). While class 3 sulfatases exist in eukaryotes, none have been identified in humans so far, and only one eukaryotic MBL sulfatase – Bds1 from *Saccharomyces cerevisiae* – has been studied to date[101]. Another member of the MBL superfamily – arylsulfatase AtsA from *Pseudoalteromonas carrageenovora* has low sequence similarity with SdsA1 and related sulfatases and hence has been proposed to form a separate class of sulfatases[98, 102]. Few studies on AtsA and related sulfatases have been reported in the literature to date.

Glx2 was the first discovered non-β-lactamase member of the MBL superfamily. Glx2 is a thioesterase involved in methylglyoxal (MG) detoxification and regulating glutathione (GSH) levels in mitochondria.[103, 104] Glx2 enzymes are ubiquitous in nature and contain a unique *C*-terminal domain comprised of 5 α-helices (shown in blue in **Figure 1.6**) in addition to the MBL domain[18]. Several bacterial species also encode an isoenzyme of Glx2 named YcbL or Glx2-2 which contains an additional α-helix, truncation of the *C*-terminal domain and a mononuclear metal centre as opposed to a binuclear centre observed in Glx2[105, 106]. YcbL enzymes have been reported to possess Glx2 activity *in vitro* and exhibit an additive effect on Glx2 activity and MG detoxification.[105, 106]

The pseudomonal quinolone signal E (PqsE) protein is another MBL thioesterase. PqsE hydrolyses 2-aminobenzoylacetyl-coenzyme A to form 2-aminobenzoylacetate during the biosynthesis of the alkylquinolone signalling molecules used by the cells to regulate expression of virulence genes.[40, 107]

Lactones are also subject to hydrolysis by members of the MBL superfamily. Several MBL lactonases have been characterised and are involved in a variety of processes including quorum quenching[108-111], as well as ascorbate[41] and vitamin B_6 degradation[112, 113]. UlaG (**Figure 1.6**) is a Mn-dependant L-ascorbate-6-P lactonase and is a founding member of a UlaG-like family consisting of, at present, poorly characterised bacterial and archaeal proteins.[114]

1.2.4 MBL fold oxidoreductases

In addition to the wide range of hydrolases, the MBL superfamily also contains oxidoreductases. Unlike most of the hydrolases which, with some exceptions, mostly bind Zn ions, oxidoreductases have one or two Fe ions bound in their active sites.

A subfamily of MBL oxidoreductases is the flavodiiron family, which includes O_2 and NO reductases and is present in all domains of life.[115] Flavodiiron proteins (FDPs) comprise flavodoxin domain (shown in green in **Figure 1.7A**) in addition to the MBL domain. Some members are demonstrated or predicted to carry additional domains at their *C*-termini, such as rubredoxin-like, Fer4_19 [FeS] cluster binding, NAD(P)H: flavin oxidoreductase-like, and NADH: rubredoxin oxidoreductase-like domains.[115] ROO, the first oxidoreductase MBL for which a 3-dimensional structure has been determined (**Figure 1.7A**), is an FDP and has been reported to possess both O_2 and NO reductase activities.[19, 116] FDPs with clear preference for O_2 (e.g. *Methanobrevibacter arboriphilus* FprA[117], and EhFdp1 of protozoan *Entamoeba histolytica*[118]) or NO (e.g. *E. coli* FlRb[42], **Figure 1.7**) have also been reported.

Cytidine monophosphate *N*-acetylneuraminic acid hydroxylase (CMAH) is a monooxygenase member of the MBL superfamily found in the deuterostomes. In addition to the MBL domain, CMAH comprises a Rieske domain containing a Rieske [2Fe2S] cluster.[119] CMAH catalyses hydroxylation of sialic acid cytidine monophosphate *N*-acetylneuraminic acid (CMP-Neu5Ac) to CMP-N-glycolylneuraminic acid (CMP-Neu5Gc) in presence of cytochrome b_5 (cyt b_5), cyt b_5 reductase and NADH.[120, 121]

CmlA is a β-hydroxylase involved in the antibiotic chloramphenicol biosynthesis and catalyses the hydroxylation of L-*p*-aminophenylalanine (**Figure 1.2**).[122] CmlA is present in *Streptomyces venezuelae*, but sequence analyses have revealed several bacterial homologues with some found in gene clusters coding for antibiotic biosynthesis pathways involving β-hydroxylation of amino acids.[122] Determination of the 3-dimensional structure of CmlA has revealed a previously unseen *N*-terminal domain (shown in magenta in **Figure 1.7A**) which has been reported to enable dimerization and hypothesised to be involved in nonribosomal peptide synthetase recognition and O_2 activation upon substrate binding.[43, 123]

While presence of additional domains is widespread within the MBL oxidoreductases, some members, such as Ethylmalonic encephalopathy 1 (ETHE1) protein, contain only an MBL domain.[124] ETHE1 has been structurally and functionally characterised in humans and *Arabidopsis thaliana*.[44, 124-126] ETHE1 is a dimeric mitochondrial enzyme with a single Fe ion bound at the active site and is involved in detoxification of H_2S (proposed mechanism is shown

in **Figure 1.7**).[124, 125] Mutations in human ETHE1 correlate with manifestation of Ethylmalonic encephalopathy, from which it gains its name.[127] *A. thaliana* ETHE1 has been reported to be crucial for endosperm and embryo development.[126]

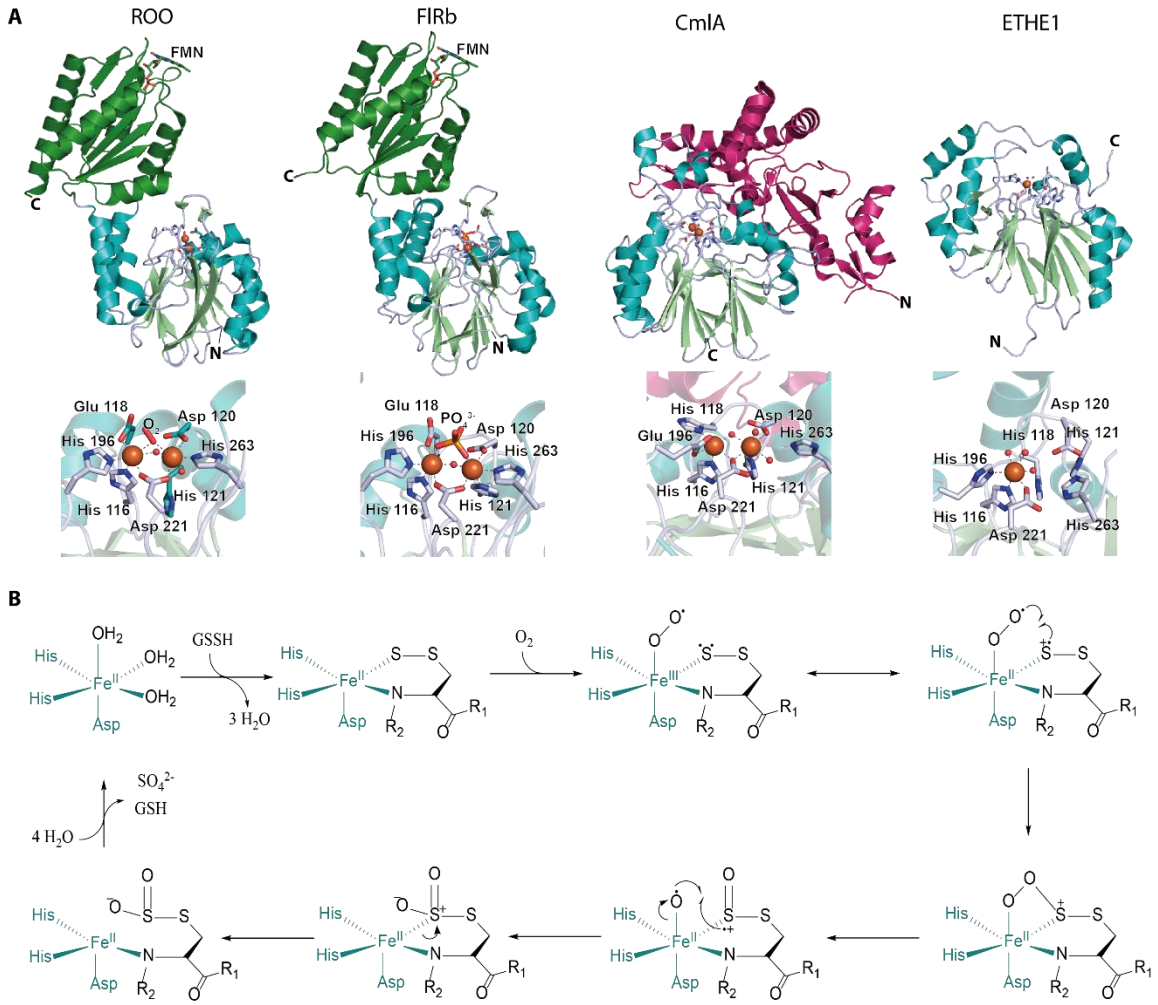


Figure 1.7: (A) Views from crystal structures of oxidoreductase members of the MBL superfamily: *D. gigas* ROO[19] (PDB: 1E5D), *E. coli* anaerobic NO reductase FlRb[42] (PDB: 4D02), *Streptomyces venezuelae* β-hydroxylase CmlA[43] (PDB: 5KIK), and human persulfide dioxygenase ETHE1[44] (PDB: 4CHL). Flavodoxin (dark green) and CmlA *N*-terminal domain (magenta) are highlighted. Active site Fe ions are shown as orange spheres, water molecules are shown as red spheres. Proteins are shown as monomers for simplicity. (B) Proposed outline EHTE1 reaction mechanism[44, 125]. Active site residues and Fe ion are shown in teal.

1.2.5 Enzymes with multiple and unknown functions

Members of the MBL superfamily have been reported to have promiscuous catalytic activities with enzymes being proposed to catalyse an average of 2.5 reactions.[128] High promiscuity may complicate identification of MBL proteins with low sequence similarity to known members and only a small fraction of the predicted members of the MBL superfamily enzymes have been studied. A large proportion of the predicted MBL proteins have no assigned function, many of

which also possess domains of unknown functions in addition to the MBL domain, making functional predictions harder.

Recently a new member of the MBL superfamily PNGM-1 (Papua New Guinea MBL-1) has been identified from deep-sea sediment metagenome.[21, 45, 129] PNGM-1 was found to possess β-lactamase and tRNase activities, despite lacking the tRNase Z domain (**Figure 1.8**).[21, 129] Another hydrolase, Tm1162 from *Thermotoga maritima*, was reported to exhibit both β-lactamase and phosphodiesterase activities and found to adopt a previously unseen within the MBL superfamily trimeric quaternary structure.[46] Sometimes dual activities may arise from the different domains. For example, Persulfide dioxygenase-rhodanese fusion protein (PRF) from *Burkholderia phytofirmans* has been proposed to be involved in sulfur assimilation with rhodanese domain catalysing sulfur transfer from thiosulfate to GSH to form sulfite and GSSH, and the MBL domain (persulfide dioxygenase) further oxidising GSSH to sulfite.[47]

In addition to proteins of dual functions, several MBL proteins of unknown functions have been characterised structurally, with crystal structures being available via the PDB with catalytic function yet to be assigned (**Figure 1.8**). One such structure that stands out is that of Tflp from a thermophilic anaerobic bacterium *Caldanaerobacter subterraneus* (subspecies *tengcongensis*), a ferredoxin-like protein carrying a unique motif consisting of 4 antiparallel β-sheets reported to form an [FeS] cluster, which is proposed to be formed from the active site Fe ions and the nearby Cys residues (PDB: 2P4Z, **Figure 1.8**).[48] Based on sequence analysis, Tflp appears to be one of few non β-lactamase MBLs to be more closely related to the B1 and B2 MBLs than the B3 subclass. In animals, PNKD (Paroxysmal non-kinesigenic dyskinesia) and MBLAC2 are examples of MBL superfamily proteins of currently unknown functions.[61] PNKD exists in four isoforms, three of which contain an MBL domain, and has been proposed to be involved in the stress response pathway and maintenance of cellular redox environment but the nature of the physiological reaction is yet to be determined.[130, 131] MBLAC2 has been reported to hydrolyse penicillin *in vitro* and in cells, but the relevance of this activity remains unclear.[132]

With the number of predicted members of the MBL superfamily far exceeding the number of characterised members, the full scope of reactions catalysed by MBL fold enzymes remains to be discovered.

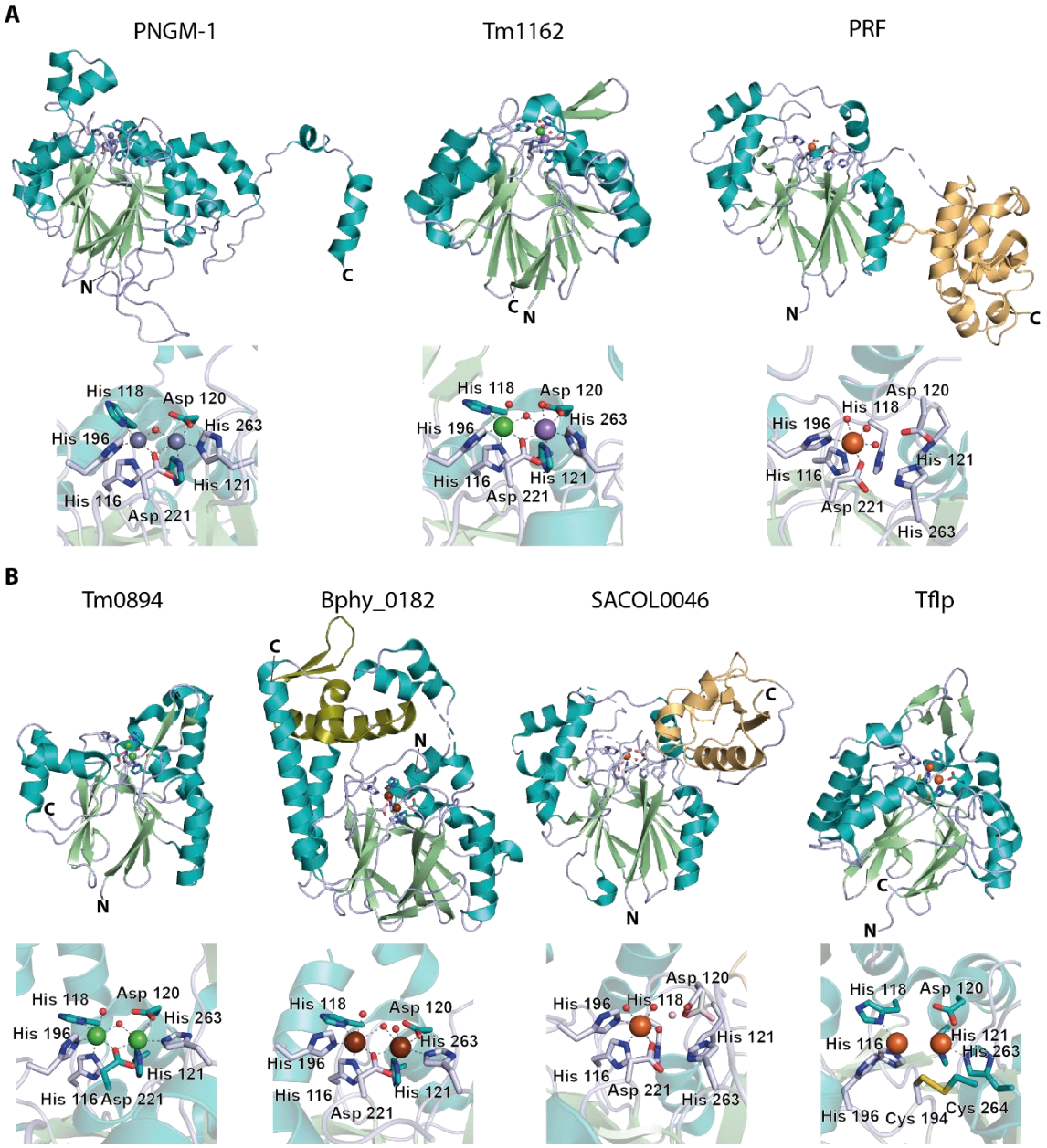


Figure 1.8: Views from crystal structures and magnified views of the active sites of (A) MBL superfamily members reported to show dual catalytic activities: PNGM-1 from deep-sea sediment uncultured bacteria (PDB: 6J4N)[45], *Thermotoga maritima* Tm1162 (PDB: 3X30)[46], and *Burkholderia phytofirmans* PRF (PDB: 5VE3)[47], and (B) MBL superfamily proteins with unknown functions: *Thermotoga maritima* Tm0894 (PDB: 1ZTC, unpublished), *Burkholderia phytofirmans* Bphy_0182 (6AO1, unpublished), *Staphylococcus aureus* SACOL0046 (PDB:3R2U, unpublished), and *Caldanaerobacter subterraneus* (subsp. *tengcongensis*) Tflp (PDB: 2P4Z)[48]. Rhodanese (beige) and winged helix (olive) domains are highlighted. For proteins where no name is available, name of the gene encoding the protein is used instead. Zn (grey), Ni (green), Mn (purple), Fe (orange), and Ca (dark brown) ions are shown as spheres, water molecules are shown as red spheres. Alternative conformations of Asp120 and water molecule in SACOL0046 are shown as faded. Residues are numbered using BBL system. Note that Cys residues in Tflp are numbered according to their position relative to the nearest active site His residue.

1.3 Metal binding within the MBL superfamily

The vast majority of the MBL superfamily proteins bind one or two metal ions in their active sites. Zn is the most common metal found in the MBL hydrolases, while Fe is found in all characterised MBL fold oxidoreductases. MBL hydrolases with different active site metals include the aforementioned UlaG, which is a Mn^{2+}-dependant lactonase[41] and a hydrolase with reported β-lactamase and phosphodiesterase functions Tm1162 which has been reported to contain a heterodinuclear metal centre containing Ni^{2+} and Mn^{2+} ions[46].

Several MBL proteins are reported to purify with a mixture of metal ions bound, especially when produced as recombinant proteins using *E. coli* expression system.[34, 133] For example, MBLAC1has been reported to bind a mixture of Zn, Fe and Mn when produced in *E. coli* cells, but show a preference for Zn when produced in mammalian cells with protein produced in both systems being catalytically active.[34] PqsE, a thioesterase, has been reported to contain two Fe ions in its active site, however it was proposed to be an artefact of heterologous expression.[40] In several different studies, *S. pneumoniae* Pce has been reported to bind solely Zn or Fe ions in the MBL domain and be catalytically active in both instances.[36, 94, 134]

In some cases, possessing a heterodinuclear centre was reported to be required for activity.[135] Glx2 from *A. thaliana* has been suggested to require the presence of both Fe and Zn ions in its active site for substrate binding, transition state stabilisation and hydrolysis. [135]

Several MBL proteins have been reported to be able to utilise different transition metals to perform their native function. Metal reconstitution experiments have been performed with 'true' MBLs BcII, VIM-2 and L1[136-139], Glx2 from *Salmonella typhimurium*[140, 141] and *E. coli*[142], *B. subtilis* phosphodiesterase YhfI[143], *P. aeruginosa* MPH, *E. coli* tRNase Z, and *Pseudoalteromonas carrageenovora* arylsulfatase AtsA[137]. In addition to varying levels of activity, different inhibition profiles were observed in 'true' MBLs and *S. typhimurium* Glx2 depending on the identity of the metal ions bound in the active site.[138, 144] Structural studies on *Pseudomonas putida* pyrroloquinoline quinone biosynthesis protein PqqB with Mn, Zn, Mg and Cu ions suggest metal binding plasticity in the MBL active site with different residues observed coordinating the mononuclear metal ion in each metal bound species.[145] In addition, a study by Baier et al.[137] has reported that binding non-native metals has, in certain cases, increased the promiscuous enzymatic activity of several MBL superfamily members. As promiscuous activities can play a role in functional divergence, promiscuity related to binding of different metal ions could have been involved in the evolution and expansion of the catalytic landscape of the MBL superfamily.[137] Numerous examples of promiscuous metal binding suggest that

while a particular metal may be better suited for a specific function, most MBL superfamily proteins are able to bind similarly sized transition metals available in their environment without the loss of their native catalytic activities.

1.4 MBL fold proteins and disease

Several MBL superfamily proteins have been linked to diseases or identified as potential therapeutic targets. The name-lending 'true' MBLs are one of the factors contributing to the antimicrobial resistance, especially due to their ability to hydrolyse carbapenems, used as last resort antibiotics.[53, 146] While there are several clinically used inhibitors for serine-β-lactamases, no 'true' MBL inhibitors are currently available for clinical use.[147] Several 'true' MBL inhibitors have been reported in the academic literature, many of which are Zn chelators as well as substrate analogues (reviewed in [147-149]). With the continued widespread use of β-lactam antibiotics, 'true' MBL inhibitors remain an area of interest in antimicrobial therapy. Additionally, other bacterial MBL proteins have been reported to play a role in virulence and infection (such as the aforementioned Pce[36, 94], PqsE[107] and quorum quenching lactonases[108-111]). However, the feasibility of these proteins as potential therapeutic targets is not yet known.

In addition to MBL proteins in pathogens, several human MBLs have reported links to human diseases (see **Table 1.2**). Functions of MBL proteins linked to disorders range from DNA repair, to tRNA processing, detoxification and some currently have unknown function.

Table 1.2: Overview of the human MBL superfamily proteins linked to diseases.

Protein	Function	Linked diseases in humans	References
SNM1A	DNA repair	Cockayne syndrome, altered sensitivity to cisplatin and ionising radiation	150, 151
SNM1B	DNA repair, telomere maintenance	Hoyeraal–Hreidarsson Syndrome	152
SNM1C	DNA repair, adaptive immunity	Omenn syndrome, T-B-severe combined immunodeficiency	153, 154
ELAC2	tRNA processing	Prostate cancer	155-157
Glx2	MG detoxification	Prostate cancer	158, 159
ETHE1	Hydrogen sulfide detoxification	Ethylmalonic encephalopathy	127
PNKD	unknown	Paroxysmal non-kinesigenic dyskinesia (Dystonia type 8)	130, 131, 160

The DNA repair nuclease SNM1A has been reported to interact with Cockayne syndrome protein B (CSB) which stimulates SNM1A activity in interstrand crosslink repair.[150] Defects in

the interaction have been suggested to contribute to the pathologies observed in Cockayne syndrome, such as increased sensitivity to UV light, developmental delay, and sensorial impairment.[150, 161] In a different study, mutations in SNM1A were reported to impair nuclear foci formation and normal interstrand crosslink repair function.[151]

SNM1B has been reported to be involved in telomere maintenance.[68, 69] The SNM1B splice variant has been reported in a patient with Hoyeraal-Hreidarsson syndrome (HH).[152] HH is a rare form of a condition linked with defects in telomere maintenance called dyskeratosis congenita and manifests in premature aging, immunodeficiency, bone marrow failure and growth retardation. [152]

The third human DNA repair nuclease SNM1C has been linked with several immunodeficiency disorders.[153, 154] Mutations of the *SNM1C* gene have been reported in patients suffering from Omenn syndrome (characterised by elevated T-cell levels or activated T-cell function with absent B-cells)[153] and T-B-severe combined immunodeficiency (caused by early arrest of both T- and B-cell maturation)[154] suggesting the involvement of SNM1C in immune cell development.

Mutations in the human homologue of tRNase Z ELAC2 have been linked to prostate cancer.[156, 157] A study of 451 men has proposed the genetic variants in *ELAC2* as potential biomarkers for risk and aggressiveness of prostate cancer.[155] Changes in ELAC2 activity have been suggested to result in abnormalities in cell division hence promoting tumorigenesis.[157]

A thioesterase Glx2 is involved in MG detoxification and regulation of GSH levels in mitochondria. Glx2 has been reported to promote prostate cancer development, tumorigenesis and progression.[158, 159] Glx2 was found to be overexpressed in prostate cancer cells and proposed as a potential biomarker for early prostate cancer detection. Different studies have reported androgen receptor[158] and PI3K/AKT/mTOR[159] signalling pathways to be involved in upregulation of Glx2 and related prostate cancer progression. Depletion of Glx2 has been reported to increase sensitivity to anticancer drugs and reduce cell viability without affecting healthy cells.[158, 159]

Persulfide dioxygenase ETHE1 gains its name from the fatal metabolic disease Ethylmalonic encephalopathy.[127] Types of mutations in the *ETHE1* gene observed in Ethylmalonic encephalopathy patients include frameshift, missense, and nonsense as well as changes in splicing and range from impairment to a complete loss of function.[125, 127, 162, 163]

Paroxysmal non-kinesigenic dyskinesia or Dystonia type 8, a movement disorder characterised by dystonia, chorea and athetosis, correlates to mutations in the *PNKD* gene encoding a protein with an unknown function.[130] PNKD isoforms have been reported to play a role in dopamine pathway, exocytosis[164], myofibrillogenesis[165], cell migration and proliferation[166] suggesting a pleiotropic outcome of *PNKD* mutations.

Whilst CMAH is not directly linked to any diseases, the *CMAH* gene has been of interest in the field of xenotransplantation.[167-169] Human CMAH is catalytically inactive due to a deletion in exon 6 which has resulted in no Neu5Gc being produced and it being immunogenic in humans.[170-172] Knocking out *CMAH* in conjunction with other genes has been reported to reduce human serum-mediated immune response *in vitro* and may be a part of a promising solution to organ shortage.[168, 169]

Given the wide range of catalytic and physiological activities of the MBL superfamily members, more currently unknown links to diseases will likely be discovered in the future and the increased understanding of similarities and differences between the MBL proteins could be beneficial for developing the best therapies and avoiding off-target effects.

1.5 Summary

Whilst first discovered as bacterial enzymes responsible for the breaking down of β-lactam antibiotics, the MBL superfamily is now known to span all domains of life and be comprised of members with various hydrolytic and redox functions. Their hydrolytic substrates include β-lactams, DNA, RNA, phospholipids, organophosphates, sulfates, thioesters, and lactones, while MBL oxidoreductases act on molecules ranging from gasses like O_2 and NO (flavidiiron subfamily) to amino and sialic acids (CmlA and CMAH respectively). Variations in substrate selectivity comes with unique structural features including additional domains involved in, among other functions, substrate binding, dimerization and activation of molecules involved in catalysis. Future structural studies may expand our understanding of substrate and inhibitor binding, differences between diverse MBL proteins and protein orthologues in different species.

Metal ions play a crucial role in MBL protein catalysis. While Zn ions are most commonly found in MBL hydrolases, enzymes with other native catalytic metal ions have been reported. MBL superfamily proteins are known promiscuous metal binders. Several MBL fold proteins have been reported to utilise non-native metal ions for native reactions as well as promiscuous activities. Different metal binding species have also been reported to differ in their inhibitory profiles and could be important in the search of inhibitors for 'true' MBLs as well as other

potential MBL targets. Another aspect to consider in future studies is the similarity of the active sites within the MBL superfamily. The impairment of functions of some MBL proteins has been linked to severe diseases in humans raising concerns that off-target inhibition of host MBL proteins may have serious side effects. Better understanding of structural differences could be beneficial in developing clinically safe MBL inhibitors.

1.6 Objectives

The work described in this thesis aimed to study two proteins belonging to the MBL superfamily – human and mouse hydroxylase (monooxygenase) CMAH and the *E. coli* thioesterase glyoxalase II (GloB).

Studies on CMAH aimed to develop a protein production system for recombinant human and mouse CMAH enzymes (hCMAH and mCMAH respectively) and to investigate and compare their functional and structural features. Cellular work focused on investigating sub-cellular localisation of hCMAH and mCMAH and scoped to detect protein-protein interactions in cells. More specifically, the aims of the work described in Chapter 2 and 3 were:

- To develop protein expression and purification systems for hCMAH and mCMAH.
- To characterise purified hCMAH and mCMAH proteins.
- To investigate activity of recombinant mCMAH.
- To study cellular localisation of endogenous and recombinant hCMAH in HeLa cells.
- To investigate protein-protein interactions of recombinant hCMAH and mCMAH in human and mouse cells.

Work on GloB aimed to investigate functional and structural features of the protein. Kinetic studies aimed to understand the effect of different metal binding and search for inhibitors, while structural work was performed to determine the GloB crystal structure and scope the possibility of obtaining the structure of the enzyme-substrate or enzyme-product complex. More specifically, the work described in Chapter 4 aimed to:

- To biochemically and kinetically characterise E. coli GloB.
- To investigate metal binding and the effect of different metals to GloB activity.
- To identify potential GloB inhibitors.
- To determine a three-dimensional crystal structure of GloB.

1.7 References

1. Richards, A. N., Production of penicillin in the United States (1941-1946). *Nature* **1964,** *201*, 441-445.

2. Versporten, A.; Bolokhovets, G.; Ghazaryan, L.; Abilova, V.; Pyshnik, G.; Spasojevic, T.; Korinteli, I.; Raka, L.; Kambaralieva, B.; Cizmovic, L.; Carp, A.; Radonjic, V.; Maqsudova, N.; Celik, H. D.; Payerl-Pal, M.; Pedersen, H. B.; Sautenkova, N.; Goossens, H., Antibiotic use in eastern Europe: a cross-national database study in coordination with the WHO Regional Office for Europe. *Lancet Infect. Dis.* **2014,** *14* (5), 381-387.

3. Donowitz, G. R.; Mandell, G. L., Beta-Lactam Antibiotics. *N. Engl. J. Med.* **1988,** *318* (8), 490-500.

4. Abraham, E. P.; Chain, E., An Enzyme from Bacteria able to Destroy Penicillin. *Nature* **1940,** *146*, 837.

5. Sabath, L. D.; Abraham, E. P., Cephalosporinase and penicillinase activity of *Bacillus cereus*. *Antimicrob. Agents Chemother.* **1965,** *5*, 392-397.

6. Sabath, L. D.; Abraham, E. P., Zinc as a Cofactor for Cephalosporinase from *Bacillus cereus* 569. *Biochem. J* **1966,** *98* (1), 11C-13C.

7. Saino, Y.; Kobayashi, F.; Inoue, M.; Mitsuhashi, S., Purification and properties of inducible penicillin β-lactamase isolated from *Pseudomonas maltophilia*. *Antimicrob. Agents Chemother.* **1982,** *22* (4), 564-570.

8. Bicknell, R.; Emanuel, E. L.; Gagnon, J.; Waley, S. G., The production and molecular properties of the zinc β-lactamase of *Pseudomonas maltophilia* IID 1275. *Biochem. J* **1985,** *229* (3), 791-197.

9. Bebrone, C., Metallo-β-lactamases (classification, activity, genetic organization, structure, zinc coordination) and their superfamily. *Biochem. Pharmacol.* **2007,** *74* (12), 1686-1701.

10. Bandoh, K.; Watanabe, K.; Muto, Y.; Tanaka, Y.; Kato, N.; Ueno, K., Conjugal transfer of imipenem resistance in *Bacteroides fragilis*. *J. Antibiot. (Tokyo)* **1992,** *45* (4), 542-547.

11. Watanabe, M.; Iyobe, S.; Inoue, M.; Mitsuhashi, S., Transferable imipenem resistance in *Pseudomonas aeruginosa*. *Antimicrob. Agents Chemother.* **1991,** *35* (1), 147-151.

12. Carfi, A.; Pares, S.; Duée, E.; Galleni, M.; Duez, C.; Frère, J. M.; Dideberg, O., The 3-D structure of a zinc metallo-β-lactamase from *Bacillus cereus* reveals a new type of protein fold. *EMBO J.* **1995,** *14* (20), 4914-4921.

13. Neuwald, A. F.; Liu, J. S.; Lipman, D. J.; Lawrence, C. E., Extracting protein alignment models from the sequence database. *Nucleic Acids Res.* **1997,** *25* (9), 1665-1677.

14. Melino, S.; Capo, C.; Dragani, B.; Aceto, A.; Petruzzelli, R., A zinc-binding motif conserved in glyoxalase II, β-lactamase and arylsulfatases. *Trends Biochem. Sci.* **1998,** *23* (10), 381-382.

15. Galleni, M.; Lamotte-Brasseur, J.; Rossolini, G. M.; Spencer, J.; Dideberg, O.; Frère, J. M., Standard numbering scheme for class B β-lactamases. *Antimicrob. Agents Chemother.* **2001,** *45* (3), 660-663.

16. Garau, G.; García-Sáez, I.; Bebrone, C.; Anne, C.; Mercuri, P.; Galleni, M.; Frère, J.-M.; Dideberg, O., Update of the standard numbering scheme for class B β-lactamases. *Antimicrob. Agents Chemother.* **2004,** *48* (7), 2347-2349.

17. Daiyasu, H.; Osaka, K.; Ishino, Y.; Toh, H., Expansion of the zinc metallo-hydrolase family of the β-lactamase fold. *FEBS Lett.* **2001,** *503* (1), 1-6.

18. Cameron, A. D.; Ridderstrom, M.; Olin, B.; Mannervik, B., Crystal structure of human glyoxalase II and its complex with a glutathione thiolester substrate analogue. *Structure* **1999,** *7* (9), 1067-1078.

19. Frazao, C.; Silva, G.; Gomes, C. M.; Matias, P.; Coelho, R.; Sieker, L.; Macedo, S.; Liu, M. Y.; Oliveira, S.; Teixeira, M.; Xavier, A. V.; Rodrigues-Pousada, C.; Carrondo, M. A.; Le Gall, J., Structure of a dioxygen reduction enzyme from *Desulfovibrio gigas*. *Nat. Struct. Biol.* **2000,** *7* (11), 1041-1045.

20. El-Gebali, S.; Mistry, J.; Bateman, A.; Eddy, S. R.; Luciani, A.; Potter, S. C.; Qureshi, M.; Richardson, L. J.; Salazar, G. A.; Smart, A.; Sonnhammer, E. L L.; Hirsh, L.; Paladin, L.; Piovesan, D.; Tosatto, S. C E.; Finn, R. D., The Pfam protein families database in 2019. *Nucleic Acids Res.* **2018,** *47* (D1), D427-D432.

21. Lee, J. H.; Takahashi, M.; Jeon, J. H.; Kang, L.-W.; Seki, M.; Park, K. S.; Hong, M.-K.; Park, Y. S.; Kim, T. Y.; Karim, A. M.; Lee, J.-H.; Nashimoto, M.; Lee, S. H., Dual activity of PNGM-1, a metallo-β-lactamase and tRNase Z, pinpoints the evolutionary origin of subclass B3 metallo-β-lactamases. *bioRxiv* **2019**, 575373.

22. Jones, P.; Binns, D.; Chang, H.-Y.; Fraser, M.; Li, W.; McAnulla, C.; McWilliam, H.; Maslen, J.; Mitchell, A.; Nuka, G.; Pesseat, S.; Quinn, A. F.; Sangrador-Vegas, A.; Scheremetjew, M.; Yong, S.-Y.; Lopez, R.; Hunter, S., InterProScan 5: genome-scale protein function classification. *Bioinformatics* **2014,** *30* (9), 1236-1240.

23. Ren, J.; Wen, L.; Gao, X.; Jin, C.; Xue, Y.; Yao, X., DOG 1.0: illustrator of protein domain structures. *Cell Res.* **2009,** *19* (2), 271-273.

24. Consortium, T. U., UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* **2018,** *47* (D1), D506-D515.

25. Brem, J.; van Berkel, S. S.; Zollman, D.; Lee, S. Y.; Gileadi, O.; McHugh, P. J.; Walsh, T. R.; McDonough, M. A.; Schofield, C. J., Structural Basis of Metallo-β-Lactamase Inhibition by Captopril Stereoisomers. *Antimicrob. Agents Chemother.* **2015,** *60* (1), 142-150.

26. Fonseca, F.; Bromley, E. H.; Saavedra, M. J.; Correia, A.; Spencer, J., Crystal structure of *Serratia fonticola* Sfh-I: activation of the nucleophile in mono-zinc metallo-β-lactamases. *J. Mol. Biol.* **2011,** *411* (5), 951-959.

27. Ullah, J. H.; Walsh, T. R.; Taylor, I. A.; Emery, D. C.; Verma, C. S.; Gamblin, S. J.; Spencer, J., The crystal structure of the L1 metallo-β-lactamase from *Stenotrophomonas maltophilia* at 1.7 Å resolution. *J. Mol. Biol.* **1998,** *284* (1), 125-136.

28. Allerston, C. K.; Lee, S. Y.; Newman, J. A.; Schofield, C. J.; McHugh, P. J.; Gileadi, O., The structures of the SNM1A and SNM1B/Apollo nuclease domains reveal a potential basis for their distinct DNA processing activities. *Nucleic Acids Res.* **2015,** *43* (22), 11047-11060.

29. Mir-Montazeri, B.; Ammelburg, M.; Forouzan, D.; Lupas, A. N.; Hartmann, M. D., Crystal structure of a dimeric archaeal cleavage and polyadenylation specificity factor. *J. Struct. Biol.* **2011,** *173* (1), 191-195.

30. Mandel, C. R.; Kaneko, S.; Zhang, H.; Gebauer, D.; Vethantham, V.; Manley, J. L.; Tong, L., Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease. *Nature* **2006,** *444* (7121), 953-956.

31. Li de la Sierra-Gallay, I.; Zig, L.; Jamalli, A.; Putzer, H., Structural insights into the dual activity of RNase J. *Nat. Struct. Mol. Biol.* **2008,** *15* (2), 206-212.

32. Pellegrini, O.; Li de la Sierra-Gallay, I.; Piton, J.; Gilet, L.; Condon, C., Activation of tRNA maturation by downstream uracil residues in *B. subtilis*. *Structure* **2012,** *20* (10), 1769-1777.

33. Levy, S.; Allerston, C. K.; Liveanu, V.; Habib, M. R.; Gileadi, O.; Schuster, G., Identification of LACTB2, a metallo-β-lactamase protein, as a human mitochondrial endoribonuclease. *Nucleic Acids Res.* **2016,** *44* (4), 1813-1832.

34. Pettinati, I.; Grzechnik, P.; Ribeiro de Almeida, C.; Brem, J.; McDonough, M. A.; Dhir, S.; Proudfoot, N. J.; Schofield, C. J., Biosynthesis of histone messenger RNA employs a specific 3' end endonuclease. *eLife* **2018,** *7*, e39865.

35. Magotti, P.; Bauer, I.; Igarashi, M.; Babagoli, M.; Marotta, R.; Piomelli, D.; Garau, G., Structure of human *N*-acylphosphatidylethanolamine-hydrolyzing phospholipase D: regulation of fatty acid ethanolamide biosynthesis by bile acids. *Structure* **2015,** *23* (3), 598-604.

36. Hermoso, J. A.; Lagartera, L.; Gonzalez, A.; Stelter, M.; Garcia, P.; Martinez-Ripoll, M.; Garcia, J. L.; Menendez, M., Insights into pneumococcal pathogenesis from the crystal structure

of the modular teichoic acid phosphorylcholine esterase Pce. *Nat. Struct. Mol. Biol.* **2005,** *12* (6), 533-538.

37. Dong, Y. J.; Bartlam, M.; Sun, L.; Zhou, Y. F.; Zhang, Z. P.; Zhang, C. G.; Rao, Z.; Zhang, X. E., Crystal structure of methyl parathion hydrolase from *Pseudomonas* sp. WBC-3. *J. Mol. Biol.* **2005,** *353* (3), 655-663.

38. Gotthard, G.; Hiblot, J.; Gonzalez, D.; Elias, M.; Chabriere, E., Structural and Enzymatic Characterization of the Phosphotriesterase OPHC2 from *Pseudomonas pseudoalcaligenes*. *PLoS One* **2013,** *8* (11), e77995.

39. Hagelueken, G.; Adams, T. M.; Wiehlmann, L.; Widow, U.; Kolmar, H.; Tummler, B.; Heinz, D. W.; Schubert, W. D., The crystal structure of SdsA1, an alkylsulfatase from *Pseudomonas aeruginosa*, defines a third class of sulfatases. *Proc. Natl. Acad. Sci. U. S. A.* **2006,** *103* (20), 7631-7636.

40. Yu, S.; Jensen, V.; Seeliger, J.; Feldmann, I.; Weber, S.; Schleicher, E.; Haussler, S.; Blankenfeldt, W., Structure elucidation and preliminary assessment of hydrolase activity of PqsE, the *Pseudomonas* quinolone signal (PQS) response protein. *Biochemistry* **2009,** *48* (43), 10298-10307.

41. Garces, F.; Fernandez, F. J.; Montella, C.; Penya-Soler, E.; Prohens, R.; Aguilar, J.; Baldoma, L.; Coll, M.; Badia, J.; Vega, M. C., Molecular architecture of the Mn^{2+}-dependent lactonase UlaG reveals an RNase-like metallo-β-lactamase fold and a novel quaternary structure. *J. Mol. Biol.* **2010,** *398* (5), 715-729.

42. Romao, C. V.; Vicente, J. B.; Borges, P. T.; Victor, B. L.; Lamosa, P.; Silva, E.; Pereira, L.; Bandeiras, T. M.; Soares, C. M.; Carrondo, M. A.; Turner, D.; Teixeira, M.; Frazao, C., Structure of *Escherichia coli* Flavodiiron Nitric Oxide Reductase. *J. Mol. Biol.* **2016,** *428* (23), 4686-4707.

43. Jasniewski, A. J.; Knoot, C. J.; Lipscomb, J. D.; Que, L., Jr., A Carboxylate Shift Regulates Dioxygen Activation by the Diiron Nonheme β-Hydroxylase CmlA upon Binding of a Substrate-Loaded Nonribosomal Peptide Synthetase. *Biochemistry* **2016,** *55* (41), 5818-5831.

44. Pettinati, I.; Brem, J.; McDonough, M. A.; Schofield, C. J., Crystal structure of human persulfide dioxygenase: Structural basis of ethylmalonic encephalopathy. *Hum. Mol. Genet.* **2015,** *24* (9), 2458-2469.

45. Park, K. S.; Hong, M. K.; Jeon, J. W.; Kim, J. H.; Jeon, J. H.; Lee, J. H.; Kim, T. Y.; Karim, A. M.; Malik, S. K.; Kang, L. W.; Lee, S. H., The novel metallo-β-lactamase PNGM-1 from a deep-sea sediment metagenome: crystallization and X-ray crystallographic analysis. *Acta Crystallogr. F. Struct. Biol. Commun.* **2018,** *74* (10), 644-649.

46. Choi, H.; Kim, H. J.; Matsuura, A.; Mikami, B.; Yoon, H. J.; Lee, H. H., Structural and functional studies of a metallo-β-lactamase unveil a new type of structurally encoded nickel-containing heterodinuclear site. *Acta Crystallogr. D Biol. Crystallogr.* **2015,** *71* (10), 2054-2065.

47. Motl, N.; Skiba, M. A.; Kabil, O.; Smith, J. L.; Banerjee, R., Structural and biochemical analyses indicate that a bacterial persulfide dioxygenase-rhodanese fusion protein functions in sulfur assimilation. *J. Biol. Chem.* **2017,** *292* (34), 14026-14038.

48. Gu, S. Y.; Yan, X. X.; Liang, D. C., Crystal structure of Tflp: a ferredoxin-like metallo-β-lactamase superfamily protein from *Thermoanaerobacter tengcongensis*. *Proteins* **2008,** *72* (1), 531-536.

49. Ambler, R. P., The Structure of β-Lactamases. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **1980,** *289* (1036), 321-331.

50. Bush, K.; Jacoby, G. A.; Medeiros, A. A., A Functional Classification Scheme for β-Lactamases and Its Correlation with Molecular Structure. *Antimicrob. Agents Chemother.* **1995,** *39* (6), 1211-1233.

51. Palzkill, T., Metallo-β-lactamase structure and function. *Ann. N.Y. Acad. Sci.* **2013,** *1277* (1), 91-104.

52. Meini, M. R.; Llarrull, L. I.; Vila, A. J., Overcoming differences: The catalytic mechanism of metallo-β-lactamases. *FEBS Lett.* **2015,** *589* (22), 3419-3432.

53. Segatore, B.; Massidda, O.; Satta, G.; Setacci, D.; Amicosante, G., High specificity of *cphA*-encoded metallo-β-lactamase from *Aeromonas hydrophila* AE036 for carbapenems and its contribution to β-lactam resistance. *Antimicrob. Agents Chemother.* **1993,** *37* (6), 1324-1328.

54. Hernandez Valladares, M.; Felici, A.; Weber, G.; Adolph, H. W.; Zeppezauer, M.; Rossolini, G. M.; Amicosante, G.; Frere, J. M.; Galleni, M., Zn(II) dependence of the *Aeromonas hydrophila* AE036 metallo-β-lactamase activity and stability. *Biochemistry* **1997,** *36* (38), 11534-11541.

55. Bebrone, C.; Delbruck, H.; Kupper, M. B.; Schlomer, P.; Willmann, C.; Frere, J. M.; Fischer, R.; Galleni, M.; Hoffmann, K. M., The structure of the dizinc subclass B2 metallo-β-lactamase CphA reveals that the second inhibitory zinc ion binds in the histidine site. *Antimicrob. Agents Chemother.* **2009,** *53* (10), 4464-4471.

56. Hall, B. G.; Salipante, S. J.; Barlow, M., Independent Origins of Subgroup B1+B2 and Subgroup B3 Metallo-β-Lactamases. *J. Mol. Evol.* **2004,** *59* (1), 133-141.

57. Moran-Barrio, J.; Lisa, M. N.; Larrieux, N.; Drusin, S. I.; Viale, A. M.; Moreno, D. M.; Buschiazzo, A.; Vila, A. J., Crystal Structure of the Metallo-β-Lactamase GOB in the Periplasmic

Dizinc Form Reveals an Unusual Metal Site. *Antimicrob. Agents Chemother.* **2016,** *60* (10), 6013-6022.

58. Crowder, M. W.; Walsh, T. R.; Banovic, L.; Pettit, M.; Spencer, J., Overexpression, purification, and characterization of the cloned metallo-β-lactamase L1 from *Stenotrophomonas maltophilia*. *Antimicrob. Agents Chemother.* **1998,** *42* (4), 921-926.

59. García-Sáez, I.; Mercuri, P. S.; Papamicael, C.; Kahn, R.; Frère, J. M.; Galleni, M.; Rossolini, G. M.; Dideberg, O., Three-dimensional Structure of FEZ-1, a Monomeric Subclass B3 Metallo-β-lactamase from *Fluoribacter gormanii*, in Native Form and in Complex with D-Captopril. *J. Mol. Biol.* **2003,** *325* (4), 651-660.

60. Moran-Barrio, J.; Gonzalez, J. M.; Lisa, M. N.; Costello, A. L.; Peraro, M. D.; Carloni, P.; Bennett, B.; Tierney, D. L.; Limansky, A. S.; Viale, A. M.; Vila, A. J., The metallo-β-lactamase GOB is a mono-Zn(II) enzyme with a novel active site. *J. Biol. Chem.* **2007,** *282* (25), 18286-18293.

61. Pettinati, I.; Brem, J.; Lee, S. Y.; McHugh, P. J.; Schofield, C. J., The Chemical Biology of Human Metallo-β-Lactamase Fold Proteins. *Trends Biochem. Sci.* **2016,** *41* (4), 338-355.

62. Callebaut, I.; Moshous, D.; Mornon, J. P.; de Villartay, J. P., Metallo-β-lactamase fold within nucleic acids processing enzymes: the β-CASP family. *Nucleic Acids Res.* **2002,** *30* (16), 3592-3601.

63. Cattell, E.; Sengerova, B.; McHugh, P. J., The SNM1/Pso2 family of ICL repair nucleases: from yeast to man. *Environ. Mol. Mutagen.* **2010,** *51* (6), 635-645.

64. Henriques, J. A.; Moustacchi, E., Isolation and characterization of *pso* mutants sensitive to photo-addition of psoralen derivatives in *Saccharomyces cerevisiae*. *Genetics* **1980,** *95* (2), 273-288.

65. Hazrati, A.; Ramis-Castelltort, M.; Sarkar, S.; Barber, L. J.; Schofield, C. J.; Hartley, J. A.; McHugh, P. J., Human SNM1A suppresses the DNA repair defects of yeast *pso2* mutants. *DNA Repair* **2008,** *7* (2), 230-238.

66. Wang, A. T.; Sengerova, B.; Cattell, E.; Inagawa, T.; Hartley, J. M.; Kiakos, K.; Burgess-Brown, N. A.; Swift, L. P.; Enzlin, J. H.; Schofield, C. J.; Gileadi, O.; Hartley, J. A.; McHugh, P. J., Human SNM1A and XPF-ERCC1 collaborate to initiate DNA interstrand cross-link repair. *Genes Dev.* **2011,** *25* (17), 1859-1870.

67. Buzon, B.; Grainger, R.; Huang, S.; Rzadki, C.; Junop, M. S., Structure-specific endonuclease activity of SNM1A enables processing of a DNA interstrand crosslink. *Nucleic Acids Res.* **2018,** *46* (17), 9057-9066.

68. Ye, J.; Lenain, C.; Bauwens, S.; Rizzo, A.; Saint-Leger, A.; Poulet, A.; Benarroch, D.; Magdinier, F.; Morere, J.; Amiard, S.; Verhoeyen, E.; Britton, S.; Calsou, P.; Salles, B.; Bizard, A.; Nadal, M.; Salvati, E.; Sabatier, L.; Wu, Y.; Biroccio, A.; Londono-Vallejo, A.; Giraud-Panis,

M. J.; Gilson, E., TRF2 and Apollo cooperate with topoisomerase 2α to protect human telomeres from replicative damage. *Cell* **2010,** *142* (2), 230-242.

69. Lam, Y. C.; Akhter, S.; Gu, P.; Ye, J.; Poulet, A.; Giruad-Panis, M.-J.; Bailey, S. M.; Gilson, E.; Legerski, R. J.; Chang, S., SNMIB/Apollo protects leading-strand telomeres against NHEJ-mediated repair. *EMBO J.* **2010,** *29*, 2230-2241.

70. Bae, J. B.; Mukhopadhyay, S. S.; Liu, L.; Zhang, N.; Tan, J.; Akhter, S.; Liu, X.; Shen, X.; Li, L.; Legerski, R. J., Snm1B/Apollo mediates replication fork collapse and S Phase checkpoint activation in response to DNA interstrand cross-links. *Oncogene* **2008,** *27* (37), 5045-5056.

71. Akhter, S.; Legerski, R. J., SNM1A Acts Downstream of ATM to Promote the G1 Cell Cycle Checkpoint. *Biochem. Biophys. Res. Commun.* **2008,** *377* (1), 236-241.

72. Ma, Y.; Pannicke, U.; Schwarz, K.; Lieber, M. R., Hairpin opening and overhang processing by an Artemis/DNA-dependent protein kinase complex in nonhomologous end joining and V(D)J recombination. *Cell* **2002,** *108* (6), 781-794.

73. Li, S.; Chang, H. H.; Niewolik, D.; Hedrick, M. P.; Pinkerton, A. B.; Hassig, C. A.; Schwarz, K.; Lieber, M. R., Evidence that the DNA endonuclease ARTEMIS also has intrinsic 5'-exonuclease activity. *J. Biol. Chem.* **2014,** *289* (11), 7825-7834.

74. Moshous, D.; Li, L.; de Chasseval, R. g.; Philippe, N.; Jabado, N.; Cowan, M. J.; Fischer, A.; de Villartay, J.-P., A new gene involved in DNA double-strand break repair and V(D)J recombination is located on human chromosome 10p. *Hum. Mol. Genet.* **2000,** *9* (4), 583-588.

75. Yan, Y.; Akhter, S.; Zhang, X.; Legerski, R., The multifunctional SNM1 gene family: not just nucleases. *Future Oncol.* **2010,** *6* (6), 1015-1029.

76. Clouet-d'Orval, B.; Rinaldi, D.; Quentin, Y.; Carpousis, A. J., Euryarchaeal β-CASP proteins with homology to bacterial RNase J Have 5'- to 3'-exoribonuclease activity. *J. Biol. Chem.* **2010,** *285* (23), 17574-17583.

77. Keller, W.; Bienroth, S.; Lang, K. M.; Christofori, G., Cleavage and polyadenylation factor CPF specifically interacts with the pre-mRNA 3' processing signal AAUAAA. *EMBO J.* **1991,** *10* (13), 4241-4249.

78. Mandel, C. R.; Bai, Y.; Tong, L., Protein factors in pre-mRNA 3′-end processing. *Cell. Mol. Life Sci.* **2008,** *65* (7-8), 1099-1122.

79. Clerici, M.; Faini, M.; Muckenfuss, L. M.; Aebersold, R.; Jinek, M., Structural basis of AAUAAA polyadenylation signal recognition by the human CPSF complex. *Nat. Struct. Mol. Biol.* **2018,** *25* (2), 135-138.

80. Yang, X.-c.; Sullivan, K. D.; Marzluff, W. F.; Dominski, Z., Studies of the 5′ Exonuclease and Endonuclease Activities of CPSF-73 in Histone Pre-mRNA Processing. *Mol. Cell. Biol.* **2009,** *29*, 31-42.

81. Kolev, N. G.; Yario, T. A.; Benson, E.; Steitz, J. A., Conserved motifs in both CPSF73 and CPSF100 are required to assemble the active endonuclease for histone mRNA 3'-end maturation. *EMBO Rep.* **2008,** *9* (10), 1013-1018.

82. Phung, D. K.; Rinaldi, D.; Langendijk-Genevaux, P. S.; Quentin, Y.; Carpousis, A. J.; Clouet-d'Orval, B., Archaeal β-CASP ribonucleases of the aCPSF1 family are orthologs of the eukaryal CPSF-73 factor. *Nucleic Acids Res.* **2013,** *41* (2), 1091-1103.

83. Silva, A. P.; Chechik, M.; Byrne, R. T.; Waterman, D. G.; Ng, C. L.; Dodson, E. J.; Koonin, E. V.; Antson, A. A.; Smits, C., Structure and activity of a novel archaeal β-CASP protein with *N*-terminal KH domains. *Structure* **2011,** *19* (5), 622-632.

84. Levy, S.; Portnoy, V.; Admon, J.; Schuster, G., Distinct activities of several RNase J proteins in methanogenic archaea. *RNA Biol.* **2011,** *8* (6), 1073-1083.

85. Hasenohrl, D.; Konrat, R.; Blasi, U., Identification of an RNase J ortholog in *Sulfolobus solfataricus*: implications for 5'-to-3' directional decay and 5'-end protection of mRNA in Crenarchaeota. *RNA* **2011,** *17* (1), 99-107.

86. Nishida, Y.; Ishikawa, H.; Baba, S.; Nakagawa, N.; Kuramitsu, S.; Masui, R., Crystal structure of an archaeal cleavage and polyadenylation specificity factor subunit from *Pyrococcus horikoshii*. *Proteins* **2010,** *78* (10), 2395-2398.

87. Baillat, D.; Hakimi, M. A.; Naar, A. M.; Shilatifard, A.; Cooch, N.; Shiekhattar, R., Integrator, a multiprotein mediator of small nuclear RNA processing, associates with the *C*-terminal repeat of RNA polymerase II. *Cell* **2005,** *123* (2), 265-276.

88. Wu, Y.; Albrecht, T. R.; Baillat, D.; Wagner, E. J.; Tong, L., Molecular basis for the interaction between Integrator subunits IntS9 and IntS11 and its functional importance. *Proc. Natl. Acad. Sci. U. S. A.* **2017,** *114* (17), 4394-4399.

89. Dominski, Z.; Yang, X. C.; Purdy, M.; Wagner, E. J.; Marzluff, W. F., A CPSF-73 homologue is required for cell cycle progression but not cell growth and interacts with a protein having features of CPSF-100. *Mol. Cell. Biol.* **2005,** *25* (4), 1489-1500.

90. Richards, J.; Belasco, J. G., Ribonuclease J: How to lead a double life. *Structure* **2011,** *19* (9), 1201-1203.

91. Takahashi, M.; Takaku, H.; Nashimoto, M., Regulation of the human tRNase ZS gene expression. *FEBS Lett.* **2008,** *582* (17), 2532-2536.

92. Rossmanith, W., Localization of human RNase Z isoforms: dual nuclear/mitochondrial targeting of the ELAC2 gene product by alternative translation initiation. *PLoS One* **2011,** *6* (4), e19152.

93. Wang, J.; Okamoto, Y.; Morishita, J.; Tsuboi, K.; Miyatake, A.; Ueda, N., Functional analysis of the purified anandamide-generating phospholipase D as a member of the metallo-β-lactamase family. *J. Biol. Chem.* **2006,** *281* (18), 12325-12335.

94. Garau, G.; Lemaire, D.; Vernet, T.; Dideberg, O.; Di Guilmi, A. M., Crystal structure of phosphorylcholine esterase domain of the virulence factor choline-binding protein e from *Streptococcus pneumoniae*: new structural features among the metallo-β-lactamase superfamily. *J. Biol. Chem.* **2005,** *280* (31), 28591-28600.

95. Gosink, K. K.; Mann, E. R.; Guglielmo, C.; Tuomanen, E. I.; Masure, H. R., Role of Novel Choline Binding Proteins in Virulence of *Streptococcus pneumoniae*. *Infect. Immun.* **2000,** *68* (10), 5690-5695.

96. Raushel, F. M., Bacterial detoxification of organophosphate nerve agents. *Curr. Opin. Microbiol.* **2002,** *5* (3), 288-295.

97. Toesch, M.; Schober, M.; Faber, K., Microbial alkyl- and aryl-sulfatases: mechanism, occurrence, screening and stereoselectivities. *Appl. Microbiol. Biotechnol.* **2014,** *98* (4), 1485-1496.

98. Barbeyron, T.; Brillet-Gueguen, L.; Carre, W.; Carriere, C.; Caron, C.; Czjzek, M.; Hoebeke, M.; Michel, G., Matching the Diversity of Sulfated Biomolecules: Creation of a Classification Database for Sulfatases Reflecting Their Substrate Specificity. *PLoS One* **2016,** *11* (10), e0164846.

99. Pogorevc, M.; Faber, K., Purification and characterization of an inverting stereo- and enantioselective sec-alkylsulfatase from the gram-positive bacterium *Rhodococcus ruber* DSM 44541. *Appl. Environ. Microbiol.* **2003,** *69* (5), 2810-2815.

100. Knaus, T.; Schober, M.; Kepplinger, B.; Faccinelli, M.; Pitzer, J.; Faber, K.; Macheroux, P.; Wagner, U., Structure and mechanism of an inverting alkylsulfatase from *Pseudomonas* sp. DSM6611 specific for secondary alkyl sulfates. *FEBS J.* **2012,** *279* (23), 4374-4384.

101. Waddell, G. L.; Gilmer, C. R.; Taylor, N. G.; Reveral, J. R. S.; Forconi, M.; Fox, J. L., The Eukaryotic Enzyme Bds1 Is an Alkyl but Not an Aryl Sulfohydrolase. *Biochem. Biophys. Res. Commun.* **2017,** *491* (2), 382-387.

102. Barbeyron, T.; Potin, P.; Richard, C.; Collin, O.; Kloareg, B., Arylsulphatase from *Alteromonas carrageenovora*. *Microbiology* **1995,** *141* (11), 2897-2904.

103. Thornalley, P. J., The glyoxalase system: new developments towards functional characterization of a metabolic pathway fundamental to biological life. *Biochem. J* **1990,** *269* (1), 1-11.

104. Armeni, T.; Cianfruglia, L.; Piva, F.; Urbanelli, L.; Luisa Caniglia, M.; Pugnaloni, A.; Principato, G., *S*-D-Lactoylglutathione can be an alternative supply of mitochondrial glutathione. *Free Radic. Biol. Med.* **2014,** *67*, 451-459.

105. Stamp, A. L.; Owen, P.; El Omari, K.; Nichols, C. E.; Lockyer, M.; Lamb, H. K.; Charles, I. G.; Hawkins, A. R.; Stammers, D. K., Structural and functional characterization of *Salmonella enterica* serovar typhimurium YcbL: An unusual type II glyoxalase. *Protein Sci.* **2010,** *19* (10), 1897-1905.

106. Reiger, M.; Lassak, J.; Jung, K., Deciphering the role of the type II glyoxalase isoenzyme YcbL (GlxII-2) in *Escherichia coli*. *FEMS Microbiol. Lett.* **2015,** *362* (2), 1-7.

107. Drees, S. L.; Fetzner, S., PqsE of *Pseudomonas aeruginosa* Acts as Pathway-Specific Thioesterase in the Biosynthesis of Alkylquinolone Signaling Molecules. *Chem. Biol.* **2015,** *22* (5), 611-608.

108. Kim, M. H.; Choi, W. C.; Kang, H. O.; Lee, J. S.; Kang, B. S.; Kim, K. J.; Derewenda, Z. S.; Oh, T. K.; Lee, C. H.; Lee, J. K., The molecular structure and catalytic mechanism of a quorum-quenching *N*-acyl-L-homoserine lactone hydrolase. *Proc. Natl. Acad. Sci. U. S. A.* **2005,** *102* (49), 17606-17611.

109. Liu, C. F.; Liu, D.; Momb, J.; Thomas, P. W.; Lajoie, A.; Petsko, G. A.; Fast, W.; Ringe, D., A Phenylalanine Clamp Controls Substrate Specificity in the Quorum-Quenching Metallo-γ-lactonase from *Bacillus thuringiensis*. *Biochemistry* **2013,** *52* (9), 1603-1610.

110. Mascarenhas, R.; Thomas, P. W.; Wu, C. X.; Nocek, B. P.; Hoang, Q. Q.; Liu, D.; Fast, W., Structural and Biochemical Characterization of AidC, a Quorum-Quenching Lactonase with Atypical Selectivity. *Biochemistry* **2015,** *54* (28), 4342-4353.

111. Bergonzi, C.; Schwab, M.; Naik, T.; Daudé, D.; Chabrière, E.; Elias, M., Structural and Biochemical Characterization of AaL, a Quorum Quenching Lactonase with Unusual Kinetic Properties. *Sci. Rep.* **2018,** *8*, 11262.

112. Funami, J.; Yoshikane, Y.; Kobayashi, H.; Yokochi, N.; Yuan, B.; Iwasaki, K.; Ohnishi, K.; Yagi, T., 4-Pyridoxolactonase from a symbiotic nitrogen-fixing bacterium *Mesorhizobium loti*: cloning, expression, and characterization. *Biochim. Biophys. Acta* **2005,** *1753* (2), 234-239.

113. Kobayashi, J.; Yoshikane, Y.; Yagi, T.; Baba, S.; Mizutani, K.; Takahashi, N.; Mikami, B., Structure of 4-pyridoxolactonase from *Mesorhizobium loti*. *Acta Crystallogr. F. Struct. Biol. Commun.* **2014,** *70* (4), 424-432.

114. Fernandez, F. J.; Garces, F.; Lopez-Estepa, M.; Aguilar, J.; Baldoma, L.; Coll, M.; Badia, J.; Vega, M. C., The UlaG protein family defines novel structural and functional motifs grafted on an ancient RNase fold. *BMC Evol. Biol.* **2011,** *11* (273).

115. Folgosa, F.; Martins, M. C.; Teixeira, M., Diversity and complexity of flavodiiron NO/O_2 reductases. *FEMS Microbiol. Lett.* **2019,** *365* (3), fnx267.

116. Rodrigues, R.; Vicente, J. B.; Felix, R.; Oliveira, S.; Teixeira, M.; Rodrigues-Pousada, C., *Desulfovibrio gigas* flavodiiron protein affords protection against nitrosative stress *in vivo*. *J. Bacteriol.* **2006,** *188* (8), 2745-2751.

117. Seedorf, H.; Dreisbach, A.; Hedderich, R.; Shima, S.; Thauer, R. K., $F_{420}H_2$ oxidase (FprA) from *Methanobrevibacter arboriphilus*, a coenzyme F_{420}-dependent enzyme involved in O_2 detoxification. *Arch. Microbiol.* **2004,** *182* (2-3), 126-137.

118. Vicente, J. B.; Tran, V.; Pinto, L.; Teixeira, M.; Singh, U., A detoxifying oxygen reductase in the anaerobic protozoan *Entamoeba histolytica*. *Eukaryot. Cell* **2012,** *11* (9), 1112-1118.

119. Schlenzka, W.; Shaw, L.; Kelm, S.; Schmidt, C. L.; Bill, E.; Trautwein, A. X.; Lottspeich, F.; Schauer, R., CMP-*N*-acetylneuraminic acid hydroxylase: The first cytosolic Rieske iron-sulphur protein to be described in Eukarya. *FEBS Lett.* **1996,** *385* (3), 197-200.

120. Shaw, L.; Schneckenburger, P.; Carlsen, J.; Christiansen, K.; Schauer, R., Mouse liver cytidine-5'-monophosphate-*N*-acetylneuraminic acid hydroxylase. Catalytic function and regulation. *Eur. J. Biochem.* **1992,** *206* (1), 269-277.

121. Shaw, L.; Schneckenburger, P.; Schlenzka, W.; Carlsen, J.; Christiansen, K.; Jürgensen, D.; Schauer, R., CMP-*N*-acetylneuraminic acid hydroxylase from mouse liver and pig submandibular glands. Interaction with membrane-bound and soluble cytochrome b_5-dependent electron transport chains. *Eur. J. Biochem.* **1994,** *219* (3), 1001-1011.

122. Makris, T. M.; Chakrabarti, M.; Munck, E.; Lipscomb, J. D., A family of diiron monooxygenases catalyzing amino acid β-hydroxylation in antibiotic biosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* **2010,** *107* (35), 15391-15396.

123. Makris, T. M.; Knoot, C. J.; Wilmot, C. M.; Lipscomb, J. D., Structure of a dinuclear iron cluster-containing β-hydroxylase active in antibiotic biosynthesis. *Biochemistry* **2013,** *52* (38), 6662-6671.

124. McCoy, J. G.; Bingman, C. A.; Bitto, E.; Holdorf, M. M.; Makaroff, C. A.; Phillips, G. N., Jr., Structure of an ETHE1-like protein from *Arabidopsis thaliana*. *Acta Crystallogr. D Biol. Crystallogr.* **2006,** *62* (9), 964-970.

125. Kabil, O.; Banerjee, R., Characterization of patient mutations in human persulfide dioxygenase (ETHE1) involved in H2S catabolism. *J. Biol. Chem.* **2012,** *287* (53), 44561-44567.

126. Holdorf, M. M.; Owen, H. A.; Lieber, S. R.; Yuan, L.; Adams, N.; Dabney-Smith, C.; Makaroff, C. A., *Arabidopsis ETHE1* encodes a sulfur dioxygenase that is essential for embryo and endosperm development. *Plant Physiol.* **2012,** *160* (1), 226-236.

127. Tiranti, V.; D'Adamo, P.; Briem, E.; Ferrari, G.; Mineri, R.; Lamantea, E.; Mandel, H.; Balestri, P.; Garcia-Silva, M. T.; Vollmer, B.; Rinaldo, P.; Hahn, S. H.; Leonard, J.; Rahman, S.; Dionisi-Vici, C.; Garavaglia, B.; Gasparini, P.; Zeviani, M., Ethylmalonic encephalopathy is caused by mutations in *ETHE1*, a gene encoding a mitochondrial matrix protein. *Am. J. Hum. Genet.* **2004,** *74* (2), 239-252.

128. Baier, F.; Tokuriki, N., Connectivity between Catalytic Landscapes of the Metallo-β-Lactamase Superfamily. *J. Mol. Biol.* **2014,** *426* (13), 2442-2456.

129. Park, K. S.; Kim, T. Y.; Kim, J. H.; Lee, J. H.; Jeon, J. H.; Karim, A. M.; Malik, S. K.; Lee, S. H., PNGM-1, a novel subclass B3 metallo-β-lactamase from a deep-sea sediment metagenome. *J. Glob. Antimicrob. Resist.* **2018,** *14*, 302-305.

130. Lee, H. Y.; Xu, Y.; Huang, Y.; Ahn, A. H.; Auburger, G. W.; Pandolfo, M.; Kwiecinski, H.; Grimes, D. A.; Lang, A. E.; Nielsen, J. E.; Averyanov, Y.; Servidei, S.; Friedman, A.; Van Bogaert, P.; Abramowicz, M. J.; Bruno, M. K.; Sorensen, B. F.; Tang, L.; Fu, Y. H.; Ptacek, L. J., The gene for paroxysmal non-kinesigenic dyskinesia encodes an enzyme in a stress response pathway. *Hum. Mol. Genet.* **2004,** *13* (24), 3161-3170.

131. Shen, Y.; Lee, H. Y.; Rawson, J.; Ojha, S.; Babbitt, P.; Fu, Y. H.; Ptáček, L. J., Mutations in PNKD causing paroxysmal dyskinesia alters protein cleavage and stability. *Hum. Mol. Genet.* **2011,** *20* (12), 2322-2332.

132. Diene, S. M.; Pinault, L.; Keshri, V.; Armstrong, N.; Khelaifia, S.; Chabrière, E.; Caetano-Anolles, G.; Colson, P.; La Scola, B.; Rolain, J. M.; Pontarotti, P.; Raoult, D., Human metallo-β-lactamase enzymes degrade penicillin. *Sci. Rep.* **2019,** *9*, 12173.

133. Marasinghe, G. P.; Sander, I. M.; Bennett, B.; Periyannan, G.; Yang, K. W.; Makaroff, C. A.; Crowder, M. W., Structural studies on a mitochondrial glyoxalase II. *J. Biol. Chem.* **2005,** *280* (49), 40668-40675.

134. Lagartera, L.; Gonzalez, A.; Hermoso, J. A.; Saiz, J. L.; Garcia, P.; Garcia, J. L.; Menendez, M., Pneumococcal phosphorylcholine esterase, Pce, contains a metal binuclear center that is essential for substrate binding and catalysis. *Protein Sci.* **2005,** *14* (12), 3013-3024.

135. Zang, T. M.; Hollman, D. A.; Crawford, P. A.; Crowder, M. W.; Makaroff, C. A., *Arabidopsis* glyoxalase II contains a zinc/iron binuclear metal center that is essential for substrate binding and catalysis. *J. Biol. Chem.* **2001,** *276* (7), 4788-4795.

136. Sabath, L. D.; Finland, M., Thiol-Group Binding of Zinc to a β-Lactamase of *Bacillus cereus*: Differential Effects on Enzyme Activity with Penicillin and Cephalosporins as Substrates. *J. Bacteriol.* **1968,** *95* (5), 1513-1519.

137. Baier, F.; Chen, J.; Solomonson, M.; Strynadka, N. C. J.; Tokuriki, N., Distinct Metal Isoforms Underlie Promiscuous Activity Profiles of Metalloenzymes. *ACS Chem. Biol.* **2015**, 1684-1693.

138. Cahill, S. T.; Tarhonskaya, H.; Rydzik, A. M.; Flashman, E.; McDonough, M. A.; Schofield, C. J.; Brem, J., Use of ferrous iron by metallo-β-lactamases. *J. Inorg. Biochem.* **2016,** *163*, 185-193.

139. Crawford, P. A.; Yang, K. W.; Sharma, N.; Bennett, B.; Crowder, M. W., Spectroscopic studies on cobalt(II)-substituted metallo-β-lactamase ImiS from *Aeromonas veronii* bv. *sobria*. *Biochemistry* **2005,** *44* (13), 5168-5176.

140. Campos-Bermudez, V. A.; Leite, N. R.; Krog, R.; Costa-Filho, A. J.; Soncini, F. C.; Oliva, G.; Vila, A. J., Biochemical and structural characterization of *Salmonella typhimurium* glyoxalase II: New insights into metal ion selectivity. *Biochemistry* **2007,** *46* (39), 11069-11079.

141. Campos-Bermudez, V. A.; Moran-Barrio, J.; Costa-Filho, A. J.; Vila, A. J., Metal-dependent inhibition of glyoxalase II: a possible mechanism to regulate the enzyme activity. *J. Inorg. Biochem.* **2010,** *104* (7), 726-731.

142. O'Young, J.; Sukdeo, N.; Honek, J. F., *Escherichia coli* glyoxalase II is a binuclear zinc-dependent metalloenzyme. *Arch. Biochem. Biophys.* **2007,** *459* (1), 20-26.

143. Na, H. W.; Namgung, B.; Song, W. S.; Yoon, S. I., Structural and biochemical analyses of the metallo-β-lactamase fold protein YhfI from *Bacillus subtilis*. *Biochem. Biophys. Res. Commun.* **2019,** *519* (1), 35-40.

144. Badarau, A.; Page, M. I., The Variation of Catalytic Efficiency of *Bacillus cereus* Metallo-β-lactamase with Different Active Site Metal Ions. *Biochemistry* **2006,** *45*, 10654-10666.

145. Tu, X.; Latham, J. A.; Klema, V. J.; Evans, R. L., 3rd; Li, C.; Klinman, J. P.; Wilmot, C. M., Crystal structures reveal metal-binding plasticity at the metallo-β-lactamase active site of PqqB from *Pseudomonas putida*. *J. Biol. Inorg. Chem.* **2017,** *22* (7), 1089-1097.

146. Babic, M.; Hujer, A. M.; Bonomo, R. A., What's new in antibiotic resistance? Focus on β-lactamases. *Drug Resistance Updates* **2006,** *9* (3), 142-156.

147. Drawz, S. M.; Papp-Wallace, K. M.; Bonomo, R. A., New β-lactamase inhibitors: a therapeutic renaissance in an MDR world. *Antimicrob. Agents Chemother.* **2014,** *58* (4), 1835-1846.

148. Drawz, S. M.; Bonomo, R. A., Three Decades of β-Lactamase Inhibitors. *Clin. Microbiol. Rev.* **2010,** *23* (1), 160-201.

149. Fast, W.; Sutton, L. D., Metallo-β-lactamase: Inhibitors and reporter substrates. *Biochim. Biophys. Acta* **2013,** *1834* (8), 1648-1659.

150. Iyama, T.; Lee, S. Y.; Berquist, B. R.; Gileadi, O.; Bohr, V. A.; Seidman, M. M.; McHugh, P. J.; Wilson, D. M., CSB interacts with SNM1A and promotes DNA interstrand crosslink processing. *Nucleic Acids Res.* **2015,** *43* (1), 247-258.

151. Ishiai, M.; Kimura, M.; Namikoshi, K.; Yamazoe, M.; Yamamoto, K.; Arakawa, H.; Agematsu, K.; Matsushita, N.; Takeda, S.; Buerstedde, J. M.; Takata, M., DNA cross-link repair protein SNM1A interacts with PIAS1 in nuclear focus formation. *Mol. Cell. Biol.* **2004,** *24* (24), 10733-10741.

152. Touzot, F.; Callebaut, I.; Soulier, J.; Gaillard, L.; Azerrad, C.; Durandy, A.; Fischer, A.; de Villartay, J. P.; Revy, P., Function of Apollo (SNM1B) at telomere highlighted by a splice variant identified in a patient with Hoyeraal-Hreidarsson syndrome. *Proc. Natl. Acad. Sci. U. S. A.* **2010,** *107* (22), 10097-10102.

153. Ege, M.; Ma, Y.; Manfras, B.; Kalwak, K.; Lu, H.; Lieber, M. R.; Schwarz, K.; Pannicke, U., Omenn syndrome due to ARTEMIS mutations. *Blood* **2005,** *105* (11), 4179-4186.

154. Moshous, D.; Callebaut, I.; de Chasseval, R.; Corneo, B.; Cavazzana-Calvo, M.; Le Deist, F.; Tezcan, I.; Sanal, O.; Bertrand, Y.; Philippe, N.; Fischer, A.; de Villartay, J. P., Artemis, a novel DNA double-strand break repair/V(D)J recombination protein, is mutated in human severe combined immune deficiency. *Cell* **2001,** *105* (2), 177-186.

155. Alvarez-Cubero, M. J.; Martinez-Gonzalez, L. J.; Saiz, M.; Carmona-Saez, P.; Alvarez, J. C.; Pascual-Geler, M.; Lorente, J. A.; Cozar, J. M., Prognostic role of genetic biomarkers in clinical progression of prostate cancer. *Exp. Mol. Med.* **2015,** *47* (8), e176.

156. Takaku, H.; Minagawa, A.; Takagi, M.; Nashimoto, M., A candidate prostate cancer susceptibility gene encodes tRNA 3′ processing endoribonuclease. *Nucleic Acids Res.* **2003,** *31* (9), 2272-2278.

157. Korver, W.; Guevara, C.; Chen, Y.; Neuteboom, S.; Bookstein, R.; Tavtigian, S.; Lees, E., The product of the candidate prostate cancer susceptibility gene *ELAC2* interacts with the γ-tubulin complex. *Int. J. Cancer* **2003,** *104* (3), 283-288.

158. Antognelli, C.; Ferri, I.; Bellezza, G.; Siccu, P.; Love, H. D.; Talesa, V. N.; Sidoni, A., Glyoxalase 2 drives tumorigenesis in human prostate cells in a mechanism involving androgen receptor and p53-p21 axis. *Mol. Carcinog.* **2017,** *56* (9), 2112-2126.

159. Talesa, V. N.; Ferri, I.; Bellezza, G.; Love, H. D.; Sidoni, A.; Antognelli, C., Glyoxalase 2 Is Involved in Human Prostate Cancer Progression as Part of a Mechanism Driven By PTEN/PI3K/AKT/mTOR Signaling With Involvement of PKM2 and ERα. *Prostate* **2017,** *77* (2), 196-210.

160. Ghezzi, D.; Viscomi, C.; Ferlini, A.; Gualandi, F.; Mereghetti, P.; DeGrandis, D.; Zeviani, M., Paroxysmal non-kinesigenic dyskinesia is caused by mutations of the MR-1 mitochondrial targeting sequence. *Hum. Mol. Genet.* **2009,** *18* (6), 1058-1064.

161. Laugel, V., Cockayne syndrome: the expanding clinical and mutational spectrum. *Mech. Ageing Dev.* **2013,** *134* (5-6), 161-170.

162. Henriques, B. J.; Lucas, T. G.; Rodrigues, J. V.; Frederiksen, J. H.; Teixeira, M. S.; Tiranti, V.; Bross, P.; Gomes, C. M., Ethylmalonic encephalopathy ETHE1 R163W/R163Q mutations alter protein stability and redox properties of the iron centre. *PLoS One* **2014,** *9* (9), e107157.

163. Tiranti, V.; Briem, E.; Lamantea, E.; Mineri, R.; Papaleo, E.; De Gioia, L.; Forlani, F.; Rinaldo, P.; Dickson, P.; Abu-Libdeh, B.; Cindro-Heberle, L.; Owaidha, M.; Jack, R. M.; Christensen, E.; Burlina, A.; Zeviani, M., *ETHE1* mutations are specific to ethylmalonic encephalopathy. *J. Med. Genet.* **2006,** *43* (4), 340-346.

164. Lee, H. Y.; Nakayama, J.; Xu, Y.; Fan, X.; Karouani, M.; Shen, Y.; Pothos, E. N.; Hess, E. J.; Fu, Y. H.; Edwards, R. H.; Ptacek, L. J., Dopamine dysregulation in a mouse model of paroxysmal nonkinesigenic dyskinesia. *J. Clin. Invest.* **2012,** *122* (2), 507-518.

165. Li, T.-B.; Liu, X.-H.; Feng, S.; Hu, Y.; Yang, W.-X.; Han, Y.; Wang, Y.-G.; Gong, L.-M., Characterization of MR-1, a Novel Myofibrillogenesis Regulator in Human Muscle. *Acta Biochim. Biophys. Sin.* **2004,** *36* (6), 412-418.

166. Ren, K.; Jin, H.; Bian, C.; He, H.; Liu, X.; Zhang, S.; Wang, Y.; Shao, R. G., MR-1 modulates proliferation and migration of human hepatoma HepG2 cells through myosin light chains-2 (MLC2)/focal adhesion kinase (FAK)/Akt signaling pathway. *J. Biol. Chem.* **2008,** *283* (51), 35598-35605.

167. Song, K. H.; Kang, Y. J.; Jin, U. H.; Park, Y. I.; Kim, S. M.; Seong, H. H.; Hwang, S.; Yang, B. S.; Im, G. S.; Min, K. S.; Kim, J. H.; Chang, Y. C.; Kim, N. H.; Lee, Y. C.; Kim, C. H., Cloning and functional characterization of pig CMP-*N*-acetylneuraminic acid hydroxylase for the synthesis of *N*-glycolylneuraminic acid as the xenoantigenic determinant in pig-human xenotransplantation. *Biochem. J* **2010,** *427* (1), 179-188.

168. Lutz, A. J.; Li, P.; Estrada, J. L.; Sidner, R. A.; Chihara, R. K.; Downey, S. M.; Burlak, C.; Wang, Z. Y.; Reyes, L. M.; Ivary, B.; Yin, F.; Blankenship, R. L.; Paris, L. L.; Tector, A. J., Double knockout pigs deficient in *N*-glycolylneuraminic acid and Galactose α-1,3-Galactose reduce the humoral barrier to xenotransplantation. *Xenotransplantation* **2013,** *20* (1), 27-35.

169. Kim, G. a.; Lee, E. M.; Jin, J.-X.; Lee, S.; Taweechaipaisankul, A.; Hwang, J. I.; Alam, Z.; Ahn, C.; Lee, B. C., Generation of *CMAH*KO/GTKO/shTNFRI-Fc/HO-1 quadruple gene modified pigs. *Transgenic Res.* **2017**, 435-445.

170. Irie, A.; Koyama, S.; Kozutsumi, Y.; Kawasaki, T.; Suzuki, A., The Molecular Basis for the Absence of *N*-Glycolylneuraminic Acid in Humans. *J. Biol. Chem.* **1998,** *273* (25), 15866-15871.

171. Chou, H.-H.; Takematsu, H.; Diaz, S.; Iber, J.; Nickerson, E.; Wright, K. L.; Muchmore, E. A.; Nelson, D. L.; Warren, S. T.; Varki, A., A mutation in human CMP-sialic acid hydroxylase occurred after the *Homo-Pan* divergence. *Proc. Natl. Acad. Sci. U. S. A.* **1998,** *95* (20), 11751-11756.

172. Varki, A., *N*-glycolylneuraminic acid deficiency in humans. *Biochimie* **2001,** *83* (7), 615-622.

Chapter 2: Biochemical and Biophysical Characterisation of CMAH

Contents

2.1 Introduction....41
2.1.1 Sialic acids....41
2.1.2 Cytidine monophosphate *N*-acetylneuraminic acid hydroxylase....42
2.1.3 Loss of CMAH activity in humans....45
2.1.4 Other inactivation events....46
2.2 Objectives....46
2.3 Production of recombinant hCMAH and mCMAH proteins....47
2.3.1 Cloning....47
2.3.2 Bacterial expression trials....48
2.3.3 Large scale bacterial expression....50
2.3.4 Mammalian expression systems....52
2.4 Characterisation....56
2.4.1 Metal content analysis....56
2.4.2 Deglycosylation experiments....57
2.4.3 Mass spectrometry and Circular Dichroism spectrosopy....58
2.4.4 Oligomerisation state analysis....61
2.5 Characterisation of FeS cluster....64
2.5.1 UV-Vis spectroscopy....64
2.5.2 Electron paramagnetic resonance....64
2.6 Hydroxylase activity assay....66
2.6.1 Assay optimisation and NADH decomposition....68
2.6.2 Reaction timecourse....69
2.7 Crystallographic studies....71
2.7.1 Thermal shift assays....71
2.7.2 hCMAH crystallisation trials....74
2.7.3 mCMAH crystallisation trials....76
2.9 Summary and discussion....79
2.10 References....80

The work in this chapter was carried out in collaboration with a number of other researchers. Dr Ilaria Pettinati advised on aspects of protein production and mammalian cell work and assisted with transient protein expression. Dr David Owen (Dr Martin Walsh group, Research Complex at Harwell) assisted with hCMAH crystallography and thermal stability analysis. Dr Michael McDonough and Dr Ian Clifton advised on crystallography and data processing. Dr Anthony Tumber assisted with running RapidFire-MS assays. Non-denaturing MS spectra were recorded by Shuang Liu. EPR experiments were run by Dr William Myers (Centre for Advanced Electron Spin Resonance, University of Oxford).

2.1 Introduction

2.1.1 Sialic acids

Sialic acids are nine-carbon α-keto acids, typically occupying terminal positions on glycan chains (**Figure 2.1**).[1, 2] There are over 50 structurally distinct sialic acids of which *N*-acetylneuraminic acid (Neu5Ac) and 2-keto-3-deoxynononic acid (Kdn) are both the most common and precursors to all other known sialic acids.[2] Variety in α-glycosidic linkages increases the diversity even further. Due to their high diversity and their terminal position on the cell surface glycoprotein and glycolipid glycan chains, sialic acids are involved in various cell-cell and cell-pathogen interactions. Sialylation plays an important role in many biological processes, including embryonic development[3], sperm[4] and egg[5] cell maturation, immunity[2], and, in certain cases, blood type[6].

Sialic acids are found primarily in animals of the Deuterostomia superphylum which includes vertebrates and some higher invertebrates. Some bacteria also sialylate their cell surface[5]. To a lesser extent, sialic acids have also been detected in some insects[7], octopuses and squids[8], some pathogenic species of fungi[9], and protozoa[10]. However, it is unclear whether these organisms produce their own sialic acids or obtain them from dietary sources or host organisms.[1] The presence of sialic acids in bacteria, fungi and protozoa is thought to help them evade detection by the host immune system[1, 11]. Multiple pathogens, including bacteria, viruses and parasites, are known to recognise host sialic acids and use them for attachment or infecting host cells.[12-15]

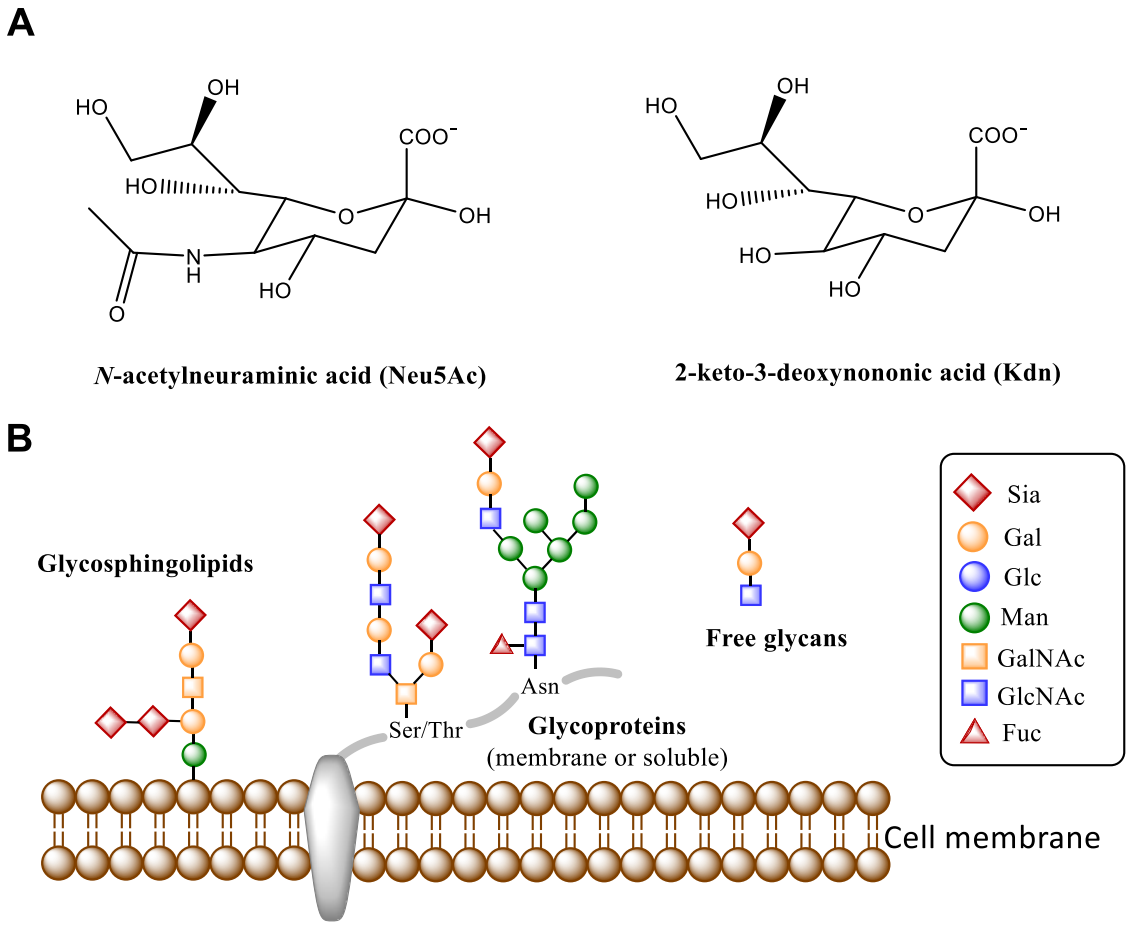


Figure 2.1: (A) Structure of the two most common sialic acids - *N*-acetylneuraminic acid (Neu5Ac) and 2-keto-3-deoxynononic acid (Kdn). (B) Location of sialic acids in glycan chains. Sia – sialic acid, Gal – galactose, Glc – glucose, Man – mannose, GalNAc – *N*-acetyl-galactosamine, GlcNAc – *N*-acetyl-glucosamine, Fuc – fucose.

2.1.2 Cytidine monophosphate *N*-acetylneuraminic acid hydroxylase

Cytidine monophosphate *N*-acetylneuraminic acid hydroxylase (CMAH, EC 1.14.18.2) catalyses the oxidation of cytidine monophosphate *N*-acetylneuraminic acid (CMP-Neu5Ac) to CMP-*N*-glycolylneuraminic acid (CMP-Neu5Gc, **Figure 2.2**). CMAH is a 560-600 amino acid protein found in the Deuterostomia superphylum. CMAH is a monooxygenase belonging to the metallo-β-lactamase (MBL) structural superfamily.[16] As described in Chapter 1: General Introduction, members of the MBL superfamily share a conserved αββα fold and bind one or two metal ions in their active site. Similarly to other MBL oxidoreductases, CMAH does not bind zinc but contains an iron ion in its active site[17]. In addition to the MBL domain, CMAH is also reported to possess a Rieske domain containing a [2Fe2S] cluster.[17]

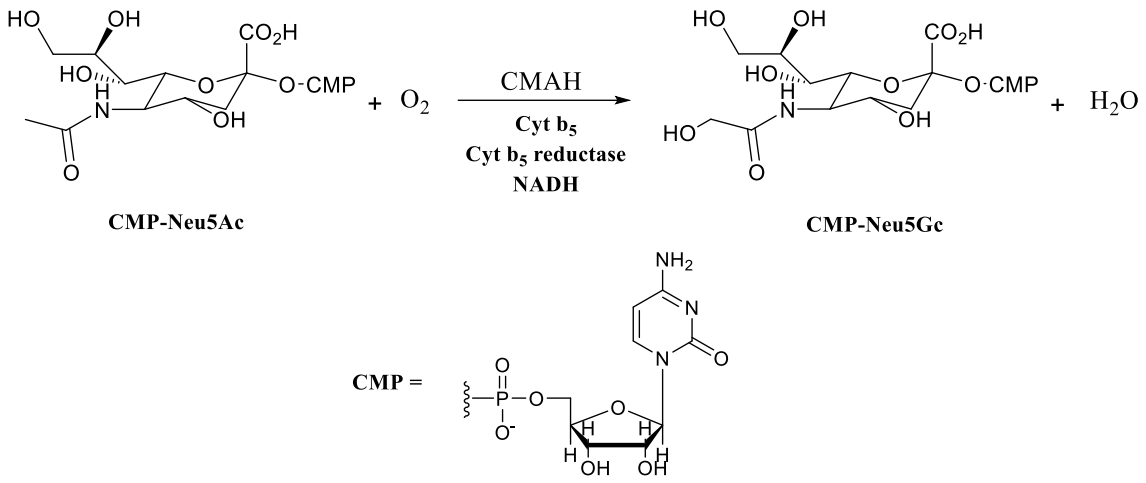


Figure 2.2: Hydroxylation of CMP-Neu5Ac as catalysed by CMAH.

Hydroxylation as catalysed by CMAH is reported to require NADH, cytochrome b_5 (cyt b_5) and cyt b_5 reductase (**Figure 2.3**).[18] The reaction is proposed to start with the substrate, CMP-Neu5Ac, binding to CMAH, followed by an interaction with cyt b_5[19]. The substrate is then converted to the hydroxylated product and free enzyme is recovered. It is unclear whether the reduction of cyt b_5 occurs elsewhere and the reduced cyt b_5 binds to the enzyme-substrate complex, or if cyt b_5 is reduced while bound to the enzyme-substrate complex. In turn, it is currently unknown whether the substrate binds to an oxidised or reduced enzyme state and at which point – i.e. at the enzyme-substrate or enzyme-product complex stage or after release of the product – the CMAH is being reduced in order to complete the catalytic cycle.

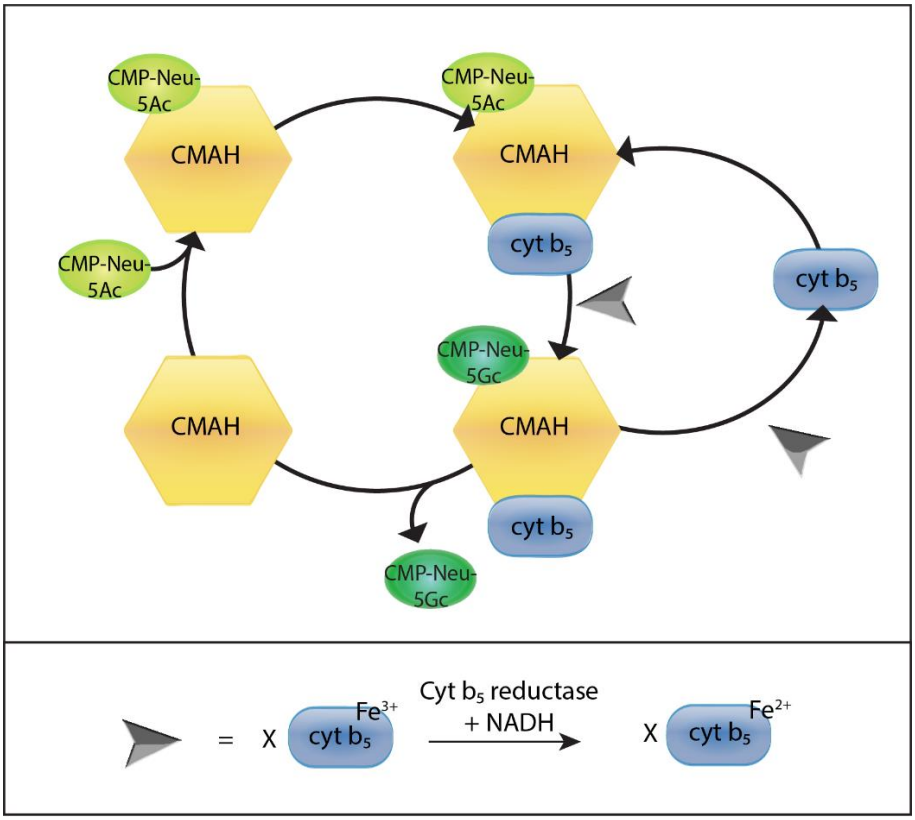


Figure 2.3: Proposed reaction scheme for CMP-Neu5AC hydroxylation as catalysed by CMAH including interactions with cyt b_5, cyt b_5 reductase and NADH. Grey arrows depict a potential electron transfer to cyt b_5 from cyt b_5 reductase and NADH. The reaction scheme was proposed by Takematsu et al.[19]

Previous studies on CMAH proteins have primarily focused on proteins in or extracted from fractionated tissues such as animal livers (murine[18-25], bovine and porcine[25] CMAH), submandibular glands (porcine[25, 26] CMAH) and gonads (starfish[27] and other echinoderms[28]).

While all of the CMAH proteins studied so far seem to require the same reaction partners (cyt b_5, cyt b_5 reductase and NADH)[17, 18, 23, 25, 27, 29], there are several differences between more distantly related organisms. Mammalian CMAHs appear to be soluble cytosolic proteins, while starfish and several other echinoderms possess membrane bound enzymes.[27, 28] Experiments with proteins extracted from endogenous sources, suggest that starfish CMAH is more tolerant to higher salt concentrations, has a lower optimal temperature (25-33 °C) and an optimal pH of 6.7[27]. In contrast, mouse CMAH loses 50% of its original activity in presence of 21 mM NaCl, while optimal reaction conditions (pH 7.4, 37 °C) are the physiological conditions in mammals[18, 27]. While mouse CMAH is the best studied of the CMAH proteins[18-22, 24, 25], other mammalian enzymes seem to behave in a very similar manner[26]. One potential difference that has not been fully explored yet is CMAH glycosylation. African green monkey CMAH produced

using a baculovirus protein expression system has been shown to be glycosylated[29]. However, potential glycosylation has not been investigated for other CMAH proteins and it remains unclear how widespread the post-translational modification is.

2.1.3 Loss of CMAH activity in humans

Unlike our closest evolutionary relatives the great apes, humans lack enzymatically active CMAH due to an *Alu*-mediated 92 base pair (bp) deletion of exon 6 in the gene encoding for CMAH.[30-32] The deletion event is thought to have occurred approximately 3 million years ago, after the last common ancestor with chimpanzees and bonobos. The human population is homozygous for this mutation, suggesting a potential evolutionary advantage of losing CMAH activity.[30, 33, 34]

Two conflicting theories exist regarding the resulting product of the mutated *CMAHP* gene. Irie et al. proposed a 104 amino acid truncation at the *N*-terminus resulting in a deletion of the Rieske domain and [2Fe2S] cluster[31], while Chou et al. suggested a frameshift mutation and early termination of the gene expression resulting in a 72 amino acid long peptide being made[30]. While the evidence of protein translation is not definitive, the region encoding *CMAHP* gene is heavily transcribed particularly in stem cells where the *CMAHP* gene expression is proposed to be involved in the Wnt/β-catenin signalling pathway.[35]

Loss of CMAH activity and lack of an alternative synthesis pathway has resulted in elimination of Neu5Gc in humans.[36] While humans do not produce their own Neu5Gc, Neu5Gc obtained from dietary sources, such as red meat and dairy products can be detected in trace amounts in various tissues with higher concentrations detectable in foetal tissues and certain types of tumours.[37] This obtained Neu5Gc is immunogenic in humans and is one of the causes of serum sickness in patients transfused with animal serum. Incorporation of Neu5Gc into human glycans can also cause a chronic inflammation termed Xenosialitis.[38]

Altered sialic acid composition resulting in changes in cell-cell and cell-pathogen interactions could explain several differences between humans and great apes. Studies have suggested that the elimination of Neu5Gc in mammals can lead to an enhanced immune response[39, 40], slower wound healing[41], and hearing loss[42]. CMAH expression is downregulated in the brains of most mammals in turn lowering the amount of Neu5Gc.[36] Loss of the only Neu5Gc biosynthesis pathway has eliminated Neu5Gc from the human brain and could have potentially been involved in the brain expansion which occurred during the evolution of the *Homo* genus.[33, 36] *CMAH*-knockout mouse models have exhibited pancreatic β-cell failure and glucose

intolerance following a high-fat diet, symptoms leading to type 2 diabetes in humans, suggesting CMAH involvement in carbohydrate metabolism in mice.[43]

Various viruses, bacteria and protozoans can use sialic acids for infection. Altered sialic acid composition has led to a change in susceptibility to pathogens. The host range of Influenza A viruses is influenced by the type of sialic acid as well as its linkage in cells' glycan chains.[13, 44]. Likewise, typhoid fever is caused by a toxin produced by the exclusively human pathogen *Salmonella enterica* serovar Typhi which binds Neu5Ac and is toxic towards cells presenting glycans with terminal Neu5Ac.[45] Malaria causing parasites *Plasmodium falciparum* and *Plasmodium reichenowi* infect humans and chimpanzees respectively. Despite both parasites being morphologically identical and sharing high genetic similarity, due to differences in the selectivity of their erythrocyte-binding antigen-175 for either Neu5Ac or Neu5Gc, cross-species infection does not happen (in case of *P. reichenowi*) or results in a very brief infection (*P. falciparum* in chimpanzees).[46]

2.1.4 Other inactivation events

Humans are not unique in their loss of CMAH function. In an unrelated inactivation approximately 30 million years ago, an inversion of exons 4 and 13 has left New World monkeys without CMAH activity and unable to synthesise endogenous Neu5Gc.[47] Due to similarities in sialic acid biology, New World monkeys can be infected by some human pathogens, like the aforementioned *P. falciparum*, that other more closely evolutionary related primates are immune to, making New World monkeys a good disease model.[47]

Ferrets – a common animal model of human-adapted Influenza A virus – are also unable to synthesise Neu5Gc.[48] The inactivating deletion of a large part of *CMAH* gene occurred between 38 and 40 million years ago and is shared with other mustelids as well as pinnipeds (seals).[48] Some breeds of dogs[49, 50] and cats possessing blood type B[51] also lack the ability to synthesise Neu5Gc. In addition, Neu5Gc has not been detected in birds (with an exception of ducks[52]), reptiles and platypus.[53] The genetic basis of this absence is yet to be determined, however it is likely there have been multiple unrelated CMAH inactivation events in animal evolution.

2.2 Objectives

The aims of the work described in this chapter were to produce and characterise the catalytically inactive human CMAH (hCMAH) and its full-length enzymatically active mouse homologue (mCMAH). Preliminary functional and structural studies are described.

2.3 Production of recombinant hCMAH and mCMAH proteins

2.3.1 Cloning

Since little is known about human CMAH (hCMAH), it was chosen for this study alongside a better understood catalytically active mouse homologue (mCMAH). The two proteins share an overall sequence similarity of 72% as mCMAH possesses an additional Rieske domain. Over the MBL domain the two proteins share 87% sequence similarity (**Figure 2.4**). Gene sequences encoding full-length hCMAH (UniProtKB: Q9Y471, isoform 2) and mCMAH (UniProtKB: Q61419) proteins were codon-optimised for *Escherichia coli* and human cell expression and purchased from Thermo Fisher Scientific. *E. coli* optimised gene sequences were excised using restriction enzymes and cloned into a pCold vector, hCMAH gene was amplified by polymerase chain reaction (PCR) to insert new restriction sites, and the sequence was cloned into a pMAL vector. Gene sequences optimised for mammalian expression were cloned from supplied cloning vectors into a pcDNA3.1 vector. Constructs produced and tested are listed in **Table 2.1**. Protein sequences are listed in Appendix 1. For experimental details see Chapter 6 (Materials and Methods).

Table 2.1: Constructs produced and tested for hCMAH and mCMAH protein expression. MBP – maltose binding protein, MW - molecular weight. Values in parentheses correspond to MW after tag cleavage. Protein sequences resulting from expression of each construct are listed in Appendix 1.

Construct	Purification tags	Predicted MW (Da)	Expression systems tested
hCMAH-pCold	*N*-His_6	60695.44 (56746.00)	*E. coli* BL21(DE3) and Rosetta gami B(DE3) pLysS
hCMAH-pMAL	*N*-MBP, *C*-His_6	101045.10 (57354.62)	*E. coli* BL21(DE3) and Rosetta gami B(DE3) pLysS
mCMAH-pCold	*N*-His_6	71122.29 (67172.85)	*E. coli* BL21(DE3) and Rosetta gami B(DE3) pLysS
hCMAH-pcDNA3.1	*C*-His_6	58215.69 (57238.68)	HEK293 (stable), ExpiCHO (transient)
mCMAH-pcDNA3.1	*C*-His_6	68642.54 (67665.53)	HEK293 (stable), Expi293 (transient), ExpiCHO (transient)

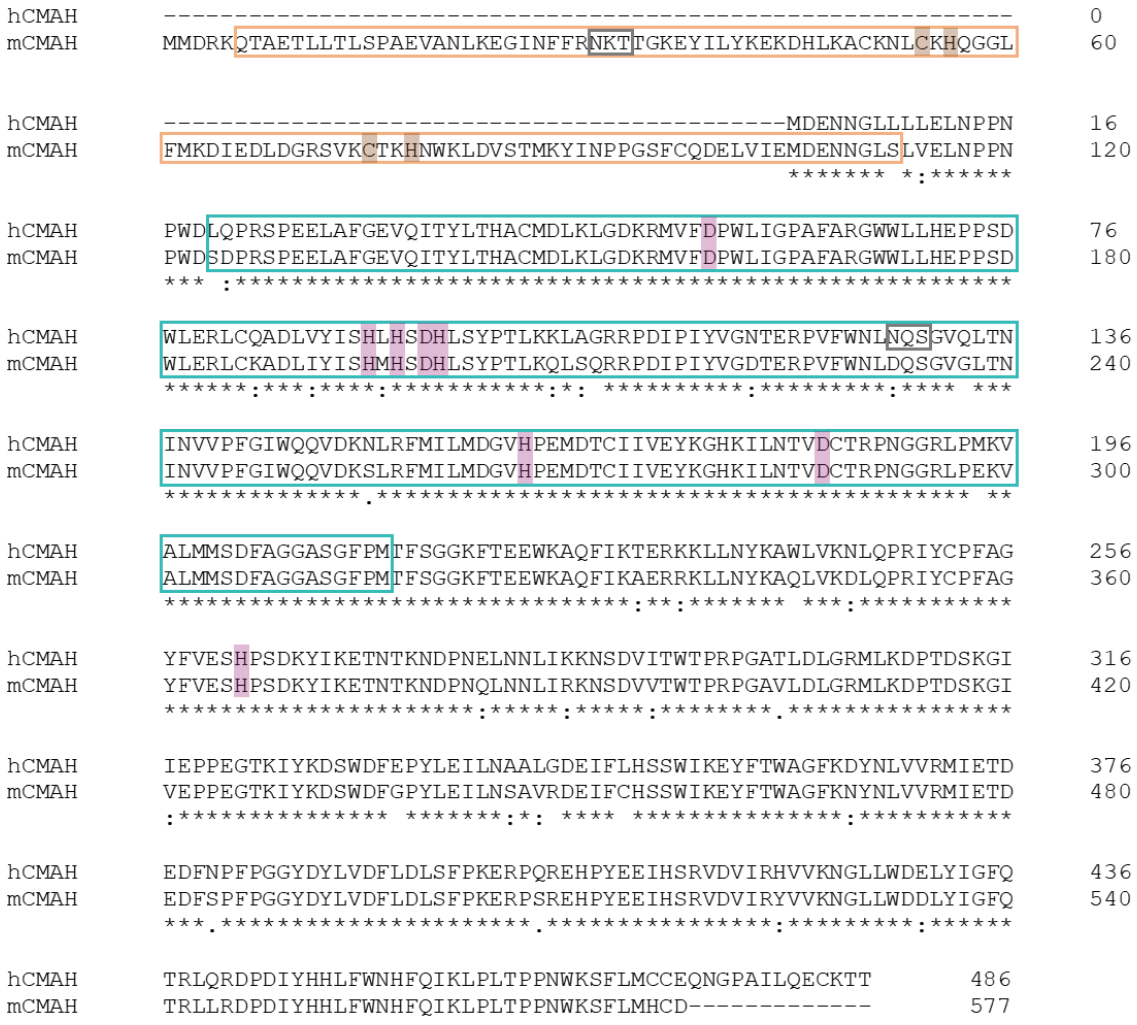


Figure 2.4: Pairwise sequence alignment of human (h) and mouse (m) CMAH proteins. The MBL domain is highlighted in cyan with MBL motifs highlighted in pink. The mCMAH Rieske domain is highlighted in orange with predicted [2Fe2S] cluster binding residues highlighted in brown. Potential *N*-glycosylation sites, as predicted using NetNGlyc server, are highlighted in grey. Alignment was performed using EMBOSS Needle software[54]. Domain boundaries were predicted by InterProScan[55]. The MBL motif positions are highlighted as proposed by Daiyasu et al.[56]

2.3.2 Bacterial expression trials

Initial hCMAH expression trials were performed in *E. coli* BL21(DE3) cells using the pCold construct in 2TY and TB media. Soluble hCMAH protein levels were, however, too poor to be detectable by Coomassie staining on sodium dodecyl sulfate – polyacrylamide gel electrophoresis (SDS-PAGE). In attempts to obtain soluble protein, expression was further tested in Rosetta-gami B(DE3) pLysS cells optimised for production of mammalian proteins (**Figure 2.5**). A significant increase in hCMAH protein levels was observed when compared to BL21(DE3) expression.

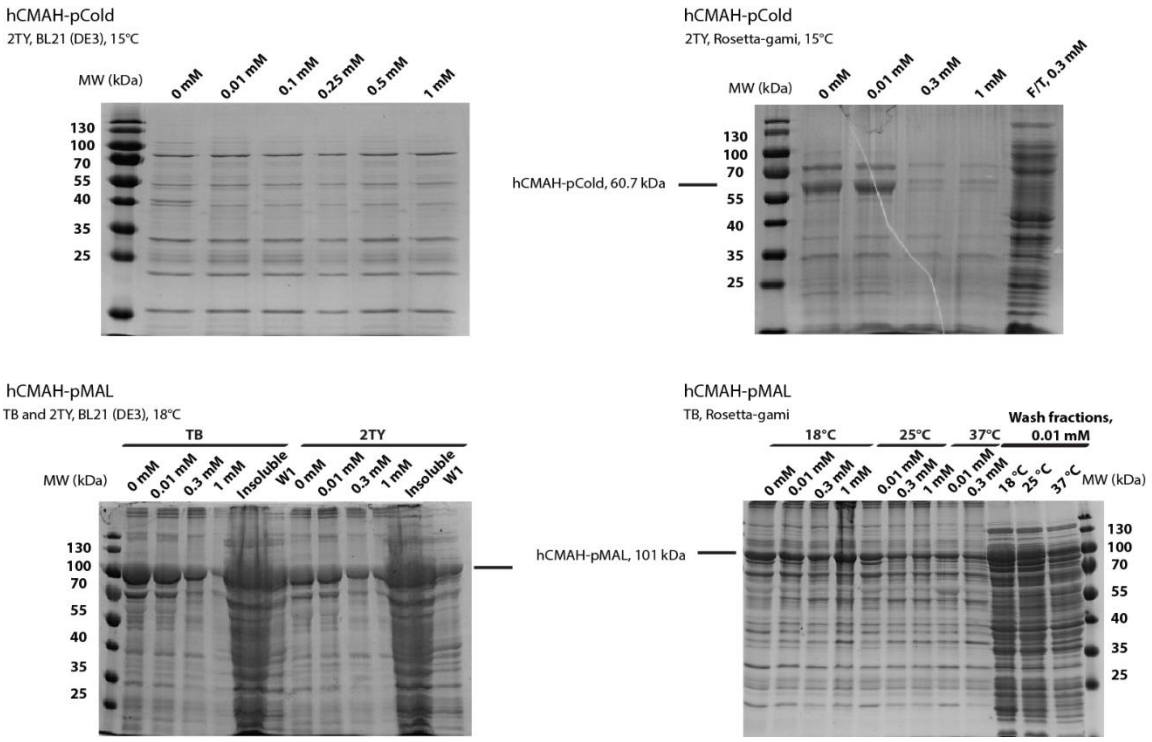


Figure 2.5: SDS-PAGE gels of expression trials for hCMAH-pCold and hCMAH-pMAL constructs in BL21(DE3) and Rosetta-gami B(DE3) pLysS cells. Cultures were grown in 2TY and TB media and protein production was induced by addition of different isopropyl β-D-1-thiogalactopyranoside (IPTG) concentrations. After induction, cultures were incubated at 15 °C (pCold construct), 18 °C, 25 °C overnight or 37 °C for 4 h. His_6-tagged proteins were isolated using Profinity™ IMAC Ni-charged resin (Bio-Rad). F/T – flowthrough, W – wash. Gels were stained using Coomassie blue stain.

In attempt to improve expression further, the hCMAH-pMAL construct was tested in BL21(DE3) and Rosetta-gami B(DE3) cells. hCMAH levels increased, particularly in BL21(DE3) cells. However, attempts to remove the *N*-terminal maltose binding protein (MBP) tag have shown incomplete tag removal (**Figure 2.6**).

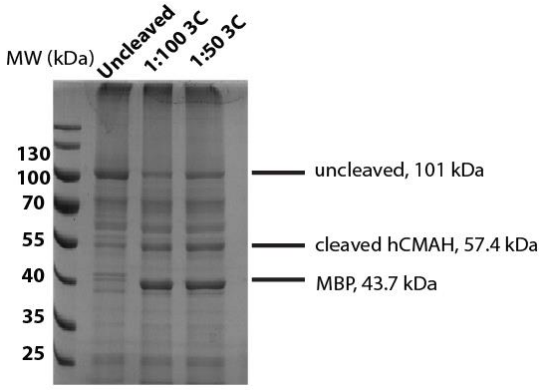


Figure 2.6: Coomassie stained SDS-PAGE gel of MBP tag cleavage by 3C protease of *N*-MBP-hCMAH-*C*-His_6 protein (hCMAH-pMAL construct). Resin-purified protein was incubated with 1:100 and 1:50 protease:protein ratio at 4 °C for 16 h.

Expression trials (**Figure 2.7**) for the mCMAH-pCold construct in BL21(DE3) cells grown in 2TY and TB media resulted in two bands observed around the predicted mCMAH molecular weight (MW, 70.1 kDa). After incubating resin-purified protein with 3C protease neither of the bands appeared to reduce in apparent MW. Expression trials were then repeated with Rosetta-gami B(DE3) cells. Two bands were consistently observed and a change in apparent MW was observed after tag cleavage. However, the levels of protein were very low, and impurities of similar MW were observed in higher abundancy than the protein of interest, so the expression was not carried out on a large scale.

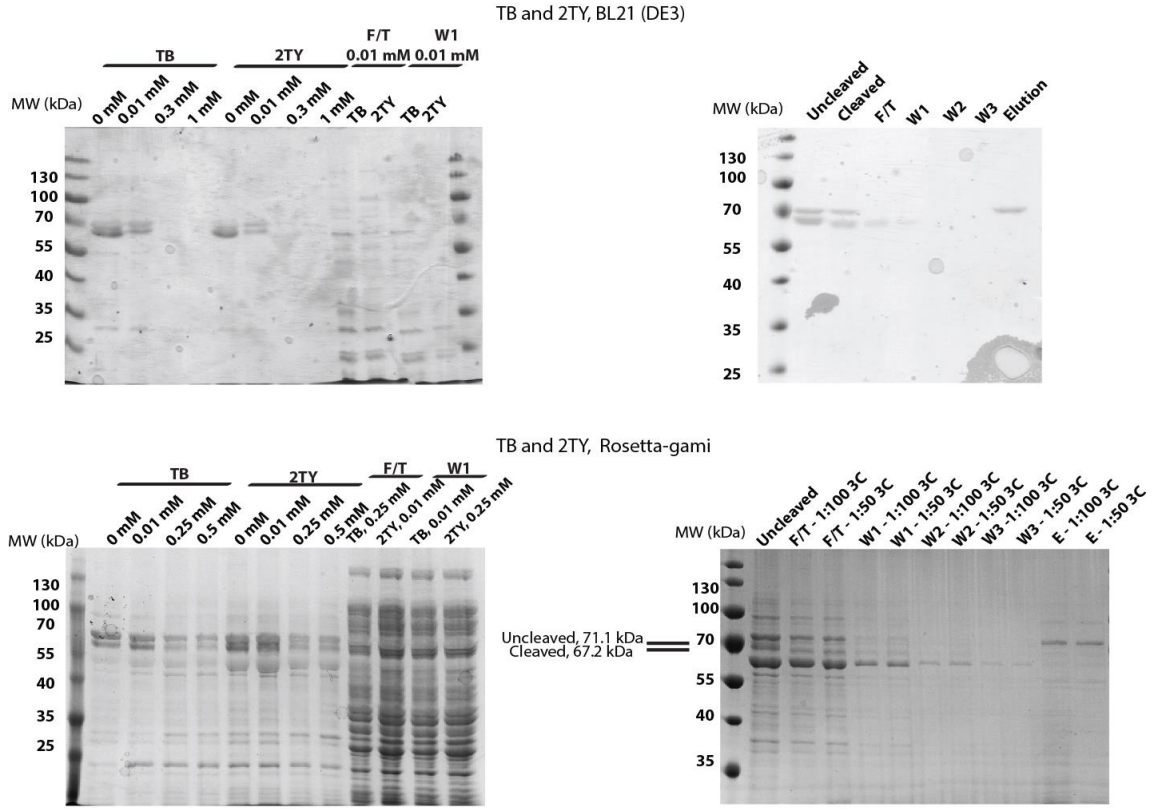


Figure 2.7: Coomassie stained SDS-PAGE gels of mCMAH expression trials in BL21(DE3) and Rosetta-gami cells grown in TB and 2TY media. Protein production was induced by addition of IPTG (concentrations shown) and cultures were incubated at 15 °C overnight. His_6-tag was cleaved by incubating the resin-purified enzyme 3C protease (1:100 protease:protein for BL21(DE3) produced protein) at 4 °C for 16 h. F/T – flowthrough, W – wash, E – elution.

2.3.3 Large scale bacterial expression

Recombinant hCMAH with an *N*-terminal His_6 tag (pCold construct, **Figure 2.8**) was produced in Rosetta gami B(DE3) cells in TB medium and purified by immobilised metal ion affinity chromatography (IMAC) followed by size exclusion chromatography, *N*-terminal tag removal by 3C protease, and a second round of IMAC. hCMAH with an *N*-terminal MBP tag and a *C*-terminal His_6 tag (pMAL construct, **Figure 2.9**) was produced on a large scale in BL21(DE3)

and Rosetta gami B(DE3) pLysS cells in TB medium and purified by IMAC followed by tag removal by 3C protease, size exclusion chromatography, and a second round of IMAC. Protein purity throughout the purification process was assessed by SDS-PAGE analysis.

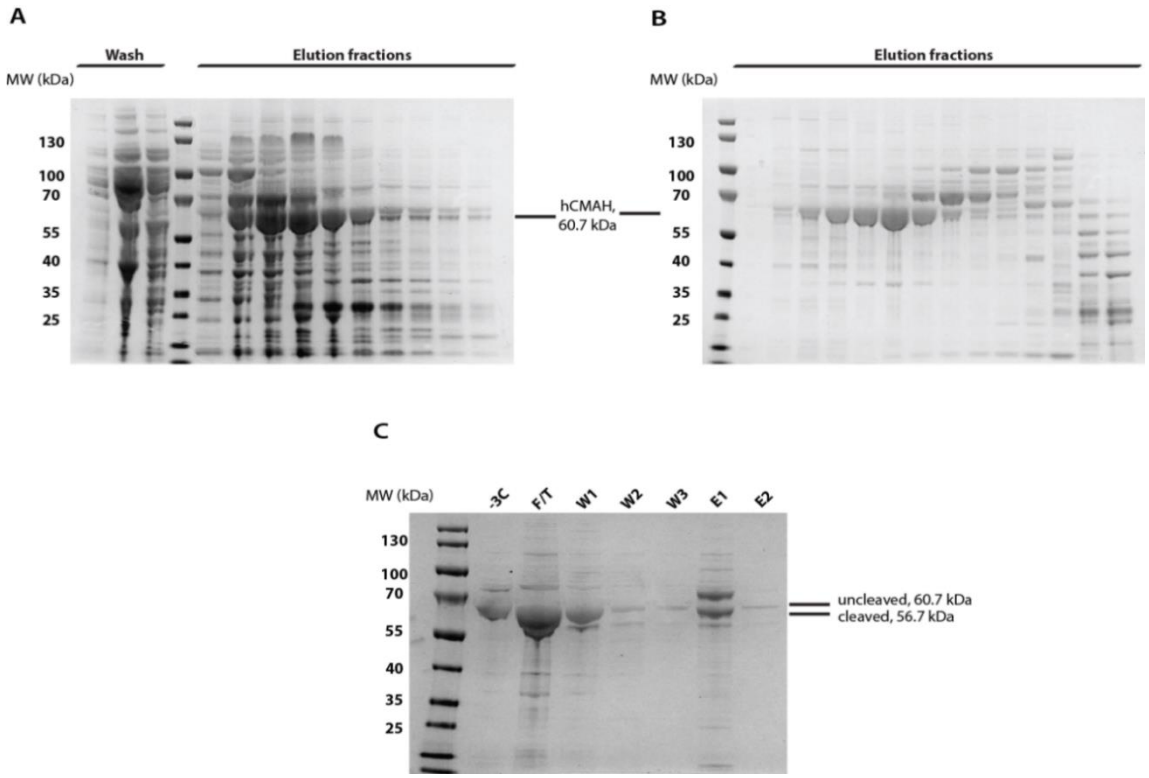


Figure 2.8: Large scale purification of hCMAH using the pCold construct. (A) hCMAH after metal ion affinity chromatography, (B) - after IMAC and size exclusion chromatography, (C) - after IMAC, size exclusion chromatography, incubation with 3C protease and a second step of IMAC. -3C – hCMAH before incubation with 3C, F/T – flowthrough, W – wash, E – elution.

Final yields of protein produced using a pCold construct were approximately 5 mg from 6 l bacterial culture. Although expression levels of MBP-tagged hCMAH were higher, final protein yields after tag removal were lower than those obtained from pCold vector, producing 4 mg and 3 mg protein when produced in BL21(DE3) and Rosetta-gami B(DE3) cells respectively.

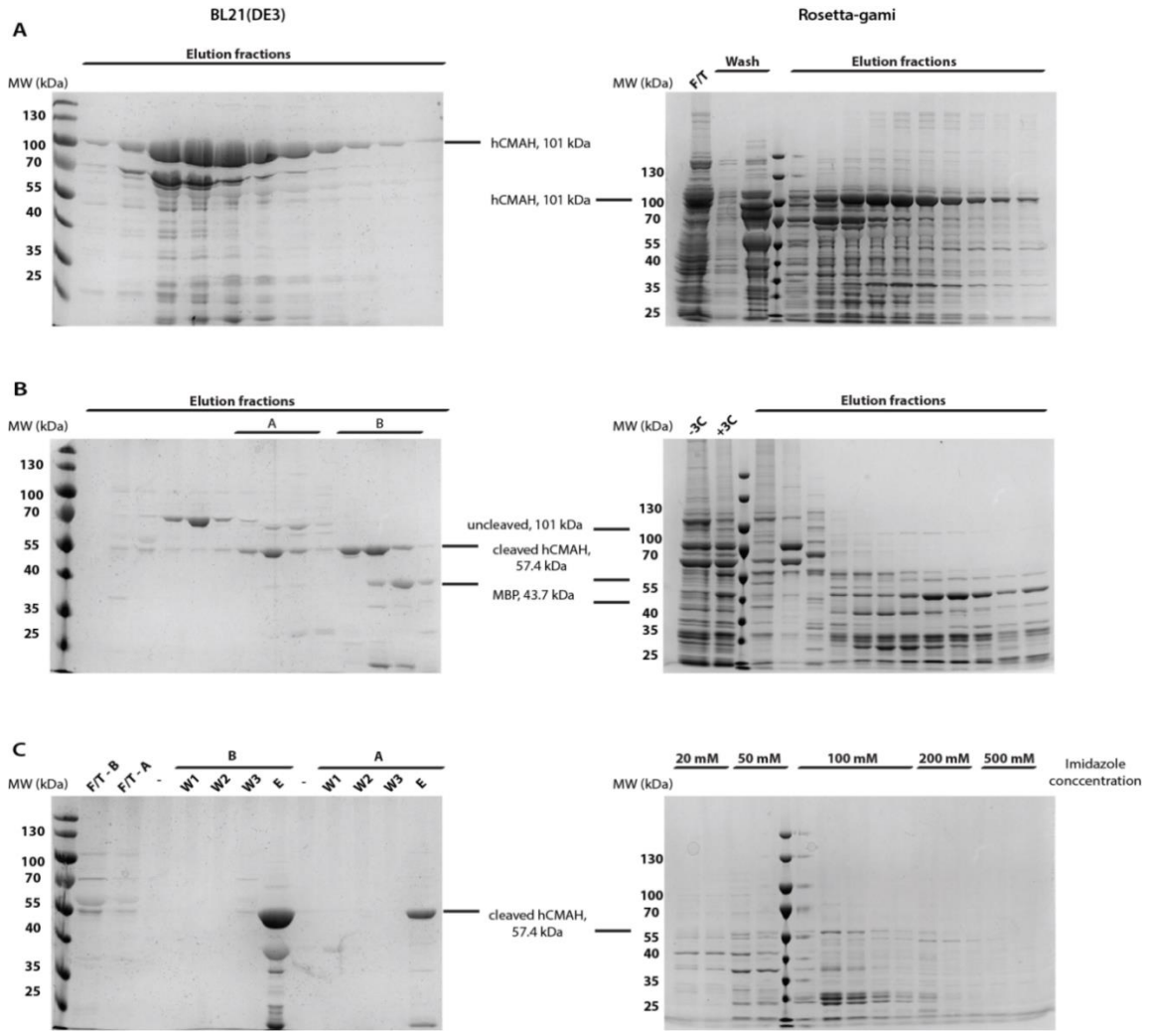


Figure 2.9: Comparison of hCMAH large scale purification from BL21(DE3) and Rosetta-gami cells. (A) – after immobilised metal ion affinity chromatography, (B) – after size exclusion chromatography (Superdex 200, 130 ml), (C) – after second immobilised metal ion affinity chromatography. F/T – flowthrough, W – wash, E – elution.

2.3.4 Mammalian expression systems

To test expression in mammalian cells, HEK293 and ExpiCHO-S cells were transfected with pcDNA3.1 vectors carrying full-length hCMAH and mCMAH coding sequences. Experimental details can be found in Chapter 6 (Materials and Methods).

The mCMAH was transiently expressed in HEK293 cells (adapted to grow in suspension by Dr Ilaria Pettinati) using the Expi293 expression system (Thermo Fisher Scientific). Media and cytosolic samples from days 4-6 post-transfection were analysed by Western blotting to assess expression levels (**Figure 2.10**). Cell viability and expression levels were reduced, compared to those observed in stably expressing HEK293 cells.

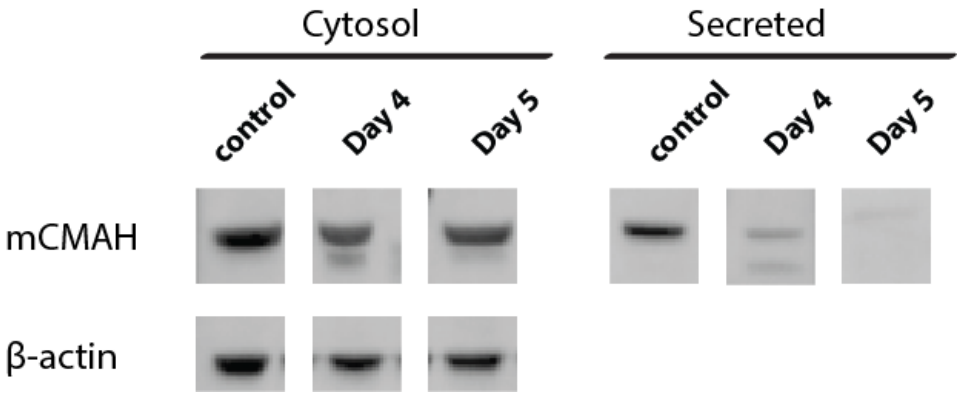


Figure 2.10: Western blot analysis of mCMAH produced using the Expi293 transient expression system. Stably expressing HEK293 cells were used as a control. β-Actin was used as a loading control for cytosolic fractions. Anti-His_6 and anti-β-actin antibodies were used in Western blot analyses.

HEK293 cells were transfected with pcDNA3.1 vectors carrying hCMAH and mCMAH encoding sequences. As pcDNA3.1 plasmid encodes for neomycin resistance, stable single cell clones were selected by treating cells with neomycin sulfate analogue G-418. Expression levels were assessed by Western blot analysis and single cell clones expressing highest levels of recombinant protein of interest were selected (**Figure 2.11A**) and adapted to grow in suspension at high density suitable for protein production[57]. Western blot analysis of cell media proteins of hCMAH and mCMAH overexpressing HEK293 cells showed both proteins to be secreted to the cell media (**Figure 2.11B**).

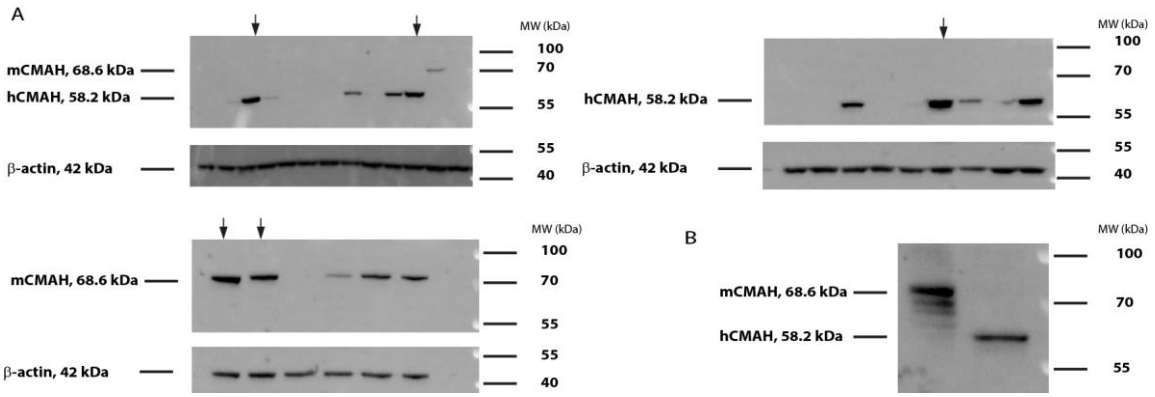


Figure 2.11: Western blot analysis of (A) single cell clones of HEK293 transfected with hCMAH and mCMAH pcDNA3.1 constructs. Transfected cells were treated with Geneticin (G-418) and 15 µg of isolated clone lysates were analysed to determine clones with highest expression levels of recombinant protein, β-actin was used as loading control. Selected clones are noted with an arrow. (B) Western blot analysis of proteins precipitated from cell growth media of selected single cell clones adapted to grow in suspension.

Recombinant hCMAH and mCMAH were purified from cell growth media by IMAC, the eluted proteins were concentrated to 100 µl and buffer exchanged to remove imidazole and reduce salt concentration. Protein purity was assessed by SDS-PAGE (**Figure 2.12**). Final protein yields

from purification were 0.3 mg and 0.5 mg per 1 l growth medium for hCMAH and mCMAH respectively.

HEK293 cells overexpressing hCMAH exhibited limited lifespan in culture with cell viability dropping below levels required for protein production within 2-3 weeks. This combined with low protein yields made HEK293 mammalian expression unsuitable for large scale protein production as required for biophysical characterisation and X-ray crystallography.

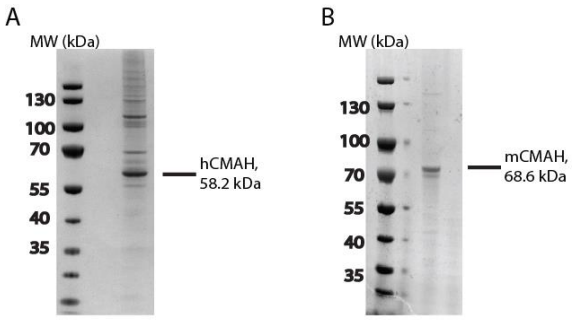


Figure 2.12: SDS-PAGE gel of hCMAH (A) and mCMAH (B) proteins produced in HEK293 cells and purified from cell growth media.

HEK293 cells overexpressing mCMAH could be maintained for longer periods of time without a significant drop in cell viability or protein expression levels. Since protein yields obtained from cell growth medium were low, proteins were purified from cell lysates by IMAC followed by *C*-terminal His_6-tag cleavage by 3C protease and a second round of IMAC. If further purification was required as determined by SDS-PAGE analysis, anion exchange chromatography was performed using Pierce Strong Anion Exchange Chromatography Spin Columns (Thermo Fisher Scientific). Protein purity was assessed by SDS-PAGE after each step (**Figure 2.13**). The purified protein solution was orange-brown in colour

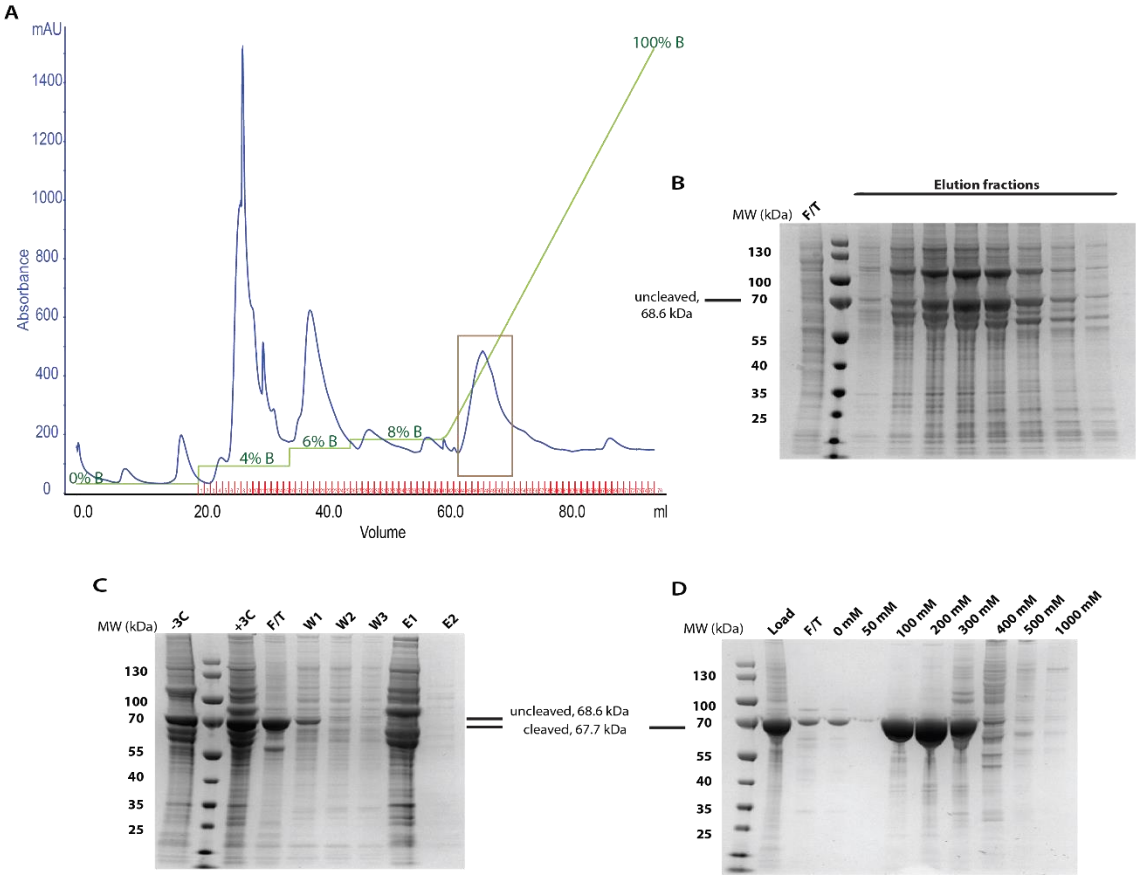


Figure 2.13: Large scale purification of mCMAH from HEK293 cell lysates. (A) – AKTA FPLC trace of mCMAH purification from mammalian cell lysates showing UV trace in blue. The green line shows concentration of HisTrap buffer B (containing 500 mM imidazole) with percentages given in green. The peak corresponding to mCMAH is indicated in brown. SDS-PAGE gels at different stages of purification: (B) - after metal ion affinity chromatography, (C) - after metal ion affinity chromatography, incubation with 3C protease and a second round of metal ion affinity chromatography, -3C mCMAH prior to incubation with 3C protease, +3C – mCMAH after the incubation with 3C protease. (D) - after additional purification using Pierce Strong Anion Exchange Chromatography Spin Columns, concentrations of NaCl are listed. F/T – flowthrough, W – wash, E – elution.

hCMAH and mCMAH (pcDNA3.1 constructs) were transiently transfected and expressed using the ExpiCHO expression system max titer protocol (Thermo Fisher sScientific). Cell media and cytosolic samples from days 4 and 5 post-transfection were analysed by Western blotting to assess recombinant protein expression levels. Cell viability of cells transfected with hCMAH fell below 70% on day 4 and protein levels were low as demonstrated by Western blot analysis (**Figure 2.14A**). mCMAH protein levels were higher (**Figure 2.14B**) than when using the Expi293 system but not as high as what was observed in stable HEK293 cell lines.

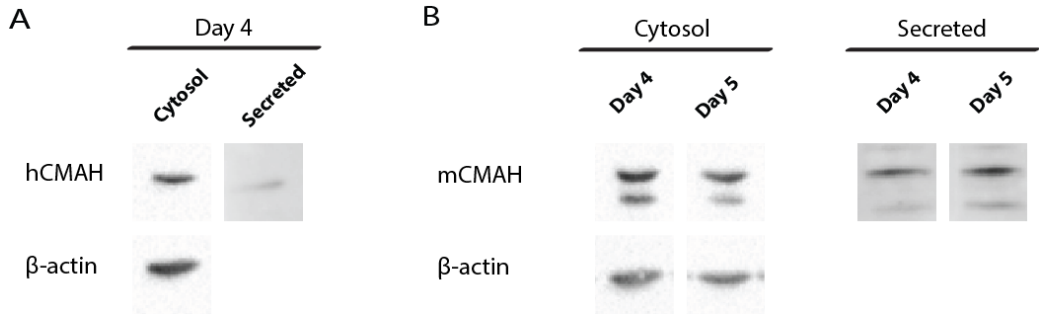


Figure 2.14: Western blot analysis of cytosolic and secreted (A) hCMAH and (B) mCMAH proteins produced using an ExpiCHO expression system. β-Actin was used as a loading control for cytosolic fractions.

Cell pellets from the transient expression experiment were lysed and recombinant mCMAH was purified by immobilised metal ion affinity chromatography. Obtained protein appeared as multiple bands on SDS-PAGE ranging from 55 to 70 kDa (**Figure 2.15**), Western blot analysis confirmed that all the bands represented different forms of mCMAH. Protein yields from ExpiCHO expression system were lower than yields from stably expressing HEK293 and the resulting protein sample was non-homogeneous making it unsuitable for further applications.

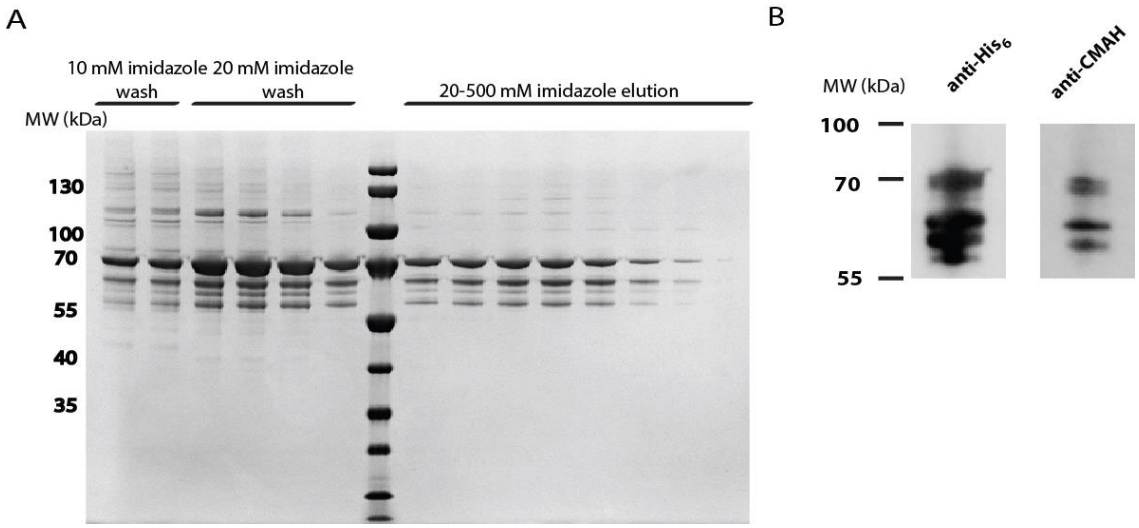


Figure 2.15: Purification of mCMAH from ExpiCHO-S cells. (A) – SDS-PAGE gel of fractions collected at different stages of purification. (B) – Western blot analysis of protein obtained from purification using anti-His_6 and anti-CMAH antibodies showing non-homogeneous mCMAH sample.

2.4 Characterisation

2.4.1 Metal content analysis

The metal content of purified hCMAH and mCMAH proteins was analysed by Particle-induced X-ray emission (microPIXE) experiments conducted by Peter Woodcock (Elspeth Garman group, Department of Biochemistry, University of Oxford).

In order to obtain a more sensitive measurement, the metal content of purified hCMAH and mCMAH was also analysed by Inductively coupled plasma mass spectrometry (ICP-MS). Samples were submitted for analysis to Department of Earth Sciences, University of Oxford.

Both analyses showed very low amounts of iron in hCMAH samples, with significantly larger amounts of iron in mCMAH samples. ICP-MS being a more sensitive technique detected trace amounts of several other metals (Ni – likely from IMAC column, Zn, Cu) which were not detectable by microPIXE. Purified hCMAH (both from bacterial and mammalian cells) contained 0.1-0.2 atoms of iron per protein molecule which was less than expected due to hCMAH still possessing an MBL domain and potentially being able to bind the mononuclear Fe ion predicted to be found in its active full-length homologues. For mCMAH, both PIXE and ICP-MS experiments have shown less than the predicted 3 iron atoms per protein molecule (approximately 1.7) suggesting that some of the protein has lost its [FeS] cluster and/or the active site Fe ion.

2.4.2 Deglycosylation experiments

Protein sequences were analysed using the NetNGlyc server[58] to check for predicted glycosylation sites. Both hCMAH and mCMAH were predicted to contain one potential *N*-glycosylation site. As bacteria lack the molecular machinery required for this post-translational modification[59] and mammalian expression of hCMAH did not yield sufficient protein, potential hCMAH glycosylation was not investigated.

To check for glycosylation, mCMAH purified from HEK293 cell lysates was treated with PNGase F under denaturing and non-denaturing conditions (New England Biolabs, according to manufacturer's instructions). No changes in apparent molecular weight were observed following SDS-PAGE analysis (**Figure 2.16A**, positive control not shown). While purified mCMAH appeared as a single band on SDS-PAGE gels, Western blot analysis of HEK293 lysates consistently showed multiple bands, suggesting the presence of multiple forms of mCMAH in cell lysates, possibly explained by heterogeneous protein glycosylation. However, only single species could be purified using the selected chromatographic techniques.

As mCMAH purified from ExpiCHO cells was observed as multiple bands on SDS-PAGE, it was also treated with PNGase F. Similarly to mCMAH purified from HEK293 cells, no changes in apparent molecular weight were observed (**Figure 2.16B**), suggesting that multiple bands observed after SDS-PAGE analysis are not a result of differences between glycosylation states.

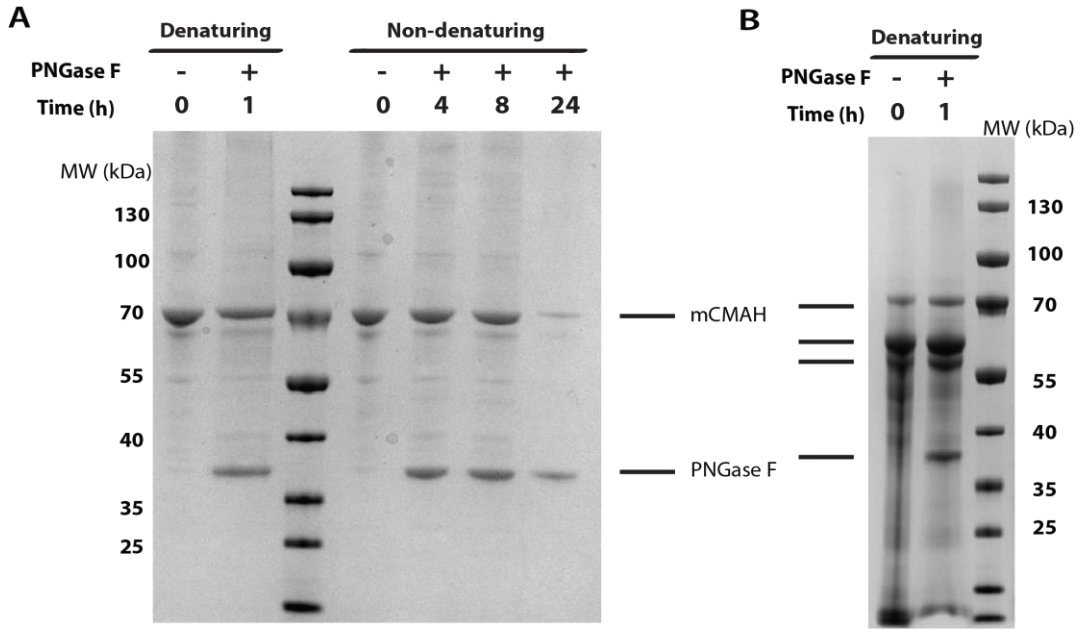


Figure 2.16: Treatment of purified mCMAH with PNGase F. (A) mCMAH purified from HEK293 cells treated with PNGase F under denaturing and non-denaturing conditions. (B) mCMAH purified from ExpiCHO-S cells treated with PNGase F under denaturing conditions.

2.4.3 Mass spectrometry and Circular Dichroism spectrosopy

Purified CMAH proteins were subjected to mass spectrometry (MS) analysis. Predicted and deconvoluted molecular weights are listed in **Table 2.2**. The extent of mCMAH protein folding and thermal stability was assessed using circular dichroism (CD) spectroscopy. Thermal stability and possibility of refolding of mCMAH were also assessed.

The deconvoluted mass spectrum of hCMAH purified from bacterial cells showed a major peak of 57198 Da (**Figure 2.17A**). However, since hCMAH obtained from HEK cells did not ionise well (peak of 58497 Da, **Figure 2.17B**), it is likely that hCMAH produced in E. coli cells did not ionise well either and the correct mass peak would be hindered by a well ionising impurity.

Table 2.2: Predicted and deconvoluted masses of hCMAH (**Figure 2.17**) and mCMAH (**Figure 2.18**) proteins. Observed masses were determined by denaturing mass spectrometry. *Mass difference for hCMAH (pCold construct) might be due to a contaminating protein. ** Ionisation of hCMAH (pcDNA3.1 construct) was too poor for accurate mass determination, multiple peaks were present. Spectra were deconvoluted using Agilent MassHunter Protein Deconvolution Software across the appropriate mass range.

Protein	Predicted mass (Da)	Deconvoluted mass (Da)
hCMAH (pCold construct)	56746	57198*
hCMAH (pcDNA3.1 construct)	58216	58497**
mCMAH (pcDNA3.1 construct)	67666	67708

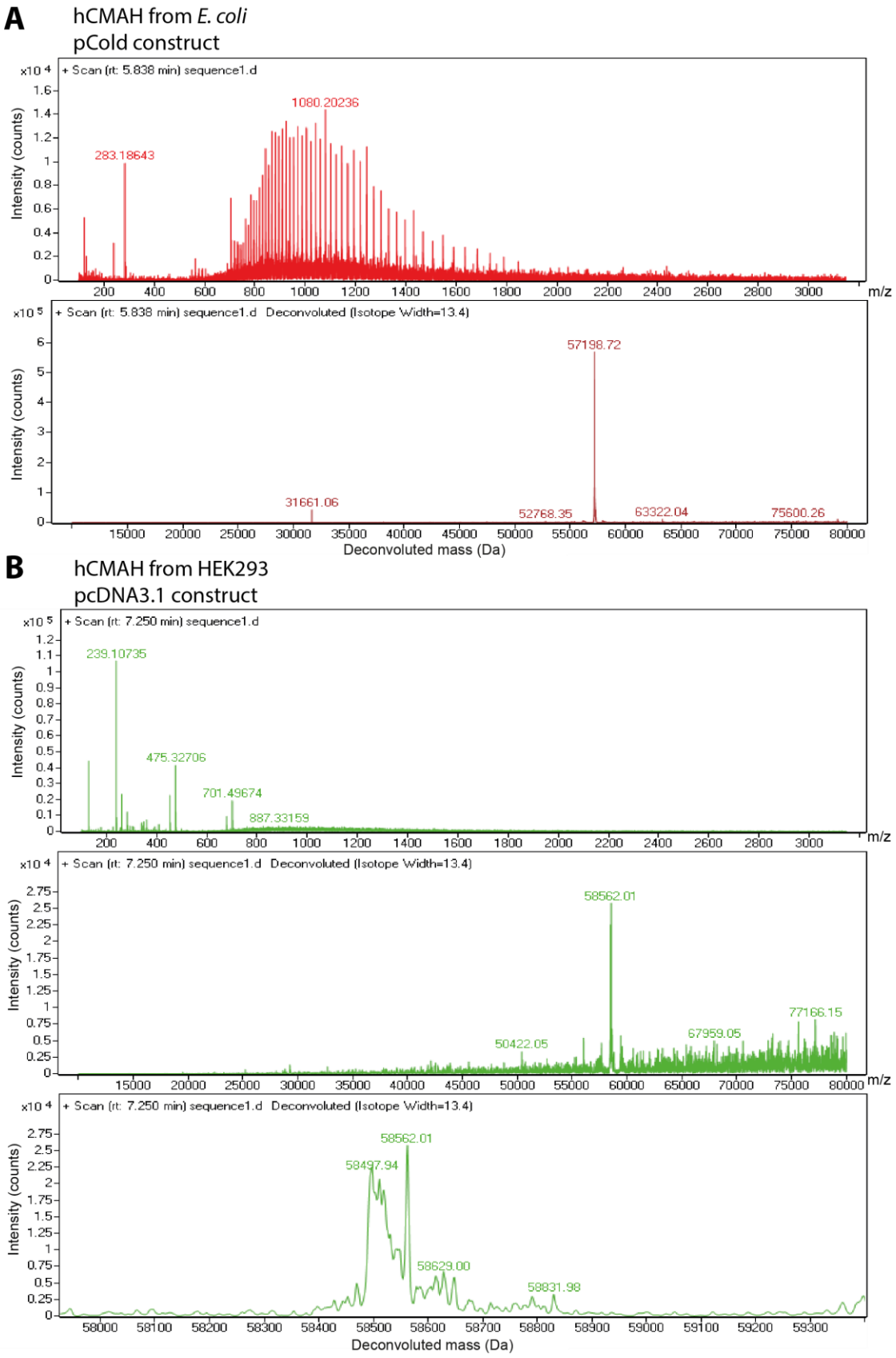


Figure 2.17: Denaturing MS spectra (acquired m/z and deconvoluted) for hCMAH purified from (A) *E. coli* and (B) HEK293 cells, including a magnified view. Note differences in scale for mass-to-charge spectra. Spectra were deconvoluted using Agilent MassHunter Protein Deconvolution Software.

Purified mCMAH was 42 Da larger than the predicted molecular weight of 67666 Da (**Figure 2.18**). The cause of the mass difference was unclear, however could be a result of a post-translational modification such as acetylation (+42 Da). No higher molecular weight adducts were observed, which is consistent with no glycosylation being detected by PNGase F treatment. Purified mCMAH was shown by CD analysis to be folded with T_m of 65.2 ± 0.9 °C (**Figure 2.19**).

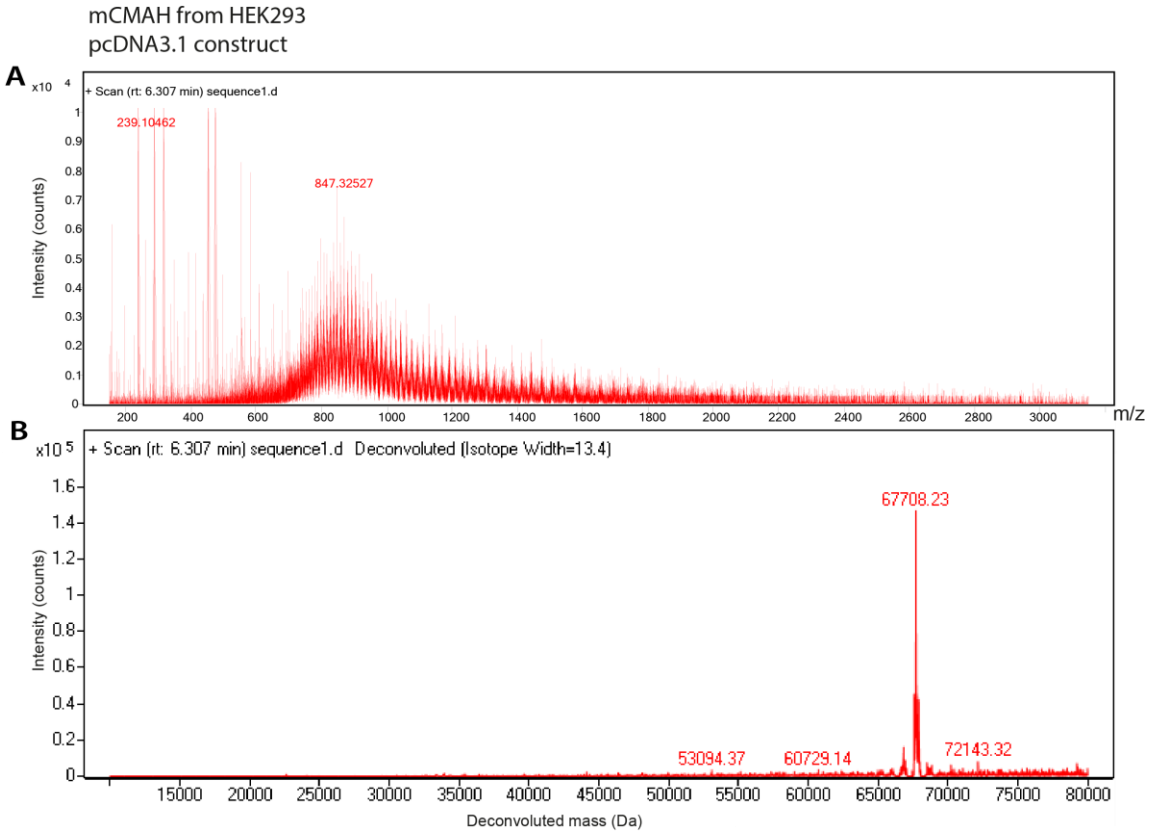


Figure 2.18: Denaturing MS spectra mCMAH. (A) Acquired m/z spectrum and (B) deconvoluted spectrum. Spectrum was deconvoluted using Agilent MassHunter Protein Deconvolution Software.

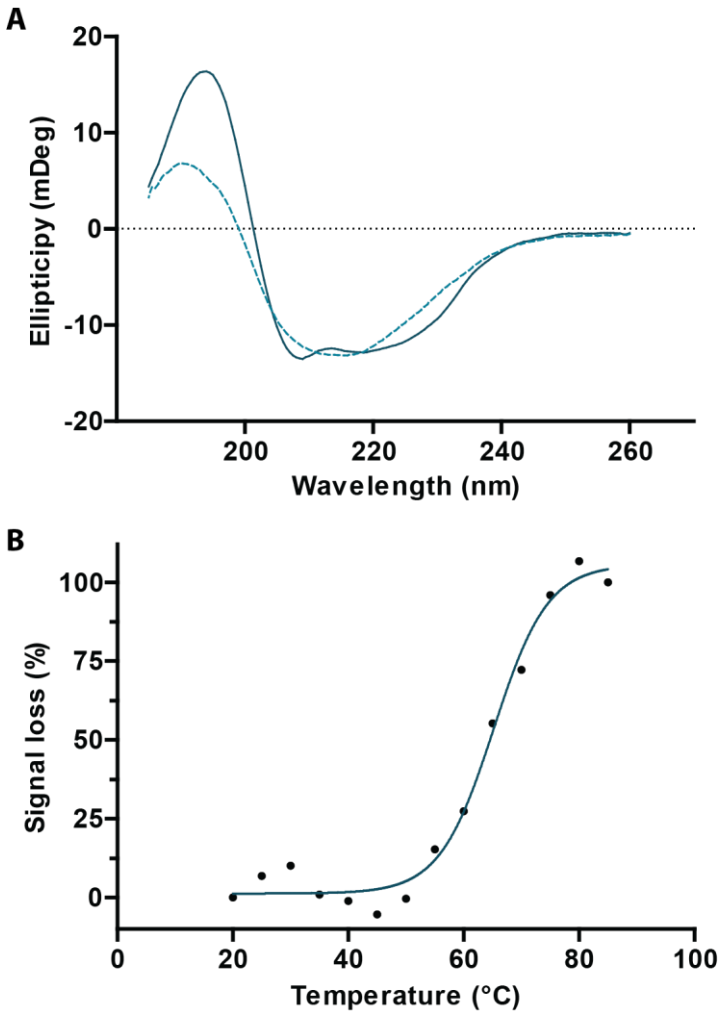


Figure 2.19: Circular dichroism spectra of purified mCMAH. (A) Spectrum of mCMAH at 20 °C (solid line) and 80 °C (dashed line). (B) Melting curve of mCMAH as determined by circular dichroism. Curve fitting was performed using Origin Pro[60].

2.4.4 Oligomerisation state analysis

Purified mCMAH was subjected to size exclusion chromatography – multiangle light scattering (SEC-MALS) analysis to investigate its oligomerisation state. SEC-MALS experiments were performed using Superdex 200 column by Dr David Staunton (Molecular Biophysics Suite, Department of Biochemistry, University of Oxford).

SEC-MALS analysis has shown that mCMAH is primarily a monomer (major peak of approximately 61.2 kDa) with a small dimeric population being observed (minor peak of approximately 107 kDa, **Figure 2.20**). As mCMAH has been previously reported to be monomeric[24], the small dimeric population observed may be an artefact of isolation and purification of mCMAH.

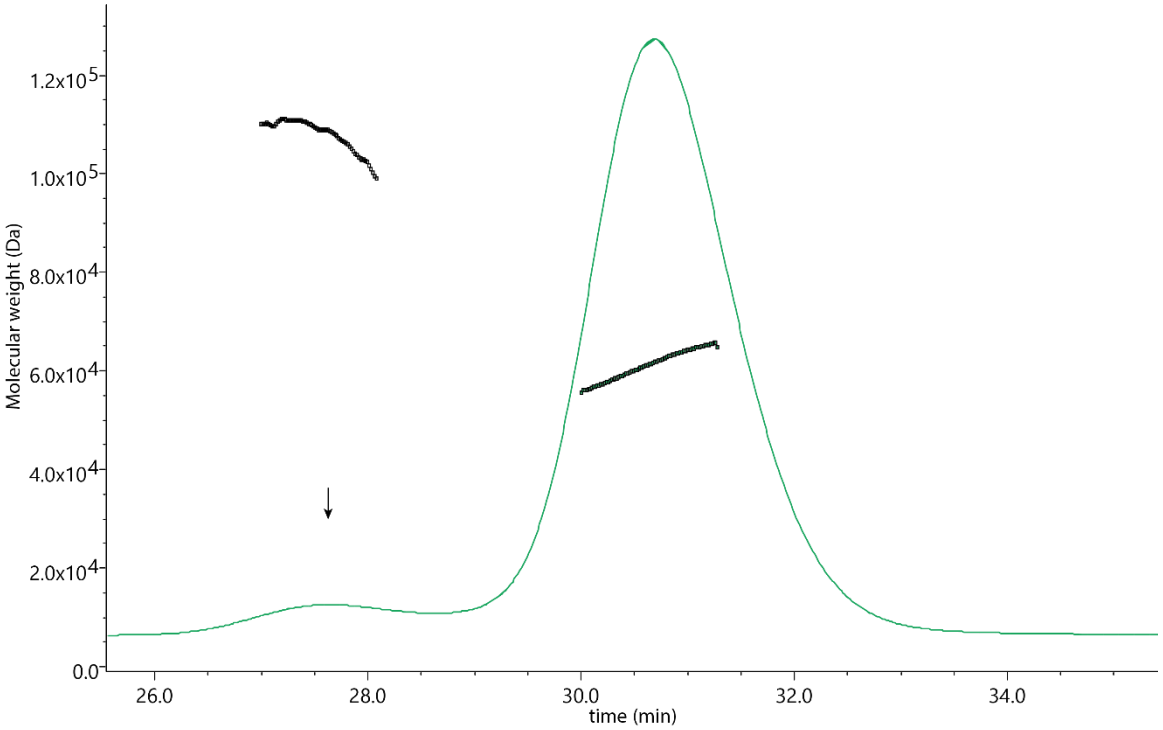


Figure 2.20: SEC-MALS analysis of purified mCMAH. Major peak of 61.2 kDa corresponds to monomeric form of mCMAH, while the minor peak (marked by an arrow) corresponds to dimeric form of mCMAH.

Purified mCMAH was further analysed using non-denaturing (native) MS employing electrospray ionisation in the positive ion mode. High resolution full scan non-denaturing MS spectra were recorded by Shuang Liu. Apo mCMAH enzyme was generated by dialysing purified mCMAH against ethylenediaminetetraacetic acid (EDTA) and phenanthroline (for experimental details see Chapter 6), and 'as-purified' and 'apo' mCMAH samples were analysed. Both samples appeared monomeric with a visible mass difference between the two samples (**Figure 2.21**). The exact mass of either 'as-purified' or 'apo' mCMAH, and the exact difference between the two could not be determined with confidence due to resolution limits, with the observed difference falling between 230-330 Da. This difference falls in range of a predicted loss of a single Fe ion and a [2Fe2S] cluster (approximately 232 Da). Further experiments and optimisation of conditions are needed to improve resolution to determine the nature of the mass difference.

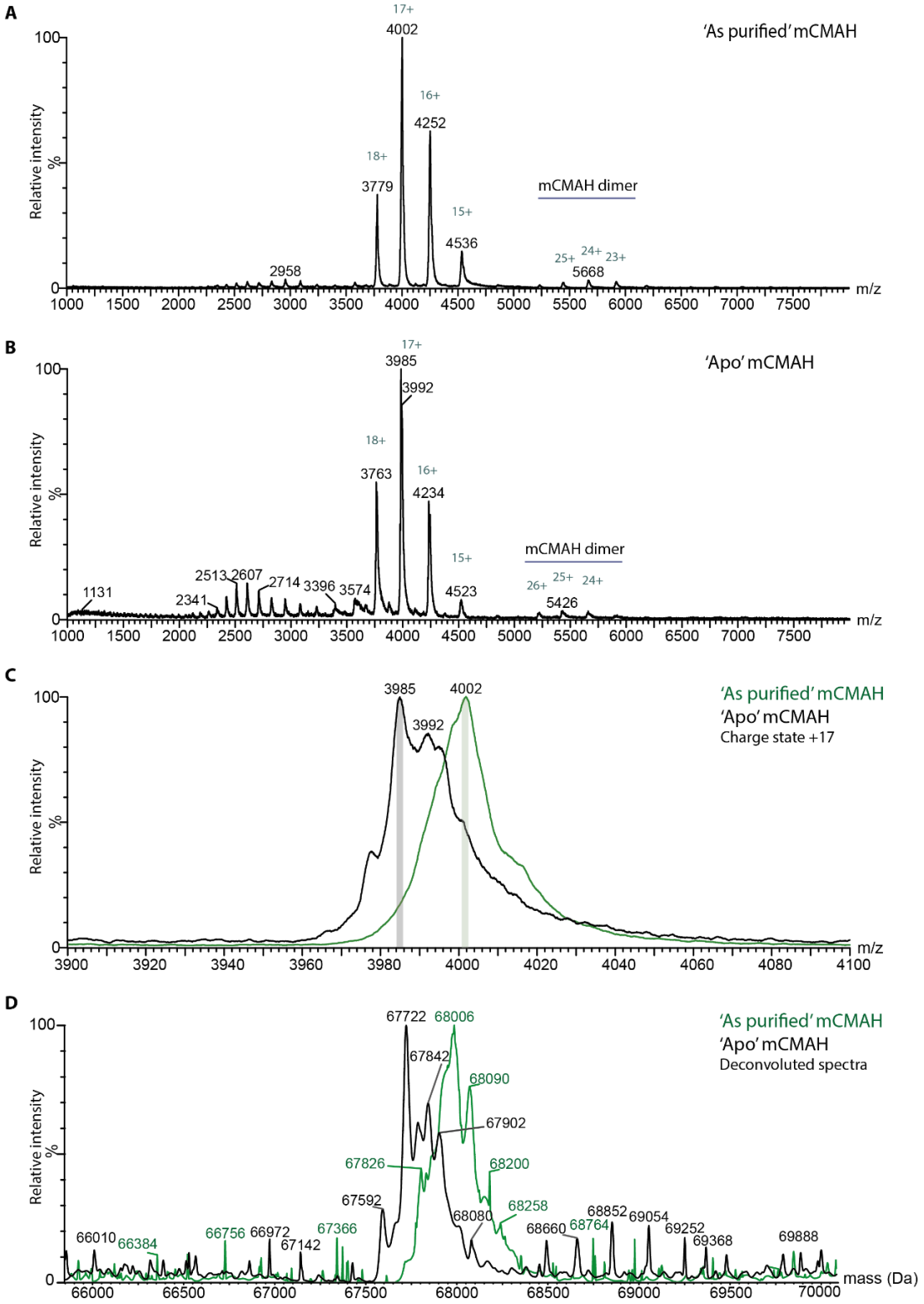


Figure 2.21: Non-denaturing MS spectra of mCMAH. (A) mCMAH as purified from HEK293 cells was primarily monomeric with a small dimeric population observed. (B) Primarily monomeric 'Apo' mCMAH with a small dimeric population observed. (C) Comparison of 17+ charge state peak for 'as purified' (green) and 'apo' mCMAH proteins showing a mass shift following a treatment with 20 mM ethylenediaminetetraacetic acid (EDTA). (D) Deconvoluted spectra of 'as purified' (green) and 'apo' mCMAH (black). The spectra were deconvoluted using MaxEnt1 feature of MassLynx 4.0 software (Waters) over an appropriate mass range.

2.5 Characterisation of FeS cluster

2.5.1 UV-Vis spectroscopy

The presence of a Rieske [2Fe2S] cluster in purified mCMAH was confirmed by UV-visible spectrometry (UV-Vis). A trace characteristic to oxidised Rieske [2Fe2S] cluster[17] with peaks at 330 nm, 460 nm and a shoulder at 570 nm was observed in mCMAH samples purified from stably expressing HEK293 lysates (**Figure 2.22**). No peaks were observed in samples purified from ExpiCHO-S cells, which in conjunction with obtained sample being non-homogenous further suggests that the ExpiCHO expression system is unsuitable for mCMAH production.

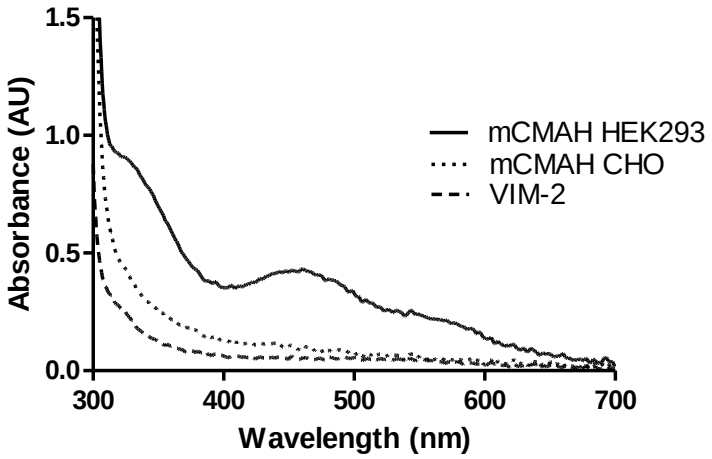


Figure 2.22: UV-vis spectrum of mCMAH proteins purified from HEK293 stably expressing cell lysastes (solid line), ExpiCHO expression system (dotted line) and a negative control bacterial MBL VIM-2 (dashed line). All proteins were at 5.6 mg/ml (80 µM for mCMAH, 200 µM for VIM-2).

2.5.2 Electron paramagnetic resonance

Electron paramagnetic resonance (EPR) is a technique used to study atoms or molecules containing unpaired electrons and is often used to study iron and iron-sulfur cluster containing proteins. As mCMAH is reported to contain [2Fe2S] cluster and a mononuclear iron, EPR was used to assess the identity and environment of mCMAH metal co-factors.

EPR measurements were run by Dr William Myers (Centre for Advanced Electron Spin Resonance, University of Oxford). Dr Emily Flashman assisted with sample preparation. A purified mCMAH protein sample was concentrated to 200 µM and supplemented with 20% glycerol to prevent ice crystal formation upon freezing. As the [2Fe2S] cluster should be in its Fe(III)-Fe(II) reduced form to be detectable by EPR, a reduced sample was prepared by treating mCMAH with 2 mM sodium dithionite for 3 min under anaerobic conditions (<10 ppm oxygen concentration, Belle Technology glove-box). Air-prepared oxidised and

dithionite-reduced samples were transferred to 4mm Standard Natural Quartz EPR Tubes (250 mm length, Norell) and flash frozen in liquid nitrogen.

Two distinct EPR signals were observed in both 'oxidised' and 'reduced' mCMAH samples. One of the observed signals (g = 4.288, **Figure 2.23**) is characteristic of mononuclear high-spin (S = 5/2) Fe(III) ions with rhombic geometry[61].

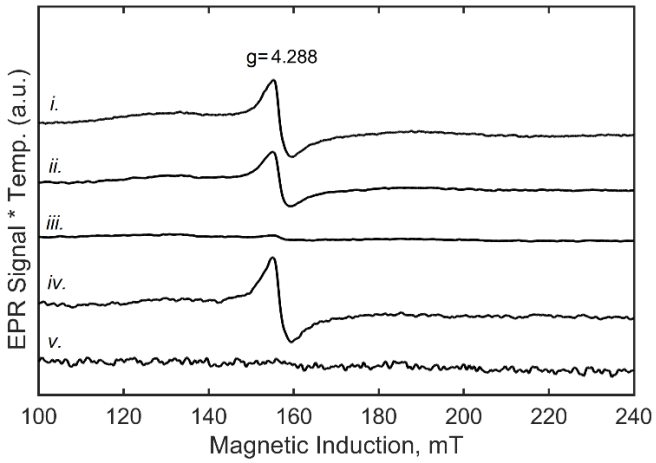


Figure 2.23: EPR signal of Fe(III) species observed in mCMAH. (i) Reduced sample, 5 K; (ii), oxidised sample 5 K; (iii) buffer control, 5 K; (iv) reduced sample, 15 K; (v) buffer 15 K. Spectra were recorded at 9.386 GHz microwave frequency and 20 μW microwave power. The figure was made by Dr William Myers.

The observed signal for the iron-sulfur cluster (g-values 1.940, 1.878, 1.866, **Figure 2.24**) did not match to the literature values characteristic for Rieske [2Fe2S] cluster of approximately 2.00-2.05, 1.85-1.95 and 1.70-1.80[62, 63], nor any other type of FeS cluster reported[17] (**Table 2.3**). Additionally, the signal observed broadened at 30 K temperature with broadening beyond detection observed at 45 K, contrary to literature examples for [2Fe2S] clusters which are reportedly the only type of iron-sulfur clusters that can be studied at temperatures exceeding 45 K.[64] In order to further investigate the unusual nature of the FeS cluster Electron Spin Echo Envelope Modulation (ESEEM) experiments were run at 5 K, g = 1.876 over 12 h. However, data quality was too poor to draw any conclusions (data not shown). Further investigation is needed to determine the cause of the unusual signal.

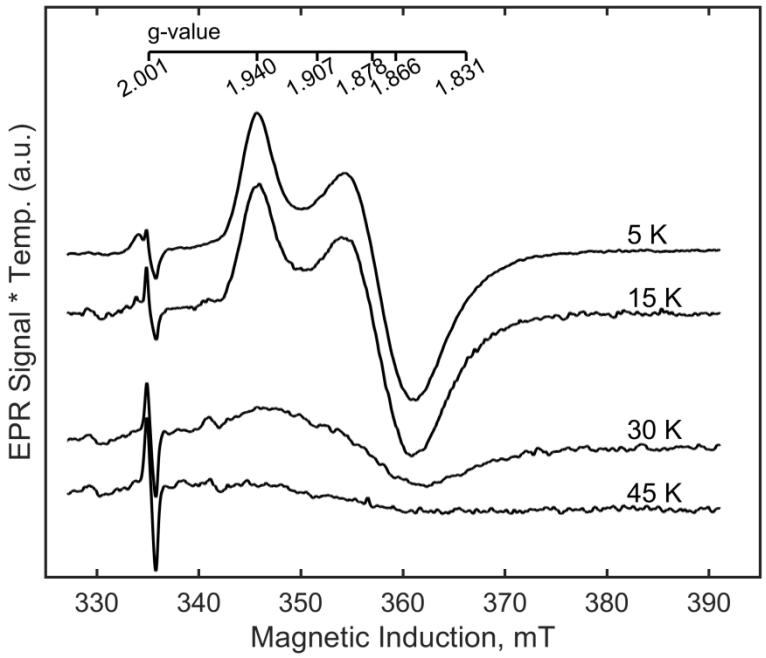


Figure 2.24: EPR analysis of the signal observed in the region of iron-sulfur cluster at increasing temperatures. Spectra were recorded at 9.386 GHz microwave frequency and 20 μW microwave power. The figure was made by Dr William Myers.

Table 2.3: Literature g-values of representative Rieske [2Fe2S] cluster proteins including the pig CMAH homologue.

Protein	g_z	g_y	g_x	Reference
Mus musculus CMAH (mCMAH)	1.940	1.878	1.831	This thesis
Sus scrofa (pig) CMAH	2.01	1.91	1.78	Schlenzka et al. [17]
Thermus thermophilus Rieske protein	2.02	1.90	1.80	Oliveira et al. [65]
Pseudomonas putida Benzoate 1,2-dioxygenase	2.01	1.91	1.77	Wolfe et al. [66]
Nostoc sp. PCC 7906 PetC	2.03	1.89	1.74	Holton et al. [63]
Rhodococcus sp. Strain NCIMB12038 Naphthalene dioxygenase	2.01	1.91	1.71	Baratto et al. [62]

2.6 Hydroxylase activity assay

Previously reported CMAH hydroxylation assays (**Figure 2.25**) were performed using ^{14}C-labelled CMP-Neu5Ac, quenched by addition of HCl or ethanol, and analysed by radio-thin-layer chromatography (radio-TLC).[20, 22] Another reported assay used non-labelled CMP-Neu5Ac and reaction products were analysed by high-performance liquid chromatography (HPLC).[21, 23] In order to eliminate the quenching step and allow for a continuous monitoring of

the mCMAH hydroxylation reaction, mCMAH catalysed hydroxylation was assayed using RapidFire sample delivery system coupled with quadrupole time-of-flight (Q-TOF) mass spectrometry (RapidFire-MS, **Figure 2.25C**). The RapidFire high-throughput delivery system replaces the conventional liquid chromatography (LC) column with an integrated solid phase extraction (SPE) cartridge that removes non-volatile buffer salts prior to elution onto the Q-TOF[67].

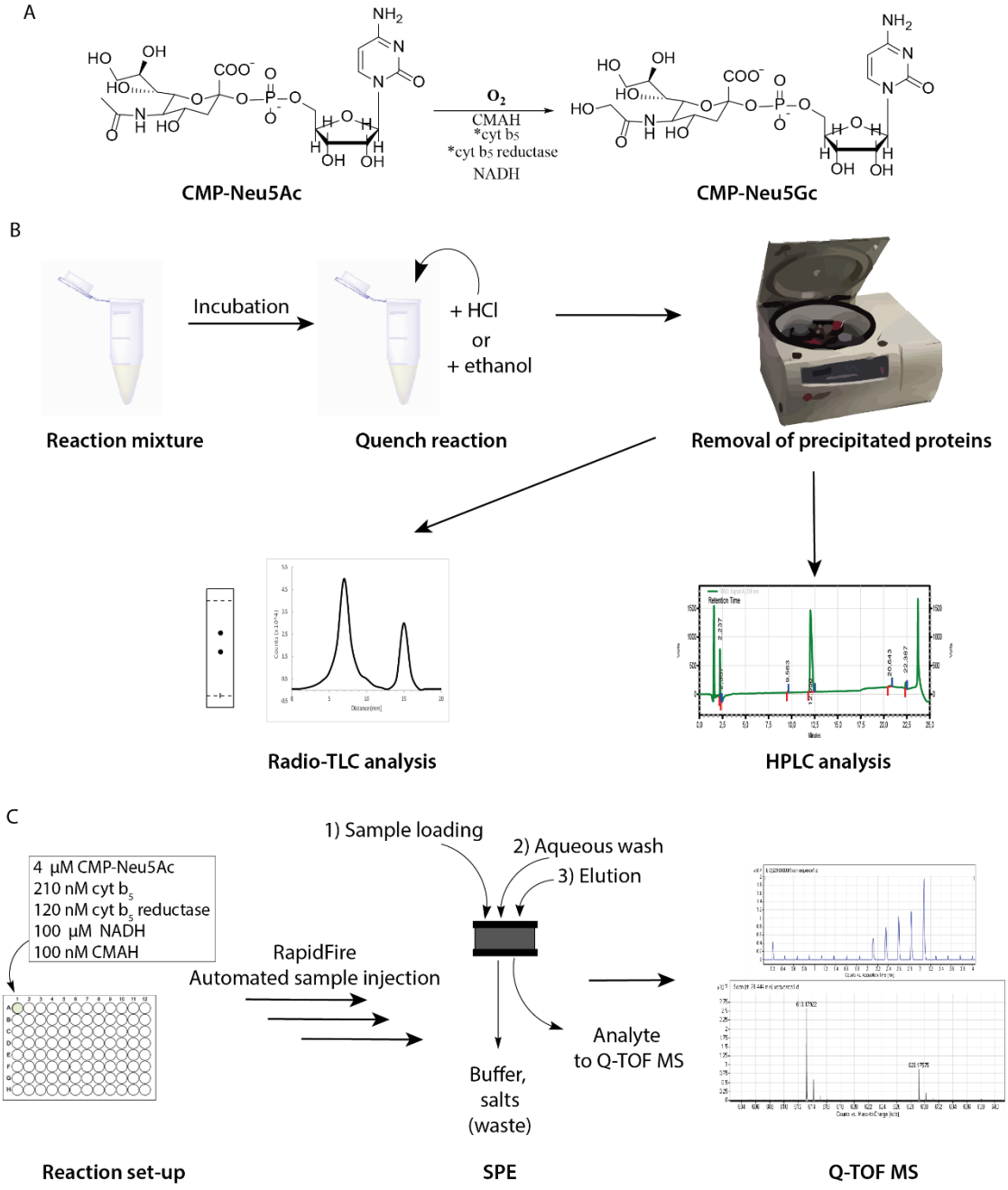


Figure 2.25: Assays for CMAH activity. (A) CMP-Neu5Ac hydroxylation reaction as catalysed by CMAH. *Microsomes or cell lysates were used as sources of cyt b_5 and cyt b_5 reductase in assays reported in literature using ^{14}C-labelled CMP-Neu5Ac (radio-TLC route)[20, 22] and non-labelled CMP-Neu5Ac (HPLC route)[21, 23]. (B) Outline of CMAH activity assays reported in literature. (C) Outline of the newly developed assay set-up using RapidFire-MS system.

2.6.1 Assay optimisation and NADH decomposition

Initial reaction conditions for CMP-Neu5Ac hydroxylation were adapted from the literature[18, 29]. To determine optimal substrate concentration and measurement parameters, increasing concentrations (0-80 μM) of CMP-Neu5Ac were run in the negative ion mode using 6 mM octylammonium acetate (OAA, synthesised by Roman Belle) buffer as the aqueous phase. OAA has an 8 carbon tail, which binds with the SPE cartridge, and a positively charged ammonium group, which forms an ion-pair with negatively charged phosphate group[68]. Optimal nozzle voltage was determined by recording increasing concentrations of CMP-Neu5Ac at varied nozzle voltages (0 – 1000 V in 250 V intervals, **Figure 2.26A**). While use of 0 V and 250 V showed little difference in signal intensity, measurements recorded using higher nozzle voltages showed decreased signal intensity. Further measurements were recorded with nozzle voltage of 0 V. The same concentrations of CMP-Neu5Ac were then run using phenyl, C8 and C18 cartridges with signal intensity being significantly higher when using C18 compared to C8 and phenyl cartridges (**Figure 2.26B**). Since detector saturation was observed using a C18 cartridge (i.e. non-linear response to increasing concentrations), lower substrate concentrations (0-10 μM) were screened (**Figure 2.26C**). Considering the signal suppression which would occur when proteins (mCMAH, cyt b_5 and cyt b_5 reductase) were added to the mixture, 4 μM final concentration of CMP-Neu5Ac in the reaction mixture was chosen for the assays.

For initial reactions, mouse liver microsomes (Sigma Aldrich) were used as a source of cyt b_5 and cyt b_5 reductase. However, it was observed that NADH was rapidly degraded by some component of the microsomes before the addition of mCMAH to the reaction mixture (**Figure 2.27**). The molecular mass observed for the degraded NADH was 558 Da which was consistent with the predicted mass of NADH following the hydrolysis of a nicotinamide group.

NADH degradation in the reaction mixture using purified human cyt b_5 and cyt b_5 reductase was not observed initially using freshly obtained proteins (purified recombinant cyt b_5 was purchased from Abcam, cyt b_5 reductase from ProSpec). However, after longer storage at -20 °C, NADH decomposition was observed as well. Several new mass peaks were observed, however, no 558 Da peak was detected suggesting a different type of degradation.

NADH degradation was overcome by increasing the NADH concentration in the reaction mixture to 100 μM and adding NADH to the reaction mixture immediately prior to the addition of mCMAH. Addition of fresh NADH to the reaction mixture already containing degraded NADH appeared to re-start the reaction.

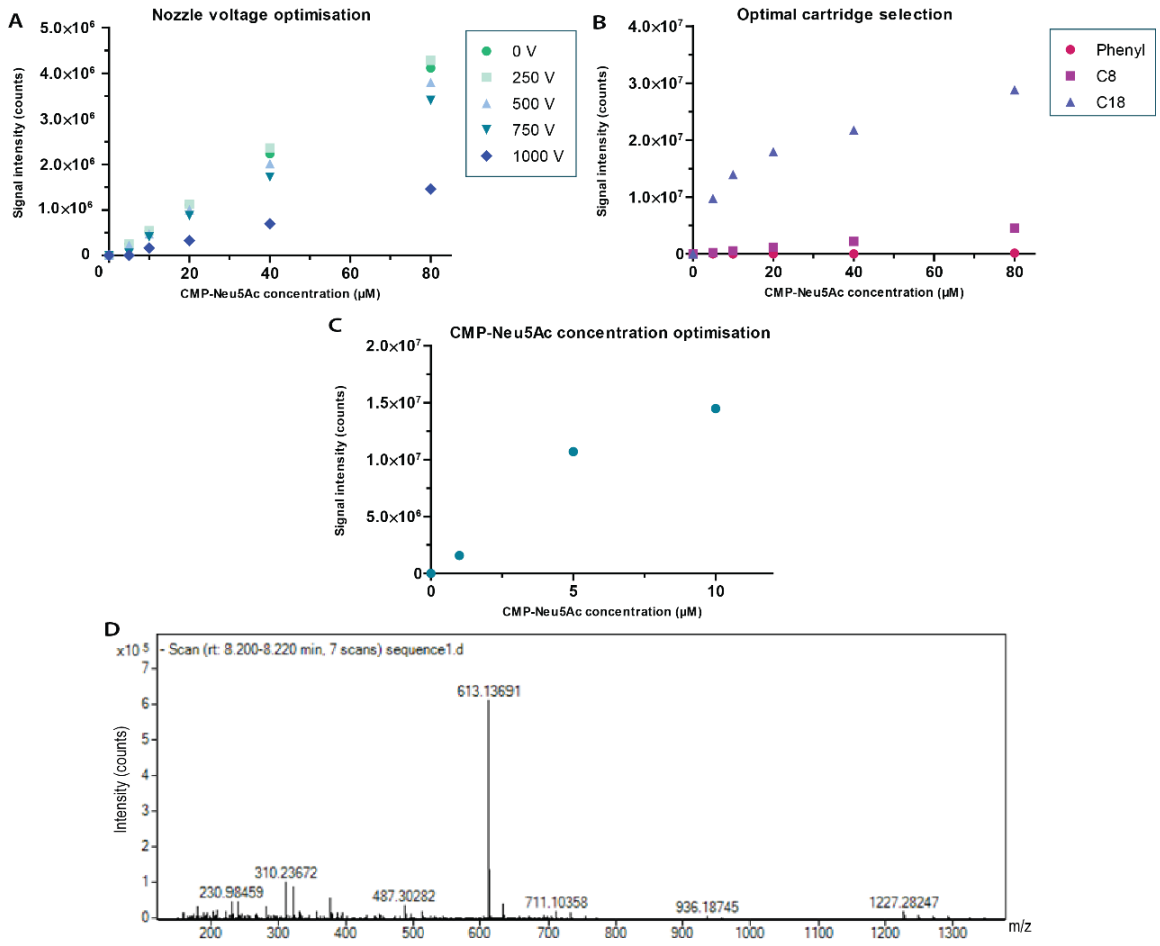


Figure 2.26: Optimisation of experimental conditions and CMP-Neu5Ac concentration for hydroxylation assays. (A) Varying CMP-Neu5Ac concentrations measured using different nozzle voltages. C8 cartridge was used for solid phase extraction. (B) Increasing CMP-Neu5Ac concentrations were run using phenyl, C8 and C18 cartridges. (C) Optimisation for optimal CMP-Neu5Ac concentration using C18 cartridge. (D) A mass spectrum of CMP-Neu5Ac (4 µM sample concentration) recorded using C18 cartridge for SPE, nozzle voltage was set to 0 V.

2.6.2 Reaction timecourse

mCMAH catalysed CMP-Neu5Ac hydroxylation reaction was followed by RapidFire-MS. Hydroxylation levels were measured every 120 s. In total ~20% conversion of CMP-Neu5Ac to CMP-Neu5Gc was observed after a reaction time of 25 min at 20 °C using freshly purified enzyme (as estimated by measuring formation of product assuming similar ionisation of both CMP-Neu5Ac and CMP-Neu5Gc, **Figure 2.28**). Freeze-thaw cycles appeared to adversely affect mCMAH activity, especially after prolonged storage times. To account for loss of activity, only freshly purified mCMAH was used for activity assays.

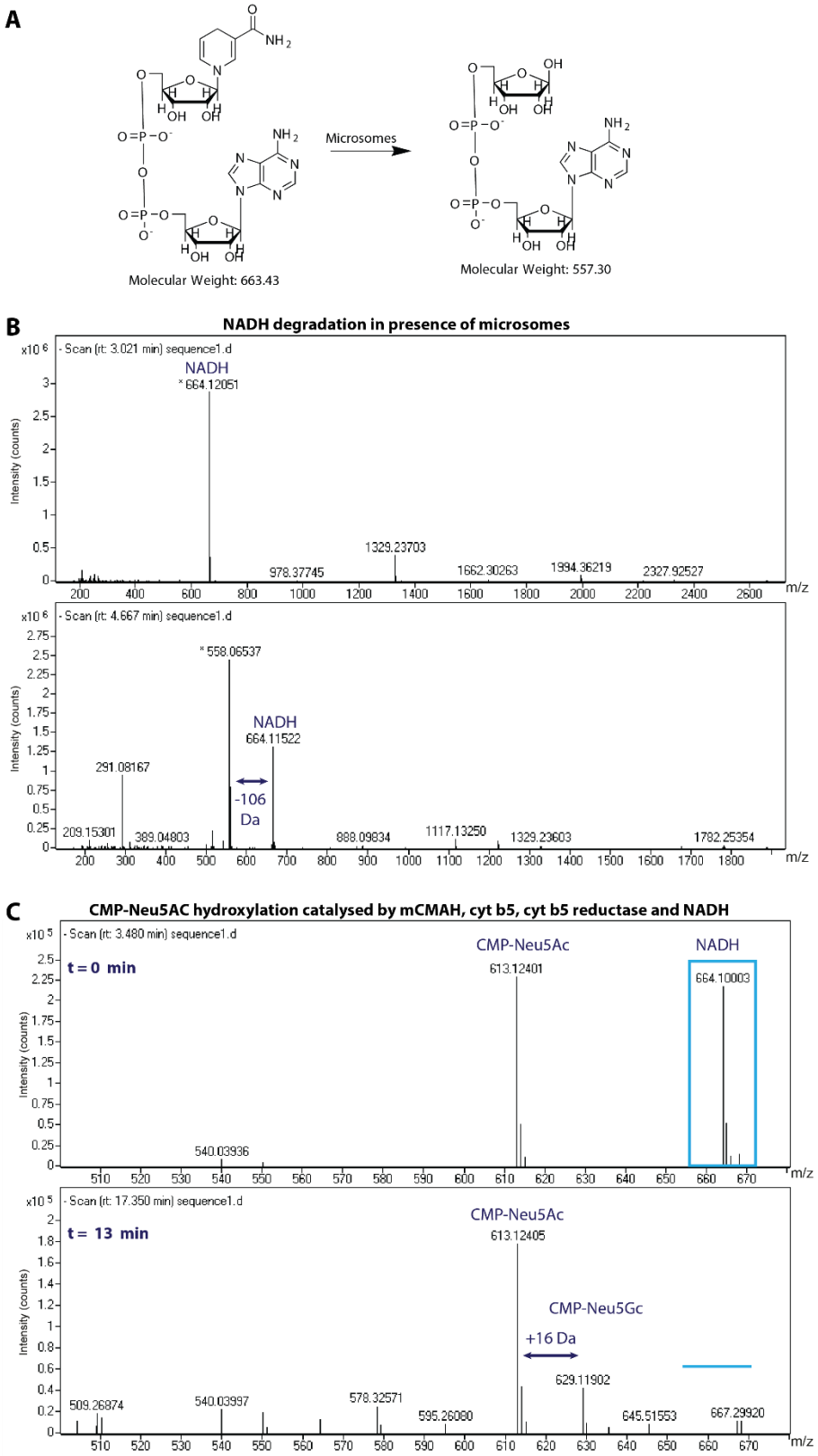


Figure 2.27: NADH degradation in the presence of co-factors observed during mCMAH hydroxylation assays. (A) Proposed reaction occurring in presence of mouse liver microsomes. (B) Mass spectra before and after 30 min incubation of NADH with microsomes. (C) NADH degradation observed in mCMAH assay conditions using purified recombinant cyt b_5 and cyt b_5 reductase upon addition of NADH (t=0 min) and after 13 min.

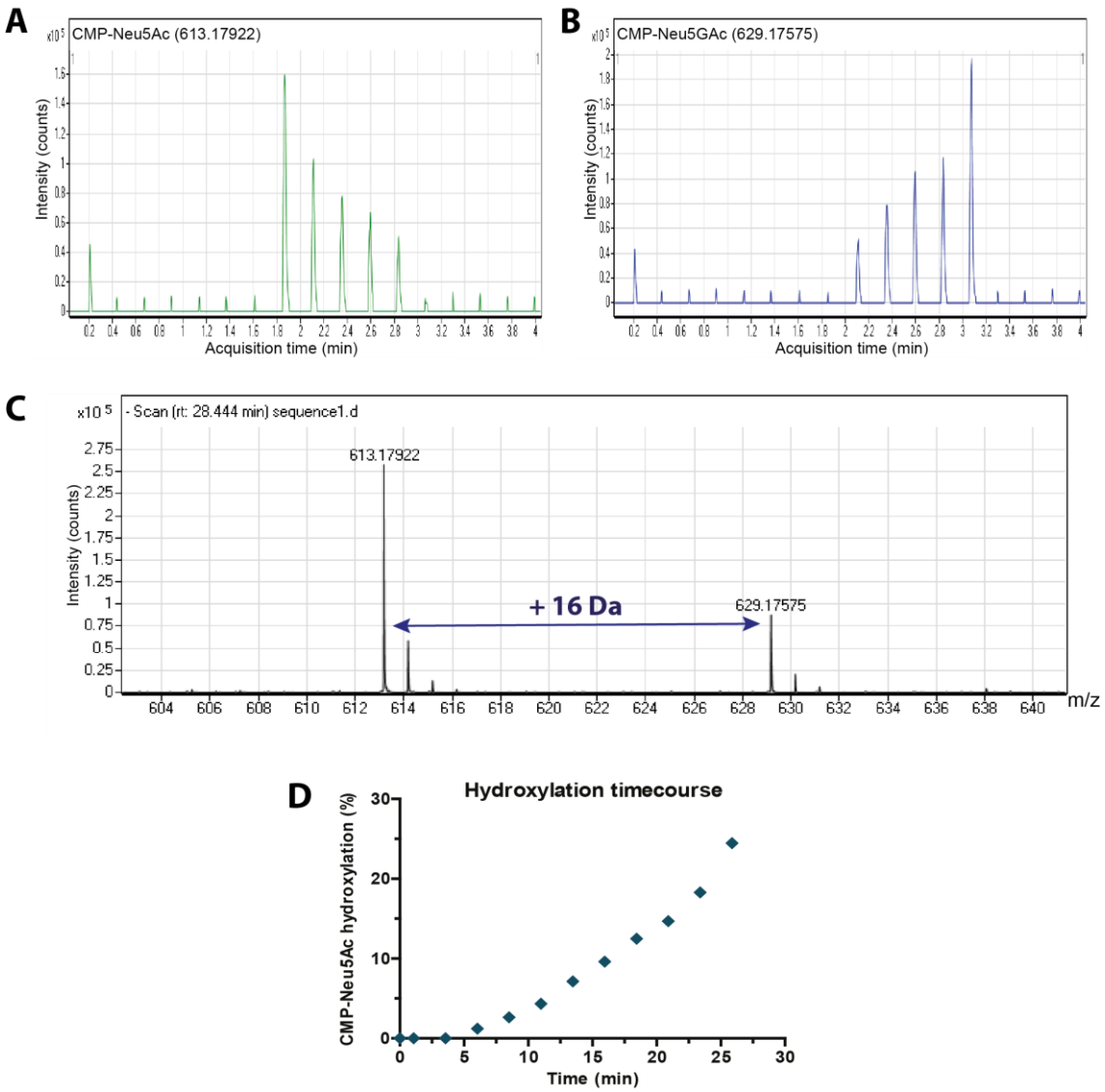


Figure 2.28: Conversion of CMP-Neu5Ac (613 Da) to CMP-Neu5Gc (629 Da) catalysed by mCMAH. (A) Depletion of CMP-Neu5Ac over time. (B) Formation of CMP-Neu5Gc over time. (C) Spectrum of CMP-Neu5Ac hydroxylation catalysed by mCMAH in the presence of cyt b_5, cyt b_5 reductase and NADH, 20 min after initiation of reaction showing CMP-Neu5Ac and CMP-Neu5Gc. (D) CMP-Neu5Ac hydroxylation timecourse as catalysed by mCMAH.

2.7 Crystallographic studies

Structural studies can play an important role in understanding function and mechanism of a protein. To date, there are no three-dimensional structures of CMAH available in the Protein Data Bank (PDB). Work towards hCMAH and mCMAH structures was therefore performed.

2.7.1 Thermal shift assays

Thermal shift assays (also known as Differential scanning fluorimetry or Thermofluor assay) were run to determine the best conditions for hCMAH (pCold construct produced in *E. coli*) and mCMAH (pcDNA3.1 construct produced in HEK293 cells) protein purification and

crystallisation. Purified proteins were diluted in Milli-Q water and mixed with SYPRO Orange dye (Thermo Fisher Scientific) and different buffers or salts and fluorescence measurements were recorded of the solution being heated to 95 °C. Fluorescence arises when SYPRO Orange dye becomes exposed to the hydrophobic environment, normally present at the core of folded soluble proteins, becoming accessible to the dye as the protein unfolds.[69] An example of a recorded melting curve and its negative derivative used for determination of the melting temperature (T_m) are shown in **Figure 2.29A** and **B** respectively.

In summary, mCMAH seemed to be most stable in sulfate containing buffers (HEPES, MOPS, ACES, PIPES, EPPS). While addition of 100 mM NaCl increased thermal stability from 57.6 °C (in 25 mM HEPES, pH 7.5) to 59.8 °C, addition of $FeCl_2$ or Neu5Ac did not have any effect (**Figure 2.29**).

Thermal shift assays for hCMAH were performed by Dr David Owen (Research Complex at Harwell, Didcot, UK). hCMAH was found to be thermally stable (T_m= 66.5 °C in Milli-Q water) with best conditions being sodium or potassium phosphate buffers (**Figure 2.30**). However, since phosphate anions tend to form insoluble salts with transition metals, use of phosphate buffers may interfere with crystallisation of transition metal containing proteins. From the commonly used buffers which do not interfere with crystallography, hCMAH was slightly (0.5 – 1 °C) more thermally stable in HEPES, pH 7.5 than Tris pH 8.0 when in presence of NaCl. Increasing NaCl concentration from 150 mM to 1.2 M resulted in an increase of T_m from 70.7 °C to 74.9 °C and 70.1 °C to 74.4 °C in HEPES and Tris respectively. However, similarly to the use of phosphate buffer, high salt concentrations can lead to salt crystal formation. For purification high salt concentration (500 mM NaCl) was used, but salt concentration was reduced to 50 mM in the buffer used for crystallisation experiments.

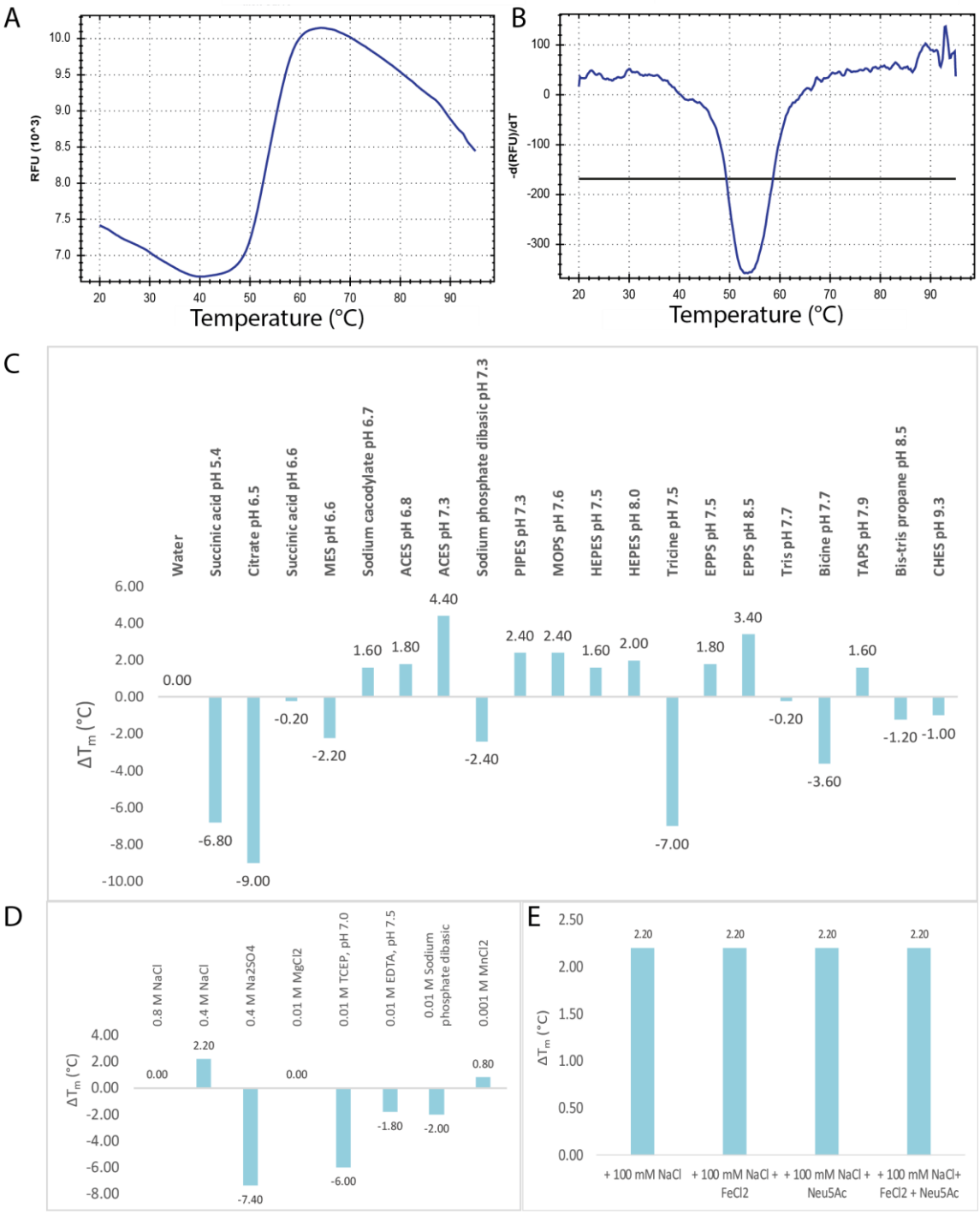


Figure 2.29: Changes in melting temperature (ΔT_m) of mCMAH as a result of different conditions. (A) A typical melting curve (mCMAH in Milli-Q water). (B) Negative first derivative of a melting curve shown in (A) used for easier determination of a T_m (negative peak). (C) Effect on T_m in different buffers as compared to water. All buffers assayed were at 0.2 M concentration. (D) Effect of select additives on thermal stability of mCMAH as compared to water. (E) Effect on mCMAH thermal stability of supplementation of 25 mM HEPES, pH 7.5 buffer with of sodium chloride, 0.1 mM $FeCl_2$, 0.1 mM Neu5Ac or 0.1 mM $FeCl_2$ and 0.1 mM Neu5Ac. Positive ΔT_m corresponds to stabilisation.

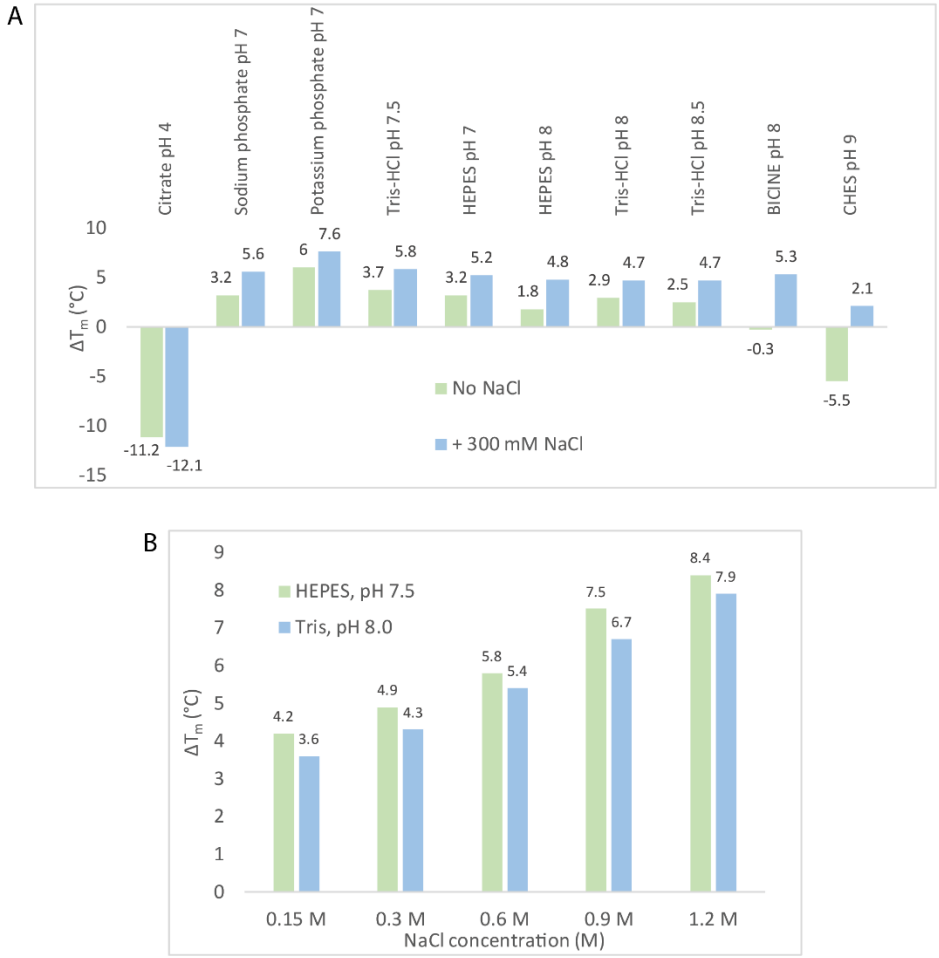


Figure 2.30: Changes in melting temperature (ΔT_m) of hCMAH in different conditions. (A) ΔT_m of hCMAH in common laboratory buffers without NaCl (green) and when supplemented with 300 mM NaCl (blue) as compared to Milli-Q water. All buffers tested were at 0.12 M. (B) ΔT_m of hCMAH as compared to Milli-Q water in 60 mM HEPES, pH 7.5 (green) and 60 mM Tris, pH 8.0 (blue) with increasing NaCl concentrations. Experiments were performed by Dr David Owen.

2.7.2 hCMAH crystallisation trials

Purified hCMAH (pCold construct, produced in *E. coli* Rosetta-gami B(DE3) cells) was subjected to crystallisation trials using broad screens (JCSG+, Morpheus, and ProPlex obtained from Molecular Dimensions; Index, Crystal Screen, PEG Ion, and SaltRx, from Hampton Research). Protein concentrations used ranged between 9 – 20 mg/ml. Small crystals of various morphologies (including microcrystals, not shown) were observed in several conditions (**Figure 2.31**). However, most of them did not diffract. A crystal of bipyramidal morphology diffracting to 3.3 Å resolution was obtained in a condition containing 100 mM MES, pH 6.5 and 1.6 M magnesium sulfate heptahydrate (**Figure 2.31B,** data collection statistics are listed in **Table 2.4**). The structure solution by molecular replacement using Phaser software[70] was attempted using several MBL fold proteins as well as models designed from closest hits detected by Phyre2 server[71]. After multiple unsuccessful attempts, the dataset was checked for

potential contamination using ContaMiner[72]. Comparing the data from presumed hCMAH crystal to 42 *E. coli* proteins using an optimised molecular replacement procedure based on MoRDa, ContaMiner analysis suggested the crystal to be *E. coli* dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex (OGDC-E2). OGDC-E2 has likely co-purified with the hCMAH (**Figure 2.31C**). There are currently three OGDC-E2 X-ray crystal structures available in the PDB (accession codes: 1E2O, 1SCZ, and 1C4T of 3 Å, 2.2 Å and 3 Å resolution respectively), hence the OGDC-E2 structure solution was not pursued further.

Table 2.4: Data collection statistics for the *E. coli* protein OGDC-E2 crystal obtained during hCMAH trials. *Values for highest resolution shell are given in parentheses.

	OGDC-E2
Source	Diamond Light Source, i03 beamline
Detector	Pilatus3 6M
Data processing	Xia2-DIALS
Resolution (Å)	38.97-3.31 (3.37 -3.31)*
Space group	*F*432
Unit cell parameters	
a, b, c (Å)	220.4, 220.4, 220.4
α, β, γ	90.0, 90.0, 90.0
Unique reflections	7296 (345)*
Multiplicity	70.5 (80.5)*
Completeness (%)	100 (100)*
I/σ(I)	10.2 (2.6)*
R_{merge}	0.719 (3.581)*
CC1/2	0.994 (0.751)*

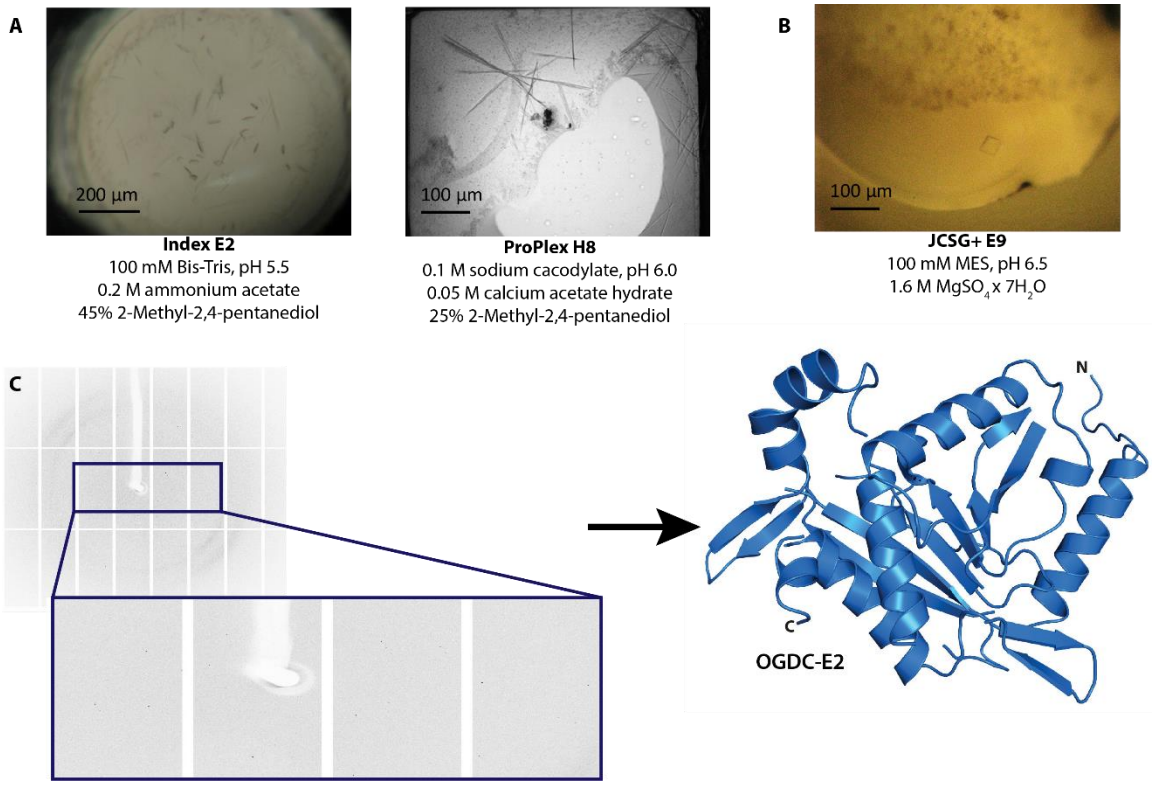


Figure 2.31: (A) Non-diffracting crystals obtained during hCMAH crystallisation trials. (B) A diffracting crystal of a contaminating *E. coli* protein OGDC-E2 obtained during hCMAH crystallisation trials. (C) Diffraction pattern of crystal shown in (B), and a structure of OGDC-E2 after 4 rounds of refinement.

2.7.3 mCMAH crystallisation trials

mCMAH crystallisation trials were performed with and without the addition of CMP-Neu5Ac using commercial broad screens (JCSG+, Morpheus, Morpheus II, Wizzard Classic 1&2, and PACT Premier obtained from Molecular Dimensions; Index, Crystal Screen, PEG Ion, SaltRx, and PEGRx from Hampton Research). Crystals were observed in several conditions (**Figure 2.32A**). A crystal diffracting to 2.3 Å resolution was obtained in conditions containing 2 M ammonium sulfate and 5% (v/v) 2-propanol (**Figure 2.32B**, data collection statistics are listed in **Table 2.5**). Phyre2[71] was used to detect proteins with highest similarity and search models for phasing by molecular replacement of the MBL and Rieske domains individually were built using Ensembler software available in the Phenix suite[73, 74]. Molecular replacement programs Phaser[70] (part of the Phenix software package[73]), MoRDa and MrBUMP[75] available as part of the CCP4 suite[76] were used. To exclude the possibility that the crystal might have originated from common contaminants that might have co-purified with mCHAH, contamination searches using ContaMiner[72] and SIMBAD[77] programs were performed. However, neither program detected any matches with common laboratory contaminants.

The dataset was re-processed using HKL-2000 (HKL Research) in P1 space group with resolution cut-off at 2.62 Å to improve data completeness. Attempts at structure solution using the re-processed dataset are ongoing.

Table 2.5: Data collection statistics for a crystal obtained from mCMAH crystallisation trials. *Values for highest resolution shell are given in parentheses. ǂPossible space group and unit cell parameters as determined by preliminary data processing and re-processing.

	Initial processing	**Reprocessed**
Source	Diamond Light Source, i04 beamline	Diamond Light Source, i04 beamline
Detector	Pilatus 6M-F	Pilatus 6M-F
Data processing	Fast DP (XDS)	HKL2000
Resolution (Å)	29.94 – 2.28 (2.32 – 2.28)*	49.63 – 2.62 (2.71– 2.62)
Space group	*P222*ǂ	*P1*ǂ
Unit cell parameters		
a, b, c (Å)	*51.6, 75.4, 166.8* ǂ	*51.9, 76.0, 168.1*ǂ
α, β, γ	*90.0, 90.0, 90.0* ǂ	*90.0 90.0 90.0*ǂ
Unique reflections	30335 (1187)*	73257 (7055)*
Multiplicity	12.4 (11.2)*	2.8 (2.5)*
Completeness (%)	98.99 (79.56)*	93.87 (86.2)
I/σ(I)	6.6 (2.3)*	4.4 (1.4)*
R_{merge}	0.435 (1.073)*	0.223 (0.554)*
CC1/2	0.972 (0.767)*	0.926 (0.505)*

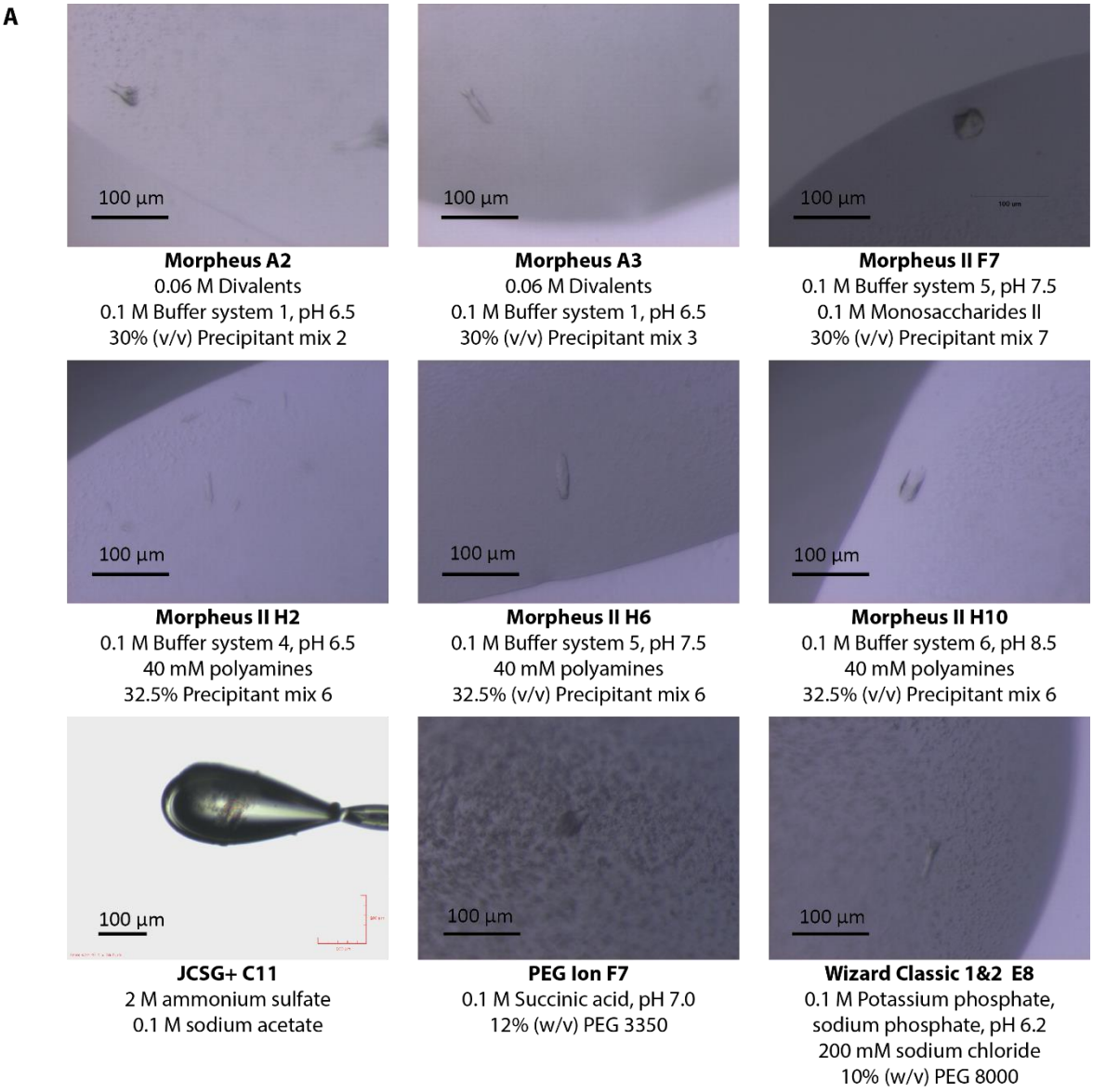


Figure 2.32: (A) Crystals obtained in various conditions during mCMAH crystallisation trials. PEG – polyethylene glycol. (B) A crystal obtained during mCMAH crystallisation trials and its diffraction pattern.

2.9 Summary and discussion

Several different expression systems were explored for production of recombinant human and mouse CMAH proteins. Although, hCMAH was successfully produced in both bacterial and mammalian expression systems, the viability of mammalian cells overexpressing recombinant hCMAH was a limiting factor for protein production. While hCMAH production in *E. coli* was more fruitful, the poor yields from its purification limited further characterisation of the protein.

Contrary to hCMAH, mCMAH expression in bacterial cells was very low with more protein being obtained from mammalian expression systems. The yields from mCMAH purification were still poor but better cell viability allowed for a more detailed characterisation. While being predicted to contain an *N*-glycosylation site[58], purified mCMAH was shown to not be glycosylated, at least in the case of recombinant mCMAH expressed in HEK293 and EpiCHO-S cells (**Figure 2.16**, **Figure 2.18**). However, a possibility of acylation, as observed by MS analysis (a +42 Da adduct), should be investigated further. The purified mCMAH protein was shown to contain a mononuclear high spin Fe(III), while the incorporation of a Rieske [2Fe2S] cluster into a recombinant mCMAH still remains inconclusive. UV-Vis analysis suggested the presence of a Rieske [2Fe2S] cluster with signal characteristic of the oxidised state (**Figure 2.22**). EPR analysis, on the other hand, has revealed a signal different from typical [2Fe2S] clusters as well as other types of iron-sulfur clusters reported in the literature (**Figure 2.24**). Despite the questionable presence of a [2Fe2S] cluster, recombinant mCMAH has been shown to possess catalytic activity in the presence of cyt b_5, cyt b_5 reductase and NADH (**Figure 2.28**). RapidFire-coupled MS assays were optimised and used instead of the radio-TLC assays reported in the literature[18], eliminating the need for isotopically-labelled Neu5Ac. Unfortunately, mCMAH enzymatic activity was shown to be significantly, or even completely, reduced following a freeze-thaw cycle.

Crystallisation attempts for recombinant hCMAH derived from *E. coli* expression did not yield a diffracting crystal of a target protein. The crystal obtained while performing hCMAH crystallisation trials was determined to be a contaminant *E. coli* OGDC-E2 protein. Crystallisation trials for mCMAH expressed in mammalian cells, yielded a single diffracting crystal. Despite considerable efforts, a structure for mCMAH has not been solved to date. Several non-diffracting crystals were also obtained for both proteins which could potentially be explored further and optimised in efforts to obtain a first CMAH crystal structure.

2.10 References

1. Angata, T.; Varki, A., Chemical Diversity in the Sialic Acids and Related α-Keto Acids: An Evolutionary Perspective. *Chem. Rev.* **2002,** *102*, 439-469.

2. Varki, A.; Schauer, R., Sialic Acids. In *Essentials of Glycobiology*, 2nd ed.; Varki, A.; Cummings, R. D.; Esko, J. D.; Freeze, H. H.; Stanley, P.; Bertozzi, C. R.; Hart, G. W.; Etzler, M. E., Eds. Cold Spring Harbor Laboratory Press: Cold Spring Harbor, 2009.

3. Schwarzkopf, M.; Knobeloch, K.-P.; Rohde, E.; Hinderlich, S.; Wiechens, N.; Lucka, L.; Horak, I.; Reutter, W.; Horstkorte, R., Sialylation is essential for early development in mice. *Proc. Natl. Acad. Sci. U. S. A.* **2002,** *99* (8), 5267-5270.

4. Ma, X.; Pan, Q.; Feng, Y.; Ma, Q.; Gagneux, P.; Ma, F., Sialylation Facilitates the Maturation of Mammalian Sperm and Affects Its Survival in Female Uterus. *Biol. Reprod.* **2016,** *94* (6), 123-123.

5. Troy, F. A., II, Polysialylation: from bacteria to brains. *Glycobiology* **1992,** *2* (1), 5-23.

6. Gandolfi, B.; Grahn, R. A.; Gustafson, N. A.; Proverbio, D.; Spada, E.; Adhikari, B.; Cheng, J.; Andrews, G.; Lyons, L. A.; Helps, C. R., A Novel Variant in *CMAH* Is Associated with Blood Type AB in Ragdoll Cats. *PLoS One* **2016,** *11* (5), e0154973.

7. Malykh, Y. N.; Krisch, B.; Gerardy-Schahn, R.; Lapina, E. B.; Shaw, L.; Schauer, R., The presence of *N*-acetylneuraminic acid in Malpighian tubules of larvae of the cicada *Philaenus spumarius*. *Glycoconj. J.* **2019,** *16* (11), 731-739.

8. Saito, M.; Kitamura, H.; Sugiyama, K., Occurrence of gangliosides in the common squid and pacific octopus among protostomia. *Biochim. Biophys. Acta* **2001,** *1511* (2), 271-280.

9. Alviano, C. S.; Travassos, L. R.; Schauer, R., Sialic acids in fungi: a minireview. *Glycoconj. J.* **1999,** *16* (9), 545-554.

10. Schauer, R.; Reuter, G.; Muhlpfordt, H.; Andrade, A. F.; Pereira, M. E., The occurrence of *N*-acetyl- and *N*-glycoloylneuraminic acid in *Trypanosoma cruzi*. *Hoppe-Seyler's Z. Physiol. Chem.* **1983,** *364* (8), 1053-1057.

11. Carrillo, M. B.; Gao, W.; Herrera, M.; Alroy, J.; Moore, J. B.; Beverley, S. M.; Pereira, M. A., Heterologous Expression of *Trypanosoma cruzi* trans-Sialidase in *Leishmania major* Enhances Virulence. *Infect. Immun.* **2000,** *68*, 2728-2734.

12. Byres, E.; Paton, A. W.; Paton, J. C.; Löfling, J. C.; Smith, D. F.; Wilce, M. C. J.; Talbot, U. M.; Chong, D. C.; Yu, H.; Huang, S.; Chen, X.; Varki, N. M.; Varki, A.; Rossjohn, J.; Beddoe, T., Incorporation of a non-human glycan mediates human susceptibility to a bacterial toxin. *Nature* **2008,** *456*, 648-652.

13. Suzuki, Y.; Ito, T.; Suzuki, T.; Holland, R. E.; Chambers, T. M.; Kiso, M.; Ishida, H.; Kawaoka, Y., Sialic acid species as a determinant of the host range of influenza A viruses. *J. Virol.* **2000,** *74* (24), 11825-11831.

14. Dankwa, S.; Lim, C.; Bei, A. K.; Jiang, R. H. Y.; Abshire, J. R.; Patel, S. D.; Goldberg, J. M.; Moreno, Y.; Kono, M.; Niles, J. C.; Duraisingh, M. T., Ancient human sialic acid variant restricts an emerging zoonotic malaria parasite. *Nat. Commun.* **2016,** *7*, 11187.

15. Iyer, J.; Grüner, A. C.; Rénia, L.; Snounou, G.; Preiser, P. R., Invasion of host cells by malaria parasites: A tale of two protein families. *Mol. Microbiol.* **2007,** *65* (2), 231-249.

16. Pettinati, I.; Brem, J.; Lee, S. Y.; McHugh, P. J.; Schofield, C. J., The Chemical Biology of Human Metallo-β-Lactamase Fold Proteins. *Trends Biochem. Sci.* **2016,** *41* (4), 338-355.

17. Schlenzka, W.; Shaw, L.; Kelm, S.; Schmidt, C. L.; Bill, E.; Trautwein, A. X.; Lottspeich, F.; Schauer, R., CMP-*N*-acetylneuraminic acid hydroxylase: The first cytosolic Rieske iron-sulphur protein to be described in Eukarya. *FEBS Lett.* **1996,** *385* (3), 197-200.

18. Shaw, L.; Schneckenburger, P.; Carlsen, J.; Christiansen, K.; Schauer, R., Mouse liver cytidine-5'-monophosphate-*N*-acetylneuraminic acid hydroxylase. Catalytic function and regulation. *Eur. J. Biochem.* **1992,** *206* (1), 269-277.

19. Takematsu, H.; Kawano, T.; Koyama, S.; Kozutsumi, Y.; Suzuki, A.; Kawasaki, T., Reaction mechanism underlying CMP-*N*-acetylneuraminic acid hydroxylation in mouse liver: formation of a ternary complex of cytochrome b_5, CMP-*N*-acetylneuraminic acid, and a hydroxylation enzyme. *J. Biochem.* **1994,** *115* (3), 381-386.

20. Shaw, L.; Schauer, R., Detection of CMP-*N*-acetylneuraminic acid hydroxylase activity in fractionated mouse liver. *Biochem. J* **1989,** *263* (2), 355-363.

21. Kozutsumi, Y.; Kawano, T.; Yamakawa, T.; Suzuki, A., Participation of cytochrome b_5 in CMP-*N*-acetylneuraminic acid hydroxylation in mouse liver cytosol. *J. Biochem.* **1990,** *108* (5), 704-706.

22. Lepers, A.; Shaw, L.; Schneckenburger, P.; Cacan, R.; Verbert, A.; Schauer, R., A study on the regulation of *N*-glycoloylneuraminic acid biosynthesis and utilization in rat and mouse liver. *Eur. J. Biochem.* **1990,** *193* (3), 715-723.

23. Kawano, T.; Kozutsumi, Y.; Takematsu, H.; Kawasaki, T.; Suzuki, a., Regulation of biosynthesis of *N*-glycolylneuraminic acid-containing glycoconjugates: characterization of factors required for NADH-dependent cytidine 5'monophosphate-*N*-acetylneuraminic acid hydroxylation. *Glycoconj. J.* **1993,** *10* (1), 109-115.

24. Schneckenburger, P.; Shaw, L.; Schauer, R., Purification, characterization and reconstitution of CMP-*N*-acetylneuraminate hydroxylase from mouse liver. *Glycoconj. J.* **1994,** *11* (3), 194-203.

25. Shaw, L.; Schneckenburger, P.; Schlenzka, W.; Carlsen, J.; Christiansen, K.; Jürgensen, D.; Schauer, R., CMP-*N*-acetylneuraminic acid hydroxylase from mouse liver and pig submandibular glands. Interaction with membrane-bound and soluble cytochrome b_5-dependent electron transport chains. *Eur. J. Biochem.* **1994,** *219* (3), 1001-1011.

26. Schlenzka, W.; Shaw, L.; Schneckenburger, P.; Schauer, R., Purification and characterization of CMP-*N*-acetylneuraminic acid hydroxylase from pig submandibular glands. *Glycobiology* **1994,** *4* (5), 675-683.

27. Schlenzka, W.; Shaw, L.; Schauer, R., Catalytic properties of the CMP-*N*-acetylneuraminic acid hydroxylase from the starfish *Asterias rubens*: comparison with the mammalian enzyme. *Biochim. Biophys. Acta* **1993,** *1161* (2-3), 131-138.

28. Gollub, M.; Schauer, R.; Shaw, L., Cytidine monophosphate-*N*-acetylneuraminate hydroxylase in the starfish *Asterias rubens* and other echinoderms. *Comp. Biochem. Physiol. B: Biochem. Mol. Biol.* **1998,** *120* (3), 605-615.

29. Takahashi, T.; Kawagishi, S.; Funahashi, H.; Hayashi, N.; Suzuki, T., Production and Purification of Secretory Simian Cytidine Monophosphate-*N*-acetylneuraminic Acid Hydroxylase Using Baculovirus-Protein Expression System. *Biol. Pharm. Bull.* **2015,** *38* (38), 1220-1226.

30. Chou, H.-H.; Takematsu, H.; Diaz, S.; Iber, J.; Nickerson, E.; Wright, K. L.; Muchmore, E. A.; Nelson, D. L.; Warren, S. T.; Varki, A., A mutation in human CMP-sialic acid hydroxylase occurred after the *Homo-Pan* divergence. *Proc. Natl. Acad. Sci. U. S. A.* **1998,** *95* (20), 11751-11756.

31. Irie, A.; Koyama, S.; Kozutsumi, Y.; Kawasaki, T.; Suzuki, A., The Molecular Basis for the Absence of *N*-Glycolylneuraminic Acid in Humans. *J. Biol. Chem.* **1998,** *273* (25), 15866-15871.

32. Hayakawa, T.; Satta, Y.; Gagneux, P.; Varki, A.; Takahata, N.; Harpending, H. C., *Alu*-mediated inactivation of the human CMP-*N*-acetylneuraminic acid hydroxylase gene. *Proc. Natl. Acad. Sci. U. S. A.* **2001,** *107* (50), 21300-21305.

33. Chou, H.-H.; Hayakawa, T.; Diaz, S.; Krings, M.; Indriati, E.; Leakey, M.; Paabo, S.; Satta, Y.; Takahata, N.; Varki, A., Inactivation of CMP-*N*-acetylneuraminic acid hydroxylase occurred prior to brain expansion during human evolution. *Proc. Natl. Acad. Sci. U. S. A.* **2002,** *99* (18), 11736-11741.

34. Hayakawa, T.; Aki, I.; Varki, A.; Satta, Y.; Takahata, N., Fixation of the human-specific CMP-*N*-acetylneuraminic acid hydroxylase pseudogene and implications of haplotype diversity for human evolution. *Genetics* **2006,** *172* (2), 1139-1146.

35. Nystedt, J.; Anderson, H.; Hirvonen, T.; Impola, U.; Jaatinen, T.; Heiskanen, A.; Blomqvist, M.; Satomaa, T.; Natunen, J.; Saarinen, J.; Lehenkari, P.; Valmu, L.; Laine, J., Human CMP-*N*-Acetylneuraminic Acid Hydroxylase is a Novel Stem Cell marker linked to Stem Cell-Specific Mechanisms. *Stem Cells* **2010,** *28* (2), 258-267.

36. Varki, A., Loss of *N*-glycolylneuraminic acid in humans: Mechanisms, consequences, and implications for hominid evolution. *Am. J. Phys. Anthropol.* **2001,** *Suppl 33*, 54-69.

37. Varki, A., *N*-glycolylneuraminic acid deficiency in humans. *Biochimie* **2001,** *83* (7), 615-622.

38. Dhar, C.; Sasmal, A.; Varki, A., From "Serum Sickness" to "Xenosialitis": Past, Present, and Future Significance of the Non-human Sialic Acid Neu5Gc. *Front. Immunol.* **2019,** *10*, 807.

39. Buchlis, G.; Odorizzi, P.; Soto, P. C.; Pearce, O. M. T.; Hui, D. J.; Jordan, M. S.; Varki, A.; Wherry, E. J.; High, K. a., Enhanced T cell function in a mouse model of human glycosylation. *J. Immunol.* **2013,** *191* (1), 228-237.

40. Okerblom, J. J.; Schwarz, F.; Olson, J.; Fletes, W.; Ali, S. R.; Martin, P. T.; Glass, C. K.; Nizet, V.; Varki, A., Loss of CMAH during Human Evolution Primed the Monocyte–Macrophage Lineage toward a More Inflammatory and Phagocytic State. *J. Immunol.* **2017**, 2366-2373.

41. Okerblom, J.; Varki, A., Biochemical, Cellular, Physiological and Pathological Consequences of Human loss of *N*-glycolylneuraminic Acid. *ChemBioChem* **2017,** *18* (13), 1151-1171.

42. Kwon, D. N.; Park, W. J.; Choi, Y. J.; Gurunathan, S.; Kim, J. H., Oxidative stress and ROS metabolism via down-regulation of sirtuin 3 expression in *Cmah*-null mice affect hearing loss. *Aging* **2015,** *7* (8), 579-594.

43. Kavaler, S.; Morinaga, H.; Jih, A.; Fan, W.; Hedlund, M.; Varki, A.; Kim, J. J., Pancreatic β-cell failure in obese mice with human-like CMP-Neu5Ac hydroxylase deficiency. *FASEB J.* **2011,** *25* (6), 1887-1893.

44. Takahashi, T.; Takano, M.; Kurebayashi, Y.; Masuda, M.; Kawagishi, S.; Takaguchi, M.; Yamanaka, T.; Minami, A.; Otsubo, T.; Ikeda, K.; Suzuki, T., *N*-glycolylneuraminic acid on human epithelial cells prevents entry of influenza A viruses that possess *N*-glycolylneuraminic acid binding ability. *J. Virol.* **2014,** *88* (15), 8445-8456.

45. Deng, L.; Song, J.; Gao, X.; Wang, J.; Yu, H.; Chen, X.; Varki, N.; Naito-Matsui, Y.; Galán, J. E.; Varki, A., Host adaptation of a bacterial toxin from the human pathogen *Salmonella typhi*. *Cell* **2014,** *159* (6), 1290-1299.

46. Martin, M. J.; Rayner, J. C.; Gagneux, P.; Barnwell, J. W.; Varki, A., Evolution of human-chimpanzee differences in malaria susceptibility: relationship to human genetic loss of *N*-glycolylneuraminic acid. *Proc. Natl. Acad. Sci. U. S. A.* **2005,** *102* (36), 12819-12824.

47. Springer, S. A.; Diaz, S. L.; Gagneux, P., Parallel evolution of a self-signal: humans and new world monkeys independently lost the cell surface sugar Neu5Gc. *Immunogenetics* **2014,** *66* (11), 671-674.

48. Ng, P. S. K.; Böhm, R.; Hartley-Tassell, L. E.; Steen, J. A.; Wang, H.; Lukowski, S. W.; Hawthorne, P. L.; Trezise, A. E. O.; Coloe, P. J.; Grimmond, S. M.; Haselhorst, T.; Von Itzstein, M.; Paton, A. W.; Paton, J. C.; Jennings, M. P., Ferrets exclusively synthesize Neu5Ac and express naturally humanized influenza A virus receptors. *Nat. Commun.* **2014,** *5*, 5750.

49. Hashimoto, Y.; Yamakawa, T.; Tanabe, Y., Further studies on the red cell glycolipids of various breeds of dogs. A possible assumption about the origin of Japanese dogs. *J. Biochem.* **1984,** *96* (6), 1777-1782.

50. Yasue, S.; Handa, S.; Miyagawa, S.; Jun, I.; Hasegawa, A.; Yamakawa, T., Difference in form of sialic acid in red blood cell glycolipids of different breeds of dogs. *J. Biochem.* **1978,** *83* (4), 1101-1107.

51. Bighignoli, B.; Niini, T.; Grahn, R. A.; Pedersen, N. C.; Millon, L. V.; Polli, M.; Longeri, M.; Lyons, L. A., Cytidine monophospho-*N*-acetylneuraminic acid hydroxylase (CMAH) mutations associated with the domestic cat AB blood group. *BMC Genet.* **2007,** *8* (1), 1-10.

52. Ito, T.; Suzuki, Y.; Suzuki, T.; Takada, A.; Horimoto, T.; Wells, K.; Kida, H.; Otsuki, K.; Kiso, M.; Ishida, H.; Kawaoka, Y., Recognition of *N*-glycolylneuraminic acid linked to galactose by the α2,3 linkage is associated with intestinal replication of influenza A virus in ducks. *J. Virol.* **2000,** *74* (19), 9300-9305.

53. Schauer, R.; Srinivasan, G. V.; Coddeville, B.; Zanetta, J. P.; Guérardel, Y., Low incidence of *N*-glycolylneuraminic acid in birds and reptiles and its absence in the platypus. *Carbohydr. Res.* **2009,** *344* (12), 1494-1500.

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

55. Jones, P.; Binns, D.; Chang, H.-Y.; Fraser, M.; Li, W.; McAnulla, C.; McWilliam, H.; Maslen, J.; Mitchell, A.; Nuka, G.; Pesseat, S.; Quinn, A. F.; Sangrador-Vegas, A.; Scheremetjew, M.; Yong, S.-Y.; Lopez, R.; Hunter, S., InterProScan 5: genome-scale protein function classification. *Bioinformatics* **2014,** *30* (9), 1236-1240.

56. Daiyasu, H.; Osaka, K.; Ishino, Y.; Toh, H., Expansion of the zinc metallo-hydrolase family of the β-lactamase fold. *FEBS Lett.* **2001,** *503* (1), 1-6.

57. Bleckwenn, N. A.; Shiloach, J., Large-scale cell culture. *Curr. Protoc. Immunol.* **2004,** *59*, A.1U.1-A.1U.44.

58. Gupta, R.; Brunak, S., Prediction of glycosylation across the human proteome and the correlation to protein function. *Pac. Symp. Biocomput.* **2002**, 310-322.

59. El-Battari, A.; Prorok, M.; Angata, K.; Mathieu, S.; Zerfaoui, M.; Ong, E.; Suzuki, M.; Lombardo, D.; Fukuda, M., Different glycosyltransferases are differentially processed for secretion, dimerization, and autoglycosylation. *Glycobiology* **2003,** *13* (12), 941-953.

60. *OriginPro*, 2017; OriginLab Corporation: Northampton, Massachusetts, USA.

61. Bou-Abdallah, F.; Chasteen, N. D., Spin concentration measurements of high-spin (g' = 4.3) rhombic iron(III) ions in biological samples: theory and application. *J. Biol. Inorg. Chem.* **2008,** *13* (1), 15-24.

62. Baratto, M. C.; Lipscomb, D. A.; Larkin, M. J.; Basosi, R.; Allen, C. C. R.; Pogni, R., Spectroscopic Characterisation of the Naphthalene Dioxygenase from *Rhodococcus* sp. Strain NCIMB12038. *Int. J. Mol. Sci.* **2019,** *20* (14), 3402-3411.

63. Holton, B.; Wu, X.; Alexandre I. Tsapin; Kramer, D. M.; Malkin, R.; Kallas, T., Reconstitution of the 2Fe-2S Center and g = 1.89 Electron Paramagnetic Resonance Signal into Overproduced *Nostoc* sp. PCC 7906 Rieske Protein. *Biochemistry* **1996,** *35* (48), 15485-15493.

64. Zhang, Y.; Yang, C.; Dancis, A.; Nakamaru-Ogiso, E., EPR studies of wild type and mutant Dre2 identify essential [2Fe–2S] and [4Fe–4S] clusters and their cysteine ligands. *J. Biochem.* **2017,** *161* (1), 67-78.

65. Tiago de Oliveira, F.; Bominaar, E. L.; Hirst, J.; Fee, J. A.; Münck, E., Antisymmetric Exchange in $[2Fe–2S]^{1+}$ Clusters: EPR of the Rieske Protein from *Thermus thermophilus* at pH 14. *J. Am. Chem. Soc.* **2004,** *126* (17), 5338-5339.

66. Wolfe, M. D.; Altier, D. J.; Stubna, A.; Popescu, C. V.; Münck, E.; Lipscomb, J. D., Benzoate 1,2-Dioxygenase from *Pseudomonas putida*: Single Turnover Kinetics and Regulation of a Two-Component Rieske Dioxygenase. *Biochemistry* **2002,** *41* (30), 9611-9626.

67. Kempa, E. E.; Hollywood, K. A.; Smith, C. A.; Barran, P. E., High throughput screening of complex biological samples with mass spectrometry - from bulk measurements to single cell analysis. *Analyst* **2019,** *144* (3), 872-891.

68. Rye, P. T.; LaMarr, W. A., Measurement of glycolysis reactants by high-throughput solid phase extraction with tandem mass spectrometry: Characterization of pyrophosphate-dependent phosphofructokinase as a case study. *Anal. Biochem.* **2015,** *482*, 40-47.

69. Boivin, S.; Kozak, S.; Meijers, R., Optimization of protein purification and characterization using Thermofluor screens. *Protein Expr. Purif.* **2013,** *91* (2), 192-206.

70. McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J., Phaser crystallographic software. *J. Appl. Crystallogr.* **2007,** *40* (Pt 4), 658-674.

71. Kelley, L. A.; Mezulis, S.; Yates, C. M.; Wass, M. N.; Sternberg, M. J. E., The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **2015,** *10*, 845.

72. Hungler, A.; Momin, A.; Diederichs, K.; Arold, S., T.,, ContaMiner and ContaBase: a webserver and database for early identification of unwantedly crystallized protein contaminants. *J. Appl. Crystallogr.* **2016,** *49* (6), 2252-2258.

73. Adams, P. D.; Afonine, P. V.; Bunkóczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L.-W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H., PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **2010,** *66* (Pt 2), 213-221.

74. Wang, X.; Snoeyink, J., Defining and computing optimum RMSD for gapped and weighted multiple-structure alignment. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2008,** *5* (4), 525-533.

75. Keegan, R. M.; Winn, M. D., Automated search-model discovery and preparation for structure solution by molecular replacement. *Acta Crystallogr. D Biol. Crystallogr.* **2007,** *63* (Pt 4), 447-457.

76. Winn, M. D.; Ballard, C. C.; Cowtan, K. D.; Dodson, E. J.; Emsley, P.; Evans, P. R.; Keegan, R. M.; Krissinel, E. B.; Leslie, A. G. W.; McCoy, A.; McNicholas, S. J.; Murshudov, G. N.; Pannu, N. S.; Potterton, E. A.; Powell, H. R.; Read, R. J.; Vagin, A.; Wilson, K. S., Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **2011,** *67* (4), 235-242.

77. Simpkin, A. J.; Simkovic, F.; Thomas, J. M. H.; Savko, M.; Lebedev, A.; Uski, V.; Ballard, C.; Wojdyr, M.; Wu, R.; Sanishvili, R.; Xu, Y.; Lisa, M.-N.; Buschiazzo, A.; Shepard, W.; Rigden, D. J.; Keegan, R. M., SIMBAD: a sequence-independent molecular-replacement pipeline. *Acta Crystallogr. D Biol. Crystallogr.* **2018,** *74* (7), 595-605.

Chapter 3: Investigating the Role of CMAH in Living Cells

Contents

3.1 Introduction....88
3.1.1 Expression of human CMAH gene....88
3.1.2 CMAH localisation and mechanism of action....89
3.1.3 Cytochrome b_5 and cytochrome b_5 reductase....89
3.2 Objectives....90
3.3 Immunofluorescence microscopy....90
3.3.1 Subcellular localisation of CMAH proteins....90
3.3.2 hCMAH does not localise to the endoplasmic reticulum or the mitochondria....91
3.3.3 Clustering of hCMAH in HeLa cells....93
3.4 Search for interacting proteins....93
3.4.1 Generating a stably mCMAH expressing TCMK-1 cell line....94
3.4.2 Pulldown experiments....95
3.4.3 Co-immunoprecipitation....96
3.5 Summary and discussion....97
3.6 References....99

The work described in this chapter was done in conjunction with Dr Ilaria Pettinati who advised on experimental design and sample preparation, and Dr Joan Gannon who assisted with microscopy and IF experimental design.

3.1 Introduction

3.1.1 Expression of human CMAH gene

As described in the introduction to Chapter 2, humans do not possess a catalytically active CMAH (cytidine monophosphate N-acetylneuraminic acid hydroxylase). It has been proposed that an *Alu*-mediated deletion event occurring approximately 3 million years ago has wiped out CMAH activity in the human population.[1-4] The deletion of 92 bp in exon 6 eliminated the region predicted to bind Rieske [2Fe2S] cluster in mouse and chimpanzee homologues.[3] Two conflicting theories of the resulting gene expression product exist. Irie et al. proposed that a catalytically inactive protein lacking the first 104 amino acids corresponding to the *N*-terminal Rieske domain is produced.[4] While Chou et al. suggested that the resulting gene product, if translated, would instead result in a 72 amino acid peptide corresponding to the *N*-terminal sequence of CMAH in mice and chimpanzees caused by early termination as a result of a frameshift mutation.[3]

Gene transcription is an energetically costly process, leading to investigations into other potential functions of human CMAH (hCMAH). Both pseudogenes transcribed to non-coding RNAs [5], and translated enzymatically inactive pseudoenzymes have been shown to possess various functions within the cells[6, 7]. The so far identified functions of non-coding RNAs include, among others, pro-inflammatory signalling[8], regulation of transcription and translation[9], and chromosomal organisation[10]. Pseudoenzymes have been proposed to play a variety of roles in cells, including but not limited to, facilitating protein complex formation[11] and regulating the activity of a catalytically active partner[12]. In addition, peptides have been shown to be produced by variety of organisms as toxins[13], antimicrobial defence[14] or signalling molecules[15]. Given the aforementioned examples, it is possible that, despite lacking enzymatic activity, human *CMAHP* gene still plays a role in other cellular functions.

CMAHP gene expression has been previously shown to be upregulated in human adult stem cells as compared to trace amounts detected in somatic HEK293 and HuH-7 cells.[16] In addition, up to a nine-fold difference in *CMAHP* expression was observed between mesenchymal and hematopoietic stem cells with the latter exhibiting higher expression levels. The major gene product observed in human umbilical cord blood (HUCB) cells was a 1972 bp messenger RNA

(mRNA) which, if translated into protein, would correspond to the CMAH sequence lacking the *N*-terminal Rieske domain. In the same study, overexpression of a stem cell variant of *CMAHP* was also proposed to be involved in the Wnt/β-catenin signalling pathway by potentially inducing nuclear translocation of synthesised β-catenin.[16]

3.1.2 CMAH localisation and mechanism of action

Previous studies of CMAH proteins from various organisms have shown that whilst echinoderms possess membrane bound CMAH enzymes[17], several catalytically active mammalian CMAH homologues have been shown to be a cytosolic[18, 19]. In addition, microscopy studies of a recombinant human stem cell green fluorescence protein (GFP)-CMAH fusion protein also seem to suggest cytosolic localisation.[16]

Enzymatic activity of CMAH has been demonstrated to require cytochrome b_5 (cyt b_5), cyt b_5 reductase and NADH.[20, 21] While the exact mode of interaction between the different components is unknown, the reaction scheme proposed by Takematsu et al. (**Figure 2.3**) suggests that binding of cyt b_5 occurs to the enzyme-substrate complex.[22] This is supported by their finding that CMAH elution from sepharose column with immobilised cyt b_5 is slower in presence of cytidine monophosphate *N*-acetylneuraminic acid (CMP-Neu5Ac) as opposed to reaction product CMP-N-glycolylneuraminic acid (CMP-Neu5Gc).[22] This then leaves two possibilities for cyt b_5 reduction by NADH dependent cyt b_5 reductase – cyt b_5 can either be reduced as a free protein and bind to the CMAH-CMP-Neu5Ac complex in its reduced form or be reduced as part of the complex.

3.1.3 Cytochrome b_5 and cytochrome b_5 reductase

Cyt b_5 is an electron transfer protein capable of accepting and transferring a single electron. Cyt b_5 was first spectroscopically observed by Keilin and Hartree in 1949 in variety of eukaryotic tissues at low temperatures.[23] It was later identified as a type b cytochrome and named cyt b_5 by Pappenheimer and Williams.[24] In mammals, two types of cyt b_5 exist – type A and type B.[25] Type A is found in soluble form, primarily in erythrocytes, and in a membrane bound form, bound to endoplasmic reticulum (ER). Type B cyt b_5 is encoded by a different gene than the type A and is found bound to mitochondrial membranes.[26] Cyt b_5 is widely conserved in nature with mammalian type A cyt b_5 proteins sharing over 80% sequence identity, mammals and birds sharing over 71% sequence identity, and mammalian and plant cyt b_5 enzymes being 31% identical.[27] In addition to being involved in CMP-Neu5Ac hydroxylation[28], cyt b_5 also plays a key role in many cellular processes requiring electron transfer, including,

among others, reduction of methaemoglobin to haemoglobin[29], fatty acid desaturation and elongation[30], and cholesterol biosynthesis[31].

Cyt b_5 is often found to be cycling between reducing and oxidising proteins.[32] Most commonly, cyt b_5 receives an electron from cyt b_5 reductase before passing it on to another protein. Cyt b_5 reductase is, similarly to cyt b_5, present in cells as soluble and membrane bound isoforms[32]. Cyt b_5 reductase contains FAD and NADH binding domains and transfers electrons from electron donor NADH via FAD to an electron receptor cyt b_5.[33] Cyt b_5 reductase can participate in one or two electron transfer reactions, meaning that two cyt b_5 molecules can be reduced using a single NADH molecule.[32]

3.2 Objectives

Since very little is known about the cellular function of hCMAH, the work described in this chapter aimed to investigate cellular localisation and binding partners of the CMAH proteins in order to gain some insights into potential role of hCMAH in cells as well as to improve understanding of the catalytically active mouse CMAH (mCMAH).

3.3 Immunofluorescence microscopy

The cellular localisation of CMAH was investigated in human HeLa and mouse TCMK-1 cells for hCMAH and mCMAH, respectively, employing immunofluorescence (IF) experiments. Cells were grown directly on coverslips, before transfecting with a relevant CMAH construct in pcDNA3.1 vector. 24 h post-transfection cells were fixed with paraformaldehyde to preserve the antigens[34], membranes were permeabilised using 0.1% Triton X-100, and samples were incubated with antibodies against CMAH and His_6-tag for detection of endogenous and recombinant CMAH proteins respectively. Experimental details are described in Chapter 6 (Materials and Methods).

3.3.1 Subcellular localisation of CMAH proteins

Wild type (WT) hCMAH was detected in low levels in the cytoplasm and appeared to be restricted to specific compartments (**Figure 3.1A**). Higher signal intensity was observed for the His_6-tagged hCMAH due to overexpression. The localisation of WT and recombinant hCMAH appeared comparable. Neither WT, nor recombinant hCMAH appeared to be localised in the nucleus. This observation was consistent with literature reports for various non-human CMAH proteins[17, 18].

Overexpressed mCMAH containing a *C*-terminal His_6-tag in TCMK-1 cells appeared to exhibit cytosolic localisation comparable to hCMAH proteins in HeLa cells with more distinct

compartmentalisation being observed (**Figure 3.1B**). WT mCMAH could not be detected in TCMK-1 due to experimental limitations as use of mouse-raised anti-CMAH antibody resulted in non-specific labelling of entire cells and no anti-CMAH antibodies raised in different hosts were commercially available at the time.

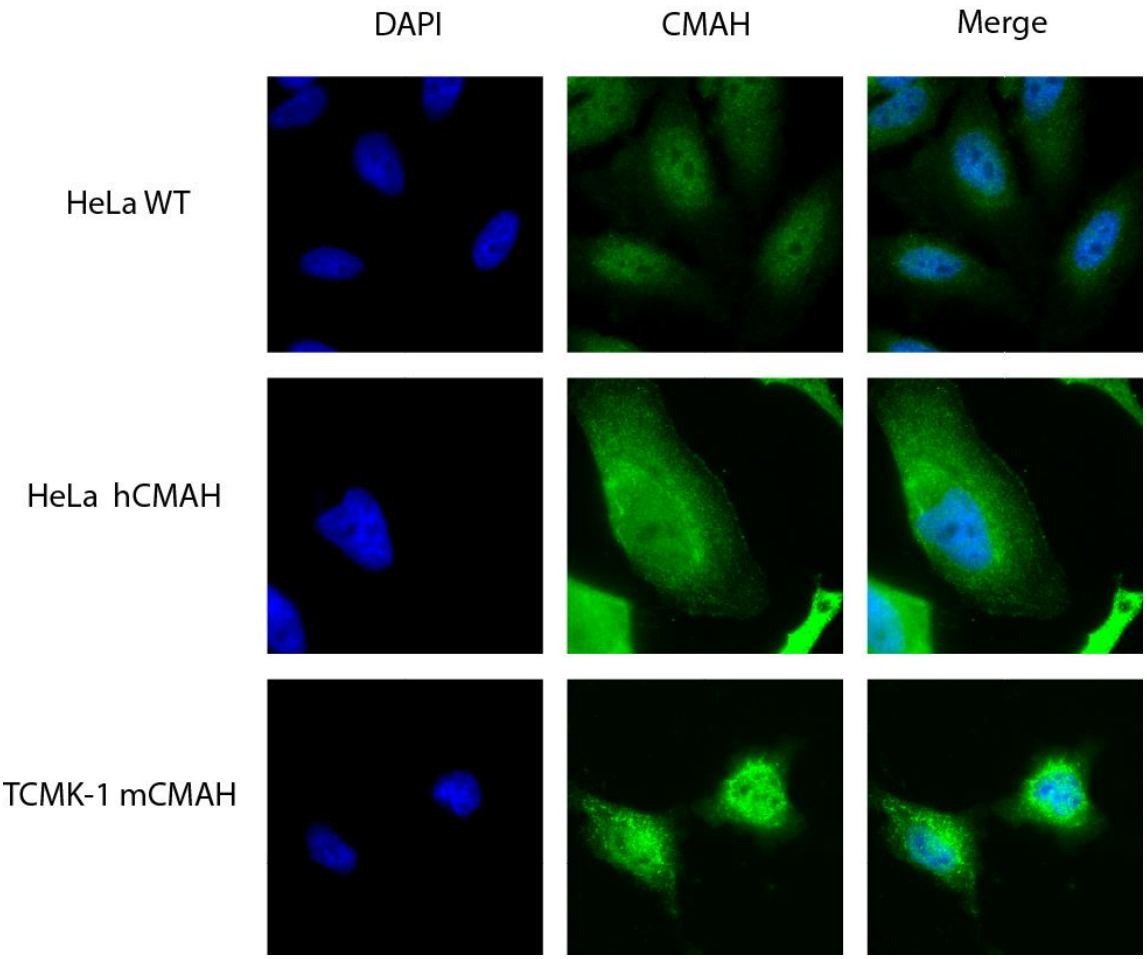


Figure 3.1: Deconvoluted immunofluorescence images showing similar localisation of hCMAH and mCMAH (green) in WT HeLa cells, HeLa cells transfected with hCMAH-*C*-His_6 protein, and TCMK-1 cells overexpressing *C*-terminally His_6 tagged mCMAH. Cell nuclei (blue) were stained with 4',6-diamidino-2-phenylindole (DAPI).

3.3.2 hCMAH does not localise to the endoplasmic reticulum or the mitochondria

In order to further investigate the potential compartmentalisation of hCMAH proteins, co-localisation IF experiments were performed.

As cyt b_5 needed for the CMAH reaction can be found bound to ER membranes[26] and catalytically active CMAH acts on a *N*-acetylneuraminic acid (Neu5Ac) which is further used for glycosylation, a process taking place in the ER[35], the possible localisation of CMAH at the ER was investigated using an ER-specific transmembrane chaperone calnexin[36] as a marker for the ER. Neither WT nor recombinant hCMAH seemed to localise to the ER (**Figure 3.2A**). Despite

some overlap of signals between recombinant hCMAH and calnexin being observed, analysis performed during data collection using Resolve3D SoftWoRx-Acquire software (Version 5.5.1, Release 3, Applied Precision Inc.) suggested no co-localisation of these proteins. Overlapping signal could be explained by high levels of protein being produced by the ribosomes located on the ER.

Type B of cyt b_5, a protein known to be required for a CMAH reaction[20], is localised at the mitochondrial membranes[26]. While the mode of interaction between CMAH and cyt b_5 is unknown and could have been lost with the loss of the Rieske domain, the potential localisation of hCMAH to the mitochondria was tested. MitoTracker Red CMXRos (Thermo Fisher Scientific) was used as a mitochondrial marker. No localisation of hCMAH to the mitochondria was observed (**Figure 3.2B**).

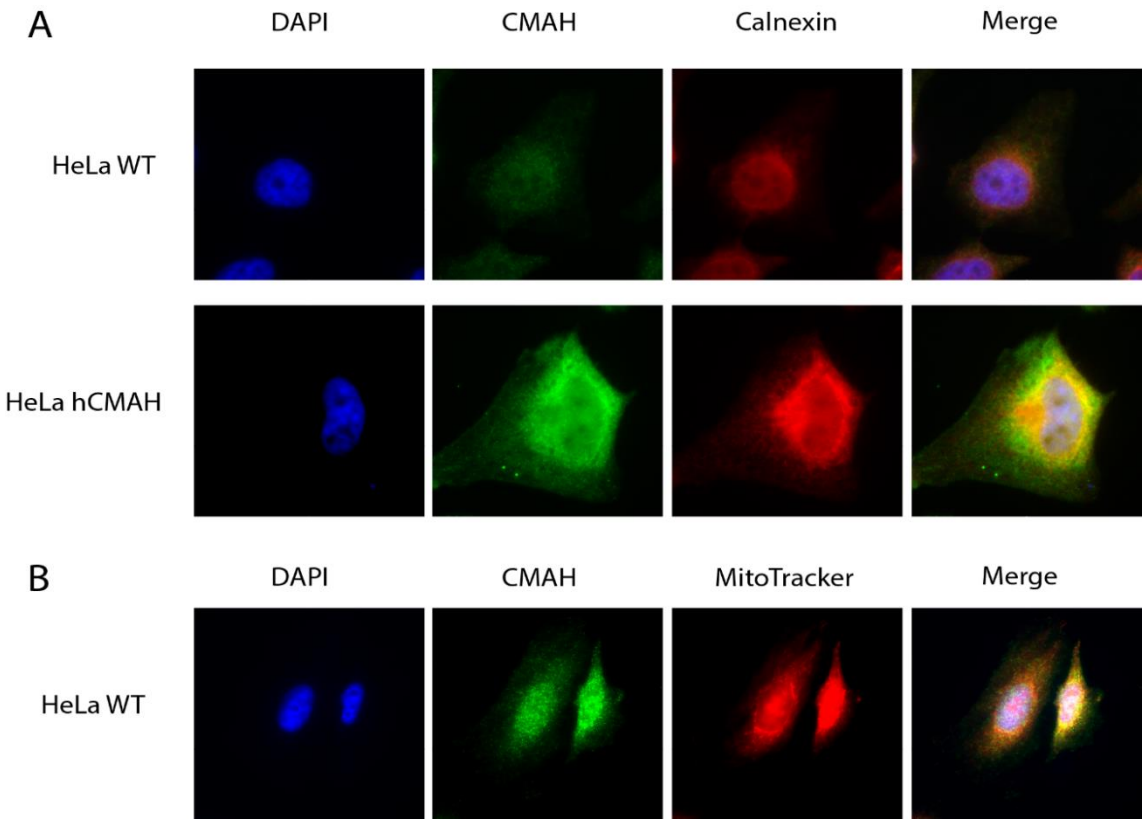


Figure 3.2: Deconvoluted immunofluorescence images for hCMAH co-localisation studies of Hela WT cells and cells overexpressing recombinant hCMAH (green) with (A) ER marker calnexin and (B) mitochondrial marker MitoTracker shown in red, showing no co-localisation. Cell nuclei (blue) were stained with DAPI.

Co-localisation experiments for mCMAH were not carried out due to a lack of possible WT control as mentioned above.

3.3.3 Clustering of hCMAH in HeLa cells

During the IF experiments when looking at CMAH or His_6-tag signal some large highly fluorescent dot-like formations were observed in both WT and hCMAH-transfected HeLa cells (**Figure 3.3**). These formations were observed in a small fraction of the cells analysed consistently throughout multiple experiments. The observed collection of hCMAH in specific compartments could potentially be a result of a particular cell cycle stage of the analysed cells or a result of protein aggregation into aggresomes caused by protein misfolding[37]. Further analysis is needed to determine at which point of the cell cycle the clustering occurs and/or whether it is an artefact.

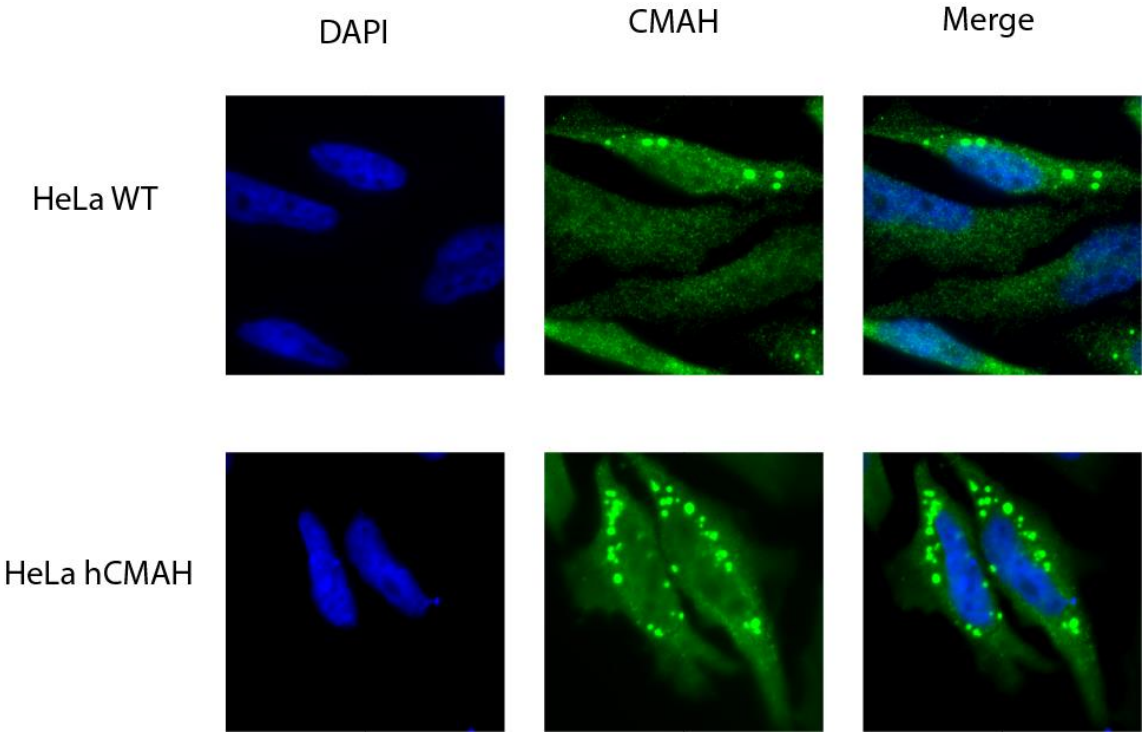


Figure 3.3: Immunofluorescence images showing dot-like formations of hCMAH (green) in WT HeLa cells as well as HeLa cells overexpressing recombinant hCMAH. Cell nuclei (blue) were stained with DAPI.

3.4 Search for interacting proteins

CMAH is reported to require cyt b_5, cyt b_5 reductase and NADH for the hydroxylase activity.[20, 21, 28] CMAH is thought to interact with cyt b_5 while cyt b_5 reductase and NADH are required for recycling of cyt b_5 (**Figure 2.3**).[22] Interacting proteins were attempted to study using pulldown and co-immunoprecipitation (co-IP) approaches. Due to low expression levels of endogenous hCMAH, HEK293 cells stably overexpressing hCMAH and mCMAH with *C*-terminal His_6 tag were used (generated as described in Chapter 2). Stably overexpressing mCMAH mouse kidney

epithelial cell line TCMK-1 was generated to study interactions of mCMAH in its native environment. For experimental procedures see Chapter 6.

3.4.1 Generating a stably mCMAH expressing TCMK-1 cell line

TCMK-1 cells were transfected with a pcDNA3.1 plasmid encoding mCMAH with a *C*-terminal His_6 tag (**Figure 3.4A**). As pcDNA3.1 plasmid encodes for neomycin resistance[38], single cell stable clones were selected by treating cells 48 h post transfection with G-418. mCMAH expression levels were assessed by Western blot analysis and single cell clones expressing high levels of recombinant protein were selected (**Figure 3.4**). Unlike in HEK293 cell lines, where mCMAH was observed primarily as a single band of approximately 70 kDa on Western blot, mCMAH in TCMK-1 was first observed as two bands around 70 kDa in size and two bands above 180 kDa (**Figure 3.4B**). This observation suggests differences in expression, processing or degradation in the HEK293 and TCMK-1 cell lines. None of the slower growing clones expressed *C*-His_6-tagged mCMAH as the same signal observed during Western blot analysis was also observed in WT non-transfected control (**Figure 3.4A&C**). The presence of signal in the WT control may be a result of non-specific binding of anti-His_6 antibodies or a detection of mouse antibodies produced in TCMK-1 cells.

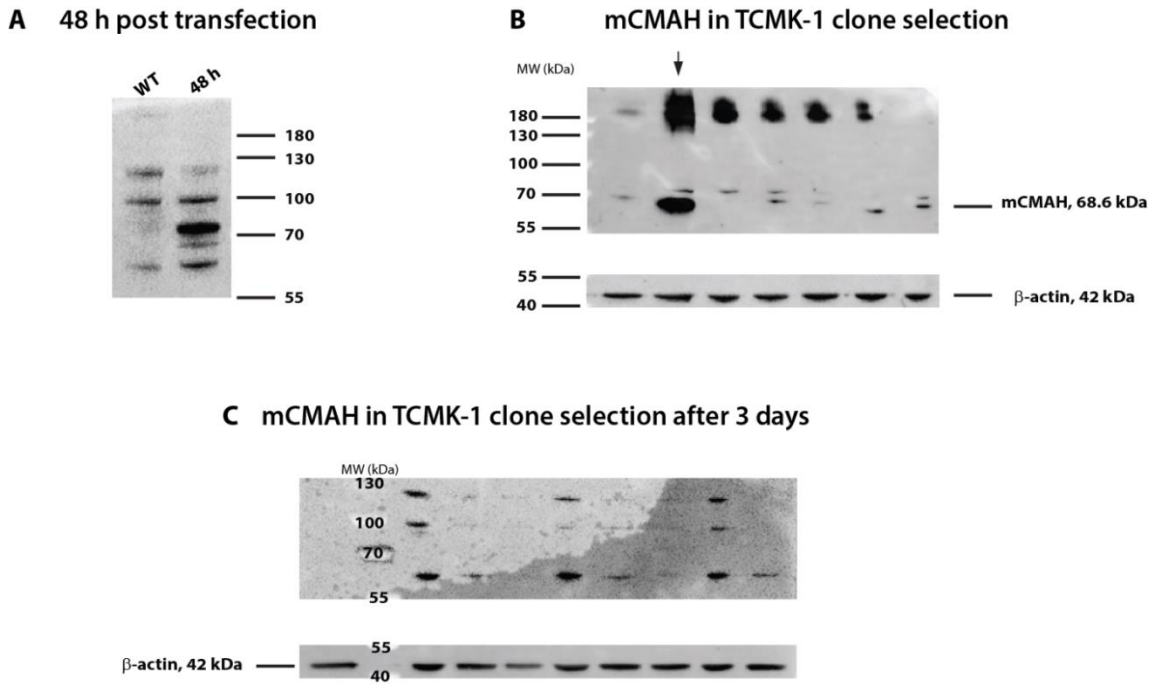


Figure 3.4: Single cell clone selection of TCMK-1 cells overexpressing recombinant mCMAH with *C*-terminal His_6-tag. (A) – Western blot analysis of TCMK-1 cells transfected with mCMAH 48 hours post transfection. (B) – Western blot analysis of the faster growing clones. (C) – Western blot analysis of the slower growing clones (grew to confluency 3 days after those shown in B). The clone selected for further studies is indicated by an arrow, β-actin was used as an internal loading control. WT – wild type. Anti-His_6 (raised in mice) and anti-β-actin antibodies were used.

3.4.2 Pulldown experiments

Pulldown experiments were focused on investigating CMAH-cyt b_5 interactions in HEK293 cells overexpressing recombinant hCMAH and mCMAH with *C*-terminal His_6 tags. Initial attempts to pulldown CMAH interactors using the Profinity IMAC resin, charged with Ni^{2+} ions (Bio-Rad) for His-tagged protein isolation proved unsuccessful (data not shown). This could be explained by CMAH-cyt b_5 interaction being short-lived and cyt b_5 only interacting with CMAH-CMP-Neu5AC complex, as proposed by Takematsu et al.[22] In order to capture short-lived interactions, cells were treated with 1% (v/v) aqueous formaldehyde to crosslink molecules interacting over short distances. His_6-tagged proteins were then pulled down and samples were incubated with 312.5 mM NaCl to decrosslink. Obtained samples were analysed using Western blot analysis for presence of both types A and B of cyt b_5. The results were inconclusive with Type A cyt b_5 being pulled down in all samples including the WT control suggesting a non-specific interaction between cyt b_5 and beads used for the experiments (**Figure 3.5**). No Type B cyt b_5 was detected in obtained samples.

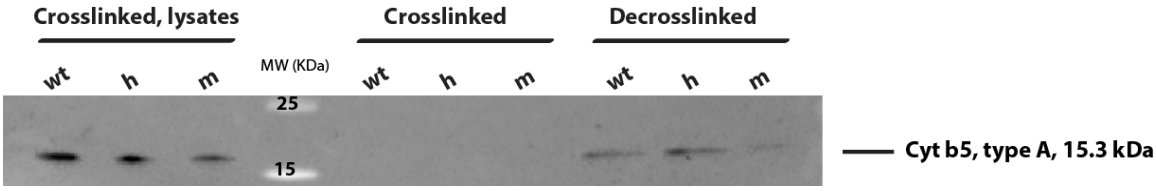


Figure 3.5: Western blot analysis of interactor studies of hCMAH (h) and mCMAH (m) recombinant proteins with type A cyt b_5. A band for cyt b_5 was observed in all samples including the WT control.

During the pulldown experiments, mCMAH was observed as two distinct bands in the crosslinked sample while only a single band, corresponding to the lower molecular weight band in the crosslinked sample, was observed after decrosslinking (**Figure 3.6**). The exact size difference between the two bands was difficult to determine due to the resolution limit of the SDS-PAGE, however, it appeared to be smaller than 30 kDa. Molecular weight of human cyt b_5 type A and type B are 15.3 kDa and 16.3 kDa respectively which could fit the observed difference. However, the said band was not observed when using antibodies against either type of cyt b_5. This leaves the results inconclusive, as the nature of the interaction could render cyt b_5 inaccessible for detection by antibodies by masking the epitope, or the observed band could be a result of a different interaction. Unfortunately, the amount of protein obtained was inadequate for identification.

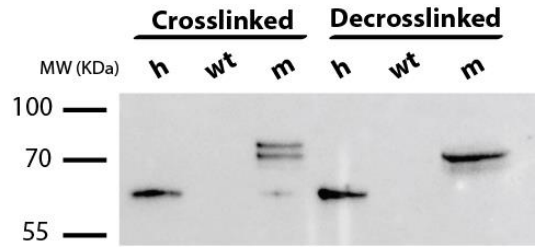


Figure 3.6: Western blot analysis of crosslinked and decrosslinked WT, and recombinant hCMAH (h) and mCMAH (m) overexpressing HEK293 cell samples obtained from pulldown experiments. Crosslinked mCMAH was observed as two bands while only a single band was detected after decrosslinking by incubating with 312.5 mM NaCl.

mCMAH interactions were also attempted to study in TCMK-1 cell line stably overexpressing recombinant mCMAH carrying a *C*-terminal His_6-tag. However, no mCMAH was detected after pulldown experiments. This could be a result of a loss of His_6-tag in cells or His_6-tag being prevented from binding Ni-charged resin.

3.4.3 Co-immunoprecipitation

Since the pulldown experiments were inconclusive, co-immunoprecipitation (co-IP) experiments were performed to investigate CMAH interactions. Dynabeads Protein G (Thermo Fisher Scientific) in conjunction with rabbit anti-His_6 antibody (Abcam) were used for the co-IP experiments.

Co-IP experiments were performed using HEK293 cells stably overexpressing recombinant hCMAH and mCMAH with *C*-terminal His_6-tag. Both hCMAH and mCMAH could be isolated in crosslinked and non-crosslinked samples, however, cyt b_5 of either type could not be detected in the eluted samples. The co-IP method was better suited for recovering hCMAH than mCMAH (**Figure 3.7**).

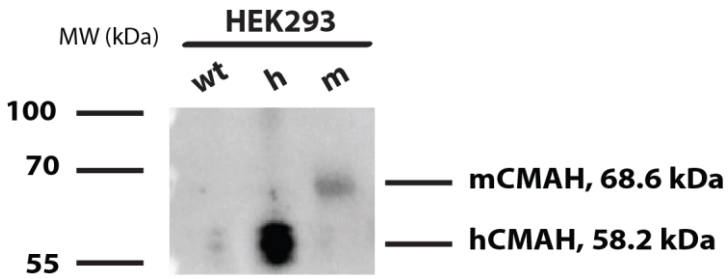


Figure 3.7: hCMAH and mCMAH proteins obtained after co-immunoprecipitation experiments using HEK293 cells, wt- wild type cells, h – cells overexpressing hCMAH, m – cells overexpressing mCMAH.

Co-immunoprecipitation experiments were also performed using TCMK-1 cells overexpressing mCMAH, but, similarly to pulldown experiments, no mCMAH was detected in the elution fractions further suggesting problems with His_6-tag accessibility.

3.5 Summary and discussion

Localisation of endogenous and recombinant hCMAH was studied in HeLa cells by immunofluorescence microscopy. While similar hCMAH cytosolic localisation was observed in WT and transfected cells, the signal observed in cells transfected with *C*-His_6-hCMAH was significantly higher. Potential compartmentalisation of hCMAH was further investigated by performing co-localisation experiments with ER chaperone calnexin and mitochondrial marker MitoTracker. hCMAH was found to not localise in the ER, or the mitochondria. Interestingly, clustering of both endogenous and recombinant hCMAH was observed in some HeLa cells throughout several different experiments, suggesting a potential localisation of hCMAH into a specific compartment at a certain stage of the cell cycle. Further work is needed in order to identify the cause of clustering and to determine exact localisation of hCMAH in HeLa cells.

As it is unclear whether hCMAH in human cells is produced as a protein lacking the Rieske domain present in other CMAH homologues[4], or a 72 amino acid peptide[3], it is also possible that the signal observed in WT HeLa cells is a result of the shorter peptide. Further work is needed to find out which of the two possibilities is actually present in WT cells.

Localisation of mCMAH was attempted to investigate in TCMK-1 cells. However, due to non-specific signal being observed in both cytoplasm and nucleus, WT TCMK-1 cells could not be used. Localisation of recombinant *C*-His_6-tagged mCMAH was found to be cytosolic with some potential compartmentalisation. However, due to a lack of a WT control, further studies were not performed.

Pulldown and co-IP experiments were performed in order to try to detect proteins interacting with hCMAH and mCMAH inside cells. HEK293 cells stably overexpressing hCMAH and mCMAH were used. Pulldown experiments failed to recover any cyt b_5, hence cells were treated with paraformaldehyde to crosslink short-distance interacting proteins. When using crosslinked samples, cyt b_5 was detected in all samples including WT control suggesting non-specific interactions with Ni-charged beads. Recombinant mCMAH was found to be pulled down with a potential interactor molecule when crosslinked samples were used, but the nature of the interactor could not be determined. Co-immunoprecipitation experiments were also attempted, but no cyt b_5 interactions were detected. Co-IP appeared to be better suited for hCMAH and not mCMAH studies based on yields of recovered proteins. As little is known about the nature of hCMAH protein, the experimental design was based on knowledge of mCMAH. However, hCMAH may play a different role in cells and interact with other proteins than what

is expected of mCMAH. Further experiments with a more open approach should be used in the future to find potential hCMAH interactors.

A mouse TCMK-1 cell line stably overexpressing recombinant mCMAH with a *C*-terminal His_6-tag was generated. Western blot analyses of cell lysates were performed using mouse-raised primary anti-His_6 antibodies followed by secondary antibodies recognising mouse IgG antibodies and resulted in multiple bands being observed in both WT and transfected TCMK-1 cell lysates. The presence of antibodies in TCMK-1 cells could be a possible explanation to both observation of multiple bands in Western blot analysis, and the non-specific staining observed in immunofluorescence experiments.

Studies of interacting proteins in TCMK-1 cells were found to be unsuccessful due to inability to detect any mCMAH after either pulldown or co-IP experiments. The reasons behind this are unclear. However, since His_6-tagged mCMAH could still be detected in cell lysates, the cause may have been the *C*-terminal His_6-tag being inaccessible to bind Ni-charged resin or anti-His_6 antibodies in solution.

3.6 References

1. Hayakawa, T.; Satta, Y.; Gagneux, P.; Varki, A.; Takahata, N.; Harpending, H. C., *Alu*-mediated inactivation of the human CMP-*N*-acetylneuraminic acid hydroxylase gene. *Proc. Natl. Acad. Sci. U. S. A.* **2001,** *107* (50), 21300-21305.

2. Hayakawa, T.; Aki, I.; Varki, A.; Satta, Y.; Takahata, N., Fixation of the human-specific CMP-*N*-acetylneuraminic acid hydroxylase pseudogene and implications of haplotype diversity for human evolution. *Genetics* **2006,** *172* (2), 1139-1146.

3. Chou, H.-H.; Takematsu, H.; Diaz, S.; Iber, J.; Nickerson, E.; Wright, K. L.; Muchmore, E. A.; Nelson, D. L.; Warren, S. T.; Varki, A., A mutation in human CMP-sialic acid hydroxylase occurred after the *Homo-Pan* divergence. *Proc. Natl. Acad. Sci. U. S. A.* **1998,** *95* (20), 11751-11756.

4. Irie, A.; Koyama, S.; Kozutsumi, Y.; Kawasaki, T.; Suzuki, A., The Molecular Basis for the Absence of *N*-Glycolylneuraminic Acid in Humans. *J. Biol. Chem.* **1998,** *273* (25), 15866-15871.

5. Grandér, D.; Johnsson, P., Pseudogene-Expressed RNAs: Emerging Roles in Gene Regulation and Disease. *Curr. Top. Microbiol. Immunol.* **2016**, 111-126.

6. Eyers, P. A.; Murphy, J. M., The evolving world of pseudoenzymes: proteins, prejudice and zombies. *BMC Biol.* **2016,** *14*, 1-6.

7. Ribeiro, A. J. M.; Das, S.; Dawson, N.; Zaru, R.; Orchard, S.; Thornton, J. M.; Orengo, C.; Zeqiraj, E.; Murphy, J. M.; Eyers, P. A., Emerging concepts in pseudoenzyme classification, evolution, and signaling. *Sci. Signal.* **2019,** *12* (594), eaat9797.

8. Rapicavoli, N. A.; Qu, K.; Zhang, J.; Mikhail, M.; Laberge, R. M.; Chang, H. Y., A mammalian pseudogene lncRNA at the interface of inflammation and anti-inflammatory therapeutics. *eLife* **2013,** *2*, e00762.

9. Kalyana-Sundaram, S.; Kumar-Sinha, C.; Shankar, S.; Robinson, D. R.; Wu, Y. M.; Cao, X.; Asangani, I. A.; Kothari, V.; Prensner, J. R.; Lonigro, R. J.; Iyer, M. K.; Barrette, T.; Shanmugam, A.; Dhanasekaran, S. M.; Palanisamy, N.; Chinnaiyan, A. M., Expressed pseudogenes in the transcriptional landscape of human cancers. *Cell* **2012,** *149* (7), 1622-1634.

10. Hacisuleyman, E.; Goff, L. A.; Trapnell, C.; Williams, A.; Henao-Mejia, J.; Sun, L.; McClanahan, P.; Hendrickson, D. G.; Sauvageau, M.; Kelley, D. R.; Morse, M.; Engreitz, J.; Lander, E. S.; Guttman, M.; Lodish, H. F.; Flavell, R.; Raj, A.; Rinn, J. L., Topological organization of multichromosomal regions by the long intergenic noncoding RNA Firre. *Nat. Struct. Mol. Biol.* **2014,** *21* (2), 198-206.

11. Walden, M.; Masandi, S. K.; Pawlowski, K.; Zeqiraj, E., Pseudo-DUBs as allosteric activators and molecular scaffolds of protein complexes. *Biochem. Soc. Trans.* **2018,** *46* (2), 453-466.

12. Zeqiraj, E.; Filippi, B. M.; Deak, M.; Alessi, D. R.; Aalten, D. M. F. v., Structure of the LKB1-STRAD-MO25 Complex Reveals an Allosteric Mechanism of Kinase Activation. *Science* **2009,** *326* (5960), 1707-1711.

13. Moyes, D. L.; Wilson, D.; Richardson, J. P.; Mogavero, S.; Tang, S. X.; Wernecke, J.; Höfs, S.; Gratacap, R. L.; Robbins, J.; Runglall, M.; Murciano, C.; Blagojevic, M.; Thavaraj, S.; Förster, T. M.; Hebecker, B.; Kasper, L.; Vizcay, G.; Iancu, S. I.; Kichik, N.; Häder, A.; Kurzai, O.; Luo, T.; Krüger, T.; Kniemeyer, O.; Cota, E.; Bader, O.; Wheeler, R. T.; Gutsmann, T.; Hube, B.; Naglik, J. R., Candidalysin is a fungal peptide toxin critical for mucosal infection. *Nature* **2016,** *532* (7597), 64-68.

14. Hancock, R. E., Cationic antimicrobial peptides: towards clinical applications. *Expert Opin. Investig. Drugs* **2000,** *9* (8), 1723-1729.

15. Lindsey, K.; Casson, S.; Chilley, P., Peptides: new signalling molecules in plants. *Trends Plant Sci.* **2002,** *7* (2), 78-83.

16. Nystedt, J.; Anderson, H.; Hirvonen, T.; Impola, U.; Jaatinen, T.; Heiskanen, A.; Blomqvist, M.; Satomaa, T.; Natunen, J.; Saarinen, J.; Lehenkari, P.; Valmu, L.; Laine, J., Human CMP-*N*-Acetylneuraminic Acid Hydroxylase is a Novel Stem Cell marker linked to Stem Cell-Specific Mechanisms. *Stem Cells* **2010,** *28* (2), 258-267.

17. Gollub, M.; Schauer, R.; Shaw, L., Cytidine monophosphate-*N*-acetylneuraminate hydroxylase in the starfish *Asterias rubens* and other echinoderms. *Comp. Biochem. Physiol. B: Biochem. Mol. Biol.* **1998,** *120* (3), 605-615.

18. Shaw, L.; Schneckenburger, P.; Schlenzka, W.; Carlsen, J.; Christiansen, K.; Jürgensen, D.; Schauer, R., CMP-*N*-acetylneuraminic acid hydroxylase from mouse liver and pig submandibular glands. Interaction with membrane-bound and soluble cytochrome b_5-dependent electron transport chains. *Eur. J. Biochem.* **1994,** *219* (3), 1001-1011.

19. Schlenzka, W.; Shaw, L.; Kelm, S.; Schmidt, C. L.; Bill, E.; Trautwein, A. X.; Lottspeich, F.; Schauer, R., CMP-*N*-acetylneuraminic acid hydroxylase: The first cytosolic Rieske iron-sulphur protein to be described in Eukarya. *FEBS Lett.* **1996,** *385* (3), 197-200.

20. Kozutsumi, Y.; Kawano, T.; Yamakawa, T.; Suzuki, A., Participation of cytochrome b_5 in CMP-*N*-acetylneuraminic acid hydroxylation in mouse liver cytosol. *J. Biochem.* **1990,** *108* (5), 704-706.

21. Shaw, L.; Schneckenburger, P.; Carlsen, J.; Christiansen, K.; Schauer, R., Mouse liver cytidine-5'-monophosphate-*N*-acetylneuraminic acid hydroxylase. Catalytic function and regulation. *Eur. J. Biochem.* **1992,** *206* (1), 269-277.

22. Takematsu, H.; Kawano, T.; Koyama, S.; Kozutsumi, Y.; Suzuki, A.; Kawasaki, T., Reaction mechanism underlying CMP-*N*-acetylneuraminic acid hydroxylation in mouse liver: formation of a ternary complex of cytochrome b_5, CMP-*N*-acetylneuraminic acid, and a hydroxylation enzyme. *J. Biochem.* **1994,** *115* (3), 381-386.

23. Keilin, D.; Hartree, E. F., Effect of Low Temperature on the Absorption Spectra of Haemoproteins; with Observations on the Absorption Spectrum of Oxygen. *Nature* **1949,** *164*, 254-259.

24. Pappenheimer, A. M., Jr.; Williams, C. M., Cytochrome b_5 and the dihydro-coenzyme I-oxidase system in the cecropia silkworm. *J. Biol. Chem.* **1954,** *209* (2), 915-929.

25. Consortium, T. U., UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* **2018,** *47* (D1), D506-D515.

26. Vergeres, G.; Waskell, L., Cytochrome b_5, its functions, structure and membrane topology. *Biochimie* **1995,** *77* (7-8), 604-620.

27. Schenkman, J. B.; Jansson, I.; Davis, G.; Tamburini, P. P.; Lu, Z.; Zhang, Z.; Rusling, J. F., Protein-Protein Interactions in the P450 Monooxygenase System. In *Molecular and Applied Aspects of Oxidative Drug Metabolizing Enzymes*, Arinç, E.; Schenkman, J. B.; Hodgson, E., Eds. Springer US: Boston, MA, 1999; pp 21-39.

28. Kawano, T.; Kozutsumi, Y.; Takematsu, H.; Kawasaki, T.; Suzuki, a., Regulation of biosynthesis of *N*-glycolylneuraminic acid-containing glycoconjugates: characterization of factors required for NADH-dependent cytidine 5'monophosphate-*N*-acetylneuraminic acid hydroxylation. *Glycoconj. J.* **1993,** *10* (1), 109-115.

29. Sannes, L. J.; Hultquist, D. E., Effects of hemolysate concentration, ionic strength and cytochrome b_5 concentration on the rate of methemoglobin reduction in hemolysates of human erythrocytes. *Biochim. Biophys. Acta* **1978,** *544* (3), 547-554.

30. Okayasu, T.; Nagao, M.; Ishibashi, T.; Imai, Y., Purification and partial characterization of linoleoyl-CoA desaturase from rat liver microsomes. *Arch. Biochem. Biophys.* **1981,** *206* (1), 21-28.

31. Fukushima, H.; Grinstead, G. F.; Gaylor, J. L., Total enzymic synthesis of cholesterol from lanosterol. Cytochrome b_5-dependence of 4-methyl sterol oxidase. *J. Biol. Chem.* **1981,** *256* (10), 4822-4826.

32. Elahian, F.; Sepehrizadeh, Z.; Moghimi, B.; Mirzaei, S. A., Human cytochrome b_5 reductase: structure, function, and potential applications. *Crit. Rev. Biotechnol.* **2014,** *34* (2), 134-143.

33. Yamada, M.; Tamada, T.; Takeda, K.; Matsumoto, F.; Ohno, H.; Kosugi, M.; Takaba, K.; Shoyama, Y.; Kimura, S.; Kuroki, R.; Miki, K., Elucidations of the catalytic cycle of NADH-cytochrome b_5 reductase by X-ray crystallography: new insights into regulation of efficient electron transfer. *Journald of Molecular Biology* **2013,** *425* (22), 4295-4306.

34. Smit, J. W.; Meijer, C. J.; Decary, F.; Feltkamp-Vroom, T. M., Paraformaldehyde fixation in immunofluorescence and immunoelectron microscopy. Preservation of tissue and cell surface membrane antigens. *J. Immunol. Methods* **1974,** *6* (1-2), 93-98.

35. Aebi, M., *N*-linked protein glycosylation in the ER. *Biochim. Biophys. Acta* **2013,** *1833* (11), 2430-2437.

36. Müller-Taubenberger, A.; Lupas, A. N.; Li, H.; Ecke, M.; Simmeth, E.; Gerisch, G., Calreticulin and calnexin in the endoplasmic reticulum are important for phagocytosis. *EMBO J.* **2001,** *20* (23), 6772-6782.

37. Johnston, J. A.; Ward, C. L.; Kopito, R. R., Aggresomes: A Cellular Response to Misfolded Proteins. *J. Cell Biol.* **1998,** *143* (7), 1883-1898.

38. pcDNA™3.1 (+) and pcDNA™3.1 (-) User Manual. Thermo Fisher Scientific: 2010; Vol. Version K.

Chapter 4: Biochemical and Structural Characterisation of *Escherichia coli* Glyoxalase II

Contents

4.1 Introduction 104
4.1.1. Formation and toxicity of methylglyoxal 104
4.1.2 Glyoxalase system 106
4.1.3 Glyoxalases as potential drug targets 108
4.2 Objectives 109
4.3 Protein production and purification 109
4.4 Characterisation of GloB 111
4.4.1 Metal content analysis 111
4.4.2 Non-denaturing MS 111
4.4.3 Circular dichroism 113
4.5 Kinetic and inhibition studies 115
4.5.1 Kinetics and metal substitution 115
4.5.2 Search for GloB inhibitors 117
4.6 Crystallography 121
4.6.1 Buffer optimisation 121
4.6.2 Crystallisation trials 122
4.6.3 Search for an alternative crystal form 127
4.6.4 GloB structure within the Glx2 family 134
4.7 Summary and conclusions 135
4.8 References 138

Some of the work in this chapter was performed in conjunction with other researchers. Non-denaturing MS spectra were recorded by Shuang Liu. Dr Karina Calvopiña advised on kinetic and inhibition assays. The inhibitors used for inhibition studies were synthesised by current and previous members of the group. Dr Michael McDonough and Dr Jürgen Brem assisted with X-ray crystallography data collection and processing. The GloB expression plasmid and an initial purification protocol was provided by Dr Madeline Kavanagh.

4.1 Introduction

4.1.1. Formation and toxicity of methylglyoxal

All cells, from smallest bacteria to the largest eukaryotes, are sites of constant biochemical activity. These dynamic processes range from breaking down nutrients to produce energy, to protein synthesis, and the maintenance of homeostasis. Some products and side products of these cellular processes can be toxic. Methylglyoxal (MG) is a reactive cytotoxic α-ketoaldehyde that is primarily produced as a side product of glycolysis and lipid metabolism.[1]

Non-enzymatic formation of MG from decomposition of triose phosphates such as dihydroxyacetone phosphate (DHAP) and glyceraldehyde-3-phosphate (GAP) has been known since the 1930s, but was considered to have no physiological significance at the time.[2] At physiological pH, DHAP and GAP can undergo deprotonation to form an enediolate phosphate, which, following the loss of a phosphate group, results in MG formation (**Figure 4.1**).[3, 4] In both eukaryotic and prokaryotic cells, enediolate phosphate occurs in triose phosphate isomerase (TIM) catalysis as an intermediate between DHAP and GAP. Occasionally, the enolate intermediate can leak out of the active site of TIM and subsequently form MG in the cytoplasm (**Figure 4.1**).[5]

MG can also be formed enzymatically by MG synthase. Enzymatic formation of MG has been mostly examined in bacteria.[6-9] MG synthase activity has been reported in goat liver[10], but further studies failed to detect the presence of a specific enzyme catalysing MG formation in mammalian cells[11, 12]. MG synthase catalysed formation of MG from DHAP has been proposed to be a part of a bacterial glycolytic bypass that converts DHAP to pyruvate.[6, 8]

Other MG formation routes include oxidation of acetone by cytochrome P450[13, 14], oxidation of aminoacetone produced in the L-threonine catabolism by semicarbazide-sensitive amine oxidase[15], oxidation of ketone bodies[16], the Maillard[17] and lipoperoxidation reactions[18].

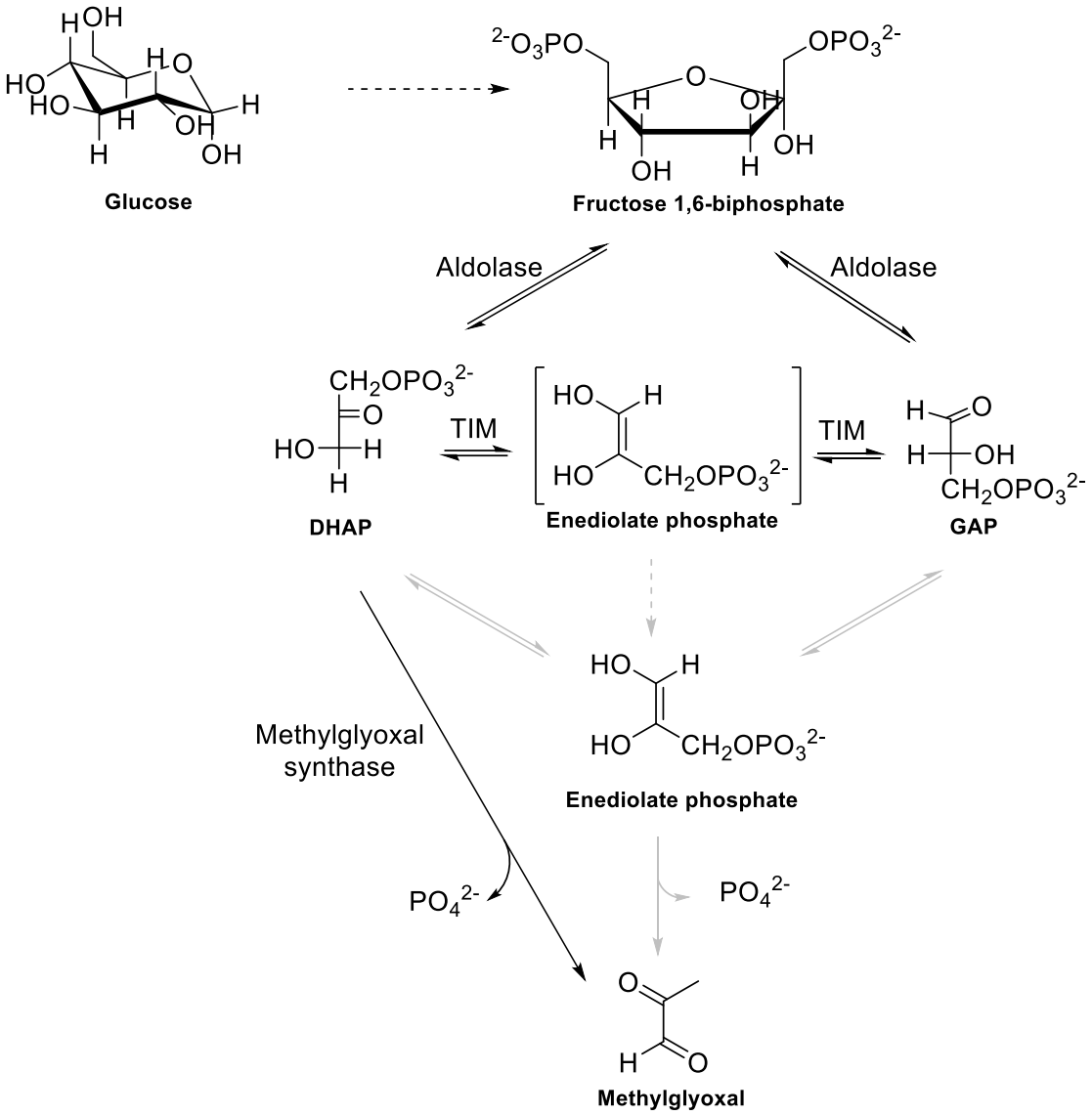


Figure 4.1: Major pathways of methylglyoxal (MG) formation from glucose. Non-enzymatic reactions are denoted by grey arrows. *Methylglyoxal synthase has only been detected in bacteria. TIM – triose phosphate isomerase, DHAP – dihydroxyacetone phosphate, GAP – glyceraldehyde-3-phosphate.

MG cytotoxicity is proposed to arise from its high reactivity as a glycating agent.[19] Glycation is a non-enzymatic process occurring when electrophilic aldehydes or ketones derived from sugars react with nucleophilic (usually amino) groups present in proteins, nucleic acids, and lipids. The initial glycation reactions are usually reversible. However, further oxidation of glycation products can lead to a formation of stable advanced glycation end products (AGEs).[19] Compared to glucose, MG and other α-oxoaldehydes are significantly more reactive and irreversible glycation products are formed faster making MG a major contributor to the formation of AGEs.[17, 20, 21] In humans, AGEs are increased in inflammatory diseases such as diabetes[16], multiple sclerosis[22], atherosclerosis[23], Alzheimer's and Parkinson's diseases[24].

In cells, MG has been reported to affect DNA, RNA and protein synthesis.[1, 9, 25] MG has also been shown to be mutagenic in bacteria[26], yeast[27] and mammalian cell lines[28], primarily causing DNA mutations in G-C base pairs[29]. The major cause of these mutations is thought to be a reaction between the guanine base and MG (**Figure 4.2A**) that alters the hydrogen bonding ability of guanine.[29] MG has also been suggested to inhibit DNA synthesis by crosslinking guanine bases of DNA with lysine and cysteine residues of DNA polymerase.[30]

MG can react with free amino acids as well as the basic side chains of lysine and arginine residues in proteins, with arginine residues being most affected (**Figure 4.2B**).[31] Protein glycation products can lead to loss of function, altered conformation and aggregation[21, 32]. AGEs have also been proposed to contribute to the formation of reactive oxygen species[33] while interactions between AGEs and receptors for AGEs may activate pro-inflammatory signalling pathways[34]. However, small amounts of MG might be beneficial for the cell as MG glycation has been reported to activate some molecular chaperones.[35-37]

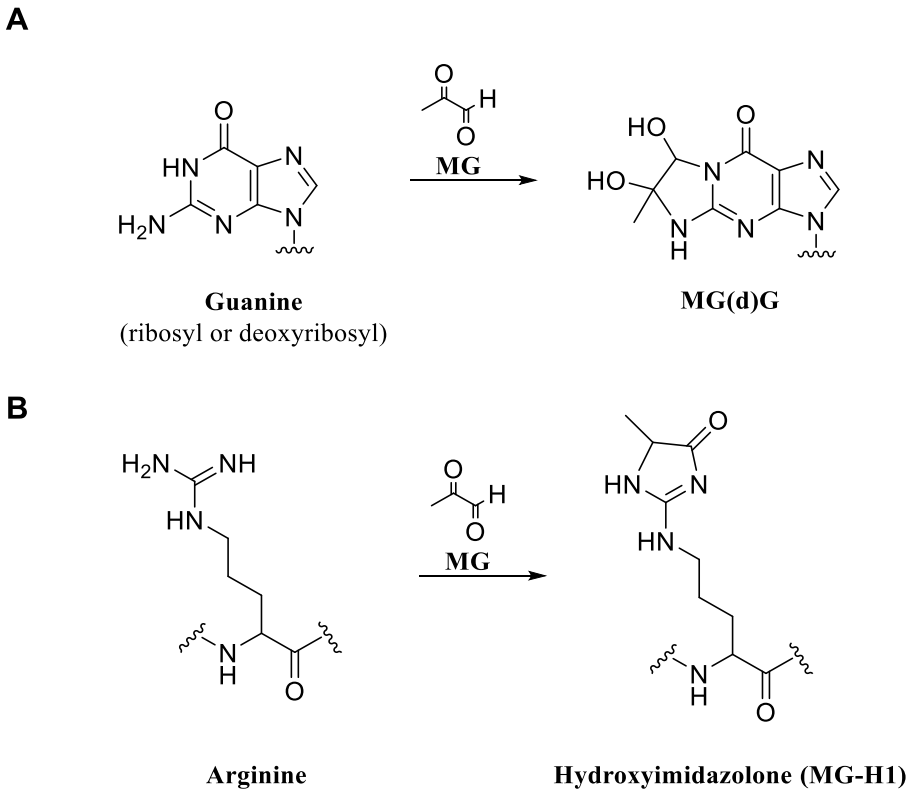


Figure 4.2: Glycation products formed from reaction of MG with (A) guanosine bases in DNA and RNA[38], and (B) arginine residues in proteins.

4.1.2 Glyoxalase system

The glyoxalase system is responsible for eliminating toxic MG and is widely distributed in nature.[9, 39] The glyoxalase system consists of two enzymes – glyoxalase I (Glx1, EC 4.4.1.5) and

glyoxalase II (Glx2, EC 3.1.2.6). A hemithioacetal formed by the non-enzymatic reaction between MG and glutathione (GSH) is converted to S-D-lactoylglutathione by Glx1 (**Figure 4.3**).[40] Glx2 then catalyses the hydrolysis of S-D-lactoylglutathione (SLG) to give D-lactate and GSH. While Glx1 and Glx2 are considered to be a part of the same pathway, they are reported to be regulated independently in *Escherichia coli*.[41] Another enzyme named glyoxalase III (Glx3, EC 4.2.1.130), that catalyses the non-reversible formation of D-lactate from MG without requiring GSH or other cofactors, has been reported in bacteria[42] and subsequently in humans, mice and the nematode *Caenorhabditis elegans*[43]. However, Glx3 is reported to possess other functions and to be less efficient than Glx1 in eliminating MG.[43, 44]

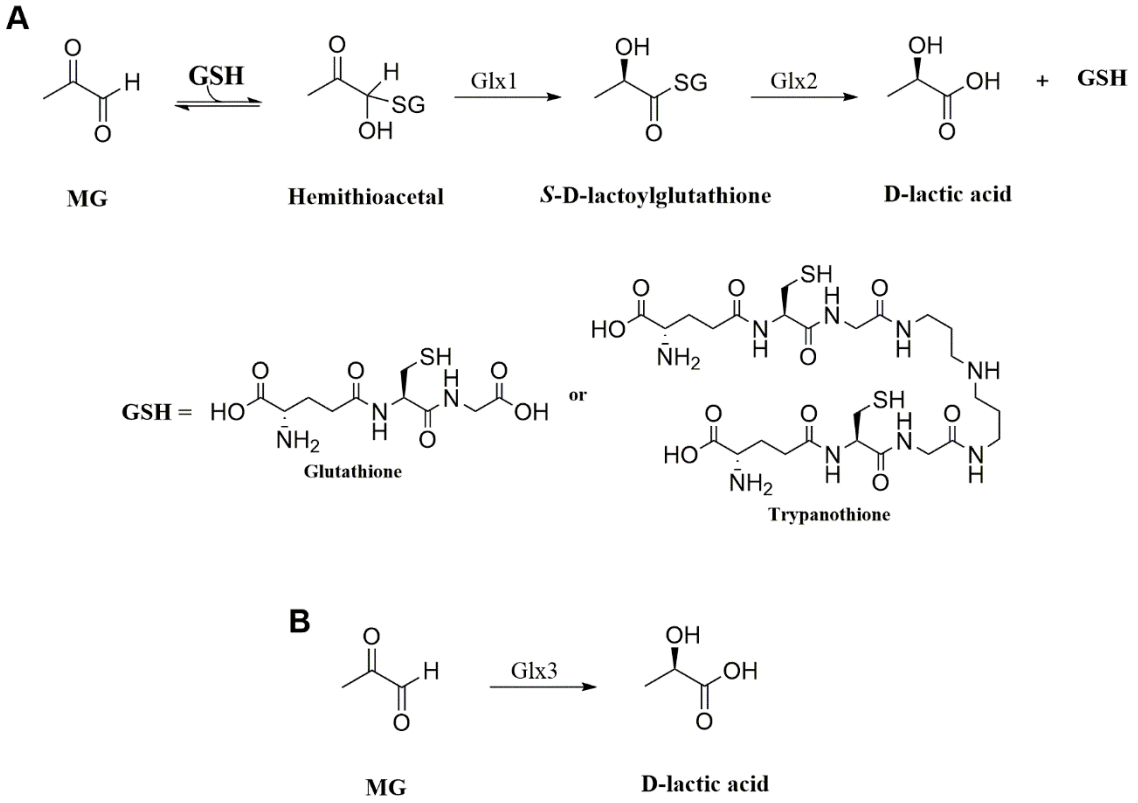


Figure 4.3: Detoxification of methylglyoxal as catalysed by (A) the glyoxalase (Glx1/Glx2) system, and (B) Glx3 enzyme. Trypanothione is preferred to glutathione by Glx1/Glx2 enzymes in trypanosomes.

Glx1 appears to be ubiquitous in all living organisms with the exception of some protozoan parasites which lack a gene encoding for Glx1.[45] Glx1 proteins are apparently mostly dimeric metalloenzymes binding Zn or Ni ions in their active sites, with their substrate binding site positioned at the interface of the two monomers.[46-48] The known exceptions to this include the *Saccharomyces cerevisiae*[49] and *Plasmodium falciparum*[50] Glx1 enzymes which are both monomeric with the latter possessing two active sites located in a single protein subunit[50]. In

most organisms, Glx1 is reported to be selective for the use of GSH as a thiol, while trypanosomes show a strong preference for the use of trypanothione (two GSH molecules linked by a spermidine molecule, shown in **Figure 4.3**).[47] Whilst the MG derived hemithioacetal is often reported as its main substrate, Glx1 can also catalyse the isomerisation of hemithioacetals derived from other α-oxoaldehydes.[51]

Like Glx1, Glx2 is reported to be ubiquitous in nature.[2] The protozoan parasites lacking Glx1 appear to have two Glx2 encoding genes[45], while yeast and higher plants have separate genes encoding for cytosolic and mitochondrial forms of the Glx2 enzyme[52, 53]. Glx2 is a monomeric member of the metallo-β-lactamase (MBL) structural superfamily and contains a highly conserved THxHxDH motif with two metal ions bound at its active site.[54] In addition to its MBL domain, Glx2 contains an α-helical domain at the *C*-terminus with the substrate binding site positioned at the interface between the two domains.[54] Similarly to Glx1, Glx2 shows specificity towards GSH derived substrates with the trypanosome Glx2 being selective for trypanothione as the thiol derivative.[40, 55] The common substrate for Glx2, SLG (**Figure 4.3**), has been reported to have biological activity. SLG has been proposed to stimulate microtubule assembly[56], and has been suggested to act as an alternative source of GSH in the mitochondria[57]. However, SLG has also been reported to be cytotoxic by causing inhibition of DNA synthesis and cell growth in human leukaemia 60 (HL60) cells[58, 59] and chemotaxis impairment in neutrophils[60]. Collectively, these reports suggest that in addition to releasing GSH, Glx2 has other roles in controlling the levels of SLG and associated cytotoxic effects.

While Glx3 catalyses the conversion of MG to D-lactate, it is not considered to be a part of the glyoxalase pathway.[2] The Glx3 enzymes identified so far, appear to be molecular chaperones (bacterial Hsp31[44] and mammalian DJ-1 proteins[61]), with human DJ-1 also being reported to possess protein deglycase activity[62]. Glx3 is a member of the DJ-1/ThiJ/PfpI superfamily[63] and is reported to have 3 potential active sites – a hydrophobic site, likely responsible for the chaperone activity, a protease-like site with Cys, His and Asp residues, and a Zn binding site comprised of two His and one Glu residues.[64] To date, little is known about the reaction mechanism of the Glx3 reaction or what role it plays in the overall MG metabolism.

4.1.3 Glyoxalases as potential drug targets

The glyoxalase system has been explored for potential drug targets as its inhibition could lead to a build-up of MG and/or SLG, cell damage and eventually apoptosis.[65] Both Glx1 and Glx2 have been reported to be overexpressed in human tumours and have been suggested to be involved in tumorigenesis and multidrug resistance.[66] Glx1 has been proposed as a target in

tumours with a high glycolytic rate[66, 67] while Glx2 has been proposed to be a novel marker for prostate cancer as well as a potential therapeutic target, as depletion of Glx2 has been reported to only affect cancerous cells by increasing their susceptibility to anticancer drugs without affecting the cell viability of healthy cells[68, 69]. In addition, high MG concentrations and reduced Glx1 levels are commonly observed in type 2 diabetes and obesity and are thought to lead to impaired metabolic and vascular health.[70] Inducers of Glx1 expression are currently in clinical trials and show promising initial results for improving glycaemic control and reducing vascular damage in overweight individuals.[71]

Glx1 and Glx2 have been proposed to be potential anti-parasitic targets for trypanosome infections.[72, 73] Due to the selectivity of trypanosome Glx enzymes for trypanothione instead of GSH as the thiol component, it could be possible to selectively target parasitic glyoxalases without affecting the host enzymes.[47, 72, 74]

In *E. coli*, knock-outs of Glx1 and Glx2 have been reported to result in reduced MG metabolism, but to have little effect on cell growth and viability under the tested conditions.[75, 76] Cells lacking Glx1 show increased sensitivity to MG at high concentrations, while those lacking Glx2 demonstrated MG sensitivity comparable to the wild-type cells.[75, 76] However, loss of Glx2 activity has resulted in an activation of the K^+ efflux system KefGB and acidification of the cytoplasm.[75] This result was attributed to build up of SLG in the cells. While this demonstrates, that on its own, Glx system may not be a sufficient antibacterial target, therapies targeting the Glx system in addition to other detoxification routes could be more promising.[75]

4.2 Objectives

The work described in this chapter aimed to produce and characterise *E. coli* glyoxalase II (GloB)*. Kinetic, inhibition and structural analyses are reported. A novel high-resolution structure of GloB is reported as well as a structure of a different crystal form of GloB and preliminary work towards substrate and/or product-bound GloB.

**Note that throughout this chapter GloB is used to denote bacterial glyoxalase II proteins, while Glx2 is used for eukaryotic enzymes and the family of homologues. The names Glo2 and HAGH (hydroxyacylglutathione hydrolase) have also been used in literature.*

4.3 Protein production and purification

A commercially obtained synthetic full-length GloB (UniProtKB: P0AC84) gene that was optimised for expression in *E. coli* was inserted into a pET22b expression vector and was provided by Dr Madeline Kavanagh. Mutagenesis PCR was run to remove a duplicate initiation

codon. Recombinant GloB was produced in BL21(DE3) cells and purified by anion exchange chromatography using a Q Sepharose column, followed by size exclusion chromatography. At each stage of purification, GloB purity was assessed by SDS-PAGE analysis (**Figure 4.4**). The final yield of purified GloB was approximately 25 mg protein from 1 l bacterial culture, the concentrated protein solution was yellow in colour. The molecular weight of purified GloB was confirmed by denaturing mass spectrometry (MS). Experimental procedures are described in Chapter 6 (Materials and Methods).

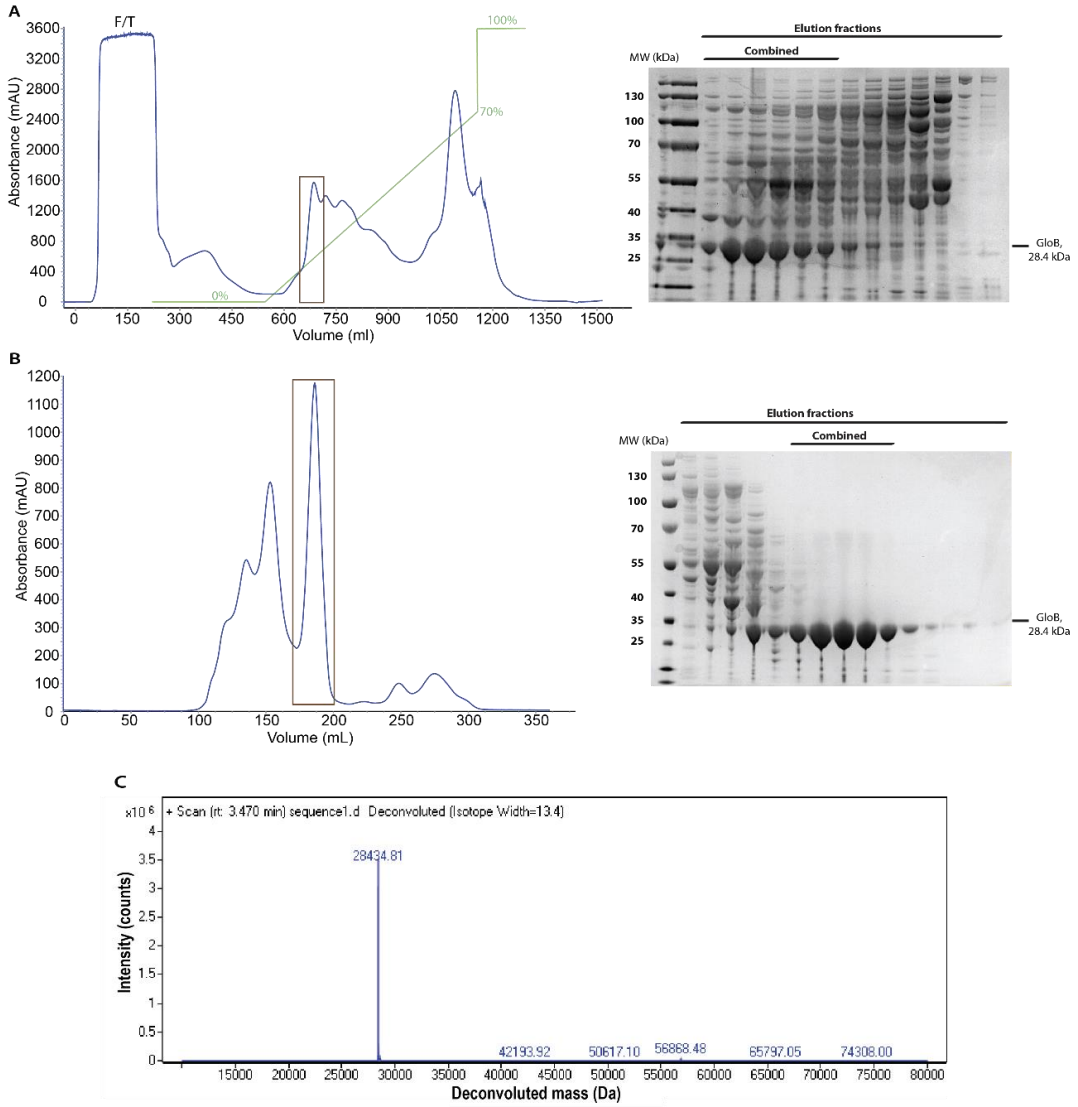


Figure 4.4: GloB purification. (A) UV trace of GloB purification by anion exchange chromatography using Q-Sepharose column and SDS-PAGE gel of selected fractions. The fractions marked as 'Combined' corresponding to the peak highlighted in brown were combined and further purified using size-exclusion chromatography. The green lines denote the gradient of GloB Ion Exchange Buffer B (50 mM Tris, 1000 mM KCl, pH 7.5) with percentages written in green. Flowthrough (F/T) peak is indicated. (B) UV trace of size exclusion chromatography and SDS-PAGE gel of selected fractions. The fractions marked as 'Combined' corresponding to the peak highlighted in brown were combined, concentrated and buffer exchanged to remove KCl. (C) The deconvoluted mass spectrum of purified GloB as determined by denaturing MS (28435 Da) consistent with the predicted molecular weight of 28434.4 Da (including the *N*-terminal Met).

4.4 Characterisation of GloB

4.4.1 Metal content analysis

The metal content of purified GloB was analysed by Inductively Coupled Plasma – Mass Spectrometry (ICP-MS, Department of Earth Sciences, University of Oxford). Purified GloB was found to contain a mixture of transition metals with the main metal present being Fe (**Figure 4.5**). This is contrary to the literature reports of recombinant *E. coli* GloB containing 1.7 Zn ions per protein molecule as analysed by ICP-MS[77]. In addition to Fe, GloB was also shown to contain Zn and Mn with trace amounts (<1%) of Cu, Ni, Cr and Co. Recombinant human, *Salmonella typhimurium* and *Arabidopsis thaliana* Glx2 enzymes have been previously reported to bind a mixture of metals (primarily Fe and Zn, additionally Mn).[78-80] In addition, MBL fold enzymes are known to be promiscuous metal binders[81], possibly explaining the ability of GloB to bind various transition metals, thus it cannot be ruled out that mixed metal ions are catalytically relevant.

The observed metal binding to GloB may have been influenced by the availability of metals during protein production. As no Zn was added to the growth medium, overproduction of a recombinant protein can potentially lead to incorporation of non-native metal ions. Fe is typically a highly bioavailable metal, especially in bacterial cultures[82, 83], consistent with the high abundance of Fe observed in the purified GloB sample. The significance of promiscuous metal binding in GloB and other MBL enzymes is currently unknown.

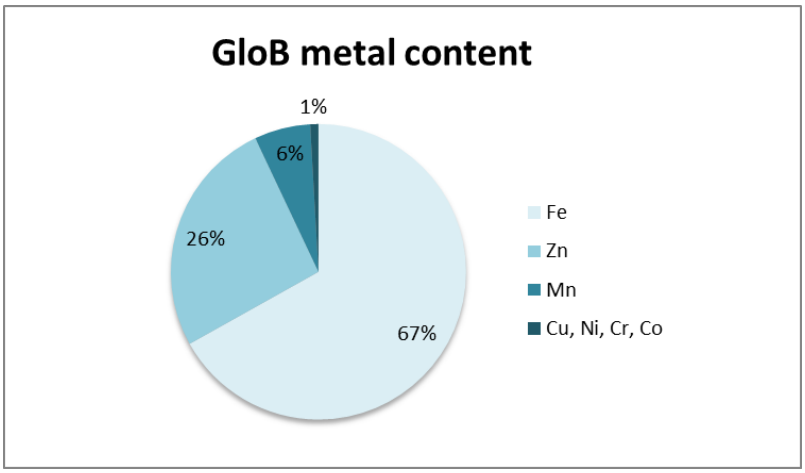


Figure 4.5: Metal ion content of purified recombinant GloB as determined by ICP-MS analysis.

4.4.2 Non-denaturing MS

GloB was subjected to non-denaturing MS analysis. High resolution full scan non-denaturing (native) MS spectra were recorded by Shuang Liu.

Non-denaturing MS spectra of purified GloB imply it is primarily a monomer with a small dimeric population (**Figure 4.6**). The deconvoluted mass of the 'as purified' GloB sample was 28550 ± 8 Da with the resolution being insufficient to tell whether two Zn, two Fe ions or the combination of Zn and Fe ion was the primary metallated state (predicted and deconvoluted masses are listed in **Table 4.1**). A small peak corresponding to a single metal ion bound state was also observed. Treatment with 2,6-pyridine-dicarboxylic acid (dipicolinic acid, DPA; as described by O'Young et al.[84]) was shown to be insufficient to remove all of the bound metals with roughly 1:1 ratio of apo and one metal ion bound species being observed in the treated sample. It is possible that the single metal bound state observed corresponds to GloB with a single Fe ion bound. Previous work in the group has demonstrated that when a 'true' MBL VIM-2 (i.e. an MBL catalysing β-lactam hydrolysis) is subjected to metal titration experiments, it can be observed in apo form or with two Zn ions bound. However, both single metal bound state and two metal ion bound species were observed during Fe^{2+} titration experiments (S. T. Cahill, DPhil thesis)[85].

Table 4.1: Predicted and observed deconvoluted masses of GloB. Predicted holo mass corresponds to di-Zn GloB. Deconvoluted Apo mass was determined using denaturing MS (spectrum is shown in **Figure 4.4C**), deconvoluted holo mass was determined by subjecting 'as purified' GloB sample to non-denaturing MS analysis.

	Predicted	**Deconvoluted**
Apo	28434.4	28434.8
Holo	28565.2 (di-Zn) 28546.0 (di-Fe)	28549.9 ± 7.9

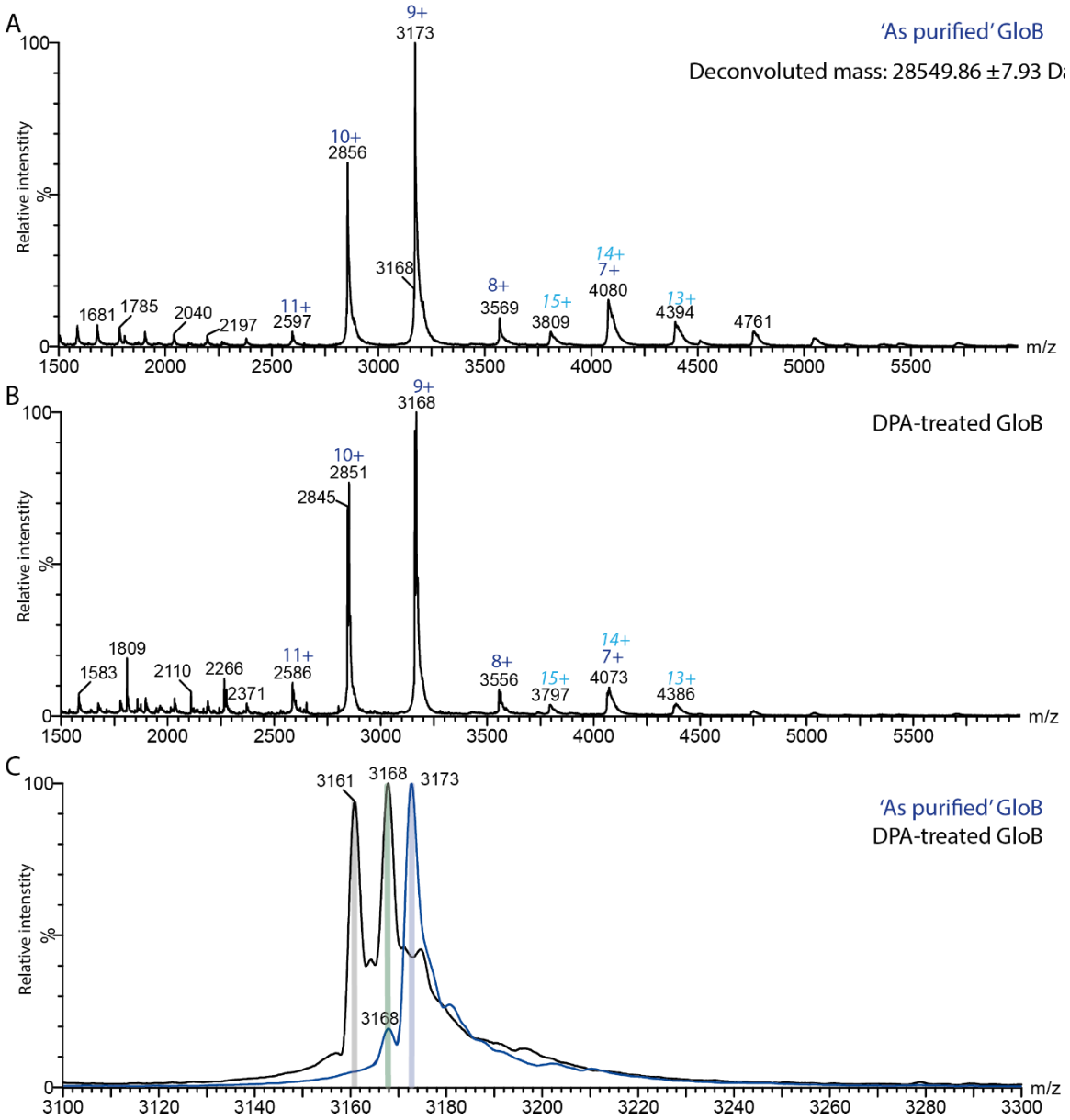


Figure 4.6: Non-denaturing mass spectra of the (A) 'as purified' GloB protein, (B) GloB protein after treatment with DPA. Charge states are given in blue, charge states corresponding to the minor dimeric population are given in light blue. (C) Magnified view of the 9+ charge state of 'as purified' and DPA-treated GloB spectra showing changes resulting from treatment with DPA, with peaks corresponding to two metal ion bound (blue), single metal ion bound (green) and no metal bound (grey) species of GloB shown. The spectra were deconvoluted using MaxEnt1 feature of MassLynx 4.0 software (Waters) over an appropriate mass range.

4.4.3 Circular dichroism

The GloB fold was then assessed using Circular dichroism (CD) spectroscopy (**Figure 4.7A**). GloB treated with 50 mM DPA ('DPA-treated') was found to adopt a similar fold to the untreated protein (**Figure 4.7B**). Thermal stability of the two GloB samples was also investigated. The melting temperature (T_m) of the 'as purified' GloB sample was determined to

be 62.3 ± 1.9 °C with a loss of signal observed at around 37 °C. 'DPA-treated' GloB was significantly less stable with T_m of 40.2 ± 0.5 °C suggesting that loss of one or two metal ions (as determined by non-denaturing MS) has a strong effect on overall thermal stability of the protein making it more prone to denaturation.

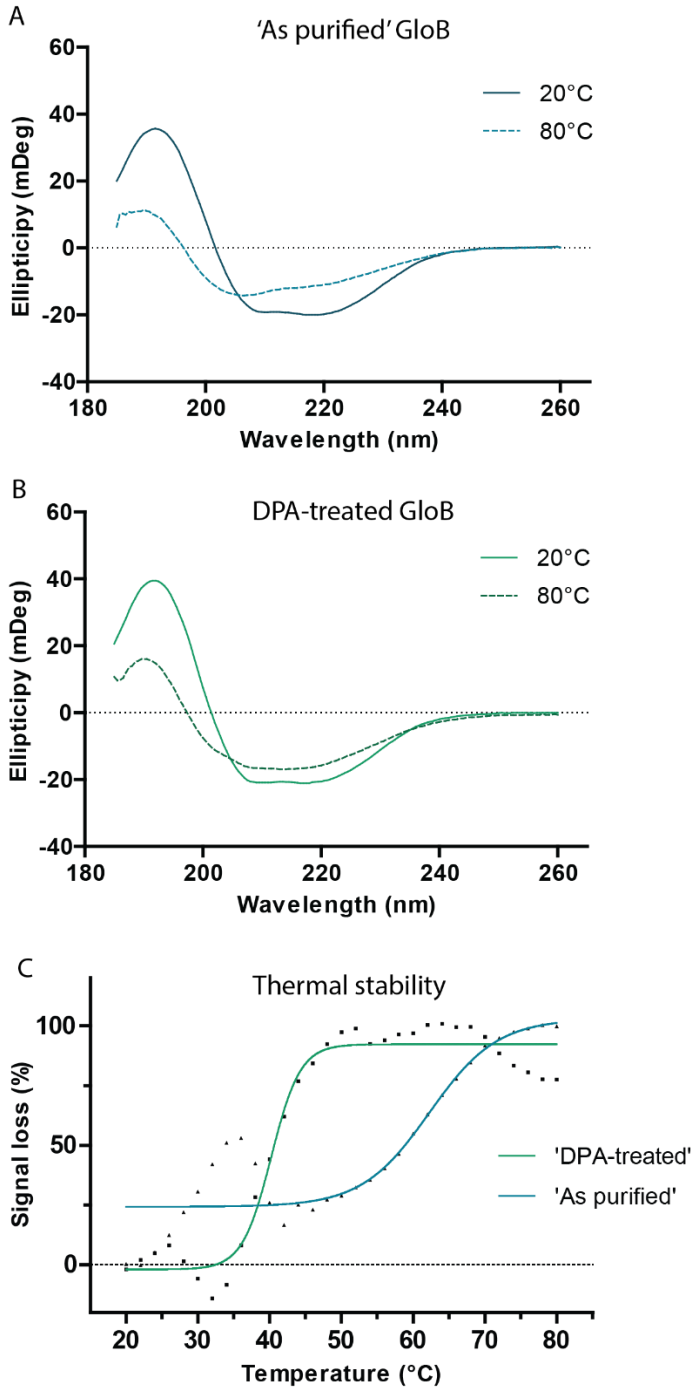


Figure 4.7: Circular dichroism (CD) analysis of GloB. (A) CD spectra of 'as purified' GloB at 20 °C (solid line) and 80 °C (dotted line). (B) CD spectra of GloB treated with 50 mM DPA at 20 °C (solid line) and 80 °C (dotted line). (C) Melting curves of 'as purified' and DPA-treated GloB as determined by CD. Curve fitting was performed using Origin Pro[86].

4.5 Kinetic and inhibition studies

There are two types of kinetic assays for Glx2 activity reported in the literature (**Figure 4.8**). The first assay, which has been previously used for characterisation of *E. coli* GloB[84], measures hydrolysis of a thioester bond of SLG by following the resulting decrease in absorbance at 240 nm (ε_{240} = 3100 M^{-1} cm^{-1}).[52, 84] A chromogenic thiol detection assay has also been used to characterise other GloB enzymes.[87] This assay uses 5,5'-dithiobis-2-nitrobenzoic acid (DTNB, also known as Ellman's reagent) to detect the thiol group of the GSH molecules produced by the GloB hydrolysis reaction. Reaction between GSH and DTNB releases 2-nitro-5-mercapto-benzoic acid (TNB) which can be detected by measuring absorbance at 412 nm (ε_{412} = 14150 M^{-1} cm^{-1}).

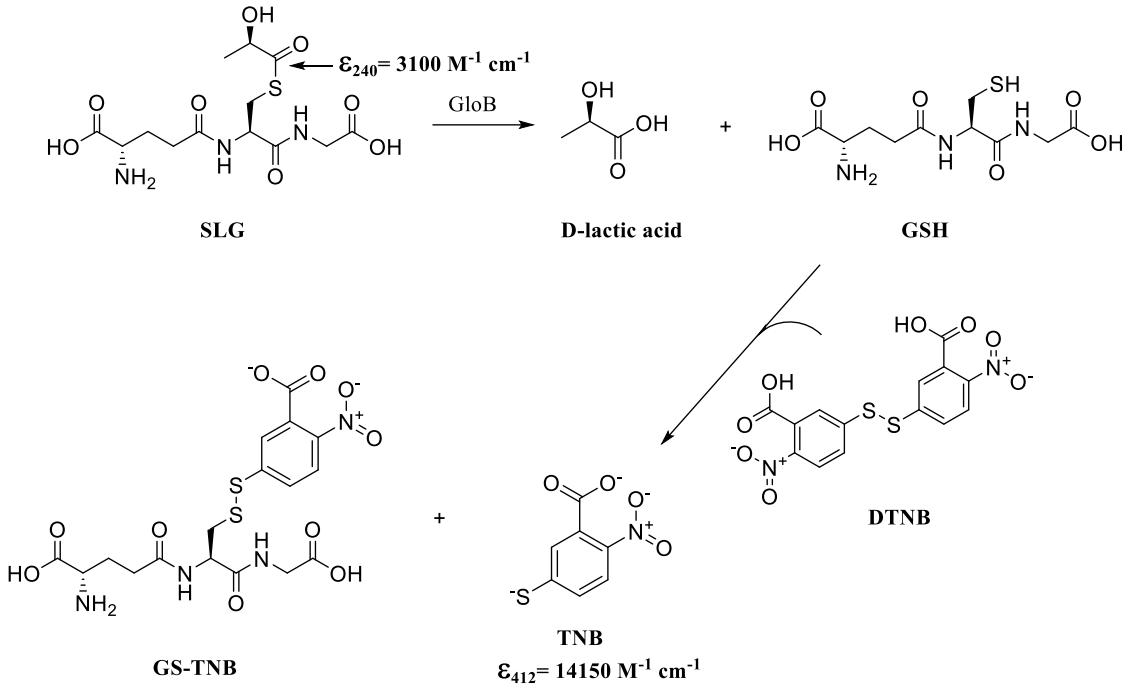


Figure 4.8: SLG hydrolysis catalysed by GloB and the reported spectrophotometric activity assays. The groups detected in the two assay types and their extinction coefficients are indicated.

4.5.1 Kinetics and metal substitution

Purified GloB was subjected to kinetic analysis using the assay conditions described by O'Young et al. (i.e. thioester hydrolysis) adapted to use in a 96-well plate format to enable a direct comparison with the reported values. SLG hydrolysis was measured at 234 nm instead of 240 nm due to the highest absorbance peak for the SLG thioester bond being observed at 234 nm. GloB was found to have a K_m of 147.8 ± 37.6 μM (**Figure 4.9**) which was comparable to the literature value of 184 ± 22 μM [84] suggesting that binding a mixture of transition metal ions (Fe, Zn, Mn, as described earlier) does not impair the enzyme's ability to catalyse SLG hydrolysis.

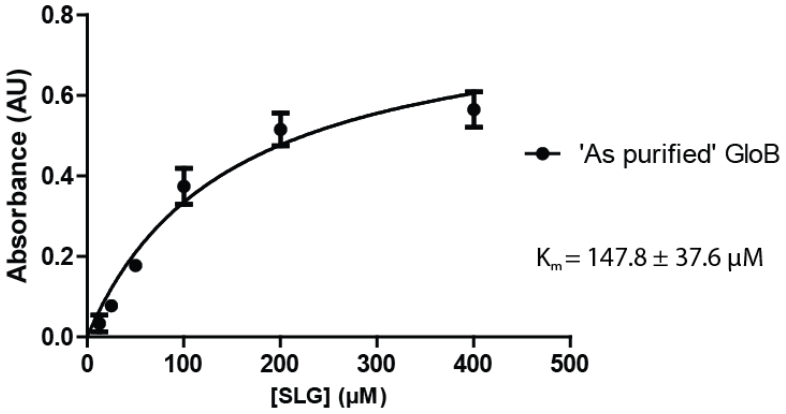


Figure 4.9: Kinetic analysis of 'as purified' GloB using the non-chromogenic assay method monitoring the decrease in absorbance at 234 nm. The initial rate of hydrolysis is plotted against the substrate concentration. Error bars represent standard error (SE), n = 4. Curves were fitted using GraphPad Prism software[88].

Preliminary metal substitution experiments were performed using 'apo' GloB (prepared as described previously). 'Apo' GloB was incubated with 6-fold excess of $CoCl_2$, $NiCl_2$, $MnCl_2$, and $FeCl_3$, and 3-fold excess of $ZnCl_2$ for 30 min at 4 °C before measuring activity. Mn ions were found to be the most efficient, restoring GloB to nearly full activity compared to the 'as purified' enzyme (**Figure 4.10**), while incubation with Zn failed to restore activity. The inability of Zn ions to restore activity is consistent with previous literature reports demonstrating inactivation of GloB by Zn[77]. This has been previously attributed to the presence of Cys residues resulting in mis-metallation of GloB away from the active site.[77] There have also been reports of other proteins reported to bind Zn as their native metal, being inhibited by Zn^{2+} salts, including human leukotriene A_4 hydrolase[89], *E. coli* deacetylase LpxC[90], and *Saccharomyces cerevisiae* phosphomannose isomerase[91]. In addition, another MBL family protein YhfI from *Bacillus subtilis* has been shown to be significantly more active when reconstituted with Mn ions instead of its native Zn ions.[92]

Interestingly, incubation with Fe^{3+} ions has shown partial restoration in GloB activity, despite several MBL enzymes being reported as inactive with ferric ions, but have hydrolase activity while bound to ferrous iron[93]. Results for GloB activity with Co are not shown due to poor reproducibility between different experiments.

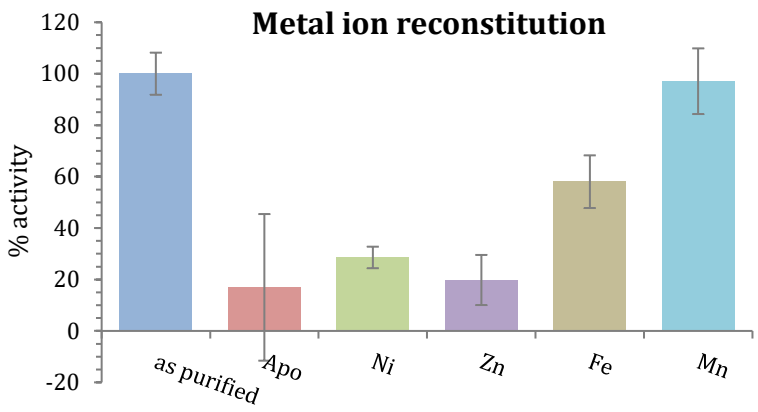


Figure 4.10: Activities of GloB with different metal ions as determined by metal ion reconstitution experiments. For metal reconstitution, 'Apo' (DPA-treated) GloB was incubated with metal chloride salts on ice for 30 min before measuring activity. Error bars represent standard error (SE), n = 4.

4.5.2 Search for GloB inhibitors

Microbial Glx2 enzymes were once proposed to be potential antibacterial targets. However, bacterial GloB gene knock-out cells exhibit no substantial difference in cell viability and growth rate, making GloB an unlikely drug target.[75] Nonetheless, understanding the mode of GloB inhibition could be useful in understanding the differences in inhibitory profiles within the MBL superfamily. Two different types of assays – thioester hydrolysis and thiol detection – were adapted to a 384-well format and were performed in order to compare the results and suitability of both assay methods.

A relatively broad scope of potential inhibitors was chosen for initial assays, including β-lactam antibiotics, transition metal chelators, known inhibitors of other MBL superfamily enzymes, and nucleotides, which have been previously shown to inhibit Glx2 enzymes in tissue homogenates[94]. A full list of inhibitors tested is provided in Appendix 1. GloB was incubated with the inhibitors for 10 min at room temperature before addition of SLG to initiate the reaction (final inhibitor concentration 40 μM). All measurements were performed in quadruplicate using a 384-well format. Hits (<40% residual activity in initial screens using both thioester hydrolysis and thiol detection assays) were then investigated further to determine IC_{50} values (40 μM highest inhibitor concentration). IC_{50} and pIC_{50} values for selected inhibitors (**Figure 4.11**) are listed in **Table 4.2**.

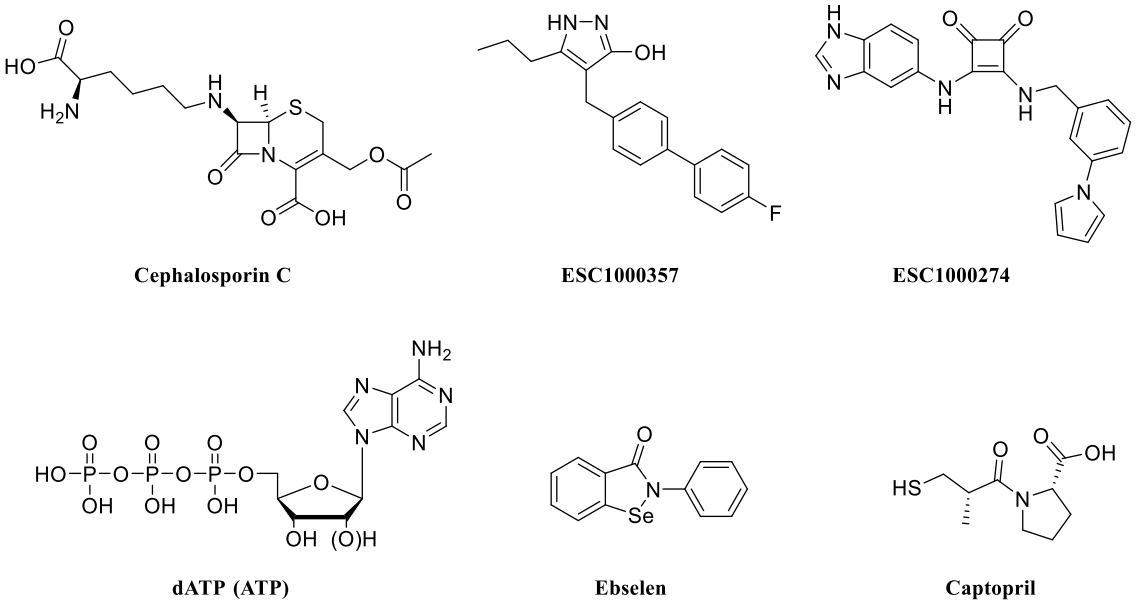


Figure 4.11: Chemical structures of inhibitors reported in **Table 4.2** screened against GloB.

The best GloB inhibitor identified was a squaramide (ESC1000274) which has an IC_{50} in the low μM range (IC_{50} curves for select compounds are shown in **Figure 4.12**, IC_{50} and pIC_{50} values are listed in **Table 4.2**). No better inhibitors have been detected during the screenings. Seven other squaramides have been screened, one of which (Sq6) inhibited GloB to <40% residual activity in the initial screen and had been selected for IC_{50} assays. Sq6 showed poorer inhibition than ESC1000274 (**Figure 4.13**, **Table 4.2**).

Table 4.2: IC_{50} and pIC_{50} values for selected compounds tested for GloB inhibition using different assays. A corresponds to measurements obtained by following absorbance at 234 nm. B shows values obtained using DTNB for thiol detection, absorbance was measured at 413 nm. Curves were fitted using GraphPad Prism[88]. ND – could not be determined. Amb – data could not be fitted unambiguously. ESC1000274 and ESC1000357 were provided by the European Lead Factory.

Compound	A		B	
	IC_{50} (μM)	**pIC_{50}**	**IC_{50} (μM)**	**pIC_{50}**
ESC1000274	16.0	4.80 ± 0.03	9.1	5.04 ± 0.04
Cephalosporin C	32.8	4.5 ± 0.1	30.4	4.52 ± 0.08
Captopril	82.9	4.1 ± 0.7	ND	ND
ESC1000357	107	4.0 ± 0.2	35.9	4.44 ± 0.08
Ebselen	109	4.0 ± 0.4	-	-
ATP	1740	3 ± 9	122	3.9 ± 0.8
dATP	Amb	Amb	92.7	4.0 ± 0.3
Sq6	Amb	Amb	-	-

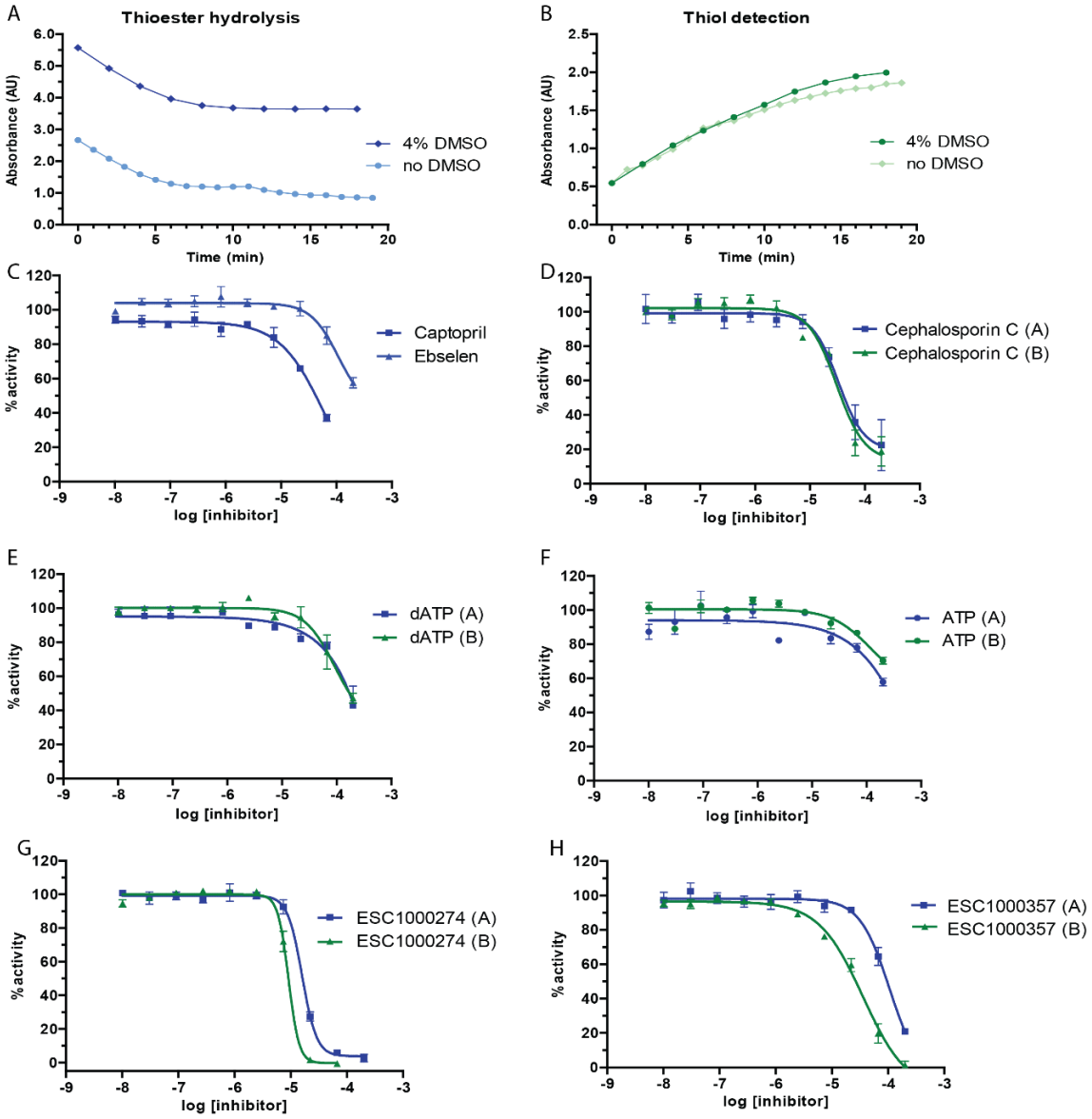


Figure 4.12: Comparison of thioester hydrolysis and thiol detection assays for GloB activity and inhibition. (A) Typical curve observed during thioester hydrolysis assays using 4 nM GloB and 400 µM in presence and absence of 4% (v/v) DMSO. (B) Typical curve observed during thioester hydrolysis assays using 2 nM GloB, 200 µM SLG and 320 µM DTNB in presence and absence of 4% (v/v) DMSO. IC_{50} traces for GloB inhibition by (C)Captopril and Ebselen, (D) Cephalosporin C, (E) dATP, (F) ATP, (G) ESC1000274, and (H) ESC2000357. Thioester hydrolysis (blue, A) was followed by measuring decrease in absorbance at 234 nm. Thiol detection assay (green, B) using DTNB was followed by measuring increase in absorbance at 413 nm. Error bars represent standard error (SE), n = 4. Curves were fitted using GraphPad Prism software[88].

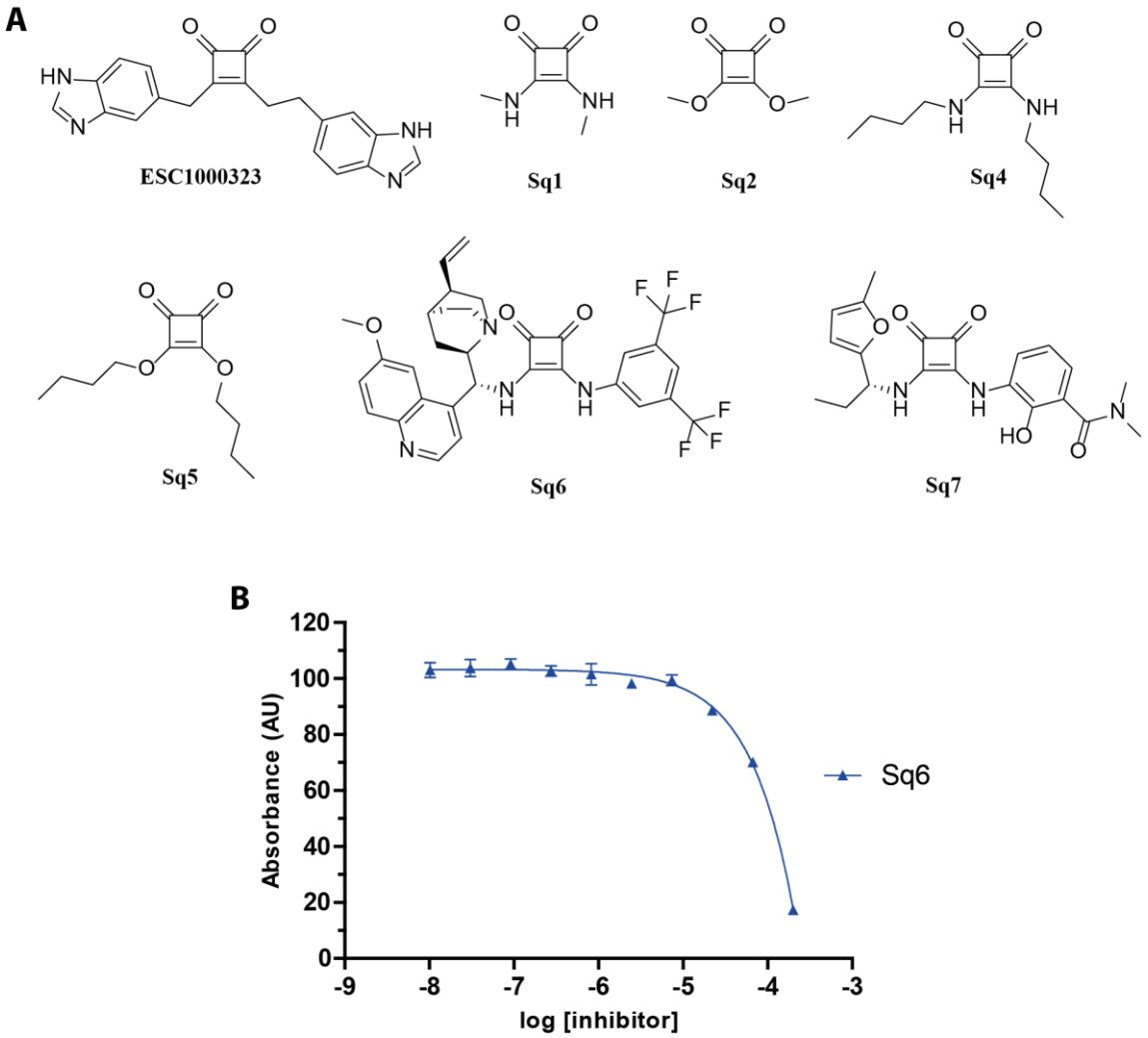


Figure 4.13: (A) Chemical structures of squaramides included in the broad screens for GloB inhibitors. Sq6 was further analysed to determine the IC_{50} value. (B) IC_{50} trace obtained from thioester hydrolysis assay using Sq6 as an inhibitor. The IC_{50} value could not be determined unambiguously. Error bars represent standard error (SE), n = 4. IC_{50} curve was fitted using GraphPad Prism software[88].

Overall, all the compounds which exhibited inhibition had several metal chelating groups present and likely interacted with the active site metals. It has been previously reported that GloB from *S. typhimurium* shows different levels of product inhibition with GSH having sub-mM IC_{50} values for Fe- and Zn-enriched GloB and 1.5 mM IC_{50} value for Mn-enriched GloB.[95] In addition, β-lactam hydrolysing MBLs have been reported to exhibit different inhibition efficiencies depending on whether Zn or ferrous Fe[93], or Zn or Cd[96] ions were bound to the active site, likely due to the metal chelating nature of the inhibitors used in these studies. Since GloB has been shown to bind a mixture of Fe, Zn and Mn ions, the poor inhibition observed could be a result of non-homogenous metal binding with inhibitors affecting only a part of the enzyme species, and not the others. However, since reconstituting GloB with Zn ions has been demonstrated to be unsuccessful in restoring enzyme activity (consistent with report from

O'Young et al.[77]), different approaches should be taken to ensure a more homogenous metal incorporation.

Inhibitor analysis employing the thioester hydrolysis detection at 234 nm had large errors, although for $pIC_{50} > 4.0$ (i.e. better inhibitors) the errors decreased. The higher errors could be explained by the absorbance wavelength used being in a relatively noisy part of the UV spectrum with many compounds absorbing light at around 220 – 250 nm, particularly the compounds containing aromatic groups[97] which could hinder analysis of the signal for the thioester bond. In addition, since not all of the compounds tested were soluble in water, dimethyl sulfoxide (DMSO) was used to dissolve the water insoluble compounds. Due to DMSO absorbance in the same region of the spectrum, the initial absorbance in presence of 4% (v/v) DMSO was approximately two times higher than the starting absorbance in absence of DMSO (**Figure 4.12A**) making the assay more suitable for use with the water soluble compounds. While the thiol detection assay was observed to have smaller errors, higher signal sensitivity, and no interference by 4% (v/v) DMSO (**Figure 4.12B**), due to its nature, it was unsuitable for use with compounds containing thiol groups (e.g. captopril). Comparison of data obtained using both assays for uninhibited enzyme and several inhibitors is shown in **Figure 4.12C-F**. Overall, the results obtained from both assays were comparable in terms of curve shape, even though the IC_{50} traces could not always be fitted unambiguously (as shown in **Table 4.2**).

4.6 Crystallography

4.6.1 Buffer optimisation

Thermal stability of GloB was tested under different conditions using thermal shift assays[98] to determine optimal buffer conditions for crystallography. GloB was determined to be a thermally stable protein with a T_m of 62.4 °C in Milli-Q water. GloB was found to be most thermally stable in HEPES and Tris buffers in the pH range of 7 - 8.5. Addition of 0.1 mM of ZnCl was found to stabilise GloB in 50 mM Tris, pH 7.5 buffer but destabilise in 25 mM HEPES, pH 7.5. Increasing the NaCl concentration had a destabilising effect on GloB protein thermal stability (**Figure 4.14**).

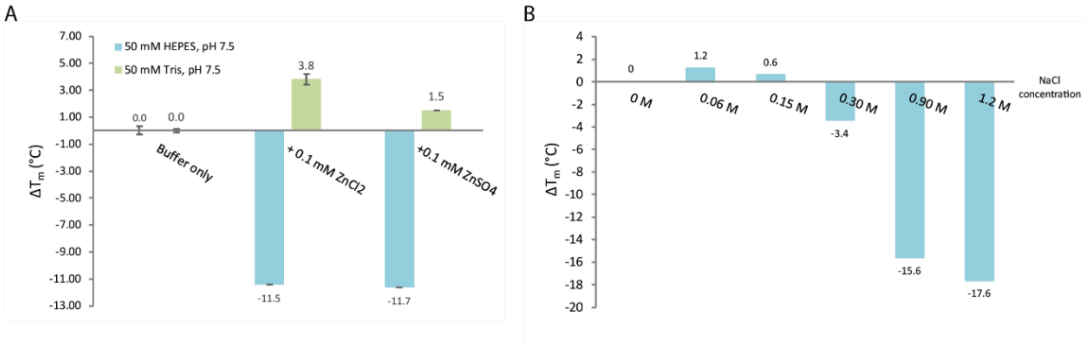


Figure 4.14: Thermal shift assays for GloB. (A) Changes in melting temperature (ΔT_m) of GloB following addition of 0.1 mM $ZnCl_2$ and 0.1 mM $ZnSO_4$ in 50 mM HEPES, pH 7.5, and 50 mM Tris, pH 7.5. Assays were run in triplicate. (B) The effect of increasing NaCl concentration to GloB thermal stability in 60 mM Tris, pH 8.0. Error bars represent standard error (SE), n = 3.

4.6.2 Crystallisation trials

Initial crystallisation trials were performed using 10 – 50 mg/ml GloB in different buffers (**Table 4.3**). Small (<30 µm) non-diffracting crystals were obtained in multiple conditions. The presence of 50 mM KCl, 50 mM NaCl or 0.1 mM $ZnCl_2$ appeared to have no positive impact on GloB crystallisation (**Figure 4.15**).

Table 4.3: GloB sample buffer conditions used for initial crystallisation experiments.

Buffer	Salt	Additive
50 mM Tris, pH 7.5	-	-
50 mM Tris, pH 7.5	-	0.1 mM $ZnCl_2$
50 mM Tris, pH 7.5	50 mM KCl	-
50 mM Tris, pH 7.5	50 mM KCl	0.1 mM $ZnCl_2$
25 mM HEPES, pH 7.5	50 mM NaCl	-
25 mM HEPES, pH 7.5	50 mM NaCl	0.1 mM $ZnCl_2$

Freshly purified GloB in 50 mM Tris, pH 7.5 was used to set up broad crystallisation screens at 4 °C and 20 °C (JCSG+, Wizard Classic 1&2, and PACT Premier obtained from Molecular Dimensions; Index, Crystal Screen, PEG Ion, SaltRx, PEGRx, Silver Bullets, Silver Bullets Bio, and Additive Screen HT from Hampton Research). Large (>100 x 200 µm) crystals appeared overnight in a condition containing 100 mM Tris base-HCl, pH 7.0, 200 mM calcium acetate, 20% PEG 3000 (Wizard Classic 1&2 F6, **Table 4.4**). Crystals were screened using an in-house X-ray source and showed diffraction to high resolution. Data were processed to 1.6 Å resolution in space group $P2_12_12_1$ (**Table 4.5**). The GloB structure was solved by molecular replacement using the structure of *S. thyphimurium* LT2 GloB (PDB code: 2QED)[78] as a search model (78% sequence identity).

Table 4.4: Crystallisation conditions yielding high-resolution GloB crystal structures.

Well	Condition	Temperature	Time to appear
Wizard Classic 1&2 F6	100 mM Tris, pH 7.0, 200 mM calcium acetate, 20% (w/v) PEG3000	4 °C and 20 °C	Overnight
Wizard Classic 1&2 E3	100 mM Tris, pH 8.5, 200 mM magnesium chloride, 20% (w/v) PEG 8000	4 °C	3 weeks
Wizard Classic 1&2 H3	100 mM CAPS, pH 10.5, 200 mM sodium chloride, 20% (w/v) PEG 8000	4 °C	3 weeks

The crystals were further analysed at Diamond Light Source I03 beamline. The original crystals, as well as crystals obtained in conditions listed in **Table 4.4**, diffracted to high resolution (1.0 – 1.1 Å space group $P2_12_12_1$, same unit cell parameters as in **Table 4.5**). The best high-resolution dataset was selected, with a resolution of 1.1 Å and high completeness. The structure was solved by molecular replacement using the model obtained from the in-house data collection. Phaser[99] was used for molecular replacement and refinements were performed using iterative cycles of Phenix Refine[100] and model building using Coot[101]. Data collection and refinement statistics for initial model and final structure (**Figure 4.16**) are shown in **Table 4.5**.

The asymmetric unit contains a single monomeric protein molecule, the final model contains all amino acid residues (1-251), two metal ions in the active site (modelled as Zn), two additional metal ions (modelled as Zn), two chloride ions, a fragment of CAPS buffer, and a glycerol molecule. The GloB structure consists of an MBL domain and a *C*-terminal domain found in the Glx2 proteins, consisting of five α-helices (**Figure 4.16B**). The symmetry related molecule was positioned right next to the active site of GloB with His56 coordinating a bridging Zn ion which also coordinated to His181 of a symmetry related molecule (**Figure 4.16C & D**).

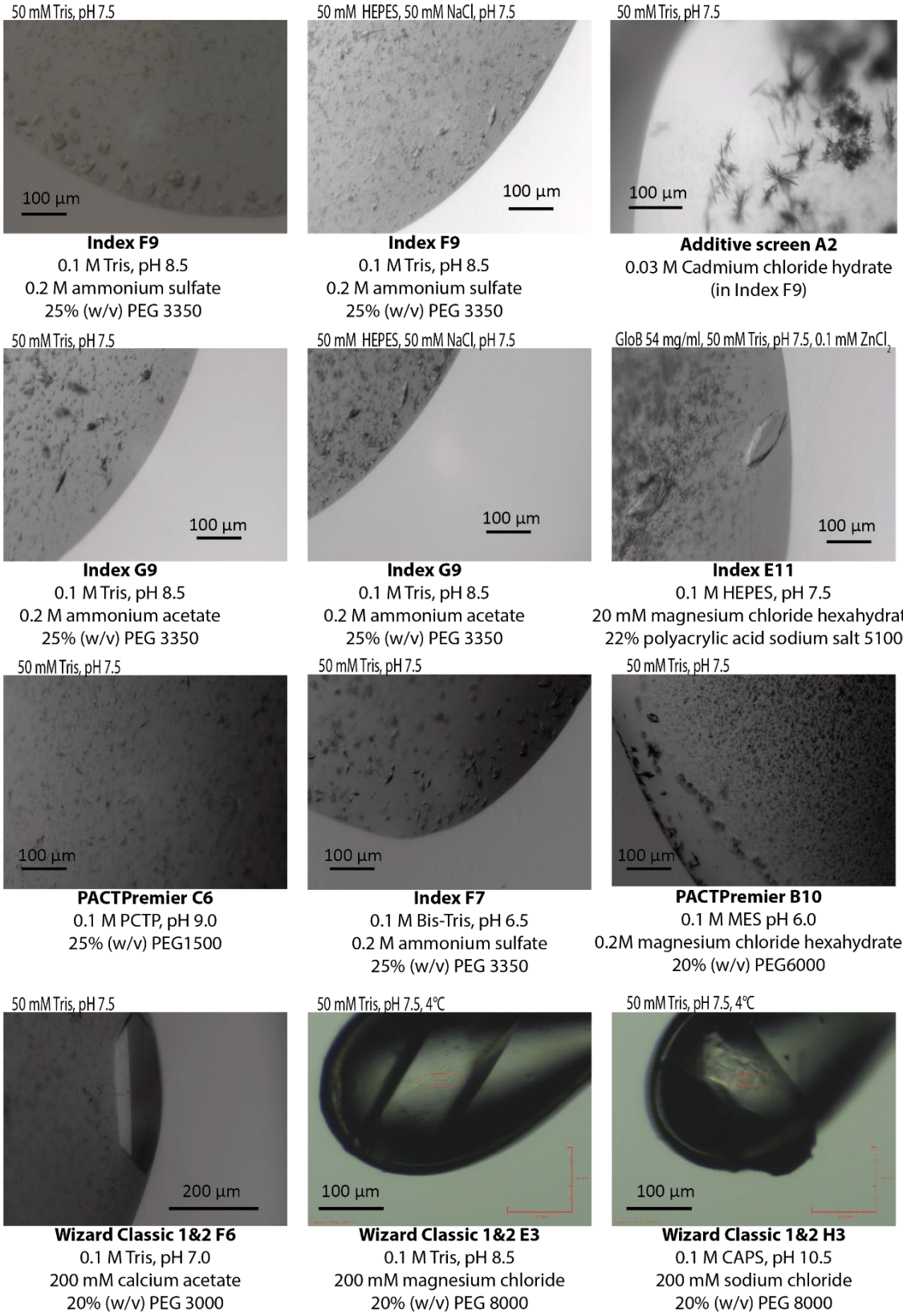


Figure 4.15: Crystals obtained during crystallisation trials of GloB. Crystals were grown at 20 °C using 20 mg/ml protein solution, unless stated otherwise. Text above each image indicate protein sample buffer conditions, crystallisation conditions are listed below.

Table 4.5: Data collection and refinement statistics for initial structure and high-resolution GloB structure. *Values for the highest resolution shell are given in parentheses. ASU – asymmetric unit. ‡ Ligands include Zn ions, glycerol molecule, acetate ion in initial model and CAPS buffer fragment in the high resolution dataset.

	Initial structure	**High resolution**
PDB code	-	6RZ0
Source	In-house (RIGAKU FR-E Superbright)	Diamond Light Source, I03 beamline
Detector	Saturn 944+	Pilatus3 6M
Data processing	HKL200	AutoPROC+STARANISO (XDS)
Wavelength (Å)	1.5419	0.9763
Resolution (Å)	31.49 – 1.60 (1.66 – 1.60)*	47.78 – 1.10 (1.14 – 1.10)*
Space group	$P2_12_12_1$	$P2_12_12_1$
Unit cell parameters		
a, b, c (Å)	59.3, 63.3, 72.6	59.4, 63.4 72.7
α, β, γ	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Molecules in ASU	1	1
Unique reflections	33183 (2690)*	111285 (10573)*
Multiplicity	4.4 (3.1)*	2.0 (2.0)*
R_{merge}	0.098 (0.158)*	0.0221 (0.0563)*
CC1/2	0.994 (0.972)*	0.999 (0.989)*
Completeness (%)	90.00 (74.06)*	99.60 (96.02)*
I/σ(I)	32.1 (11.4)*	37.42 (17.08)*
Wilson B factor (Å²)	11.16	7.05
Refinement	Phenix Refine	Phenix Refine
Number of atoms		
Protein	2024	2264
Ligands‡	21	21
Water	429	496
R_{work}/R_{free}	0.1508 / 0.1853	0.1208 / 0.1293
RMSD		
Bonds (Å)	0.06	0.06
Angles (°)	1.14	1.25
Average B factors (Å²)		
Protein	14.28	9.46
Ligands‡	31.05	20.19
Water	26.79	22.78
Ramachandran plot		
Favoured (%)	98.0	98.0
Allowed (%)	2.0	2.0
Outliers (%)	0.0	0.0

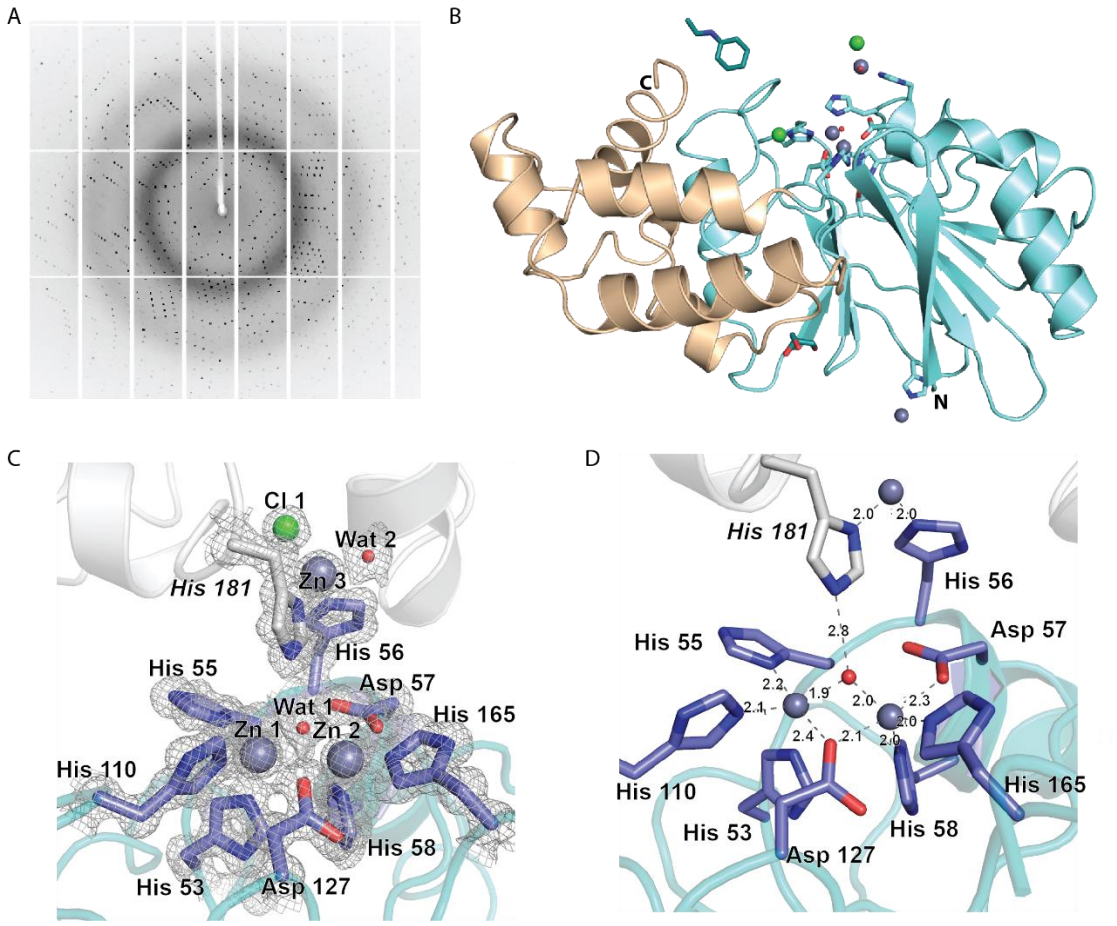


Figure 4.16: High resolution crystal structure analysis of GloB. (A) X-ray diffraction pattern of GloB. Data was collected at Diamond Light Source I03 beamline. (B) A view from the crystal structure of GloB showing the MBL domain (cyan) and *C*-terminal Glx2 domain (brown). Bound Zn ions (grey spheres), chloride ions (green spheres), CAPS buffer and glycerol molecules (teal), and water molecules (red spheres) coordinating to Zn ions are also shown. (C) Enlarged view of the active site with 2mFo-dFc electron density maps (contoured to 1σ) showing the position of symmetry related molecule (white). (D) Enlarged view of the active site showing distances in Å between the active site residues, metals and the symmetry related molecule.

Further experiments were limited by poor reproducibility in obtaining GloB crystals. Efforts to reproduce the crystals were only successful when using the same 96-well block as for the originally obtained highly diffracting crystals. Use of a 96-well block freshly prepared from commercially available 10 ml stocks failed and optimisation attempts resulted in smaller irregular crystals, that only diffracted to approximately 3 Å resolution appearing after 1-4 weeks at 20 °C (**Figure 4.17**) as opposed to previous crystals in the same condition appearing overnight. Smaller crystals (< 100 μm) of similar morphology could be obtained using seeds from the original crystals.

^{1}H Nuclear Magnetic Resonance (NMR) analysis of the F6 condition from the original block has shown some trace amounts of impurities (including isopropanol) present in addition to the original mixture which may aid crystallisation of GloB (data not shown). Additionally,

inorganic salts, which would not be detectable by ^{1}H NMR could also be present. Additives were screened in efforts to obtain crystal of similar properties (Additive screen HT, Silver Bullets, Silver Bullets Bio, and Detergent Screen, all Hampton Research, using Wizard Classic 1&2 F6 condition in all wells), but no diffracting protein crystals were obtained.

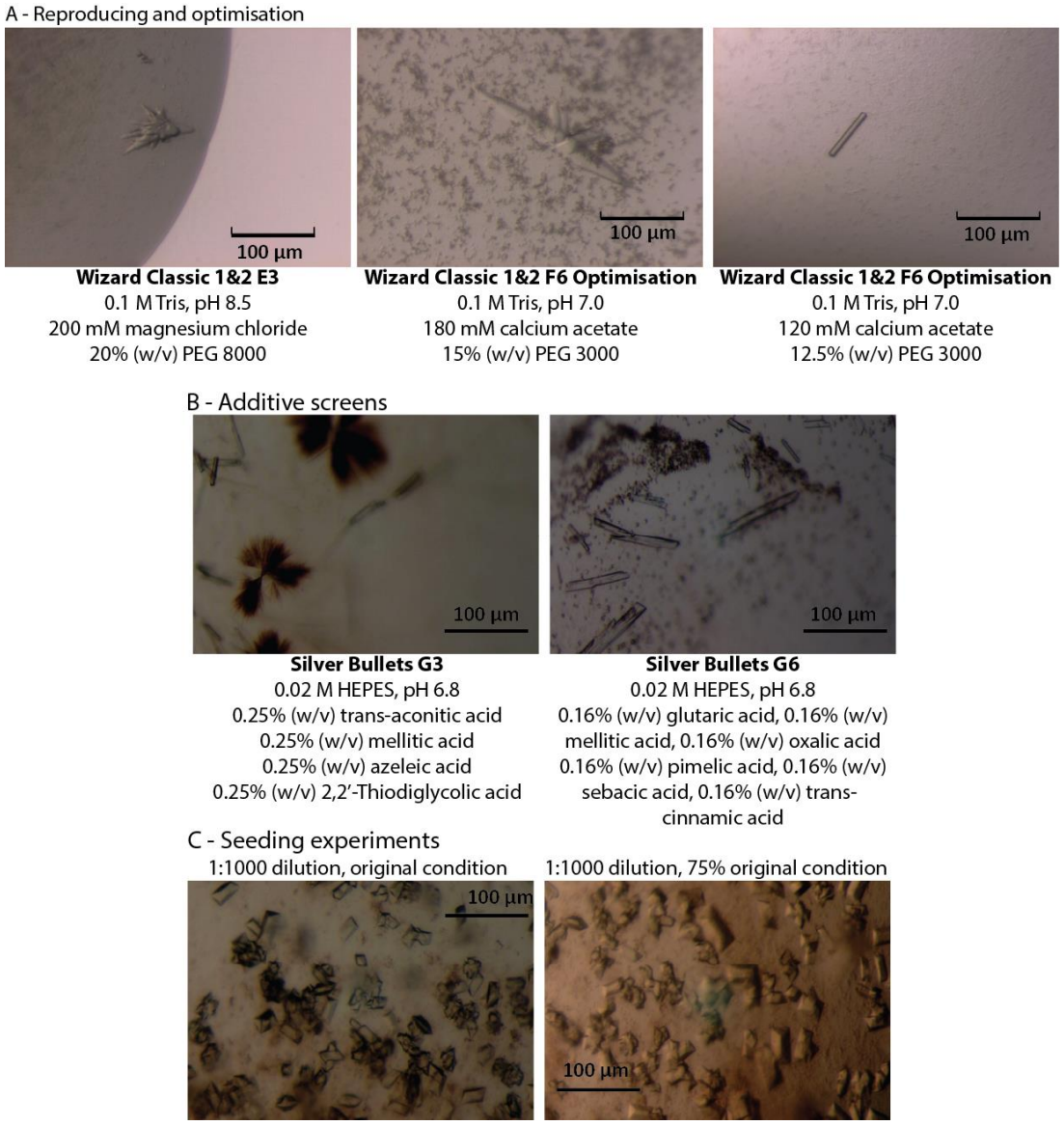


Figure 4.17: Crystals resulting from efforts to reproduce well diffracting GloB crystals obtained in conditions listed in **Table 4.4**. (A) Crystals obtained when trying to reproduce original crystals and during the optimisation of Wizard Classic 1&2 F6 condition. (B) Crystals obtained using additive screens with Wizard Classic 1&2 F6 condition in wells. (C) Crystals obtained from seeding experiments, using seeds from original crystals (Wizard F6). Crystallisation experiments were performed at 20 °C. Crystallisation conditions are listed under photos.

4.6.3 Search for an alternative crystal form

A crystal structure of human Glx2 in complex with a slow reacting substrate *S*-(*N*-hydroxy-*N*-bromophenylcarbamoyl)glutathione and a hydrolysis product GSH has been

reported (PDB ID: 1QH5).[54] A structure of the *Leishmania infantum* Glx2 bound to the second product D-lactate has also been solved (PDB ID: 2P1E).[55] No enzyme-substrate or enzyme-product complexes for bacterial GloB enzymes have been reported to date.

Soaking experiments using crystals described previously were performed in an attempt to obtain a crystal structure of an enzyme-substrate or enzyme-product complex. However, no electron density for bound molecules was observed in the active site, likely caused by the positioning of a symmetry related molecule partly blocking the active site (**Figure 4.16C & D**). This, combined with difficulty in reproducing crystals, rendered this crystal form unsuitable for soaking experiments.

Further screening experiments were performed in order to identify a new crystal form with a more accessible active site in the crystalline state. Crystallisation trials were set up with the Morpheus (I and II) and Wizard Classic 1&2 screens (all from Molecular Dimensions) at 4 °C and 20 °C. Crystals were obtained in conditions containing 0.1 M Buffer system 2 (sodium HEPES, MOPS acid), pH 7.5 , 0.1 M carboxylic acids (mixture of sodium formate, ammonium acetate, sodium citrate tribasic dihydrate, potassium sodium tartrate tetrahydrate, and sodium oxamate, each at 0.02 M) and precipitant mix 4 (12.5% MPD, 12.5% PEG 1000, and 12.5% PEG 3350) at 4 °C. The crystals diffracted X-rays to 1.8 Å resolution and had space group $P2_12_12_1$ (**Figure 4.18**). The crystal structure was solved by molecular replacement (Phaser[99]) using the high resolution GloB structure as a search model. Structure refinements were performed using Phenix Refine[100]. Coot[101] was used for model building. Data collection and processing statistics are given in **Table 4.6**.

The active site of GloB in the new crystal form was more accessible (**Figure 4.18F**). However, the structure contained an L-tartrate molecule bound to the active site (modelled with 50% occupancy) with one of the active site metals shifting 1.7 Å due to this interaction (**Figure 4.18D**). In addition, binding of tartrate was found to displace residue Asp57 so it no longer coordinates the active site metal. In addition, a fragment of HEPES buffer was observed bound near the active site.

Comparison of the two crystal forms of GloB reveals that the two structures are very similar (RMSD of 0.52 Å for backbone Cα atoms, **Figure 4.18E**) with the largest difference arising from the positioning of *C*-terminal helices α8 and α9 (RMSD of 0.53 Å for backbone Cα atoms of residues 223-251). The *C*-terminal domain in the Glx2 family of proteins is comprised of five α-helices (α5 – α9) which are located near the active site and are proposed to be involved in substrate binding[54, 78]. While no substrate is bound in either of the structures obtained, one of

the structures contains L-tartrate bound at the active site. There is also a HEPES molecule near the active site with its sulfate group positioned in a proposed binding site for the carboxylate group of GSH, as based on a comparison with the structure of human Glx2 bound to GSH (**Figure 4.18C**)[54], potentially contributing to the shift of α8 and α9 helices. Overall, the most significant difference observed was in the crystal packing

Table 4.6: Data collection and refinement statistics for L-tartrate (TLA) bound GloB. *Values for the highest resolution shell are given in parentheses. ‡ Ligands include Zn ions, L-tartrate and HEPES buffer fragment.

	GloB-TLA
PDB code	6S0I
Source	Diamond Light Source, I24 beamline
Detector	Pilatus3 6M
Data processing	FastDP (XDS)
Wavelength (Å)	0. 9686
Resolution (Å)	29.11 - 1.803 (1.868 - 1.803)*
Space group	$P2_12_12_1$
Unit cell parameters	
a, b, c (Å)	39.1, 43.7, 159.1
α, β, γ	90.0, 90.0, 90.0
Molecules in ASU	1
Unique reflections	25706 (2322)*
Multiplicity	2.0 (1.9)
R_{merge}	0.02629 (0.2651)
CC1/2	0.999 (0.835)
Completeness (%)	98.93 (91.31)
I/σ(I)	17.27 (2.48)
Wilson B factor (Å²)	21.08
Refinement	Phenix Refine
Number of atoms	
Protein	1971
Ligands‡	22
Water	230
R_{work}/R_{free}	0.1614 / 0.1975
RMSD	
Bonds (Å)	0.004
Angles (°)	0.71
Average B factors (Å²)	
Protein	26.3
Ligands	33.2
Water	37.6
Ramachandran plot	
Favoured (%)	96.8
Allowed (%)	3.2
Outliers (%)	0.0

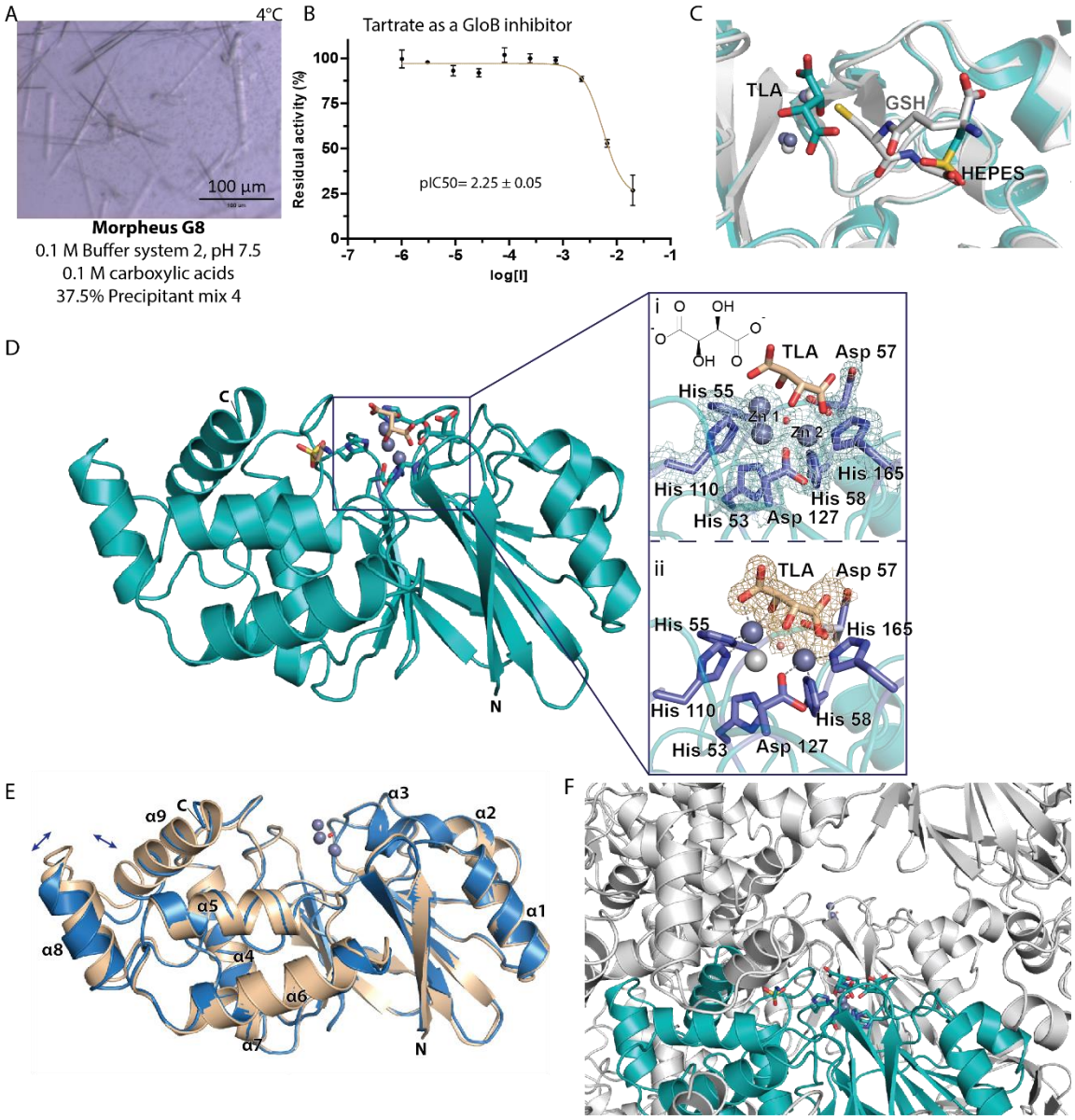


Figure 4.18: The new crystal form obtained during GloB crystallisation trials. (A) Crystals obtained during crystallisation trials. (B) IC_{50} trace for GloB inhibition by tartrate. Assay was performed using thiol-detection by DTNB assay conditions, trace was fitted using GraphPad Prism software[88]. (C) A view from GloB crystal structure with HEPES fragment and TLA bound (cyan), overlaid with human Glx2 structure bound to GSH (PDB: 1QH5, grey). (D) A view from the crystal structure of GloB showing L-tartrate (TLA) and a fragment of HEPES buffer bound (brown), with a magnified view of the active site with a L-tartrate (TLA) molecule bound. 2mFo-DFc maps for active site residues (i, contoured to 1 σ) are shown in cyan. Polder map for tartrate is shown in brown (ii, contoured to 3 σ). (E) Aligned GloB structures for the two crystal forms with the high-resolution structure shown in blue and tartrate-bound structure in brown (RMSD = 0.52 Å for Cα atoms). (F) A view from GloB crystal structure showing positions of symmetry related molecules (white) relative to the active site (cyan, shown in stick representation, Zn ions shown as grey spheres).

Following the structural analysis of GloB showing electron density corresponding to L-tartrate in the active site, GloB inhibition by L-tartrate was tested by kinetic analysis using thiol detection assay and serial dilutions of L-tartrate starting with the highest concentration of 20 mM (**Figure 4.18B**). Tartrate was found to be a weak inhibitor with an IC_{50} value of 5.6 mM

(pIC_{50} = 2.25 ± 0.05). Given its relatively high IC_{50} value, it was deemed possible for L-tartrate to be displaced by a substrate or an inhibitor if used at a sufficiently high concentration. Attempts to optimise crystallisation conditions showed tartrate was important for crystallisation as no crystals were observed in conditions not containing tartrate. L-Tartaric acid is a naturally occurring dicarboxylic acid, commonly found in fruit. *E. coli* cells are able to ferment L-tartrate under anaerobic conditions.[102] While tartrate could be present in cells together with GloB, it is unlikely to reach mM concentrations required for inhibition, hence the binding observed is unlikely to have any biological significance. However, the tartrate binding mode observed in the crystal could potentially be exploited in the design of GloB inhibitors.

Attempts to remove tartrate from the active site were performed by soaking the crystals in a newly prepared well condition lacking tartrate. The crystals were observed to begin to slowly degrade after 60 min (**Figure 4.19A**). After 60-90 min the crystals were either harvested or transferred to a condition containing SLG or *S*-methylglutathione and soaked for 30 s – 30 min before harvesting. The soaked crystals diffracted to 2.0 – 2.6 Å resolution. Preliminary electron density maps show that soaking crystals in a condition with tartrate absent successfully removed tartrate from the active site. Residual electron density was observed around the active site in several datasets, but the occupancies of the bound molecules were too low to define the position and the identity of the ligand (2 Å dataset shown in **Figure 4.19B**). Consequentially, further refinement of the structures has not been pursued. Data collection statistics are given in **Table 4.7**.

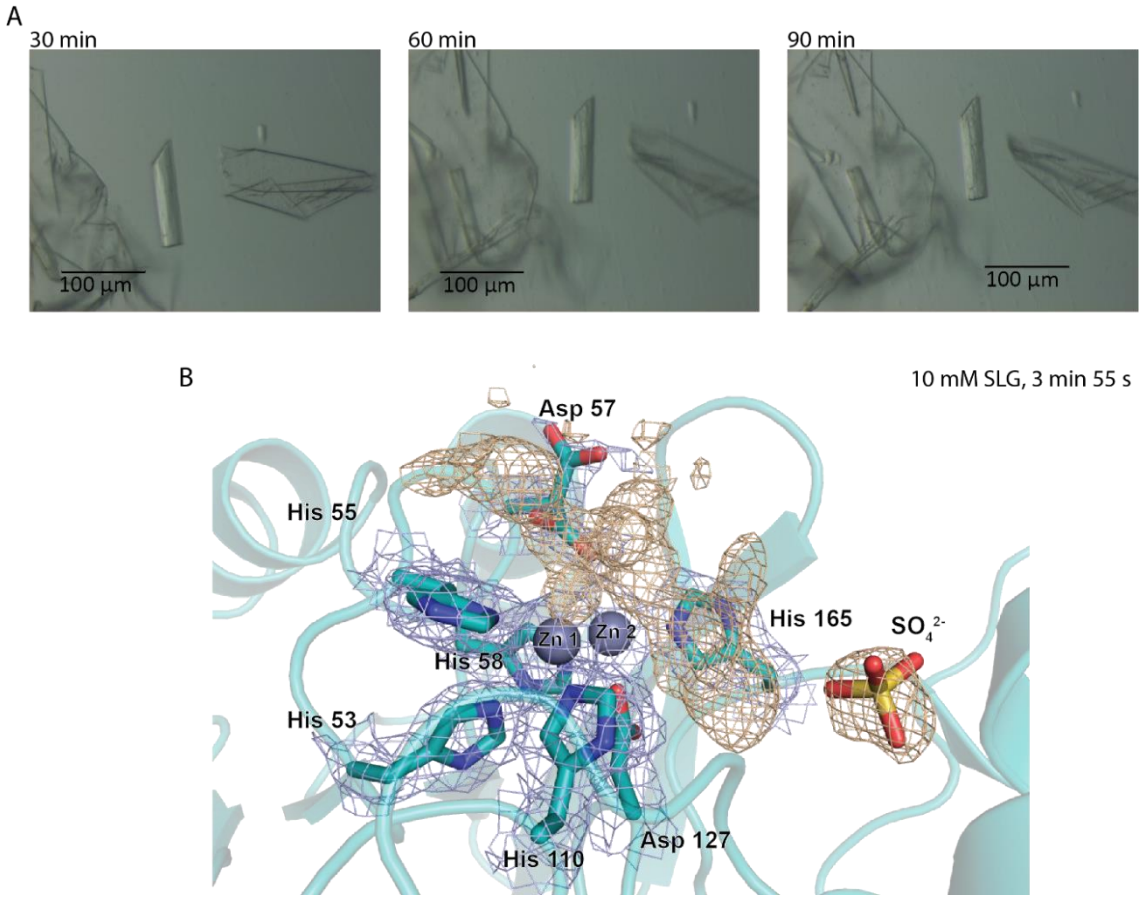


Figure 4.19: Soaking experiments of GloB with SLG. (A) GloB crystal after 30, 60 and 90 min of soaking in no-tartrate condition. (B) Active site of GloB with electron density map (2mFo-DFc, 1 σ, blue) of the active site residues and Polder map (3 σ, orange) showing electron density near the active site with a sulfate group (likely from HEPES buffer in the crystallisation condition) following soaking experiments. Crystals were obtained by soaking crystal (tartrate-bound) in condition lacking tartrate for 90 min, then SLG was added to 10 mM. Crystals were harvested and plunged in liquid N_2 after 3 min 55 s. Diffraction data were collected at the I03 beamline, Diamond Light Source.

Plate-shaped crystals were also obtained during crystallisation trials. They diffracted to 2.51 Å resolution in space group $P2_12_12_1$, but with unit cell parameters different to those observed in both high resolution and tartrate-bound crystals (data collection statistics are given in **Table 4.7**). In the new crystal form (crystal form 3, CF3) a symmetry related molecule was found to be located near the active site, as in the high-resolution GloB structure, but His181 or other residues were not found to be directly interacting with the active site (**Figure 4.20**). Due to the low resolution of the dataset and tartrate-bound structure having a more accessible active site in the crystalline state, no further studies were pursued using CF3.

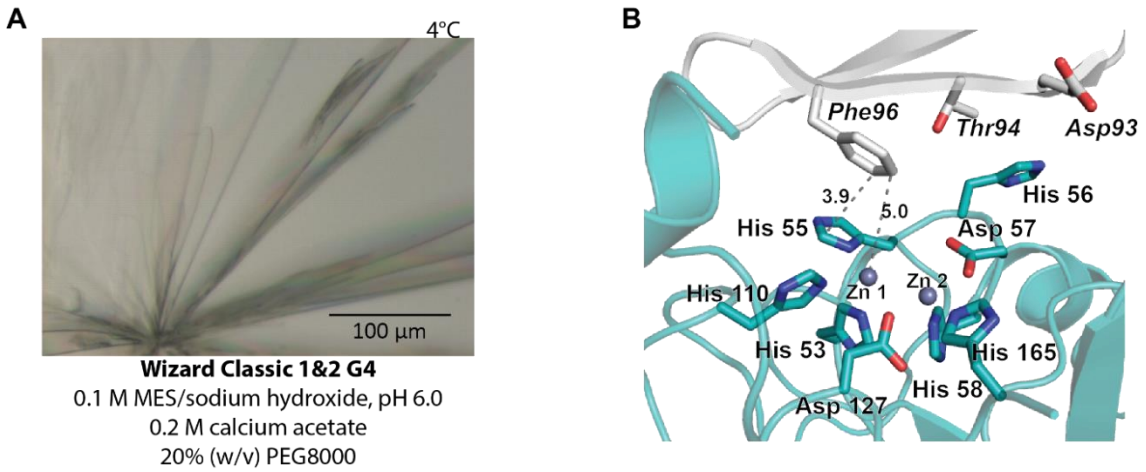


Figure 4.20: The third GloB crystal form (CF3) obtained during crystallisation broad screening experiments. (A) CF3 Crystals were obtained at 4 °C. (B) A view from the crystal structure of CF3 showing the active site and the position of the symmetry related molecule (white) near the active site with distances in Å.

Table 4.7: Data processing statistics for GloB crystal soaked with SLG (GloB-S) and CF3 crystal form of GloB obtained during crystallisation trials (GloB-CF3). *Values for the highest resolution shell are given in parentheses.

	GloB-S	**GloB-CF3**
Source	Diamond Light Source, I03 Beamline	Diamond Light Source, I24 Beamline
Detector	Eiger2 XE 16M	Pilatus3 6M
Data processing	AutoPROC+STARANISO (XDS)	Xia2-DIALS
Wavelength (Å)	0.97933	0.96877
Resolution (Å)	42.23 - 2.0 (2.072 - 2.0)*	56.32 - 2.51 (2.6 - 2.51)*
Space group	$P2_12_12_1$	$P2_12_12_1$
Unit cell parameters		
a, b, c (Å)	39.2, 43.8, 158.6	39.2, 60.5, 112.6
α, β, γ	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Molecules in ASU	1	1
Unique reflections	19080 (1725)*	9680 (956)*
R_{merge}	0.07384 (0.3615)*	0.1041 (0.3544)*
Multiplicity	2.0 (2.0)*	2.0 (2.0)*
CC1/2	0.992 (0.706)*	0.974 (0.553)*
Completeness (%)	98.91 (90.84)*	99.98 (100.00)*
I/σ(I)	7.21 (2.12)*	5.09 (2.04)*

4.6.4 GloB structure within the Glx2 family

To date, Glx2 structures from four organisms (*S. typhimurium*, *A. thaliana*, *Homo sapiens*, and *L. infantum*) have been reported and are available in the PDB[54, 55, 78, 103]. GloB structures were compared to other Glx2 enzymes with known structures. All the structures show the same fold with an MBL domain and a smaller *C*-terminal domain specific to the Glx2 family of enzymes (**Figure 4.21**). Differences between the structures were small with sequence identities ranging from 78% (*S. typhimurium* GloB) to 22% (*L. infantum* Glx2) and RMSD ranging from 0.60 Å to 2.81 Å for Cα atoms (**Table 4.8**). The active sites of Glx2 enzymes were found to be highly conserved with RMSD values falling below 0.21 Å for all active site residues in the structures analysed. The largest difference was observed between *E. coli* GloB and *L. infantum* Glx2. This could potentially be explained by selectivity of *L. infantum* Glx2 for trypanothione-derived substrate instead of a glutathione-derived substrate that the remaining proteins utilise.

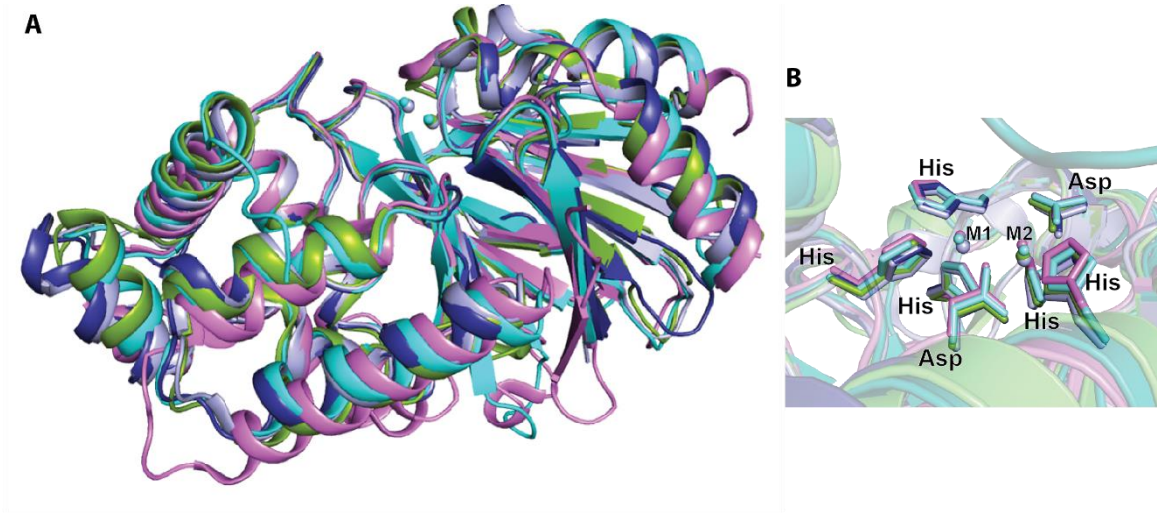


Figure 4.21: Overlay of Glx2 proteins with known structures. (A) Full-length proteins, and (B) active sites. *E. coli* GloB (dark blue), *S. typhimurium* (pale purple, PDB ID: 2QED), *A. thaliana* (green, PDB ID: 1XM8), human (cyan, PDB ID: 1QH5), and *L. infantum* (magenta, PDB ID: 2P18) proteins were aligned using PyMOL 'super' command. Sequence identity and RMSD values are given in **Table 4.8**.

Table 4.8: Comparison of *E. coli* GloB with other Glx2 enzymes with known structures. RMSD values for backbone Cα atoms are given when aligned to the high-resolution GloB structure. RMSD values were compared using 'super' command in PyMOL. TLA – L-tartaric acid. *Numbers in parentheses show number of atoms aligned.

Protein	PDB ID	Sequence similarity	RMSD (Å)	
			Full protein	Active site
E. coli GloB (high res.)	6RZ0	-	-	-
E. coli GloB (TLA-bound)	6S0I	100%	0.52 (253)*	0.21 (38)*
S. typhimurium GloB [78]	2QED	78%	0.60 (252)*	0.09 (38)*
A. thaliana mitochondrial Glx2 [103]	1XM8	30%	1.51(247)*	0.17 (38)*
H. sapiens Glx2 [54]	1QH5	28%	1.40 (244)*	0.20 (38)*
L. infantum Glx2 [55]	2P18	22%	2.81 (239)*	0.18 (28)*

4.7 Summary and conclusions

Recombinant GloB was produced and purified from *E. coli* cells by anion exchange and size exclusion chromatographic steps. The purified GloB was found to be in a monomeric state and to contain a mixture of metal ions bound, with iron and zinc being the most abundant. Non-denaturing MS analysis has shown that GloB is primarily a monomer in solution with a small dimeric population being observed. This observation is consistent with the previous literature reports of other Glx2 enzymes which were also found to be monomers. The resolution of the non-denaturing MS experiments was insufficient to determine the different metal bound species.

Protein folding and thermal stability of the purified GloB were investigated using circular dichroism spectroscopy. GloB was found to be folded with a melting temperature of 62.3 ± 1.9 °C. Protein treated with DPA appeared to adopt a similar fold, but thermal stability was reduced. While, as shown by non-denaturing MS experiments, removal of bound metal ions was incomplete and generated approximately 1:1 mixture of apo and one-metal-bound species, removal of part of the metals had an obvious effect on thermal stability.

Purified GloB was subjected to kinetic studies and found to be active with activity comparable to the literature reports for *E. coli* GloB[84]. As literature studies were performed using enzyme demonstrated to only contain Zn ions, it can be concluded that presence of other transition metals in the active site, does not appear to reduce GloB enzymatic activity. Initial metal reconstitution experiments have shown that incubating 'DPA-treated' GloB with Mn^{2+} for 30 min restored its enzymatic activity almost to 100%, while incubating with Zn^{2+} did not restore

the enzymatic activity (**Figure 4.10**). Interestingly, incubation with Fe^{3+} resulted in partial re-activation of GloB activity.

The inhibition studies of GloB did not yield inhibitors with low IC_{50} values. The best inhibitor found was a squaramide with a low μM IC_{50} value. Both thioester hydrolysis and thiol detection assays gave comparable results with the thiol detection assay resulting in smaller errors being observed (**Figure 4.12**, **Table 4.2**). Overall, both assays were found to have their limitations. The thioester hydrolysis assay was better suited for the use with water soluble compounds, as DMSO absorbs UV light in a similar region of the spectrum increasing the background signal and noise. Thiol detection assays were not affected by the presence of DMSO, but due to the nature of the assay cannot be used for compounds containing thiol groups. These observations should be taken into account in the further search for GloB inhibitors to minimise errors and eliminate false negative results of thiol-containing inhibitors.

Crystallisation trials of GloB resulted in multiple crystals. The best crystals diffracted to 1.1 Å resolution, the highest resolution reported for Glx2 proteins so far. The GloB crystal structure was solved by molecular replacement, refined and deposited in the PDB (PDB ID: 6RZ0). The crystal form producing the well diffracting crystals was found to be unsuitable for soaking experiments due to a reproducibility issue and potentially inaccessible active site in the crystal lattice caused by an interaction between His56 near the active site and His181 from a symmetry related molecule via a bridging metal ion (**Figure 4.16C**).

An additional crystal form of GloB with a more accessible active site was obtained that contained a tartrate molecule bound in the active site (deposited in the PDB, PDB ID: 6S0I). While tartrate binding is unlikely to be physiologically relevant due to high concentrations required for inhibition, the binding mode observed may be exploited in the design of better inhibitors. Crystal soaking experiments were performed in an attempt to obtain a substrate/product bound structure. These experiments were successful at removing tartrate, but residual density surrounding the active site was insufficient to identify the binding of ligand. However, this shows that tartrate can be removed from the active site making this crystal form more suitable for future soaking experiments.

In addition to differences in crystal packing, the two GloB structures obtained differed slightly in the position of Asp57 and one of the Zn ions at the active site as well as the two *C*-terminal helices. Despite varying sequence similarities with Glx2 enzymes from other organisms, the *E. coli* GloB structure was found to exhibit same fold with the MBL and the Glx2 *C*-terminal domains (**Figure 4.21**, **Table 4.8**). The GloB structure differs the most from *L. infantum* Glx2

enzyme, likely due to the latter requiring a trypanothione rather than GSH derived substrate. The active site geometry of all Glx2 enzymes with reported 3-D structures appears to be highly conserved.

Overall, the results of experiments described in this chapter could be further explored in the search for selective inhibitors to Glx2/GloB proteins and understanding the underlying enzymatic reaction mechanism (proposed reaction mechanism is outlined in **Figure 4.22**).

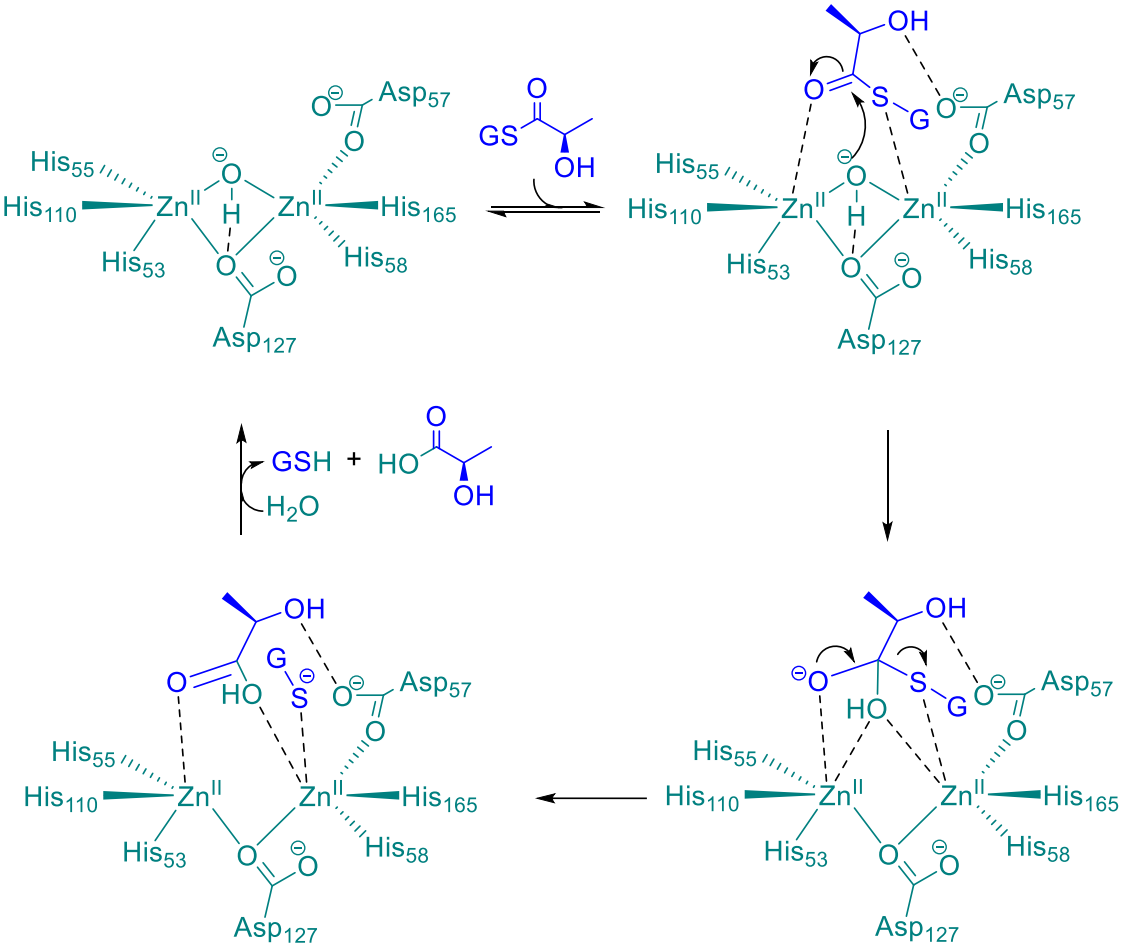


Figure 4.22: Proposed SLG hydrolysis reaction mechanism catalysed by GloB. Residue numbers correspond to *E. coli* GloB.

4.8 References

1. Thornalley, P. J., Pharmacology of methylglyoxal: formation, modification of proteins and nucleic acids, and enzymatic detoxification - a role in pathogenesis and antiproliferative chemotherapy. *Gen. Pharmacol.* **1996,** *27* (4), 565-573.

2. Sousa Silva, M.; Gomes, R. A.; Ferreira, A. E.; Ponces Freire, A.; Cordeiro, C., The glyoxalase pathway: the first hundred years... and beyond. *Biochem. J* **2013,** *453* (1), 1-15.

3. Richard, J. P., Acid-base catalysis of the elimination and isomerization reactions of triose phosphates. *J. Am. Chem. Soc.* **1984,** *106* (17), 4926-2936.

4. Richard, J. P., Mechanism for the formation of methylglyoxal from triosephosphates. *Biochem. Soc. Trans.* **1993,** *21* (2), 549-553.

5. Richard, J. P., Kinetic parameters for the elimination reaction catalyzed by triosephosphate isomerase and an estimation of the reaction's physiological significance. *Biochemistry* **1991,** *30* (18), 4581-4585.

6. Cooper, R. A.; Anderson, A., The formation and catabolism of methylglyoxal during glycolysis in *Escherichia coli*. *FEBS Lett.* **1970,** *11* (4), 273-276.

7. Hopper, D. J.; Cooper, R. A., The purification and properties of *Escherichia coli* methylglyoxal synthase. *Biochem. J* **1972,** *128* (2), 321-329.

8. Hopper, D. J.; Cooper, R. A., The regulation of *Escherichia coli* methylglyoxal synthase; a new control site in glycolysis? *FEBS Lett.* **1971,** *13* (4), 213-216.

9. Cooper, R. A., Metabolism of methylglyoxal in microorganisms. *Annu. Rev. Microbiol.* **1984,** *38*, 49-68.

10. Ray, S.; Ray, M., Isolation of methylglyoxal synthase from goat liver. *J. Biol. Chem.* **1981,** *256* (12), 6230-6233.

11. Phillips, S. A.; Thornalley, P. J., The formation of methylglyoxal from triose phosphates. Investigation using a specific assay for methylglyoxal. *Eur. J. Biochem.* **1993,** *212* (1), 101-105.

12. Sato, J.; Wang, Y.-m.; van Eys, J., Methylglyoxal Formation in Rat Liver Cells. *J. Biol. Chem.* **1980,** *255* (5), 2046-2050.

13. Casazza, J. P.; Felver, M. E.; Veech, R. L., The metabolism of acetone in rat. *J. Biol. Chem.* **1984,** *259* (1), 231-236.

14. Koop, D. R.; Casazza, J. P., Identification of ethanol-inducible P-450 isozyme 3a as the acetone and acetol monooxygenase of rabbit microsomes. *J. Biol. Chem.* **1985,** *260* (25), 13607-13612.

15. Lyles, G. A.; Chalmers, J., The metabolism of aminoacetone to methylglyoxal by semicarbazide-sensitive amine oxidase in human umbilical artery. *Biochem. Pharmacol.* **1992,** *43* (7), 1409-1414.

16. Aleksandrovskii, Y. A., Antithrombin III, C1 inhibitor, methylglyoxal, and polymorphonuclear leukocytes in the development of vascular complications in diabetes mellitus. *Thromb. Res.* **1992,** *67* (2), 179-189.

17. Thornalley, P. J.; Langborg, A.; Minhas, H. S., Formation of glyoxal, methylglyoxal and 3-deoxyglucosone in the glycation of proteins by glucose. *Biochem. J* **1999,** *344* (Pt 1), 109-116.

18. Esterbauer, H.; Cheeseman, K. H.; Dianzani, M. U.; Poli, G.; Slater, T. F., Separation and characterization of the aldehydic products of lipid peroxidation stimulated by ADP-Fe^{2+} in rat liver microsomes. *Biochem. J* **1982,** *208* (1), 129-140.

19. Singh, R.; Barden, A.; Mori, T.; Beilin, L., Advanced glycation end-products: a review. *Diabetologia* **2001,** *44* (2), 129-146.

20. Thornalley, P. J., Dicarbonyl intermediates in the Maillard reaction. *Ann. N.Y. Acad. Sci.* **2005,** *1043*, 111-117.

21. Thornalley, P. J., Protein and nucleotide damage by glyoxal and methylglyoxal in physiological systems - role in ageing and disease. *Drug Metabol. Drug Interact.* **2008,** *23* (1-2), 125-150.

22. Wetzels, S.; Wouters, K.; Schalkwijk, C. G.; Vanmierlo, T.; Hendriks, J. J. A., Methylglyoxal-Derived Advanced Glycation Endproducts in Multiple Sclerosis. *Int. J. Mol. Sci.* **2017,** *18* (2), 421.

23. Cantero, A. V.; Portero-Otin, M.; Ayala, V.; Auge, N.; Sanson, M.; Elbaz, M.; Thiers, J. C.; Pamplona, R.; Salvayre, R.; Negre-Salvayre, A., Methylglyoxal induces advanced glycation end product (AGEs) formation and dysfunction of PDGF receptor-β: implications for diabetic atherosclerosis. *FASEB J.* **2007,** *21* (12), 3096-3106.

24. Li, J.; Liu, D.; Sun, L.; Lu, Y.; Zhang, Z., Advanced glycation end products and neurodegenerative diseases: mechanisms and perspective. *J. Neurol. Sci.* **2012,** *317* (1-2), 1-5.

25. Kalapos, M. P., Methylglyoxal toxicity in mammals. *Toxicol. Lett.* **1994,** *73* (1), 3-24.

26. Kasai, H.; Kumeno, K.; Yamaizumi, Z.; Nishimura, S.; Nagao, M.; Fujita, Y.; Sugimura, T.; Nukaya, H.; Kosuge, T., Mutagenicity of methylglyoxal in coffee. *Gan* **1982,** *73* (5), 681-683.

27. Bronzetti, G.; Corsi, C.; Del Chiaro, D.; Boccardo, P.; Vellosi, R.; Rossi, F.; Paolini, M.; Cantelli Forti, G., Methylglyoxal: genotoxicity studies and its effect in vivo on the hepatic microsomal mono-oxygenase system of the mouse. *Mutagenesis* **1987,** *2* (4), 275-277.

28. Murata-Kamiya, N.; Kamiya, H.; Kaji, H.; Kasai, H., Methylglyoxal induces G:C to C:G and G:C to T:A transversions in the *supF* gene on a shuttle vector plasmid replicated in mammalian cells. *Mutat. Res.* **2000,** *468* (2), 173-182.

29. Murata-Kamiya, N.; Kaji, H.; Kasai, H., Deficient nucleotide excision repair increases base-pair substitutions but decreases TGGC frameshifts induced by methylglyoxal in *Escherichia coli*. *Mutat. Res.* **1999,** *442* (1), 19-28.

30. Murata-Kamiya, N.; Kamiya, H., Methylglyoxal, an endogenous aldehyde, crosslinks DNA polymerase and the substrate DNA. *Nucleic Acids Res.* **2001,** *29* (16), 3433-3438.

31. Rabbani, N.; Thornalley, P. J., Methylglyoxal, glyoxalase 1 and the dicarbonyl proteome. *Amino Acids* **2012,** *42* (4), 1133-1142.

32. Nevin, C.; McNeil, L.; Ahmed, N.; Murgatroyd, C.; Brison, D.; Carroll, M., Investigating the Glycating Effects of Glucose, Glyoxal and Methylglyoxal on Human Sperm. *Sci. Rep.* **2018,** *8* (1), 1-12.

33. Guimaraes, E. L.; Empsen, C.; Geerts, A.; van Grunsven, L. A., Advanced glycation end products induce production of reactive oxygen species via the activation of NADPH oxidase in murine hepatic stellate cells. *J. Hepatol.* **2010,** *52* (3), 389-397.

34. Kim, W.; Hudson, B. I.; Moser, B.; Guo, J.; Rong, L. L.; Lu, Y.; Qu, W.; Lalla, E.; Lerner, S.; Chen, Y.; Yan, S. S.; D'Agati, V.; Naka, Y.; Ramasamy, R.; Herold, K.; Yan, S. F.; Schmidt, A. M., Receptor for advanced glycation end products and its ligands: a journey from the complications of diabetes to its pathogenesis. *Ann. N.Y. Acad. Sci.* **2005,** *1043*, 553-561.

35. Gomes, R. A.; Vicente Miranda, H.; Sousa Silva, M.; Graca, G.; Coelho, A. V.; do Nascimento Ferreira, A. E.; Cordeiro, C.; Freire, A. P., Protein glycation and methylglyoxal metabolism in yeast: finding peptide needles in protein haystacks. *FEMS Yeast Res.* **2008,** *8* (1), 174-181.

36. Puttaiah, S.; Biswas, A.; Staniszewska, M.; Nagaraj, R. H., Methylglyoxal inhibits glycation-mediated loss in chaperone function and synthesis of pentosidine in α-crystallin. *Exp. Eye Res.* **2007,** *84* (5), 914-921.

37. Padival, A. K.; Crabb, J. W.; Nagaraj, R. H., Methylglyoxal modifies heat shock protein 27 in glomerular mesangial cells. *FEBS Lett.* **2003,** *551* (1-3), 113-118.

38. Shapiro, R.; Cohen, B. I.; Shiuey, S. J.; Maurer, H., On the reaction of guanine with glyoxal, pyruvaldehyde, and kethoxal, and the structure of the acylguanines. A new synthesis of N2-alkylguanines. *Biochemistry* **1969,** *8* (1), 238-245.

39. Hopkins, F. G.; Morgan, E. J., On the distribution of glyoxalase and glutathione. *Biochem. J* **1945,** *39* (4), 320-324.

40. Thornalley, P. J., The glyoxalase system: new developments towards functional characterization of a metabolic pathway fundamental to biological life. *Biochem. J* **1990,** *269* (1), 1-11.

41. Freedberg, W. B.; Kistler, W. S.; Lin, E. C. C., Lethal Synthesis of Methylglyoxal by *Escherichia coli* During Unregulated Glycerol Metabolism. *J. Bacteriol.* **1971,** *108* (1), 137-144.

42. Misra, K.; Banerjee, A. B.; Ray, S.; Ray, M., Glyoxalase III from *Escherichia coli*: a single novel enzyme for the conversion of methylglyoxal into D-lactate without reduced glutathione. *Biochem. J* **1995,** *305 (Pt 3)*, 999-1003.

43. Lee, J. Y.; Song, J.; Kwon, K.; Jang, S.; Kim, C.; Baek, K.; Kim, J.; Park, C., Human DJ-1 and its homologs are novel glyoxalases. *Hum. Mol. Genet.* **2012,** *21* (14), 3215-3225.

44. Subedi, K. P.; Choi, D.; Kim, I.; Min, B.; Park, C., Hsp31 of *Escherichia coli* K-12 is glyoxalase III. *Mol. Microbiol.* **2011,** *81* (4), 926-936.

45. Sousa Silva, M.; Ferreira, A. E.; Gomes, R.; Tomas, A. M.; Ponces Freire, A.; Cordeiro, C., The glyoxalase pathway in protozoan parasites. *Int. J. Med. Microbiol.* **2012,** *302* (4-5), 225-229.

46. Cameron, A. D.; Olin, B.; Ridderström, M.; Mannervik, B.; Jones, T. A., Crystal structure of human glyoxalase I - evidence for gene duplication and 3D domain swapping. *EMBO J.* **1997,** *16* (12), 3386-3395.

47. Ariza, A.; Vickers, T. J.; Greig, N.; Armour, K. A.; Dixon, M. J.; Eggleston, I. M.; Fairlamb, A. H.; Bond, C. S., Specificity of the trypanothione-dependent *Leishmania major* glyoxalase I: structure and biochemical comparison with the human enzyme. *Mol. Microbiol.* **2006,** *59* (4), 1239-1248.

48. He, M. M.; Clugston, S. L.; Honek, J. F.; Matthews, B. W., Determination of the structure of *Escherichia coli* glyoxalase I suggests a structural basis for differential metal activation. *Biochemistry* **2000,** *39* (30), 8719-8727.

49. Marmstal, E.; Aronsson, A. C.; Mannervik, B., Comparison of glyoxalase I purified from yeast (*Saccharomyces cerevisiae*) with the enzyme from mammalian sources. *Biochem. J* **1979,** *183* (1), 23-30.

50. Iozef, R.; Rahlfs, S.; Chang, T.; Schirmer, H.; Becker, K., Glyoxalase I of the malarial parasite *Plasmodium falciparum*: evidence for subunit fusion. *FEBS Lett.* **2003,** *554* (3), 284-288.

51. Jagt, D. L. V.; Han, L.-P. B.; Lehman, C. H., Kinetic evaluation of substrate specificity in the glyoxalase-I-catalyzed disproportionation of α-ketoaldehydes. *Biochemistry* **1972,** *11* (20), 3735-3740.

52. Maiti, M. K.; Krishnasamy, S.; Owen, H. A.; Makaroff, C. A., Molecular characterization of glyoxalase II from *Arabidopsis thaliana*. *Plant Mol. Biol.* **1997,** *35* (4), 471-481.

53. Bito, A.; Haider, M.; Hadler, I.; Breitenbach, M., Identification and phenotypic analysis of two glyoxalase II encoding genes from *Saccharomyces cerevisiae*, GLO2 and GLO4, and

intracellular localization of the corresponding proteins. *J. Biol. Chem.* **1997,** *272* (34), 21509-21519.

54. Cameron, A. D.; Ridderstrom, M.; Olin, B.; Mannervik, B., Crystal structure of human glyoxalase II and its complex with a glutathione thiolester substrate analogue. *Structure* **1999,** *7* (9), 1067-1078.

55. Sousa Silva, M.; Ferreira, E. N.; Roma, S.; Toma, A. M.; Ponces Freire, A.; Cordeiro, C., Catalysis and Structural Properties of *Leishmania infantum* Glyoxalase II: trypanothione specificity and phylogeny. *Biochemistry* **2008,** *47* (1), 195-204.

56. Clelland, J. D.; Thornalley, P. J., The potentiation of GTP-dependent assenbly of microtubules by *S*-D-lactoylglutathione. *Biochem. Soc. Trans.* **1993,** *21* (2), 160S.

57. Armeni, T.; Cianfruglia, L.; Piva, F.; Urbanelli, L.; Luisa Caniglia, M.; Pugnaloni, A.; Principato, G., *S*-D-Lactoylglutathione can be an alternative supply of mitochondrial glutathione. *Free Radic. Biol. Med.* **2014,** *67*, 451-459.

58. Edwards, L.; Clelland, J. D.; Thornalley, P. J., Characteristics of the inhibition of human promyelocytic leukaemia HL60 cell growth by *S*-D-lactoylglutathione in vitro. *Leuk. Res.* **1993,** *17* (4), 305-310.

59. Clelland, J. D.; Allen, R. E.; Thornalley, P. J., Inhibition of growth of human leukaemia 60 cells by *S*-2-hydroxyacylglutathiones and monoethyl ester derivatives. *Biochem. Pharmacol.* **1992,** *44* (10), 1953-1959.

60. Allen, R. E.; Thornalley, P. J., The effect of *S*-D-lactoylglutathione on the movement of neutrophils. *Biochem. Soc. Trans.* **1993,** *21* (2), 161S.

61. Shendelman, S.; Jonason, A.; Martinat, C.; Leete, T.; Abeliovich, A., DJ-1 is a redox-dependent molecular chaperone that inhibits α-synuclein aggregate formation. *PLoS Biol.* **2004,** *2* (11), e362.

62. Richarme, G.; Dairou, J., Parkinsonism-associated protein DJ-1 is a bona fide deglycase. *Biochem. Biophys. Res. Commun.* **2017,** *483* (1), 387-391.

63. Lee, S. J.; Kim, S. J.; Kim, I. K.; Ko, J.; Jeong, C. S.; Kim, G. H.; Park, C.; Kang, S. O.; Suh, P. G.; Lee, H. S.; Cha, S. S., Crystal structures of human DJ-1 and *Escherichia coli* Hsp31, which share an evolutionarily conserved domain. *J. Biol. Chem.* **2003,** *278* (45), 44552-44559.

64. Zhao, Y.; Liu, D.; Kaluarachchi, W. D.; Bellamy, H. D.; White, M. A.; Fox, R. O., The crystal structure of *Escherichia coli* heat shock protein YedU reveals three potential catalytic active sites. *Protein Sci.* **2003,** *12* (10), 2303-2311.

65. Antognelli, C.; Palumbo, I.; Aristei, C.; Talesa, V. N., Glyoxalase I inhibition induces apoptosis in irradiated MCF-7 cells via a novel mechanism involving Hsp27, p53 and NF-κB. *Br. J. Cancer* **2014,** *111* (2), 395-406.

66. Thornalley, P. J.; Rabbani, N., Glyoxalase in tumourigenesis and multidrug resistance. *Semin. Cell Dev. Biol.* **2011,** *22* (3), 318-325.

67. Taniguchi, H.; Horinaka, M.; Yoshida, T.; Yano, K.; Goda, A. E.; Yasuda, S.; Wakada, M.; Sakai, T., Targeting the glyoxalase pathway enhances TRAIL efficacy in cancer cells by downregulating the expression of antiapoptotic molecules. *Mol. Cancer Ther.* **2012,** *11* (10), 2294-2300.

68. Antognelli, C.; Ferri, I.; Bellezza, G.; Siccu, P.; Love, H. D.; Talesa, V. N.; Sidoni, A., Glyoxalase 2 drives tumorigenesis in human prostate cells in a mechanism involving androgen receptor and p53-p21 axis. *Mol. Carcinog.* **2017,** *56* (9), 2112-2126.

69. Talesa, V. N.; Ferri, I.; Bellezza, G.; Love, H. D.; Sidoni, A.; Antognelli, C., Glyoxalase 2 Is Involved in Human Prostate Cancer Progression as Part of a Mechanism Driven By PTEN/PI3K/AKT/mTOR Signaling With Involvement of PKM2 and ERα. *Prostate* **2017,** *77* (2), 196-210.

70. Rabbani, N.; Thornalley, P. J., Dicarbonyl stress in cell and tissue dysfunction contributing to ageing and disease. *Biochem. Biophys. Res. Commun.* **2015,** *458* (2), 221-226.

71. Xue, M.; Weickert, M. O.; Qureshi, S.; Kandala, N. B.; Anwar, A.; Waldron, M.; Shafie, A.; Messenger, D.; Fowler, M.; Jenkins, G.; Rabbani, N.; Thornalley, P. J., Improved Glycemic Control and Vascular Function in Overweight and Obese Subjects by Glyoxalase 1 Inducer Formulation. *Diabetes* **2016,** *65* (8), 2282-2294.

72. Vickers, T. J.; Greig, N.; Fairlamb, A. H., A trypanothione-dependent glyoxalase I with a prokaryotic ancestry in *Leishmania major*. *Proc. Natl. Acad. Sci. U. S. A.* **2004,** *101* (36), 13186-13191.

73. Padmanabhan, P.; Mukherjee, A.; Madhubala, R., Characterization of the gene encoding glyoxalase II from *Leishmania donovani*: a potential target for anti-parasite drugs. *Biochem. J* **2006,** *393* (1), 227-234.

74. Wyllie, S.; Fairlamb, A. H., Methylglyoxal metabolism in trypanosomes and leishmania. *Semin. Cell Dev. Biol.* **2011,** *22* (3), 271-277.

75. Ozyamak, E.; Black, S. S.; Walker, C. A.; Maclean, M. J.; Bartlett, W.; Miller, S.; Booth, I. R., The critical role of *S*-lactoylglutathione formation during methylglyoxal detoxification in *Escherichia coli*. *Mol. Microbiol.* **2010,** *78* (6), 1577-1590.

76. MacLean, M. J.; Ness, L. S.; Ferguson, G. P.; Booth, I. R., The role of glyoxalase I in the detoxification of methylglyoxal and in the activation of the KefB K+ efflux system in *Escherichia coli*. *Mol. Microbiol.* **1998,** *27* (3), 563-571.

77. O'Young, J.; Sukdeo, N.; Honek, J. F., *Escherichia coli* glyoxalase II is a binuclear zinc-dependent metalloenzyme. *Arch. Biochem. Biophys.* **2007,** *459* (1), 20-26.

78. Campos-Bermudez, V. A.; Leite, N. R.; Krog, R.; Costa-Filho, A. J.; Soncini, F. C.; Oliva, G.; Vila, A. J., Biochemical and structural characterization of *Salmonella typhimurium* glyoxalase II: New insights into metal ion selectivity. *Biochemistry* **2007,** *46* (39), 11069-11079.

79. Rai, S.; Yadav, S.; Rai, R.; Chatterjee, A.; Singh, S.; Rai, L. C., Molecular and biochemical characterization of All0580 as a methylglyoxal detoxifying glyoxalase II of *Anabaena* sp. PCC7120 that confers abiotic stress tolerance in *E. coli*. *Int. J. Biol. Macromol.* **2019,** *124*, 981-993.

80. Limphong, P.; McKinney, R. M.; Adams, N. E.; Bennett, B.; Makaroff, C. A.; Gunasekera, T.; Crowder, M. W., Human glyoxalase II contains an Fe(II)Zn(II) center but is active as a mononuclear Zn(II) enzyme. *Biochemistry* **2009,** *48* (23), 5426-5434.

81. Baier, F.; Chen, J.; Solomonson, M.; Strynadka, N. C. J.; Tokuriki, N., Distinct Metal Isoforms Underlie Promiscuous Activity Profiles of Metalloenzymes. *ACS Chem. Biol.* **2015**, 1684-1693.

82. Ratledge, C.; Dover, L. G., Iron metabolism in pathogenic bacteria. *Annu. Rev. Microbiol.* **2000,** *54*, 881-941.

83. Messenger, A. J. M.; Barclay, R., Bacteria, iron and pathogenicity. *Biochem. Educ.* **1983,** *11* (2), 54-63.

84. O'Young, J.; Sukdeo, N.; Honek, J. F., *Escherichia coli* glyoxalase II is a binuclear zinc-dependent metalloenzyme. *Arch. Biochem. Biophys.* **2007,** *459* (1), 20-26.

85. Cahill, S. T. Structural and Kinetic Studies on β-Lactamase Mechanism and Inhibition. DPhil Thesis, University of Oxford, 2017.

86. *OriginPro*, 2017; OriginLab Corporation: Northampton, Massachusetts, USA.

87. Ridderstrom, M.; Saccucci, F.; Hellman, U.; Bergman, T.; Principato, G.; Mannervik, B., Molecular cloning, heterologous expression, and characterization of human glyoxalase II. *J. Biol. Chem.* **1996,** *271* (1), 319-323.

88. Software, G. *GraphPad Prism*, v5.04; GraphPad Software: San Diego, California, USA.

89. Wetterholm, A.; Macchia, L.; Haeggstrom, J. Z., Zinc and other divalent cations inhibit purified leukotriene A_4 hydrolase and leukotriene B_4 biosynthesis in human polymorphonuclear leukocytes. *Arch. Biochem. Biophys.* **1994,** *311* (2), 263-271.

90. Jackman, J. E.; Raetz, C. R.; Fierke, C. A., UDP-3-*O*-(R-3-hydroxymyristoyl)-*N*-acetylglucosamine deacetylase of *Escherichia coli* is a zinc metalloenzyme. *Biochemistry* **1999,** *38* (6), 1902-1911.

91. Wells, T. N.; Coulin, F.; Payton, M. A.; Proudfoot, A. E., Phosphomannose isomerase from *Saccharomyces cerevisiae* contains two inhibitory metal ion binding sites. *Biochemistry* **1993,** *32* (5), 1294-1301.

92. Na, H. W.; Namgung, B.; Song, W. S.; Yoon, S. I., Structural and biochemical analyses of the metallo-β-lactamase fold protein YhfI from *Bacillus subtilis*. *Biochem. Biophys. Res. Commun.* **2019,** *519* (1), 35-40.

93. Cahill, S. T.; Tarhonskaya, H.; Rydzik, A. M.; Flashman, E.; McDonough, M. A.; Schofield, C. J.; Brem, J., Use of ferrous iron by metallo-β-lactamases. *J. Inorg. Biochem.* **2016,** *163*, 185-193.

94. Gillis, G. S. Nucleotide Inhibition of Glyoxalase II. PhD Thesis, University of North Texas, 1999.

95. Campos-Bermudez, V. A.; Moran-Barrio, J.; Costa-Filho, A. J.; Vila, A. J., Metal-dependent inhibition of glyoxalase II: a possible mechanism to regulate the enzyme activity. *J. Inorg. Biochem.* **2010,** *104* (7), 726-731.

96. Badarau, A.; Page, M. I., The Variation of Catalytic Efficiency of *Bacillus cereus* Metallo-β-lactamase with Different Active Site Metal Ions. *Biochemistry* **2006,** *45*, 10654-10666.

97. Taniguchi, M.; Lindsey, J. S., Database of Absorption and Fluorescence Spectra of >300 Common Compounds for use in PhotochemCAD. *Photochem. Photobiol.* **2018,** *84*, 290-327.

98. Boivin, S.; Kozak, S.; Meijers, R., Optimization of protein purification and characterization using Thermofluor screens. *Protein Expr. Purif.* **2013,** *91* (2), 192-206.

99. McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J., Phaser crystallographic software. *J. Appl. Crystallogr.* **2007,** *40* (4), 658-674.

100. Adams, P. D.; Afonine, P. V.; Bunkóczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L.-W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H., PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **2010,** *66* (Pt 2), 213-221.

101. Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K., Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **2010,** *66* (4), 486-501.

102. Kim, O. B.; Unden, G., The L-Tartrate/Succinate Antiporter TtdT (YgjE) of L-Tartrate Fermentation in *Escherichia coli*. *J. Bacteriol.* **2007,** *189* (5), 1597-1603.

103. Marasinghe, G. P.; Sander, I. M.; Bennett, B.; Periyannan, G.; Yang, K. W.; Makaroff, C. A.; Crowder, M. W., Structural studies on a mitochondrial glyoxalase II. *J. Biol. Chem.* **2005,** *280* (49), 40668-40675.

Chapter 5: Summary and Conclusions

Contents

5.1 Biochemical and Biophysical Characterisation of CMAH 147
5.2 Investigating the Role of CMAH in Living Cells 148
5.3 Biochemical and Structural Studies of *Escherichia coli* Glyoxalase II 149
5.4 References 151

The metallo-β-lactamase (MBL) superfamily has member proteins across all domains of life. [1] Members of the MBL superfamily share a conserved αββα structural fold and an active site able to bind one or two metal ions. The first MBL discovered (*Bacillus cereus* β-lactamase II, BcII) was a bacterial Zn-binding enzyme that catalyses hydrolysis of the β-lactam antibiotics. In addition to β-lactam hydrolysis, members of the MBL superfamily can hydrolyse esters, thioesters, sulfates and phosphodiesters as well as catalyse redox reactions.[1-4]

The work described in this thesis aimed to investigate structural and functional properties of two MBL superfamily proteins – cytidine monophosphate *N*-acetylneuraminic acid hydroxylase (CMAH) and glyoxalase II (GloB). The work addressed previously reported and predicted features of the CMAH and GloB proteins and extended to structural studies.

5.1 Biochemical and Biophysical Characterisation of CMAH

CMAH is a monooxygenase member of the MBL superfamily that catalyses oxidation of cytidine monophosphate *N*-acetylneuraminic acid (CMP-Neu5Ac) to CMP-*N*-glycolylneuraminic acid (CMP-Neu5Gc).[5] Human CMAH is enzymatically inactive due to a deletion event in exon 6 which occurred approximately 3 million years ago, after the last common ancestor with bonobos and chimpanzees.[6, 7]

The work in Chapter 2 details cloning and production of human and mouse CMAH proteins (hCMAH and mCMAH respectively) in bacterial and mammalian expression systems. Crystallisation attempts of recombinant hCMAH produced in *Escherichia coli* Rosetta-gami B(DE3) cells were unsuccessful in producing any diffracting hCMAH crystals. One of the diffracting crystals was found to be *E. coli* dihydrolipoyllysine-residue succinyltransferase component of 2 oxoglutarate dehydrogenase complex (OGDC-E2, **Figure 2.31**). Further characterisation of hCMAH was limited due to low purification yields.

Further work on hCMAH could explore different expression systems, including yeast cells and baculovirus expression in insect cells, in attempts to develop a more robust method for protein production. Further characterisation could then be performed.

Recombinant mCMAH was produced in stably expressing HEK293 cells. A mass spectrometry (MS) based CMP-Neu5Ac hydroxylation assay was developed and purified recombinant mCMAH was shown to be catalytically active (**Figure 2.28**). Purified mCMAH was demonstrated to be monomeric (**Figure 2.20** and **Figure 2.21**) and possess a mononuclear Fe^{3+} ion (**Figure 2.23**). Incorporation of the Rieske [2Fe2S] cluster was found to be inconclusive with UV-Visible spectroscopy analysis being consistent with presence of the Rieske cluster (**Figure 2.22**) while

electron paramagnetic resonance (EPR) analysis exhibited a signal unusual not only to Rieske [2Fe2S] clusters, but also to other iron-sulfur cluster reported in literature (**Figure 2.24**). mCMAH crystallisation trials have yielded a single diffracting crystal. Structure solution is currently ongoing.

Further work is needed to determine the nature of the iron-sulfur cluster of mCMAH, including exploring different EPR methods. In addition, avoiding freezing and thawing mCMAH samples could be beneficial as freeze-thaw cycles have been reported to contribute to protein damage and denaturation[8]. As purified recombinant mCMAH was found to be catalytically active, further studies could focus on understanding the reaction mechanism, substrate and oxygen binding, and electrochemistry. The position of the observed hydroxylation should be confirmed by MS-MS. Potential post-translational modification of mCMAH could be further investigated using MS-based approaches. Additional work is needed to solve the mCMAH structure, either from the dataset obtained, or via different methods. With cryogenic electron microscopy (cryo-EM) improving rapidly, in the near future it may be possible to routinely image proteins of ~ 70 kDa.[9] This could be explored as a potential method for mCMAH structural studies.

5.2 Investigating the Role of CMAH in Living Cells

Human CMAH is enzymatically inactive[6, 7], resulting in little being known about the underlying reasons of high gene expression being observed in various tissues[10]. The work described in Chapter 3 focused on investigating presence, location and interactors of hCMAH and its active mouse homologue (mCMAH) in living cells.

Immunofluorescence (IF) studies have shown that both endogenous and recombinant *C*-His_6-tagged hCMAH proteins are localised in the cytoplasm of HeLa cells with some compartmentalisation being observed (**Figure 3.1**). Co-localisation experiments have shown that hCMAH does not localise to the endoplasmic reticulum or the mitochondria (**Figure 3.2**). However, some large dot-like formations were observed in multiple experiments (**Figure 3.3**), suggesting possible compartmentalisation of hCMAH under specific conditions, such as different stage of the cell cycle or protein aggregation. Recombinant mCMAH was shown to exhibit a similar localisation in TCMK-1 cells to the hCMAH proteins, with some more defined co-localisation being observed. However, IF studies of mCMAH have been inconclusive, as endogenous mCMAH could not be detected in TCMK-1 cells, likely due to a non-specific binding of anti-CMAH antibodies used.

Attempts to study proteins interacting with hCMAH and mCMAH in living cells were made using pulldown and co-immunoprecipitation (co-IP) experiments. Pulldown experiments using recombinant mCMAH in HEK293 cells resulted in an adduct being observed (**Figure 3.6**) in samples treated with formaldehyde to crosslink molecules interacting over a short distance. However, low yields limited further identification of the interactor being observed. No proteins interacting with hCMAH were detected. Co-IP experiments have been shown to be more suitable for studies of hCMAH rather than mCMAH. Interactor studies using the TCMK-1 cell line were unsuccessful since no detectable mCMAH has been recovered after pulldown and co-IP experiments.

Further studies are needed to determine the cause of the dot-like formations observed in hCMAH experiments. The use of synchronised cells could help determine whether the observed clustering is cell cycle dependant. Co-localisation experiments for other organelles, including peroxisomes and lysosomes, could also be performed to further study the cellular localisation. In order to investigate the localisation of mCMAH proteins, a cell line derived from a closely related organism possessing catalytically active CMAH (e.g. rat or hamster) could be used for immunofluorescence studies. This could potentially eliminate the non-specific binding observed in TCMK-1 cell line originally used for studies. Co-IP and pulldown studies could be expanded to detect other potential interacting proteins. A different mouse cell line could be used as an alternative to TCMK-1. Additionally, different methods to study protein-protein interactions, including crosslinking methods coupled with MS, such as BioID[11, 12] or TAP[13] tagging, followed by MS, could be explored.

5.3 Biochemical and Structural Studies of *Escherichia coli* Glyoxalase II

Methylglyoxal is a cytotoxic compound produced as a side product of several cellular processes.[14] The glyoxalase (Glx) system consisting of glyoxalase I and glyoxalase II enzymes, has been reported to be responsible for glutathione (GSH) dependent detoxification of methylglyoxal.[15] Glyoxalase II (Glx2 or GloB) is a member of the MBL superfamily catalysing the hydrolysis of *S*-D-lactoylglutathione to D-lactate and GSH.

The work described in Chapter 4 details purification and characterisation of *E. coli* GloB. Purified recombinant GloB was found to be a thermally stable protein that binds a mixture of metal ions, primarily Fe, Zn, Mn, contrary to the literature reports showing only Zn ions being present.[16] Despite different metal constitution, GloB activity was found to be comparable to that reported in literature[16] (**Figure 4.9**). Metal ion reconstitution experiments using first row

transition metals have shown that incubation with Mn^{2+} was able to almost fully restore activity of GloB treated with 2,6-pyridinedicarboxylic acid, while incubation with Zn^{2+} did not increase the activity of treated GloB (**Figure 4.10**). Interestingly, Fe^{3+} appeared to restore GloB activity to around 55% of the initial activity. Inhibition studies of 'as purified' GloB did not yield any strong inhibitors but resulted in several low to mid μM inhibitors being observed (**Table 4.2**). Squaramide ESC1000274 showed the best inhibition with pIC50 value of 4.80 ± 0.03 or 5.04 ± 0.04 depending on the assay type. Two assay methods – one measuring the hydrolysis of a thioester bond and the other detecting formation of a thiol group of the hydrolysis product – were assessed and advantages and limitations of each discussed.

GloB was demonstrated to be structurally similar to Glx2 of other organisms with active sites being highly conserved across the family. Structural studies resulted in 1.1 Å structure of GloB, the highest resolution of all Glx2 structures currently available. However, the active site was found to be hindered by the symmetry related molecule (**Figure 4.16**). In the search for alternative crystal form, 1.8 Å GloB structure with a tartrate molecule bound at the active site was solved (**Figure 4.18**). Preliminary soaking experiments were performed and resulted in the removal of the tartrate molecule from the active site suggesting suitability of the tartrate-bound crystal form for future studies to obtain structures of enzyme-substrate and/or inhibitor complex. GloB crystal structures were submitted to Protein Data Bank (PDB, accession codes 6RZ0 and 6S0I for high resolution and tartrate-bound structures respectively).

Further experiments should be performed to investigate the effect of different metal ions to GloB activity with studies being expanded to Fe^{2+} as well. Further work is also needed to find potent GloB inhibitors. In addition, coupled assays detecting the second product D-lactate could be explored. Structural studies focusing on obtaining a substrate or product bound GloB structure as well as any identified inhibitors could help to understand the reaction mechanism as well as mode of inhibition. This may also be beneficial in understanding the functional differences between the members of the MBL superfamily as well as differences between GloB and other Glx2 proteins.

5.4 References

1. Bebrone, C., Metallo-β-lactamases (classification, activity, genetic organization, structure, zinc coordination) and their superfamily. *Biochem. Pharmacol.* **2007,** *74* (12), 1686-1701.

2. Pettinati, I.; Brem, J.; Lee, S. Y.; McHugh, P. J.; Schofield, C. J., The Chemical Biology of Human Metallo-β-Lactamase Fold Proteins. *Trends Biochem. Sci.* **2016,** *41* (4), 338-355.

3. Baier, F.; Tokuriki, N., Connectivity between Catalytic Landscapes of the Metallo-β-Lactamase Superfamily. *J. Mol. Biol.* **2014,** *426* (13), 2442-2456.

4. Daiyasu, H.; Osaka, K.; Ishino, Y.; Toh, H., Expansion of the zinc metallo-hydrolase family of the β-lactamase fold. *FEBS Lett.* **2001,** *503* (1), 1-6.

5. Shaw, L.; Schauer, R., Detection of CMP-*N*-acetylneuraminic acid hydroxylase activity in fractionated mouse liver. *Biochem. J* **1989,** *263* (2), 355-363.

6. Irie, A.; Koyama, S.; Kozutsumi, Y.; Kawasaki, T.; Suzuki, A., The Molecular Basis for the Absence of *N*-Glycolylneuraminic Acid in Humans. *J. Biol. Chem.* **1998,** *273* (25), 15866-15871.

7. Chou, H.-H.; Takematsu, H.; Diaz, S.; Iber, J.; Nickerson, E.; Wright, K. L.; Muchmore, E. A.; Nelson, D. L.; Warren, S. T.; Varki, A., A mutation in human CMP-sialic acid hydroxylase occurred after the *Homo-Pan* divergence. *Proc. Natl. Acad. Sci. U. S. A.* **1998,** *95* (20), 11751-11756.

8. Cao, E.; Chen, Y.; Cui, Z.; Foster, P. R., Effect of freezing and thawing rates on denaturation of proteins in aqueous solutions. *Biotechnol. Bioeng.* **2003,** *82* (6), 684-690.

9. Lyumkis, D., Challenges and opportunities in cryo-EM single-particle analysis. *J. Biol. Chem.* **2019,** *294* (13), 5181-5197.

10. Nystedt, J.; Anderson, H.; Hirvonen, T.; Impola, U.; Jaatinen, T.; Heiskanen, A.; Blomqvist, M.; Satomaa, T.; Natunen, J.; Saarinen, J.; Lehenkari, P.; Valmu, L.; Laine, J., Human CMP-*N*-Acetylneuraminic Acid Hydroxylase is a Novel Stem Cell marker linked to Stem Cell-Specific Mechanisms. *Stem Cells* **2010,** *28* (2), 258-267.

11. Kim, D. I.; Jensen, S. C.; Noble, K. A.; Kc, B.; Roux, K. H.; Motamedchaboki, K.; Roux, K. J., An improved smaller biotin ligase for BioID proximity labeling. *Mol. Biol. Cell* **2016,** *27* (8), 1188-1196.

12. Roux, K. J.; Kim, D. I.; Raida, M.; Burke, B., A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *J. Cell Biol.* **2012,** *196* (6), 801-810.

13. Rigaut, G.; Shevchenko, A.; Rutz, B.; Wilm, M.; Mann, M.; Séraphin, B., A generic protein purification method for protein complex characterization and proteome exploration. *Nat. Biotechnol.* **1999,** *17*, 1030-1032.

14. Thornalley, P. J., Pharmacology of methylglyoxal: formation, modification of proteins and nucleic acids, and enzymatic detoxification - a role in pathogenesis and antiproliferative chemotherapy. *Gen. Pharmacol.* **1996,** *27* (4), 565-573.

15. Sousa Silva, M.; Gomes, R. A.; Ferreira, A. E.; Ponces Freire, A.; Cordeiro, C., The glyoxalase pathway: the first hundred years... and beyond. *Biochem. J* **2013,** *453* (1), 1-15.

16. O'Young, J.; Sukdeo, N.; Honek, J. F., *Escherichia coli* glyoxalase II is a binuclear zinc-dependent metalloenzyme. *Arch. Biochem. Biophys.* **2007,** *459* (1), 20-26.

Chapter 6: Materials and Methods

Contents

6.1 Materials 156
6.1.1 Cloning reagents 156
6.1.2 Competent cells and cell growth media 156
6.1.2 Mammalian cell lines and cell growth media 156
6.1.3 Protein purification columns 156
6.2 General procedures 157
6.2.1 pH measurements 157
6.2.2 Centrifugation 157
6.2.3 DNA and protein concentration measurements 157
6.2.4 Filtering solutions 157
6.3 Bacterial growth media, stock solution and buffer preparation 157
6.3.1 LB-agar medium preparation 157
6.3.2 Bacterial media preparation 158
6.3.3 Antibiotic stock preparation 158
6.3.4 IPTG stock solution preparation 158
6.3.5 Buffer preparation 158
6.4 Electrophoretic techniques 159
6.4.1 Agarose gel electrophoresis 159
6.4.2 SDS-PAGE 160
6.4.3 Western blotting 162
6.5 Molecular Cloning 164
6.5.1 Primers and vectors 164
6.5.2 Codon-optimised gene sequences 165
6.5.3 Polymerase chain reaction 165
6.5.4 Purification of PCR products 166
6.5.5 DNA digestion and ligation 166
6.5.6 Bacterial transformations 167
6.5.7 Plasmid extraction 167
6.6 Bacterial growth and protein production 167
6.6.1 Starter culture preparation 167
6.6.2 Expression trials 167
6.6.3 Large scale culture 168

6.7 Mammalian cell culture 168
6.7.1 General techniques 168
6.7.2 Maintenance of HeLa and TCMK-1 cells 168
6.7.3. Maintenance of adherent HEK293 168
6.7.4 Suspension adaptation for HEK293 169
6.7.5 Maintenance of HEK293 in suspension 169
6.7.6 Freezing and thawing cells 169
6.7.7 Cell transfection and stable cell line generation 170
6.7.8 Transient expression using Expi293 system 170
6.7.9 Transient expression using ExpiCHO system 170
6.7.10 Sample preparation for Western blotting 170
6.8 Protein purification 171
6.8.1 Cell lysis 171
Bacterial cells *171*
Expression trials *171*
Mammalian cells *171*
6.8.2 General procedures 172
Tag cleavage by 3C protease *172*
Protein concentrating *172*
Buffer exchange and protein storage *172*
6.8.3 Chromatographic techniques 172
Ni-affinity chromatography *172*
Size exclusion chromatography *172*
Elution of untagged protein and removal of His_6-tag and 3C protease *174*
Ion exchange chromatography *174*
Further ion exchange using spin columns *174*
6.9 X-ray crystallography 174
6.9.1 Broad screen and optimisation conditions 174
6.9.2 Crystallisation experiments 174
6.9.3 Soaking experiments 175
6.9.4 Crystal harvesting 175
6.9.5 Diffraction data collection and processing 175
6.9.6 Phasing and structure refinements 176
6.9.7 Data deposition 176
6.10 Thermal shift assays 176

6.11 Metal content analysis 176
6.11.1 ICP-MS 176
6.11.2 MicroPIXE 176
6.12 Generation of Apo enzymes 177
6.13 UV-visible spectroscopy 177
6.14 Circular Dichroism spectroscopy 177
6.15 Mass spectrometry 177
6.15.1 Non-denaturing MS 177
6.15.2 RapidFire High-throughput MS 178
6.16 mCMAH assays 178
6.17 GloB assays 179
6.17.1 Assay conditions 179
6.17.2 Measurements 179
6.17.3 Data analysis 180
6.18 Protein-protein interactions 180
6.18.1 Crosslinking and decrosslinking 180
6.18.2 Pulldowns 180
6.18.3 Co-immunoprecipitation 181
6.19 Immunofluorescence 181
6.19.1 Coverslip preparation 181
6.19.2 Image collection and processing 182
6.20 References 183

6.1 Materials

All chemicals and reagents were purchased from Sigma-Aldrich (Merck) at highest available purity unless stated otherwise.

Milli-Q water was purified using a Millipore Elix Reverse Osmosis system followed by filtration through a 0.22 μm Millipore Milli-Q filter.

6.1.1 Cloning reagents

Codon-optimised genes for *Escherichia Coli* and human cell expression were purchased from Thermo Fisher Scientific. Restriction enzymes and DNA ligase were obtained from New England Biolabs and *PfuUltra* High Fidelity DNA polymerase AD was purchased from Agilent.

6.1.2 Competent cells and cell growth media

E. coli strains XL10-Gold (Agilent), BL21(DE3) (New England Biolabs), and Rosetta-gami (DE3) pLysS (Novagen) were used for cloning and protein expression.

Bacto tryptone, yeast extract, agar and Terrific broth (TB) for bacterial cell growth media and antibiotic plates were obtained from Thermo Fisher Scientific.

Isopropyl β-D-1-thiogalactopyranoside (IPTG) was purchased from Apollo Scientific.

6.1.2 Mammalian cell lines and cell growth media

HEK293, HeLa, TCMK-1 cells were obtained from ATCC. ExpiCHO-S cells were obtained from Thermo Fisher Scientific.

FreeStyle293, Expi293 and ExpiCHO expression media, OptiMEM and Glutamax were obtained from Thermo Fisher Scientific.

FuGENE HD transfection reagent was obtained from Promega.

6.1.3 Protein purification columns

Affinity chromatography (HisTrap HP, HisTrap FF and HisTrap FF crude), size exclusion chromatography (Superdex 200 and Superdex 75) resin and columns and ion exchange Q-Sepharose FF resin were obtained from GE Healthcare Life Sciences.

6.2 General procedures

6.2.1 pH measurements

The pH of solutions was determined using Jenway pH Meter 3305, with an Aldrich glass/calomel combination electrode. Prior to measurements, calibration was carried out using phthalate (pH 4.0), phosphate (pH 7.0) and borate (10.0) buffer solutions (Thermo Fisher Scientific). When not in use, the electrode was stored in 4 M potassium chloride solution.

6.2.2 Centrifugation

Samples with a volume of less than 1.5 ml were centrifuged using 1-14 Microfuge (Sigma, room temperature) and Microfuge 22R bench top centrifuge (Beckman Coulter Ltd., 4 °C). Samples with a volume less than 50 ml were centrifuged and protein concentrating was performed using an Allegra X-30R centrifuge (Beckman Coulter Ltd), while an Avanti J-25 (Beckman Coulter Ltd) was used to pellet bacterial cell growth cultures and cell lysates at 4 °C. Mammalian cell cultures were pelleted using a Centrifuge 5702 (Eppendorf).

6.2.3 DNA and protein concentration measurements

DNA and protein concentrations were measured using a NanoDrop One Microvolume UV-Vis spectrophotometer (Thermo Fisher Scientific).

Protein extinction coefficients were determined using the ExPASy Protparam online application (http://www.expasy.org/).

6.2.4 Filtering solutions

Large volume (buffer) solutions were filtered under vacuum through a 0.22 μm filter. Filtropur S 0.2 and 0.45 (Sarstedt) syringe filters were used to filter cell lysates and solutions with 2 ml – 100 ml volumes. To filter small volumes (less than 2 ml), Corning Costar Spin-X Centrifuge Tube filters (0.22 μm and 0.45 μm) were used.

6.3 Bacterial growth media, stock solution and buffer preparation

6.3.1 LB-agar medium preparation

LB-agar medium was prepared according the recipe in **Table 6.1**. Solid components were resuspended in 500 ml Milli-Q water and sterilised by autoclaving. The autoclaved mixture was stored at room temperature until use.

Table 6.1: Recipe with required amounts for 1 l LB-agar

Reagent	Mass
Agar	7.5 g
Tryptone	5 g
Yeast extract	5 g
Sodium chloride	2.5 g

6.3.2 Bacterial media preparation

2TY and Terrific broth (TB) media were prepared following the protocol (**Table 6.2**). Components were dissolved in 900 ml Milli-Q water before adjusting volume to 1 l. The solution was autoclaved and cooled down to room temperature before use.

Table 6.2: Recipes for 1 l 2TY and TB media.

Medium	Reagent	Amount
2TY	Tryptone	16 g
	Yeast extract	10 g
	Sodium chloride	5 g
TB	TB medium powder	47.6 g
	glycerol	8 ml

6.3.3 Antibiotic stock preparation

Ampicillin and kanamycin 1000x stocks were made by dissolving 500 mg and 340 mg, respectively, of antibiotic in 10 ml Milli-Q water. The solutions were then filtered (0.22 µm filter) and aliquots were stored at -20 °C.

6.3.4 IPTG stock solution preparation

IPTG stock (100 mM) was prepared by dissolving 23.9 mg of IPTG in 1 ml Milli-Q water. The solution was filtered (0.22 µm) before use. Leftover stocks were frozen and stored at -20 °C to be used for expression trials.

6.3.5 Buffer preparation

Buffer compositions are listed in **Table 6.3**. All components were dissolved in 900 ml Milli-Q water and pH was adjusted to the required value. Volume was then adjusted to 1 l and buffers were filtered (0.22 µm). All buffers were stored at 4 °C.

Table 6.3: Buffers used for protein purification and mammalian cell-based experiments.

Buffer	Composition	pH
HisTrap Buffer A	25 mM HEPES, 500 mM NaCl, 5 mM imidazole	7.5
HisTrap Buffer B	25 mM HEPES, 500 mM NaCl, 500 mM imidazole	7.5
HisTrap Buffer C	25 mM HEPES, 500 mM NaCl, 20 mM imidazole	7.5
Gel filtration buffer	25 mM HEPES, 500 mM NaCl	7.5
Ion exchange buffer A	50 mM Tris	7.8
Ion exchange buffer B	50 mM Tris, 1 M NaCl	7.8
GloB ion exchange buffer A	50 mM Tris	8.0
GloB ion exchange buffer B	50 mM Tris, 1 M KCl	8.0
GloB Gel filtration buffer	50 mM Tris, 100 mM KCl	7.5
Phosphate buffed saline (PBS)	137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4	7.4

6.4 Electrophoretic techniques

6.4.1 Agarose gel electrophoresis

Tris, acetate, Ethylenediaminetetraacetic acid (EDTA) buffer (TAE) was prepared as a 50x stock according to protocol in **Table 6.4**. Solid components were dissolved in 700 ml Milli-Q water before adding glacial acetic acid. Volume was then adjusted to 1 l and buffer was stored at room temperature. Buffer was diluted to make 1x buffer as required.

Agarose gels were prepared using Clarit-E horizontal electrophoreses system (Alpha Laboratories). Agarose (amounts listed in **Table 6.4**) was suspended in 1x TAE buffer and heated in a microwave until fully dissolved. Solution was cooled down slightly and SYBR Safe Stain (Thermo Fisher Scientific) was added. The resulting solution was poured into a mould using 8- or 16-well comb and allowed to set.

Table 6.4: Recipes for TAE buffer and 0.8% agarose gel.

	Reagent	Amount
50x TAE	Tris base	242 g
	EDTA disodium salt	18.6 g
	Acetic acid (glacial)	57 ml
0.8% agarose gel	Agarose	0.4 g
	1x TAE buffer	50 ml
	SYBR Safe Stain	2 µl

DNA samples were mixed with 6x Purple Loading Buffer (Thermo Fisher Scientific) and loaded to the wells. GeneRuler 1 kb DNA ladder (Thermo Fisher Scientific) was loaded to the first well as a size reference. The gel was then run at 100 V for 30 min. DNA bands were visualised under UV light.

6.4.2 SDS-PAGE

Precast NuPAGE 4-12% Bis-Tris protein gels (1 mm, 10 and 15 wells), NuPAGE MOPS Running Buffer (20x) and a Mini Gel Tank were obtained from Thermo Fisher Scientific.

SDS-PAGE gels were made according to the recipes in **Table 6.5** using the Mini-PROTEAN Tetra Cell system (Bio-Rad). When making gels, the running gel layer was topped with propan-2-ol to prevent drying out. Propan-2-ol was removed before pouring the stacking gel layer. 10- and 15-well combs were used. Tetramethyleneethylenediamine (TEMED) was added last to both running and stacking gel mixtures.

Table 6.5: Recipe for running and stacking gel components to make four SDS-PAGE gels.

Gel	Reagent	Amount for 4 gels
Running gel	Milli-Q water	6.06 ml
	30% Bis-acrylamide	5.67 ml
	1.5 M Tris-HCl, pH 8.8	4 ml
	10% (w/v) Ammonium persulfate	267 µl
	10% (w/v) Sodium dodecyl sulfate	160 µl
	TEMED	13.3 µl
Stacking gel	Milli-Q water	3.05 ml
	30% Bis-acrylamide	1.25 ml
	0.5 M Tris-HCl, pH 6.8	670 µl
	10% (w/v) Ammonium persulfate	25 µl
	10% (w/v) Sodium dodecyl sulfate	50 µl
	TEMED	5 µl

10x SDS-PAGE running buffer was prepared according to the protocol below. Solution was stored at room temperature and diluted before use.

6x SDS-PAGE sample buffer was prepared using components listed in **Table 6.6** to make 50 ml solution in Milli-Q water. Solution was aliquoted (1 ml) and stored at -20 °C. Defrosted aliquots were stored at room temperature in between uses.

SDS-PAGE staining solution was made by mixing ingredients listed in **Table 6.6**. Solution was stored protected from light at room temperature.

Samples were prepared by mixing 10 µl of protein solution with 2 µl of 6x SDS-PAGE sample buffer and heating it at 95 °C on a hot block for 3 min. 10 µl of the sample was loaded onto a gel. The electrophoresis was run at constant voltage (180 V, 50 min for SDS-PAGE gels and 200 V, 40 min for NuPAGE gels). Gels were then stained using SDS-PAGE staining solution for 10 min and de-stained in water.

Table 6.6: Recipes for solutions used for SDS-PAGE.

Solution	Reagent	Amount
10x SDS-PAGE running buffer	Tris base	30 g
	Glycine	144 g
	Sodium dodecyl sulfate	10 g
6x SDS-PAGE sample buffer	Sodium dodecyl sulfate	6 g
	Dithiotreitol	4.65 g
	Bromophenol blue	30 mg
	Glycerol	23.5 ml
	1 M Tris-HCl, pH 6.8	3 ml
SDS-PAGE staining solution	Coomassie brilliant blue G250	0.4 g
	Ethanol	50 ml
	70% perchloric acid	34 ml
	Milli-Q water	800 ml

6.4.3 Western blotting

Prepared samples (lysates or purified proteins) were run on NuPAGE Bis-Tris Precast gels at 30 mA for 110 min. Proteins were transferred from gels to a nitrocellulose membrane (Amersham Protran 0.2 μm NC, GE Healthcare Life Sciences) at 100 V for 60 min in ice cold transfer buffer (25 mM Tris base, 192 mM glycine, 20% (v/v) methanol). Membranes were then saturated with 5% (w/v) Blotting-Grade Blocker (Bio-Rad) in PBS-T (PBS buffer supplemented with 0.1% (v/v) Tween-20) for 1 h at room temperature and incubated with primary antibody diluted in 5% Blotting-Grade Blocker in PBS-T for 1 h at room temperature or overnight at 4 °C with gentle shaking. To remove excess antibody, membranes were washed with PBS-T (1x 10 min, 2x 5 min) and membranes were incubated with the appropriate secondary antibody for 1 h at room temperature. Excess antibody was washed with PBS-T (1x 10 min, 2x 5 min) and membranes were incubated with ECL™ Prime Western Blotting Detection Reagent (GE Healthcare Life Sciences) for 5 min before imaging on ChemiDoc MP Imaging System with ImageLab software (BioRad). Antibodies used and their dilutions are listed in **Table 6.7**.

Table 6.7: Antibodies used for Western blot (WB) and immunofluorescence (IF) experiments.

Antibody	Epitope	Host species	Dilution for WB	Dilution for IF	Supplier
Anti-CMAH (E-7), monoclonal	3-242 of mouse CMAH	mouse	1:1000	1:25	Santa Cruz Biotechnologies, sc-365023
Anti-6X His tag, monoclonal	His_6 tag	mouse	1:1000 (lysates), 1:5000 (pulldown)	1:250	Abcam, ab18184
Anti-6X His tag, polyclonal	His_6 tag	rabbit	1:2000	1:300	Abcam, ab9108
Anti-cytochrome b_5, polyclonal	1-100 of human cytochrome b_5, Type A	rabbit	1:500	-	Abcam, ab186863
Anti-cytochrome b_5, type B (F-5), monoclonal	86-130 of human cytochrome b_5, Type B	mouse	1:500	-	Santa Cruz Biotechnologies, sc-390876
Anti-calnexin, polyclonal	*C*-terminal sequence of dog calnexin	rabbit	-	1:50	Enzo Life Sciences, ADI-SPA-860-D
Anti-beta actin [AC-15], monoclonal (HRP conjugated)	*N*-terminal sequence of β-actin	mouse	1:50000	-	Abcam, ab49900
Anti-mouse secondary (HRP conjugated)	Mouse IgG	goat	1:5000	-	Invitrogen, G21040
Anti-mouse secondary (AlexaFluor488)	Mouse IgG	goat	-	1:1000	Invitrogen, A11029
Anti-rabbit secondary (HRP conjugated)	Rabbit IgG	goat	1:15000	-	Promega, W4011
Anti-rabbit secondary (AlexaFluor488)	Rabbit IgG	goat	-	1:1000	Invitrogen, A11034
Anti-rabbit secondary (AlexaFluor594)	Rabbit IgG	goat	-	1:1000	Invitrogen, R37117

6.5 Molecular Cloning

GloB was cloned to pET22b-HA-CPDSalI[1] by Dr Madeline Kavanagh from a synthetic gene codon-optimised for expression in *E. coli* using restriction sites NdeI and BamHI.

Microcentrifuge tubes and pipette tips used in the procedures described here were sterilised by autoclaving (121 °C, 11 min) and Milli Q water was filtered (0.22 μm filter) before use.

6.5.1 Primers and vectors

Oligonucleotides were designed using ApE (A Plasmid Editor) software[2] (see **Table 6.8**) and obtained as lyophilised pellets. Primers were dissolved in filtered (0.22 μm) Milli-Q water to prepare a 100 μM stock solution and then diluted to 10 μM prior to use in the Polymerase chain reaction (PCR).

Table 6.8: DNA primers used for PCR and sequencing reations. Melting temperature (T_m) were estimated using ApE - A plasmid Editor software[2].

Primer name	Primer sequence	T_m (°C)
hCMAH-pMAL Fw 1	GTTCCAGGGGCCCGATGAAAATAATGGTCTGCTG	69
hCMAH-MAL Fw 2	CATATGCTGGAAGTTCTGTTCCAGGGGCCCGATG	70
hCMAH-pMAL Rv 1	GTGGTGGTGGTGGTGGGTGGTTTTACATTCTTGC	69
hCMAH-pMAL Rv 2	GAATTCTTAGTGGTGGTGGTGGTGGTGGGTGG	69
hCMAH gene seq Fw	CCAACACCAAAAACGATCCGAACG	61
mCMAH-pMAL Fw long	CATATGCTGGAAGTTCTGTTCCAGGGGCCCATGGATCGTAAACAGACC	73
mCMAH pMAL Rv long	GAATTCTTAGTGGTGGTGGTGGTGGTGGTCGCAATGCATCAGAAAGC	73
mCMAH gene seq Fw	CCGGAAATGGATACCTGCATTATCGTG	61
GloB-M Fw	GAAGGAGATATACATATGAACCTGAATAGC	57
GloB-M Rv	GCTATTCAGGTTCATATGTATATCTCCTTC	57

Vectors used in experiments described in this thesis are listed in **Table 6.9**.

Table 6.9: Expression vectors used for work described in this thesis.

Vector	Resistant to	Obtained from
pcold-6xHis-HRV3Csite-HsTRBP	Ampicillin	Phillip Zamore (Addgene plasmid # 41096[3])
pMAL-c5x	Ampicillin	New England Biolabs
pcDNA3.1(+)	Ampicillin, neomycin	Thermo Fisher Scientific

6.5.2 Codon-optimised gene sequences

Codon-optimised genes used are listed in **Table 6.10**.

Table 6.10: Codon-optimised genes used in experiments with restriction sites for cloning.

Protein encoded	Codon-optimised for expression in	Restriction sites
hCMAH	*E. coli*	SalI and NotI
mCMAH	*E. coli*	SalI and NotI
hCMAH	*H. sapiens*	HindIII and XhoI
mCMAH	*H. sapiens*	HindIII and XhoI

6.5.3 Polymerase chain reaction

The polymerase chain reaction (PCR) was run on TC-312 thermal cycler (Techne). A typical PCR reaction (final volume 50 µL, **Table 6.11**) for *PfuUltra* High Fidelity DNA Polymerase AD was set up.

Table 6.11: Typical PCR reaction composition.

Reagent	Volume (µl)
10 x reaction buffer	5
10 µM Forward primer	1
10 µM Reverse primer	1
50 ng/µl DNA template	2
10 mM dNTP mix	1
DNA polymerase	1
Milli-Q water	39

For high GC content primers (hCMAH and mCMAH PCR reactions) dimethyl sulfoxide (DMSO, 0.2 µl) was added to the reaction mixture. To insert restriction sites and 3C-protease cleavage site into hCMAH and mCMAH, two rounds of PCR were run. For site-directed mutagenesis – a single round of PCR was run. Typical thermocycler parameters are listed below.

	Restriction site insertion		Site-directed mutagenesis	
Initial denaturation	95 °C, 2min		95 °C, 2min	
Denaturation	95 °C, 30 s		95 °C, 30 s	
Annealing *	55 °C, 30 s	5 x	53 °C, 1 min	25 x
Elongation	72 °C, 2min		72 °C, 14 min	
Denaturation	95 °C, 30 s			
Annealing *	66 °C, 30 s	25 x		
Elongation	72 °C, 2 min			
Final elongation	72 °C, 10 min		72 °C, 10 min	

*Annealing temperatures were set 3-4 °C lower than the melting temperatures (T_m) of primers used as estimated by ApE software[2]. For restriction site insertion PCR annealing temperatures for the first round were chosen based on T_m region of the primer complementary to the template DNA.

6.5.4 Purification of PCR products

PCR products were run on 0.8% agarose gel (as described in Section 6.4.1 Agarose gel electrophoresis) and visualised under UV light. Bands containing amplified DNA were excised and purified according to manufacturer's instructions using GeneJET Gel Extraction Kit (Thermo Fisher Scientific).

The PCR product for site-directed mutagenesis was purified according to the manufacturer's instructions using the GeneJET PCR Purification Kit (Thermo Fisher Scientific).

6.5.5 DNA digestion and ligation

Gene insert DNA sequences (500 ng) and vectors (1000 ng) were digested with appropriate restriction enzymes (see **Table 6.10**, for pMAL constructs EcoRI and NdeI restriction enzymes were used) at 37 °C overnight. Digested vectors were then run on an agarose gel and purified using GeneJET gel Extraction Kit while linear fragments were purified using GeneJET PCR Purification Kit (Thermo Fisher Scientific).

The digested vector (40 ng) was incubated with digested gene insert DNA in 1:10 ratio and T4 DNA ligase at room temperature for 16-48 h.

The PCR product for site-directed mutagenesis reaction was treated with DpnI (37 °C, 3 h) in FastDigest buffer (Thermo Fischer Scientific) to remove methylated template DNA.

6.5.6 Bacterial transformations

Competent cells were incubated on ice with plasmid DNA (**Table 6.12**) before heat-shocking at 42 °C for 30 s. Cells were incubated on ice for 2 min before adding SOC medium and incubating at 37 °C for 1 h while shaking (200 rpm). 100-200 µl of cell culture were then plated on LB-agar containing ampicillin (XL10-Gold and BL21(DE3) cells) or ampicillin and kanamycin (Rosetta-gami (DE3) pLysS) and incubated at 37 °C overnight.

Table 6.12: Transformation conditions for bacterial strains used for cloning and protein production.

Strain	DNA sample used	Incubation time	SOC medium added
XL10-Gold	5 µl ligation/ DpnI digestion mixture	20 min	250 µl
BL21(DE3)	100 ng plasmid DNA	20 min	250 µl
Rosetta-gami (DE3) pLysS	50 ng plasmid DNA	5 min	80 µl

6.5.7 Plasmid extraction

Single colonies (XL10-Gold) were picked and grown in 5-10 ml of 2TY medium supplemented with 50 µg/ml ampicillin overnight (37 °C, 200 rpm). Plasmid DNA was purified using GeneJET Plasmid Miniprep Kit (Thermo Fisher Scientific) according to the manufacturer's guidelines. Constructs were checked by standard Sanger sequencing (Source Bioscience or Eurofins Genomics).

6.6 Bacterial growth and protein production

The work outlined here was performed inside a HERASafe microbiological safety cabinet (Thermo Fisher Scientific) using sterile pipettes and pipette tips.

6.6.1 Starter culture preparation

A single colony of freshly transformed cells was picked and grown in 100 ml of 2TY medium supplemented with 50 µg/ml ampicillin (BL21(DE3) cells) or 50 µg/ml ampicillin and 34 µg/ml kanamycin (Rosetta-gami (DE3) pLysS) at 37 °C, 200 rpm.

6.6.2 Expression trials

Flasks containing 100 ml TB and 2TY media were supplemented with 50 µg/ml ampicillin (BL21(DE3)) or 50 µg/ml ampicillin and 34 µg/ml kanamycin (Rosetta-gami (DE3) pLysS) were inoculated with 10 ml of overnight culture and grown at 37 °C, 180 rpm. Maltose binding protein (MBP) tagged constructs were additionally supplemented with 10 mM glucose. Cells were grown to OD_{600} of 0.5 – 0.6, temperatures were then changed to 15 °C (pCold constructs),

18 °C, 20 °C, 25 °C or 37 °C and cells were induced using varied isopropyl IPTG concentrations (0 – 1 mM). Cells were incubated at the induction temperature overnight (or 4 h at 37 °C). 50 ml of culture from each flask was spun down (3082 x g, 10 min, 4 °C) and pellets were stored at -80 °C.

6.6.3 Large scale culture

Successful conditions identified during expression trials were scaled up to 12 l. Cells were grown in 1 l flasks containing TB (hCMAH) supplemented appropriate antibiotics, and 10 mM glucose (pMAL construct) at 37 °C to OD_{600} of 0.6 – 0.7. Temperature was changed to 15 °C (pCold constructs) or 18 °C (pMAL construct), protein production was induced with 0.01 mM IPTG and cells were incubated overnight. Cells were pelleted (11900 x g, 10 min, 4 °C) and pellets were stored at -80 °C.

GloB expressing BL21(DE3) cells were grown in 2TY medium. Expression was induced by addition of 0.5 mM IPTG following a 4 h incubation at 37 °C before harvesting.

6.7 Mammalian cell culture

6.7.1 General techniques

Adherent cells were grown at 37 °C in a humidified incubator (5% CO_2). Cells in suspension were grown at 37 °C in a humidified incubator (8% CO_2). Sterile pipettes and culture flasks were used and all the work was performed inside HERAsafe microbiological safety cabinet.

To count the cells 10 µl of cell suspension were mixed with 10 µl Trypan Blue Solution (0.4%, Thermo Fisher Scientific). 10 µl of the mixture were transferred to a Countess cell counting chamber slide (Thermo Fisher Scientific) and cells were counted using Countess II FL cell counter (Thermo Fisher Scientific).

6.7.2 Maintenance of HeLa and TCMK-1 cells

HeLa and TCMK-1 cells were grown in T75 flasks until 70-80% confluent. Medium was removed and cells were rinsed with 6 ml PBS before detaching cells by incubating with 0.25% trypsin-EDTA (1.5 ml) at 37 °C for approximately 3 min. Cells were resuspended in fresh medium to final volume of 10 ml and 1-5 ml of suspension was transferred to a new flask and topped up to a final volume of 12 ml with fresh medium.

6.7.3. Maintenance of adherent HEK293

Cells were grown in T25 or T75 flasks till 70 - 80% confluent. The growth medium was removed and cells were detached by washing from the culture flask with DMEM. 1-6 ml of the

cell suspension were transferred to a new flask and topped up with DMEM to a final volume of 12 ml in T75 flasks.

6.7.4 Suspension adaptation for HEK293

HEK293 cells were transferred to T75 flask and topped up with 1:1 mixture of DMEM and FreeStyle 293. Once confluent, cells were expanded to 3 x T75 flasks. Cells were grown in 1:1 mixture of DMEM and FreeStyle 293 for 7 – 10 days before detaching and transferring to 20 ml FreeStyle 293 medium in 125 ml Corning Erlenmeyer Shake Flask.

6.7.5 Maintenance of HEK293 in suspension

Cells were grown at 37 °C, 8% CO_2 while shaking at 150 rpm. Cell medium was changed every 2-3 days. Cells were pelleted (0.2 g, 5 min), supernatant was collected and stored at -20 °C (when containing recombinant protein) and pellet resuspended in fresh FreeStyle 293 medium. The suspension volume was increased every time cell count reached 3 million cells/ml.

6.7.6 Freezing and thawing cells

Cells were counted, pelleted and resuspended in growth medium, DMSO and Fetal Bovine Serum (FBS) were added to the final concentrations listed in **Table 6.13**. Cells were aliquoted to 1 ml and frozen in Nalgene Mr. Frosty freezing container placed at -80 °C overnight. Frozen cells were stored at -80 °C or in LS4800 cryogenic sample storage (Worthington Industries) equipped with Liquid Nitrogen Low Level Alarm - M284CE (Taylor Wharton).

Table 6.13: Freezing conditions for different mammalian cell lines.

	HEK293	HEK293 (suspension)	HeLa	TCMK-1	ExpiCHO-S
Cell count ($x10^6$ cells/ml)	3	10	3	3	10
DMSO	10%	10%	10%	5%	10%
FBS	20%	-	20%	20%	-

Adherent cells were thawed at 37 °C and pelleted to remove DMSO-containing medium before resuspending in 4 ml of medium and transferring to T25 flask. Suspension cells were thawed at 37 °C and transferred to 30 ml of medium. Cells were grown as described above.

6.7.7 Cell transfection and stable cell line generation

Cells were transfected with hCMAH-pcDNA3.1(+) and mCMAH-pcDNA3.1(+) plasmids using final DNA concentration of 0.02 µg/µl in OptiMEM reduced serum medium and 1:3 FuGENE HD:DNA ratio in 12-well format.

Stable single cell clones were selected by treating transfected cells (48 h after transfection) with 1 mg/ml G-418, Geneticin (Thermo Fisher Scientific) in 96-well format. Grown single cell clones were then transferred to 46-well, 12-well and 6-well plates using medium supplemented with G-418. Cell clones from 6-well plates were detached in 2 ml of DMEM medium. 1 ml was transferred to T25 flask and topped up with 3 ml medium and the remaining 1 ml was used for Western blot analysis to select cell clones expressing high levels of recombinant protein of interest.

6.7.8 Transient expression using Expi293 system

HEK293 cells adapted to grow in suspension by Dr Ilaria Pettinati (as described in section 6.7.7 Cell transfection and stable cell line generation) were used for transfection. Cells were transfected with hCMAH- and mCMAH-pcDNA3.1 plasmids according to the Expi293 Expression System User Guide (Thermo Fisher Scientific).

Cells were counted and 1 ml samples for WB analysis were collected every two days. Cells were harvested when viability fell below 65%.

6.7.9 Transient expression using ExpiCHO system

Cells were transfected with hCMAH- and mCMAH-pcDNA3.1 plasmids according to the ExpiCHO Expression System User Guide (Thermo Fisher Scientific) in 100 ml culture volume using the Max Titer protocol.

Cells were counted and 1 ml samples for WB analysis were collected every day starting day 4. Cells were harvested when viability fell below 70%.

6.7.10 Sample preparation for Western blotting

Cell suspensions were spun down (200 x g, 5 min, room temperature) and supernatant was transferred to a clean microcentrifuge tube.

Pelleted cells were washed with PBS and resuspended in 30-300 µl radioimmunoprecipitation assay (RIPA) modified lysis buffer (PBS, 0.05% Triton X-100, 0.02% SDS, 0.2% sodium deoxycholate) supplemented with 100x Protease Inhibitor Cocktail (Thermo Fisher Scientific)

and incubated for 30 min on ice. Cell debris was removed by centrifugation (18000 x g, 10 min, 4 °C) and supernatants were stored at -80 °C.

Collected cell media samples were spun down (16100 x g, 10 min, room temperature) and proteins were precipitated with aqueous trichloroacetic acid (TCA, 10% (v/v) final concentration) for 1 h at room temperature. Precipitated proteins were pelleted (18000 x g, 10 min) and pellets were washed twice in cold acetone. Pellets were allowed to air dry for 5-10 min before resuspending in 30-40 µl RIPA buffer.

Protein concentrations of lysates and solubilised proteins were determined by mixing 1 ml of Milli-Q water with 2 µl of protein sample (or RIPA buffer for blank measurement), 1 ml of Bradford reagent was then added and absorbance at 595 nm was measured. Bovine serum albumin (BSA) was used to construct a standard curve.

6.8 Protein purification

6.8.1 Cell lysis

Bacterial cells

Bacterial pellets were resuspended in HisTrap Buffer A or GloB Ion Exchange Buffer A (4 ml of buffer for 1 g of pellet) and supplemented with DNaseI and cOmplete protease inhibitor cocktail tablet (Roche). Cells were then lysed by sonication (Sonics Vibra-cell, 60% amplitude, 9.9 s pulse and 9.9 s rest cycle, total sonication time 5 min). Cell debris was removed by centrifugation (48400 x g, 2x 20 min, 4 °C) and the supernatant was filtered (0.22 µm filter).

Expression trials

Cells were lysed as described above and 1 ml of clarified lysate was incubated with 200 µl 1:1 mixture of Profinity™ IMAC Ni-charged resin (Bio-Rad) and HisTrap Buffer A at 4 °C for 45 min. The resin was washed with 3 x 200 µl of HisTrap Buffer A (1000 x g, 3 min, 4 °C) and bound proteins were eluted with 200 µl of HisTrap Buffer B following a 30 min incubation at 4 °C. Levels of protein of interest were assessed by SDS-PAGE.

Mammalian cells

Cells were resuspended in HisTrap Buffer A supplemented with benzonase nuclease (Santa Cruz Biotechnology), cOmplete protease inhibitor tablet, 1 mM Tris(2-carboxyethyl)phosphine hydrochloride (TCEP-HCl) and 0.05% IGEPAL CA-630 (100 ml buffer for pellets from 3 l culture). Cells were lysed by sonication (Sonics Vibra-cell 60% amplitude, 9.9 s pulse, 9.9 s rest, total sonication time 7 min). Cell debris was removed by

centrifugation (48400 x g, 2 x 20 min, 4 °C). Supernatant was diluted with equal volume of HisTrap A buffer.

6.8.2 General procedures

Tag cleavage by 3C protease

C-Terminal His_6 and MBP tags were cleaved by incubating with 1:100 (w:w) 3C protease with an *N*-terminal-His_6 tag at 4 °C for 16 h.

Protein concentrating

Proteins were concentrated using an Amicon Ultra 15 ml centricon with 30 kDa (hCMAH and mCMAH) or 10 kDa (GloB) molecular weight cut-off (3082 x g, 4 °C) or 0.5 ml centricon (18000 x g, 4 °C).

Buffer exchange and protein storage

Proteins were concentrated and buffer exchanged to crystallography buffers (see **Table 6.14**). Small sample volumes (less than 100 µl) were buffer exchanged using Micro-Bio-Spin 6 columns (Bio-Rad). For larger sample volumes proteins were buffer exchanged and concentrated using Amicon Ultra centricons. Buffer exchanged proteins were then aliquoted, flash-frozen in liquid N_2 and stored at -80 °C.

Table 6.14: Storage and crystallography buffers for different proteins.

Protein	Final buffer
hCMAH	25 mM HEPES, 50 mM NaCl, pH 7.5
mCMAH	25 mM HEPES, 100 mM NaCl, pH 7.5
GloB	50 mM Tris, pH 7.5

6.8.3 Chromatographic techniques

Ni-affinity chromatography

Clarified lysates were loaded onto a HisTrap column (as in **Table 6.15**) followed by an extensive wash with HisTrap Buffer A, followed by wash with HisTrap Buffer A with 4% HisTrap Buffer B and elution with a gradient of HisTrap Buffer B (20 – 500 mM imidazole). Fractions containing the protein of interest were identified by SDS-PAGE, combined and concentrated as described earlier.

Size exclusion chromatography

Size exclusion Superdex 200 (120 ml, hCMAH and mCMAH) and Superdex 75 (300 ml, GloB) columns were equilibrated with respective Gel Filtration Buffers and concentrated samples

(2 ml for 120 ml column and 8 ml for 300 ml column) were injected onto the columns. Proteins were eluted with 1 column volume of respective Gel Filtration Buffer. Protein purity was assessed by SDS-PAGE.

Table 6.15: Protein purification steps with chromatography columns used for each protein construct.

Construct	Tag(s)	Purification steps and columns
hCMAH-pCOLD	*N*-His_6	1 – HisTrap HP 5 ml 2 – Size exclusion (Superdex 200, 120 ml) 3 – 3C tag cleavage 4 – HisTrap HP 1 ml
hCMAH-pMAL	*N*-MBP, *C*-His_6	1 – HisTrap HP 5 ml 2 – 3C tag cleavage 3 – Size exclusion (Superdex 200, 120 ml) 4 – HisTrap HP 1 ml
hCMAH-pcDNA3.1 (secreted)	*C*-His_6	1 – HisTrap HP 1 ml
mCMAH-pcDNA3.1 (secreted)	*C*-His_6	1 – HisTrap HP 1 ml
mCMAH-pcDNA3.1 (from lysed cells, small scale)	*C*-His_6	1 – HisTrap FF crude 1 ml 2 – 3C tag cleavage 3 – HisTrap HP 1 ml *4 – Pierce Strong Anion Exchange Spin Column, Mini, if needed*
mCMAH-pcDNA3.1 (from lysed cells, large scale)	*C*-His_6	1 – HisTrap FF crude 5 ml 2 – Size exclusion chromatography (Superdex 200, 120 ml) or Anion exchange (Q-sepharose FF, 10 ml) 3 – 3C tag cleavage 4 – HisTrap HP 1 ml
GloB-pET22b	none	Adapted from O'Young *et al*.[4] 1 – Anion exchange (Q-sepharose FF, 60 ml) 2 – Size exclusion chromatography (Superdex 75, 300 ml)

Elution of untagged protein and removal of His_6-tag and 3C protease

To remove the His_6 tag and His_6-tagged 3C protease protein sample was supplemented with imidazole to final concentration of 20 mM (or buffer exchanged if already contained imidazole) and loaded onto a HisTrap HP 1 ml column pre-equilibrated with HisTrap Buffer C. The untagged protein was eluted by washing the column with 3 x 3 ml of HisTrap Buffer C. Protein purity was assessed by SDS-PAGE and protein containing fractions were combined and concentrated.

Ion exchange chromatography

Q-Sepharose FF columns (10 ml for mCMAH and 60 ml for GloB) were equilibrated with Ion Exchange Buffer A before loading the sample. Column was then washed with 3 column volumes of Ion Exchange Buffer A and bound proteins were eluted with a gradient of Ion Exchange Buffer B (0-1000 mM NaCl or KCl). Fractions containing the protein of interest were identified by SDS-PAGE, combined and concentrated.

Further ion exchange using spin columns

Pierce Strong Anion Exchange Spin Columns (Thermo Fisher Scientific) were used according to manufacturer's instructions to test ion exchange conditions and further purify small volumes of protein.

6.9 X-ray crystallography

6.9.1 Broad screen and optimisation conditions

Solutions for broad screens and optimisation screens were stored in 96 well polypropylene V-bottomed Masterblocks (Greiner Bio One) with 2 ml of solution per well. Blocks were sealed with Starseal aluminium foil sealing tape (Starlab) and stored at 4 °C.

6.9.2 Crystallisation experiments

Crystal plates were set up in 96 well Intelli-Plates (Art Robbins Instruments) with three sitting drop vapour diffusion experiments per condition using Phoenix/RE instrument (Art Robbins Instruments) equipped with a 96-syringe head and a single nano dispenser. 1:2, 1:1 and 2:1 ratios of condition:protein solutions were used. Crystals were grown at 4 °C or 20 °C. Protein concentration ranges and additives used for co-crystallisation are listed in **Table 6.16**. Proteins samples in storage buffer (see **Table 6.14**) were used for crystallography experiments unless stated otherwise in corresponding chapters.

Table 6.16: Protein concentration ranges and additives used for crystallisation experiments. CMP Neu5Ac – cytidine monophosphate *N*-acetylneuraminic acid.

Protein	Concentration range	Additive
hCMAH	10 – 20 mg/ml	none
mCMAH	3 – 16 mg/ml	None or 3-fold excess of CMP-Neu5Ac
GloB	10 – 24 mg/ml	None or 0.1 mM $ZnCl_2$

6.9.3 Soaking experiments

GloB crystals were transferred to a well containing a crystallisation condition supplemented with a small molecule compound at a stated concentration and 20% glycerol (for non-cryoprotected condition) and left for a time stated in **Table 6.17**.

After soaking in no tartrate condition, crystals were transferred to a condition containing *S*-D-lactoylglutathione (SLG) or S-methylglutathione for 30 s – 30 min.

Table 6.17: Compounds used in soaking experiments, their concentrations and soaking times.

Compound soaked	Concentration	Time range
SLG	3 mM – 20 mM	30 s – overnight
Glutathione (GSH)	5 mM	2 min – 30 min
S-methylglutathione	5 mM	5 min – 30 min
Tartrate (soaked out)*	0 mM (-20 mM from original condition)	30 min – 120 min

6.9.4 Crystal harvesting

Cryoprotectant solutions were made up in the original condition by supplementing it with glycerol to a final concentration of 20% (v/v). No cryoprotectant was used for conditions already containing glycerol, 2-methyl-2,4-pentanediol or low molecular weight polyethylene glycols (PEGs). Crystals were soaked in cryoprotectant solution before removing from the solution using a nylon loop and plunging in liquid N_2. Cryogenically cooled crystals were stored under liquid N_2 until data collection.

6.9.5 Diffraction data collection and processing

Data were collected in-house using CuKα d±1.58 Å source coupled with CCD Saturn 944+ detector (Rigaku) and equipped with Cryostream Controller 700 (Oxford Cryosystems), and at the Diamond Light Source synchrotron beamlines I03, I04 and I24. Data collected in house were integrated, scaled, indexed and merged using HKL-3000. Synchrotron data were processed using the CCP4i2 pipeline[5].

6.9.6 Phasing and structure refinements

GloB crystal structures were solved by molecular replacement using Phaser[6]. The initial GloB structure was phased using *Salmonella thyphimurium* LT2 glyoxalase II structure as a model (Protein Data Bank (PDB) accession code 2QED, 78% sequence identity). All further structures were phased using the initial model. Models were refined using iterative cycles of Phenix[7] Refine and model building in Coot[8].

In addition to Phaser, molecular replacement programmes available in CCP4 software package[5] were used for hCMAH and mCMAH phasing attempts. ContaMiner[9] and SIMBAD[10] were used to check for contaminant crystals.

6.9.7 Data deposition

Refined models and density maps for *E. coli* GloB were deposited to the PDB Europe (PDBe, accession codes 6S0I and 6RZ0).

6.10 Thermal shift assays

Protein thermal stability was assessed using thermal shift assays[11]. RUBIC Buffer Screen, Durham pH Screen and Durham Salt Screen (Molecular Dimensions) were used to find optimal conditions and optimise protein stability. Assays were set up on a 25 µL scale in a 96-well white PCR plate (StarLabs). Experiments were run from 20 °C to 95 °C in steps of 0.2 °C using CFX96 Touch Real-Time PCR Detection System (Bio-Rad). Excitation wavelength was 492 nm and emission was measured at 610 nm Data were analysed using Bio-Rad CFX Manager 3.1 software.

6.11 Metal content analysis

6.11.1 ICP-MS

Samples for inductively coupled plasma mass spectrometry (ICP-MS) measurements were buffer exchanged to 25 mM HEPES, 100 mM NaCl, pH 7.5 buffer prepared in Ultrapure water and submitted for analysis. ICP-MS was run by Phil Holdship, Department of Earth Sciences, University of Oxford.

6.11.2 MicroPIXE

Samples were prepared in 20 mM tris-base, 10% Glycerol, pH adjusted to 7.5 with acetic acid at a final concentration of 3.2 mg/ml (mCMAH) and 5.0 mg/ml (hCMAH). MicroPIXE experiments were performed by Peter Woodcock and Elspeth Garman, Department of Biochemistry, University of Oxford.

6.12 Generation of Apo enzymes

Apo-GloB and apo-mCMAH were generated by dialysing purified enzymes against 3 changes of buffer containing chelating agents. Samples were placed in 0.5 ml or 3 ml Slide-A-Lyzer Dialysis Cassettes (Thermo Scientific) and dialysed against >100 volumes of 50 mM Tris, pH 7.5 buffer containing 50 mM 2,6-pyridinedicarboxylic acid (DPA, GloB) or 20 mM EDTA (mCMAH) and 5 mM 1,10-phenantroline over 20-24 h (three changes of buffer). DPA was removed from apo-GloB sample by a second round of dialysis against >100 volumes of 50 mM Tris, pH 7.5 while apo-mCMAH was dialysed against > 100 volumes of 200 mM ammonium acetate, pH 7.5 over 20 – 24 h (three changes of buffer). Dialysis was performed at 4 °C. All buffers were treated with Chelex-100 prior to use.

6.13 UV-visible spectroscopy

Presence of Rieske [2Fe2S] cluster was confirmed by measuring UV-visible spectrum of 2 µl samples (3 – 15 mg/ml) in the range of 190 nm to 850 nm at 0.5 nm intervals using NanoDrop One UV-vis spectrophotometer.

6.14 Circular Dichroism spectroscopy

Circular dichroism spectroscopy was performed using a Chirascan CD/Fluorimeter spectrometer (Applied Photophysics). Enzyme samples were buffer exchanged to 10 mM phosphate buffer, pH 7.5 and diluted to 0.2 mg/ml final concentration. Samples were transferred to quartz cuvettes with 0.1 cm path length. Spectra were recorded in triplicate at 20 °C for wavelength range from 260 nm to 185 nm at 0.5 nm intervals. The baseline was corrected by subtracting the spectrum of the buffer. Thermal denaturation experiments were run between 20 °C and 80 °C. Full spectra or ellipticity at 218 nm or 222 nm were measured. Data for a single wavelength were fitted to a Boltzmann sigmoidal curve and apparent T_m was estimated using OriginPro software[12].

6.15 Mass spectrometry

6.15.1 Non-denaturing MS

Non-denaturing mass spectrometry (MS) experiments were performed in collaboration with Shuang Liu. Enzymes were buffer exchanged to 200 mM ammonium acetate, pH 7.5 using Micro-Bio-Spin 6 columns or by dialysis (apo-mCMAH). Spectra were acquired using quadrupole time-of-flight (Q-TOF) mass spectrometer (Synapt HD MS, Waters). Samples were transferred to the wells on a sample plate of an automated chip-based nano-electrospray ion source (TriVersa Nanomate, Advion) and then sprayed through a nozzle in a Nanomate chip at

a spray voltage of 1.85 kV (spray backing gas pressure 0.8 psi, inlet pressure 3.7 mbar). Sample and extractor cone voltages were kept at 100 V and 5.2 V respectively. Calibration and sample acquisition were performed in the positive ion mode. Data collection and processing were performed using MassLynx 4.0 (Waters).

6.15.2 RapidFire High-throughput MS

RapidFire high-throughput MS experiments were run in collaboration with Dr Anthony Tumber. Spectra were acquired using an Agilent Technologies 6550 iFunnel Q-TOF LC/MS spectrometer paired with RapidFire RF365 high-throughput sampling robot (Agilent). Samples were aspirated under vacuum for 400 ms and applied to a solid phase extraction (SPE) cartridge. A C4 cartridge was used for protein measurements, while C18 cartridge was used for mCMAH hydroxylation assays. The SPE was then washed with an aqueous phase for 5500 ms at a flow rate of 1.5 ml/min and bound analytes of interest were eluted from the SPE in an organic phase for 5500 ms at a flow rate of 1.25 ml/min. Between sample injections the SPE was re-equilibrated with LCMS grade water. The entire cycle of sample injection, wash, elution and re-equilibration taking 12 – 15 seconds

Protein spectra were recorded in positive ion mode, using 0.1% (v/v) formic acid in LC-MS grade water as aqueous phase and 85% (v/v) acetonitrile in LC-MS grade water with 0.1 % (v/v) formic acid as organic phase. Spectra were deconvoluted using Agilent MassHunter Protein Deconvolution Software across the appropriate mass range. Hydroxylation assays for mCMAH were run in negative ion mode, using 6 mM octylammonium acetate (OAA)[13] as aqueous phase and 80% acetonitrile in water as organic phase. OAA is not available commercially and was synthesized by Roman Belle. In brief, 0.1 mol of octylamine and 0.1 mol of acetic acid were added gently to 100 ml of diethyl ether on ice. The solution was gently stirred and cooled on dry ice to allow the salt to crystallize the diethyl ether was removed and the salt washed twice with 50 ml of hexane. For the MS ionization source parameters, the drying gas was set to 225 °C (protein intact mass determination) and 280 °C (hydroxylation assays) and a gas flow rate of 13 l/min. Ion chromatogram data were extracted and peak-area data for extracted ion chromatograms were integrated using RapidFire Integrator software (Agilent).

6.16 mCMAH assays

Activity assays for mCMAH were performed at room temperature and activity was measured using RapidFire High-throughput MS. Reactions were set up with a final volume of 1 ml. Buffer, substrate (CMP-Neu5Ac), cytochrome b_5 (cyt b_5, Abcam) and cyt b_5 reductase 3 (ProSpec) were

mixed and transferred to a 96-well deep well block (Greiner Bio-One). NADH was added and the no-enzyme measurement was recorded. Enzyme was added to initiate the reaction and measurements were taken every 120 s to follow the reaction progress. Concentrations of each reaction component are listed in **Table 6.18**.

Table 6.18: Standard reaction conditions for mCMAH activity assays.

Component	Concentration
HEPES, pH 7.4	20 mM
CMP-Neu5Ac, sodium salt	4 µM
Recombinant human cytochrome b_5	210 nM
Recombinant human cytochrome b_5 reductase 3	120 nM
NADH	100 µM
mCMAH	100 nM

6.17 GloB assays

6.17.1 Assay conditions

All reactions were run in 10 mM MOPS, pH 7.4. Enzyme concentrations were 4 nM and 2 nM for thioester hydrolysis and thiol detection assays respectively. Kinetic parameters were determined using varied substrate concentrations. **Table 6.19** shows amounts of each component in the reaction mixture for inhibition studies.

Table 6.19: Reaction compositions for GloB inhibition assays.

Component	Thioester hydrolysis assay	Thiol detection assay
GloB	5 µl of 20 nM stock	5 µl of 10 nM stock
SLG	5 µl of 1.6 mM stock	10 µl of 500 µM SLG + 800 µM DTNB mixture
Inhibitor	1 µl	1 µl
10 mM MOPS, pH 7.4	14 µl	9 µl

Inhibitor stocks were prepared in DMSO or Milli-Q water. 1 µl of DMSO or Milli-Q water was used in no-inhibitor control reactions. Enzyme was incubated with inhibitor for 10 min prior to addition of the substrate to initiate the reaction.

6.17.2 Measurements

Measurements were carried out at 25 °C using CLARIOstar or PHERAstar FS multi-mode plate readers (BMG Labtech). Thioester hydrolysis was measured at 234 nm while thiol detection by 5,5'-dithiobis-2-nitrobenzoic acid (DTNB) was measured at 413 nm. K_m assays were carried

out using UV-star 96 well format μclear transparent plates (Greiner Bio-One) with a final reaction volume of 100 μl. Inhibition assays were done in quadruplicates using UV-star 384 well format μclear transparent plates (Greiner Bio-One) with a final reaction volume of 25 μl.

6.17.3 Data analysis

Kinetic parameters and IC_{50} values were determined using GraphPad Prism 5 software[14]. Kinetic parameters were calculated by fitting initial rates to Michaelis-Menten equation. IC_{50} values were determined by fitting calculated residual activities (see formula below) to non-linear curves.

Residual activities were calculated using the following formula:

$$\%RA = \frac{V_{inhibited} - V_{background}}{V_{uninhibited} - V_{background}} \times 100.$$

6.18 Protein-protein interactions

6.18.1 Crosslinking and decrosslinking

HEK293 wild type (WT), hCMAH and mCMAH transfected cells were grown in T75 flasks to approximately 80-100% confluency. Cells were crosslinked by incubating with 1% formaldehyde for 10 min, 37 °C, 50 rpm. The reaction was quenched by adding glycine to final concentration of 0.125 M (5 min, 37 °C, 50 rpm), the medium was removed, and cells were washed with 10 ml cold PBS. Fresh PBS (10 ml) was added and cells were detached by scraping.

To revert the crosslinking reactions, samples were incubated with 312.5 mM NaCl at 37 °C overnight. Identity of interacting proteins was assessed by Western blot analysis.

6.18.2 Pulldowns

Cells (either crosslinked or just detached) were pelleted (500 g, 5 min, 4 °C), resuspended in 300 μl of mild lysis buffer (20 mM Tris HCl, 137 mM NaCl, 0.5% IGEPAL CA-630, pH 8.0, with protease inhibitor cocktail) and incubated on ice for 30 min vortexing every 10 min. Cell debris was pelleted (18000 x g, 5 min, 4 °C) and protein concentration in lysate was measured using the Bradford assay. Samples containing equivalent total protein concentrations were diluted to 1 ml in pulldown buffer (3.25 mM Sodium phosphate, 70 mM NaCl, 0.01% Tween-20, pH 7.4) and incubated with 50 μl of Dynabeads His-tag Isolation and Pulldown (Thermo Fisher Scientific), pulldown was performed at 4 °C according to manufacturer's manual. Proteins were eluted in 50 mM sodium phosphate, 300 mM NaCl, 300 mM imidazole, 0.01% Tween-20, pH 8.0. Identity of interacting proteins was checked by Western blot.

6.18.3 Co-immunoprecipitation

Cells (whether crosslinked or only detached) were pelleted (500 g, 5 min, 4 °C), rinsed with PBS and resuspended in lysis buffer (10 mM Tris, 10 mM NaCl, 2 mM EDTA, 0.5% Triton X-100, 1x protease inhibitor cocktail, pH 8.0) and incubated on ice for 10 min. NaCl (5 M stock) was added to final concentration of 150 mM and cell debris were pelleted (18000 x g, 15 min, 4 °C). Supernatants were incubated with 2 µg of rabbit anti-His_6 antibody at 4 °C overnight. Sample was then incubated with gentle agitation with 30 µl Dynabeads Protein G (Thermo Fisher Scientific) pre-equilibrated in wash buffer (50 mM Tris, 150 mM NaCl, 0.1% IGEPAL CA-630) at 4 °C for 2 h. Beads were washed with wash buffer 5 times and bound proteins were eluted in 200 mM glycine, pH 3.0. Identity of interacting proteins was checked by Western blot.

6.19 Immunofluorescence

6.19.1 Coverslip preparation

HeLa and TCMK-1 cells were used for immunofluorescence experiments. Glass coverslips (18 mm diameter) were placed in wells of a 12-well plate and 1 ml of 50,000 cells/ml cell suspension were added. Cells were grown to about 50% confluency (approximately 16-18 h) before transfection with hCMAH-pcDNA3.1 or mCMAH-pcDNA3.1 (as described in earlier), 50 µl of OptiMEM medium was used for wild type (WT) controls. Cells were incubated at 37 °C, 5% CO_2 for 3 h, then medium was removed, cells were washed with PBS (3 times) and 1 ml of fresh DMEM added. Cells were then incubated for further 21 h at 37 °C, 5% CO_2. For mitochondrial staining, 23.5 h after transfection cells were incubated with 1:1000 dilution of MitoTracker™ Red CMXRos (Thermo Fisher Scientific) in growth medium for 30 min at 37 °C.

The medium was then removed and cells were rinsed with filtered PBS (3 times). Cells were fixed by incubating with 3% paraformaldehyde dissolved in PBS for 20 min and rinsed with filtered PBS (3 times). Cell membranes were permeabilised with 0.1% Triton-X 100 in PBS for 4 min, washed with PBS (3 times, 5 min) and blocked with 0.2% (w/v) fish skin gelatine (FSG) dissolved in PBS for 30 min.

Coverslips were incubated cell-face down on 50 µl of primary antibody(-ies) diluted in 0.2% FSG in PBS overnight at 4 °C, then washed with 0.2% FSG in PBS (3 times, 5 min) and incubated with 1 ml of secondary antibody(-ies) diluted in 0.2% FSG in PBS for 30 min in the dark (antibodies used are listed in **Table 6.7**). Coverslips were washed with 0.2% FSG in PBS (3 times, 5 min) and cell nuclei were stained with 0.1 µg/ml 4',6-diamidino-2-phenylindole (DAPI) for 2 min and excess DAPI was washed out with 0.2% FSG in PBS (3 times, 5 min).

Coverslips were mounted on slides using ProLong™ Gold Antifade Mountant (Thermo Fisher Scientific). Samples were stored in the dark at 4 °C before data collection.

6.19.2 Image collection and processing

Images were collected at the Micron Advanced Bioimaging Unit (University of Oxford, UK) by Dr Joan Gannon using an inverted DVcore (Olympus) fluorescence microscope with 100X/1.35 objective. Resolve3D SoftWoRx-Acquire (Version 5.5.1, Release 3, Applied Precision Inc, Issaquah, Washington) software was used for image acquisition and deconvolution. Images were processed using Fiji-ImageJ software[15].

6.20 References

1. Shen, A.; Lupardus, P. J.; Morell, M.; Ponder, E. L.; Sadaghiani, A. M.; Garcia, K. C.; Bogyo, M., Simplified, Enhanced Protein Purification Using an Inducible, Autoprocessing Enzyme Tag. *PLoS One* **2009,** *4* (12), e8119.

2. Davis, M. W. *ApE- A plasmid Editor*, v2.0.49.10.

3. Fukunaga, R.; Han, Bo W.; Hung, J.-H.; Xu, J.; Weng, Z.; Zamore, Phillip D., Dicer Partner Proteins Tune the Length of Mature miRNAs in Flies and Mammals. *Cell* **2012,** *151* (3), 533-546.

4. O'Young, J.; Sukdeo, N.; Honek, J. F., *Escherichia coli* glyoxalase II is a binuclear zinc-dependent metalloenzyme. *Arch. Biochem. Biophys.* **2007,** *459* (1), 20-26.

5. Winn, M. D.; Ballard, C. C.; Cowtan, K. D.; Dodson, E. J.; Emsley, P.; Evans, P. R.; Keegan, R. M.; Krissinel, E. B.; Leslie, A. G. W.; McCoy, A.; McNicholas, S. J.; Murshudov, G. N.; Pannu, N. S.; Potterton, E. A.; Powell, H. R.; Read, R. J.; Vagin, A.; Wilson, K. S., Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **2011,** *67* (4), 235-242.

6. McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J., Phaser crystallographic software. *J. Appl. Crystallogr.* **2007,** *40* (4), 658-674.

7. Adams, P. D.; Afonine, P. V.; Bunkóczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L.-W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H., PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **2010,** *66* (Pt 2), 213-221.

8. Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K., Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **2010,** *66* (4), 486-501.

9. Hungler, A.; Momin, A.; Diederichs, K.; Arold, S., T.,, ContaMiner and ContaBase: a webserver and database for early identification of unwantedly crystallized protein contaminants. *J. Appl. Crystallogr.* **2016,** *49* (6), 2252-2258.

10. Simpkin, A. J.; Simkovic, F.; Thomas, J. M. H.; Savko, M.; Lebedev, A.; Uski, V.; Ballard, C.; Wojdyr, M.; Wu, R.; Sanishvili, R.; Xu, Y.; Lisa, M.-N.; Buschiazzo, A.; Shepard, W.; Rigden, D. J.; Keegan, R. M., SIMBAD: a sequence-independent molecular-replacement pipeline. *Acta Crystallogr. D Biol. Crystallogr.* **2018,** *74* (7), 595-605.

11. Boivin, S.; Kozak, S.; Meijers, R., Optimization of protein purification and characterization using Thermofluor screens. *Protein Expr. Purif.* **2013,** *91* (2), 192-206.

12. *OriginPro*, 2017; OriginLab Corporation: Northampton, Massachusetts, USA.

13. Rye, P. T.; LaMarr, W. A., Measurement of glycolysis reactants by high-throughput solid phase extraction with tandem mass spectrometry: Characterization of pyrophosphate-dependent phosphofructokinase as a case study. *Anal. Biochem.* **2015,** *482*, 40-47.

14. *GraphPad Prism*, v5.04; GraphPad Software: San Diego, California, USA.

15. Schindelin, J.; Arganda-Carreras, I.; Frise, E.; Kaynig, V.; Longair, M.; Pietzsch, T.; Preibisch, S.; Rueden, C.; Saalfeld, S.; Schmid, B.; Tinevez, J. Y.; White, D. J.; Hartenstein, V.; Eliceiri, K.; Tomancak, P.; Cardona, A., Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **2012,** *9* (7), 676-682.

Appendix 1: Protein Sequences

Molecular weights (MW) and extinction coefficients (ε) were calculated using ProtParam tool (https://web.expasy.org/protparam/). 3C protease cleavage sites are underlined.

hCMAH-pCold construct (MW = 60695.44 Da, ε = 115320 M^{-1} cm^{-1})

MNHKVHHHHHHIEGRHMELGTLEGSEFLEVLFQGPVDDENNGLLLLELNPPNPWDLQPRSPEELA
FGEVQITYLTHACMDLKLGDKRMVFDPWLIGPAFARGWWLLHEPPSDWLERLCQADLVYISHLHS
DHLSYPTLKKLAGRRPDIPIYVGNTERPVFWNLNQSGVQLTNINVVPFGIWQQVDKNLRFMILMDG
VHPEMDTCIIVEYKGHKILNTVDCTRPNGGRLPMKVALMMSDFAGGASGFPMTFSGGKFTEEWKA
QFIKTERKKLLNYKAWLVKNLQPRIYCPFAGYFVESHPSDKYIKETNTKNDPNELNNLIKKNSDVIT
WTPRPGATLDLGRMLKDPTDSKGIIEPPEGTKIYKDSWDFEPYLEILNAALGDEIFLHSSWIKEYFT
WAGFKDYNLVVRMIETDEDFNPFPGGYDYLVDFLDLSFPKERPQREHPYEEIHSRVDVIRHVVKNG
LLWDELYIGFQTRLQRDPDIYHHLFWNHFQIKLPLTPPNWKSFLMCCEQNGPAILQECKTT

hCMAH-pMAL construct (MW = 101045.10 Da, ε = 181670 M^{-1} cm^{-1})

MKIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHD
RFGGYAQSGLLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEEIP
ALDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFKYENGKYDIKDVGVDNAGAKAGLTFLVDLI
KNKHMNADTDYSIAEAAFNKGETAMTINGPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSA
GINAASPNKELAKEFLENYLLTDEGLEAVNKDKPLGAVALKSYEEELVKDPRIAATMENAQKGEIMP
NIPQMSAFWYAVRTAVINAASGRQTVDEALKDAQTNSSSNNNNNNNNNNLGIEGRISHMLEVLFQ
GPDENNGLLLLELNPPNPWDLQPRSPEELAFGEVQITYLTHACMDLKLGDKRMVFDPWLIGPAFAR
GWWLLHEPPSDWLERLCQADLVYISHLHSDHLSYPTLKKLAGRRPDIPIYVGNTERPVFWNLNQSG
VQLTNINVVPFGIWQQVDKNLRFMILMDGVHPEMDTCIIVEYKGHKILNTVDCTRPNGGRLPMKVA
LMMSDFAGGASGFPMTFSGGKFTEEWKAQFIKTERKKLLNYKAWLVKNLQPRIYCPFAGYFVESHP
SDKYIKETNTKNDPNELNNLIKKNSDVITWTPRPGATLDLGRMLKDPTDSKGIIEPPEGTKIYKDSW
DFEPYLEILNAALGDEIFLHSSWIKEYFTWAGFKDYNLVVRMIETDEDFNPFPGGYDYLVDFLDLSFP
KERPQREHPYEEIHSRVDVIRHVVKNGLLWDELYIGFQTRLQRDPDIYHHLFWNHFQIKLPLTPPN
WKSFLMCC EQNGPAILQE CKTTHHHHHH

hCMAH-pcDNA3.1 construct (MW = 58215.69 Da, ε = 115320 M^{-1} cm^{-1})

MDENNGLLLLELNPPNPWDLQPRSPEELAFGEVQITYLTHACMDLKLGDKRMVFDPWLIGPAFAR
GWWLLHEPPSDWLERLCQADLVYISHLHSDHLSYPTLKKLAGRRPDIPIYVGNTERPVFWNLNQSG
VQLTNINVVPFGIWQQVDKNLRFMILMDGVHPEMDTCIIVEYKGHKILNTVDCTRPNGGRLPMKVA

LMMSDFAGGASGFPMTFSGGKFTEEWKAQFIKTERKKLLNYKAWLVKNLQPRIYCPFAGYFVESHP
SDKYIKETNTKNDPNELNNLIKKNSDVITWTPRPGATLDLGRMLKDPTDSKGIIEPPEGTKIYKDSW
DFEPYLEILNAALGDEIFLHSSWIKEYFTWAGFKDYNLVVRMIETDEDFNPFPGGYDYLVDFLDLSFP
KERPQREHPYEEIHSRVDVIRHVVKNGLLWDELYIGFQTRLQRDPDIYHHLFWNHFQIKLPLTPPN
WKSFLMCCEQNGPAILQECKTTLEVLFQGPHHHHHH

mCMAH-pCold construct (MW = 71122.29 Da, ε = 121405 M^{-1} cm^{-1})

MNHKVHHHHHHIEGRHMELGTLEGSEFLEVLFQGPVDMDRKQTAETLLTLSPAEVANLKEGINFF
RNKTTGKEYILYKEKDHLKACKNLCKHQGGLFMKDIEDLDGRSVKCTKHNWKLDVSTMKYINPPGS
FCQDELVIEMDENNGLSLVELNPPNPWDSDPRSPEELAFGEVQITYLTHACMDLKLGDKRMVFDP
WLIGPAFARGWWLLHEPPSDWLERLCKADLIYISHMHSDHLSYPTLKQLSQRRPDIPIYVGDTERPV
FWNLDQSGVGLTNINVVPFGIWQQVDKSLRFMILMDGVHPEMDTCIIVEYKGHKILNTVDCTRPNG
GRLPEKVALMMSDFAGGASGFPMTFSGGKFTEEWKAQFIKAERRKLLNYKAQLVKDLQPRIYCPFA
GYFVESHPSDKYIKETNTKNDPNQLNNLIRKNSDVVTWTPRPGAVLDLGRMLKDPTDSKGIVEPPE
GTKIYKDSWDFGPYLEILNSAVRDEIFCHSSWIKEYFTWAGFKNYNLVVRMIETDEDFSPFPGGYDY
LVDFLDLSFPKERPSREHPYEEIHSRVDVIRYVVKNGLLWDDLYIGFQTRLLRDPDIYHHLFWNHFQ
IKLPLTPPNWKSFLMHCD

mCMAH-pcDNA3.1 construct (MW = 68642.54 Da, ε = 121405 M^{-1} cm^{-1})

MMDRKQTAETLLTLSPAEVANLKEGINFFRNKTTGKEYILYKEKDHLKACKNLCKHQGGLFMKDIE
DLDGRSVKCTKHNWKLDVSTMKYINPPGSFCQDELVIEMDENNGLSLVELNPPNPWDSDPRSPEEL
AFGEVQITYLTHACMDLKLGDKRMVFDPWLIGPAFARGWWLLHEPPSDWLERLCKADLIYISHMH
SDHLSYPTLKQLSQRRPDIPIYVGDTERPVFWNLDQSGVGLTNINVVPFGIWQQVDKSLRFMILMDG
VHPEMDTCIIVEYKGHKILNTVDCTRPNGGRLPEKVALMMSDFAGGASGFPMTFSGGKFTEEWKAQ
FIKAERRKLLNYKAQLVKDLQPRIYCPFAGYFVESHPSDKYIKETNTKNDPNQLNNLIRKNSDVVTW
TPRPGAVLDLGRMLKDPTDSKGIVEPPEGTKIYKDSWDFGPYLEILNSAVRDEIFCHSSWIKEYFTW
AGFKNYNLVVRMIETDEDFSPFPGGYDYLVDFLDLSFPKERPSREHPYEEIHSRVDVIRYVVKNGLL
WDDLYIGFQTRLLRDPDIYHHLFWNHFQIKLPLTPPNWKSFLMHCDLEVLFQGPHHHHHH

GloB-pET22b corrected sequence (MW = 28434.38 Da, ε = 28795 M^{-1} cm^{-1})

MNLNSIPAFDDNYIWVLNDEAGRCLIVDPGDAEPVLNAIAANNWQPEAIFLTHHHHDHVGGVKEL
VEKFPQIVVYGPQETQDKGTTQVVKDGETAFVLGHEFSVIATPGHTLGHICYFSKPYLFCGDTLFSGG
CGRLFEGTASQMYQSLKKLSALPDDTLVCCAHEYTLSNMKFALSILPHDLSINDYYRKVKELRAKNQ
ITLPVILKNERQINVFLRTEDIDLINVINEETLLQQPEERFAWLRSKKDRF

Appendix 2: Compounds Tested for GloB Inhibition

In addition to the 155 compounds listed below, 163 compounds obtained from the European Lead Factory were screened.

Plate AS 1:

Well	Compound	Classification	Source
A01	Penicillin G	penicillin (1st gen)	Sigma-Aldrich (Merck)
A02	Penicillin V	penicillin (1st gen)	LKT Labs
A03	Dicloxacillin	penicillin (2nd gen)	British Pharmacopoeia Reference Standard
A04	Flucloxacillin	penicillin (2nd gen)	Beecham Research Laboratories
A05	Oxacillin	penicillin (2nd gen)	Sigma-Aldrich (Merck)
A06	Ticarcillin	penicillin (4th gen)	Global Life Sciences
A07	Piperacillin	penicillin (4th gen)	MP Biomedicals
A08	Temocillin	penicillin (4th gen)	Carbosynth
A09	Carbenicillin	penicillin (4th gen)	BioVision
A10	Mecillinam	penicillin (extended spectrum)	Fluorochem
B01	Cefazolin	cephalosporin (1st gen)	Fluorochem
B02	Cephalexin	cephalosporin (1st gen)	Sigma-Aldrich (Merck)
B03	Cephradine C-I	cephalosporin (1st gen)	Sigma-Aldrich (Merck)
B04	Cephalothin	cephalosporin (1st gen)	TOKU-E
B05	Cephalosporin C	cephalosporin (1st gen)	Sigma-Aldrich (Merck)
B06	Cefminox	cephalosporin (2nd gen)	TOKU-E
B07	Moxolactam	cephalosporin (2nd gen)	Cayman Chemical Company
B08	Cefaclor	cephalosporin (2nd gen)	Sigma-Aldrich (Merck)
B09	Cefprozil	cephalosporin (2nd gen)	AK Scientific
B10	Ceforanide	cephalosporin (2nd gen)	USP
C01	Cefuroxime	cephalosporin (2nd gen)	Abcam
C02	Cefoxitin	cephalosporin (2nd gen)	Sigma-Aldrich (Merck)
C03	Cefmetazole	cephalosporin (2nd gen)	TOKU-E
C04	Cefdinir C-I	cephalosporin (3rd gen)	Sigma-Aldrich (Merck)
C05	Cefditoren	cephalosporin (3rd gen)	TOKU-E
C06	Ceftazidime	cephalosporin (3rd gen)	TOKU-E
C07	Ceftriaxone	cephalosporin (3rd gen)	Cayman Chemical Company
C08	Cefotaxim	cephalosporin (3rd gen)	Sigma-Aldrich (Merck)
C09	Ceftiofur	cephalosporin (3rd gen)	Molekula
C10	Cefepime	cephalosporin (4th gen)	Alfa Aesar
D01	Ceftaroline fosamil	cephalosporin (5th gen)	Teflaro
D02	7-aminocephalosporanic acid	cephalosporin (precursor)	Sigma-Aldrich (Merck)
D03	Biapenem	carbapenem	Global Life Sciences

Well	Compound	Classification	Source
D04	Doripenem	carbapenem	Global Life Sciences
D05	Imipenem	carbapenem	Molekula
D06	Meropenem	carbapenem	Combi-blocks
D07	Tebipenem	carbapenem	Ark Pharm
D08	Panipenem	carbapenem	Ontario Chemicals
D09	Faropenem	penem	Fluorochem
D10	Aztreonam	monobactam	Molekula
E01	Bis(methylamino)cyclobut-3-ene-1,2-dione (Sq1)	squaramide	Enamine
E02	3,4-dimethoxy-3-cyclobutene-1,2-Dione (Sq2)	squaramide	Sigma-Aldrich (Merck)
E03	3,4-bis(butylamino) cyclobut-3-ene-1,2-dione (Sq4)	squaramide	Sigma-Aldrich (Merck)
E04	3,4-Dibutoxycyclobut-3-ene-1,2-dione (Sq5)	squaramide	Sigma-Aldrich (Merck)
E05	Sq6 (CAS: 1256245-79-0)	squaramide	AChemBlock
E06	CXCR2 Antagonist IV (Sq7)	squaramide	Sigma-Aldrich (Merck)
E07	ESC1000274	squaramide	European Lead Factory
E08	ESC1000323	squaramide	European Lead Factory
E09	OP-0595	beta-lactamase inhibitor	Activate Scientific
E10	Avibactam	beta-lactamase inhibitor	Sigma-Aldrich (Merck)
F01	Tazobactam	beta-lactamase inhibitor	Sigma-Aldrich (Merck)
F02	Sulbactam	beta-lactamase inhibitor	Molekula
F03	Clavulanic acid (potassium clavulanate:cellulose)	beta-lactamase inhibitor	Sigma-Aldrich (Merck)
F04	GC-01-27A	indole acylhydrazine	CJS Compound library*
F05	GC-03-213BB	indole tetrazole	CJS Compound library*
F06	GC-04-299B	indole tetrazole	CJS Compound library*
F07	GC-04-308B	indole hydroxamate	CJS Compound library*
F08	CL-05-613-4-1	other	CJS Compound library*
F09	AR 413	boronate	CJS Compound library*
F10	AK-4	boronate	Alen Krajnc
G01	AK-6-6	boronate	Alen Krajnc
G02	AK-29	boronate	Alen Krajnc
G03	AK-46	boronate	Alen Krajnc
G04	AK-61	boronate	Alen Krajnc
G05	AK-63	boronate	Alen Krajnc
G06	AK-91	boronate	Alen Krajnc
G07	AK-110	boronate	Alen Krajnc
G08	AK-131	boronate	Alen Krajnc
G09	IS-41	hydroxamate	CJS Compound library*
G10	IS-42	hydroxamate	CJS Compound library*

Well	Compound	Classification	Source
H01	IS-43	hydroxamate	CJS Compound library*
H02	IS-47	hydroxamate	CJS Compound library*
H03	IS-99	hydroxamate	CJS Compound library*
H04	LBE-273-5-42	hydroxamate	CJS Compound library*
H05	LBE-322-5-93	hydroxamate	CJS Compound library*
H06	LBE-332-6-03	hydroxamate	CJS Compound library*
H07	Ebselen	antioxidant	Sigma-Aldrich (Merck)
H08	3-hydroxy-1,2-dimethylpyridine-4(1H)-one	Fe chelator	Fluorochem
H09	Deferasirox	Fe chelator	BioNet
H10	Deferoxamine	Fe chelator	Sigma-Aldrich (Merck)

Plate AS 2:

Well	Compound	Classification	Source
A01	Captopril		Sigma-Aldrich (Merck)
A02	SUB04-16a	captopril derivative	CJS Compound library*
A03	SUB04-16b	captopril derivative	CJS Compound library*
A04	C2	captopril derivative	CJS Compound library*
A05	RH01-026 (L-Cys)	thiol	CJS Compound library*
A06	RH01-066	thiol	CJS Compound library*
A07	RH01-029 (L-Cys)	thiol	CJS Compound library*
A08	RH01-038 (D-Cys)	thiol	CJS Compound library*
A09	RH01-064	thiol	CJS Compound library*
A10	LBE-435-7-02	thiol	CJS Compound library*
B01	LBE-443-7-10	thiol	CJS Compound library*
B02	LBE-370-6-41	thiol	CJS Compound library*
B03	LBE-111-3-83	thiol	CJS Compound library*
B04	LBE-521-7-90-8	thiol	CJS Compound library*
B05	LBE-527-7-96	thiol	CJS Compound library*
B06	LBE-465-7-32	thiol	CJS Compound library*
B07	Cysteamine HCl	thiol	Sigma-Aldrich (Merck)
B08	L-glutathione	glutahione	Sigma-Aldrich (Merck)
B09	6-aminopenicillanic acid	penicillin (core)	Sigma-Aldrich (Merck)
B10	folic acid	metal chelator	Sigma-Aldrich (Merck)
C01	N-hydroxymethyl-phthalimide	bicyclic	Sigma-Aldrich (Merck)
C02	2,3-PDA	pyridine carboxylate	Sigma-Aldrich (Merck)
C03	2,4-PDA	pyridine carboxylate	Sigma-Aldrich (Merck)
C04	2,5-PDA	pyridine carboxylate	Sigma-Aldrich (Merck)
C05	3,4-PDA	pyridine carboxylate	Sigma-Aldrich (Merck)
C06	BRW-BG-024-01	pyridine carboxylate	Dr Lennart Brewitz
C07	BRW-BG-044-01	pyridine carboxylate	Dr Lennart Brewitz

Well	Compound	Classification	Source
C08	BRW-BG-086-01	pyridine carboxylate	Dr Lennart Brewitz
C09	BMRL08030	pyridine carboxylate	CJS Compound library*
C10	p-aminosalicylic acid	salicylic acid	Sigma-Aldrich (Merck)
D01	salicylic acid	salicylic acid	Sigma-Aldrich (Merck)
D02	125370014	Misc metal chelator	Otava Chemicals
D03	109850011	Misc metal chelator	Otava Chemicals
D04	7217890083	Misc metal chelator	Otava Chemicals
D05	1045869	Misc metal chelator	Otava Chemicals
D06	1779335	Misc metal chelator	Otava Chemicals
D07	1794806	Misc metal chelator	Otava Chemicals
D08	1905597	Misc metal chelator	Otava Chemicals
D09	2917417	Misc metal chelator	Otava Chemicals
D10	4215808	Misc metal chelator	Otava Chemicals
E01	4802103	Misc metal chelator	Otava Chemicals
E02	6238583	Misc metal chelator	Otava Chemicals
E03	6328802	Misc metal chelator	Otava Chemicals
E04	4-hydroxyphenyl-pyruvic acid	carboxylic acid	Sigma-Aldrich (Merck)
E05	Sparfloxacin	antibiotic	Sigma-Aldrich (Merck)
E06	Ofloxacin	antibiotic	Sigma-Aldrich (Merck)
E07	Gemifloxacin mesylate	quinone antibiotic	LKT Laboratories
E08	Norfloxacin	antibiotic	Sigma-Aldrich (Merck)
E09	-	-	-
E10	-	-	-
F01	*ATP*	*nucleoside*	Sigma-Aldrich (Merck)
F02	*dATP*	*nucleoside*	Sigma-Aldrich (Merck)
F03	*cAMP*	*nucleotide*	Sigma-Aldrich (Merck)
F04	*CGMP*	*nucleotide*	Sigma-Aldrich (Merck)
F05	*NADH*	*nucleotide*	Sigma-Aldrich (Merck)
F06	*SAM*	*nucleotide*	Sigma-Aldrich (Merck)
F07	*Tartrate*	*carboxylic acid*	Sigma-Aldrich (Merck)
F08	*Oxamate*	*carboxylic acid*	Sigma-Aldrich (Merck)
F09	*allylmalonic acid*	*carboxylic acid*	Alfa Aesar
F10	*malonate*	*carboxylic acid*	Sigma-Aldrich (Merck)
G01	*Fumarate*	*carboxylic acid*	Sigma-Aldrich (Merck)
G02	*6-aminocaproic acid*	*carboxylic acid/amine*	Sigma-Aldrich (Merck)
G03	*Citrate*	*carboxylic acid/TCA intermediate*	Sigma-Aldrich (Merck)
G04	*succinate*	*carboxylic acid/TCA intermediate*	Sigma-Aldrich (Merck)
G05	*pyruvate*	*carboxylic acid/TCA intermediate*	Sigma-Aldrich (Merck)
G06	*a-ketoglutarate*	*carboxylic acid/TCA intermediate*	Sigma-Aldrich (Merck)

Well	Compound	Classification	Source
G07	*Citrate*	*carboxylic acid/TCA intermediate*	Sigma-Aldrich (Merck)
G08	*D-2-hydroxyglutarate*	*TCA intermediate*	Sigma-Aldrich (Merck)
G09	*N-acetylglycine*	*amino acid*	Sigma-Aldrich (Merck)
G10	*N-acetyl-L-glutamic acid*	*amino acid*	Sigma-Aldrich (Merck)
H01	*L-glutamate*	*amino acid*	Sigma-Aldrich (Merck)
H02	*ascorbate*	*antioxidant*	Sigma-Aldrich (Merck)
H03	*succinamide*	*amide*	Sigma-Aldrich (Merck)
H04	*Streptomycin*	*antibiotic sugar*	Sigma-Aldrich (Merck)
H05	*Dopamine HCl*	*monoamine*	Sigma-Aldrich (Merck)
H06	*4-nitrophenyl phosphate*	*dopamine derivative*	Molekula
H07	*4-nitrocatechol sulfate*	*dopamine derivative*	Sigma-Aldrich (Merck)
H08	-	-	-
H09	-	-	-
H10	-	-	-

Columns 11 and 12 were filled with 100% DMSO (or Milli-Q water). PDA – pyridine dicarboxylic acid.

*Compounds were synthesised by previous members of the group.

Compounds dissolved in water are listed in italics, Milli-Q water was used in wells 11 and 12 in corresponding rows.