Structural studies of peptide binding to chicken class I MHC molecules, the CD6-CD166 interaction and enzymes involved in bacterial sulfur metabolism

Paul Elliott Chappell



Sir William Dunn School of Pathology
and
Brasenose College

A thesis submitted to the University of Oxford for the degree of Doctor of Philosophy in Pathology

Trinity Term 2015

Acknowledgements

Firstly, I'd like to express my sincere thanks to my supervisor, Susan Lea, for her continual encouragement, guidance and interesting discussions throughout my DPhil.

I'd like to also thank Marion Brown and Lee Garner for our collaboration on the CD6/CD166 project. Thanks to Jim Kaufman for involving me with the MHC project and the many highly interesting discussions. I would like to thank Ben Berks and Daniel Grabarczyk for our collaboration on the SoxBYZ and TsdA projects.

Special mention must go to Pietro Roversi who I spent countless hours with at the Diamond Light Source collecting data and learning about X-ray crystallography. Thank you to Steve Johnson for his expert technical assistance, discussions and for kindly reviewing chapters of this thesis.

Thanks also go to the BBSRC for funding my 4 year PhD departmental studentship at the Sir William Dunn School of Pathology.

Thanks to everyone in the Lea lab, past and present. Especially Matthijs Jore and Matthew Cottee for providing the table tennis opposition on our breaks. I must also thank my undergraduate college tutor, André Furger, for guiding me through my undergraduate degree at Pembroke College and introducing me to the fascinating world of Biochemistry.

Finally, I'd like to thank my family for all of their support over the years.

Abstract

T cell activation is critical for the initiation of the adaptive immune response against both pathogens and malignant cells. This process depends on a complex set of finely tuned interactions, not only between the T cell receptor (TCR) and peptide-MHC complexes, but also between numerous accessory molecules. Chapter 2 of this thesis examines CD6, a transmembrane glycoprotein predominantly found on the surface of T cells, and its interaction with CD166, another transmembrane glycoprotein found on the surface of antigen presenting cells (APCs), amongst others. This interaction is required for optimal T cell activation. I have solved the X-ray crystal structures of the three extracellular SRCR domains of CD6 and the two membrane distal IgSF domains of CD166. In collaboration with colleagues, the ligand binding sites on both CD6 and CD166 have been further defined. We have also shown that a single nucleotide polymorphism (SNP) in CD6 caused glycosylation, hindering the CD6/CD166 interaction. Finally, native mass spectrometry revealed competition between the heterophilic CD6/CD166 interaction and the homophilic CD166 interaction.

Chapter 3 investigates the mechanism of peptide binding to chicken class I major histocompatibility complex (MHC) molecules. These molecules are expressed by almost every cell type and are constantly sampling the cellular environment, binding peptide fragments in the endoplasmic reticulum (ER) and presenting them on the cell surface to T cells. The chicken MHC features a single, dominantly expressed class I MHC molecule. For some haplotypes, sequencing of the peptides bound to this molecule reveals an obvious peptide-binding motif. However, the dominantly expressed class I molecules from the B2, B14 and B21 haplotypes appear to be more promiscuous in their peptide-binding capabilities, with peptide sequencing revealing no obvious peptide-binding motif. I have solved a number of X-ray crystal structures of these promiscuous class I MHC molecules bound to different peptides. These structures have extended our understanding of the dominantly expressed class I MHC molecule of the B21 haplotype and revealed the mechanism of promiscuous peptide binding for the dominantly expressed molecules of the B2 and B14 haplotypes.

The final two chapters focus on a number enzymes involved in bacterial sulfur metabolism. Chapter 4 investigates the interaction between the carrier protein and a partner enzyme of the Sox pathway, an important route for the oxidation of inorganic sulfur compounds. Solution of the X-ray crystal structure of the T. thermophilus SoxYZ carrier protein in complex with T. thermophilus SoxB, a partner enzyme, revealed the molecular details of the interaction and offered insight into how SoxYZ is able to make specific, yet promiscuous interactions with other partner enzymes. Chapter 5 presents the X-ray crystal structure of Allochromatium vinosum thiosulfate dehydrogenase (TsdA), an enzyme that catalyses the oxidation of two thiosulfate molecules to form tetrathionate. This structure confirmed an unusual cysteine-ligated heme, identified a thiosulfate binding site in the active site cavity and revealed an electron transfer route. TsdA is homologous to the SoxAX enzyme of the Sox pathway and this structure offered insight into its mechanism.

Publications resulting from this thesis

Chappell, P.E., Garner, L.I., Yan, J., Metcalfe, C., Hatherley, D., Johnson, S., Robinson, C.V., Lea, S.M., and Brown, M.H. (2015). Structures of CD6 and Its Ligand CD166 Give Insight into Their Interaction. Struct. Lond. Engl. 1993 23, 1426–1436.

Chappell, P., Meziane, E.K., Harrison, M., Magiera, Ł., Hermann, C., Mears, L., Wrobel, A.G., Durant, C., Nielsen, L.L., Buus, S., et al. (2015). Expression levels of MHC class I molecules are inversely correlated with promiscuity of peptide binding. eLife 4, e05345.

Grabarczyk, D.B., Chappell, P.E., Johnson, S., Stelzl, L.S., Lea, S.M., and Berks, B.C. (2015). Structural basis for specificity and promiscuity in a carrier protein/enzyme system from the sulfur cycle. Proc. Natl. Acad. Sci. U.S.A. 112, E7166–E7175.

Grabarczyk, D.B., Chappell, P.E., Eisel, B., Johnson, S., Lea, S.M., and Berks, B.C. (2015). Mechanism of thiosulfate oxidation in the SoxA family of cysteine-ligated cytochromes. J. Biol. Chem. 290, 9209–9221.

Abbreviations

ABC	ATP-binding cassette
ACP	Acyl carrier protein
AIDS	Acquired immune deficiency syndrome
ALCAM	Activated leukocyte cell adhesion molecule
APBS	Adaptive Poisson-Boltzmann solver
ATP	Adenosine triphosphate
CD	Cluster of differentiation
CHO	Chinese hamster ovary
CNX	Calnexin
CRT	Calreticulin
CV	Column volume
Da	Daltons
DAMP	Danger-associated molecular pattern
DC	Dendritic cell
DMEM	Dulbecco's Modified Eagle's Medium
DNA	Deoxyribonucleic acid
DRiPs	Defective ribosomal products
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
EM	Electron microscopy
ER	Endoplasmic reticulum
ERAAP	Endoplasmic reticulum aminopeptidase associated with antigen processing
FAS	Fatty acid synthase
FCS	Fetal calf serum
FFAS	Fold & function assignment
GalNAc	N-acetylgalactosamine
GPCR	G-protein-coupled receptor
GS	Glutamine synthetase
GVHD	Graft versus host disease
HBS	HEPES buffered saline
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HIV	Human immunodeficiency virus
HLA	Human leukocyte antigen

IgSF	Immunoglobulin superfamily
ITAM	Immunoreceptor tyrosine-based activation motif
ITC	Isothermal titration calorimetry
Kd	Dissociation constant
LC-MS-MS	Liquid chromatography tandem mass spectrometry
M	Molar
mAb	Monoclonal antibody
MALS	Multi angle light scattering
MHC	Major histocompatibility complex
mRNA	Messenger RNA
MS	Multiple sclerosis or Mass spectrometry
MS	Mass spectrometry
MWCO	Molecular weight cut off
NBD	Nucleotide binding domain
NiNTA	Nickel nitrilotriacetic acid
NK cell	Nautral killer cell
NKT cell	Natural killer T cell
NMR	Nuclear magnetic resonance
PAGE	Polyacrylamide gel electrophoresis
PAMP	Pathogen-associated molecular pattern
PDB	Protein data bank
PEG	Poly ethylene glycol
PISA	Proteins, interfaces, sturctures and assemblies server
PLC	Peptide-loading complex
PRR	Pattern recognition receptor
PSI-BLAST	Position-specific iterative basic local alignment search tool
RAG	Recombination-activating gene
RCSB	Research Collaboratory for Structural Bioinformatics
RNA	Ribonucleic acid
rpm	Revolutions per minute
S2	Schneider 2
SDS-PAGE	Sodium dodecyl sulfate - polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
SNP	Single nucleotide polymorphism
SPR	Surface plasmon resonance

SRCR	Scavenger receptor cysteine-rich
TAP	Transported associated with antigen processing
TCR	T cell receptor
TCS	Tissue culture supernatant
tRNA	Transfer RNA
UGT1	UDP-glucose glycoprotein transferase I
UV	Ultraviolet
WT	Wild type

Table of Contents

Chapter 1: Introduction 25

1.1 Structure determination to atomic resolution 28

1.2 Thesis aims 35

Chapter 2: CD6 and its ligand, CD166. 37

2.1 Introduction 37

2.1.1 The immune system 37

2.1.2 CD6, a group B SRCR superfamily member 42

2.1.3 Activated leukocyte cell adhesion molecule (ALCAM) / CD166 44

2.1.4 The interaction between CD6 and CD166 45

2.1.5 CD6 single nucleotide polymorphisms (SNPs) 46

2.1.6 Therapeutic antibodies 47

2.2 Aims 47

2.3 Protein expression & purification 48

2.3.1 Plasmid, cloning & stable cell lines 48

2.3.2 Expression and purification 49

2.3.3 Multi angle light scattering (MALS) 51

2.4 Crystallisation, data collection and processing 53

2.4.1 Deglycosylation 53

2.4.2 Crystallisation 54

2.4.3 Data collection 56

2.4.4 CD6 data processing 56

2.4.5 CD166 data processing 58

2.5 Structure determination and refinement 60

2.5.1 CD6 structure determination 60

2.5.2 CD6 model building and refinement 61

2.5.3 CD166 structure determination 62
2.5.4 CD166 model building & refinement 63
2.6 A low resolution CD6 dataset 64
2.7 The Structures of Human CD6 & CD166 66
2.7.1 Human CD6 66
2.7.2 Human CD166 67
2.8 Mutagenesis of binding sites 69
2.9 Cross-species binding 71
2.10 Electrostatic analysis of binding sites 73
2.11 CD6 single nucleotide polymorphisms 74
2.12 Native Mass Spectrometry 77
2.13 Discussion 79
2.14 Summary 83
2.14.1 Future directions 83
Chapter 3: Chicken MHC class I molecules 85
3.1 Introduction 85
3.2 Aims 96
3.3 Dominantly expressed class I MHC molecule of the B21 haplotype 98
3.3.1 Crystallisation 99
3.3.2 Data collection 99
3.3.3 Data processing 100
3.3.4 Structure determination 101
3.3.5 Model building and refinement 102
3.3.6 Structural analysis 103
3.3.7 Thermostability of BF2*2101-peptide complexes 106
3.4 Dominantly expressed class I MHC molecule of the B2 haplotype 109

3.4.1 Crystallisation 110
3.4.2 Data collection 110
3.4.3 Data processing 111
3.4.4 B2 peptide D structure determination 113
3.4.5 B2 peptide C structure determination 114
3.4.6 Model building and refinement 116
3.4.7 Structural analysis 117
3.5 Dominantly expressed class I MHC molecule of the B14 haplotype 120
3.5.1 Crystallisation 120
3.5.2 Data collection 121
3.5.3 Data processing 121
3.5.4 B14 peptide F structure determination 122
3.5.5 Model building and refinement 123
3.6 Discussion 126
3.6.1 Future work 128
Chapter 4: The structure of SoxYZ in complex with SoxB 129
4.1 Introduction 129
4.1.1 Carrier proteins in metabolic pathways 129
4.1.2 The Sox pathway 130
4.2 Aims 137
4.3 Initial crystallisation, data collection and processing 138
4.3.1 Initial crystallisation of SoxBYZ 139
4.3.2 Data collection 140
4.3.3 Data processing 140
4.4 Initial structure determination 142
4.4.1 Structure determination 142

4.5 Final crystallisation, data collection and processing 146

4.5.1 Crystallisation of SoxBYZ 146

4.5.2 Final data collection 146

4.5.3 Data processing 147

4.6 Final structure determination and refinement 149

4.6.1 Structure determination 149

4.6.2 Model building and refinement 151

4.7 Structure of SoxBYZ 154

4.7.1 Y-patch interactions 156

4.7.2 Z-patch interactions 159

4.7.3 Conformational changes in SoxB mobile loop 161

4.7.4 Models of manganese ion coordination at SoxB active site 163

4.8 Discussion 165

4.8.1 Future work 168

Chapter 5: The structure of *Allochromatium vinosum* TsdA 169

5.1 Introduction 169

5.2 Aims 175

5.3 Native crystallisation, data collection and processing 176

5.3.1 Crystallisation of native TsdA 176

5.3.2 Data collection 176

5.3.3 Data processing 177

5.4 Native structure determination & refinement 178

5.5 Tetrathionate co-crystallisation of TsdA 181

5.5.1 Data collection 181

5.5.2 Data processing 182

5.6 TsdA-tetrathionate structure determination, refinement & model validation 183

5.6.1 Structure determination 183

5.6.2 Model building & refinement 184

5.7 Structural analysis 185

5.7.1 TsdA active site 188

5.7.2 Anomalous TsdA tetrathionate dataset 191

5.8 Discussion 198

5.9 Future work 202

Chapter 6: Appendix 203

6.1 X-ray crystallography 203

6.1.1 Protein crystals and diffraction 203

6.1.2 Data collection 206

6.1.3 Structure solution and generation of electron density maps 209

6.1.4 Refinement 213

References 215

List of Tables

Table 2.1 - Commercial crystallisation screens 55
Table 2.2 - Data collection parameters for CD6 and CD166 56
Table 2.3 – CD6 data processing statistics 57
Table 2.4 - CD166 data processing statistics 58
Table 2.5 – CD6 cell content analysis 60
Table 2.6 – CD6 molecular replacement 61
Table 2.7 - CD6 refinement statistics 62
Table 2.8 - CD166 cell content analysis 62
Table 2.9 – CD166 molecular replacement 63
Table 2.10 - CD166 refinement statistics 63
Table 2.11 - Low resolution CD6 data collection 64
Table 2.12 - Low resolution CD6 data processing 64
Table 2.13 - Low resolution CD6 molecular replacement 65
Table 2.14 - Native mass spectrometry calculated masses 78
Table 3.1 - Comparison of chicken and mammalian MHC 92
Table 3.2 – B21 peptide F data collection 100
Table 3.3 - B21 peptide F data processing statistics 100
Table 3.4 - B21 peptide F cell content analysis 101
Table 3.5 - B21 peptide F molecular replacement 101
Table 3.6 - B21 refinement statistics 102
Table 3.7 - B21 melting temperatures 106
Table 3.8 – B2 peptide D and C data collection 110
Table 3.9 - B2 peptides D and C data processing statistics 112
Table 3.10 – B2 peptide D cell content analysis 113
Table 3.11 - B2 peptide D molecular replacement 114
Table 3.12 - B2 peptide C cell content analysis 114
Table 3.13 - B2 peptide C self-rotation function peaks 115
Table 3.14 - B2 peptide C molecular replacement 115
Table 3.15 - B2 peptides D & C refinement statistics 116
Table 3.16 - B14 peptide F data collection 121
Table 3.17 - B14 peptide F data processing statistics 122
Table 3.18 - B14 peptide F cell content analysis 122
Table 3.19 - B14 peptide F molecular replacement 123
Table 3.20 - B14 peptide F refinement statistics 123

Table 4.1 - Initial data collection 140
Table 4.2 - Initial data processing 141
Table 4.3 - Potential molecular replacement search models 142
Table 4.4 - Cell content analysis of initial dataset 142
Table 4.5 - Top self-rotation function peaks of initial dataset 143
Table 4.6 - Molecular replacement hits with SoxB model in initial dataset 144
Table 4.7 - Molecular replacement hits with refined SoxB model in initial dataset 145
Table 4.8 - Molecular replacement hits with SoxYZ model in initial dataset 145
Table 4.9 - Data collection parameters for final dataset 147
Table 4.10 - Data processing statistics for final dataset 148
Table 4.11 - Cell content analysis of final dataset 149
Table 4.12 - Self-rotation function peaks for final dataset 150
Table 4.13 - Molecular replacement hits with SoxBYZ model in final dataset 151
Table 4.14 - SoxBYZ refinement statistics 152
Table 4.15 – Collaborators' ITC binding experiments 158
Table 5.1 - Initial data collection 177
Table 5.2 - Data processing 177
Table 5.3 - Cell content analysis 178
Table 5.4 - Top self-rotation function peaks 179
Table 5.5 - Initial molecular replacement scores 180
Table 5.6 - Refinement statistics 181
Table 5.7 - Initial data collection 182
Table 5.8 - Data processing 182
Table 5.9 - Cell content analysis 183
Table 5.10 - Molecular replacement with tetrathionate dataset 184
Table 5.11 - Refinement statistics 184
Table 5.12 - Long wavelength anomalous data collection 191
Table 5.13 - Long wavelength anomalous data processing 192
Table 5.14 - Kinetic analysis of TsdA mutants 196

List of Figures

Figure 2.1 - Schematic representation of CD6. 43
Figure 2.2 - Schematic representation of ALCAM / CD166 44
Figure 2.3 – Schematic representation of the CD6/CD166 heterophilic interaction 46
Figure 2.4 - Construct Sequences 49
Figure 2.5 - Purification of CD6 and CD166 50
Figure 2.6 - MALS analysis of recombinant proteins 53
Figure 2.7 – Deglycosylation of CD6 and CD166 54
Figure 2.8 - CD6 and CD166 crystals 55
Figure 2.9 - CD6 data processing analysis 57
Figure 2.10 - CD6 systematic absences 57
Figure 2.11 – CD166 data processing analysis 58
Figure 2.12 - CD166 dataset systematic absences 59
Figure 2.13 - CD6 structure 66
Figure 2.14 - CD166 structure 68
Figure 2.15 - Homophilic CD166 Crystal Contact 69
Figure 2.16 - CD6 Mutagenesis 70
Figure 2.17 - Cross species binding of CD6 to human CD166 72
Figure 2.18 - Electrostatic analysis of binding sites 73
Figure 2.19 - CD6 non-synonymous single nucleotide polymorphisms 74
Figure 2.20 - The CD6 SNP S351N reduces affinity for CD166 75
Figure 2.21 - Glycosylation of the CD6 SNP, S351N 76
Figure 2.22 - Native Mass Spectrometry 77
Figure 3.1 - Structure of murine MHC class I in complex with peptide. 89
Figure 3.2 - Peptide binding motifs 90
Figure 3.3 - Cell surface expression levels of the dominantly expressed class I molecules from different chicken haplotypes 95
Figure 3.4 - Promiscuous peptide binding by BF2*2101 95
Figure 3.5 - B2, B14 & B21 sequence alignment 97
Figure 3.6 - B21 peptide sequences 98
Figure 3.7 – B21 peptide F data processing analysis 101
Figure 3.8 - B21 peptide binding 104
Figure 3.9 - Thermal stability data with Boltzmann sigmoidal fits 107
Figure 3.10 - B21 peptide complex melting temperatures 108
Figure 3.11 - Order of MHC peptide complex thermostability 108

Figure 3.12 - B2 peptide sequences 109
Figure 3.13 - L test for acentric data (B2 peptide D) 111
Figure 3.14 - B2 peptides D and C data processing analysis 112
Figure 3.15 – B2 peptide D self-rotation function 113
Figure 3.16 - B2 peptide C self-rotation function 115
Figure 3.17 – Overall structure comparison of BF2*0201 to BF*2101 117
Figure 3.18 - B2 peptide binding 118
Figure 3.19 - B14 peptide sequences 120
Figure 3.20 - B14 peptide F data processing analysis 122
Figure 3.21 - Comparison of BF2*0201, BF2*1401 & BF2*2101 structures 124
Figure 3.22 - B14 peptide binding 125
Figure 4.1 - Sox pathway model 132
Figure 4.2 - PAGE & SEC-MALS analysis of SoxBYZ complex 134
Figure 4.3 - Binding of SoxB to SoxYZ derivatives 136
Figure 4.4 - Analysis of disulfide-linked SoxBYZ complex 138
Figure 4.5 - Initial SoxBYZ crystal 139
Figure 4.6 - Initial data processing analysis 141
Figure 4.7 - Systematic absences present in initial dataset 141
Figure 4.8 - Self-rotation function peaks in initial dataset 143
Figure 4.9 - Structural alignment of three placed copies of SoxBYZ 146
Figure 4.10 - SoxBYZ crystal 147
Figure 4.11 - Data processing analysis of final dataset 148
Figure 4.12 - Self-rotation function for final dataset 150
Figure 4.13 - Polygon analysis of final refinement 153
Figure 4.14 – Overall SoxBYZ structure 154
Figure 4.15 - Buried surfaces and CONSURF analysis 155
Figure 4.16 - T. thermophilus and P. pantotrophus SoxYZ backbone alignment 156
Figure 4.17 - SoxB-SoxY interface 157
Figure 4.18 - SoxB-SoxZ interface 160
Figure 4.19 - Electrostatic analysis of Z-patch 161
Figure 4.20 - Conformational changes in SoxB mobile loop 162
Figure 4.21 - SoxB side chain conformational alterations 163
Figure 4.22 - Models of manganese ion coordination 164
Figure 4.23 - SoxYZ interacting surfaces of partner enzymes 167
Figure 5.1 - Sequence alignment showing homology between TsdA and SoxA 170

Figure 5.2 - Comparison of the proposed reactions catalysed by SoxAX and TsdA . 170
Figure 5.3 - Modification of Cys123 at pH 8.0 172
Figure 5.4 - Reactivity of TsdA with thiosulfate and tetrathionate at pH 4.5 174
Figure 5.5 - Data processing analysis 178
Figure 5.6 - Self-rotation function 179
Figure 5.7 - Tetrationate soaked TsdA protein crystal 181
Figure 5.8 - TsdA data processing analysis 183
Figure 5.9 - Structure of TsdA and comparison with SoxAX 185
Figure 5.10 - Accessibility and environment of heme groups in TsdA 187
Figure 5.11 - Electron density at Cys123 189
Figure 5.12 - TsdA-tetrathionate active site 190
Figure 5.13 - Residual log likelihood gradient map 193
Figure 5.14 - TsdA and SoxA active site comparison 194
Figure 5.15 - Visible spectra of TsdA active site mutants 196
Figure 5.16 - ESI-MS of TsdA active site mutants 197
Figure 5.17 - Proposed reaction mechanism 200
Figure 5.18 – Active site electron density observed by Brito et al. 201
Figure 6.1 - Phase diagram showing crystal growth by vapour diffusion 203
Figure 6.2 - Unit cell dimensions 204
Figure 6.3 - Bragg's law 205
Figure 6.4 - Lattice planes and Miller indices 206
Figure 6.5 - Molecular replacement process 212
Figure 6.6 - Self-rotation function illustration 213

Chapter 1: Introduction

Proteins are a very diverse group of molecules, involved in almost all cellular activities and processes. Many facilitate important cellular functions including DNA replication, metabolism and pathogen recognition (Call and Wucherpfennig, 2005; Hübscher et al., 2002; Printz et al., 1993), while others, such as cytoskeletal and viral capsid proteins, undertake structural roles (Cabeen and Jacobs-Wagner, 2010; Roos et al., 2007). Proteins are composed of one or more long chains of amino acids with each position in the chain occupied by one of 20 standard amino acids. Each particular amino acid features a unique side chain that possesses distinct chemical properties. The amino acid sequence of a protein, encoded in DNA molecules, is known as the primary structure and, providing there is a stable, kinetically accessible free energy minimum under the folding conditions, determines the native three-dimensional protein structure (Anfinsen, 1973). The central dogma of molecular biology states that sequence information stored in DNA is transferred to RNA in a process known as transcription, and from DNA to protein during translation (Crick, 1970). Sequence information can also be transferred from DNA to DNA during DNA replication (Hübscher et al., 2002), from RNA to RNA during viral replication (White et al., 2011), and from RNA to DNA during reverse transcription, a process carried out by retroviruses to generate DNA that can integrate into the host genome (Sarafianos et al., 2009). In contrast, sequence information transfer from the protein level to RNA or DNA has not been observed.

The human proteome is much more complex than its genome, a result of processes such as alternative splicing and post-translational modification which operate on messenger RNA (mRNA) and protein, respectively. Alternative splicing of precursor mRNA allows the generation of multiple protein isoforms from a single gene and is one of the major mechanisms that contributes to proteome diversity in

eukaryotic organisms (Nilsen and Graveley, 2010). Post-translational modification can also dramatically increase proteome diversity by providing mechanisms to modulate various protein properties including enzymatic activity, protein degradation, interaction affinities/kinetics for other molecules and even cellular localisation (Beck-Sickinger and Mörl, 2006). Post-translational modification of proteins occurs in both prokaryotic and eukaryotic organisms and includes the covalent addition of specific chemical groups to amino acid residues as well as proteolytic cleavage of peptide backbones (Walsh et al., 2005). Approximately 300 different post-translational modifications have been identified to-date and while most require enzymatic activity, some can occur without (Wright, 1991).

Protein phosphorylation, for example, is a rapid, reversible and prevalent post-translational modification, which regulates many diverse cellular processes such as metabolism, apoptosis and immunity (Humphrey et al., 2015; Smith-Garvin et al., 2009). Phosphorylation is mediated by protein kinases and is observed on serine, threonine, tyrosine and histidine amino acid residues. The reverse reaction, dephosphorylation, is catalysed by protein phosphatases (Endicott et al., 2012). Glycogen phosphorylase, a metabolic enzyme that converts glycogen to glucose-1-phosphate, is an example of a protein whose activity is regulated by phosphorylation. Phosphorylation of a specific serine residue causes the enzyme to switch from its inactive form, phosphorylase b, into its active form, phosphorylase a, allowing a rapid response to nutrient status (Fischer and Krebs, 1955; Krebs and Fischer, 1956). Protein phosphorylation also plays a crucial role in T cell receptor signalling and is discussed in Chapter 2.

Disulfide bond formation is another common post-translational modification, which involves the oxidation of two cysteine sulfhydryl groups to form a persulfide bond between the two cysteine residue side chains (Fass, 2012). There are three major types of disulfide bond, structural, allosteric and catalytic. Structural disulfide

bonds are usually found in the core of protein domains and help protect the extracellular regions of membrane proteins from proteolysis and denaturation by stabilising the overall structure. This increased stabilisation is achieved, according to classic polymer theory, by decreasing the entropy of the unfolded state. However, under physiological conditions, the effects of disulfide bond formation on the solvation and residual structure of the unfolded protein may have both enthalpic and entropic effects (Betz, 1993; Flory, 1956). Allosteric disulfide bonds, in contrast to structural disulfide bonds, are more accessible and are found predominantly at the surface of protein structures. Reduction and re-oxidation of these disulfide bonds may regulate protein function, with such bonds thought to be relatively common among leukocyte cell surface proteins, for example (Fan et al., 2009; Metcalfe et al., 2011). Catalytic disulfides participate in catalytic reactions that result in the formation, reduction or isomerization of disulfide bonds.

Protein glycosylation, a process in which carbohydrate moieties are covalently attached to amino acid residues by the enzymatic actions of glycosidases and glycosyltransferases, is one of the most diverse and abundant post-translational modifications studied to-date. Glycan residues are most frequently attached via the amide nitrogen atom of asparagine side chains (N-linked) or via the hydroxyl oxygen of the side chains of either serine or threonine residues (O-linked). N-linked glycosylation occurs at asparagine residues in the consensus sequence Asn-X-Ser/Thr (X = any amino acid except proline) and features oligomeric glycans. In contrast, O-linked glycosylation has no consensus sequence and features much simpler glycans (Schwarz and Aebi, 2011; Spiro, 2002). Protein glycosylation is essential for many diverse cellular processes in both prokaryotic (Nothaft and Szymanski, 2010) and eukaryotic organisms (Moremen et al., 2012). In particular, covalently attached glycan residues are directly involved in protein folding and quality control mechanisms in the endoplasmic reticulum (ER) of eukaryotic cells (Xu and Ng, 2015) and play an

important role in the proper folding and trafficking of MHC molecules to the cell surface (Ryan and Cobb, 2012).

As the function of a protein is dependent on its three-dimensional structure, atomic resolution structure determination of protein molecules can provide useful insights into their function.

1.1 Structure determination to atomic resolution

The three most common experimental techniques for solving protein and other macromolecular structures to atomic resolution are X-ray crystallography, nuclear magnetic resonance (NMR) and more recently, cryo-electron microscopy (cryo-EM). Of the ~110,000 macromolecular structures currently deposited in the RCSB Protein Data Bank (www.rcsb.org) approximately 89% were determined by X-ray crystallography, 9.9% by NMR, 0.7% by EM and 0.4% by other methods.

The structure solution of sperm whale myoglobin in 1958 by X-ray crystallography was the first protein structure to be determined at atomic resolution (Kendrew et al., 1958). This milestone achievement along with subsequent technical advances in both X-ray crystallography and molecular biology led to an explosion in the number of structures determined by this technique. X-ray crystallography experiments involve measuring the diffraction pattern generated by exposure of a protein crystal to an X-ray beam. The intensity and phase of each reflection can be used to calculate electron density maps into which an atomic model of the protein is built. Unfortunately, it is not possible to measure the phase of each reflection and this information has to be deduced by either experimental phasing experiments or molecular replacement. The introduction of brighter synchrotron radiation sources, crystal freezing and hybrid photon counting detectors now allows much faster and more accurate data collection from much smaller crystals (E. Eikenberry, 2002;

Garman and Owen, 2007; Helliwell and Mitchell, 2015). Coupled with improvements in macromolecular crystallographic software (Adams et al., 2010; Winn et al., 2011) and automated data processing pipelines (Winter et al., 2013), the time in which it takes to solve a structure, once a well-diffracting crystal has been obtained, has dramatically decreased.

The generation of well-diffracting protein crystals is one of the major limitations of structure determination by X-ray crystallography. Despite numerous studies into the thermodynamics and physics of protein crystallisation and the revelation that the process is predominantly entropically driven, it is not possible to predict the precise conditions that will result in crystal formation for a given protein (Derewenda and Vekilov, 2006; Li et al., 2008). To find suitable conditions, systematic, high-throughput crystallisation trials are necessary (Hui and Edwards, 2003) (Appendix 6.1.1). Today, this involves screening vast numbers of different crystallisation conditions, varying precipitant type, precipitant concentration, protein concentration and temperature as well as testing various additives (Chayen and Saridakis, 2008). Protein engineering experiments, such as mutagenesis to increase overall thermostability, increase solubility or to reduce surface entropy, can increase the chances of obtaining well-diffracting crystals in some cases (Derewenda, 2004; Derewenda and Vekilov, 2006). For example, deriving thermostable mutants of G-protein-coupled receptors (GPCRs) reduced the conformational heterogeneity and increased the stability of the receptors, aiding their crystallisation (Vaidehi et al., 2014; Warne et al., 2008). Mutagenesis to improve protein solubility was critical for the generation of protein crystals that led to the X-ray crystal structure determination of the catalytic domain of HIV-1 integrase (Dyda et al., 1994). The mutation of residues with high conformational entropy (glutamic acid and lysine residues) to residues with lower conformational entropy enhanced both the crystallisation success and protein crystal quality of RhoGDI (Longenecker et al., 2001; Mateja et al., 2002).

The end result of structure determination by X-ray crystallography is a molecular model that has been built into an experimentally determined electron density map that represents the average electron density for the asymmetric unit in the crystal (Appendix 6.1.3). X-ray crystallography can be used to solve the structure of a wide variety of macromolecules, from small molecules such as glycine (Destro et al., 2000) to large viral particles such as the 150 MDa human adenovirus, which contains nearly 1 million amino acid residues (Reddy et al., 2010).

The emergence of NMR in the 1980s as a technique for determining the structure of protein molecules and its subsequent development has led to it now accounting for ~10% of all deposited structure in the RSCB Protein Data Bank (www.rcsb.org). A framework for protein structure determination by NMR was outlined in 1982 (Wüthrich et al., 1982) and in 1984 the first de novo structure of a globular protein was determined (Williamson et al., 1985). Many more structures followed and the award of a Nobel Prize in 1998 highlighted the significance of this achievement.

Structure determination by NMR involves placing a protein sample in a fixed magnetic field, causing nuclei with quantum mechanical spin of ½ to split into two distinct energy levels. The difference in energy between the two levels is dependent on the local environment of the nuclei and corresponds to the frequency at which nuclei precess about the applied magnetic field (the Larmor frequency). A short radio frequency pulse is applied to the sample and the relaxation back to equilibrium measured in a receiver coil. Transfer of magnetization through bonds or space is detected allowing both dihedral angles and distances between magnetic nuclei to be determined. These experimentally determined through-bond (J-coupling) (Hahn and Maxwell, 1952) and through-space (Nuclear Overhauser Effect) (Anderson and Freeman, 1962) relationships are then used as restraints along with other known geometric information (such as bond angles and bond lengths) in a restrained molecular dynamics simulation. The end result is an ensemble of protein structures,

each satisfying the input restraints, which can reveal information about protein conformational flexibility and dynamics (Marion, 2013). Due to the low sensitivity of NMR, protein concentrations in the μM to mM range are required. There is also an upper protein size limit for structure determination by NMR. Increased protein size leads to more complex NMR spectra making the process of assigning magnetic nuclei to their corresponding resonance peaks more difficult. Line-width also increases with larger proteins, leading to spectral overlap. Technical developments such as multidimensional experiments and isotopic labeling have decreased the complexity of NMR spectra and simplified the assignment process resulting in an increase in the size limit to ~100 kDa (Frueh et al., 2013). It is possible to combine NMR with other techniques to determine the structure of larger macromolecular structures. For example, the structure of an archeal box C/D ribonucleoprotein enzyme bound to substrate RNA (~390kDa) was determined using a combination of NMR and small-angle neutron scattering (Lapinaite et al., 2013).

A third technique that can be used to determine macromolecular structures is cryo-EM. Traditionally, electron microscopy has been used to determine low-resolution structures of large macromolecular complexes in the MDa range. For example, a 20 Å electron microscopy reconstruction was obtained for an intact E. coli activator-dependent transcription initiation complex (Hudson et al., 2009). X-ray or NMR structures can also be fitted into low-resolution EM reconstruction maps to generate pseudo-atomic resolution structures of larger complexes (Allen and Stokes, 2013). Recently there has been rapid progression in the cryo-EM field resulting in the determination of near-atomic resolution structures of both large macromolecular complexes and smaller macromolecules. This has been possible with the development of high-voltage field emission gun electron sources, more stable lens systems, improved vacuums and more stable sample stages. Advances in direct

electron detectors and image-processing software has also had a dramatic impact (Kühlbrandt, 2014).

Neither crystals nor highly concentrated samples are required for cryo-EM. Instead, samples are flash frozen in a thin film of aqueous solution on a support grid and EM projection images collected of the molecules in random orientations (Grassucci et al., 2007; Lepault et al., 1983). Due to the inherent sensitivity of biological molecules to radiation damage, low electron dosages are used to collect the images, resulting in a low signal-to-noise ratio. To improve the signal-to-noise ratio, 10,000s or 100,000s of particle projection images are aligned and averaged, with the quality of the final map depending on the accuracy of these alignments. The use of field emission gun electron sources and higher acceleration voltages has led to an increase in the temporal and spatial coherence of the electron beam. This permits the collection of high-resolution images at particular defocus values required to visualise unstained samples. Direct electron detectors are much more sensitive than older CCD detectors or photographic film and are able to detect individual electrons, greatly enhancing the signal to noise ratio (Li et al., 2013; McMullan et al., 2009; Ruskin et al., 2013). A consequence of the increased sensitivity of direct electron detectors is that even low electron doses are capable of saturating the detector. To overcome this issue, long exposures are fractionated into multiple frames of shorter exposures (~0.2s per frame). In combination with the development of maximum likelihood-based image-processing software, this allows beam-induced movements to be tracked and corrected for by aligning the individual frames before summing them into a final image (John L. Rubinstein, 2014; Scheres, 2014). This reduces blurring in the final image that would otherwise result from beam-induced movements (Campbell et al., 2012). Removing frames that are affected by radiation damage can also improve the final averaged two dimensional projection image (Baker and Rubinstein, 2010). In addition, image-processing software has become more advanced, allowing *in silico* purification

where different molecules or different conformers of the same molecule are identified and separated. This further improves the two dimensional particle alignments and ultimately results in improved and higher resolution three dimensional EM reconstructions (Scheres, 2012). These technical improvements have resulted in several ribosome structures being solved to resolutions approaching 3 Å (Amunts et al., 2014; Voorhees et al., 2014; Wong et al., 2014). It has even been possible to solve the structure of a 170kDa protein, human γ-secretase to 4.32 Å and 4.5 Å (Lu et al., 2014; Sun et al., 2015), a protein size previously thought to be too small for structure determination by electron microscopy.

With further technical improvements it may be possible to reduce the lower limit of particle size and increase resolution. For example, new grid structures may help reduce beam-induced movement of the sample (Russo and Passmore, 2014a, 2014b) and the introduction of phase plates will allow the generation of phase contrast without having to collect defocused images (Danev et al., 2014; Murata et al., 2010). Ultimately, improvements in the signal-to-noise ratio of the collected two dimensional projection images and alignment algorithms will increase the quality of particle alignments. This will lead to higher resolution EM reconstructions as well as allowing the high quality alignment of smaller molecules, thereby reducing the lower size limit of the technique.

X-ray crystallography has been particularly instrumental in advancing the understanding of fundamental biological process at the molecular level and, as a result, several Nobel prizes have been awarded for significant technical advances and the structure solutions of important biological molecules. The huge number of macromolecular structures solved by X-ray crystallography has informed the biology of many different fields. For example, first structure of a human MHC class I molecule (HLA-A2) (Bjorkman et al., 1987a) revealed the structural basis of antigen presentation by MHC molecules, contributing significantly to the field of immunology. The structure

revealed a peptide-binding groove, in which electron density was observed for an unknown antigen. Experiments prior to this suggested that HLA molecules recognize enormous numbers of foreign antigens by binding short peptides at a single site (Guillet et al., 1986). The structure showed that most of the polymorphic residues of the HLA-A2 molecule cluster to the peptide-binding groove, offering a plausible mechanism as to how different MHC molecules could recognize different sets of peptide antigens. Two subsequent structures, in which the peptides were modelled (Fremont et al., 1992), revealed that MHC molecules were able to bind such a wide range of peptides by making a small number of specific interactions with peptide side chains at so-called anchor points.

The award of the 2009 Nobel Prize in Chemistry for studies of the structure and function of the ribosome, a ribonucleoprotein, further illustrated the importance of X-ray crystallography in understanding fundamental biological processes. The 2.8 Å structure of the *Thermus thermophilus* 70S ribosome in complex with both tRNA and mRNA revealed the molecular basis of translation, a fundamental cellular process which mediates protein synthesis (Selmer et al., 2006). Although lower-resolution X-ray crystal structures of the bacterial 70S ribosome had been previously solved (Schuwirth et al., 2005; Yusupov et al., 2001), this was the first structure that revealed molecular details of its interaction with mRNA and tRNA. Many of the biochemical experiments performed prior to this structure determination could be rationalised and the structure is being used to design future experiments.

X-ray crystallography has also played a critical role in structure-based drug design, a process that utilises structural information to guide the rational design of potential therapeutic reagents. HIV treatment has greatly improved due to the introduction of numerous new drugs that inhibit different stages of the viral lifecycle, many of which were developed using structure-based drug design. For example, the X-ray crystal structures of HIV-1 protease aided the design of Saquinavir, the first HIV

protease inhibitor drug to be approved for treatment of the disease. The structure of the HIV-1 protease in complex with a peptide transition-state mimic showed the mechanism of binding and guided modifications that ultimately led to a molecule with increased potency for HIV protease inhibition (Roberts et al., 1990).

As this thesis investigates a number of different areas of biology, each of the following chapters features an introduction to the relevant area of biology.

1.2 Thesis aims

The overall aim of this thesis is to determine the atomic structure of a number of different proteins using X-ray crystallography to further understand their biological functions and mechanisms. The technique will be first applied to study the human proteins CD6 and CD166, which are known to participate in an interaction important for the the activation of T-cells during the initiation of the adaptive immune response. Next, peptide-binding to chicken MHC class I molecules will be examined to investigate the mechanisms of peptide binding that are so critical for antigen presentation to the immune system. The interaction of sox pathway proteins SoxYZ (a carrier-protein) and SoxB (a partner enzyme) is then studied to elucidate the molecular details of a carrier-protein / partner enzyme interaction. Finally, TsdA, a bacterial enzyme involved in sulfur metabolism, which also features homology to SoxB, is subjected to structure determination by X-ray crystallography with the aim of determining its overall molecular structure and catalytic mechanism.

Chapter 2: CD6 and its ligand, CD166.

2.1 Introduction

2.1.1 The immune system

The vertebrate immune system has been traditionally divided into two arms, innate and adaptive immunity. The innate response is evolutionary more ancient than the adaptive response and offers a rapid, non-specific response to invading pathogens (Pancer and Cooper, 2006).

Innate immunity is critical for early defense and encompasses anatomical barriers, soluble factors and cellular responses. Anatomical barriers include physical barriers that prevent pathogen entry such as skin and other epithelial surfaces, as well as chemical barriers such as stomach acid that can inhibit the growth of invading pathogens. Epithelial cells, amongst other cell types, secrete antimicrobial peptides such as defensins, offering an additional innate defense mechanism at these surfaces. Defensins are a diverse family of cationic antimicrobial peptides that disrupt the membranes of invading organisms. These molecules are also important for the antimicrobial action of granulocytes (Ganz, 2003).

Innate immune system components such as innate immune cells (including macrophages, neutrophils and dendritic cells) and the complement system can be activated by danger signals such as pathogen-associated molecular patterns (PAMPs) (usually conserved microbial components) (West et al., 2006) or damage-associated molecular patterns (DAMPs) (Chen and Nuñez, 2010). These signals are recognised by germline-encoded pattern recognition receptors (PRRs) and can result in direct phagocytosis or complement-mediated destruction of the invading organism as well as the induction of the inflammation response through the release of various cytokines, chemokines and other pro-inflammatory mediators. The induction of inflammation

leads to the recruitment of additional phagocytes and professional antigen presenting cells such as neutrophils and dendritic cells (DCs) (Guermonprez et al., 2002; Janeway and Medzhitov, 2002; Lea and Johnson, 2012; Takeuchi and Akira, 2010).

Natural killer (NK) cells are another type of innate immune cell, which exhibit cytotoxic effector functions and cytokine release upon detection of signs of viral infection or transformation of host cells. NK cell activation is regulated by a wide variety of both activatory and inhibitory germline-encoded receptors (Lanier, 2008). The major ligand of inhibitory receptors of the killer cell immunoglobulin-like receptor (KIR) family are MHC class I molecules, which are expressed on the surface of host cells (Vivier et al., 2008). Decreased expression of these molecules is often a sign of viral infection because many viruses interfere with the class I MHC antigen presentation pathway to prevent the presentation of virally derived peptides to the adaptive immune system (Hewitt, 2003).

Although innate immune responses are traditionally described as having no memory, there is growing evidence for limited memory-like responses in innate immune cells such as macrophages, monocytes and natural killer (NK) cells (Gardiner and Mills; Sun et al., 2014).

In contrast to the innate immune response, the adaptive response, although much slower to respond, is capable of generating highly targeted and specific responses with long-lasting immunological memory (Pancer and Cooper, 2006). Adaptive immunity is mediated by T- and B-lymphocytes and its activation is highly dependent on DCs, professional antigen presenting cells that form a bridge between innate and adaptive immunity. Upon exposure to PAMPs, DCs become potent activators of naïve T cells via presentation of antigenic peptides in the context of MHC molecules (MHC restriction), initiating the adaptive immune response (Banchereau and Steinman, 1998). MHC molecules will be discussed in greater detail in Chapter 3.

T lymphocytes (T cells) are directly responsible for the cell-based effector functions of the adaptive immune system and can indirectly promote B cell antibody responses through their helper functions. T cells originate from hematopoietic stem cells in the bone marrow which migrate to the thymus where they undergo further development and maturation (Ciofani and Zúñiga-Pflücker, 2010). During their development and maturation, each αβ T cell generates a T cell receptor (TCR) with different antigen specificity through the recombination of V (variable), D (diversity) and J (joining) somatic gene segments, a process initiated by RAG recombinases (Fugmann et al., 2000; Shinkai et al., 1992). This process allows the generation of an almost unlimited number of TCRs, each of which is specific for one antigen. TCRs can also be cross-reactive, further increasing the number of peptide antigens they can recognize. However, the extent of cross-reactivity is unclear as it may be difficult to maintain the affinities required for T cell activation (Birnbaum et al., 2014a; Mazza et al., 2007). Due to the stochastic nature of the generation of TCRs, T cells undergo positive and negative selection in the thymus, a process in which developing T cells are exposed to self-peptide MHC complexes on the surface of thymic antigen-presenting cells. This process assesses the strength of TCR engagement and leads to the elimination, by apoptosis, of T cells carrying TCRs that either have extremely low or extremely high affinities for self-peptide MHC. The resulting naïve T cell repertoire is MHC-restricted and lacks TCRs with high affinity for self-peptide MHC complexes that could otherwise lead to autoimmunity (Moran and Hogquist, 2012; Starr et al., 2003; Wiegers et al., 2011). Once mature, these naïve T cells recirculate through the blood and lymph nodes, sampling different tissue environments, increasing the chance of encountering a specific antigen (Mackay et al., 1990; Sallusto et al., 1999). When DCs encounter specific antigen at sites of inflammation, they express CCR7 and migrate through lymphatic vessels to secondary lymphoid organs where they subsequently activate naïve T cells inducing their clonal expansion (Randolph et al., 2005).

The vast majority of T cells feature highly diverse TCRs composed of α and β chains (αβ T cells) (Ciofani and Zúñiga-Pflücker, 2010) and include $CD4^+$ T helper cells, $CD8^+$ cytotoxic T cells and natural killer T (NKT) cells. $CD4^+$ T helper cells recognize exogenous antigenic peptides in complex with class II MHC molecules and depending on the context of their activation (Nakayama and Yamashita, 2010) can differentiate into a number functional subsets such as T_H1, T_H2, T_H9, T_H17, T_{FH} and T_{Reg} (Bluestone et al., 2009). T_H1 cells are capable of producing interferon-γ, which stimulates immune cells such as natural killer (NK) cells and macrophages, among other functions (Mosmann et al., 1986; Schroder et al., 2004; Welsh, 1984). Follicular B helper T cells (T_{FH} cells) provide help to B cells, resulting in the stimulation of antibody production (Crotty, 2015; Johnston et al., 2009; Nurieva et al., 2009; Yu et al., 2009). Regulatory T (T_{Reg}) cells are essential for maintaining immunological tolerance, pivotal to the prevention of autoimmune disease (Vignali et al., 2008). $CD8^+$ cytotoxic T cells are capable of directly killing target cells through induction of apoptosis via a perforin/granzyme/IFN-γ1 mediated mechanisms or ligation of FAS by FASL (Harty et al., 2000). NKT cells recognise lipid molecules bound to CD1d (a non-polymorphic MHC-like molecule), with their subsequent activation resulting in the release of cytokines and chemokines which enhance $CD4^+$ and $CD8^+$ T cell responses through dendritic cell activation (Fujii et al., 2003; Hermans et al., 2003; Koch et al., 2005; Macho-Fernandez and Brigl, 2015; Zajonc et al., 2005).

In addition to αβ T cells, there is a smaller subset of T cells that feature a TCR composed of γ and δ chains known as γδ T cells. This subset of T cells is not as well studied as αβ T cells, but it is known that γδ T cells are rapid responders, capable of recognizing a broad and distinct set of antigens from αβ TCRs. They appear to protect specific tissues, produce a large range of soluble factors and mediate crucial responses to specific pathogens (Vantourout and Hayday, 2013).

Due to the immense effector functions of T cells, their activation and differentiation needs to be carefully controlled, allowing appropriate responses to real infections while preventing catastrophic damage to the host. T cell activation occurs upon recognition of an antigenic peptide MHC complex by the TCR, against an extremely noisy background of self-peptides, and is therefore a tightly regulated process.

Signalling through the TCR is not only important for T-cell activation but also for development and maturation. Depending on the affinity of the TCR for its ligand, a diverse range of responses can be triggered. For example, the γδ TCR isoform is expressed on the surface of developing T cells, with high levels of signalling through this receptor committing the cell to the γδ T cell lineage. In contrast, low levels of signalling through the γδ TCR favours commitment to the αβ T cell lineage (Ciofani and Zúñiga-Pflücker, 2010). As mentioned earlier, levels of TCR signalling also determine survival or deletion during thymic selection (Starr et al., 2003).

In the periphery, TCR signalling is never in an "off-state", with TCRs constantly engaging in weak interactions with MHC molecules in complex with endogenous self-peptides. This process is known as T cell tickling and maintains T cells in a state of heightened antigenic reactivity (Hochweller et al., 2010). Stronger interactions with foreign peptide MHC complexes result in full T cell activation (Malissen and Bongrand, 2015). This recognition is extremely sensitive with the presence of a single / small number of antigenic peptide-MHC complex molecules being sufficient to induce T cell activation (Huang et al., 2013; Manz et al., 2011; O'Donoghue et al., 2013).

The TCR αβ heterodimer is associated with CD3γε, CD3δε and CD3ζζ, which contain immunoreceptor tyrosine-based activation motifs (ITAMs) in their cytoplasmic domains. T cell activation leads to the phosphorylation of CD3 ITAMs by the tyrosine kinase LCK, which results in the recruitment and activation of ZAP-70. ZAP-70

subsequently phosphorylates the LAT and SLP-76 adaptor proteins, thereby amplifying the signal. T cell activation, in addition to the TCR peptide MHC interaction, also relies on various costimulatory interactions such as that between CD28 molecule on the T cell surface and CD80 or CD86 on the surface of antigen presenting cells (Chen and Flies, 2013).

The mechanism by which the TCR is capable of transmitting antigenic peptide MHC binding to intracellular signalling is still controversial. Conformational changes have been proposed to initiate signalling, however, various crystal structures of TCR molecules and peptide MHC complexes have not revealed any obvious structural changes (Birnbaum et al., 2014b; Garcia and Adams, 2005; Mitra et al., 2004). The kinetic segregation model proposes that inhibitory tyrosine phosphatases such as CD45 and CD148 are passively excluded from T-cell APC contact zones due to their large extracellular domains (van der Merwe and Dushek, 2011). This exclusion of phosphatases would result in long-lived phosphorylation and subsequent recruitment and activation of ZAP-70. Another model suggests that CD3 ITAMs become phosphorylated by LCK molecules associated with the intracellular domains of the co-receptors CD4 and CD8, the extracellular domains of which also bind the MHC peptide complex (Yin et al., 2012). Furthermore, it has been proposed that the TCR may act as a mechanotransducer which could result in a piston-like movement exposing CD3 ITAMs (Kim et al., 2012; Ma and Finkel, 2010).

2.1.2 CD6, a group B SRCR superfamily member

CD6, a member of the SRCR superfamily, is a transmembrane receptor protein expressed on the surface of T cells and a subset of B cells (Aruffo et al., 1997). It includes an extracellular region that features three consecutive group B SRCR domains, followed by a membrane proximal stalk, a transmembrane domain and a

long (244 amino acid) cytoplasmic tail that contains multiple signal transduction motifs (Figure 2.1) (Aruffo et al., 1997). SRCR domains are primitive recognition domains with no identified unifying function (Sarrias et al., 2004). Examples of molecular recognition by SRCR domains include the binding of hemoglobin-haptoglobin complexes to CD163 (Kristiansen et al., 2001) and pathogen binding to MARCO (Ojala et al., 2007).

The interaction between CD6 and its ligand (CD166) is important for both thymocyte development (Patel et al., 1995) and T-cell activation (Bott et al., 1993; Gangemi et al., 1989; Hassan et al., 2004; Osorio et al., 1994), with both proteins found to colocalise with T cell receptor (TCR) at the centre of the immunological synapse (Gimferrer et al., 2004). The vast majority of identified CD6 isoforms result from the alternative splicing of exons that code for the cytoplasmic domain, however, despite their identification, functional effects are unknown. In contrast, an isoform that lacks the CD166 binding domain (CD6Δd3) in the extracellular region becomes up-regulated on T cell activation and is unable to localise at the immunological synapse (Castro et al., 2007). This could provide a regulatory mechanism to control signalling.

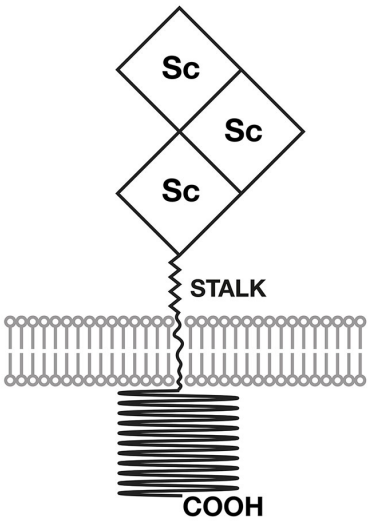

Figure 2.1 - Schematic representation of CD6.
Cartoon diagram illustrating CD6 topology. Extracellular region comprises three scavenger receptor cysteine-rich domains (Sc) and a membrane-proximal stalk region. The cytoplasmic domain is extremely long (244 amino acids) and contains multiple signalling motifs.

Expression of CD6 in the absence of its ligand, CD166, raises the threshold of activation of T cells. In contrast, ligation of CD6 by CD166 results in the co-stimulation of T cells, which is mediated by phosphorylation-dependent binding of SLP-76 to tyrosine-662 found in the cytoplasmic tail of CD6 (Hassan et al., 2006).

2.1.3 Activated leukocyte cell adhesion molecule (ALCAM) / CD166

Activated leukocyte cell adhesion molecule (ALCAM or CD166) is a member of the immunoglobulin superfamily (IgSF) of proteins, featuring two V-set and three C2-set IgSF extracellular domains (Figure 2.2) (Bowen et al., 1995). The expression profile of CD166 is much wider than CD6, being found on the surface of activated leukocytes, hematopoietic stem cells and neuronal cells to name just a few. The cytoplasmic domain is relatively short and regulates adhesion through a link with the cytoskeleton (Te Riet et al., 2014).

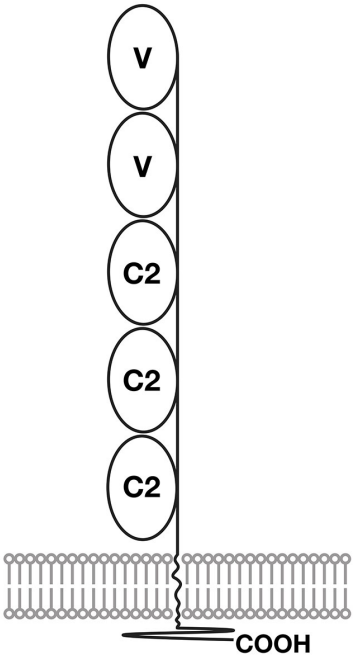

Figure 2.2 - Schematic representation of ALCAM / CD166
Cartoon diagram illustrating CD166 topology. Extracellular region comprises three C2-set and two V-set IgSF domains. The cytoplasmic domain is extremely short and lacks obvious signalling motifs.

In addition to binding CD6, CD166 is capable of forming homophilic interactions (van Kempen et al., 2001) which regulate both normal and neoplastic cells (Chitteti et al., 2013; Weidle et al., 2010). CD166 has been reported to participate in a number of migratory processes. It has been shown that CD166 over-expression on blood-brain barrier cells in multiple sclerosis patients promotes leukocyte trafficking and that blocking CD166 restricts the transmigration of lymphocytes (Cayrol et al., 2008; Swart, 2002). There is also evidence that over-expression of CD166 leads to increased recruitment of lymphocytes to the synovium of rheumatoid arthritis patients (Levesque et al., 1998) and exocrine glands of Sjögren's syndrome patients (Abidi et al., 2006). Furthermore, there is great interest in CD166 due to its expression patterns being correlated with both tumor progression and regression (Kijima et al., 2012; van Kilsdonk et al., 2008).

2.1.4 The interaction between CD6 and CD166

The interaction between CD6 and CD166 has been characterised biochemically, with experiments revealing that the interaction is dependent on the membrane proximal SRCR domain of CD6 (domain 3) and the N-terminal, membrane distal IgSF domain of CD166 (domain 1) (Figure 2.3) (Bowen et al., 1996; Whitney et al., 1995). Human CD6 is reported to bind to human CD166 with an affinity typical of many leukocyte membrane protein interactions (K_D = 0.4-1.0 μM) (Hassan et al., 2004). The homophilic CD166 interaction is weaker than the heterophilic interaction (K_D = 29-48 μM (Hassan et al., 2004)) and is also dependent on the membrane distal N-terminal IgSF domain of CD166 (van Kempen et al., 2001).

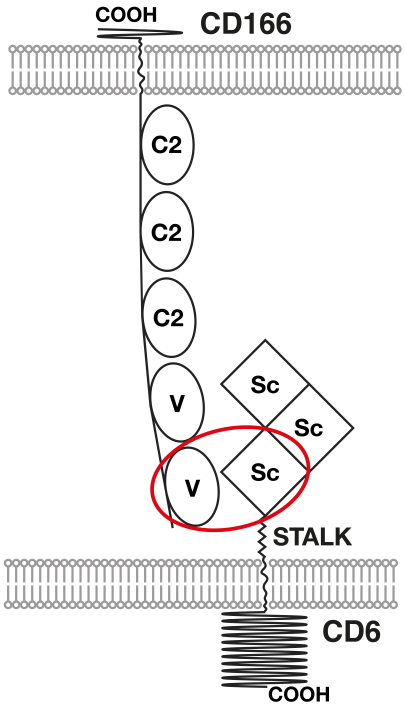

Figure 2.3 – Schematic representation of the CD6/CD166 heterophilic interaction
Proteins illustrated as in Figures 2.1 and 2.2. The membrane proximal SRCR domain of CD6 (domain 3) and the membrane distal IgSF domain of CD166 (domain 1) mediate the heterophilic interaction – highlighted by the red oval.

2.1.5 CD6 single nucleotide polymorphisms (SNPs)

A number of CD6 SNPs have been associated with susceptibility to multiple sclerosis. The CD6 SNPs most strongly associated with multiple sclerosis were located in intronic regions with the intronic SNP rs17824933 found to be associated with progression of the disease. This SNP did not alter the total overall expression level of CD6 on the surface of $CD4^+$ and $CD8^+$ T cells, however, it did increase the expression level of a CD6Δd3 isoform relative to the full-length protein (Kofler et al., 2011). Two non-synonymous SNPs, with slightly weaker associations, R225W (rs11230563) and A257V (rs2074225) were also identified. The protective haplotype (R225/A257) showed higher cell surface expression of CD6 compared to the mild risk (W225/V257) and risk haplotypes (R225/V257) (Swaminathan et al., 2013). It is likely

that the risk conferred by CD6 SNPs is complicated and may involve modulation of expression levels, ligand binding and signal transduction.

2.1.6 Therapeutic antibodies

There are numerous examples of the use of anti-CD6 antibodies in the clinic, all of which are specific for the membrane distal SRCR domain of CD6 (domain 1). Anti T-12, a murine anti-CD6 monoclonal antibody (mAb), has been used to deplete T cells from donor bone marrow before transplantation to reduce both acute and chronic graft versus host disease (GVHD) (Reinherz et al., 1982). The same antibody was used experimentally in the treatment of multiple sclerosis (Hafler et al., 1986).

Biocon, a biotech company, have released the first humanised anti-CD6 mAb (Itolizumab) to the market. This antibody is being used as an immunosuppressive therapeutic drug for treating autoimmune diseases. Treatment of psoriasis patients with the Itolizumab has resulted in positive clinical outcomes (Aira et al., 2014; Budamakuntla et al., 2015; Jayaraman, 2013) and there is evidence that this antibody could also provide beneficial effects to rheumatoid arthritis patients (Rodriguez et al., 2012).

2.2 Aims

The primary aim of this chapter is to solve the crystal structures of the extracellular regions of CD6 and CD166, both individually and in complex, to reveal the molecular details of an unusual interaction between an IgSF domain and an SRCR domain. This may also provide insight into how SNPs and therapeutic antibodies may perturb the interaction.

2.3 Protein expression & purification

Bacterial expression systems are often the expression system of choice for crystallisation experiments due to their controllability, short generation time and high yield of recombinant protein. However, expressing mammalian disulfide bond-containing proteins is particularly challenging when using a bacterial system as they lack the cellular machinery to form such bonds. In order to generate the correct pattern of disulfide bonds, refolding experiments are often required. Each of the extracellular domains of human CD6 and CD166 contain 4 and 1 disulfide bond(s), respectively. As such, a eukaryotic expression system in the form of the Chinese hamster ovary (CHO) Lec 3.2.8.1 cell line was chosen as the expression system of choice. There are several advantages of using this system. Firstly, eukaryotic cells contain the machinery for disulfide bond formation. Furthermore, recombinant protein can be secreted directly into the culture medium, negating the need for cell lysis. One important disadvantage of using eukaryotic expression systems for crystallisation experiments is that recombinant proteins containing glycosylation sites will become glycosylated due to the presence of the appropriate glycosylation machinery in eukaryotic cells. This reduces sample homogeneity, however, glycosylation of such sites is sometimes necessary for protein solubility and function. The CHO Lec 3.2.8.1 cell line was specifically chosen because it features four independent mutations in both the N- and O-glycosylation pathways (Stanley, 1989) resulting in homogeneous O-linked glycosylation (a single GalNAc residue per site) and N-linked glycosylation of the high mannose type, which can be enzymatically trimmed back to a single N-linked N-acetylglucosamine residue using endoglycosidase H.

2.3.1 Plasmid, cloning & stable cell lines

Collaborators in Neil Barclay's group amplified the DNA sequences of the three extracellular SRCR domains of human CD6, excluding the membrane proximal stalk

(residues 1 to 364) [Uniprot ID: P30203] and the three N-terminal membrane distal IgSF domains of human CD166 (residues 1 to 335) [Uniprot ID: Q13740] from plasmid DNA. These DNA fragments, containing their native signal sequences, were cloned into the PEE14 vector as His tagged proteins (Figure 2.4).

A MWLFFGITGLLTAALSGHPSPAPPDQLNTSSAESELWEPGERLPVRLTNG
SSSCSGTVEVRLEASWEPACGALWDSRAAEAVCRALGCGGAEAASQLAPP
TPELPPPPAAGNTSVAANATLAGAPALLCSGAEWRLCEVVEHACRSDGRR
ARVTCAENRALRLVDGGGACAGRVEMLEHGEWGSVCDDTWDLEDAHVVCR
QLGCGWAVQALPGLHFTPGRGPIHRDQVNCSGAEAYLWDCPGLPGQHYCG
HKEDAGVVCSEHQSWRLTGGADRCEGQVEVHFRGVWNTVCDSEWYPSEAK
VLCQSLGCSASRGHHHHHH

B MESKGASSCRLLFCLLISATVFRPGLGWYTVNSAYGDTIIIPCRLDVPQN
LMFGKWKYEKPDGSPVFIAFRSSTKKSVQYDDVPEYKDRLNLSENYTLSI
SNARISDEKRFVCMLVTEDNVFEAPTIVKVFKQPSKPEIVSKALFLETEQ
LKKLGDCISEDSYPDGNITWYRNGKVLHPLEGAVVIIFKKEMDPVTQLYT
MTSTLEYKTTKADIQMPFTCSVTYYGPSGQKTIHSEQAVFDIYYPTEQVT
IQVLPPKNAIKEGDNITLKCLGNGNPPPEEFLFYLPGQPEGIRSSNTYTL
MDVRRNATGDYKCSLIDKKSMIASTAITVHYLDLSTRHHHHHH

Signal sequence
His-tag

Figure 2.4 - Construct Sequences
CD6 construct sequence (A) and CD166 construct sequence (B). Signal sequences predicted using SignalP 4.1 Server. Theoretical molecular weights of the native proteins, excluding the signal sequence and potential glycosylation, were calculated using ProtParam (ExPASy) - CD6: 32,084 Da, CD166: 36,018 Da

Our collaborators performed transfections with these plasmids containing the CD6 or CD166 extracellular fragments and generated two stable CHO Lec 3.2.8.1 cell lines that expressed the recombinant proteins.

2.3.2 Expression and purification

Stable CHO Lec 3.2.8.1 cell lines provided by our collaborators were thawed and pelleted by centrifugation at 4000rpm to remove the storage media. Cell pellets were resuspended in Dulbecco's Modified Eagle's Medium (DMEM -glutamine) (Life technologies) + 15μM MSX + 10% fetal calf serum (FCS) + GS supplement (Life

technologies) + sodium pyruvate (Life technologies) (1mM) + Penicillin (Life technologies) (100 U/mL) + Streptomycin (Life technologies) (100 µg/mL). Cells were grown initially in 75cm^2 flasks until confluent, subsequently split into 175cm^2 flasks and then eventually into cell factories. Once cell factories became confluent, sodium butyrate (2mM) was added to the media and the cells allowed to secrete recombinant protein for two weeks.

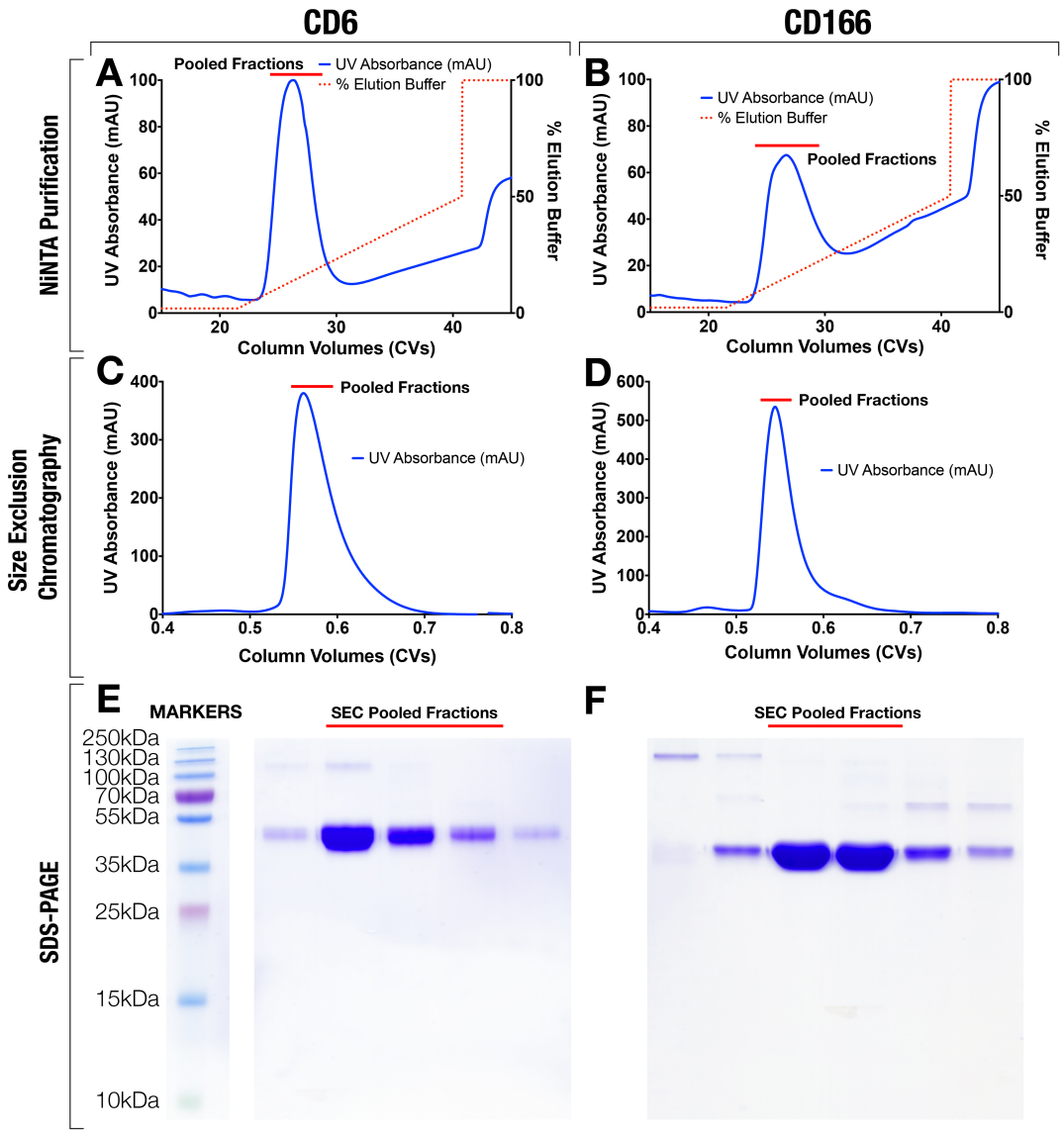

Figure 2.5 - Purification of CD6 and CD166
Panels A & B: Chromatogram of NiNTA column elution. UV_{280} absorbance is shown in blue and the imidazole elution gradient in red. Elution buffer contained running buffer (10 mM HEPES, 150 mM NaCl, pH 7.4) plus 1M imidazole. Panels C & D: Chromatograms of the size exclusion elution using an S200 column. Panels E & F: Reducing SDS-PAGE (15%) gel showing final purification and fractionation of recombinant proteins. Red bar shows final pooled fractions.

Tissue culture supernatant (TCS) was filtered using a Millipore Stericup 0.22µm filter to remove cellular debris and then concentrated to ~50mL with a 10,000 MWCO Sartorius Vivablock. Recombinant proteins were purified from the concentrated supernatant using 1 mL NiNTA superflow cartridges (Qiagen, USA). NiNTA cartridges were washed with 20 column volumes (CVs) of water and then 10 CVs of running buffer (10 mM HEPES, 150 mM NaCl, pH 7.4) before concentrated TCS was applied at a rate of 1 mL/min. After application of the TCS, the column was washed with 20 CVs of running buffer plus 20 mM imidazole. Proteins were eluted over the next 20 CVs with an imidazole gradient that increased the imidazole concentration from 20 mM to 500 mM over this volume (Figure 2.5A & B). Following elution, proteins were immediately dialysed in 10mM HEPES, 150mM NaCl, pH 7.4 to remove imidazole. Fractions were pooled and concentrated to 5mL in a 10,000 MWCO spin concentrator (Millipore, USA) and, to further improve purity, subjected to a final purification step using a 26/60 Superdex-200 size exclusion column (SEC) (GE Healthcare, USA) using the same running buffer with a flow rate of 2 mL/min (Figure 2.5C & D). Resulting fractions were subjected SDS-PAGE analysis using a 15% polyacrylamide gel to confirm purity (Figure 2.5E & F). Sample fractions were pooled as indicated by the red bars in Figure 2.5.

2.3.3 Multi angle light scattering (MALS)

To assess the molecular species present and the homogeneity of the purified samples, MALS analysis was carried out on the individual proteins and a 1:1 mixture (Figure 2.6). A Superdex-200 10/30 column (GE Healthcare) equilibrated in 50 mM Tris HCl, pH 7.5, 150 mM NaCl at 0.4 mL/min, was used to perform SEC. The column was followed by an in-line Dawn Heleos-II light scattering detector (Wyatt Technologies) and Optilab-Rex refractive index monitor (Wyatt Technologies). The

ASTRA 5.3.4.14 (Wyatt Technologies) software package was used to calculate molecular mass assuming a dn/dc value of 0.186 ml/g. Samples were diluted approximately 10-fold when applied to the column.

For the CD6 sample (39 μM protein sample injected onto the column, resulting in an approximate total concentration of 3.9 μM on the column assuming a 10-fold dilution factor) the molecular mass calculations indicated a mass of ~50-60 kDa (Figure 2.6A). Although the theoretical molecular weight of the CD6 construct was ~37.8 kDa (excluding any glycosylation), when taking into account the SDS-PAGE analysis that suggested a molecular weight of ~50 kDa (Figure 2.5E), this species was most likely predominantly monomer. The slope in the molecular weight observed across this peak may indicate self-association or sample heterogeneity due to glycosylation. Analysis of the CD166 sample (injected sample: 28 μM, approximate concentration on the column: 2.8 μM) identified two species, one at ~45 kDa and another at ~70 kDa (Figure 2.6B). The theoretical molecular weight of the CD166 construct was ~36 kDa, with SDS-PAGE analysis suggesting a glycosylated mass of ~45 kDa (Figure 2.5F). This inferred that the ~45 kDa species observed in the MALS analysis was indeed a monomer. Although the theoretical CD166 dimer mass was ~90 kDa, a second species of ~70 kDa observed in the MALS analysis was most likely a dimer, with the slightly lower calculated molecular mass due to an equilibrium between monomers and dimers in the SEC peak. The equimolar mix of CD6 and CD166 (injected sample: both components at 27 μM each, approximate concentration on the column: 2.7 μM) revealed a new SEC peak. This peak could be explained by the formation of a heterodimer, however it was not possible to rule out that this was a homodimer of CD6 (Figure 2.6C). The calculated mass of the peak was less than the theoretical heterodimer or homodimer masses, again, and was likely due to equilibrium between monomer and dimers in this SEC peak. It was not possible to concentrate the proteins any further than those used in these experiments due to issues with protein

solubility so it was not possible to attempt to obtain calculated molecular masses closer to the true dimer values.

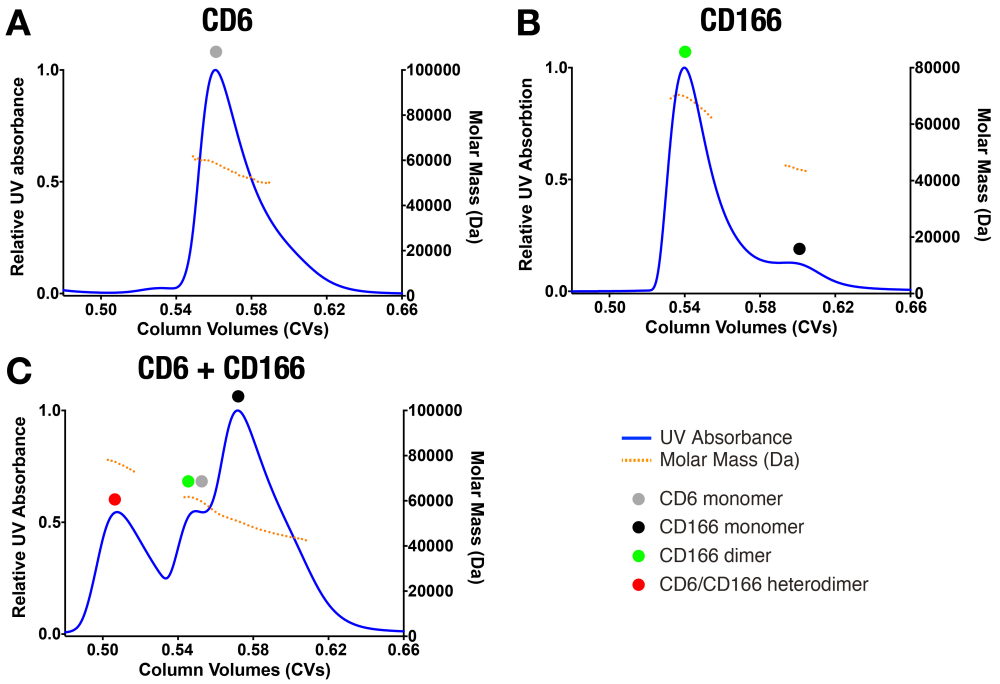

Figure 2.6 - MALS analysis of recombinant proteins
SEC-MALS analysis of individual proteins and a 1:1 mixture. Panel A: CD6 existed primarily as a monomer at an injection concentration of 39 μM. Panel B: CD166 existed in equilibrium between a monomer and a dimer at an injection concentration of 28 μM. Panel C: Mixing equimolar amounts (27 μM) of CD6 and CD166 resulted in a more complicated profile with an additional peak that could be explained by a CD6/CD166 heterodimer.

2.4 Crystallisation, data collection and processing

2.4.1 Deglycosylation

In order to increase the homogeneity of the protein samples, high mannose N-linked glycosylation moieties were enzymatically cleaved to leave a single N-acetylglucosamine residue at each glycosylation site using endoglycosidase H (New England Biolabs Inc.). Deglycosylation conditions were optimized to remove the sugar residues with the minimum amount of enzyme and with the greatest yield of soluble deglycosylated protein. The optimal conditions for deglycosylation were using 2.5U

enzyme per µg of protein, with glycosylated protein concentrations of ~0.3 mg/mL in total reaction volumes of 200 µL for 2 hours at 37°C (Figure 2.7).

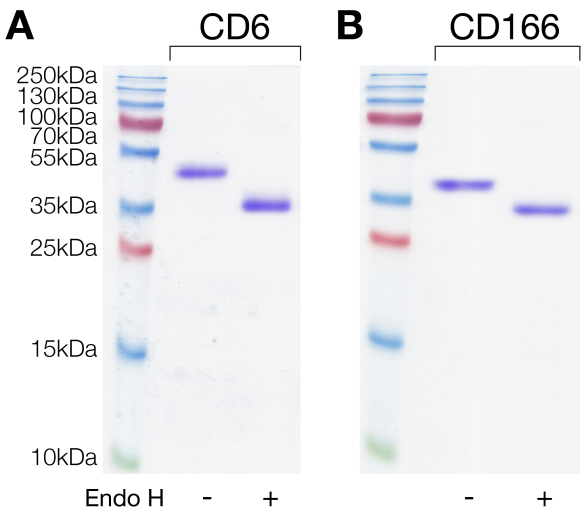

Figure 2.7 – Deglycosylation of CD6 and CD166
15% SDS-PAGE analysis of deglycosylation. 2.5U endoglycosidase H per µg protein was added where indicated by (+).

2.4.2 Crystallisation

Protein solutions were subjected to crystallisation trials using the sitting drop method with commercially available sparse matrix and systematic crystallisation screens purchased from Molecular Dimensions (Table 2.1). A Hydra 96 (Alpha Biotech Ltd) robot was used to transfer 70 µL of mother liquor into the reservoirs of MRC 2-drop crystallisation plates. An Oryx Nano-crystallisation robot (Douglas Instruments, UK) was used to set up drops that contained both protein solution and mother liquor. Plates were then sealed using a StarSeal Advanced Polyolefin Film (STARLAB) and incubated at 21°C.

Commercial crystallisation screens
Structure Screen 1 & 2
JCSG-plus
PACT premier
Morpheus
Stura & Macrosol
Clear Strategy
MIDAS
Proplex

Table 2.1 - Commercial crystallisation screens

The extracellular SRCR domains of human CD6 were crystallised in 0.1M ammonium sulfate, 0.3M sodium formate, 0.1M sodium cacodylate, 3% w/v PGA-LM, 20% MPD, pH 6.5. Drops contained 2.5 nL of a 1:1 protein mixture of CD6 and CD166 at a total protein concentration of 37.5 µM and 2.5 nL of mother liquor. Trays were stored at 20°C and crystals grew within 4 weeks.

Crystals of the extracellular portion of human CD166 were obtained in 0.1M sodium HEPES, 25% w/v PEG-2000 MME, pH 7.5. Drops contained 2 nL CD166 protein solution at 112 µM and 2 nL mother liquor and were again stored at 20°C. Crystals grew within 4 weeks (Figure 2.8).

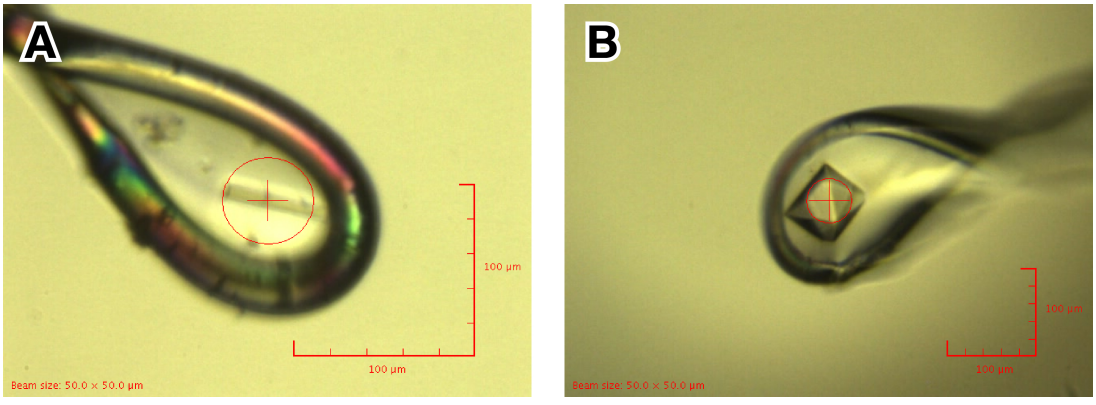

Figure 2.8 - CD6 and CD166 crystals
Crystals of CD6 (A) and CD166 (B) mounted at Diamond Light Source's i04-1 beamline. Red circle shows beam width. Scale bars indicate a 100µm x 100µm square.

2.4.3 Data collection

Both crystals were cryoprotected in mother liquor plus 20% (v/v) ethylene glycol before being flash frozen in liquid nitrogen. X-ray diffraction data for CD6 and CD166 were collected at Diamond Light Source's i04-1 beamline after initial test images were collected to determine optimal data collection strategies (Table 2.2). Additional images were collected in case of incorrect determination of crystal symmetry at this stage.

Data Collection	CD6	CD166
Diffraction source	Diamond Light Source, i04-1	Diamond Light Source, i04-1
Detector	Pilatus 2M	Pilatus 2M
Energy (eV)	13476	13476
Wavelength (Å)	0.92	0.92
Temperature (K)	100	100
Detector resolution (Å)	3.0	1.8
Exposure per image (s)	0.75	0.25
Ω start (º)	1	138
Ω end (º)	51	318
Ω osc (º)	0.25	0.30
Number of images	200	600

Table 2.2 - Data collection parameters for CD6 and CD166

2.4.4 CD6 data processing

The CD6 dataset was processed using XIA2 (Winter et al., 2013) in the 3daii mode, indexing in a primitive tetragonal cell (a = 161.48 Å, b = 161.48 Å, c = 93.85 Å, $\alpha = 90^{\circ}$, $\beta = 90^{\circ}$, $\gamma = 120^{\circ}$) with data to a resolution of 3.15Å in the P 6_3 2 2 space group (Table 2.3 & Figure 2.9). Systematic absences are consistent with the chosen space group (Figure 2.10).

Data Collection	CD6
Resolution range (Å)	77.93-3.15 (3.23-3.15)
Space group	P 6_3 2 2
Unit cell dimensions	
a, b, c (Å)	161.48, 161.48, 93.85
α, β, γ (°)	90, 90, 120
Total reflections	69680 (5176)
Unique reflections	12884 (933)
Multiplicity	5.4 (5.5)
Completeness (%)	99.6 (99.8)
Mean I/sigma(I)	7.2 (2.2)
R_{merge} (%)	23.2 (80.0)
R_{pim} (%)	10.9 (37.1)
CC(1/2)	0.975 (0.673)

Table 2.3 – CD6 data processing statistics

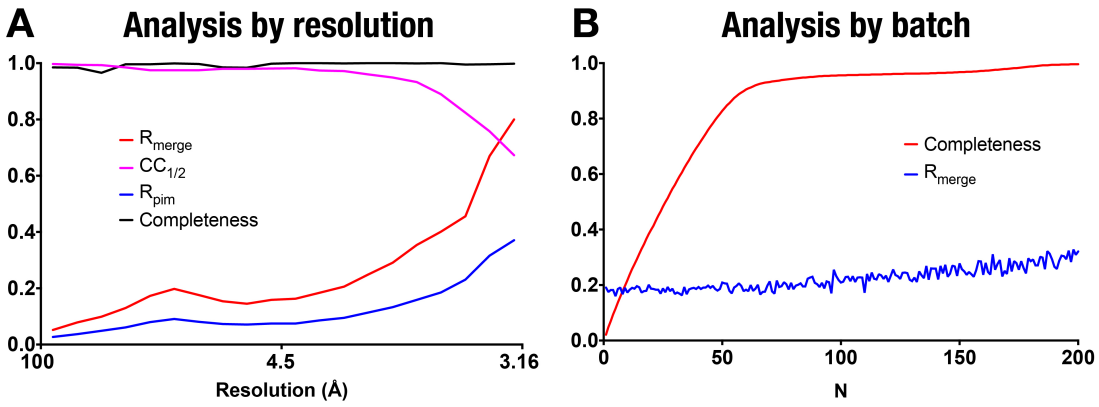

Figure 2.9 - CD6 data processing analysis
Panel A: Analysis by resolution. Panel B: Analysis by batch.

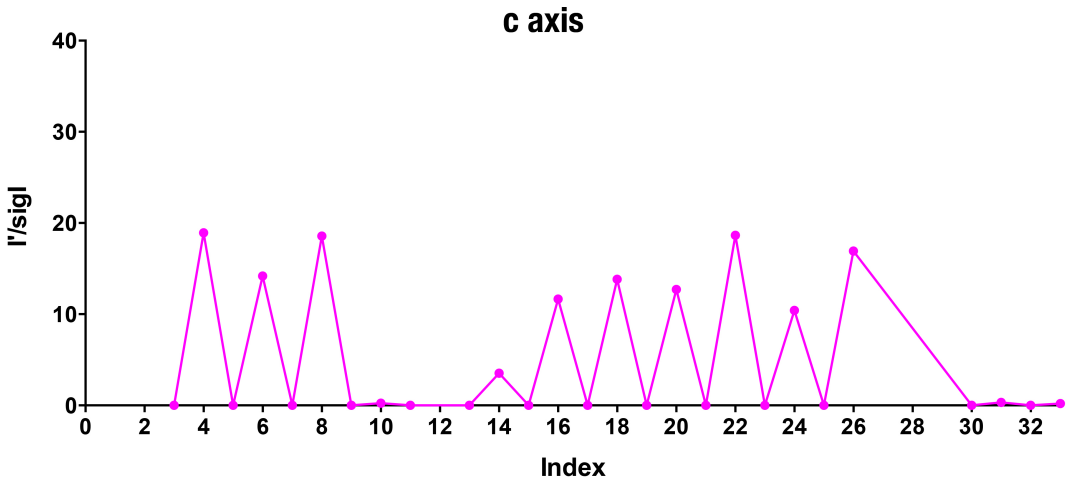

Figure 2.10 - CD6 systematic absences
Systematic absences are consistent with the P6_3 2 2 space group (l=2n).

2.4.5 CD166 data processing

The CD166 data were processed with XIA2 (Winter et al., 2013) in 3daii mode, indexing in a primitive tetragonal cell (a = 72.32, b = 72.32, c = 105.04, α = 90, β = 90, γ = 90) with data to a resolution of 1.86 Å in the $P4_3\ 2_1\ 2$ space group (Table 2.4, Figure 2.11 & 2.12).

Data Collection	CD166
Resolution range (Å)	51.14-1.86 (1.91-1.86)
Space group	P $4_3\ 2_1\ 2$
Unit cell	
a, b, c (Å)	72.32, 72.32, 105.04
α, β, γ (°)	90, 90, 90
Total reflections	313566 (22938)
Unique reflections	24110 (1731)
Multiplicity	13.0 (13.3)
Completeness (%)	99.9 (99.9)
Mean I/sigma(I)	20.1 (3.8)
R_{merge} (%)	8.0 (65.9)
R_{pim} (%)	2.3 (18.5)
$CC_{1/2}$	0.999 (0.926)

Table 2.4 - CD166 data processing statistics

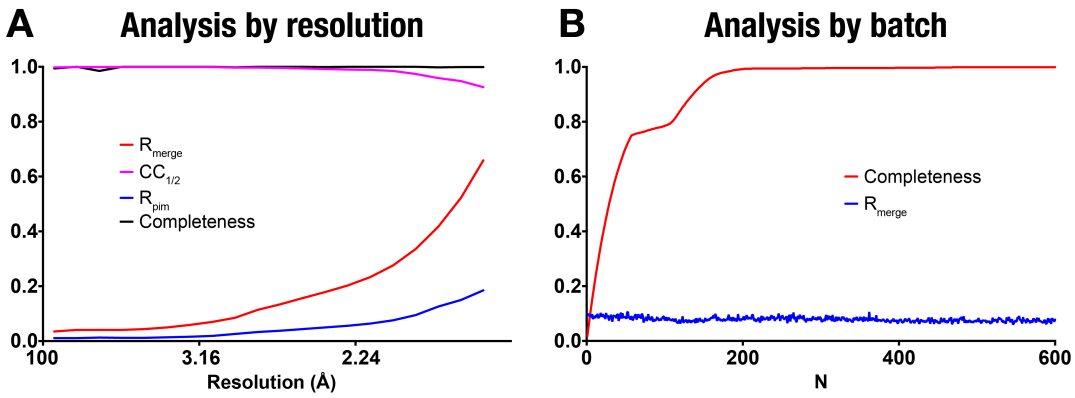

Figure 2.11 – CD166 data processing analysis
Panel A: Analysis by resolution. Panel B: Analysis by image batch.

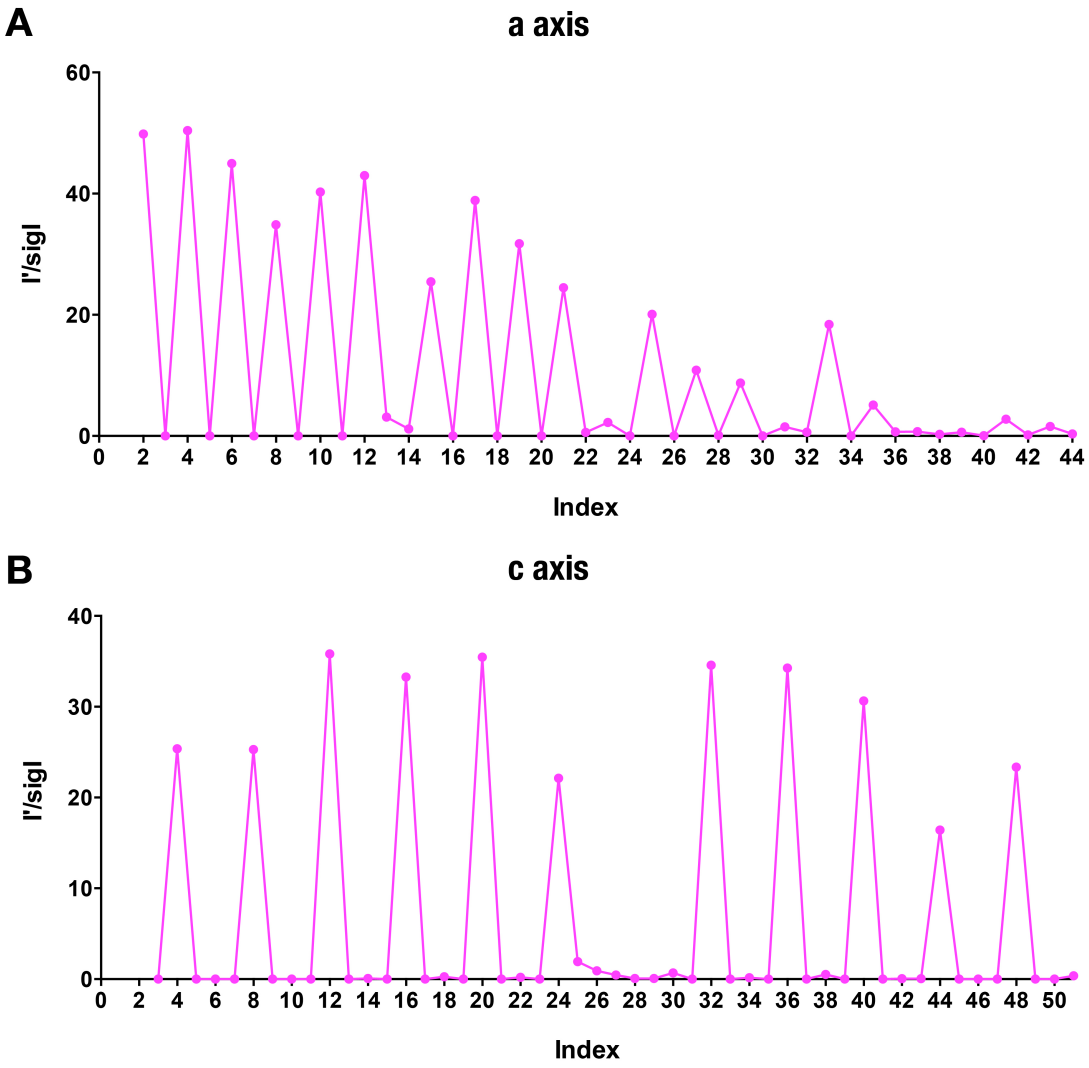

Figure 2.12 - CD166 dataset systematic absences
Systematic absences are consistent with $P4_3\,2_1\,2$ space group (h =2n and l=4n).

2.5 Structure determination and refinement

Due to the availability of appropriate search models in the Protein Data Bank (PDB), molecular replacement was attempted to determine initial phases for both structures.

2.5.1 CD6 structure determination

Processed data were analyzed with PHENIX.XTRIAGE (Adams et al., 2010) with no evidence of twinning or translational non-crystallographic symmetry. Cell content analysis by MATTHEWS (Kantardjieff and Rupp, 2003) suggested the presence of two copies of CD6 in the asymmetric unit with a solvent content of 49.86% (Table 2.5). Inspection of the self-rotation functions for the data revealed no significant non-crystallographic peaks.

Nmol / asym	Matthews coefficient	% solvent	P(3.15)	P(tot)
1	4.90	74.93	0.02	0.01
2	2.45	49.86	0.98	0.99
3	1.63	24.79	0.00	0.00

Table 2.5 – CD6 cell content analysis

The sequence of the three extracellular domains of CD6 was submitted to the FFAS03 server for analysis by the profile-profile alignment and fold recognition algorithm FFAS (Rychlewski et al., 2000). This method allowed the identification of remote homologies that other sequence comparison methods could not detect. The top FFAS hit was the SRCR domain of Mac-2 binding protein (PDBID: 1by2) (Hohenester et al., 1999), with 26.7% sequence identity to the CD6 domain 1, 44.5% to domain 2 and 26.1% to domain 3. The CCP4 (Winn et al., 2011) program CHAINSAW (Stein, 2008) was used to generate search models for each of the three SRCR domains of CD6. The non-conserved residues were pruned from the MAC-2

binding protein coordinates in a manner that retained all atoms common to the model and target sequence.

Molecular replacement was performed using PHASER (McCoy et al., 2007), with a total of three SRCR domains being placed in the asymmetric unit (Table 2.6).

	RFZ	TFZ	LLG
1st copy	3.9	7.1	122
2nd copy	3.0	12.3	293
3rd copy	2.9	10.4	376

Table 2.6 – CD6 molecular replacement
Rotation function (RFZ) & translation function (TFZ) Z-scores and log-likelihood gains (LLG) for sequential placement of three SRCR domains

It was only possible to place a total of three SRCR domains during molecular replacement, consistent with only a single copy of the protein present in the asymmetric unit. Analysis of this molecular replacement solution revealed crystal contacts that would allow the crystal to pack.

2.5.2 CD6 model building and refinement

The molecular replacement solution was subjected to rounds of model building and refinement using COOT (Emsley et al., 2010) and AUTOBUSTER (Bricogne et al.). After inspecting the electron density map after the first round of refinement, there was no evidence of any electron density for additional copies of CD6 in the asymmetric unit. Domain assignment was carried out by both the inspection of electron density at the N- and C-termini of each SRCR domain, identifying connectivity between domains, and also by inspection of electron density at certain amino acid positions that featured amino acid side chains that differed greatly in size and shape between domains. The model was subjected to further rounds of refinement and rebuilding reaching a final model with an R_{work} of 25.31% and R_{free} of 28.31% (Table 2.7). Model geometry was validated using Molprobity (Chen et al., 2010).

Refinement	CD6
Resolution range (Å)	77.93-3.15 (3.45-3.15)
Number of reflections (work / test set)	12883 / 629 (2851 / 156)
Number of atoms (protein, glycans, ligands, waters)	4399, 27, 60, 43
Protein residues	324
Mean B-factors (protein, glycans, ligands, waters)	37.85, 61.24, 35.33, 10.28
R_{work} (%)	25.31 (28.48)
R_{free} (%)	28.31 (30.76)
R.M.S.D. from idea values (bonds, Å)	0.0061
R.M.S.D. from idea values (angles, º)	0.76
Ramachandran favoured (%)	95.05
Ramachandran outliers (%)	0
Molprobity score	1.06 (100th percentile)

Table 2.7 - CD6 refinement statistics

2.5.3 CD166 structure determination

Data were analysed with PHENIX.XTRIAGE (Adams et al., 2010) with no indication of twinning or translational non-crystallographic symmetry. Cell content analysis using MATTHEWS (Kantardjieff and Rupp, 2003) suggested a single copy of the CD166 construct (Table 2.8) and subsequent inspection of self-rotation function showed no significant non-crystallographic peaks.

The CD166 structure was solved using the molecular replacement program PHASER (McCoy et al., 2007) and search models derived from the second IgSF domain of RAGE (PDBID: 3cjj) (Koch et al., 2010), which had a sequence identity of 16.4% to domain 1 and 30.3% to domain 2 of CD166 (Table 2.9). Although the construct used for crystallisation contained the three membrane distal IgSF extracellular domains of CD166, only two IgSF domains could be placed in the asymmetric unit. Inspection of the two-domain solution in COOT (Emsley et al., 2010), revealed insufficient space for a third connected IgSF domain.

Nmol / asym	Matthews coefficient	% solvent	P(3.15)	P(tot)
1	1.91	35.53	1.00	1.00

Table 2.8 - CD166 cell content analysis

	RFZ	TFZ	LLG
1st copy	3.8	6.8	53
2nd copy	3.4	7.6	95

Table 2.9 – CD166 molecular replacement
Rotation function (RFZ) & translation function (TFZ) Z-scores and log-likelihood gains (LLG) for sequential placement of two IgSF domains.

2.5.4 CD166 model building & refinement

In order to determine which of the two domains of CD166 were present in the crystal, initial rounds of model building and refinement of the molecular replacement solution using COOT (Emsley et al., 2010) and AUTOBUSTER (Bricogne et al.) enabled calculation of electron density maps that could be inspected at various amino acid side chain positions. Due to the relatively high resolution of this structure, it was trivial to identify side chains by their electron density and hence it was possible to determine that the crystal contained CD166 domains 1 & 2. The final model was built using BUCCANEER (Cowtan, 2006) with the molecular replacement solution from PHASER (McCoy et al., 2007) used to place and seed chain growing. Final rounds of refinement and model building were carried out using AUTOBUSTER (Bricogne et al.) and COOT (Emsley et al., 2010), yielding a model with R_{work} = 22.17% and R_{free} = 23.91% (Table 2.10).

Refinement	CD166
Resolution range (Å)	21.09-1.86
Number of reflections (work / test set)	24029 / 1229
Number of atoms (protein, glycans, ligands, waters)	3496, 54, 160, 186
Protein residues	218
Mean B-factors (protein, glycans, ligands, waters)	32.53, 47.28, 59.05, 38.80
R_{work} (%)	22.17
R_{free} (%)	23.91
R.M.S.D. from idea values (bonds, Å)	0.0061
R.M.S.D. from idea values (angles, º)	0.93
Ramachandran favoured (%)	96.79
Ramachandran outliers (%)	0
Molprobity score	0.80 (100th percentile)

Table 2.10 - CD166 refinement statistics

2.6 A low resolution CD6 dataset

A second crystal form of CD6 was grown in a 500 nL drop containing 75% 1:1 protein mixture of CD6 and CD166 (total protein concentration of 38.5 μM) and 25% mother liquor (0.2M Magnesium chloride, 0.1M sodium acetate, pH 5.0, 8% (w/v) γ-PGA (Na^+ form, LM)). The crystal was cryoprotected in 20% (v/v) ethelyne glycol and flash frozen in liquid nitrogen. A native dataset was collected at Diamond Light Source's i04 beamline (Table 2.11).

Data Collection	CD6
Diffraction source	Diamond Light Source, i04
Detector	Pilatus 6M
Energy (eV)	12658
Wavelength (Å)	0.9795
Temperature (K)	100
Detector resolution (Å)	3.50
Exposure per image (s)	0.20
Ω start (°)	132
Ω end (°)	312
Ω osc (°)	0.20
Number of images	900

Table 2.11 - Low resolution CD6 data collection

Data were processed using XIA2 (Winter et al., 2013) in 3daii mode, indexing a trigonal cell (a = 140.4 Å, b = 140.4 Å, c = 190.7 Å, α = 90°, β = 90°, γ = 120°) with data to a resolution of 5.19 Å in the P 3 space group (Table 2.12).

Data Collection	CD6
Resolution range (Å)	75.01-5.19 (5.33-5.19)
Space group	P 3
Unit cell	
a, b, c (Å)	140.4, 140.4, 190.7
α, β, γ (°)	90, 90, 120
Total reflections	85042 (6158)
Unique reflections	16216 (1202)
Multiplicity	5.2 (5.1)
Completeness (%)	99.7 (99.2)
Mean I/sigma(I)	7.5 (2.1)
R_{merge} (%)	13.7 (74.4)
R_{pim} (%)	10.4 (57.2)
$CC_{1/2}$	0.998 (0.901)

Table 2.12 - Low resolution CD6 data processing

Although this dataset was at a much lower resolution than the dataset presented in section 2.4.4, molecular replacement was attempted because observation of the same overall structure of CD6 in two different crystal forms would increase the confidence that this was the physiologically relevant structure and not a structure influenced by crystal packing contacts.

PHENIX.XTRIAGE (Adams et al., 2010) suggested no evidence of translational non-crystallographic symmetry, but did suggest possible twinning. Molecular replacement using PHASER (McCoy et al., 2007) was able to place 4 copies of the refined CD6 structure determined earlier in section 1.4.1 (Table 2.13).

	RFZ	**TFZ**	**LLG**
Copy 1	3.9	6.3	44
Copy 2	4.0	13.4	151
Copy 3	4.0	8.5	145
Copy 4	4.0	12.0	112

Table 2.13 - Low resolution CD6 molecular replacement

Inspection of this molecular replacement solution showed suitable crystal contacts such that it was possible for the crystal to pack. A single round of refinement with REFMAC using intensity-based twin refinement (Murshudov et al., 1997) resulted in a model with an R_{work} of 24.96% and R_{free} of 29.83%.

2.7 The Structures of Human CD6 & CD166

2.7.1 Human CD6

The X-ray crystal structure of CD6 was the first of consecutive SRCR domains, with earlier structures featuring either single, isolated SRCR domains (Garza-Garcia et al., 2008; Hohenester et al., 1999; Rodamilans et al., 2007) or single SRCR domains as part of larger polypeptide chains comprised of other protein domains (Herter et al., 2005; Katz et al., 2004; Roversi et al., 2011; Somoza et al., 2003). The most striking structural feature in this new crystal structure was that of the non-linear organisation of the consecutive SRCR domains, which resulted in a molecule that spanned a shorter distance (78.1 Å) than if the domains were organised in a linear manner (~90 Å) (Figure 2.13). This distinct structural feature was also observed in a lower resolution dataset collected from a different crystal form (Section 2.5), consistent with this being the native structure and not an artifact of crystal packing.

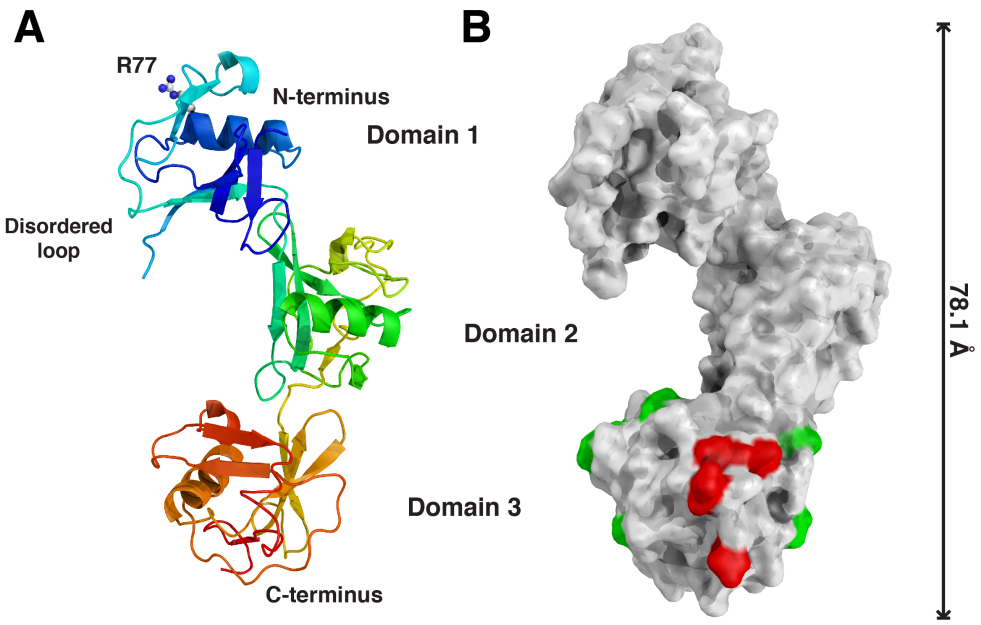

Figure 2.13 - CD6 structure
Panel A: Cartoon representation coloured from blue at the N-terminus to red at the C-terminus There is a disordered loop in a similar position to a flexible loop observed in CD5 domain 1. Panel B: Surface representation with mutations affecting CD166 binding in red and not affecting binding in green.

When examining the two interfaces between the consecutive SRCR domains of CD6 (between CD6 domains 1 & 2 and domains 2 & 3), a similar overall arrangement was observed, however the residues involved in the inter-domain interactions were not conserved. Taken with the variation in sequence length of linker regions found between consecutive SRCR domains in other proteins, it is unclear if this non-linear arrangement of domains is a common structural feature.

Previously published mutagenesis experiments identified a number of residues that are critically important for the interaction with CD166. These, along with residues whose mutation does not affect the binding, were mapped onto the surface representation of the CD6 structure (Figure 2.13) (Bodian et al., 1997; Skonier et al., 1997). The residues that reduced CD166 binding clustered to one region and identified part of the binding site.

2.7.2 Human CD166

Although the construct used for crystallisation experiments contained the three membrane-distal IgSF domains of CD166, determination of the crystal structure only revealed electron density for the two membrane-distal domains in an extended conformation (Figure 2.14), with no space in the asymmetric unit for a third domain. This third domain was most probably proteolytically cleaved during the crystallisation process but this could not be shown by SDS-PAGE because there was insufficient crystallised material.

Previously published CD166 mutagenesis experiments were far more extensive compared to those performed with CD6, identifying both the face and mapping the boundaries of the CD6 interaction site (Figure 2.14B) (Skonier et al., 1996a, 1996b). Mutated residues that affected CD6 binding clustered to the AGFCC'C'' face of domain 1, a face used by many IgSF domains for interaction with other molecules.

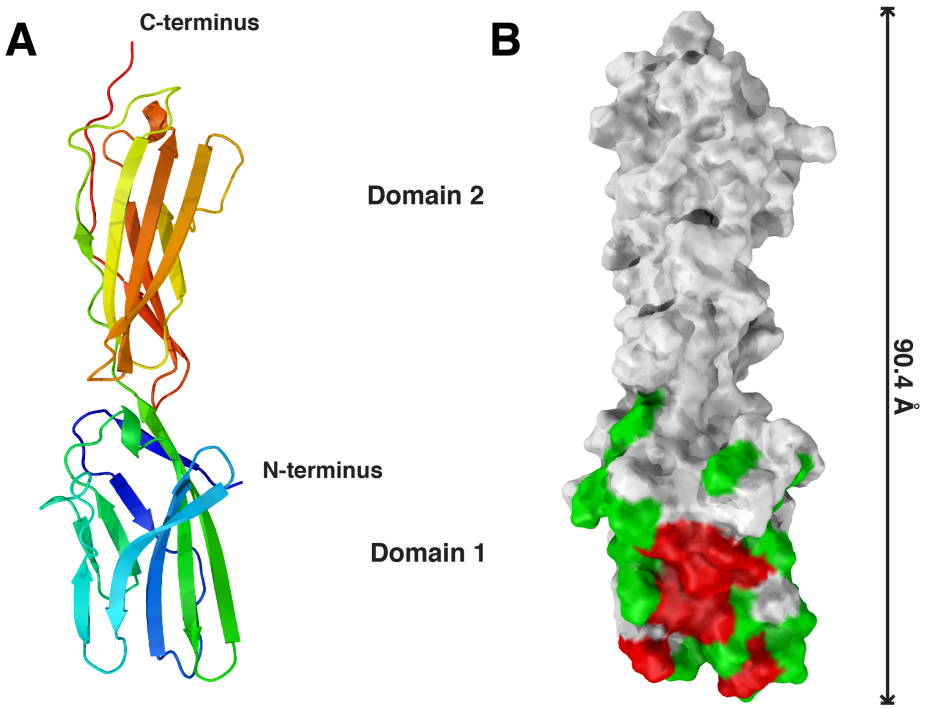

Figure 2.14 - CD166 structure
Panel A: Cartoon representation coloured from blue at the N-terminus to red at the C-terminus. Panel B: Surface representation with mutations affecting CD6 binding in red and not affecting binding in green.

Analysis of crystal packing by PISA revealed a potential homophilic interface that buried 12739 Å^2 (Figure 2.15), consistent with the MALS analysis that showed CD166 existed as both monomer and dimer in solution (Figure 2.6B). Interestingly, this putative homophilic binding site overlapped with residues observed to affect binding to CD6, suggesting that there may be competition between the heterophilic and homophilic interactions.

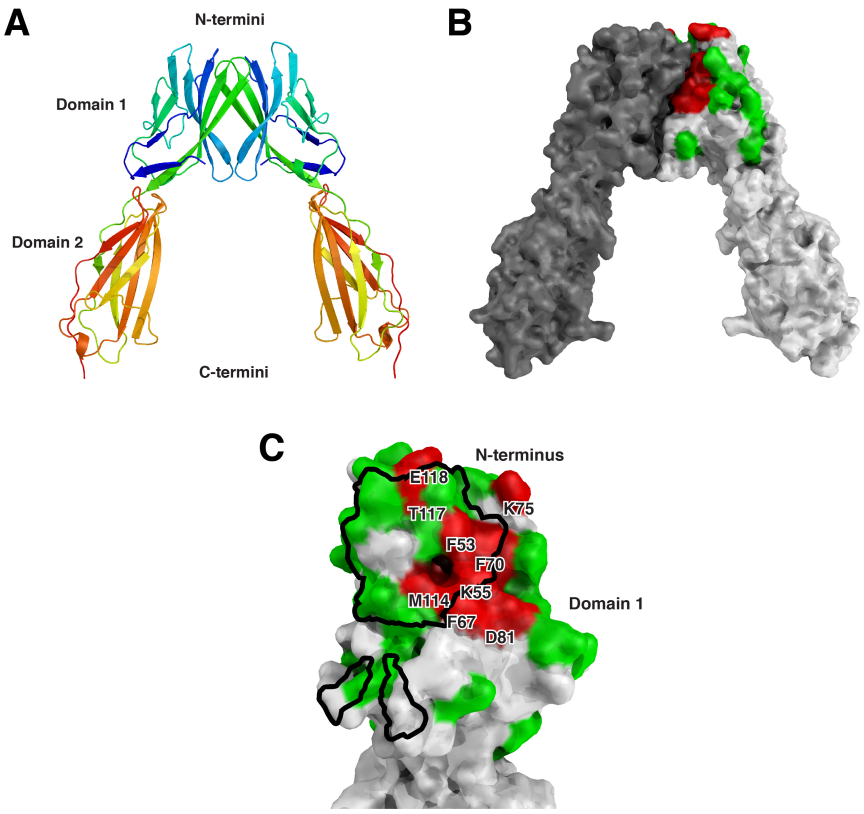

Figure 2.15 - Homophilic CD166 Crystal Contact
Panel A: Cartoon representation coloured from blue at the N-terminus to red at the C-terminus. Panel B: Surface representation with mutations affecting CD6 binding in red and not affecting binding in green. Panel C: Black outline highlights the homophilic interface with mutations highlighted as in panel B.

2.8 Mutagenesis of binding sites

Previously published mutagenesis of CD6 domain 3 and CD166 domain 1 identified critical residues involved in the interaction between the two proteins (Bodian et al., 1997; Skonier et al., 1996a, 1996b, 1997). While the CD166 mutagenesis experiments were comprehensive enough to not only identify the interaction site but also define the boundaries, the CD6 experiments only identified three residues important for the heterophilic interaction and certainly did not identify the boundaries of the interaction site.

Taking into account the new structural information, additional CD6 mutants were designed to perform a more comprehensive mutagenesis experiment with the goal of defining the boundaries of the CD166 interaction site. Surface-exposed

residues surrounding those previously identified to be involved in the interaction with CD166 were chosen for mutation to alanine to assess the contribution of specific residues to the interaction.

Mutagenesis primers were designed and mutants generated using the Q5 Site-Directed Mutagenesis Kit (New England Biolabs). CD6 mutant proteins were transiently expressed in 293T cells as fusion proteins with rat CD4 domains 3 & 4 at their C-termini. Binding of CD6 mutants to wild-type CD166 domains 1-5 expressed by Cho 3.2.8.1 cells was assessed using surface plasmon resonance experiments on a BIAcore 3000 instrument in collaboration with Lee Garner in the Brown group (Sir William Dunn School of Pathology, University of Oxford). Mutants were immobilised via a rat CD4 domains 3 & 4 mAb (OX68) and soluble CD166 injected at approximately the K_D (0.1 μM). Experiments were all performed in HBS-EP running buffer at 20 μL/min at 25°C.

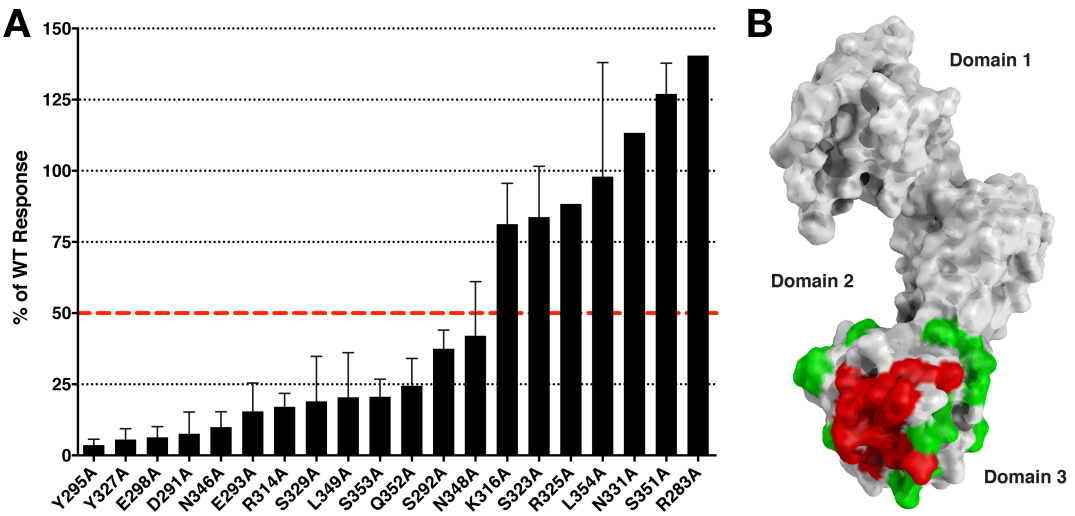

Figure 2.16 - CD6 Mutagenesis
Panel A: Equilibrium binding experiments performed at 25°C with soluble CD166d1-5 injected at 0.1 μM (approximately the K_D) over CD6-CD4d3+4 chimeric mutants immobilised by OX68 mAb. The equilibrium binding data were plotted as the mean percentage (±S.E.M.) of the wild type response, normalised to 100%. Equivalent levels of mutant protein were immobilised with levels were monitored by binding of OX126, a CD6 domain 3 mAb. A response of 50% or lower was used to define residue involvement in binding. Panel B: Data from panel A and previously published data are mapped to on a surface representation of the CD6 structure. Residues affecting CD166 binding are shown in red and those that had no effect in green.

Mutants that resulted in a binding response of 50% or lower than the wild type response were defined as being involved with CD166 binding. Mapping these residues along with those that did not perturb CD166 binding on to the CD6 structure surface revealed the boundaries of the CD166 interaction site on the surface of CD6 (Figure 2.16).

2.9 Cross-species binding

It has been previously reported that human CD166 binds human, rat and mouse CD6 with the same affinity (Bowen et al., 1997, Brown unpublished). These sequences were subjected to a multiple sequence alignment with Clustal Omega (Sievers et al., 2011) with sequence differences mapped to a surface representation of the CD6 structure (Figure 2.17). There was extensive sequence conservation in CD6 domain 3 that clustered to the region identified by mutagenesis experiments as the CD166 interaction site, further supporting the key role these residues play in the CD6/CD166 interaction.

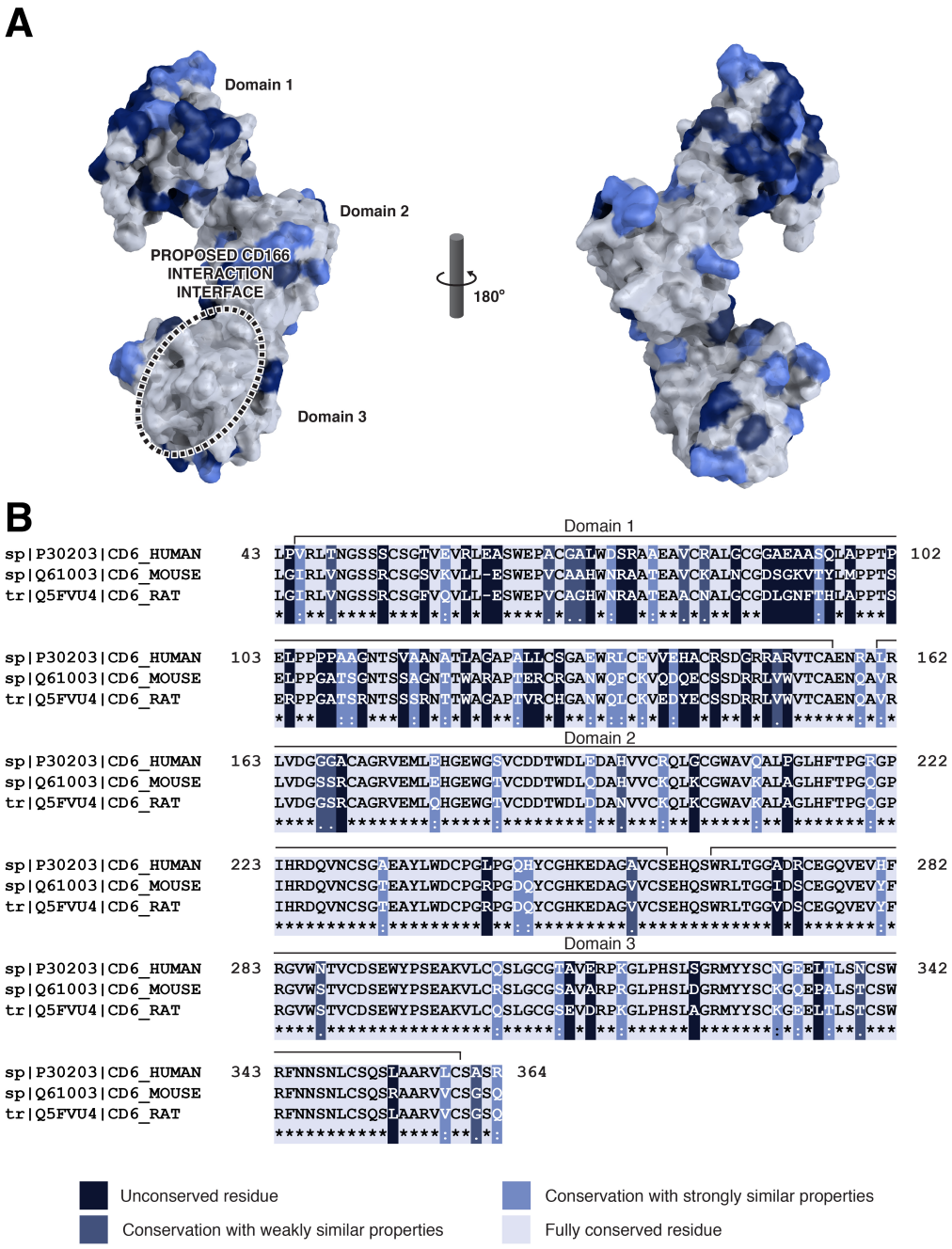

Figure 2.17 - Cross species binding of CD6 to human CD166
Panel A: Sequence conservation between species clusters to the proposed CD166 interaction surface. Panel B: Clustal Omega (Sievers et al., 2011) sequence alignment of human, mouse and rat CD6.

2.10 Electrostatic analysis of binding sites

The electrostatic potentials of the heterophilic interaction sites on CD6 and CD166, calculated by APBS (Dolinsky et al., 2007), revealed charge complementarity of the proposed binding sites (Figure 2.18). CD6 featured a stripe of negative charge flanked by positive charge and CD166, a complementary positively charged stripe flanked by negative charge. This charge complementarity suggested electrostatic potential might be important for the interaction, consistent with the observation that charge reversal mutations in the binding sites of both CD6 and CD166 disrupt the interaction (Bodian et al., 1997; Skonier et al., 1996a, 1996b, 1997 and data presented in this chapter).

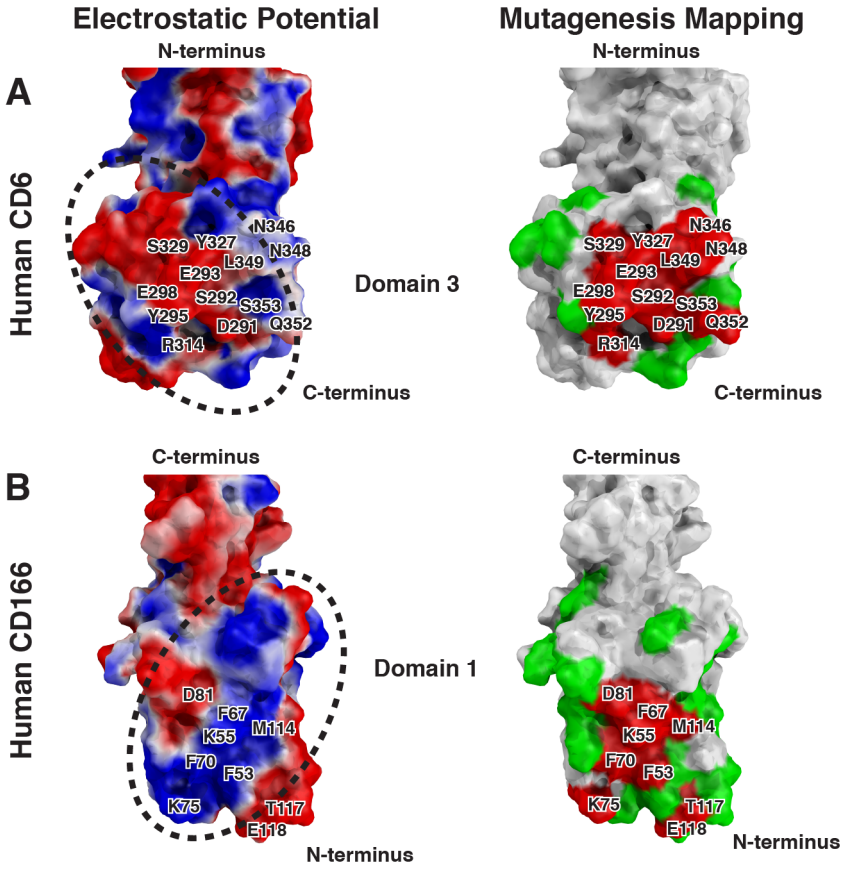

Figure 2.18 - Electrostatic analysis of binding sites
Surface representations showing complementary electrostatic potential between the binding faces of CD6 and CD166 as calculated with APBS [positive charge in blue and negative charge in red] (Panels A & B). Panels on the right show mutations affecting the heterophilic interaction in red and those that do not in green.

2.11 CD6 single nucleotide polymorphisms

Numerous CD6 SNPs have been identified with a number found to be associated with susceptibility to multiple sclerosis (MS). The CD6 SNPs with the strongest MS associations are found in intronic DNA sequences, apparently affecting the ratio of full length CD6 to CD6Δd3 (Kofler et al., 2011). However, weaker associations with susceptibility to MS have been found with two of the five non-synonymous CD6 SNPs (R225W and A257V). These two residues are either buried, or predominantly buried, in the domain structure and therefore sequence variation may disrupt the overall stability of the molecule (Figure 2.19).

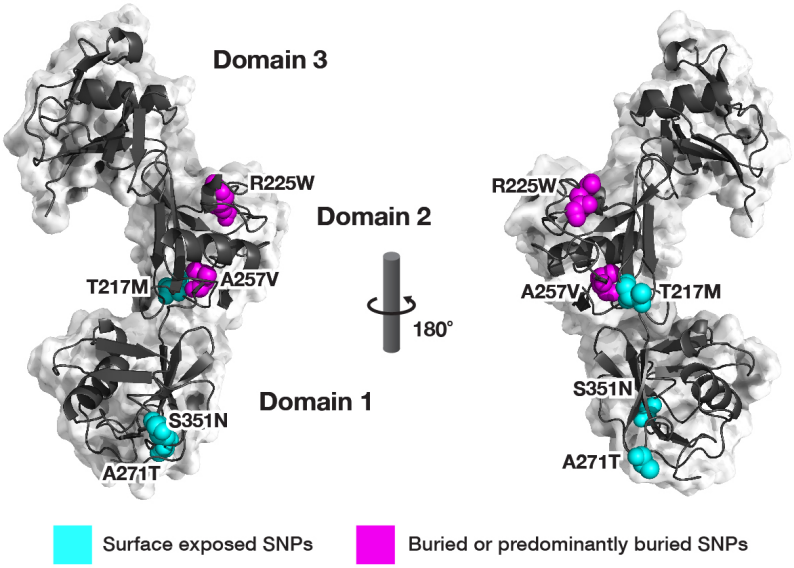

Figure 2.19 - CD6 non-synonymous single nucleotide polymorphisms
Non-synonymous SNPs associated with disease are not located in the CD166 binding interface. SNP residues are displayed as spheres with those surface exposed in cyan and buried in magenta.

A CD6 SNP, found in 10% of genomes analysed (1000 genomes.org) [S351N], was located within the proposed CD166 binding site and, although there are currently no known disease associations, it appeared to have the potential to affect CD166 binding. To test if this SNP did indeed affect CD166 binding, two additional CD6 mutants (S351N and S315A) were generated and expressed at fusion proteins with rat

CD4 domains 3 & 4 as in the previous mutagenesis experiments. SPR binding experiments were carried out in collaboration with Lee Garner in the Brown Group (Sir William Dunn School of Pathology, University of Oxford). Analysis of the SPR data showed that the S351A variant had a similar affinity for CD166 as the wild-type protein (K_D = ~0.1 μM), whereas the S351N variant had a 5-fold reduced affinity (K_D = ~0.5 μM) (Figure 2.20).

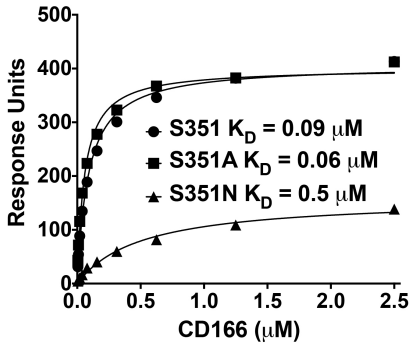

Figure 2.20 - The CD6 SNP S351N reduces affinity for CD166
Panel A: Equilibrium binding analysis of CD166 to CD6 S351, S351A and S351N. K_D values were calculated using curve fitting in Graphpad Prism.

The mutation of serine to asparagine at this sequence position in CD6 introduced a consensus N-linked glycosylation site (NQS). Given that a serine to alanine mutation did not affect CD166 binding, it is possible that the difference in affinity for the S315N variant was due to steric hindrance resulting from glycosylation of this site. In order to determine whether or not this site was glycosylated as a result of the SNP, mass spectrometry was carried out in collaboration with Clive Metcalf (Barclay Group, Sir William Dunn School of Pathology, University of Oxford). Due to the unknown mass of the N-linked carbohydrate, glycosylated peptides were not identifiable by LC-MS-MS. In order to determine whether this site was glycosylated or not, the precursor ion intensity obtained from extracted ion chromatograms of the FNNSNLC**S**QSLAAR and FNNSNLC**N**QSLAAR tryptic peptides before and after deglycosylation were quantified. The S351 peptide was detected both before and after deglycosylation therefore indicating the absence of glycosylation. In contrast, the

S351N peptide was only detected after deglycosylation, indicating that this peptide is indeed glycosylated (Figure 2.21).

Taking into account both the SPR and mass spectrometry data, it was likely that the interaction with CD166 S351N was perturbed due to steric hindrance caused by the presence of a carbohydrate attached to the N-linked glycosylation site introduced by the SNP.

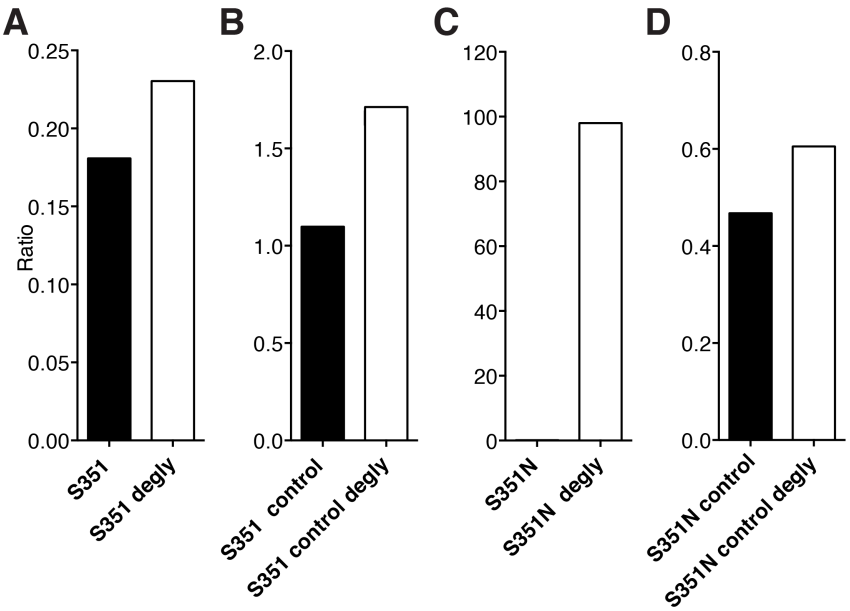

Figure 2.21 - Glycosylation of the CD6 SNP, S351N
Average ratios of total extracted precursor ion area of peptides to precursor ion areas of CD6 control peptides that do not contain glycosylation sites. Panels A & B: S351 [FNNSNLCSQSLAAR] and S351 control peptide. Panels C & D: S351N [FNNSNLCNQSLAAR] and S351N control peptide. Black bars indicate before deglycosylation and white bars after.

2.12 Native Mass Spectrometry

Native mass spectrometry experiments were carried out in collaboration with Jun Yan in Carol Robinson's group (Department of Chemistry, University of Oxford) to investigate the possibility of competition between the heterophilic (CD6/CD166) and homophilic (CD166/CD166) interactions suggested by the overlap of the heterophilic binding site on CD166 with a homophilic binding interface observed in the CD166 crystal structure. If the homophilic interface was physiologically relevant and not a crystal packing artifact, homophilic CD166 interactions would be disrupted by the formation of heterodimers with CD6.

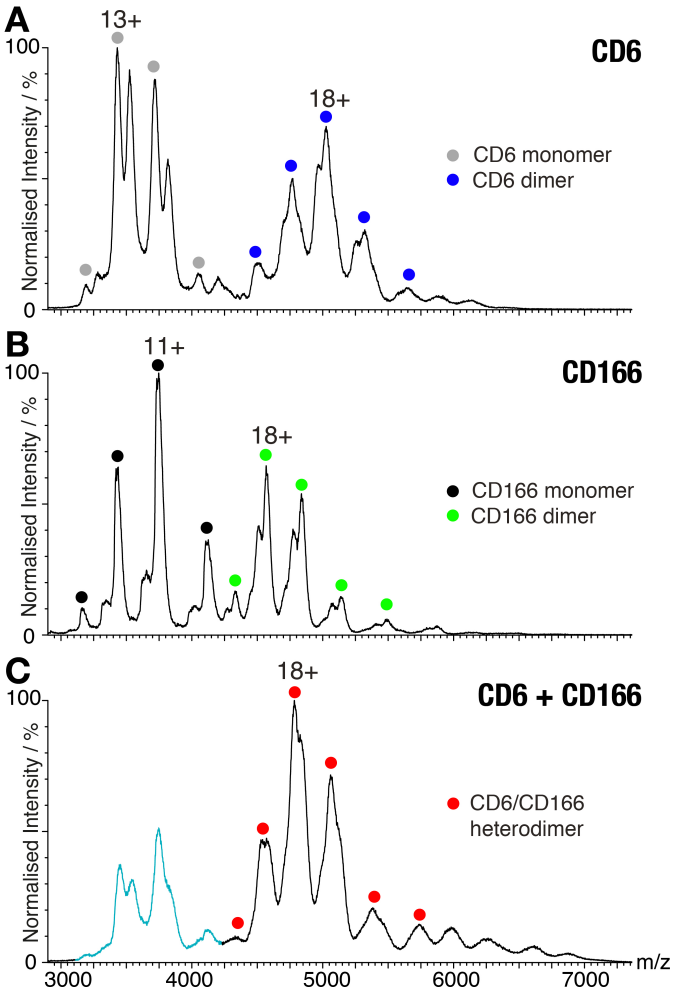

Figure 2.22 - Native Mass Spectrometry
Heterodimers are formed at the expense of CD6 and CD166 homodimers. Native mass spectra shown for CD6 (A), CD166 (B) and CD6/CD166 1:1 mixture (C). Where more than one mass was found for a species due to glycosylation, only the most abundant one is indicated for clarity. Calculated masses are shown in Table 9.

CD6	**Calculated Mass (Da)**	**Mass Error (Da)**
Monomer 1	44509	±3
Monomer 2	45721	±4
Dimer 1	89315	±13
Dimer 2	90173	±16
Dimer 3	91436	±13
CD166		
Monomer 1	39751	±8
Monomer 2	40139	±12
Monomer 3	40975	±4
Monomer 4	41198	±6
Dimer 1	80045	±9
Dimer 2	81046	±24
Dimer 3	82112	±24
CD6 + CD166		
Dimer 1	84557	±40
Dimer 2	85592	±32
Dimer 3	86770	±24

Table 2.14 - Native mass spectrometry calculated masses

Native mass spectrometry experiments were carried out with each protein separately and a 1:1 mixture of both (Figure 2.22 and Table 2.14). Individual proteins were at concentrations of 20 μM and the 1:1 mixture was produced by incubating equal amounts of CD6 and CD166, both at 20 μM, for 30 minutes at room temperature. The samples were then subjected to buffer exchange from 10 mM HEPES (pH 7.4), 150 mM NaCl into 250 mM ammonium acetate (pH 7.6) using Biospin-6 columns (Bio-Rad). Analysis of individual protein samples revealed that, individually, both CD6 and CD166 were predominantly in monomer-dimer equilibrium. This was consistent with the CD166 MALS analysis that showed both monomer and dimer species (Figure 2.6B). It also suggested that the slope in the molecular weight observed across the CD6 peak was likely due to self-association (Figure 2.6A). The 1:1 mixture revealed that upon mixing the two proteins, heterodimers were formed at the expense of CD6 and CD166 homodimers. The additional peak observed in the MALS analysis of the two proteins in solution together was therefore likely a heterodimer (Figure 2.6C).

2.13 Discussion

The structures presented in this chapter provide molecular insights into the heterophilic CD6/CD166 interaction as well as a possible CD166 homophilic binding interface. The CD6 structure is the first experimentally determined molecular structure of consecutive SRCR domains and reveals a distinctive non-linear domain organisation. Mapping previously published mutagenesis experiments to the newly solved X-ray crystal structures revealed the surfaces involved in the heterophilic interaction on both molecules (Bodian et al., 1997; Skonier et al., 1996a, 1996b, 1997). The CD6 binding site on the membrane distal extracellular domain of CD166 (domain 1) was clearly identified with defined boundaries, however the CD166 binding site on the membrane proximal domain of CD6 (domain 3) was not so well defined. By performing additional and more extensive CD6 mutagenesis experiments utilising the new structural information, the binding site for CD166 was more comprehensively delineated. Electrostatic analysis of the two binding site surfaces revealed complementary charge patterns consistent with the observations in the mutagenesis experiments that charge reversal mutations disrupted heterophilic binding. The ligand binding interactions of the SRCR domain-containing proteins MARCO and CD163 also involves complementary electrostatic potential (Nielsen et al., 2013; Ojala et al., 2007).

CD166, or activated leukocyte cell adhesion molecule (ALCAM), in addition to binding CD6, can participate in homophilic interactions that mediate cell adhesion (van Kempen et al., 2001; Nelissen et al., 2000; Te Riet et al., 2007). The potential homophilic binding interface identified in the CD166 crystal structure may be important for these trans CD166 homophilic interactions, especially as the interface involves the membrane distal IgSF domain of CD166, which is proposed to mediate the homophilic interaction (van Kempen et al., 2004). Significant conformational flexibility in the extracellular domains of CD166 would be required to allow the molecules to adopt a

conformation that would allow this homophilic interface to form between two different cells. Interestingly, this homophilic interface overlaps with the CD6 binding site.

The CD6/CD166 heterophilic interaction, mediated by the membrane proximal SRCR domain of CD6 and the membrane distal IgSF domain of CD166, occurs in trans between apposing cells and is important for thymocyte development and T cell activation (Bott et al., 1993; Bowen et al., 1996; Gangemi et al., 1989; Osorio et al., 1994; Patel et al., 1995; Whitney et al., 1995). The heterophilic interaction has a higher affinity (K_D = 0.4-1.0 μM) than the homophilic CD166 interaction (K_D = 29-48 μM) (Hassan et al., 2004). Native mass spectrometry experiments showed that CD6/CD166 heterodimers were formed at the expense of both CD6 and CD166 homodimers, suggesting that the CD166 homodimeric interface may be physiologically relevant, although steric hindrance cannot be ruled out. Trans homophilic CD166 interactions were blocked by preventing clustering on the same cell surface with an antibody specific for the membrane proximal extracellular domains of CD166 and also by a truncated form of CD166 that lacked the N-terminal IgSF domain (van Kempen et al., 2004, 2001; Weidle et al., 2010). This suggests that weak interactions in cis at the same cell surface might promote the homophilic interactions in trans that lead to ALCAM-mediated cell adhesion. The projection of the membrane distal SRCR domain that results from the non-linear organisation of the consecutive SRCR domains may be important for preventing the clustering of CD166 in cis and provide a mechanism to disrupt CD166 trans homodimers and promote the heterophilic interaction.

The CD6 homophilic interaction detected by native mass spectrometry most likely occurs in cis on the same cell surface, as is the case for the closely related protein, CD5 (Brown and Lacey, 2010). The non-linear arrangement of the extracellular SRCR domains of CD6 may allow individual domains to participate in distinct interactions in different orientations. For example, in addition to interaction of the membrane proximal domain of CD6 with the membrane distal domain of CD166,

other SRCR domains may be involved in homophilic interactions. A truncated form of CD6, lacking the membrane distal N-terminal domain of CD6, was not stable at the cell surface. However, in contrast, a form of CD6 lacking the membrane proximal domain that is implicated in CD166 binding, was stably expressed at the cell surface (Castro et al., 2007; Kofler et al., 2011; M. H. Brown unpublished data). This suggests that CD6 homophilic interactions may be important for the stabilisation of CD6 as dimers at the cell surface.

The roles of the two membrane distal extracellular SRCR domains of CD6 (domains 1 and 2) remain to be definitively determined. As mentioned, the projection of domain 1 may help disrupt trans homophilic CD166 interactions, thereby promoting formation of heterophilic interaction. The role of domain 2 may be to act as a spacer to maintain this projection of domain 1, however there are also other alternatives. The presence of a conserved triacidic motif (D187/D188/E253) in domain 2 suggests there may be a divalent cation-dependent interaction that has been as yet unidentified (Nielsen et al., 2013) because the interaction with CD166 occurs in the absence of cations (Hassan et al., 2004; Patel et al., 1995; L. I. Garner unpublished data). It has also been reported that CD6 is capable of direct pathogen binding (Sarrias et al., 2007). The consensus sequence of a bacterial binding peptide motif first identified in an SRCR domain of DMBT1 is found in domain 2 of CD6 and is surface exposed opposite to the CD166 binding face (Bikker et al., 2004; Madsen et al., 2010). It should be noted, however, that mutation of this consensus sequence in the WC1 protein did not affect bacterial ligand binding (Hsu et al., 2015). There are also data suggesting subsidiary roles for domains other than CD6 domain 1 and CD166 domain 1 in the heterophilic interaction but these are below the detection threshold of SPR analysis (Bowen et al., 1996).

During T cell activation both CD6 and CD166 have been observed to move into areas of close contact between T cells and antigen presenting cells (Castro et al.,

2007; Gimferrer et al., 2004). The dimensions of interacting receptors at these close contacts are restricted to no more than the ~140 Å spanned by the TCR and associated receptors (Dushek et al., 2012). The non-linear organization of the consecutive SRCR domains of CD6 resulted in an overall length of 78.1 Å (excluding the 37 amino acid membrane proximal stalk) and may be important for its accommodation in these areas of close contact. The two IgSF V-set domains of CD166 spanned 90.4 Å and by extrapolating (using the dimensions of C-set IgSF domains), the overall linear array of IgSF domains in a fully extended conformation of CD166 would span ~200 Å. Flexibility in these domains would therefore be required in order to accommodate this molecule in areas of close proximity, as is proposed for E-cadherins (Wu et al., 2011).

The strongest association of a CD6 SNPs with multiple sclerosis is located in an intron. Although this SNP does not cause a decrease in overall CD6 levels on the cell surface, it does causes a reduction in the ratio of full-length CD6 to a form of CD6 that lacks domain 3 (the CD166 binding domain), suggesting the effect is mediated by decreased ligand binding (Kofler et al., 2011). There are also MS associations with the non-synonymous SNPs R225W and A257V that in combination reduce expression levels, thereby lowering the threshold of activation and ligand. These two residues are either completely buried or predominantly buried in the CD6 protein structure and therefore their mutation would be expected to have an effect on protein domain stability. The mutagenesis and mass spectrometry experiments have shown that the less common non-synonymous SNP, S351N, reduces the affinity for CD166 due to the introduction of an N-linked glycosylation site and subsequent glycosylation at S315N. This SNP will presumably have functional effects due to the effect on ligand binding.

The non-linear organisation of the three SRCR domains of CD6 may have implications for other proteins that contain consecutive SRCR domains. Although the

residues at the interfaces between domains are not conserved, the general orientation of interfacing domains in CD6 is similar. If this non-linear arrangement is a common structural feature of consecutive SRCR domains, longer arrays of SRCR domains will form helical structures. One such example is AIM (Spα, CD5L), a group B SRCR family member, whose interaction with IgM depends on avidity with contributions from each SRCR domain (Arai et al., 2013).

2.14 Summary

Complementing the new structural information with mutagenesis, SPR and mass spectrometry experiments, the heterophilic binding sites on both CD6 and CD166 have been well defined. The crystal structure of CD166 identified a potential homodimer interface that overlaps with the heterophilic binding site, suggesting competition between the two interactions. Native mass spectrometry analysis revealed that heterodimers of CD6/CD166 are indeed formed at the expense of weaker homodimeric interactions. This is likely due to the overlapping binding sites, however steric hindrance cannot be ruled out. Non-synonymous SNPs associated with susceptibility to MS are either completely buried or partially buried in domain 2 of CD6, consistent with their alteration disrupting domain structure thereby reducing stability and expression levels on the cell surface. The CD6 SNP S351N affects CD166 binding through glycosylation of the N-linked glycosylation site introduced by the SNP. This SNP was found in 10% of genomes analysed, but there are no current known disease associations.

2.14.1 Future directions

The next stage in this project is to determine the crystal structure of the heterodimer complex of CD6/CD166. This would reveal the molecular details of an unusual interaction between an SRCR domain and IgSF V-set domain that has

important functional effects in the adaptive immune system. The original CD166 construct containing the three IgSF domains was most likely proteolytically cleaved during the first set of crystallisation trials, thereby reducing the homogeneity of the sample and likelihood of crystals of the heterodimer forming. To reduce the problem of sample heterogeneity, recent attempts to obtain crystals of the heterodimer included the generation of stable S2 (Drosophila) cell lines (Moraes et al., 2012) expressing the two membrane distal IgSF domains of CD166. This expression system also had a considerably greater yield of soluble CD166 protein. The most promising crystal lead came from the use of CD166 protein purified from the S2 cells with a 2 times excess of CD6 expressed in the original Cho expression system. A new crystal form was grown that had unit cell dimensions compatible with a heterodimer, however the data were of insufficient resolution to determine a structure.

Chapter 3: Chicken MHC class I molecules

3.1 Introduction

T cells recognise protein antigens through their TCRs as short peptides bound to highly polymorphic class I or class II MHC molecules. These molecules are capable of binding both self and non-self peptide antigens allowing them to activate T cell responses against both invading pathogens and transformed cells. Class I MHC molecules present predominantly cytoplasmic derived, endogenous peptides whereas class II molecules present predominantly exogenous peptide fragments. Class I MHC molecules are also capable of presenting exogenous peptides through a process called cross-presentation (Shen and Rock, 2006). Other non-classical MHC proteins, such as CD1d, are capable of presenting lipid antigens to T cells (Cohen et al., 2009).

Class I MHC peptide antigen originate in the cytosol and result from degradation by the proteasome (Maupin-Furlow, 2012). The majority of these peptides are derived from polypeptides that never reach their native structures due to translational or post-translational errors, so called defective ribosomal products (DRiPs) (Yewdell, 2011). DRiPs account for ~30% of all newly synthesized proteins and are rapidly degraded by the proteasome (Schubert et al., 2000). Peptides can also be derived from fully folded proteins, for example, the listerolysin protein secreted by Listeria monocytogenes following internalisation by macrophages can be detected by $CD8^+$ T cells (Villanueva et al., 1995). The C-termini of most peptides bound to class I MHC molecules on the cell surface are generated by proteasomal cleavage (Cascio et al., 2001).

A number of cytosolic protein chaperones have been identified that are crucial for antigen processing and presentation by class I MHC molecules. Knockdown of Hsp90α, a chaperone protein, or CHIP (carboxyl terminus of Hsp70 interacting

protein), which ubiquitinates Hsp90α-associated proteins, inhibited antigen presentation and recognition by $CD8^+$ T cells in a pre-proteasomal manner (Kunisawa and Shastri, 2006). Other chaperone proteins, such as the group II chaperonin TRiC, appear to be important for protecting proteolytic intermediates from further degradation after proteasomal cleavage (Kunisawa and Shastri, 2003).

Following their generation in the cytosol, peptide fragments are transported into the ER by TAP (transporter associated with antigen presentation), a member of the ATP-binding cassette (ABC) family of proteins (Oancea et al., 2009). The heterodimer TAP consists of two monomers, TAP1 and TAP2, with each subunit consisting of a central core domain of 6 transmembrane helices that form a channel, a nucleotide binding domain (NBD) and a number of other N-terminal domains not seen in other ABC transporters (Gaudet and Wiley, 2001; Hinz and Tampé, 2012). TAP1/2 negative cells express very low levels of cell surface class I MHC molecules due to the absence of translocated peptides in the ER (Gabathuler et al., 1994; Lampen et al., 2010). The TAP heterodimer associates with a number of other protein components to form the peptide-loading complex (PLC). Two molecules of tapasin interact with each TAP subunit within the membrane and act as a bridge to recruit MHC class I heterodimers and calreticulin (CRT) (Koch et al., 2004; Leonhardt et al., 2005; Procko et al., 2005; Sadasivan et al., 1996; Tan et al., 2002). Tapasin negative cells have reduced cell surface expression of class I MHC molecules, however the expression level reduction is less extreme than for TAP1/2 negative cells (Garbi et al., 2000; Grandea et al., 2000; Peh et al., 2000; Purcell et al., 2001). Each tapasin molecule is disulfide bonded to ERp57, a protein disulfide isomerase homolog (Dong et al., 2009). ERp57, along with CRT and calnexin (CNX) are normally involved in a quality control mechanism that mediates disulfide bond isomerisation to assist in the folding of newly synthesised glycoproteins (Hebert et al., 2005). Binding to CRT or CRT is dependent on a single glucose residue that can be either removed by glucosidase II or replaced by UDP-

glucose glycoprotein transferase I (UGT1) (Caramelo et al., 2004; Ritter et al., 2005; Soldà et al., 2007). In the specific case of class I MHC peptide loading, tapasin and CRT play a role in peptide editing and optimising the binding of high affinity peptides at the expense of low affinity peptides (Howarth et al., 2004; Wearsch and Cresswell, 2007; Wearsch et al., 2011; Williams et al., 2002; Zhang et al., 2011). Although ERp57 has disulfide isomerase activity, this is not important for class I MHC peptide loading. The structure of the luminal region of human tapasin conjugated to ERp57 (Dong et al., 2009) showed that tapasin obstructs both of the active sites of ERp57 and cysteine to serine mutations had no effect on peptide loading (Peaper and Cresswell, 2008). Mutagenesis experiments revealed that, in addition to CRT dependent interactions between ERp57 and the MHC class I molecule, the MHC molecule also interacts directly with tapasin (Dong et al., 2009). The current data suggests a mechanism in which monoglucosylated MHC class I molecules are recruited to the PLC by CRT in an interaction that is sensitive to peptide occupancy. If a high affinity peptide is bound, the affinity of the MHC molecule for tapasin is reduced and the single glucose residue is removed by glucosidase II, allowing the complex to be transported through the golgi to the cell surface. If, instead, a low affinity peptide is bound, the peptide either dissociates, resulting in reglucosylation by UGT1, or the low-affinity bound peptide causes a conformational change that is detected by UGT1, again resulting in reglucosylation by UGT1. The reglucosylation event causes the MHC class I molecule to reassociate with CRT and the PLC where peptide binding can occur once again.

MHC class I molecules usually bind peptides of 8 to 10 amino acids in length, however, TAP is capable of translocating much longer peptides into the ER (Koopmann et al., 1996). These longer peptides are trimmed in the ER at their N-termini by ERAAP (ER aminopeptidase associated with antigen processing) (ERAP1/2 in humans) (Saric et al., 2002; Saveanu et al., 2005; Serwold et al., 2002). ERAP1

does not over trim peptides due to its activity requiring a structural change that can only be induced by longer peptides (Nguyen et al., 2011). ERAAP knockout mice feature much less stable class I MHC peptide complexes with much longer bound peptides. Interestingly, wild type and knockout mice on the same background are histoincompatible due to a drastic alteration in the repertoire of bound peptides (Hammer et al., 2007).

Peptide loading onto class II molecules occurs via a different mechanism to class I molecules. Class II MHC molecules in the ER associate with an invariant chain that binds in the peptide binding groove and targets the molecule's release from the ER into late endosomes (Romagnoli and Germain, 1994). The invariant chain is then processed by proteases such as cathepsins leaving a peptide fragment, known as CLIP, occupying the peptide binding groove (Ghosh et al., 1995). Association of the class II MHC-CLIP complex with the DM chaperone protein facilitates the exchange of the CLIP peptide for a high affinity peptide generated by late endosomal proteases and peptidases (Denzin and Cresswell, 1995; Pos et al., 2012; Roche and Cresswell, 1991).

Class I MHC complexes are comprised of a highly polymorphic heavy (α-) chain, an invariant chain (β_2-microglobulin) and a peptide fragment (Figure 3.1). The heavy chain features three alpha domains, of which the α-1 and α-2 domains make up the peptide-binding groove. The vast majority of the sequence polymorphisms are clustered to the peptide-binding groove, which consists of a beta sheet that forms the floor of the peptide-binding groove and a number of alpha helices that enclose the groove from either side (Figure 3.1A & 3.1B). Although electron density, discrete from the protein, was observed in the peptide-binding groove of the first MHC class I X-ray crystal structure, it could not be modelled due to a heterogeneous mixture of peptides present in the peptide-binding grooves of the molecules in the crystal (Bjorkman et al., 1987a, 1987b). The crystal structures of the murine class I MHC molecule H-2KB, in

complex with either an octamer or nonamer peptide, offered the first insight into the mechanism of peptide binding to the peptide-binding groove (Fremont et al., 1992). Both peptides feature the same amino- and carboxyl-terminal interactions, with the nomamer peptide featuring a central bulge allowing accommodation of the extra amino acid residue in the peptide. These structures also showed that most of the specific interactions occur between the side chain atoms of the MHC molecule and main chain atoms of the peptide (Fremont et al., 1992).

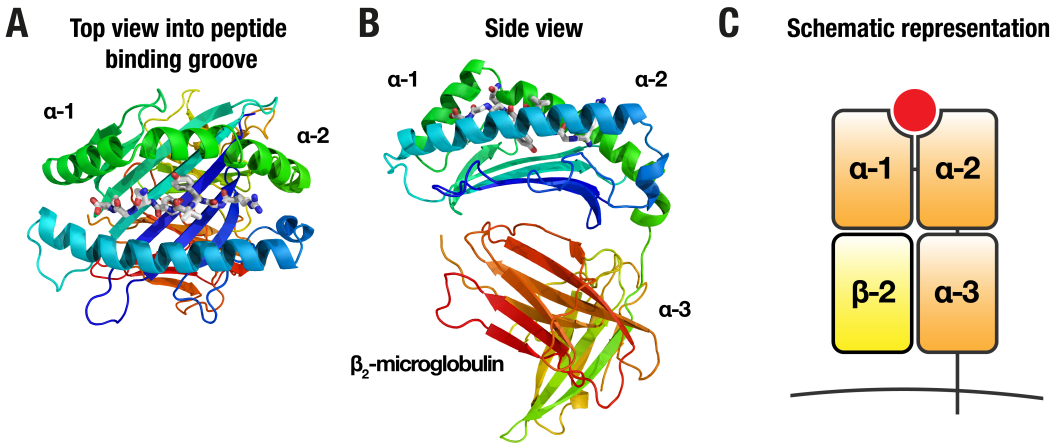

Figure 3.1 - Structure of murine MHC class I in complex with peptide.
The first crystal structure of an MHC class I molecule in which the bound peptide could be modelled (PDBID: 2vaa) (Fremont et al., 1992). Panels A & B: MHC protein molecule shown in cartoon representation with bound peptide shown in stick representation. Panel C: Schematic representation of MHC molecule and bound peptide (red circle).

Following the first structure determination of an MHC molecule, the general principles of peptide binding were established. Further structural analysis showed that mammalian class I MHC binding grooves can be divided into six distinct pockets, A to F, into which bound peptides make conserved and polymorphic interactions. Peptides are anchored into the binding groove through interactions of their side chains with MHC residues in two or three of these pockets. Interactions usually occur between the side chain of peptide position 2 with pocket B, position 5 with pocket C or position

6 with pocket D and the C-terminal peptide position with pocket F. Class I MHC molecules usually bind octamers or nonamers, however, it is possible for them to bind peptides up to 11 amino acids in length. In such cases, the central region of the peptide bulges out of the groove while still maintaining the conserved interactions and the N- and C-termini. These studies also highlighted the importance of water molecules in peptide binding (Guo et al., 1992; Matsumura et al., 1992; Natarajan et al., 1999; Schumacher et al., 1991; Speir et al., 2001; Wilson and Fremont, 1993).

Sequence polymorphisms located within the peptide-binding groove result in different MHC molecules binding different repertoires of peptides. In order for a peptide to bind to a particular MHC molecule, the peptide must be compatible with the molecule's peptide binding motif. This motif specifies at least two so-called anchor positions where peptide amino acids are restricted to a single or very restricted set of amino acids. It is these anchor positions that the peptide side chains interact with pocket residues of the MHC molecule, tethering the peptide into the groove (Figure 3.2) (Rammensee et al., 1995).

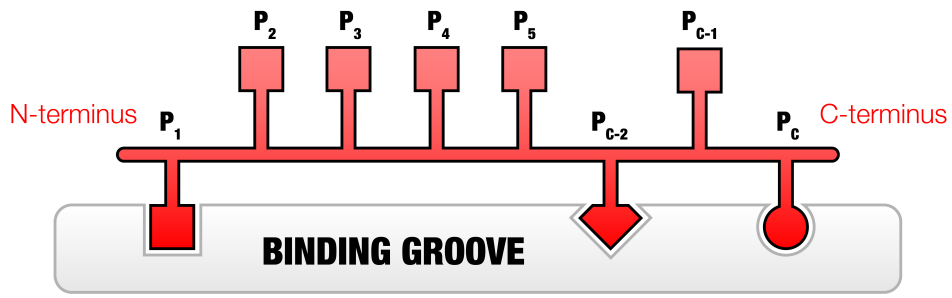

Figure 3.2 - Peptide binding motifs
Schematic representation of peptide binding to a class I MHC binding groove. Peptide coloured in red with side chains represented as shapes attached to a peptide backbone. Peptide residues are numbered from the N-terminus. Due to differences in peptide length and bulging in the central portion of some peptides, peptide positions at the C-terminus are labeled relative to their position from the most C-terminal residue.

Due to antigen presentation to T-cells requiring peptide presentation in complex with MHC molecules, prediction of peptide binding to MHC molecules is very important for vaccine design and immunotherapies. For example, many vaccination trials attempting to boost immune responses against cancer use peptides predicted to bind HLA-A2, a class I MHC molecule which 60% of the Caucasian population are thought to possess (Jorritsma et al., 2009). Peptide binding prediction methods rely on statistical, structural or artificial machine learning systems (e.g. artificial neural networks, a type of model inspired by biological neural networks) (Brusic et al., 2004; Firebaugh, 1988; Schalkoff, 1992; Trost et al., 2007). The majority of these prediction methods require fixed peptide lengths, a restraint that is inappropriate for class I molecules as they can accommodate longer peptides that bulge out of the peptide-binding groove (Guo et al., 1992). Such prediction methods may fail to predict shorter or longer peptides that are capable of binding. Some more recent methods do not depend on peptide length and are therefore more appropriate for predicting class I peptide binding (Meydan et al., 2013).

The highly polymorphic MHC class I heavy (α-) chain, part of which forms the peptide-binding groove, is encoded within the MHC locus. Comparison of this locus in chicken and mammalian genomes revealed several key differences (Kaufman et al., 1999a; The MHC sequencing Consortium, 1999) (Table 3.1).

The chicken MHC is much smaller than the human MHC, containing only two classical class I genes (BF1 and BF2) in comparison to three in the mammalian locus (HLA-A, HLA-B and HLA-C). The chicken MHC also undergoes very little recombination in comparison to its human counterpart. Another important difference is in the arrangement of the two MHC loci that results in the TAP and tapasin genes being located in close proximity to the two class I genes of the chicken MHC, which is not the case in the human MHC. As a result, only one of the two chicken MHC class I genes (BF2) is predominantly expressed due to co-evolution with the TAP transporter

(Kaufman et al., 1999b; Shaw et al., 2007; Wallny et al., 2006). Chicken TAP genes are highly polymorphic and it has been shown that the TAP peptide translocating specificity matches the peptide motif of the dominantly expressed class I MHC molecule (Walker et al., 2011). In contrast, the mammalian TAP transporter has evolved to provide peptides to all three classical MHC class I molecules (Kaufman et al., 1995, 1999a; The MHC sequencing Consortium, 1999).

Characteristic	Chicken MHC	Mammalian MHC
Size	92kb	3.6mb
Total number of genes	19 genes	130 genes
Recombination events	Very little / none	Many
Classical class I MHC genes	2 (BF1 and BF2) with BF1 predominantly expressed, and BF2 not expressed at all in some haplotypes, low level of expression in others.	3 (HL-A, HL-B, HL-C) of which all are expressed.
Genomic organisation	Class III region is located outside of the class I and class II regions. TAP genes are located between the two class I MHC genes (BF1 and BF2).	Class III region located between the class I and class II regions. Transport associated with antigen presentation (TAP) genes are located in the extended class II region, separated from the class I genes by significant physical and recombinational distance.
Transport associated with antigen presentation (TAP)	Polymorphic sequences.	Monomorphic sequences.

Table 3.1 - Comparison of chicken and mammalian MHC

The dominant expression of a single class I chicken MHC molecule leads to striking associations with either resistance or susceptibility to infectious pathogens. Determination of the peptide motifs for the dominantly expressed chicken MHC class I molecule (BF) from the B4, B12 and B15 haplotypes explained the striking MHC-determined resistance and susceptibility of chickens to Rous sarcoma virus. Using the resistant B12 haplotype's peptide-binding motif, it was predicted that this class I MHC molecule would bind 10 times more peptides than the class I MHC molecules of the

susceptible B4 and B15 haplotypes (Wallny et al., 2006). Marek's disease, an economically important disease that is estimated to cost the agricultural industry up to $2 billion USD each year (Morrow, 2004), is another example pathogen-associated disease with strong MHC-associated resistance and susceptibility. The B2, B6, B14 and B21 haplotypes have been identified as conferring resistance to the disease in contrast to the B4, B12, B15 and B19 haplotypes which confer susceptibility (Plachy et al., 1992). Interestingly, the relative expression levels of the MHC class I molecules from different haplotypes were found to vary significantly. The dominantly expressed class I molecule from the B21 haplotype (resistant to Marek's disease) exhibited a 10-fold lower expression level compared to the class I molecules of the B4, B12, B15 and B19 haplotypes (susceptible to Marek's disease) (Kaufman et al., 1995) and led to the proposal that the basis of MHC-determined resistance or susceptibility to Marek's disease was due to the cell surface expression level polymorphism (Kaufman and Salomonsen, 1997; Kaufman et al., 1995).

It is not surprising that striking associations of resistance and susceptibility to infectious disease have been mapped to the chicken MHC, given the single dominantly expressed class I MHC molecule. In heterozygous chickens, resistance to a particular virus or other intracellular pathogen would depend on antigen presentation of pathogen-derived peptides by one or both of these two MHC class I molecules. A heterozygous chicken with only two class I MHC molecules offers much less protection to infectious pathogens than a heterozygous mammal expressing six class I MHC molecules. This may also explain why there are fewer reported associations with infectious diseases in the mammalian MHC due to the increased probability of binding an antigenic peptide that results from increased numbers of class I MHC molecules. However, the functional redundancy that results from multiple copies of the class I genes may also be making it difficult to tease out associations with individual genes. The greater number of MHC molecules in the mammalian MHC also increases the

probability of binding self-peptides and may explain associations with autoimmune disease (Blackwell et al., 2009; Hill, 1998; Tiwari and Terasaki, 1981).

One particular well-known human MHC association with infectious disease is that of progression or non-progression to acquired immune deficiency syndrome (AIDS) following human immunodeficiency virus (HIV) infection. Human class I MHC alleles such as B*35:01 lead to the rapid onset of AIDS following HIV infection, in contrast to HLA-B*57:01 and HLA-B*27:05, which confer long-term non-progression (Carrington et al., 1999; Goulder and Walker, 2012; The International HIV Controllers Study, 2010). A number of different explanations for these associations have been proposed, including the presentation of particularly effective peptide, number of peptides or peptide-independent recognition by natural killer (NK) cells (Carrington et al., 1999; Košmrlj et al., 2010; Martin et al., 2002).

As mentioned, the peptide motifs of the dominantly expressed class I MHC molecules of the B4, B12, B15 and B19 haplotypes were easily determined by sequence analysis of peptides found bound to them. The peptide-binding motifs of these molecules are highly fastidious, even more so than known human and mouse motifs (Kaufman et al., 1995; Wallny et al., 2006). A crystal structure of the dominantly expressed class I molecule of the B4 haplotype (BF2*0401) revealed positively-charged residues in a narrow binding groove as expected from the determined peptide motif (Zhang et al., 2012). In contrast, similar investigations into the dominantly expressed class I molecule of the B21 haplotype revealed that it was capable of binding peptides with many different amino acids found in every peptide position and as such it was not possible to determine a peptide binding motif. Our collaborators also identified two other haplotypes (B2 and B14) for which it was also not possible to trivially determine a peptide-binding motif. Interestingly, they found that the dominantly expressed class I molecules of these two haplotypes were also expressed at a lower level, similar to that of the B21 haplotype (Figure 3.3).

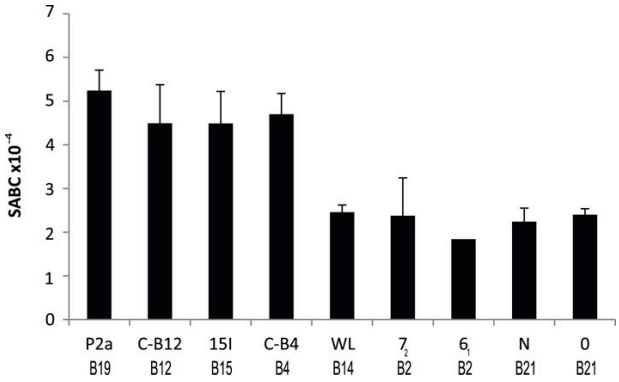

Figure 3.3 - Cell surface expression levels of the dominantly expressed class I molecules from different chicken haplotypes
Experiment carried out by our collaborators (Kaufman Lab, Department of Pathology, University of Cambridge). Quantitative flow cytometry experiment with spleen cells from inbred chicken lines stained with F21-2 (mAb against chicken class I MHC heavy chain).

Two crystal structures of the dominantly expressed class I molecule from the B21 haplotype bound to two peptides of different sequence revealed remodelling of the peptide-binding groove, primarily through the conformational flexibility of the MHC residue Arg9 (Figure 3.4). This allows peptides of very different sequence to bind and therefore makes the determination of a peptide motif for this molecule very difficult (Koch et al., 2007).

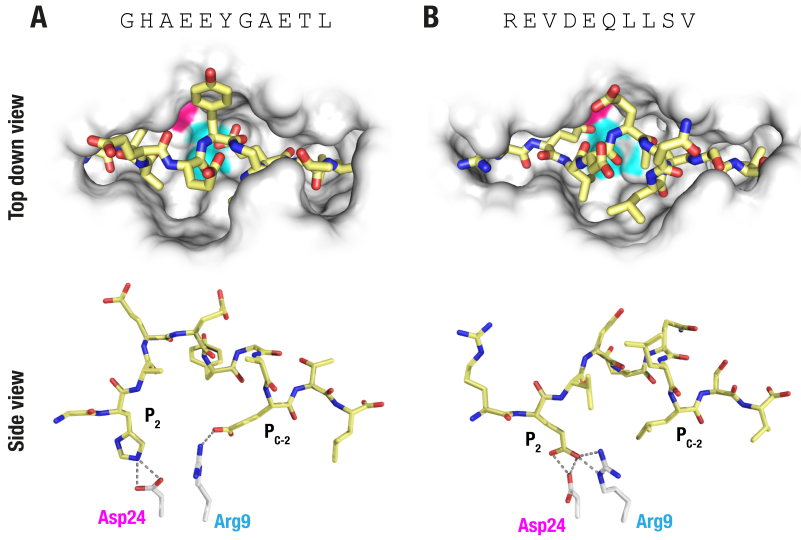

Figure 3.4 - Promiscuous peptide binding by BF2*2101
Peptides shown as cartoon representations in yellow. Top row of illustrations show MHC heavy chain in grey surface representation, with Arg9 coloured in cyan and Asp24 in magenta. Bottom row of illustrations show the Arg9 and Asp24 MHC residue side chains in stick representations in white. Conformational flexibility allows remodelling of peptide-binding groove.

3.2 Aims

The aim of this chapter is to solve the x-ray crystal structures of the dominantly expressed Chicken MHC class I molecules from the B2 (BF2*0201), B14 (BF2*1401) and B21 (BF2*2101) haplotypes in the presence of different peptides (Figure 3.5). In the case of B21 haplotype, although two structures have been solved, additional structures will help to further understand the peptide binding mechanism and may help to define a peptide-binding motif that could be used to predict peptide binding to the molecule. In contrast, there are no available structures for the dominantly expressed class I molecule of the B2 and B14 haplotype and any structural information will help to elucidate their mechanisms of peptide binding.

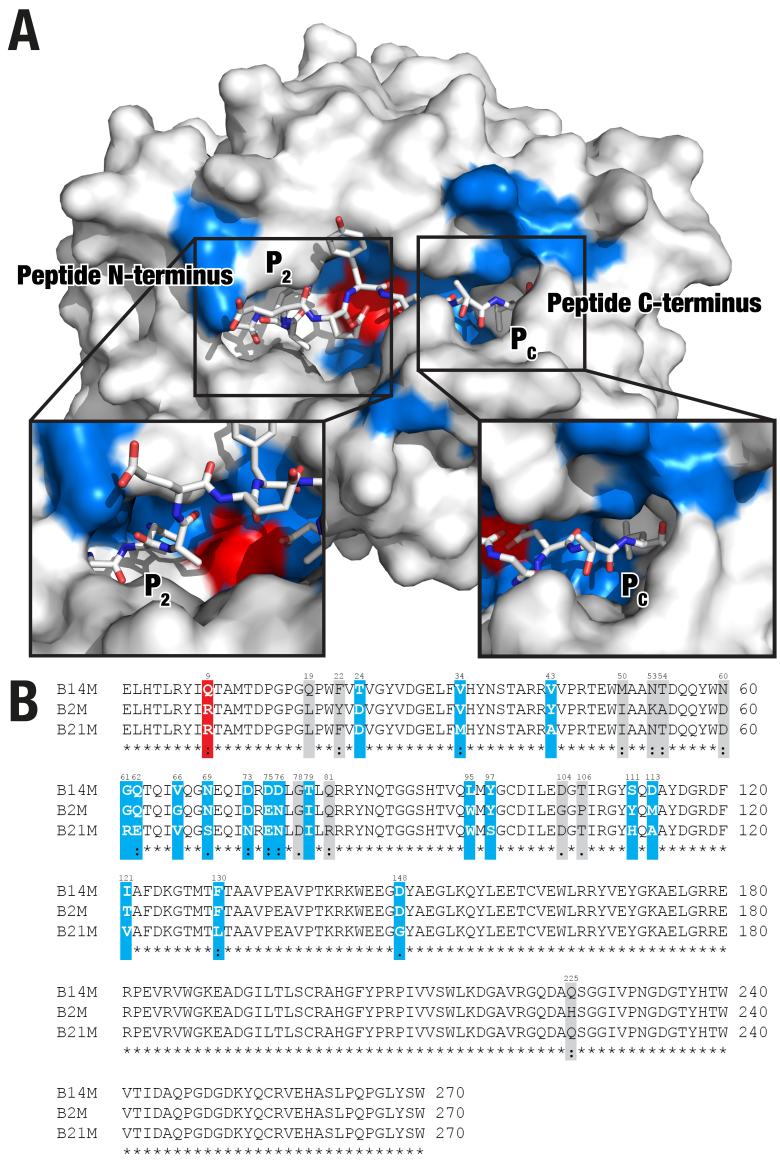

Figure 3.5 - B2, B14 & B21 sequence alignment

Panel A: MHC heavy chain displayed in surface representation with sequence differences near to or in the peptide binding groove highlighted in blue or red in the case of Arg9. Panel B: Sequence alignment. Grey = sequence difference outside peptide binding groove. Blue = sequence difference in peptide binding groove. Red = sequence difference at B21 Arg9 position.

3.3 Dominantly expressed class I MHC molecule of the B21 haplotype

Although the structure of the dominantly expressed class I MHC molecule of the B21 haplotype (BF2*2101) has already been solved in the presence of two different peptides (Koch et al., 2007), these structures did not allow rationalisation of many peptide sequences known to bind this molecule. The two initial structures of this molecule, in complex with either GHAEEYGAETL or REVDEQLLSV (Figure 3.4, 3.6A & 3.6B), identified three anchor residue positions, P_2, P_{C-2} and P_C (highlighted in blue in Figure 3.6). Inspection of these positions in the other peptides known to bind BF2*2101 revealed a number of peptide sequences that needed further investigation. For example, the peptides TAGQEDYDRL and TAGQSNYDRL (Figure 3.6D & E) both feature an alanine residue at the P_2 anchor position. Additionally, it is not known how this MHC molecule is able to accommodate two acidic residues in peptide positions P_2 and P_{C-2}, something that would not have been predicted given the acidic Asp24 and basic Arg9 MHC residues thought to form interactions with these peptide positions.

	P_1	P_2	P_3	P_4	P_5	P_6	P_{C-4}	P_{C-3}	P_{C-2}	P_{C-1}	P_C
A	G	H	A	E	E	Y	G	A	E	T	L
B	R	E	V	D	E	Q	-	L	L	S	V
C	T	N	P	E	S	K	-	V	F	Y	L
D	T	A	G	Q	E	D	-	Y	D	R	L
E	T	A	G	Q	S	N	-	Y	D	R	L
F	Y	E	L	D	E	K	-	F	D	R	L
G	G	H	A	E	E	Y	G	A	D	T	L
H	G	R	A	E	E	Y	G	A	D	T	L

N-terminus → C-terminus

Figure 3.6 - B21 peptide sequences
Peptides A & B are the peptides bound in the previously solved structures with PDBIDs 3BEV & 3BEW, respectively (Koch et al., 2007).

In order to further understand the mechanism of peptide binding to this molecule, our collaborator provided us with a number of additional peptide-MHC complexes for crystallisation trials (Figure 3.6). Peptides A to F feature sequences eluted from the MHC molecules of B21 cells, whereas peptides G and H are synthetic derivatives of peptides A and B. Crystallisation was successful for complexes C, D, E, G and H, with native datasets collected for each during my undergraduate part II research project. These structures were solved using molecular replacement with PHASER (McCoy et al., 2007) with a search model derived from the previously determined structure of this molecule (PDBID: 3BEV), minus bound peptide (Koch et al., 2007). Prior to deposition in the Protein Data Bank (Berman et al., 2000), all of these structures were re-refined during the course of my DPhil and as such are included in this chapter. In addition to these previous datasets, an additional crystal and subsequent dataset was obtained for the B21 class I MHC molecule in complex with the peptide YELDEKFDRL during my DPhil (Figure 3.6F).

3.3.1 Crystallisation

Collaborators in the Kaufman laboratory (Department of Pathology, University of Cambridge) provided us with recombinantly expressed and purified peptide-MHC complexes. A crystal grew within 4 weeks in a drop containing 60% protein and 40% mother liquor. Protein solution contained the peptide MHC complex at 11 mg/mL and the mother liquor was composed of 0.1M sodium acetate, pH 5.0, 1.5M ammonium sulfate.

3.3.2 Data collection

The new BF2*2101 peptide F complex crystal was cryoprotected in mother liquor plus 8M sodium formate before being flash frozen in liquid nitrogen. A native dataset was collected at Diamond Light Source's i04-1 beamline (Table 3.2).

Data Collection	B21 peptide F
Diffraction source	Diamond Light Source, i04-1
Detector	Pilatus 6M
Energy (eV)	13477
Wavelength (Å)	0.92
Temperature (K)	100
Detector resolution (Å)	2.5
Exposure per image (s)	0.4
Ω start (º)	30
Ω end (º)	210
Ω osc (º)	0.2
Number of images	900

Table 3.2 – B21 peptide F data collection

3.3.3 Data processing

Data were processed using XIA2 (Winter et al., 2013) running in 3daii mode, indexing an orthorhombic cell (a = 60.52 Å, b = 69.03 Å, c = 93.74 Å, α = 90°, β = 90°, γ = 90°), in the P 2_1 2_1 2_1 space group with data to a resolution of 2.39 Å (Table 3.3 & Figure 3.7).

	B21 peptide F
Resolution range (Å)	50.84-2.39 (2.45-2.39)
Space group	P 2_1 2_1 2_1
Unit cell dimensions	
a, b, c (Å)	60.52, 69.03, 93.74
α, β, γ (º)	90, 90, 90
Total reflections	102783 (7576)
Unique reflections	16087 (1171)
Multiplicity	6.4 (6.5)
Completeness (%)	99.9 (100.0)
Mean I/sigma(I)	11.9 (2.1)
R_{merge} (%)	15.4 (90.9)
R_{pim} (%)	6.6 (38.7)
CC(1/2)	0.994 (0.707)

Table 3.3 - B21 peptide F data processing statistics

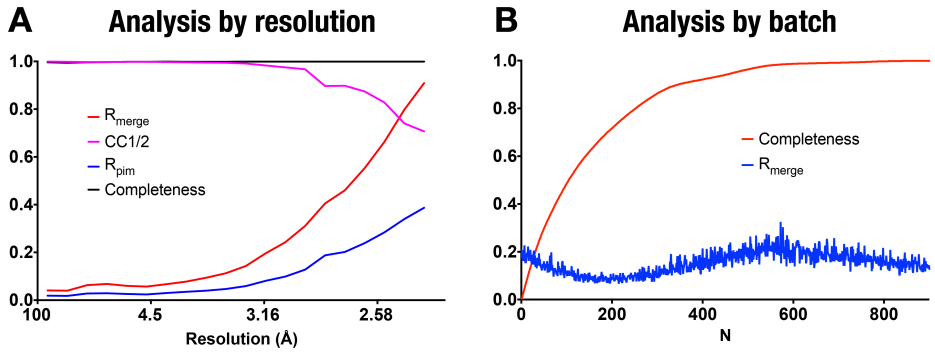

Figure 3.7 – B21 peptide F data processing analysis
Panel A: Analysis by resolution. Panel B: Analysis by batch.

3.3.4 Structure determination

PHENIX.XTRIAGE (Adams et al., 2010) suggested no evidence of twinning or translational non-crystallographic symmetry. Cell content analysis by MATTHEWS (Kantardjieff and Rupp, 2003) strongly suggested the presence of a single copy of the MHC-peptide complex (Table 3.4), assuming a total molecular weight of 47.8 kDa as calculated by PROTPARAM (Gasteiger et al., 2005).

Nmol / asym	Matthews coefficient	% solvent	P(2.39)	P(tot)
1	2.05	39.90	1.00	1.00

Table 3.4 - B21 peptide F cell content analysis

Analysis of the self-rotation function map calculated by MOLREP (Vagin and Teplyakov, 2010) revealed no significant non-crystallographic peaks. As there were a number of crystal structure already available for the dominantly expressed class I molecule from the B21 haplotype (BF2*2101), this structure was solved by molecular replacement using PHASER (McCoy et al., 2007). One copy of a search model derived from the atomic coordinates of the B21 class I MHC molecule (PDBID: 3BEV), excluding the peptide, (Koch et al., 2007) was placed in the asymmetric unit (Table 3.5).

Copy	RFZ	TFZ	LLG
3BEV model	16.9	28.0	1031

Table 3.5 - B21 peptide F molecular replacement

3.3.5 Model building and refinement

Rounds of model building and refinement were carried out using COOT (Emsley et al., 2010) and AUTOBUSTER (Bricogne et al.) and resulted in a final model with an R_{work} of 25.9% and an R_{free} of 26.9% (Table 3.6 in red). The results of the re-refinement of the data collected during my undergraduate part II project are also shown in Table 3.6 in black.

Refinement statistics	B21-F	B21-D	B21-E	B21-G	B21-H	B21-C
Resolution range (Å)	50.84-2.39 (2.56-2.39)	56.03-2.80 (3.13-2.80)	50.92-2.81 (2.88-2.81)	52.1-2.84 (3.17-2.84)	51.31-2.69 (2.76-2.69)	41.03-2.90 (3.24-2.90)
Number of reflections (work / test)	15275 / 811 (2732 / 139)	8432 / 929 (2340 / 265)	8546 / 421 (2198 / 103)	8906 / 449 (2268 / 105)	9903 / 474 (728 / 44)	8474 / 439 (2375 / 131)
Number of atoms (protein, ligand, water)	3079, 24, 42	3061, 0, 5	3057, 4, 31	3022, 9, 32	3023, 4, 4	3091, 0, 30
Protein residues	384	383	383	378	378	384
Mean B-factors (protein, ligands, waters)	30.23, 46.25, 24.38	30.30, N/A, 14.33	34.24, 26.32, 12.88	39.02, 40.37, 17.49	30.92, 34.69, 23.97	32.22, N/A, 8.61
R_{work} (%)	25.90 (32.80)	27.45 (32.27)	27.50 (33.12)	23.62 (25.64)	23.88 (33.30)	25.10 (28.83)
R_{free} (%)	26.90 (34.19)	29.60 (30.23)	28.20 (35.08)	27.26 (27.12)	26.70 (42.80)	27.27 (31.37)
R.M.S.D. from ideal values (bonds, Å)	0.007	0.007	0.007	0.007	0.007	0.007
R.M.S.D. from ideal values (angles, °)	0.85	0.85	0.86	0.83	1.13	0.83
Ramachandran favoured (%)	98.15	94.43	95.76	96.24	98.12	96.04
Ramachandran outliers (%)	0	0	0	0.27	0	0.26
Molprobity score	0.86 (100th percentile)	1.96 (99th percentile)	1.19 (100th percentile)	1.44 (100th percentile)	1.16 (100th percentile)	1.46 (100th percentile)

Table 3.6 - B21 refinement statistics
Refinement statistics for the structure crystallised in this chapter shown in red and those from re-refinements shown in black.

Examiners were aware that the B21-C, B21-D, B21-E, B21-G and B21-H datasets featured in my undergraduate masters dissertation. The data were included in this thesis because the models were re-refined during my DPhil using updated refinement software, which allowed a number of model errors to be corrected.

3.3.6 Structural analysis

The previously determined structures of the dominantly expressed class I MHC molecule of the B21 haplotype in the presence of two peptides of very different sequence (Figure 3.8A & B) revealed the importance of a flexible arginine residue at position 9 of the MHC heavy chain (Koch et al., 2007). This conformational flexibility of Arg9 was a result of a large central cavity in the peptide-binding groove due to a number of small amino acid residues such as Gly68, Ser69, Ser97 and Gly152. These structures also revealed the presence of three peptide anchor residues at P_2 (peptide position 2), P_{C-2} and P_C. The first structure (Figure 3.8A) featured an 11-mer peptide with a histidine residue at P_2 that interacted with the MHC heavy chain residue Asp24, a glutamic acid residue at P_{C-2} that formed interactions with Arg9 residue and a leucine residue at P_C that packed into a hydrophobic pocket. In contrast, the 10-mer peptide (Figure 3.8B) featured a hydrophobic leucine residue at P_{C-2}. This peptide bound the peptide-binding groove through interactions between the glutamic acid residue at P_2 and both Arg9 and Asp24. Arg9 adopted a different conformation to that observed in the first 11-mer structure causing a hydrophobic binding pocket to be exposed, into which the leucine residue at P_{C-2} could bind. This peptide also featured a hydrophobic residue at P_C that packed into the same hydrophobic pocket as the 11-mer.

The first of the new set of structures, in complex with a peptide of sequence TNPESKVFYL (B21 peptide C / Figure 3.8C), exhibited a similar binding mechanism to that of the REVDEQLLSV (Figure 3.8B) peptide. The asparagine residue at P_2 interacts with both Arg9 and Asp24, causing the Arg9 residue to occupy a conformation that resulted in the exposure of a hydrophobic pocket in which the phenylalanine peptide residue at P_{C-2} packed into.

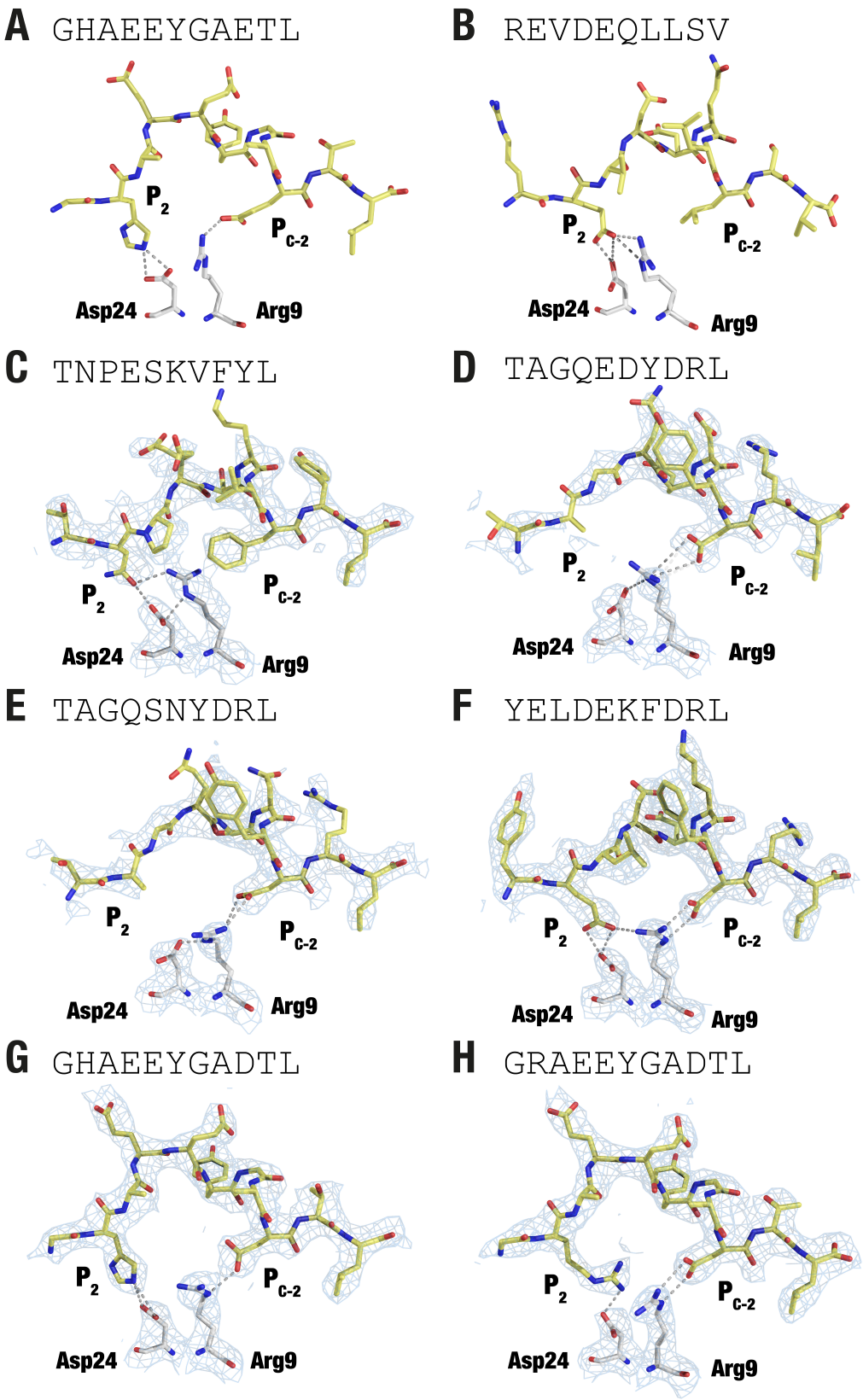

Figure 3.8 - B21 peptide binding

Side views with peptide and MHC residues Arg9 & Asp24 displayed as sticks. Interactions shown as grey dotted lines. Peptide residues are coloured yellow and MHC residues, white. Panels A & B: Previously determined structures (PDBID: 3BEV & 3BEW).

The next two structures with bound peptides TAGQEDYDRL (Figure 3.8D) and TAGQSNYDRL (Figure 3.8E) showed that the alanine residue at P_2 did not form any interactions with the peptide-binding groove, revealing for the first time, that an interaction of P_2 with the peptide-binding groove is not always necessary for peptide binding to BF2*2101. Instead, peptide binding was mediated by the aspartic acid residue at P_{C-2} that formed interactions with Arg9, which in turn interacted with Asp24. The hydrophobic leucine residue at P_C packed into the hydrophobic pocket as seen previously. The C-terminal portions of both of these peptides exhibited less-well defined electron density suggesting possible conformational flexibility. The structure of the YELDEKFDRL (Figure 3.8F) peptide revealed how this MHC class I molecule was able to bind peptides with acidic residues present at both P_2 and P_{C-2}. Asp24 interacted with the glutamic acid residue at P_2 and Arg9 interacted with both the glutamic acid residue at P_2 and the aspartic acid residue at P_{C-2}. The two structures in complex the GHAEEYGADTL (Figure 3.8G) & GRAEEYGADTL peptides (Figure 3.8H) revealed a similar mechanism of binding to that observed in the first 11-mer structure (Figure 3.8A), with either a histidine or arginine residue at P_2 forming interactions with Asp24 and an aspartic acid residue at P_{C-2} interacting with Arg9.

These structures have revealed the enormous plasticity of this class I MHC molecule in binding to a wide variety of peptides of very different sequence. Unfortunately, it was not possible to use this new structural data to rationalise our collaborators peptide sequencing data and determine a peptide motif for this molecule. Peptide positions outside of the anchor positions seemed to influence the residues allowed at certain anchor positions in a very complex manner that was not possible to predict based on the currently available structures.

3.3.7 Thermostability of BF2*2101-peptide complexes

In order to assess the stability of peptide binding, the thermal stabilities of the class I MHC-peptide complexes were assessed using the ThermoFluor assay (Matulis et al., 2005). 20 µL of each protein solution at 1 mg/mL was added to 5 µL of SPYRO orange dye (Life Technologies) (1000-fold dilution) in a 96-well plate. The temperature was increased at a rate of 1.5°C per minute and fluorescence monitored in a Stratagene MX3005P real time PCR machine. Fluorescence measurements were normalised and then fitted to a Boltzmann sigmoidal curve using GraphPad Prism (Figure 3.9) to calculate melting temperatures (Figure 3.10 & Table 3.7).

Analysis of the melting temperatures suggested that the most stable complexes were those that featured a hydrophobic residue at either P_2 or P_{C-2}, with complexes that featured a hydrophobic residue at P_{C-2} slightly more stable than those at P_C (Figure 3.11). In contrast, complexes that featured charged residues at P_2 or P_{C-2} were significantly less thermostable. The presence of two acidic residues at positions P_2 and P_{C-2} versus one acidic and one basic residue did not appear to influence the thermostability of the complex. Instead, the least thermostable of these complexes (B21 peptide H) featured a bulky arginine residue at P_2, whereas the most stable complexes featured slightly smaller amino acids.

Unfortunately, this data, in combination with the structural analyses and our collaborators sequencing data, gives us insufficient information to make specific predictions about whether a given peptide sequence will bind the BF2*2101 molecule.

	Melting temperature (ºC)	Standard Error	N
B21 peptide A	53.2	0.4	7
B21 peptide B	61.2	0.6	4
B21 peptide C	63.9	0.5	8
B21 peptide D	58.9	0.6	8
B21 peptide E	58.2	0.4	8
B21 peptide F	54.4	0.4	4
B21 peptide G	55.7	0.5	7
B21 peptide H	44.3	0.2	6

Table 3.7 - B21 melting temperatures

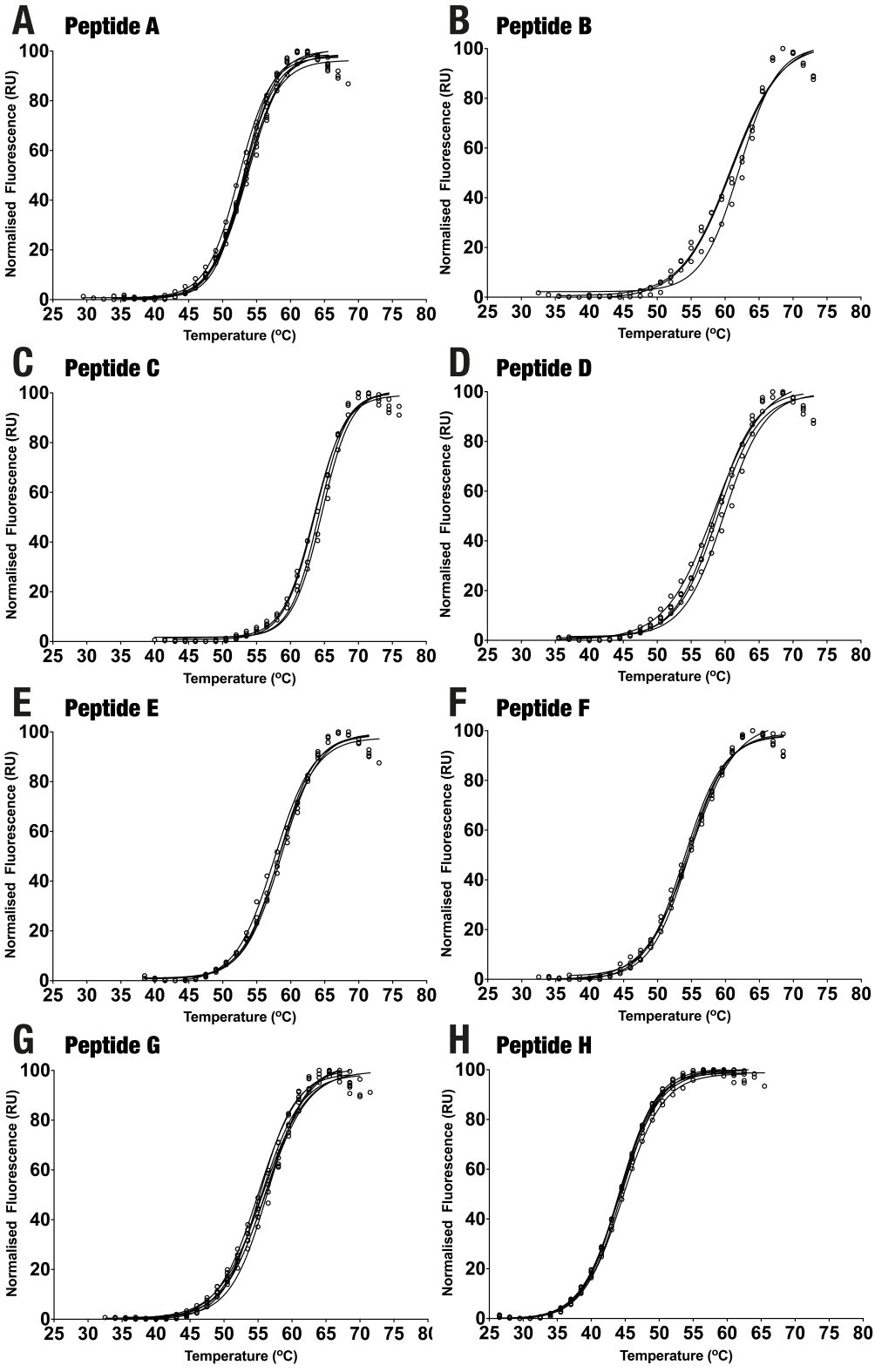

Figure 3.9 - Thermal stability data with Boltzmann sigmoidal fits
Normalised fluorescence traces fitted to Boltzmann sigmoidal curves. Determined melting temperatures shown in Table 6 and Figure 11.

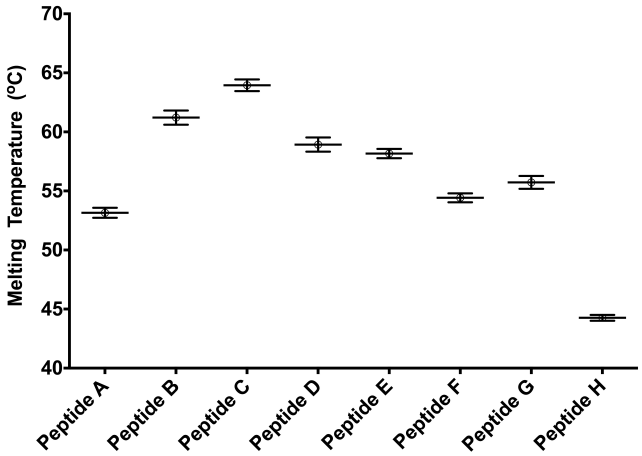

Figure 3.10 - B21 peptide complex melting temperatures

Melting temperatures of B21 class I MHC complexes as determined by the ThermoFluor assay. Related to Figure 10 & Table 6.

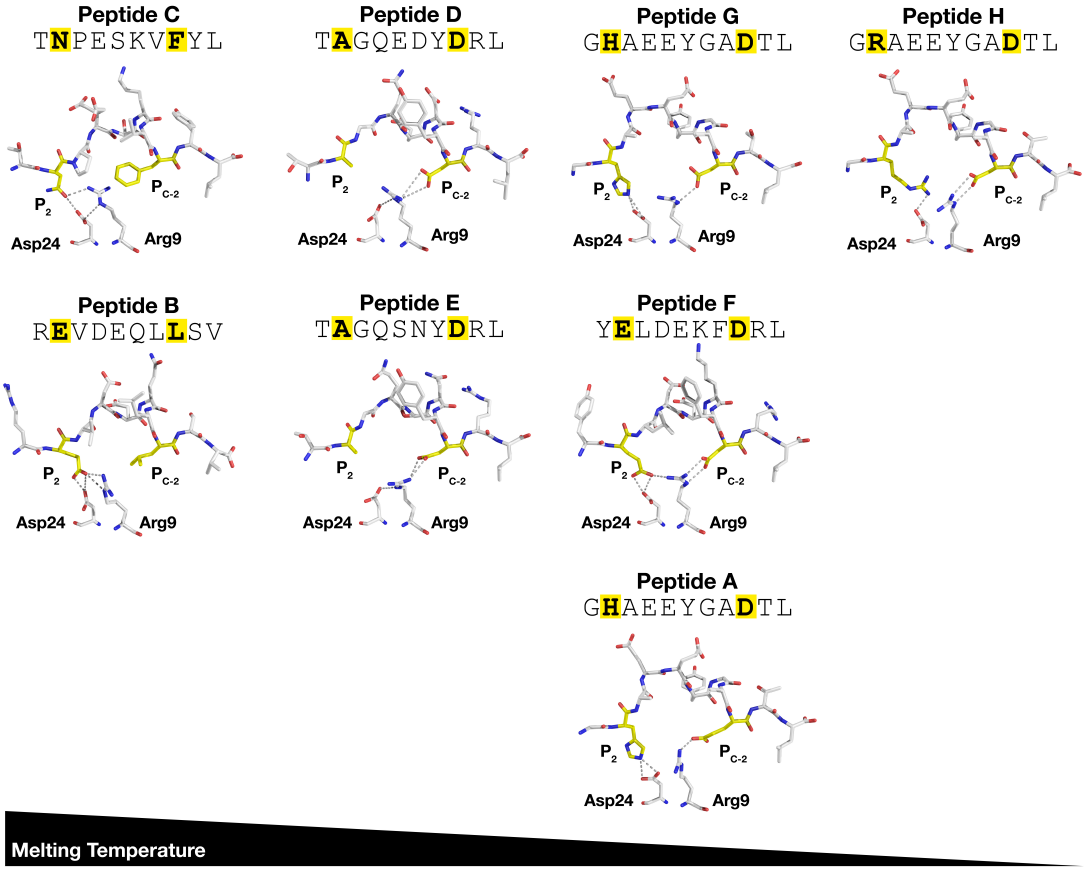

Figure 3.11 - Order of MHC peptide complex thermostability

3.4 Dominantly expressed class I MHC molecule of the B2 haplotype

Our collaborators provided us with a number of class I MHC peptide complexes from the B2 haplotype (BF2*0201) (Figure 3.12). Their analysis of peptide sequences isolated from cell lines of this haplotype revealed two groups. One group of peptides showed strong preference for arginine residues at P_2 and tyrosine residues at P_C. These peptides were likely bound to the poorly expressed, minor BF1*0201 gene because this class I molecule was similar to the dominantly expressed class I molecule from the B15 haplotype which binds similar peptides. The second group of peptides featured broadly hydrophobic residues at P_2 and P_C (Figure 3.12). This class I MHC molecule features the Arg9 and Asp24 residues that are responsible for the promiscuous peptide-binding mechanism of the dominantly expressed class I MHC molecule of the B21 haplotype, suggesting that this molecule may employ a similar mechanism capable of achieving promiscuous peptide binding.

	P_1	P_2	P_3	P_4	P_5	P_{C-3}	P_{C-2}	P_{C-1}	P_C
A	G	I	L	E		S	F	T	V
B	T	A	W	L		G	F	L	V
C	V	I	F	P		A	K	S	L
D	Y	P	Y	L	G	P	N	T	L
E	Q	T	F	L	Q	P	L	T	V
F	D	L	T	P	K	V	Q	V	Y
G	R	R	G	G		V	K	R	I

N-terminus → C-terminus

Figure 3.12 - B2 peptide sequences

3.4.1 Crystallisation

Collaborators in the Kaufman laboratory (Department of Pathology, University of Cambridge) provided us with recombinantly expressed and purified BF2*0201 MHC class I protein-peptide complexes with each peptide sequence one of those isolated from cells of the B2 haplotype (Figure 3.12). Crystals were obtained for two of the provided complexes. Both crystals grew within 4 weeks in drops containing 60% protein and 40% mother liquor. Both grew in drops with a protein concentration of ~10 mg/mL and mother liquor that was composed of 0.1M MMT buffer, pH 7.0, 25% PEG 1500.

3.4.2 Data collection

Crystals were cryoprotected in mother liquor plus 15% (v/v) ethylene glycol before being flash frozen in liquid nitrogen. Native datasets were collected at Diamond Light Source's i04 beamline (Table 3.8).

Data Collection	B2 peptide D	B2 peptide C
Diffraction source	Diamond Light Source, i04	Diamond Light Source, i04
Detector	ADSC	ADSC
Energy (eV)	12658	12658
Wavelength (Å)	0.9795	0.9795
Temperature (K)	100	100
Detector resolution (Å)	2.7	3.2
Exposure per image (s)	0.4	1
Ω start (º)	184	0
Ω end (º)	304	115
Ω osc (º)	0.2	0.5
Number of images	600	230

Table 3.8 – B2 peptide D and C data collection

3.4.3 Data processing

Both datasets were processed using XIA2 (Winter et al., 2013) running in 3da mode. Although both crystals grew under the same crystallisation conditions, two different crystal forms were obtained. The B2 peptide D dataset initially indexed in a hexagonal cell (a = 174.0 Å, b = 174.0 Å, c = 87.59 Å, α = 90°, β = 90°, γ = 120°) with data to 3.32 Å in the P 6 2 2 space group. Analysis in PHENIX.XTRIAGE (Adams et al., 2010) suggested twinning with the L test for acentric data strongly suggesting perfect twinning (Figure 3.13). Due to the absence of twinning operators in the P 6 2 2 space group, the data were reprocessed with XIA2 (Winter et al., 2013) in a hexagonal cell (a = 173.9 Å, b = 173.9 Å, c = 87.51 Å, α = 90°, β = 90°, γ = 120°) with data to 3.30 Å in the P 6_5 space group (Table 3.9 & Figure 3.14A & B). The B2 peptide C dataset indexed in an orthorhombic cell (a = 88.08 Å, b = 92.54 Å, c = 223.6 Å, α = 90°, β = 90°, γ = 90°) with data to 3.13 Å in the P 2_1 2_1 2_1 space group (Table 3.9 & Figure 3.14C & D).

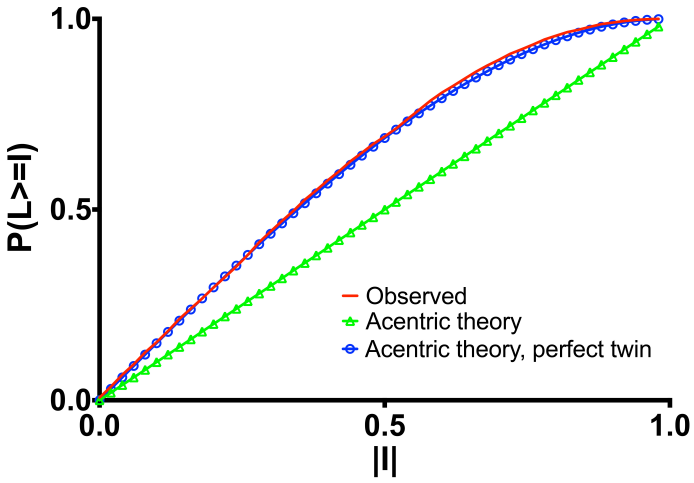

Figure 3.13 - L test for acentric data (B2 peptide D)

	B2 peptide D	B2 peptide C
Resolution range (Å)	75.66-3.30 (3.56-3.30)	81.95-3.13 (3.21-3.13)
Space group	P 6_5	P 2_1 2_1 2_1
Unit cell dimensions		
a, b, c (Å)	173.9, 173.9, 87.51	88.08, 92.54, 223.6
α, β, γ (°)	90, 90, 120	90, 90, 90
Total reflections	168124 (34983)	153040 (10553)
Unique reflections	22882 (4672)	32916 (2374)
Multiplicity	7.3 (7.5)	4.6 (4.4)
Completeness (%)	100 (99.9)	99.6 (98.8)
Mean I/sigma(I)	11.1 (3.8)	8.8 (2.1)
R_{merge} (%)	16.5 (55.6)	17.7 (69.4)
R_{pim} (%)	6.5 (21.8)	9.0 (36.1)
CC(1/2)	0.993 (0.889)	0.987 (0.670)

Table 3.9 - B2 peptides D and C data processing statistics

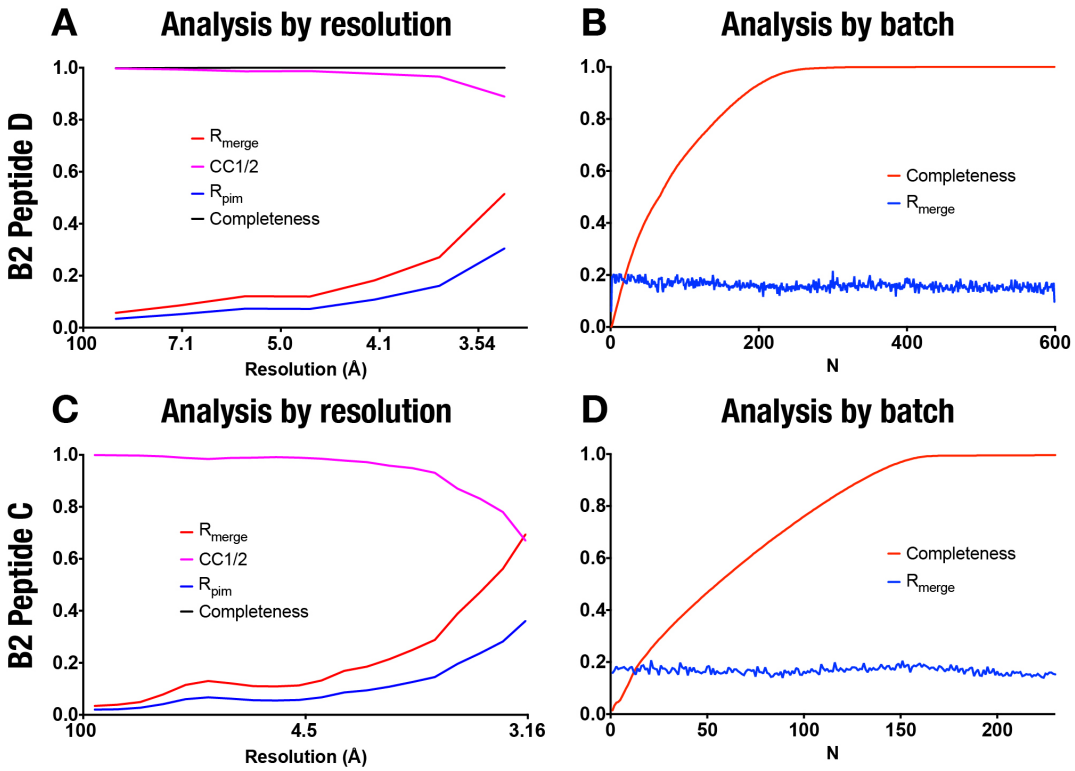

Figure 3.14 - B2 peptides D and C data processing analysis
Panel A & C: Analysis by resolution. Panel B & D: Analysis by batch.

3.4.4 B2 peptide D structure determination

PHENIX.XTRIAGE (Adams et al., 2010) suggested no evidence of translational non-crystallographic symmetry but indicated the crystal was merohedrally twinned. Cell content analysis by MATTHEWS (Kantardjieff and Rupp, 2003) suggested between three or four copies of the MHC-peptide complex per asymmetric unit (Table 3.10).

Nmol / asym	Matthews coefficient	% solvent	P(3.30)	P(tot)
1	7.98	84.56	0.00	0.00
2	3.99	69.18	0.06	0.04
3	2.66	53.76	0.67	0.59
4	1.99	38.35	0.26	0.37
5	1.60	22.94	0.00	0.00
6	1.33	7.53	0.00	0.00

Table 3.10 – B2 peptide D cell content analysis

Inspection of the self-rotation function calculated by MOLREP (Vagin and Teplyakov, 2010) revealed only peaks consistent with 2-fold axes generated by the crystal twinning and explains why the data were initially processed in the P 6 2 2 space group (Figure 3.15).

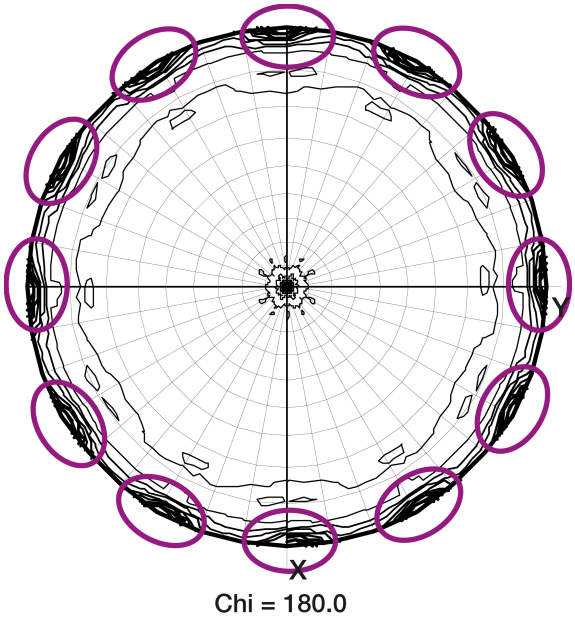

Figure 3.15 – B2 peptide D self-rotation function
Self-rotation function peaks calculated by MOLREP (Vagin and Teplyakov, 2010). Purple ovals indicate 2-fold axes generated by crystal twinning.

A search model for molecular replacement was generated using the previously deposited structure of the dominantly expressed chicken class I MHC molecule from the B21 haplotype (PDBID: 3BEV) (Koch et al., 2007), using CHAINSAW (Stein, 2008) to remove any atoms of non-conserved residues that were not common to the target sequence. Molecular replacement using this model was then carried out using PHASER (McCoy et al., 2007), which placed only two copies of the search model (Table 3.11). Inspection of the molecular replacement solution showed that the two copies of the placed MHC-peptide complex formed crystal contacts that would allow crystal packing.

Copy	RFZ	TFZ	LLG
3BEV chainsaw model	10.4	17.4	264
3BEV chainsaw model	7.2	27.5	882

Table 3.11 - B2 peptide D molecular replacement

3.4.5 B2 peptide C structure determination

PHENIX.XTRIAGE (Adams et al., 2010) suggested no evidence of twinning or translational non-crystallography symmetry. Cell content analysis by MATTHEWS (Kantardjieff and Rupp, 2003) suggested between three and five copies of the MHC-peptide complex per asymmetric unit (Table 3.12). Inspection of the self-rotation function calculated using MOLREP (Vagin and Teplyakov, 2010) indicated the presence of a 2-fold non-crystallographic symmetry axis (Figure 3.16 & Table 3.13).

Nmol / asym	Matthews coefficient	% solvent	P(3.13)	P(tot)
1	9.52	87.09	0.00	0.00
2	4.76	74.18	0.01	0.01
3	3.17	61.27	0.22	0.17
4	2.38	48.35	0.67	0.68
5	1.90	35.44	0.09	0.14
6	1.59	22.53	0.00	0.00
7	1.36	9.62	0.00	0.00

Table 3.12 - B2 peptide C cell content analysis

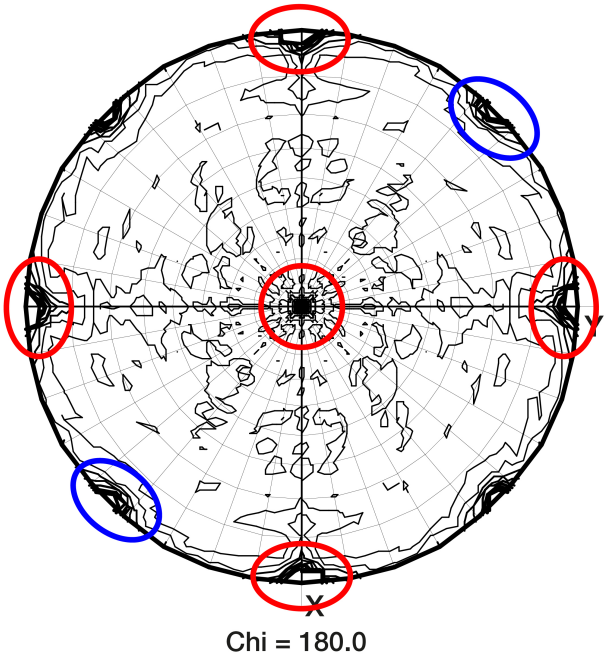

Figure 3.16 - B2 peptide C self-rotation function
Self-rotation function peaks calculated by MOLREP (Vagin and Teplyakov, 2010). Crystallographic and non-crystallographic peaks are highlighted with red and blue circles/ovals, respectively.

θ	Φ	χ	P(i)/P(0)
90.00	-44.41	180.00	0.37

Table 3.13 - B2 peptide C self-rotation function peaks
Self-rotation function peaks above 30% of the origin. Related to Figure 3.17.

Molecular replacement was carried out using PHASER (McCoy et al., 2007) and a search model derived from the heavy and light chains of the dominantly expressed class I MHC molecule of the B21 haplotype (PDBID: 3BEV) (Koch et al., 2007) which was processed using CHAINSAW (Stein, 2008) to remove any atoms from non-conserved residues that were not common to both sequences. Four copies of this model were placed (Table 3.14).

Copy	RFZ	TFZ	LLG
3BEV chainsaw model	6.7	9.0	89
3BEV chainsaw model	5.9	20.4	337
3BEV chainsaw model	4.4	21.3	573
3BEV chainsaw model	3.3	16.6	496

Table 3.14 - B2 peptide C molecular replacement

3.4.6 Model building and refinement

Rounds of model building and refinement were carried out using COOT (Emsley et al., 2010) and REFMAC (Murshudov et al., 1997) or AUTOBUSTER (Bricogne et al.). The B2 peptide D structure was refined with REFMAC (Murshudov et al., 1997) with intensity based twin refinement, automatically generated non-crystallographic symmetry restraints and a fixed X-ray weighting term to retain correct model geometry. The final B2 peptide D model was refined to an R_{work} of 23.68% and an R_{free} of 26.17%, with twin fractions of 0.6145 and 0.3855. The B2 peptide C model was refined with AUTOBUSTER (Bricogne et al.) with automatically generated non-crystallographic symmetry restraints and a fixed X-ray weighting term to preserve model geometry. The final B2 peptide C model was refined to an R_{work} of 29.25% and R_{free} of 29.86% (Table 3.15).

Refinement statistics	B2 peptide D	B2 peptide C
Resolution range (Å)	75.66-3.30 (3.39-3.30)	81.95-3.13 (3.23-3.13)
Number of reflections (work / test)	20557 / 1150 (1582 / 92)	31095 / 1665 (2720 / 151)
Number of atoms (protein, ligand, water)	6052, 0, 8	12102, 0, 9
Protein residues	754	1509
Mean B-factors (protein, ligands, waters)	77.86, N/A, 34.87	47.03, N/A, 3.812
R_{work} (%)	23.68 (44.0)	29.25 (36.08)
R_{free} (%)	26.17 (45.0)	29.86 (36.13)
R.M.S.D. from ideal values (bonds, Å)	0.004	0.007
R.M.S.D. from ideal values (angles, °)	0.759	0.82
Ramachandran favoured (%)	95.69	96.3
Ramachandran outliers (%)	0	0
Molprobity score	1.06 (100th percentile)	1.26 (100th percentile)

Table 3.15 - B2 peptides D & C refinement statistics

3.4.7 Structural analysis

As expected, the two new structures of the dominantly expressed class I MHC molecule from the B2 haplotype exhibit an overall structure that is essentially identical to that of the previously determined structure of the same molecule from the B21 haplotypes (Figure 3.17).

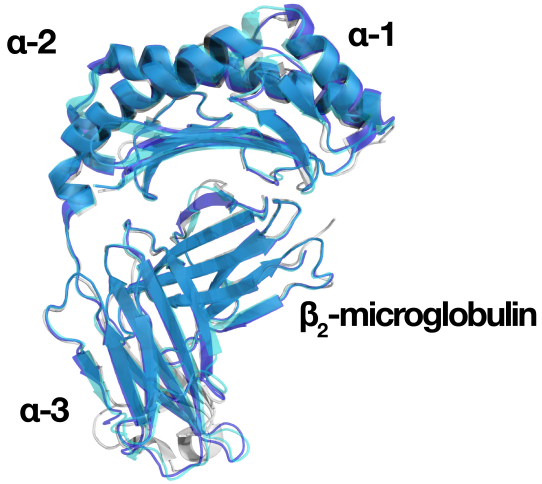

Figure 3.17 – Overall structure comparison of BF2*0201 to BF*2101
Structures of the dominantly expressed class I MHC molecules from the B2 and B21 haplotypes shown in cartoon representation. B21 in grey, peptide D (B2) in blue and peptide C (B2) in cyan.

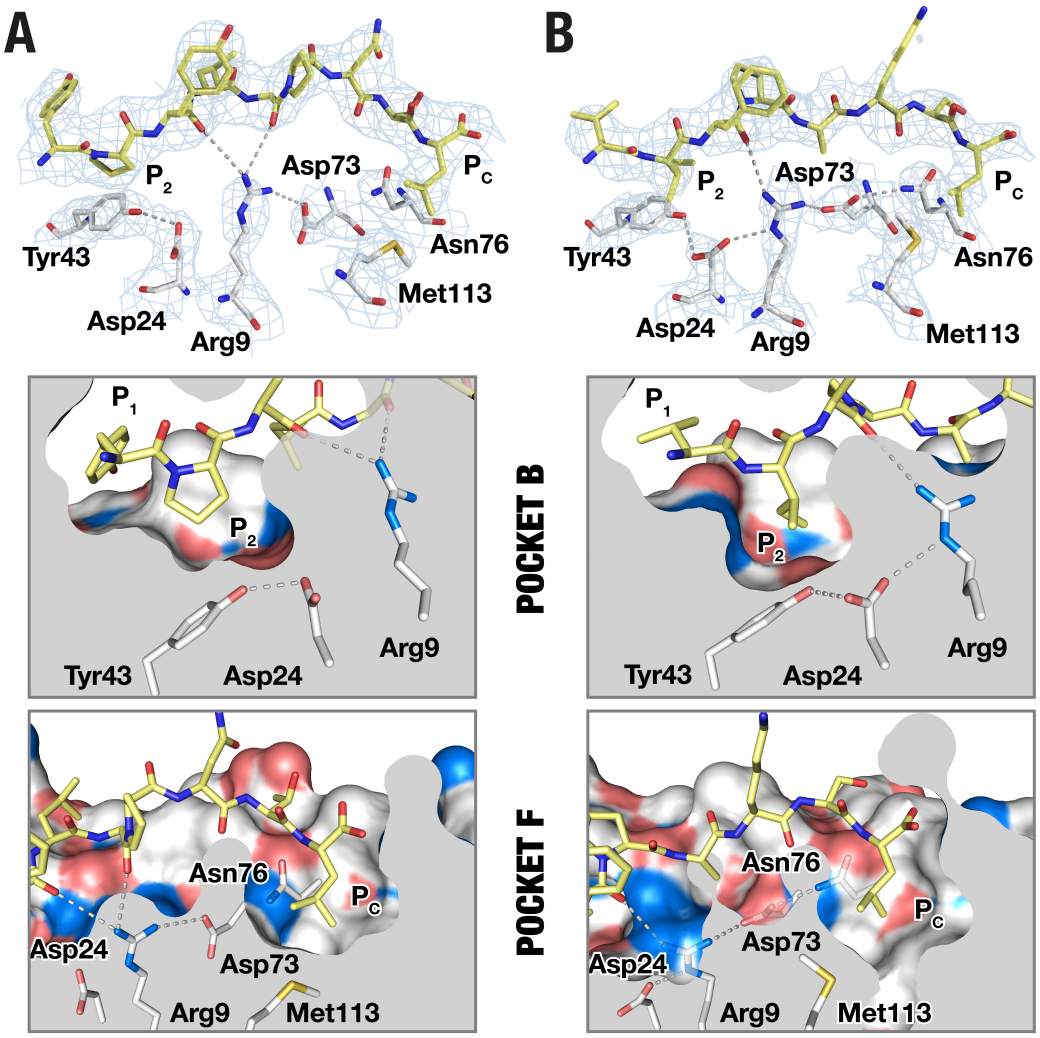

Figure 3.18 - B2 peptide binding
Peptide shown as yellow sticks. Important MHC residues shown as white sticks with surface representations illustrating binding groove pockets. 2Fo-Fc electron density maps contoured at 1σ are shown in blue in the top two panels.

The two new structures of the dominantly expressed class I MHC molecule of the B2 haplotype in the presence of different peptides reveals a completely new mechanism of promiscuous peptide-binding, despite conservation of Arg9 and Asp24, residues that were instrumental in the promiscuous binding mechanism of the B21 haplotype. The first B2 structure in complex with the VPYLGPNTL peptide (Figure 3.18A) revealed two shallow hydrophobic pockets into which the proline residue at P_2 and leucine residue at P_C packed. Asp24 interacts with Tyr43 and Arg9 with peptide main chain carbonyl oxygen atoms. The second structure in complex with the VIFPAKSL peptide (Figure 3.18B) revealed a similar binding mechanism with the

isoleucine residue at P_2 and leucine residue at P_C packing into the same shallow hydrophobic pocket as the first structure.

3.5 Dominantly expressed class I MHC molecule of the B14 haplotype

Our collaborators provided us with a number of class I MHC peptide complexes from the B14 haplotype (BF2*1401) (Figure 3.19). Inspection of these peptide sequences, each isolated from cells of the B14 haplotype, suggested a strong preference for lysine / arginine at the C-terminal peptide position, but apart from at this position, there were no other obvious anchor residues.

	P_1	P_2	P_3	P_4	P_5	P_{C-3}	P_{C-2}	P_{C-1}	P_C
A	L	Y	K	A		I	R	R	K
B	S	M	Y	Y		I	K	R	R
C	Y	F	K	P		G	I	R	R
D	Y	A	K	K		N	E	P	K
E	S	V	H	R		I	I	G	R
F	S	W	F	R	K	P	M	T	R
G	A	I	K	K	N	N	P	R	K

N-terminus → C-terminus

Figure 3.19 - B14 peptide sequences
Peptide-MHC complexes provided by our collaborators. Solid blue bar indicates likely anchor residue.

In order to assess the mechanism of peptide-binding to this MHC class I molecule, attempts were made to crystallise a number of these complexes and solve their X-ray structures.

3.5.1 Crystallisation

A single crystal was obtained for one of the provided complexes (Figure 3.19F). This crystal grew within 4 weeks in a drop containing 60% protein and 40% mother

liquor. The protein solution was at a concentration of ~10 mg/mL and the mother liquor composed of 0.1M MIB buffer, pH 5.0, 25% PEG 1500.

3.5.2 Data collection

The crystal was cryoprotected in mother liquor plus 15% (v/v) ethylene glycol before being flash frozen in liquid nitrogen. A native dataset was collected at the European Synchrotron Radiation Facility's ID29 beamline (Table 3.16).

Data Collection	B14 peptide F
Diffraction source	European Synchrotron Radiation Facility, ID29
Detector	Pilatus
Energy (eV)	12660
Wavelength (Å)	0.9793
Temperature (K)	100
Detector resolution (Å)	2.5
Exposure per image (s)	0.1
Ω start (º)	30
Ω end (º)	210
Ω osc (º)	0.1
Number of images	1800

Table 3.16 - B14 peptide F data collection

3.5.3 Data processing

Data processing was carried out using XIA2 (Winter et al., 2013) running in 3daii mode, indexing in an orthorhombic cell (a = 62.20 Å, b = 90.58 Å, c = 144.8 Å, α = 90°, β = 90°, γ = 90°) with data to 2.58 Å resolution in the P 2_1 2_1 2_1 space group (Table 3.17 & Figure 3.20).

	B14 peptide F
Resolution range (Å)	72.38-2.58 (2.65-2.58)
Space group	P 2_1 2_1 2_1
Unit cell dimensions	
a, b, c (Å)	62.20, 90.58, 144.8
α, β, γ (°)	90, 90, 90
Total reflections	145517 (7809)
Unique reflections	26279 (1865)
Multiplicity	5.5 (4.2)
Completeness (%)	99.9 (98.4)
Mean I/sigma(I)	8.8 (2.3)
R_{merge} (%)	13.5 (59.3)
R_{pim} (%)	6.2 (31.2)
CC(1/2)	0.922 (0.817)

Table 3.17 - B14 peptide F data processing statistics

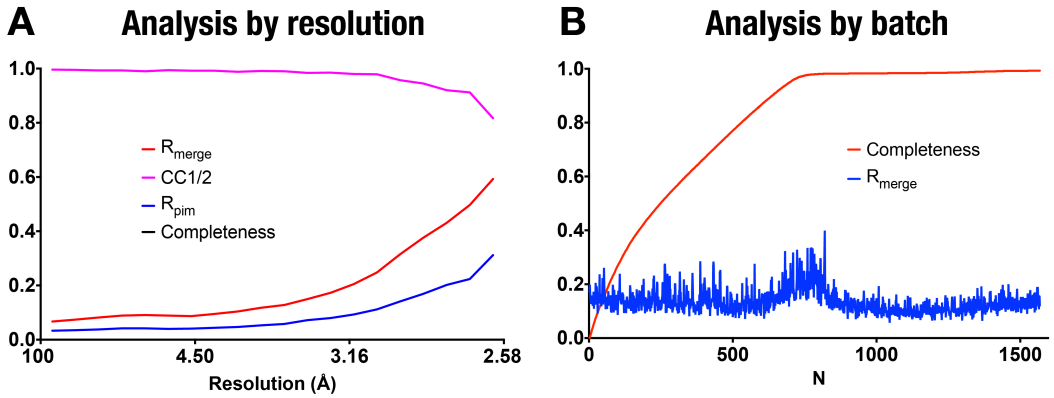

Figure 3.20 - B14 peptide F data processing analysis
Panel A: Analysis by resolution. Panel B: Analysis by batch.

3.5.4 B14 peptide F structure determination

PHENIX.XTRIAGE (Adams et al., 2010) suggested no evidence of twinning, but a peak in the Patterson map with an intensity 45.6% of the origin suggested the presence of translational non-crystallographic symmetry. Cell content analysis by MATTHEWS (Kantardjieff and Rupp, 2003) suggested the presence of two copies of the complex (Table 3.18).

Nmol / asym	Matthews coefficient	% solvent	P(2.58)	P(tot)
1	4.26	71.14	0.02	0.03
2	2.13	45.28	0.97	0.97
3	1.42	13.43	0.00	0.00

Table 3.18 - B14 peptide F cell content analysis

Analysis of the self-rotation function map calculated by MOLREP (Vagin and Teplyakov, 2010) revealed no non-crystallographic peaks above 10% of the origin intensity. Molecular replacement was carried out using PHASER (McCoy et al., 2007) and a search model derived from the heavy and light chains of the dominantly expressed class I MHC molecule of the B21 haplotype (PDBID: 3BEV) (Koch et al., 2007) which was processed using CHAINSAW (Stein, 2008) to remove any atoms from non-conserved residues that were not common to both sequences. A molecular replacement solution was found that placed two copies of this search model and inspection revealed crystal contacts that would allow the crystal to pack (Table 3.19).

Copy	RFZ	TFZ	LLG
3BEV model	8.5	7.9	40
3BEV model	3.7	34.3	2268

Table 3.19 - B14 peptide F molecular replacement

3.5.5 Model building and refinement

Rounds of model building and refinement were carried out using COOT (Emsley et al., 2010) and AUTOBUSTER (Bricogne et al.), resulting in a final model with an R_{work} of 28.57% and an R_{free} of 29.09% (Table 3.20).

Refinement statistics	P479
Resolution range (Å)	72.38-2.58 (2.69-2.58)
Number of reflections (work / test)	24895 / 1333 (2741 / 155)
Number of atoms (protein, ligand, water)	6042, 0, 77
Protein residues	755
Mean B-factors (protein, ligands, waters)	40.04, N/A, 29.50
R_{work} (%)	28.57 (33.43)
R_{free} (%)	29.09 (34.78)
R.M.S.D. from ideal values (bonds, Å)	0.007
R.M.S.D. from ideal values (angles, º)	0.84
Ramachandran favoured (%)	98.25
Ramachandran outliers (%)	0
Molprobity score	0.91 (100th percentile)

Table 3.20 - B14 peptide F refinement statistics

As expected, the dominantly expressed class I MHC molecule from the B14 haplotype exhibits an overall structure that is almost identical to that of the same molecules from the B2 and B21 haplotypes (Figure 3.21).

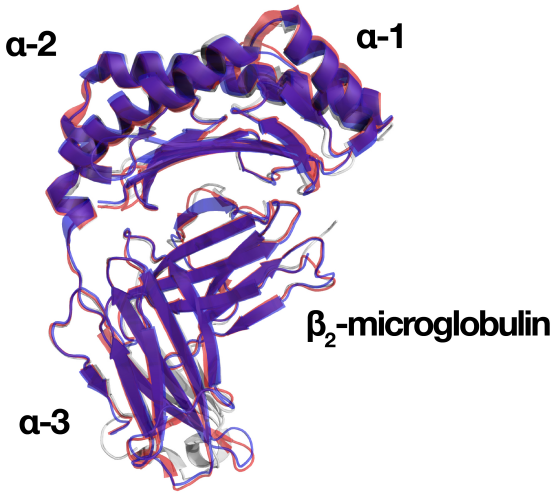

Figure 3.21 - Comparison of BF2*0201, BF2*1401 & BF2*2101 structures
Structures of the dominantly expressed class I MHC molecules from the B2, B14 and B21 haplotypes shown in cartoon representation. B2 in blue, B14 in red and B21 in grey.

The structure of the dominantly expressed class I MHC molecule from the B14 haplotype (BF2*1401) in the presence of the SWFRKPMTR (Figure 3.22) peptide revealed insight into the peptide binding mechanism of this molecule. The peptide tryptophan residue at P_2 packed into a deep hydrophobic pocket forming hydrophobic interactions along with a hydrogen bond with the MHC heavy chain residue Thr24. The arginine residue at P_C also packed into a deep pocket and interacted with the MHC heavy chain residues Asp73, Asp76 and Asp113. These pockets were much deeper than those in the equivalent molecule from the B2 haplotype and explain the larger amino acids seen at these peptide positions.

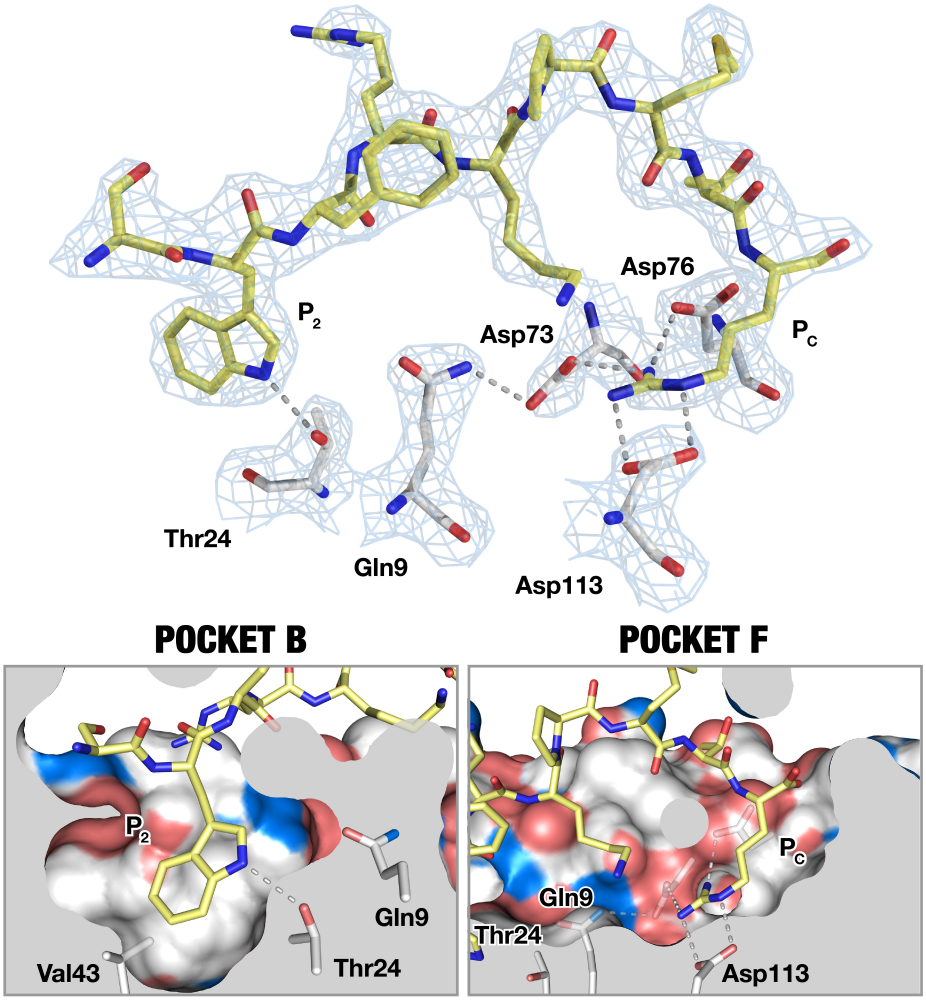

Figure 3.22 - B14 peptide binding
Peptide shown as yellow sticks. Important MHC residues shown as white sticks with surface representations illustrating binding groove pockets. 2Fo-Fc electron density map contoured at 1σ shown in blue.

3.6 Discussion

The structures presented in this chapter offer insight into the peptide binding mechanisms of three dominantly expressed chicken class I MHC molecules (BF2*0201, BF2*1401, BF2*2101) which our collaborators found to bind a much wider variety of peptides than the equivalent molecules of other fastidious chicken haplotypes (BF2*0401, BF2*1201, BF2*1501 & BF2*1901) and human class I MHC molecules. These so-called promiscuous class I MHC molecules were observed by our collaborators to have a 10-fold lower expression level on the cell surface compared to the more fastidious molecules.

The BF2*2101 molecule featured a flexible Arg9 residue that enabled remodelling of the peptide binding groove, allowing very different amino acid residues at peptide positions P_2 and P_{C-2} (Koch et al., 2007). The new structures presented in this chapter showed that, for some peptides, an anchor residue at position P_2 is not required for binding. Thermostability analysis of numerous peptide complexes showed that these complexes without an anchor residue at P_2 are as stable, if not more so than those that utilise three peptide anchor residues. It also revealed that peptides that contained charged residues at both P_2 and P_{C-2} were less stable than those containing a single charged residue at the same positions. In the case of the B21 peptide H, a larger charged residue at P_2 (arginine) dramatically reduced the whole complex stability compared to a smaller histidine residue at the same position as in B21 peptide G. The two new BF2*2101 structures illustrated the plasticity of its binding groove and its ability to bind a wide variety of amino acid residues at anchor positions. Given that it is possible to predict the chicken response to infectious pathogens or vaccines by the peptide-binding motif of the dominantly expressed class I MHC molecules in some cases (Wallny et al., 2006), it would be extremely useful if the peptide-binding motif for BF2*2101 could be determined. Unfortunately this was not

possible with peptide residues outside of the anchor positions appearing to influence the allowed residues at peptide positions P_2 and P_{C-2}.

The BF2*0201 and BF2*1401 structures revealed that these class I MHC molecules bind peptides with a similar mechanism to that employed by mammalian class I molecules. They both bound peptides through anchor residues at peptide positions P_2 and P_{C-2}, with the side chains packing into hydrophobic pockets in the peptide-binding groove. Despite the BF2*0201 molecule featuring the Arg9 and Asp24 residues instrumental for the promiscuous binding mechanism seen for the BF2*2101 molecule, this mechanism does not appear to be utilised. This is likely due to the replacement of amino acid residues with smaller side chains in BF2*2101 that surround the Arg9 and Asp24 residues with larger ones in BF2*0201 (Asn69 versus Ser69 and Tyr97 versus Ser97), thereby reducing their conformational flexibility. Instead of interacting with peptide side chains as in the case of BF2*2101, Arg9 of the BF2*0201 molecule and Gln9 of BF2*1401 interact with peptide main chain oxygen atoms and as such, these interactions are independent of peptide sequence at this position.

The human class I molecule HLA-A2 is known to bind the widest variety of peptides of any human class I molecule and the mechanism of peptide binding is similar to that of the chicken molecules BF2*0201 and BF2*1401. Although human HLA-A2 has one of the widest peptide-binding repertoires of human class I MHC molecules, its anchor residue requirements are much stricter than those observed for the chicken molecules (Paul et al., 2013) and explains its more fastidious peptide-binding motif compared to the chicken molecules.

Interestingly, our collaborators were also able to demonstrate an inverse correlation between peptide binding promiscuity and cell surface expression level for human class I MHC molecules, suggesting that this property first seen with the chicken molecules, may be a general property of class I MHC molecules. Presentation of a

wider variety of peptides, either through increased numbers of MHC molecules or greater peptide binding promiscuity by individual molecules, is thought to lead to greater negative selection in the thymus, ultimately leading to a reduced number of T-cell clones (Nowak et al., 1992; Vidović and Matzinger, 1988). The inverse correlation of peptide binding promiscuity and cell surface expression level may offer a mechanism that allows similar numbers of T-cell clones to survive negative selection.

3.6.1 Future work

The structures presented in this chapter reveal the underlying mechanisms of peptide binding to these three promiscuous chicken class I MHC molecules. Unlike the fastidious class I molecules for which peptide-binding motifs could be determined simply by inspecting the sequences of bound peptides, it was not possible to determine their peptide-binding motifs. Further crystal structures of these molecules in complex with different peptides along with additional peptide sequence information may help elucidate the intricate mechanisms of binding and establish robust peptide-binding motifs for these molecules. Additionally, as our collaborators have also identified an inverse correlation between peptide-binding promiscuity and cell-surface expression level with human class I molecules, it would be interesting to explore in more detail the peptide binding mechanisms of mammalian class I molecules.

Chapter 4: The structure of SoxYZ in complex with SoxB

4.1 Introduction

Many metabolic pathways utilise carrier proteins to shuttle covalently linked intermediates between the active sites of enzymes. This arrangement confers a number of advantages, in certain situations, over systems that do not employ carrier proteins. For example, carrier proteins can facilitate the transfer of otherwise insoluble intermediates, control their reactivity and increase local concentrations (Crosby and Crump, 2012). Despite their importance, there is very little information about the molecular details of such carrier protein-partner interactions.

The bacterial Sox pathway is important for the oxidation of inorganic sulfur compounds and is an example of a metabolic pathway that employs a carrier protein to shuttle intermediates between the active sites of different partner enzymes (Sauvé et al., 2007). This chapter focuses on the interaction between the carrier protein SoxYZ complex and partner enzyme SoxB, which is inferred to catalyse the hydrolytic cleavage of thiosulfonate groups.

4.1.1 Carrier proteins in metabolic pathways

In order for carrier proteins to transfer covalently bound catalytic intermediates between enzymes, the intermediate is usually conjugated to the end of a long flexible arm (Perham, 2000). Carrier proteins and enzyme partners can form both transient and permanent complexes. Some even exist as separate protein domains in a single polypeptide chain.

Animal type I fatty acid synthases (FASs) contain all functional domains in a single ~2,500 amino acid polypeptide. These FASs feature acyl carrier protein (ACP) domains that form an integral part of the catalytic machinery, shuttling covalently

attached intermediates between the catalytic domains, maintaining high local concentrations and short diffusion times of the attached substrates (Leibundgut et al., 2008). This arrangement is also seen in the case of SrfA-C, a non-ribosomal peptide synthetase with a peptidyl carrier domain flexibly tethered to a catalytic platform. Association of the adenylation and condensation domains forms the catalytic platform and the carrier domain moves substrate between the domains (Tanovic et al., 2008). In contrast, other carrier proteins and partner enzymes form transient complexes to achieve catalysis. Type II FAS enzyme complexes, typically observed in bacteria, are functionally similar to the type I systems described, but the enzymes involved and the APC are discrete, dissociated proteins. Acyl intermediates are covalently attached to the discrete APC, which moves between enzymes (Crosby and Crump, 2012; Nguyen et al., 2014).

4.1.2 The Sox pathway

Many sulfur-oxidizing bacteria (both photosynthetic and non-photosynthetic) possess a periplasmic metabolic pathway capable of oxidising thiosulfate, an environmentally abundant and stable sulfur compound. During the oxidation process, electrons are released and subsequently used for respiratory energy generation or carbon fixation (Frigaard and Dahl, 2008). This pathway is known as the Sox pathway and is an example of a carrier protein-dependent pathway.

The Sox pathway utilises the discrete heterodimeric carrier protein complex SoxYZ. Mass spectrometry analysis of SoxY and SoxZ proteins from *Paracoccus pantotrophus* revealed that intermediates were covalently attached to a cysteine residue in the GGCGG pentapeptide, a sequence highly conserved in SoxY orthologs (Quentmeier and Friedrich, 2001). X-ray crystal structures of the *Paracoccus pantotrophus* SoxYZ complex showed that this conserved pentapeptide forms part of a swinging arm, located at the C-terminus of SoxY. In close proximity to this was a

conserved deep, apolar pocket proposed to protect labile reaction intermediates (Sauvé et al., 2007).

The oxidation of thiosulfate to sulfate by the Sox pathway involves the SoxAX, SoxB and SoxCD enzymes (Figure 4.1). Although it was not possible to co-purify proteins of the Sox pathway in complex with each other from cell extracts, when individual components from *Paracoccus pantotrophus* were purified and mixed, Sox activity was reconstituted *in vitro*. This suggested that interactions between SoxYZ and partner proteins were weak (Rother et al., 2001).

The structures of *Rhodovulum sulfidophilum* SoxAX (Bamford et al., 2002), *Thermus thermophilus* SoxB (Sauvé et al., 2009) and *Paracoccus pantorophus* SoxCD (Zander et al., 2011) revealed that each active site access channel was sufficiently wide to accommodate substrates conjugated to the carrier arm of SoxYZ. Interaction of these partner enzymes with SoxYZ was proposed to be mediated by a flexible surface loop called the 'Z-loop' found immediately adjacent to the carrier arm of SoxYZ (Sauvé et al., 2007).

The current model for oxidation of thiosulfate by the Sox pathway is as follows (Figure 4.1). The first step involves the oxidative linkage of thiosulfate to the SoxYZ carrier protein complex. This reaction is catalysed by the heterodimeric SoxAX enzyme, resulting in the formation of a cysteine S-thiosulfonate derivative of SoxYZ (Bamford et al., 2002; Dambe et al., 2005). Subsequently the terminal sulfone of the S-thiosulfonate adduct is removed as sulfate in a reaction catalysed by SoxB, before SoxCD catalyses the oxidation of the terminal sulfane to sulfone. SoxB then catalyses the removal of the terminal sulfone group of SoxYZ to regenerate a substrate-free SoxYZ species. This cyclical pathway allows thiosulfate to be completely oxidised to sulfate; however some sulfur-oxidising bacteria do not possess SoxCD (Hensen et al., 2006). In this case it is proposed that SoxL mediates the transfer of the sulfane

intermediate from SoxYZ to the sulfur globules before being subsequently oxidised to sulfate in a cytoplasmic pathway (Welte et al., 2009).

It has been reported that the Sox pathway may also be utilised for the oxidation of other reduced sulfur species such as sulfide, elemental sulfur, sulfite and tetrathionate (Appia-Ayme et al., 2001; Mukhopadhyaya et al., 2000; Wodara et al., 1997). The modular nature of the pathway would allow compatible intermediates to be fed in at appropriate stages. Repeating the reactions catalysed by SoxCD and SoxB would allow oxidation of sulfur intermediates of various lengths.

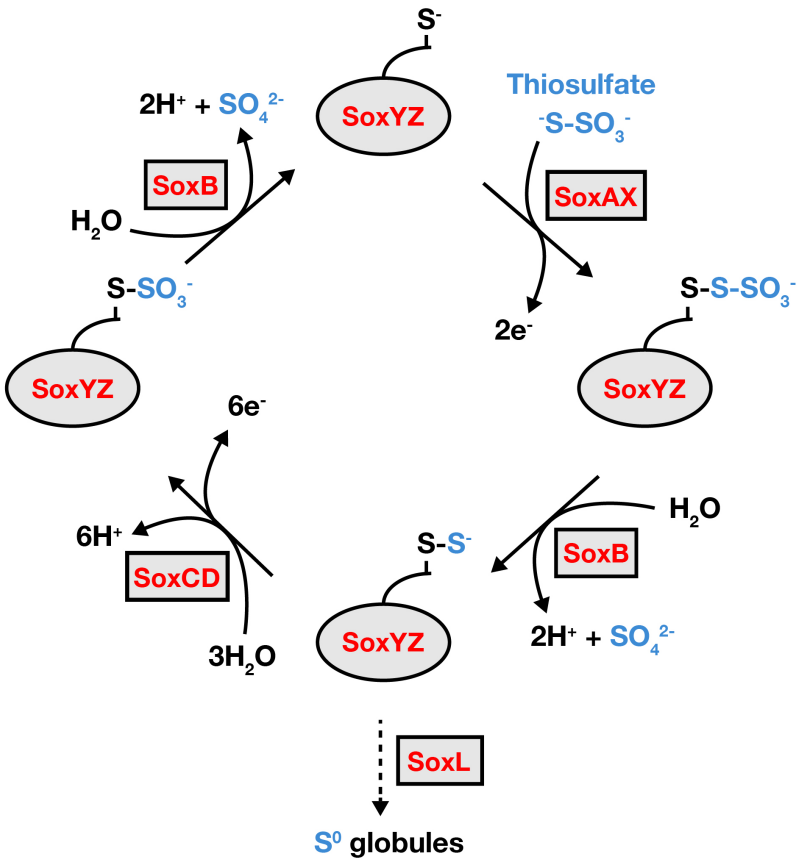

Figure 4.1 - Sox pathway model
Oxidation of thiosulfate by the Sox pathway. Pathway intermediates are carried by SoxYZ with the cysteine sulfur on the SoxY swinging arm shown as a black -S. Solid arrows show the cyclical pathway that results in the complete oxidation of thiosulfate to sulfate in some sulfur-oxidising bacteria. Other bacteria lack the SoxCD enzyme with the sulfane intermediate used to produce sulfur globules that are oxidised to sulfate by a cytoplasmic pathway (dashed arrow).

The carrier protein complex, SoxYZ, is involved in every step of the Sox pathway (Quentmeier and Friedrich, 2001), carrying sulfur compounds that not only contain variable numbers of sulfur atoms, but also feature different chemical groups at their termini. Reconstitution experiments suggest weak interactions between SoxYZ and known partner proteins (SoxAX, SoxB, SoxCD), however, specific interactions have not been demonstrated experimentally. It is likely that SoxYZ also interacts with other proteins encoded in the *sox* locus such as SoxEF/FccAB, whose precise function is not well understood (Bardischewsky et al., 2006; Chen et al., 1994). SoxYZ is also proposed to interact with as yet unidentified partner proteins to mediate sulfite oxidation in *Allochromatium vinosum* (Dahl et al., 2013).

Collaborators in the Berks laboratory (Biochemistry Department, University of Oxford) have performed a number of biochemical experiments investigating the activity of *Thermus thermophilus* SoxB and its interaction with SoxYZ from the same species. There has been no published biochemical evidence for the hydrolytic cleavage of thiosulfonate groups by SoxB and although our collaborators were unable to provide evidence for this reaction, they did show that it was capable of catalysing the hydrolysis of trithionate, a small molecule substrate analogue ($^{-}O_3S\text{-}S\text{-}SO_3^{-}$). A co-crystal structure of Thermus thermophilus SoxB (PDBID: 2WDE) showed a molecule of thiosulfate (a non-hydrolysable substrate analogue) bound to a pair of Mn(II) ions (Sauvé et al., 2009). Consistent with this observation, our collaborators were able to show that enzyme activity of the native protein was abolished in the presence of EDTA but enhanced on the addition of Mn^{2+} ions. The crystal structure showed the thiosulfate molecule bound to Arg416, a residue proposed to be important for substrate binding and transition state stabilisation (Sauvé et al., 2009). Our collaborators mutated this residue to glycine, resulting in an enzyme that exhibited no detectable hydrolase activity.

Next, our collaborators performed a series of experiments that demonstrated, for the first time, specific protein-protein interactions between SoxB and SoxYZ. They first mutated the cysteine residue in the swinging arm of SoxY to serine so that an assessment of the protein-protein interactions could be made with no contribution from the substrate-binding residue. Native PAGE analysis of SoxB and SoxY(C151S)Z at room temperature revealed SoxBYZ complex formation (Figure 4.2A). Subsequent SEC-MALS analysis also suggested complex formation, albeit in a concentration-dependent manner. The calculated molecular masses for the individual protein components were consistent with their expected masses and did not vary with sample concentration. In contrast, the calculated molecular masses for the SoxBYZ sample increased with protein concentration, indicative of an association. The highest concentration of the SoxBYZ sample resulted in a calculated molecular mass of 68 kDa for the SEC peak, not the expected 90 kDa. This suggested that at this SEC peak there was an equilibrium between the individual protein components and the SoxBYZ complex and explained why a mass higher than either of the individual components is observed and why it does not reach the expected complex molecular mass (Figure 4.2B).

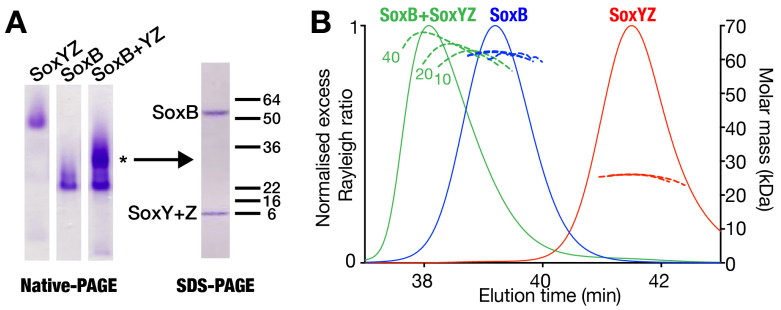

Figure 4.2 - PAGE & SEC-MALS analysis of SoxBYZ complex
Experiments carried out by collaborators. Panel A: Native PAGE samples contained 10 μM of each protein component and used the Laemmli buffer system with a 7% polyacrylamide gel. Mixture sample shows an additional major band in the native PAGE gel (*). This band was excised and subjected to SDS-PAGE analysis after boiling the sample in an SDS-containing buffer. Proteins visualised with Coomassie stain. Panel B: SEC-MALS analysis of individual proteins and mixture. Proteins loaded at 10 μM, 20 μM and 40 μM per component.

In order to quantitatively assess the interaction between SoxB and SoxYZ, our collaborators performed a series of ITC experiments (Figure 4.3A-D). The first experiment implied a simple 1:1 association of SoxB and SoxY(C151S)Z with a K_D of 3 μM (Figure 4.3A). They next generated a series of SoxYZ derivative species to investigate the influence of bound substrate to the SoxY carrier arm. A S-thiosulfonate derivative, SoxY(SSO_3)Z, did not significantly alter the K_D, however the ΔH of the interaction with the S-thiosulfonate derivative switched from being exothermic for the SoxY(C151S)Z interaction to endothermic, suggesting the presence of the thiosulfonate adduct perturbs the interaction mechanism (Figure 4.3A & 3B). Next, a S-carboxymethyl derivative, SoxY(Ac)Z, was produced and subsequently converted to its amide derivative, SoxY(Am)Z. The SoxY(Am)Z derivative would be unable to form interactions with the Mn(II) ions present in the SoxB active site. ITC experiments showed that SoxY(Ac)Z binding was similar to that of SoxY(SSO_3)Z whereas SoxY(Am)Z exhibited an exothermic enthalpy change (Figure 4.3C & D), similar to that of the SoxY(Cys151S)Z interaction. Taken together these results suggest that while the K_D is not affected by the species conjugated to the end of the carrier arm, the mechanism of interaction is. SoxYZ adducts that would be unable to coordinate the Mn(II) ions exhibited enthalpically-driven interactions, whereas those able to coordinate the Mn(II) ions exhibited entropically-driven interactions.

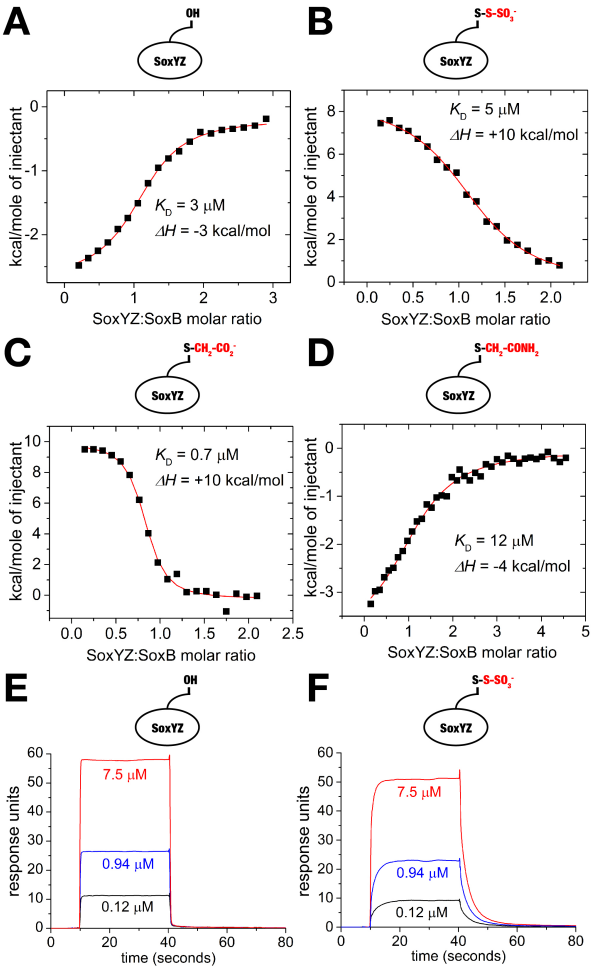

Figure 4.3 - Binding of SoxB to SoxYZ derivatives
Panels A-D: Isothermal calorimetry experiments with SoxYZ species indicated above integrated heat plots. Panels E & F: Surface plasmon resonance sensorgrams showing association and dissociation of SoxYZ derivatives and SoxB.

SPR experiments were carried out by our collaborators to investigate the reaction kinetics of SoxB and SoxY(C151S)Z or SoxY(SSO_3). Equilibrium binding experiments yielded similar K_D values to those obtained by ITC (2.7μM and 2.4 μM respectively). Interestingly, their kinetic analysis showed that the association and dissociation rates for the SoxB-SoxY(C151S)Z interaction were extremely rapid (Figure 4.3E) in contrast to those for the SoxB-SoxY(SSO_3)Z interaction (Figure 4.3F). They were unable to reliably simultaneously fit the kinetic parameters due to the speed of the association and dissociation phases, but fitting the dissociation phases to a simple exponential decay model gave an apparent k_{off} value of 5.7 ± 0.8 s^{-1} for the

dissociation of SoxBY(C151S)Z, 0.34 ± 0.01 s^{-1} for the major dissociation phase of SoxBY(SSO_3)Z and 4.5 ±0.7 s^{-1} for a minor burst phase of SoxBY(SSO_3)Z.

4.2 Aims

The aim of this chapter is to determine the x-ray crystal structure of the *Thermus thermophilus* SoxBYZ complex. This would reveal molecular details of the interaction between the SoxYZ carrier protein and SoxB partner enzyme. It may also provide insight into how SoxYZ is able to achieve both specific yet promiscuous binding to numerous other structurally distinct partner proteins.

4.3 Initial crystallisation, data collection and processing

Previous crystallisation attempts of the *Thermus thermophilus* SoxBYZ complex using purified components were unsuccessful. Given the relatively low affinity of the interaction between SoxYZ and SoxB (K_D = 3 μM), our collaborators in the Berks Lab (Biochemistry Department, University of Oxford) were able to stabilise the otherwise weak and transient SoxBYZ interaction using a 'suicide enzyme' strategy. Trp175 in the active site of SoxB was mutated to cysteine allowing it to form a disulfide bond to the cysteine residue in the carrier arm of SoxY. Due to the conformational flexibility of the carrier arm, protein-protein interactions outside of the active site should not be affected. Our collaborators confirmed the formation of this disulfide cross-linked complex using SDS-PAGE (Figure 4.4) and mass spectrometry.

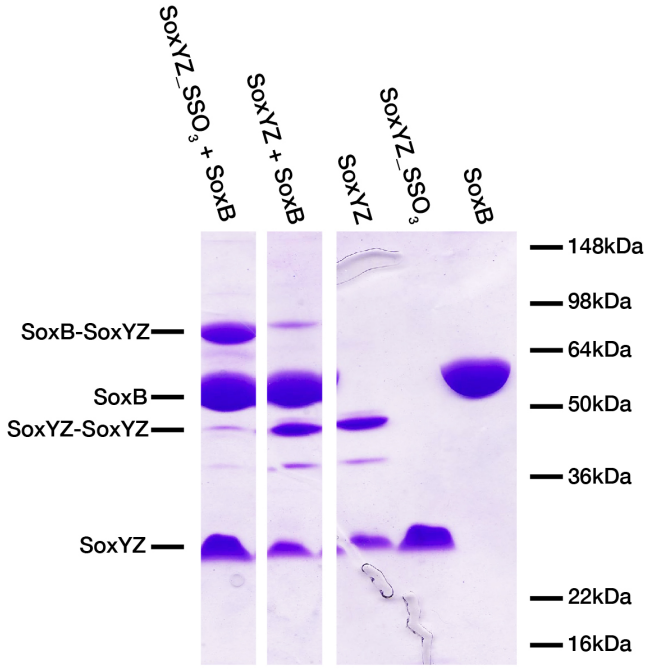

Figure 4.4 - Analysis of disulfide-linked SoxBYZ complex
SDS-PAGE analysis of SoxB incubation with either SoxYZ or SoxY(SSO_3)Z along with individual protein samples. Carried out using a 12% polyacrylamide gel under non-reducing conditions. Experiment carried out by collaborators in the Berks laboratory (Biochemistry Department, University of Oxford).

4.3.1 Initial crystallisation of SoxBYZ

Crystallisation trials of the disulfide-linked SoxBYZ complex were carried out using a number of commercially available screens (Molecular Dimensions), purchased in 96 deep-well plates. A Hydra 96 (Alpha Biotech Ltd) robot was used to transfer 70 µL of each crystallisation condition into the reservoirs of MRC 2-drop crystallisation plates. The sitting-drop vapour-diffusion method was chosen to attempt crystallisation of the sample and was set up using an Oryx Nano-crystallisation robot (Douglas Instruments, UK). Plates were subsequently sealed using a StarSeal Advanced Polyolefin Film (STARLAB) and incubated at 21°C.

Crystals grew at 20°C within 1 week in condition E11 (0.16M calcium acetate hydrate, 0.08M sodium cacodylate, pH6.5, 14.4% PEG 8000) of the JCSG-plus screen (Molecular Dimensions) (Figure 4.5). Drops of 0.55 µL solution contained 50% protein (~70 µM) and 50% mother liquor.

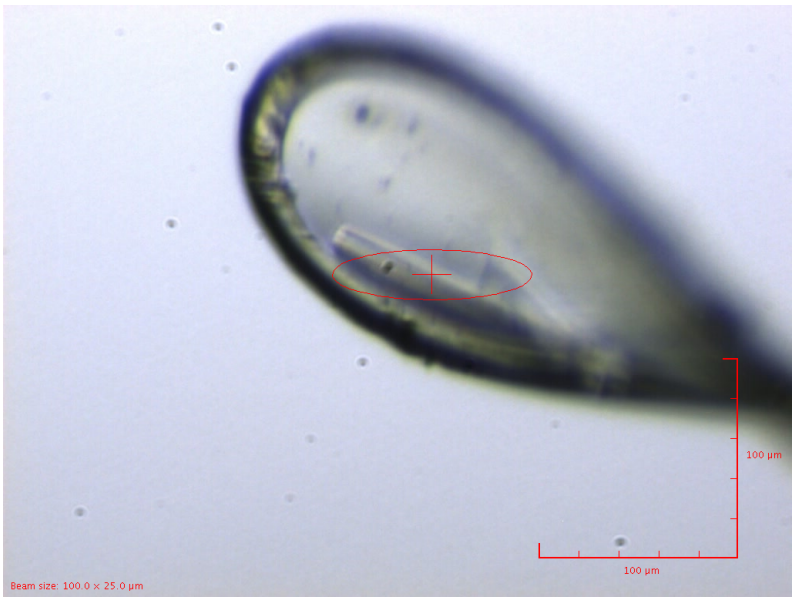

Figure 4.5 - Initial SoxBYZ crystal
SoxBYZ crystal mounted at Diamond Light Source's i02 beamline. Red oval shows beam dimensions and scale bars in the bottom right indicate a 100µm x 100µm square.

4.3.2 Data collection

Crystals were cyroprotected with mother liquor plus 20% (v/v) ethylene glycol and flash frozen in liquid nitrogen. A dataset was collected at Diamond Light Source's i02 beamline. In case of incorrect space group determination at this stage, 180 degrees of data were collected (Table 4.1).

Data Collection	
Diffraction source	Diamond Light Source, i02
Detector	Pilatus 6M
Energy (eV)	12658
Wavelength (Å)	0.9795
Temperature (K)	100
Detector resolution (Å)	2.8
Exposure per image (s)	0.2
Ω start (º)	0
Ω end (º)	180
Ω osc (º)	0.2
Number of images	900

Table 4.1 - Initial data collection

4.3.3 Data processing

Data were processed using XIA2 (Winter et al., 2013) running in 3dai mode, which indexed with a monoclinic cell (a = 69.9 Å, b = 223.1 Å, c = 120.1 Å, α = 90.00°, β = 97.27 °, γ = 90.00 °). The last 190 images were removed from the processing as analysis by batch revealed they had high R_{merge} values and the data were already complete at this point. This gave data to 3.9 Å in the $P2_1$ spacegroup (Table 4.2 and Figures 4.6 & 4.7).

Data Collection	
Resolution range (Å)	81.43-3.90 (4.36-3.90)
Space group	P1 2_1 1
Unit cell dimensions	
a, b, c (Å)	69.9, 223.1, 120.1
α, β, γ (°)	90.00, 97.27, 90.00
Total reflections	89814 (25797)
Unique reflections	32401 (9222)
Multiplicity	2.8 (2.8)
Completeness (%)	97.6 (97.9)
Mean I/sigma(I)	5.5 (2.4)
R_{merge} (%)	19.6 (53.2)
R_{pim} (%)	14.1 (38.4)
CC(1/2)	0.978 (0.802)

Table 4.2 - Initial data processing

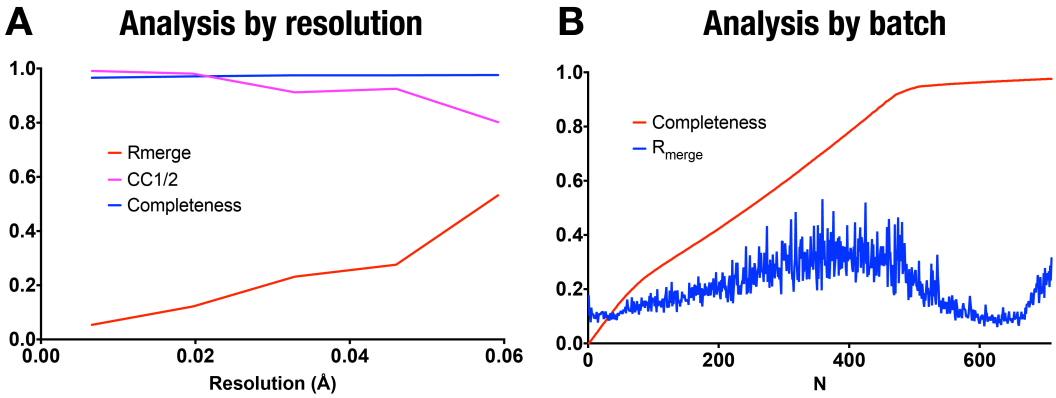

Figure 4.6 - Initial data processing analysis
Panel A: Analysis by resolution. Panel B: Analysis by batch.

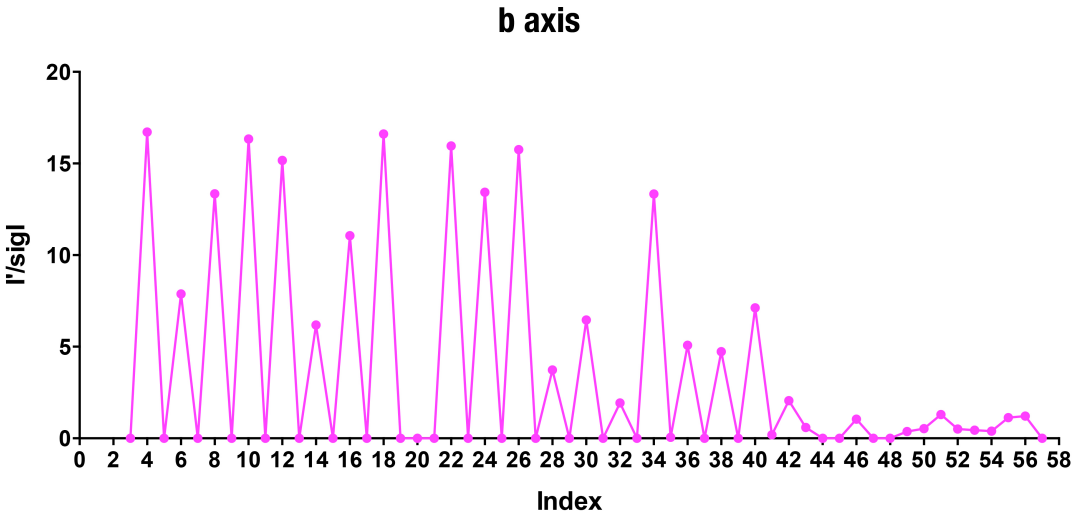

Figure 4.7 - Systematic absences present in initial dataset
Systematic absences are consistent with the $P2_1$ spacegroup (k = 2n)

4.4 Initial structure determination

Although the resolution of this dataset was low (3.9 Å), the availability of a number of relevant x-ray crystal structures (Table 4.3) made a structure solution by molecular replacement possible. Although such a solution would be difficult to refine, it would enable possible verification that the crystals did indeed contain all of the expected protein components and would also give insight into their arrangement in the SoxBYZ complex.

Protein	Organism	PDBID
SoxB	*Thermus thermophilus*	2WDF
SoxB-glycerol	*Thermus thermophilus*	2WDC
SoxB-sulfate	*Thermus thermophilus*	2WDD
SoxB-thiosulfate	*Thermus thermophilus*	2WDE
SoxZ	*Thermus thermophilus*	1V8H
SoxYZ	*Paracoccus pantotrophus*	2OX5, 2OXG, 2OXH

Table 4.3 - Potential molecular replacement search models

4.4.1 Structure determination

PHENIX.XTRIAGE (Adams et al., 2010) suggested no evidence of twinning or translational non-crystallographic symmetry. Cell content analysis was carried out using MATTHEWS (Kantardjieff and Rupp, 2003), suggesting the presence of between three to five copies of the SoxBYZ complex assuming a total molecular weight of 88.9kDa (Table 4.4).

Nmol / asym	Matthews coefficient	% solvent	P(3.90)	P(tot)
1	10.45	88.23	0.00	0.00
2	5.22	76.47	0.01	0.00
3	3.48	64.70	0.12	0.08
4	2.61	52.93	0.52	0.47
5	2.09	41.17	0.34	0.43
6	1.74	29.40	0.01	0.01
7	1.49	17.63	0.00	0.00
8	1.31	5.86	0.00	0.00

Table 4.4 - Cell content analysis of initial dataset

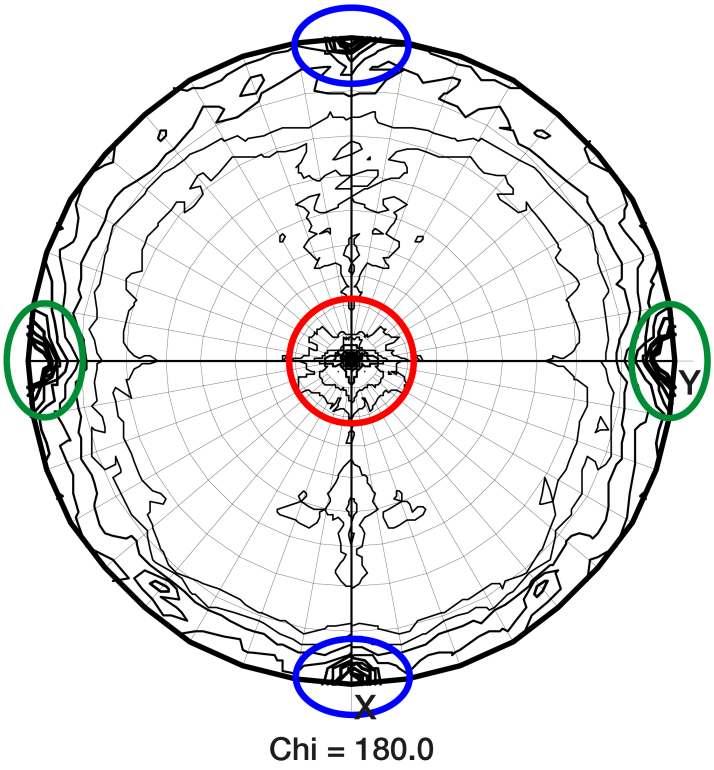

Figure 4.8 - Self-rotation function peaks in initial dataset
Visual representation of the self-rotation function peaks generated by MOLREP (Vagin and Teplyakov, 2010). Peaks with intensities over 30% of the origin peak are highlighted. Green ovals highlight the crystallographic 2-fold axis along b. Red and blue ovals / circles correspond to peaks in Table 4.5.

θ	Φ	χ	P(i)/P(0)
0	0	0	1
179.42	**1.80**	**180**	**0.45**
90.00	**0**	**180**	**0.44**

Table 4.5 - Top self-rotation function peaks of initial dataset

Analysis of the self-rotation function calculated with MOLREP (Vagin and Teplyakov, 2010) revealed two non-origin peaks with intensities over 30% of the origin peak intensity. Both peaks occurred when Chi = 180.0, suggesting the presence of two 2-fold non-crystallographic symmetry axes (Figure 4.8 and Table 4.5). These identified non-crystallographic axes along the a and c crystallographic axes suggesting a pseudo-orthorhombic system when considered with the crystallographic 2-fold axis down b.

The molecular replacement program PHASER (McCoy et al., 2007) was able to place four copies of *Thermus thermophilus* SoxB (PDBID: 2WDF) (Table 4.6).

Copy	RFZ	TFZ	LLG
SoxB (2WDF)	16.2	14.3	330
SoxB (2WDF)	12.1	32.5	1129
SoxB (2WDF)	11.9	41.7	2208
SoxB (2WDF)	7.4	38.6	3307

Table 4.6 - Molecular replacement hits with SoxB model in initial dataset

Attempts to locate 4 copies of search models derived from *Paracoccus pantotrophus* SoxY (PDBID: 2OX5) and *Thermus thermophilus* SoxZ (PDBID: 1V8H) were unsuccessful after initially placing the 4 copies of *Thermus thermophilus* SoxB. A SoxYZ search model was generated using the *Paracoccus pantotrophus* SoxYZ structure as a template. The coordinates of *Thermus thermophilus* SoxZ were aligned to the SoxZ copy in the *Paracoccus pantotrophus* SoxYZ complex structure using SUPERPOSE (Krissinel and Henrick, 2004). A model of *Thermus thermophilus* SoxY was generated with SCWRL (Krivov et al., 2009) and used to replace the SoxY copy in the *Paracoccus pantotrophus* SoxYZ structure. Flexible terminal regions were then removed along with a large loop in SoxZ. It was only possible to find a single copy of this SoxYZ model using PHASER (McCoy et al., 2007).

In an attempt to improve the molecular replacement solution, the model containing 4 copies of SoxB was rebuilt and subjected to a single round of refinement using COOT (Emsley et al., 2010) and AUTOBUSTER (Bricogne et al.) (R_{work} = 26.92% and R_{free} = 33.18%). A single copy of SoxB from the refined model was used in another molecular replacement search with PHASER (McCoy et al., 2007) (Table 4.7).

Copy	RFZ	TFZ	LLG
Refined SoxB	18.2	13.6	480
Refined SoxB	13.6	39.2	1487
Refined SoxB	14.0	45.7	2826
Refined SoxB	7.7	34.7	4122

Table 4.7 - Molecular replacement hits with refined SoxB model in initial dataset

With four copies of the refined SoxB model placed, it was possible to locate three copies of the SoxYZ model (Table 4.8). Inspection of these molecular replacement hits revealed that although the three SoxYZ models were placed in similar positions with respect to their SoxB partners, these placements were not identical (Figure 4.9). It was not possible to locate a fourth copy of SoxYZ or indeed individual copies of SoxY and SoxZ using PHASER, likely due to the poor quality of the low-resolution data.

Copy	RFZ	TFZ	LLG
SoxYZ model	3.8	9.5	4740
SoxYZ model	3.8	11.0	4719
SoxYZ model	3.0	5.7	5597

Table 4.8 - Molecular replacement hits with SoxYZ model in initial dataset

Self-rotation function peaks (Figure 4.8 & Table 4.5) could now be explained by investigating the rotational element of the superposition of each SoxB copy onto the other SoxB copies, including those generated by the crystallographic 2-fold axis, using SUPERPOSE (Krissinel and Henrick, 2004). This showed that the first non-origin peak (θ = 179.42, Φ = 1.80, χ = 180) appeared as a result of a 2-fold non-crystallographic axis relating the first placed copy of SoxB with a symmetry copy of the third, as well as the second placed copy of SoxB with a symmetry copy of the fourth. The second non-origin peak (θ = 90.00, Φ = 0, χ = 180) resulted from a 2-fold non-crystallographic axis relating the first placed copy of SoxB to the third and also the second placed copy of SoxB to the fourth copy.

This partial solution was not refined and new crystals of the complex were grown in order to obtain high resolution and better quality data.

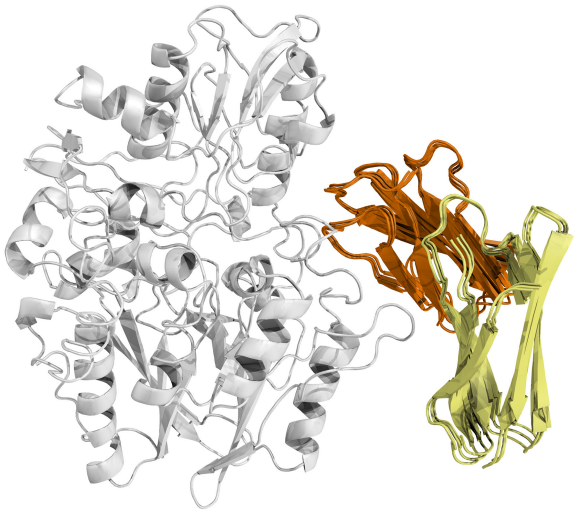

Figure 4.9 - Structural alignment of three placed copies of SoxBYZ
Proteins displayed as cartoons with SoxB in grey, SoxY in orange and SoxZ in yellow. Three copies of the SoxBYZ complex, aligned by SoxB, show similar positions of SoxYZ relative to SoxB.

4.5 Final crystallisation, data collection and processing

4.5.1 Crystallisation of SoxBYZ

Crystallisation trials were set up as described in section 1.3.1. New crystals grew at 20°C within 1 week in a condition containing 0.1M Tris-HCl, 0.2M ammonium sulfate, 8.55% PEG 8000, pH 8.3 (Figure 4.10). The drop contained 70% protein solution at a concentration of ~93 µM and 30% mother liquor.

4.5.2 Final data collection

Crystals were cyroprotected in mother liquor plus 30% (v/v) ethylene glycol and flash frozen in liquid nitrogen before a dataset was collected at Diamond Light Source's i04 beamline (Table 4.9).

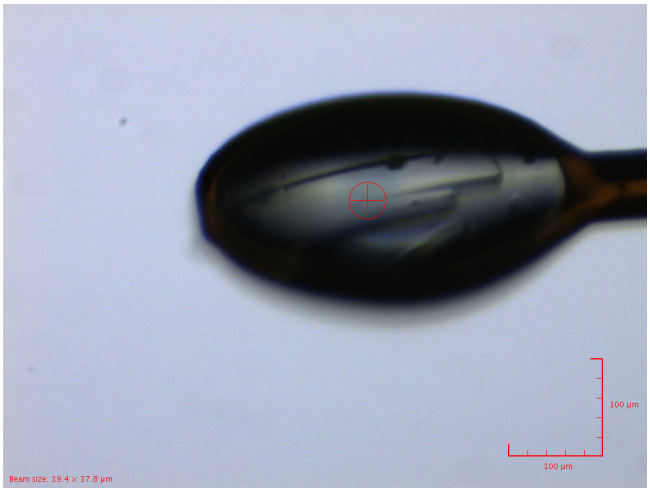

Figure 4.10 - SoxBYZ crystal
SoxBYZ crystal mounted at Diamond Light Source's i04 beamline. Red circle shows beam dimensions and scale bars in the bottom right indicate a 100µm x 100µm square.

Data Collection	
Diffraction source	Diamond Light Source, i04
Detector	ADSC Quantum 315R
Energy (eV)	12400
Wavelength (Å)	0.9999
Temperature (K)	100
Detector resolution (Å)	2.8
Exposure per image (s)	1.0
Ω start (º)	0
Ω end (º)	180
Ω osc (º)	0.5
Number of images	360

Table 4.9 - Data collection parameters for final dataset

4.5.3 Data processing

Data were processed using XIA2 (Winter et al., 2013) running in 3dai mode, indexing a triclinic cell (a = 70.22 Å, b = 115.9 Å, c = 120.9 Å, α = 86.51°, β = 83.21°, γ = 89.77°) with data to a resolution of 3.28 Å in the P1 space group (Table 4.10 and Figure 4.11).

Data Collection	
Resolution range (Å)	59.81-3.28 (3.37-3.28)
Space group	P1
Unit cell dimensions	
a, b, c (Å)	70.22, 115.9, 120.9
α, β, γ (°)	86.51, 83.21, 89.77
Total reflections	112492 (8416)
Unique reflections	57115 (4256)
Multiplicity	2.0 (2.0)
Completeness (%)	98.7 (98.4)
Mean I/sigma(I)	7.9 (1.9)
R_{merge} (%)	8.8 (30.2)
R_{pim} (%)	8.8 (30.2)
CC(1/2)	0.99 (0.74)

Table 4.10 - Data processing statistics for final dataset

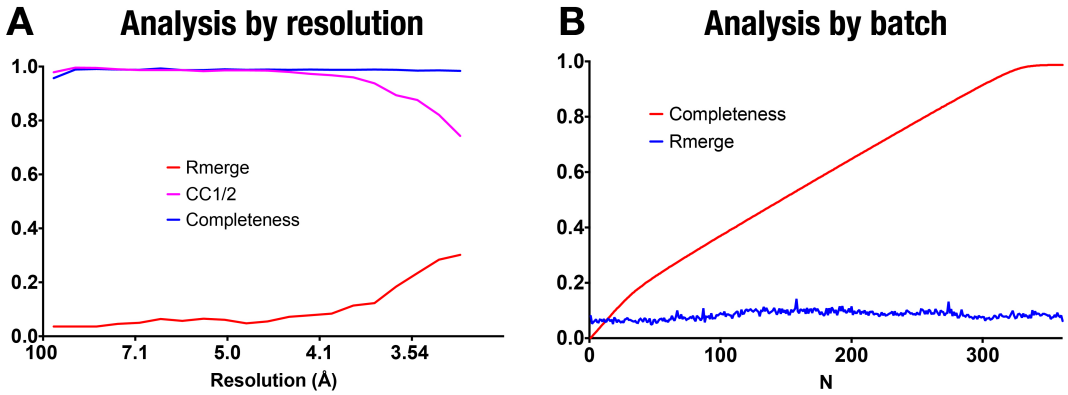

Figure 4.11 - Data processing analysis of final dataset
Panel A: Analysis by resolution. Panel B: Analysis by batch.

Analysis of the data processing statistics showed that the new dataset is of higher resolution (3.28 Å) and had higher mean I/sigma(I), CC(1/2) and lower R_{pim} values than the initial dataset. There was no obvious deterioration in image quality as the data collection progressed, as evident by consistent R_{merge} by batch. This offers a significant improvement over the initial dataset.

4.6 Final structure determination and refinement

As molecular replacement was successful in placing 4 copies of SoxB and 3 copies of the SoxYZ model in the lower resolution dataset, the same approach was taken with the new dataset.

4.6.1 Structure determination

The three placed copies of the SoxBYZ complex in the lower resolution dataset featured similar placements of the SoxYZ model relative to the SoxB copy. This suggested that the molecular replacement solutions were likely real and that these models could be used as search models for the new dataset.

Cell content analysis by MATTHEWS (Kantardjieff and Rupp, 2003) suggested the presence of either 3 to 5 copies of SoxBYZ in the asymmetric unit (Table 4.11).

Nmol / asym	Matthews coefficient	% solvent	P(3.15)	P(tot)
1	10.97	88.80	0.00	0.00
2	5.49	77.60	0.00	0.00
3	3.66	66.39	0.08	0.05
4	2.74	55.19	0.41	0.35
5	2.19	43.99	0.47	0.54
6	1.83	32.79	0.03	0.05
7	1.57	21.59	0.00	0.00
8	1.37	10.39	0.00	0.00

Table 4.11 - Cell content analysis of final dataset

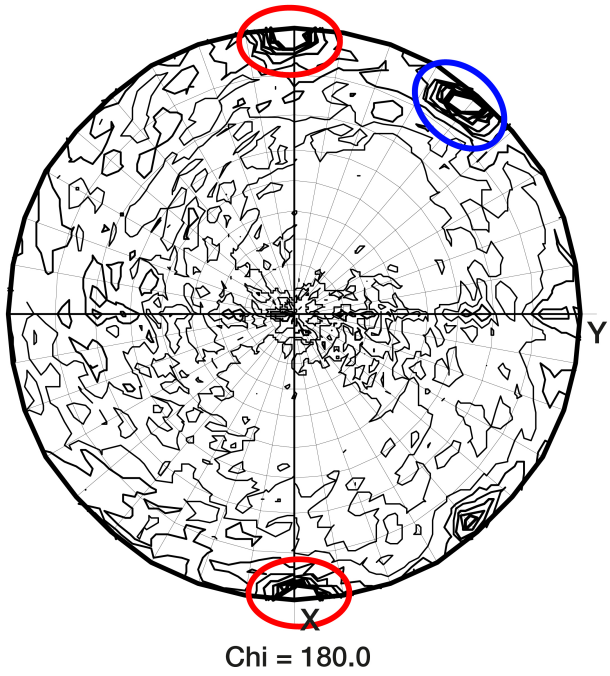

Figure 4.12 - Self-rotation function for final dataset
Visual representation of the self-rotation function peaks generated with MOLREP (Vagin and Teplyakov, 2010). Peaks with intensities over 30% of the origin peak are highlighted.

θ	Φ	χ	P(i)/P(0)
0	0	0	1
90	**0.44**	**180**	**0.6**
86.95	**141.55**	**179.91**	**0.31**

Table 4.12 - Self-rotation function peaks for final dataset
Self-rotation function peaks over 30% of the origin peak. Related to Figure 4.12.

Analysis of the self-rotation function revealed two peaks with intensities over 30% of the origin peak intensity. Both of these peaks appeared where Chi = 180.0 suggesting the presence of two 2-fold non-crystallographic symmetry axes (Figure 4.12 and Table 4.12).

Molecular replacement was carried out with PHASER (McCoy et al., 2007) using a model arbitrarily chosen from the initial molecular replacement solution of the lower resolution dataset. Four copies of this SoxBYZ model were placed with Z-scores and LLGs shown in Table 4.13. It was not possible to find any additional

copies of the SoxBYZ complex and inspection of this molecular replacement solution confirmed that there were sufficient crystal contacts for the crystal to pack.

Copy	RFZ	TFZ	LLG
SoxBYZ – 1	12.6	-	119
SoxBYZ – 2	9.5	17.1	293
SoxBYZ – 3	7.6	16.8	334
SoxBYZ – 4	7.9	15.6	275

Table 4.13 - Molecular replacement hits with SoxBYZ model in final dataset

Aligning the four copies placed of SoxBYZ onto one another using SUPERPOSE (Krissinel and Henrick, 2004) allowed rationalisation of the self-rotation function peaks (Figure 4.12 and Table 4.12). The strongest non-origin peak (θ = 90, Φ = 0.44, χ = 180) corresponded to a 2-fold rotation axis that related the first placed copy of the complex (chains A, B & C) and the third placed copy (chains G, H & I) as well as the second placed copy (chains D, E & F) and the fourth placed copy (chains J, K & L). The second strongest non-origin peak (θ = 86.95, Φ = 141.55, χ = 179.91) corresponded to a 2-fold axis relating the first placed copy (chains A, B & C) to the fourth placed copy (chain J, K & L). A 2-fold rotation axis related the second placed copy (chains D, E & F) with the third placed copy (chains G, H & I) and explained a weaker peak (0.15 of the origin) not highlighted in Figure 4.12 or Table 4.12 (θ = 90, Φ = 0.44, χ = 180).

4.6.2 Model building and refinement

Due to the slight differences in the placement of SoxYZ copies relative to SoxB in the molecular replacement solution for the lower resolution dataset, an initial rigid body refinement was carried out using AUTOBUSTER (Bricogne et al.). Each chain was defined as a rigid body, allowing refinement of the positions of individual domains relative to one another. Cycles of model building and refinement were then carried out

using COOT (Emsley et al., 2010) and AUTOBUSTER (Bricogne et al.). Due to the relatively low number of reflections in this dataset and the large number of atoms in the model, refinements were carried out with non-crystallographic symmetry restraints to improve the parameters-to-observations ratio. Automatic NCS detection by AUTOBUSTER (Smart et al., 2012) was carried out and inspection of the log file confirmed the expected 18 NCS relations and generation of appropriate local structure similarity restraints (LSSR). These restraints were defined by considering pairs of non-bonded atoms within 5.5 Å of each other with the difference between these distances, in the target structure and the structure being refined, used to impose penalties. In addition, the high-resolution structure of *Thermus thermophilus* SoxB (PDB: 2WDF) was used as a target structure, generating additional LSSR restraints to further condition the refinement. The x-ray weight was set to 1 and was not automatically adjusted to prevent poor model geometry quality. The structure was refined to an R_{work} of 26.96% and R_{free} of 27.47% (Table 4.14). The final model statistics are in line with other deposited structures at this resolution (Figure 4.13).

Refinement statistics	
Resolution range (Å)	59.81-3.28
Number of reflections (work / test set)	54219/2894 (3898/232)
Number of atoms (protein, glycans, ligands, waters)	23062, 0, 8 Mn^{2+}, 11
Protein residues	2952
Mean B-factors (protein, glycans, ligands, waters)	92.8
R_{work} (%)	26.96 (27.76)
R_{free} (%)	27.47 (30.04)
R.M.S.D. from ideal values (bonds, Å)	0.007
R.M.S.D. from ideal values (angles, °)	0.84
Ramachandran favoured (%)	94.29
Ramachandran outliers (%)	0.24
Molprobity score	1.48 (100th percentile)

Table 4.14 - SoxBYZ refinement statistics

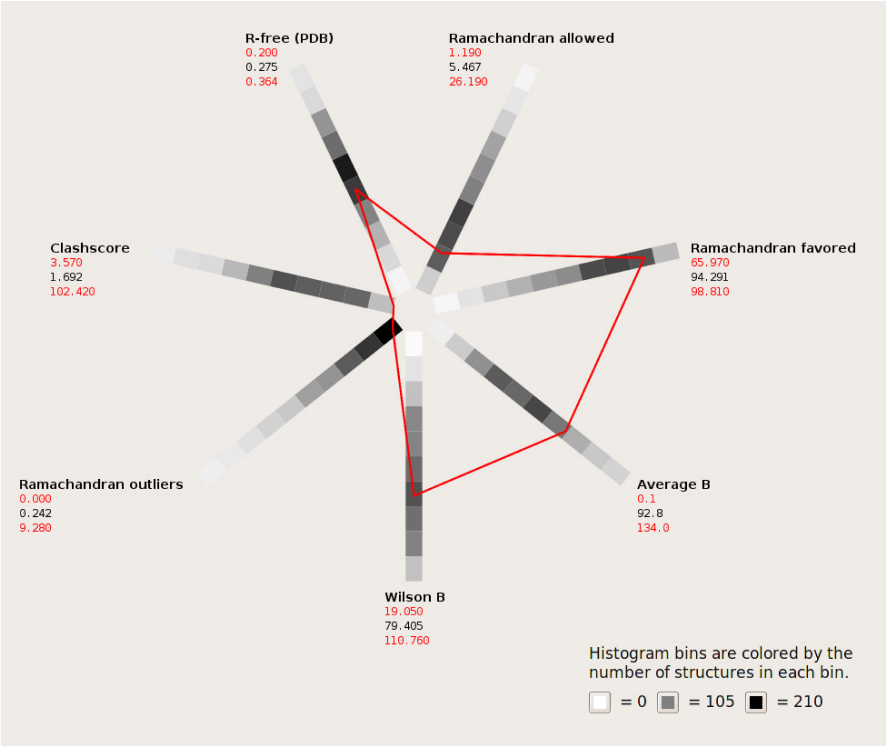

Figure 4.13 - Polygon analysis of final refinement
Polygon analysis in PHENIX (Adams et al., 2010) shows various how various statistics of the final refinement of SoxBYZ compare to other deposited structures of similar resolution.

Comparison of the two crystal forms suggests they were related. Inspection of unit cell dimensions of both crystals showed that the a and b dimensions were almost identical and that the b dimension of the initial P 2_1 crystal was approximately double that of the final P 1 crystal. The γ angles were very similar and the β angles were equivalent, both approximately 7° from 90°. There was however, a slightly more significant difference in the α angles of approximately 3.5°. The P 1 crystal form contained 4 copies of the SoxBYZ complex, whereas the P 1 2_1 1 crystal form contained 8, consistent with the unit cell being approximately double the volume. When the two structures were aligned and symmetry copies displayed in COOT (Emsley et al., 2010), the SoxBYZ copies did indeed occupy similar positions, confirming the two crystals were related.

4.7 Structure of SoxBYZ

The final 3.28 Å SoxBYZ complex structure featured four well-ordered copies of SoxB with unambiguous, continuous electron density. In contrast, the SoxY and SoxZ protein components exhibited high B-factors and ambiguous, non-continuous electron density, except for in one copy. Despite this, each of the four copies of the complex exhibited the same overall arrangement. The most well ordered copy was used for subsequent structural analyses.

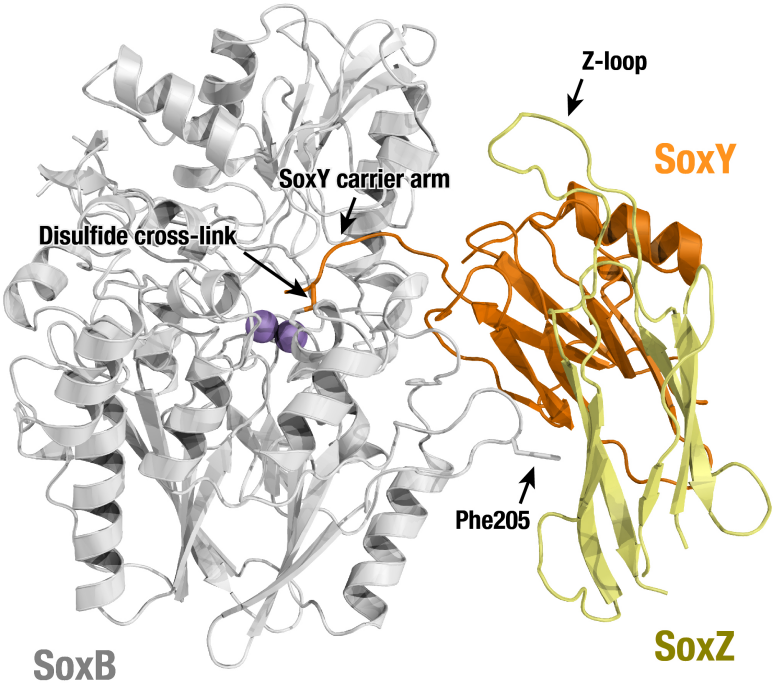

Figure 4.14 – Overall SoxBYZ structure
SoxBYZ complex with protein displayed as cartoon representations, manganese ions as purple spheres and the disulfide bond between the SoxY carrier arm and SoxB as sticks. The Phe205 side chain that is part of the Z-patch is shown as sticks.

The structure solution revealed the organisation of the protein components involved in this carrier protein partner enzyme interaction (Figure 4.14). Structural analysis using PISA (Krissinel and Henrick, 2007), showed two main sites of interaction between SoxB and SoxYZ (Figure 4.15A). The SoxB-SoxYZ interaction buried a total surface area of 1,220 Å^2 with the SoxB/SoxY (Y-patch) and SoxB/SoxZ (Z-patch) interactions burying total surface areas of 990 Å^2 and 230 Å^2, respectively. The Z-loop

of SoxZ, which was predicted to be involved in interactions with partner enzymes, made no contact with SoxB, strongly suggesting it was not involved in the interaction (Figure 4.14). The total surface buried by the interaction is relatively small and is consistent with its low affinity.

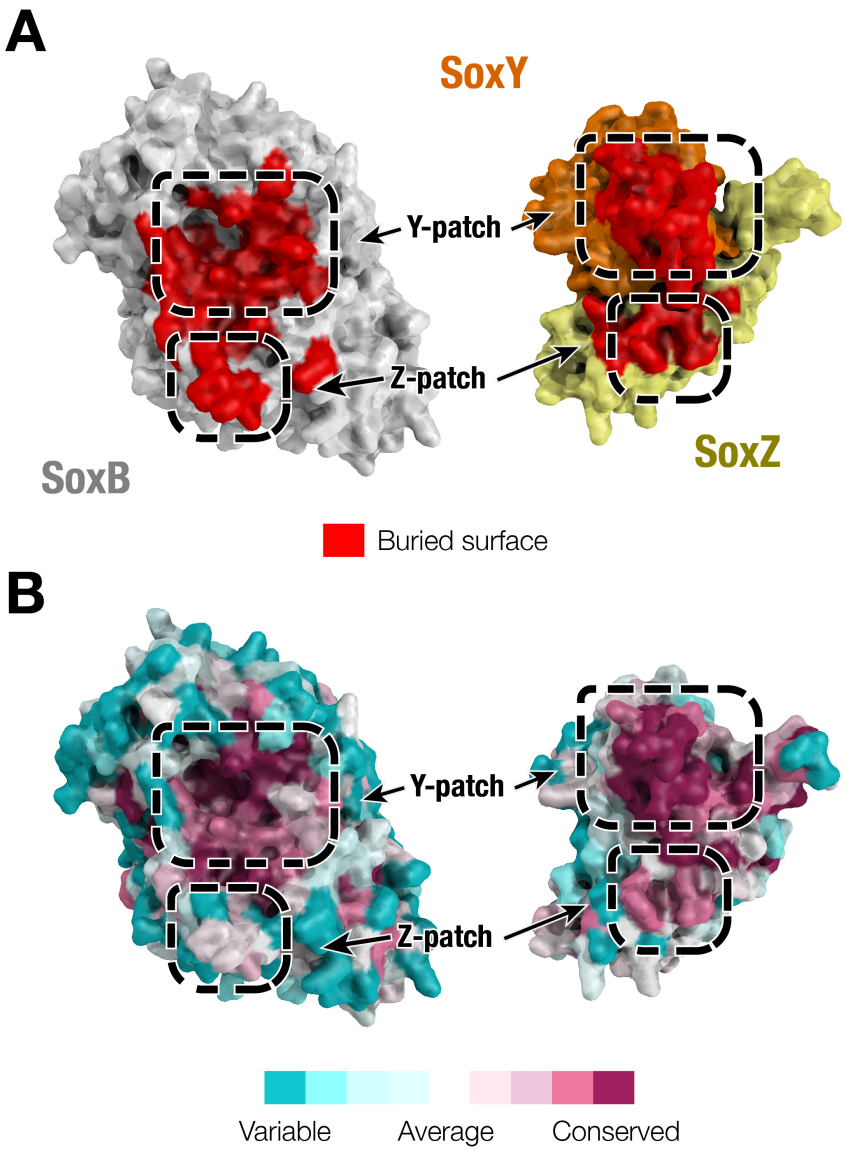

Figure 4.15 - Buried surfaces and CONSURF analysis
Surface representations of separated SoxB and SoxYZ proteins. Y- and Z-patches highlighted with dashed outlines. Panel A: Surfaces that become buried when the proteins interact are coloured red. Panel B: CONSURF (Ashkenazy et al., 2010) analysis displaying sequence conservation. Cyan represents the most variable regions and purple represents the most highly conserved regions. Z-patch conservation may be underestimated due to insertions and deletions in this region causing difficulties in the sequence alignment process.

Sequence conservation analysis of each protein component was carried out using CONSURF (Ashkenazy et al., 2010). Sequences used in this analysis were chosen by performing a context-specific iterated BLAST search and selecting those with sequence identity between 35% and 95% to the target sequence. A total of 150, 102 and 55 unique protein sequences were considered for SoxB, SoxY and SoxZ, respectively. The analysis showed that the residues involved in the SoxB-SoxYZ interaction were some of the most well conserved surface residues (Figure 4.15), as would be expected if they were involved in protein-protein interactions.

4.7.1 Y-patch interactions

The most striking structural difference between the protein components of the SoxBYZ complex and previously determined x-ray crystal structures was observed in the carrier arm of SoxY. The C-terminal SoxY carrier arm exhibited a large change in conformation when compared to the *Paracoccus pantotrophus* SoxYZ structure in the absence of SoxB (Figure 4.16). The carrier arm in the *Paracoccus pantotrophus* structure packed against a pocket formed between SoxY and SoxZ, whereas in the new *Thermus thermophilus* SoxBYZ complex structure, it occupied the active site channel of SoxB.

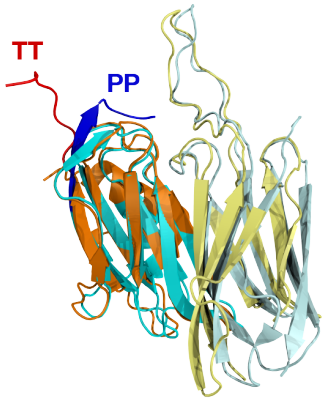

Figure 4.16 - T. thermophilus and P. pantotrophus SoxYZ backbone alignment
Proteins are shown in cartoon representation. T. thermophilus SoxY coloured orange and SoxZ in yellow. P. pantotrophus SoxYZ structure (PDBID: 2OX5) shown in shades of blue. Carrier arms of the T. thermophilus and P. pantotrophus structures are coloured in red and blue, respectively.

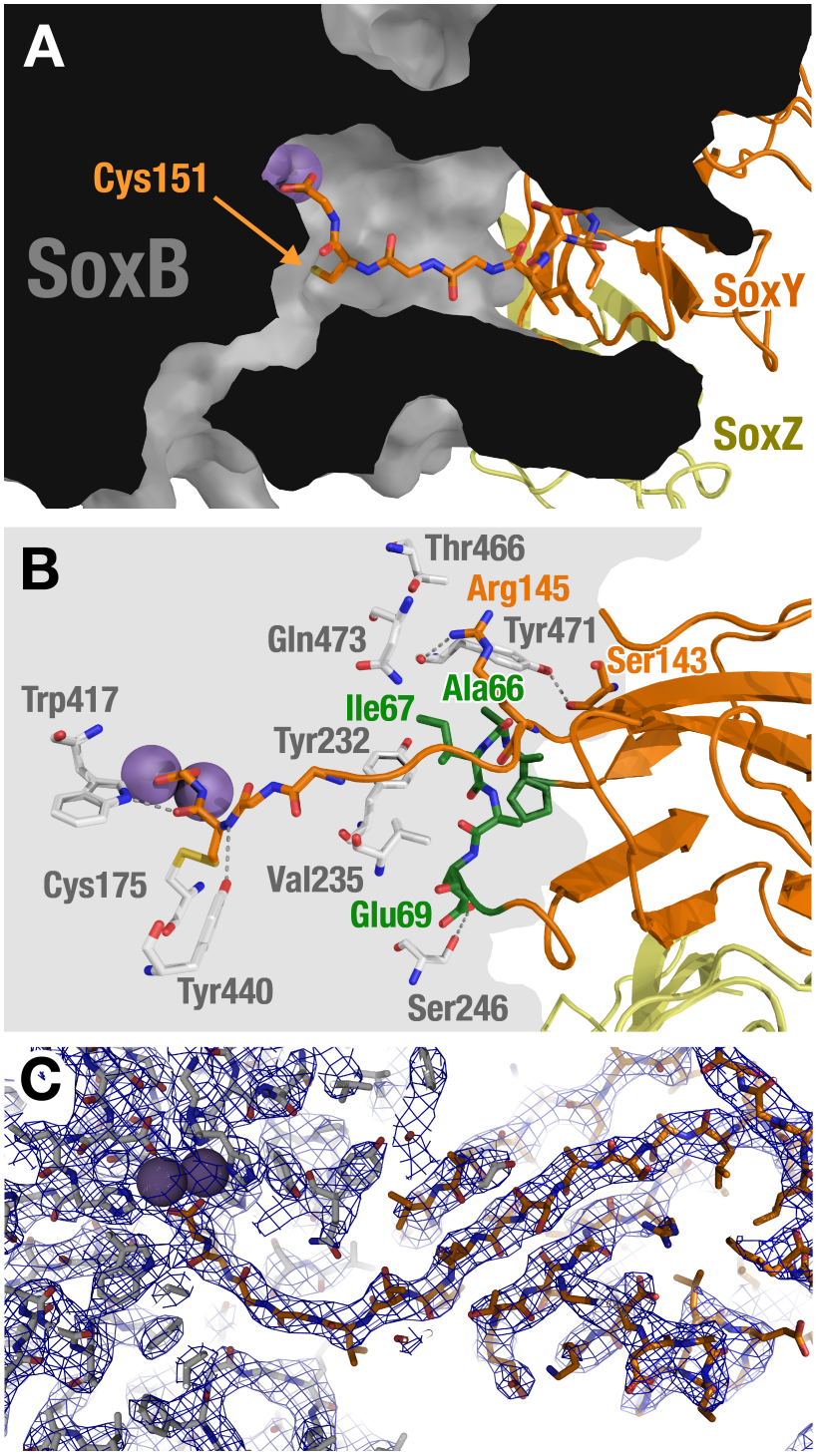

Figure 4.17 - SoxB-SoxY interface

Panel A: SoxY carrier arm in SoxB active site channel. SoxB is shown in a surface representation (grey). SoxY and SoxZ shown as cartoons in orange and yellow, respectively. Manganese ions are shown as purple spheres. Panel B: Colour scheme as in panel A. The SoxY ^{65}PAIAES^{70} loop is highlighted in green. Panel C: Electron density contoured at 1.0 sigma for this region.

As expected, a disulfide bond between Cys151 of SoxY and the cysteine residue at position 175 of SoxB, mutated from tryptophan, was observed in the complex structure. The C-terminal carboxylate group of the SoxY carrier arm coordinated the Mn(II) ions in the active site in a similar manner to that of the sulfonate group of the thiosulfate molecule bound in the *Thermus thermophilus* SoxB co-crystal structure (PDBID: 2WDE) (Figure 4.17). In addition to its interactions with the SoxB active site tunnel, SoxY carrier arm residues Arg-145 and Ser-143 and ^{66}AIAES^{70} loop contact SoxB around the active site channel mouth (Figure 4.17B). Additional interactions mediated by water molecules may exist, however it was not possible to see these at this resolution.

Our collaborators confirmed the importance of the C-terminal carrier arm in the interaction between SoxB and SoxYZ, with its removal resulting in a dramatic increase from ~3 μM for the wild-type interaction to between 100 and 150 μM, as measured by ITC experiments (Table 4.15).

		K_D (μM) and ΔH (kcal/mol) measured for interaction with:			
Protein	**Activity (U/mg)**	SoxY(C151S)Z	SoxY(SSO_3)Z	SoxY(Ac)Z	SoxY(Am)Z
SoxB (WT)	*1.8 ± 0.5*	*K_D = 3 ΔH = -3*	*K_D = 5 ΔH = +10*	*K_D = 0.7 ΔH = +10*	*K_D = 12 ΔH = -3*
SoxB (CO)	*1.2 ± 0.5*	*Not determined*	*Not determined*	*Not determined*	*Not determined*
SoxB (R416G)	*0.0*	*K_D = 17 ΔH = -4*	*K_D = 75 ΔH = -2*	*Not determined*	*Not determined*
SoxB (F205S)	*1.4 ± 0.2*	*No signal*	*Not determined*	*K_D > 150 ΔH = +ve*	*Not determined*
SoxB (D207R)	*1.6 ± 0.7*	*No signal*	*Not determined*	*Not determined*	*Not determined*
		SoxB (CO)			
SoxY(Ac)Z (ΔZ-loop)	*N/A*	*K_D = 0.4 ΔH = +8*			
SoxYZ (Δ carrier arm)	*N/A*	*K_D = 100-150 ΔH = -6*			

Table 4.15 – Collaborators' ITC binding experiments
SoxB enzyme activity measured as trithionate hydrolysis. WT = wild type, CO = codon optimised gene. Experiments performed by collaborators in the Berks laboratory (Biochemistry Department, University of Oxford).

4.7.2 Z-patch interactions

The interaction at the Z-patch involved a hydrophobic face on SoxZ, formed by Ile-6, Val-25 and Leu-74 of SoxZ, which was surrounded by two arginine residues, Arg-8 of SoxZ and Arg-117 of SoxY. SoxB interacted with this face through a surface loop that formed a complementary hydrophobic bulge surrounded by the negatively charged residues (Figure 4.18A). Leu-204 and Phe-205 of the SoxB hydrophobic bulge inserted into the hydrophobic pocket and Arg-8 of SoxZ formed both a salt-bridge with Asp-207 of SoxB and carbon-pi interactions with Phe205 of SoxB. A sequence alignment of *Thermus thermophilus* and *Paracoccus pantotrophus* SoxB showed conservation of both an aromatic residue at *Thermus thermophilus* position 205 and an acidic residue at position 207 (Figure 4.18B). A similar alignment of SoxZ showed conservation of a basic residue at *Thermus thermophilus* position 8 (Figure 4.18C). Our collaborators confirmed the importance of SoxB residues 205 and 207 with mutagenesis and ITC binding experiments. Both the SoxB F205S and D207R mutants abolished binding to SoxY(C151S)Z (Table 4.15). They also showed that the K_D of the SoxB interaction with SoxY(Ac)Z, a species in which the cysteine residue of the carrier arm was carboxymethlated, increased from ~0.7 μM with wild-type SoxB to > 150 μM with the F205S SoxB mutant (Table 4.15). This indicated that SoxYZ binding to SoxB occurs irrespective of the covalently attached species at the end of the carrier arm.

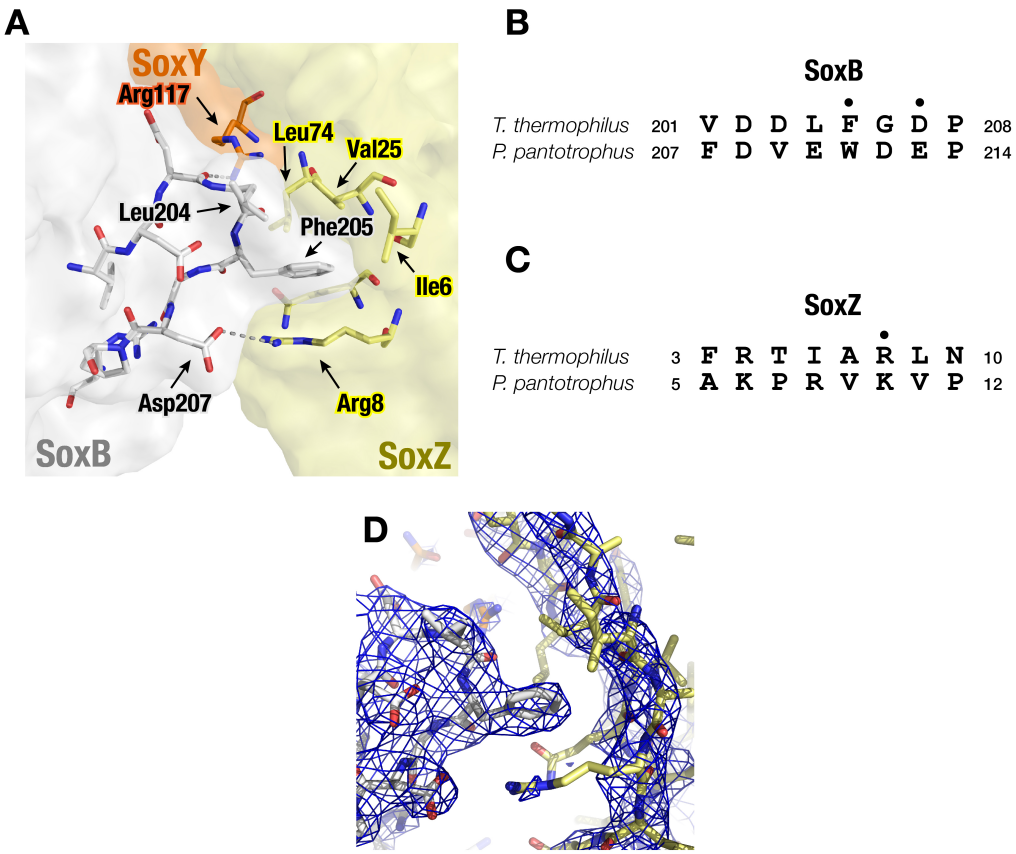

Figure 4.18 - SoxB-SoxZ interface
Panel A: Proteins shown as surface representations with residues involved at the interface shown as sticks. SoxB is coloured in grey, SoxY in orange and SoxZ in yellow. Panel B: Sequence alignment showing conservation of an aromatic residue and acidic residue at Thermus thermophilus positions 205 and 207, respectively. Panel C: Sequence alignment showing conservation of basic residue at Thermus thermophilus position 8. Panel D: Electron density contoured at 1.0 sigma for this region. Density for SoxZ Arg-8 becomes visible when the map is contoured at a lower level.

The electrostatic potential of the Z-patch of *Thermus thermophilus* SoxYZ *and* SoxB in the new complex structure was calculated using APBS (Dolinsky et al., 2004) in PyMOL (Schrödinger, LLC) (Figure 4.19). The Z-patch of SoxYZ was positively charged, with SoxB exhibiting a complementary, negatively charged surface. This was likely important for the interaction as the residues responsible for the electrostatic potential were conserved.

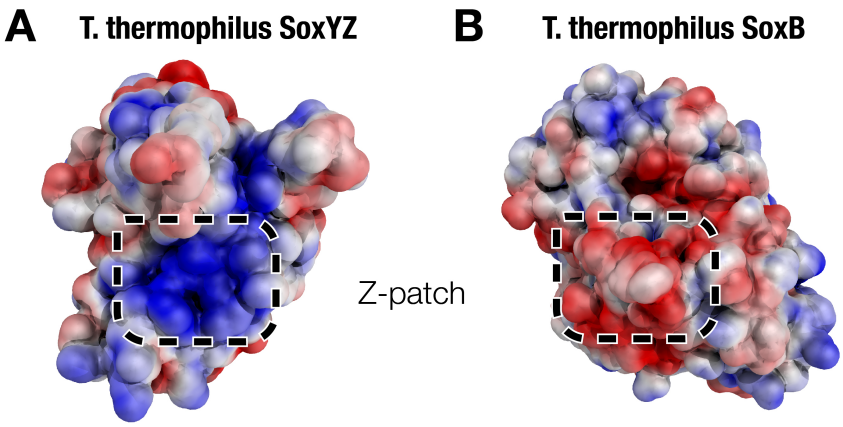

Figure 4.19 - Electrostatic analysis of Z-patch
Proteins displayed as surfaces with colours representing electrostatic potential as calculated by APBS (Dolinsky et al., 2004).

4.7.3 Conformational changes in SoxB mobile loop

The previously determined x-ray structures of *Thermus thermophilus* SoxB in absence of ligand (PDBID: 2WDF) and in complex with thiosulfate (PDBID: 2WDE), a substrate analogue, showed that binding of a substrate analogue resulted in the widening of the active site channel due to the movement of a loop formed by residues 463-478 (Figure 4.20A). The SoxB structure in absence of a ligand showed that the mobile loop was tethered close to the active site by a salt bridge between Asp476 and the catalytically essential Arg416. In contrast, the thiosulfate-bound structure showed Arg416 coordinating the thiosulfate molecule instead of forming a salt bridge with Asp476, which instead formed a salt bridge with Arg385 (Figure 4.20B).

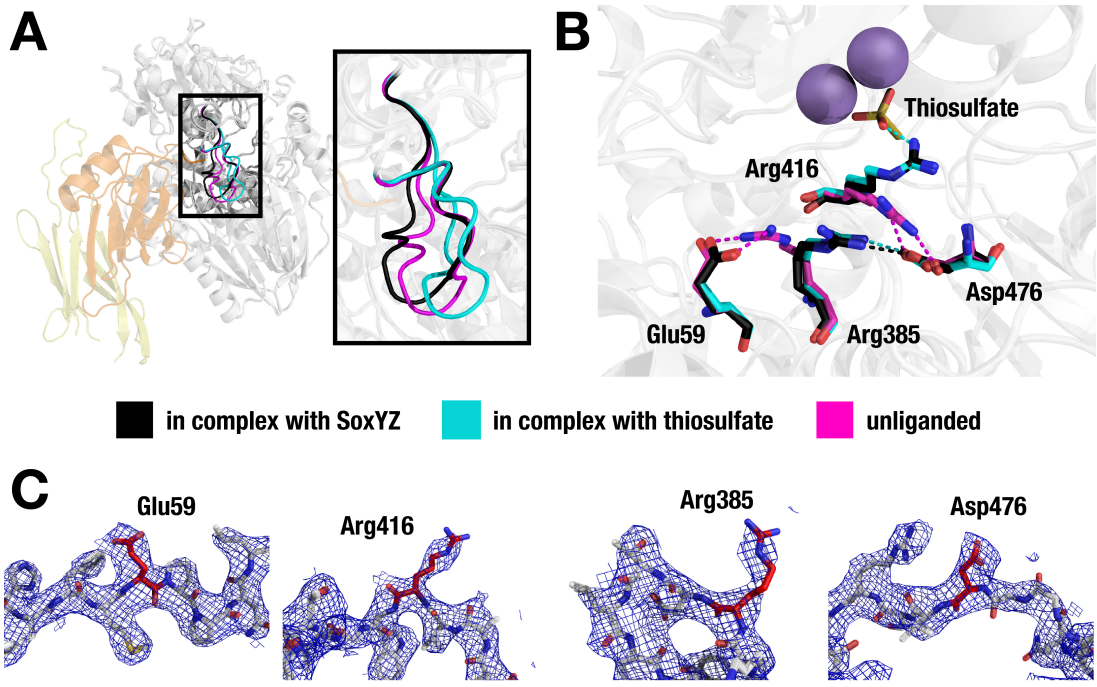

Figure 4.20 - Conformational changes in SoxB mobile loop
Thermus thermophilus SoxB in complex with SoxZY (black), thiosulfate (cyan) and in the absence of ligand (magenta). Panel A: Mobile loop displayed in cartoon representation. Panel B: Important SoxB residues and thiosulfate shown in cartoon representation. Manganese ions shown as purple spheres. Panel C: Electron density contoured at 1.0 sigma for important SoxB residues in the SoxBYZ complex structure.

The conformational change of this mobile loop induced by SoxYZ binding is less extreme and in a different direction to that caused by thiosulfate binding. This change is likely required to prevent steric clashes that would exist in the ligand-free conformation (Figure 4.21).

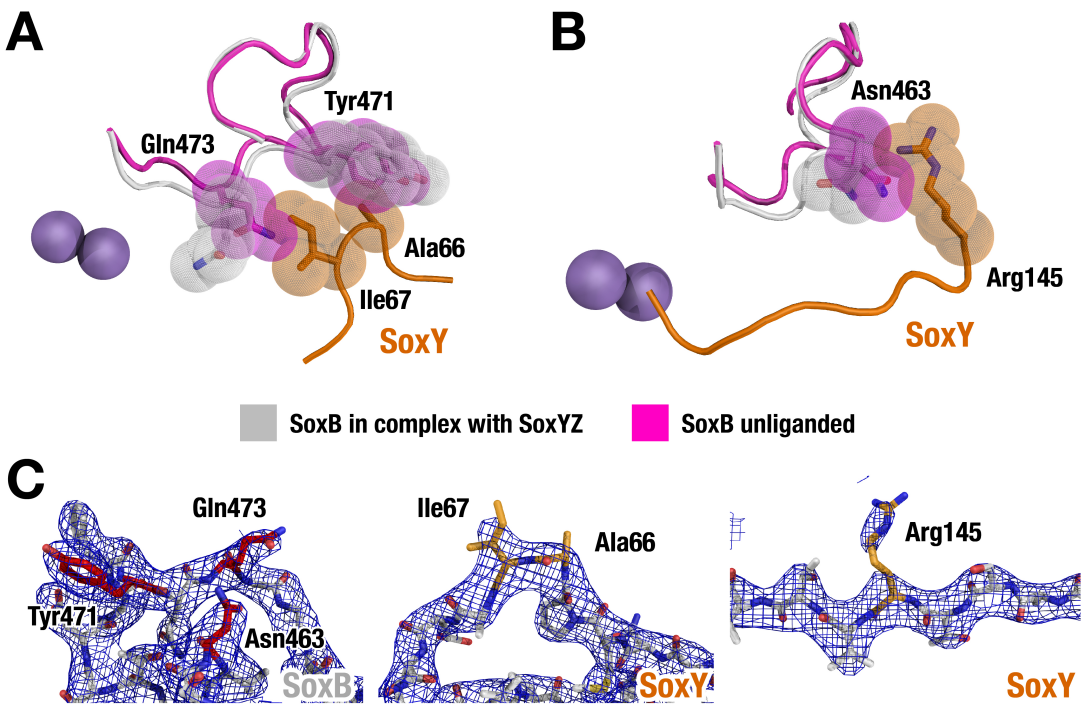

Figure 4.21 - SoxB side chain conformational alterations
Changes in loop conformations relieve steric clashes with SoxY. Panels A & B: SoxB from the SoxBYZ structure is coloured in white, SoxB in absence of ligand in magenta and SoxY from the SoxBYZ complex in orange. Panel C: Electron density contoured at 1.0 sigma for highlighted residues.

Formation of the SoxBYZ complex also disrupts the interaction between Arg416 and Asp476 (Figure 4.20B). Arg416 does not form any new interactions, suggesting that breaking of the Arg416-Asp476 interaction is driven by the conformational change in the loop. This would allow Arg416 to ligate substrate bound to the end of the SoxY carrier arm and stabilise the interaction.

4.7.4 Models of manganese ion coordination at SoxB active site

The thiosulfate co-crystal structure of *Thermus thermophilus* SoxB (PDBID: 2WDE) suggested that the manganese ions in the SoxB active site must be coordinated by the sulfonate group of the substrate for catalysis to occur (Sauvé et al., 2009). The C-terminal carboxylate group of the SoxY carrier arm could compete with the substrate for binding to these ions. Our collaborators generated models for the

two possible coordination states of the manganese ions. They first mutated the SoxB residue 175 back to tryptophan and then conjugated a S-thiosulfonate group to the cysteine residue of the SoxY carrier arm in our new complex structure. In a second model they manually positioned the S-thiosulfonate group into a similar position that the thiosulfate molecule occupied in the thiosulfate SoxB co-crystal structure. These two models were then subjected to energy minimisation and position restrained molecular dynamics equilibration.

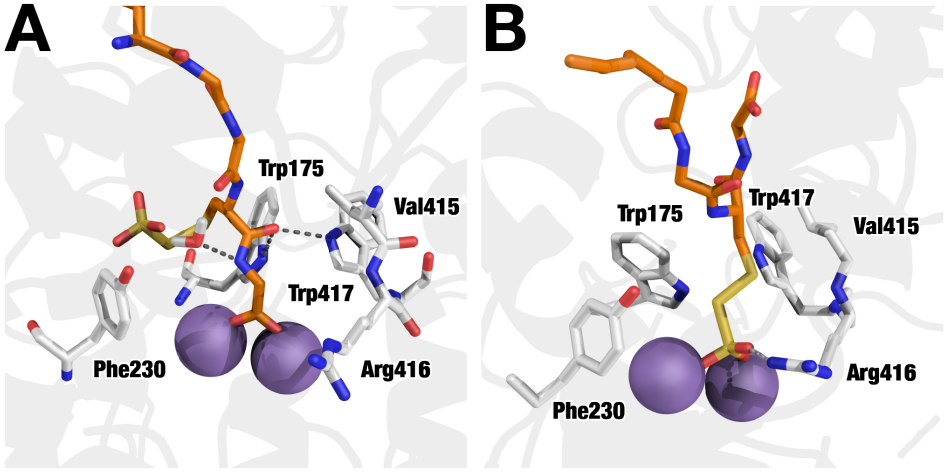

Figure 4.22 - Models of manganese ion coordination
Protein residues displayed at sticks with SoxB in white and SoxY in orange. Manganese ions displayed as purple spheres. Panel A: C-terminal carboxylate coordination of manganese ions. Panel B: S-thiosulfonate coordination of manganese ions.

Inspection of the models suggested that in the case of coordination of the manganese ions by the C-terminal carboxylate group (Figure 4.22A), the Cys-Gly peptide forms hydrogen bonds with a ring of hydrophobic residues in close proximity to the manganese ions. In the case of coordination by the S-thiosulfonate substrate group (Figure 4.22B), the hydrophobic side chain of the SoxY cysteine residue forms predominantly hydrophobic interactions with the ring residues and the catalytically important Arg416 SoxB residue interacts with a terminal oxygen atom of the substrate. This observation would explain the enthalpically-driven interactions of SoxB with

SoxYZ species that do not feature adducts capable of coordinating the manganese ions, as well as the entropically-driven interactions with SoxYZ species that feature adducts capable of coordination of the manganese ions (Table 4.15). The interaction between SoxB and SoxY(SSO_3)Z switches from entropically-driven in the case of wild-type SoxB protein, to enthalpically-driven in the case of the SoxB R416G mutant protein. This suggested that this arginine residue was required for ligation of the manganese ions by the S-thiosulfonate group, helping to position it at the active site.

4.8 Discussion

The new SoxBYZ complex structure revealed, for the first time, the molecular details of an interaction between the SoxYZ carrier protein and a partner enzyme of the sox pathway. The structure showed that SoxB interacted with SoxYZ through two separate interaction surfaces, termed the Y- and Z-patches. The Y-patch accounted for ~80% of the total surface area buried by the overall interaction, with the interaction at this point mediated by a sparse network of hydrogen bonds. It was in this region that the SoxY carrier arm entered the SoxB active site channel. The Z-patch comprised ~20% of the total buried surface area and was located away from the SoxB active site channel and mouth. This interaction involved a SoxB surface loop and a surface on SoxZ that featured compatible hydrophobic and electrostatic properties. The presence of the two interacting surfaces is common for proteins that engage in promiscuous interactions (Meenan et al., 2010; Schreiber and Keating, 2011). It would allow the carrier arm of SoxY to make limited specific contacts with the SoxB active site channel, with the Z-patch compensating for the limited interactions at the enzyme active site, tunnel and mouth. This, in addition to its flexibility, would allow the carrier arm to be accommodated in the active site tunnels of other partner enzymes. It also explains how a mutation in the Z-patch, away from the active site of a progenitor

enzyme, could have been sufficient to allow interaction with the carrier protein. The Z-loop previously predicted to be important for the interaction did not make any contact with SoxB (Sauvé et al., 2007).

The presented crystal structure was likely representative of the native, physiological structure because a similar organisation of protein components was observed in all copies of the complex in the higher resolution structure and in the three placed copies in the lower-resolution dataset. Furthermore, our collaborators were able to experimentally confirm the importance of both Y- and Z-patches with mutagenesis and ITC binding experiments confirming that interactions at both regions were required for SoxBYZ complex formation (Table 4.15).

CONSURF (Ashkenazy et al., 2010) sequence conservation analysis of *Thermus thermophilus SoxB and SoxYZ* (Ashkenazy et al., 2010) revealed that the most conserved surface residues corresponded to those involved in the interactions at the Y- and Z-patches (Figure 4.15). This sequence conservation suggested that these sites could also be important in the interactions of SoxYZ with other partner enzymes. CONSURF (Ashkenazy et al., 2010) analysis of the available crystal structures of *Paracoccus pantotrophus* SoxAX (PDBID: 2C1D) and SoxCD (PDBID: 2XTS) revealed that the surfaces surrounding both enzyme active sites were also highly conserved in addition to a region that corresponded to the Z-patch of SoxB. Interestingly, these putative Z-patches were formed by protruding loops that featured surface-exposed aromatic side chains surrounded by acidic residues, highly similar to the Z-patch of *Thermus thermophilus* SoxB and suggesting a similar mechanism of binding (Figure 4.23).

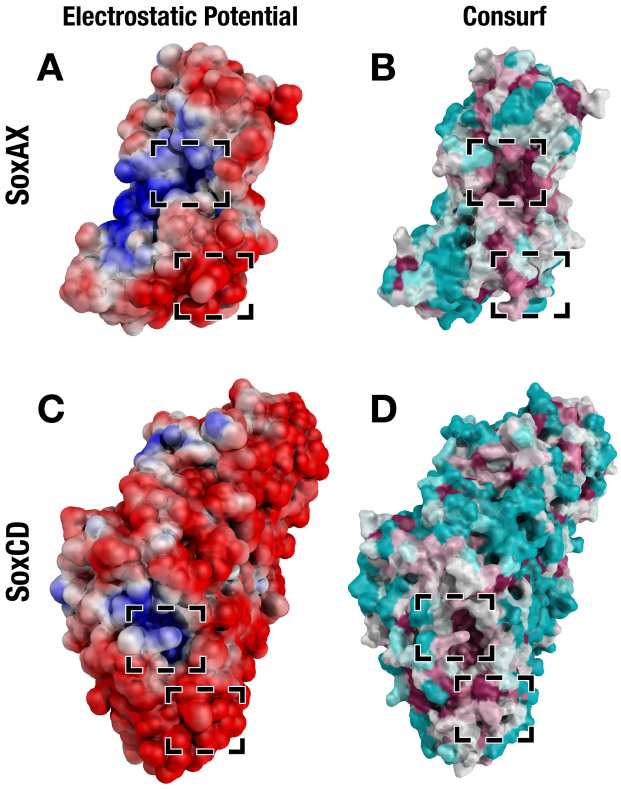

Figure 4.23 - SoxYZ interacting surfaces of partner enzymes
SoxYZ partner proteins of Paracoccus pantotrophus shown as surface representations. The crystal structures of SoxAX and Panels A & C: Surfaces are coloured to illustrate the electrostatic potential as calculated by APBS (Dolinsky et al., 2004) in PyMOL (Schrödinger, LLC). Panels B & D: Surfaces coloured to show sequence conservation with the same colour scheme as in Figure 15.

When considered with the previously available structures of *Thermus thermophilus* SoxB in the absence and presence of thiosulfate, the new structural data suggested a mechanism of induced-fit at the SoxB active site. This mechanism explained our collaborators findings that SoxB selectivity for substrate-bound SoxYZ was kinetically determined with the dissociation rate for the substrate-bound form of SoxYZ significantly lower than the substrate-free form (Figure 4.3E & F). The ligand-free structure of SoxB showed the catalytically essential Arg416 residue tethered to a mobile SoxB loop. SoxYZ binding appeared to distort the conformation of this loop resulting in the release of Arg416 that was then free to bind substrate, cooperatively coupling the binding of SoxYZ to substrate binding at the active site. After catalysis

occurs, the substrate group would be lost and the mobile loop no longer stabilised, resulting in an accelerated rate of dissociation. The structures also suggested that thiosulfate, which would be present during the utilisation of the sox pathway, would not act as a significant competitive inhibitor. The binding of thiosulfate, just as the binding of substrate bound to SoxYZ, required a conformational change in the side chain of Arg416 and the mobile loop. In the absence of SoxYZ binding this would present a greater energetic barrier.

Finally, our collaborators modelling experiments based on the new SoxBYZ crystal structure suggested how the SoxB active site manganese ions could be coordinated by either the C-terminal carboxylate group or the S-thiosulfonate substrate group of the SoxY carrier arm. The modelling suggested that most energetically favourable conformation was that of S-thiosulfonate coordination.

4.8.1 Future work

In order to assess whether other partner proteins do indeed utilise the same mechanism of binding as seen between SoxB and SoxYZ, it would be highly informative to solve the crystal structures of additional complexes. In the absence of such structural information, it would be possible to utilise mutagenesis experiments to try to disrupt the interaction of SoxYZ with partner enzymes. The data presented in this chapter could be used as a starting point to predict residues in SoxAX and SoxCD that are involved in their interaction with SoxB.

Chapter 5: The structure of *Allochromatium vinosum* TsdA

5.1 Introduction

As discussed in the introduction of the previous chapter, the widely distributed Sox pathway allows some bacteria to completely oxidise thiosulfate ($S_2O_3^{2-}$) to sulfate (SO_4^{2-}), with the released electrons utilised for respiratory chain energy generation or carbon fixation (Frigaard and Dahl, 2009).

Allochromatium vinosum, a purple sulfur bacterium, possesses two distinct pathways for thiosulfate oxidation. Under alkaline conditions, thiosulfate is completely oxidised to sulfate. This process proceeds using the Sox pathway (Friedrich et al., 2001; Frigaard and Dahl, 2009; Gregersen et al., 2011) but the absence of SoxCD in *Allochromatium vinosum* leads to the formation of sulfur globules. Further oxidation of these sulfur globules requires genes encoded in the dissimilatory sulfite reductase (dsr) locus (Dahl et al., 2005). In contrast, under acidic conditions, thiosulfate is oxidised to tetrathionate ($S_4O_6^{2-}$) in a reaction catalysed by thiosulfate dehydrogenase (TsdA) (Denkmann et al., 2012; Hensen et al., 2006).

Allochromatium vinosum TsdA is a periplasmic 27.2 kDa diheme c-type cytochrome that catalyses the oxidative conjugation of two thiosulfate molecules, forming a single molecule of tetrathionate and releasing two electrons (Hensen et al., 2006). The normal catalytic reaction can be reversed in the presence of a suitable electron donor, reducing tetrathionate to two molecules of thiosulfate, however the rate of this reverse reaction is much slower than of the forward (Liu et al., 2013).

The N-terminal domain of *Allochromatium vinosum* TsdA has been proposed to be related to the SoxA catalytic domain (Denkmann et al., 2012). Although the overall sequence identity between the SoxA catalytic domain and the TsdA N-terminal domain is extremely low, both share a conserved Cys-X_2-Cys-His heme c binding

motif. In addition, both feature a conserved arginine residue (TsdA Arg119) and a conserved catalytic cysteine residue (TsdA Cys123) at equivalent positions relative to the heme-binding motif (Figure 5.1). It is likely that the reaction catalysed by TsdA is analogous to the proposed reaction of SoxAX (Bamford et al., 2002), with SoxYZ replacing a second molecule of thiosulfate (Figure 5.2). As there is currently no evidence for the proposed cysteine-mediated reaction of SoxAX due to the protein complex not being very amenable to mechanistic experimentation, TsdA could potentially be used as a model to understand SoxAX catalysis.

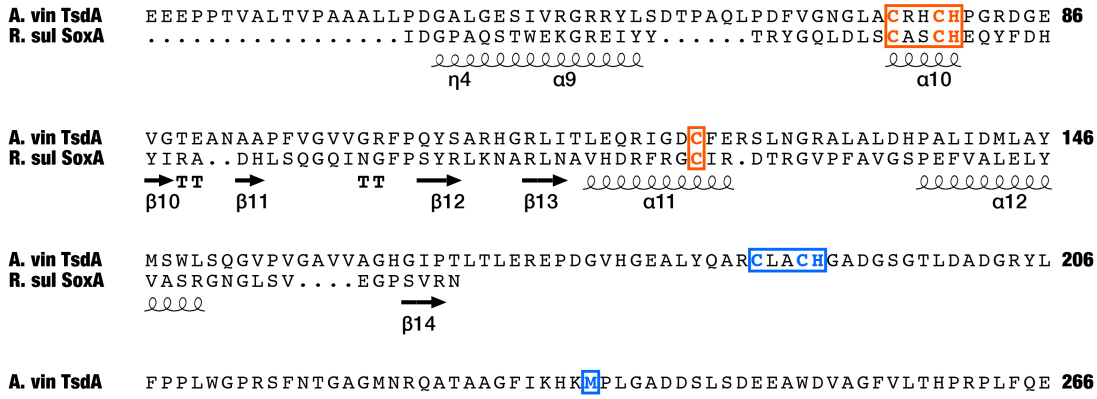

Figure 5.1 - Sequence alignment showing homology between TsdA and SoxA
Sequence alignment of the N-terminal domain of Allochromatium vinosum TsdA and the C-terminal domain of Rhodovulum sulfidophilum SoxA calculated by Clustal Omega (Sievers et al., 2011). Heme binding residues in the catalytic domain are coloured orange and those in the electron transfer domain in blue.

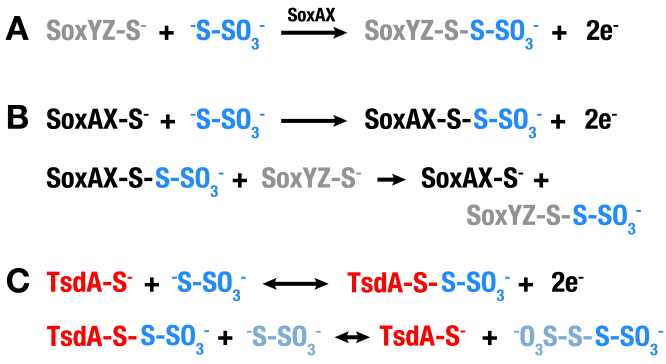

Figure 5.2 - Comparison of the proposed reactions catalysed by SoxAX and TsdA
Panel A: Overall SoxAX reaction. Panel B: The two proposed reactions of SoxAX. Panel C: Analogous reaction of TsdA.

The catalytic domain of SoxA contains a heme that is coordinated by a modified cysteine residue that features an additional terminal sulfur atom. A number of other c-type cytochromes exhibit cysteine coordination (Alric et al., 2004; Grein et al., 2010; Suga et al., 2013; Ukita et al., 2010; Van Driessche et al., 2006), but SoxA is unusual, featuring a cysteine S-sulfane as the coordinating residue. The modification of this cysteine residue suggests that it may be involved in the catalytic mechanism of SoxA.

A mechanism for the catalysis by SoxAX has been proposed based on the similarity of its active site with rhodanese, a sulfur transferase (Bordo et al., 2000; Gliubich et al., 1996). Both enzyme active sites feature a cysteine persulfide residue that coordinates the heme, in addition to an adjacent conserved arginine residue and a strong positively charged electrostatic potential (Bamford et al., 2002). SoxAX catalysis is proposed to occur in two steps. Firstly, a thiosulfate molecule becomes conjugated to the active site cysteine residue in an oxidative reaction. The subsequent final step involves a thiol-disulfide exchange reaction with the S-thiosulfonate adduct and a cysteine residue of the SoxZY carrier arm (Bamford et al., 2002). It is assumed that the observed extra sulfur atom attached to the cysteine residue is a result of an off-pathway reaction.

Sequence analysis identified Cys123 of *Allochromatium vinosum* TsdA as the equivalent catalytic heme ligand in SoxA (Figure 5.1) (Denkmann et al., 2012), with mutagenesis in TsdA showing that it is essential for catalysis. However, it has been reported that SoxAX is still able to perform catalysis when the cysteine heme ligand is mutated to methionine. In this case, a copper ion-based mechanism has been proposed due to EPR experiments indicating copper binding near to the active site heme (Kappler et al., 2008; Kilmartin et al., 2011).

Our collaborators in the Berks laboratory (Biochemistry Department, University of Oxford) have been investigating the catalytic mechanism of *Allochromatium vinosum*

TsdA through a number of biochemical experiments. Electrospray ionisation mass spectrometry (ESI-MS) analysis of the purified protein showed a predominant peak of 28,228 Da (only two Daltons more than the predicted mass of the native protein with two bound c-type hemes) strongly suggesting that the active site residue, Cys123, was in an unmodified form (Figure 5.3A). This is in contrast to native and recombinant SoxAX proteins that are reported to exhibit S-sulfane modification of the active site cysteine residue.

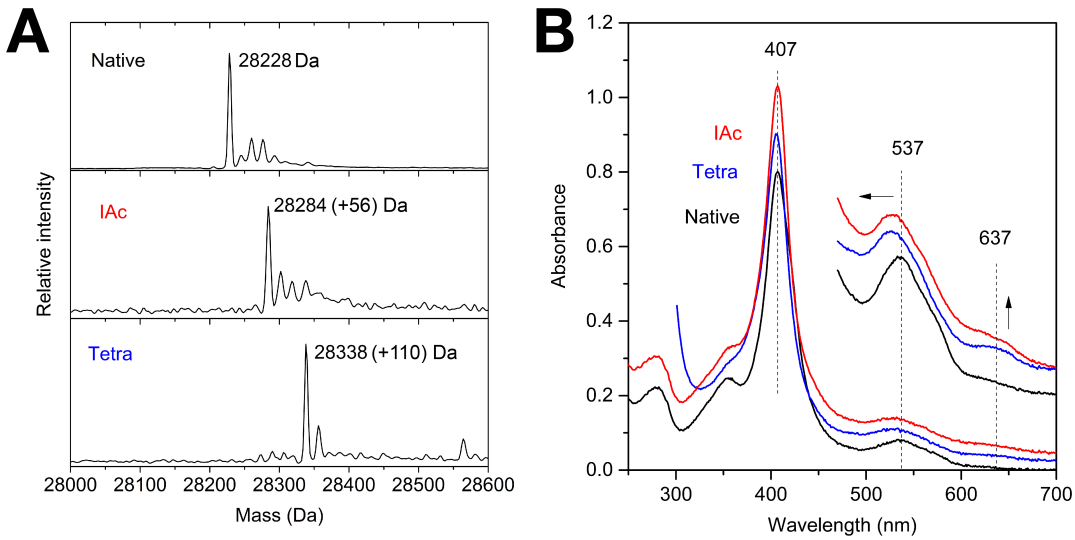

Figure 5.3 - Modification of Cys123 at pH 8.0
Panel A: ESI-MS analysis of native protein, 1 hour 20mM sodium iodoacetate incubation and 1 hour 50mM tetrathionate incubation. Panel B: UV-visible spectra of 4 μM TsdA (main spectra) or 20 μM TsdA (inset). Expected mass increase is 57 Da for carboxymethylation and 112 Da for thiosulfonation. Experiments were carried out by collaborators in the Berks Group (Biochemistry Department, University of Oxford).

The proposed reaction mechanism for TsdA (Figure 5.2) requires Cys123 to dissociate from the heme iron atom. It also suggests that the reverse reaction would involve the active site cysteine residue taking part in a nucleophilic attack on a molecule of tetrathionate. If this was the case, Cys123 should be accessible to electrophilic thiolate-reactive reagents such as iodoacetate, an alkylating agent. Our collaborators' ESI-MS analysis of a TsdA sample incubated with iodoacetate showed a mass increase of 56 Da, consistent with S-carboxymethylation of Cys123 (Figure

5.3A). Analysis of the UV/visible spectra showed broadening of the Soret band (407nm) to a lower wavelength, a shift of the oxidised heme α and β bands at 537nm to a lower wavelength and an increase in intensity at 637nm. These three changes are characteristic of an increase in the content of ferric high spin c-type heme, suggesting that S-carboxymethylation of Cys123 either prevents coordination of the heme iron or acts as a weak field ligand (Figure 5.3B).

Another prediction based on the proposed reaction mechanism is that the S-thiosulfonate intermediate would accumulate via the reverse reaction in the presence of tetrathionate and absence of an electron donor. Our collaborators incubated TsdA with tetrathionate either at the optimum catalytic, pH 4.5 (Figure 5.4), or at pH 8.0, where there was very little catalytic activity (Figure 5.3). Samples were subsequently subjected to ESI-MS analysis (Figure 5.3A and Figure 5.4C). Complete S-thiosulfonation was achieved within the mixing time at pH 4.5 and within 1 hour at pH 8.0, as expected due to the respective pH conditions. Both samples showed a mass increase of 112 Da, consistent with S-thiosulfonate formation. Inspection of the UV/visible spectra showed that at pH 8.0, carboxymethylation and thiosulfonation of Cys123 exhibited similar changes to heme ligation (Figure 5.3B). However, at pH 4.5, the intensity and sharpness of the heme α and β bands (553nm and 524nm) indicate the reduction of heme groups, likely due to oxidation of thiosulfate released in the tetrathionate modification reaction (Figure 5.4A). They also observed that formation of the S-thiosulfonate adduct prevented subsequent alkylation by iodoacetate.

Treatment of TsdA with thiosulfate allows the forward reaction to be investigated. Incubation with a 1:1 molar ratio enzyme to thiosulfate yielded a single ESI-MS mass peak consistent with complete thiosulfonation (Figure 5.4C) and inspection of the visible spectra indicated a large proportion of reduced heme groups (Figure 5.4B). However, addition of a molar excess of thiosulfate did not remove the S-thiosulfonate group as predicted and TsdA instead remained fully thiosulfonated

(Figure 5.4C) with some further heme reduction (Figure 5.4B). This result suggested that either the second step of the reaction could not take place until the heme was reoxidised or that the S-thiosulfonate adduct was not a genuine catalytic intermediate.

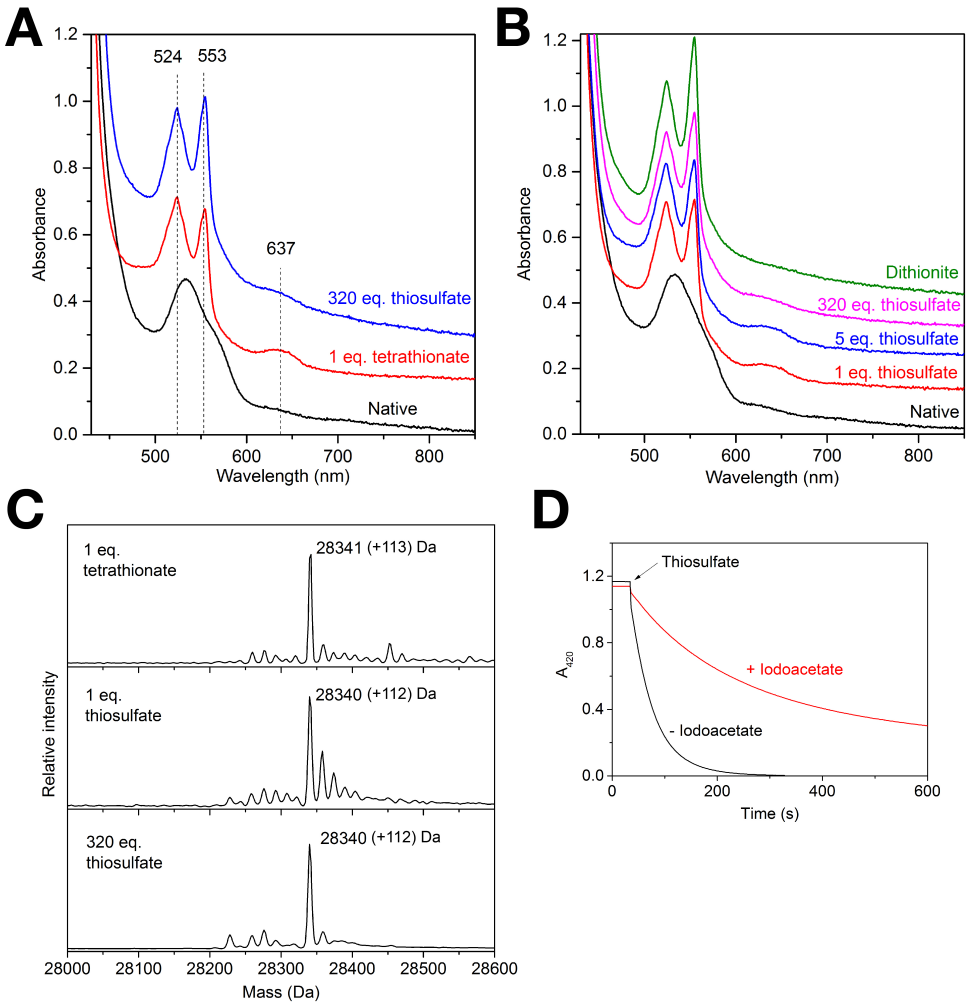

Figure 5.4 - Reactivity of TsdA with thiosulfate and tetrathionate at pH 4.5
Panels A – C: 25 μM TsdA native, 3 minute incubation with thiosulfate or tetrathionate. Panels A & B: Visible spectra of samples, with dithionite added in B until no further spectral changes were seen. Panel C: ESI-MS analysis of samples. Expected mass increase of 112 Da for thiosulfonation. Panel D: Enzyme turnover results in thiosulfonate group removal from Cys123. 1nM thiosulfonated TsdA, 8mM thiosulfate and 1mM potassium ferricyanide in presence and absence of 10mM iodoacetate. Oxidation of ferricyanide was monitored at 420nm. Panel E: Instability of S-thiosulfonate group. Thiosulfonated TsdA was generated by incubation with tetrathionate at pH 8.0 for 1 hour at 37°C in the presence of 20mM sodium iodoacetate to capture the formation of any thiol groups. Expected mass of thiosulfonated TsdA is 28340 Da and S-carboxylmethylated cysteine persulfide is 28317 Da. Experiments were carried out by collaborators in the Berks Group (Biochemistry Department, University of Oxford).

To investigate whether the S-thiosulfonate adduct was compatible with enzyme turnover, our collaborators generated the Cys123-S-thiosulfonate intermediate by incubation with tetrathionate, before removing excess tetrathionate by desalting. They then monitored the oxidation of ferricyanide by UV/visible absorption at 420nm to measure TsdA enzyme activity. This intermediate showed no reduction in activity compared to a control sample, whereas incubation with iodoacetate abolished activity. This showed the sensitivity of the reaction to modification of Cys123 while suggesting the S-thiosulfonate adduct is compatible with the proposed mechanism.

Furthermore, our collaborators compared TsdA enzymatic activity in the presence and absence of iodoacetate (Figure 5.4D). As mentioned, thiosulfonation of Cys123 blocks alkylation by iodoacetate, therefore in order for alkylation to occur, the S-thiosulfonate group must be removed from Cys123. This is indeed the case, further implicating the S-thiosulfonate adduct as an intermediate in the TsdA catalytic reaction cycle.

5.2 Aims

The aim of this chapter is to solve the x-ray crystal structure of *Allochromatium vinosum* TsdA in both its native and S-thiosulfonated state. This will identify important active site residues involved in substrate binding and may help further inform the catalytic mechanism.

5.3 Native crystallisation, data collection and processing

5.3.1 Crystallisation of native TsdA

Collaborators in the Berks laboratory (Biochemistry, Oxford) provided us with protein material to solve the x-ray crystal structure of *Allochromatium vinosum* TsdA. A structure would provide insight into the catalytic mechanism and complement our collaborators' biochemical analysis.

Commercial crystallisation screens were purchased from Molecular Dimensions and a Hydra 96 (Alpha Biotech Ltd) robot was used to transfer 70 µL of each condition into the reservoirs of MRC 2-drop crystallisation plates. The sitting-drop vapour-diffusion approach to protein crystallisation was chosen and was set up with a Mosquito LCP robot (TTP Labtech). Plates were sealed using StarSeal Advanced Polyolefin Film (STARLAB) and incubated at 20 °C.

Crystals grew overnight in condition G5 of the Stura & Macrosol screen (Molecular Dimensions). This drop contained 70% protein and 30% mother liquor. Protein solution contained TsdA at a concentration of ~350 µM and the mother liquor was composed of 12% PEG 8K and 0.2M ammonium sulfate.

5.3.2 Data collection

The crystal was cyroprotected with mother liquor plus 20% (v/v) ethylene glycol and flash frozen in liquid nitrogen. A dataset was collected at Diamond Light Source's i04-1 beamline. In the possible scenario where structure solution was not possible by molecular replacement, 720 degrees of data were collected in an attempt to obtain a dataset with high multiplicity that could be used to attempt experimental phasing with the iron atoms present in the heme groups of TsdA (Table 5.1).

Data Collection	Native dataset
Diffraction source	Diamond Light Source, i04-1
Detector	Pilatus 6M
Energy (eV)	13477
Wavelength (Å)	0.92
Temperature (K)	100
Detector resolution (Å)	1.80
Exposure per image (s)	0.2
Ω start (º)	0
Ω end (º)	720
Ω osc (º)	0.20
Number of images	3600

Table 5.1 - Initial data collection

5.3.3 Data processing

Data were processed using XIA2 (Winter et al., 2013) running in 3da mode, indexing a triclinic cell (a = 48.52 Å, b = 53.78 Å, c = 55.43 Å, α = 93.55°, β = 112.40°, γ = 105.40 °) in the P1 spacegroup with final data to a resolution of 1.60 Å (Table 5.2).

	Native dataset
Resolution range (Å)	17.22-1.60 (1.64-1.60)
Space group	P1
Unit cell dimensions	
a, b, c (Å)	48.52, 53.78, 55.43
α, β, γ (º)	93.55, 112.40, 105.40
Total reflections	436149 (26863)
Unique reflections	59636 (3741)
Multiplicity	7.3 (7.2)
Completeness (%)	92.1 (78.4)
Mean I/sigma(I)	15.1 (2.6)
R_{merge} (%)	8.5 (69.2)
R_{pim} (%)	3.4 (27.6)
CC(1/2)	99.7 (87.4)

Table 5.2 - Data processing

Data processing statistics indicated good overall data quality and a resolution of 1.6 Å (Table 5.2 and Figure 5.5). The completeness of the data to 1.6 Å never reached 100% because the detector was set to 1.8 Å, an underestimation of the resolution to which the crystals diffracted.

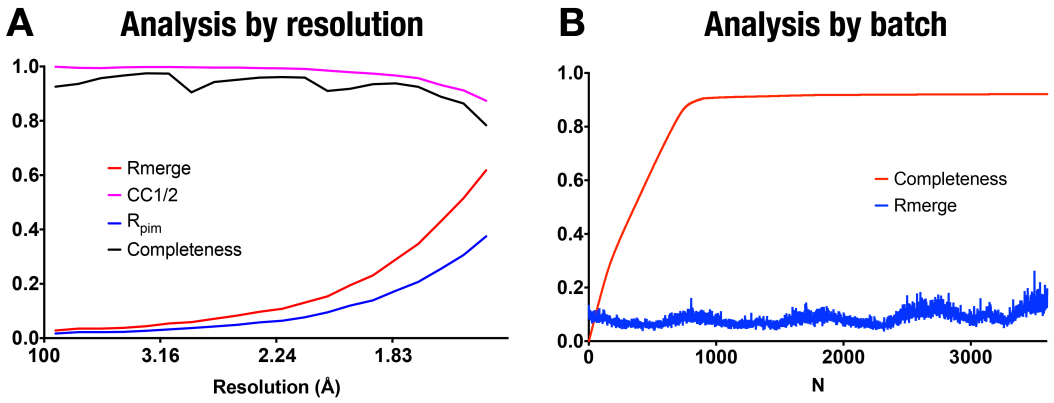

Figure 5.5 - Data processing analysis
Panel A: Analysis by resolution. Panel B: Analysis by batch.

5.4 Native structure determination & refinement

No evidence of twinning or translational non-crystallographic symmetry was indicated by PHENIX.XTRIAGE (Adams et al., 2010). Cell content analysis by MATTHEWS (Kantardjieff and Rupp, 2003) suggested a 99% probability of two molecules of TsdA per asymmetric unit (Table 5.3). Analysis of self-rotation function peaks using MOLREP (Vagin and Teplyakov, 2010) revealed a single non-origin peak (53% of the origin) with Chi = 180 (Figure 5.6 and Table 5.4). This was consistent with a 2-fold non-crystallographic symmetry axis that would describe the relationship between two molecules of TsdA in the asymmetric unit.

Nmol / asym	Matthews coefficient	% solvent	P(1.60)	P(tot)
1	4.91	74.97	0.00	0.01
2	2.46	49.93	0.99	0.99
3	1.64	24.90	0.01	0.00

Table 5.3 - Cell content analysis
MATTHEWS (Kantardjieff and Rupp, 2003) calculation assuming a molecular weight of 28226Da.

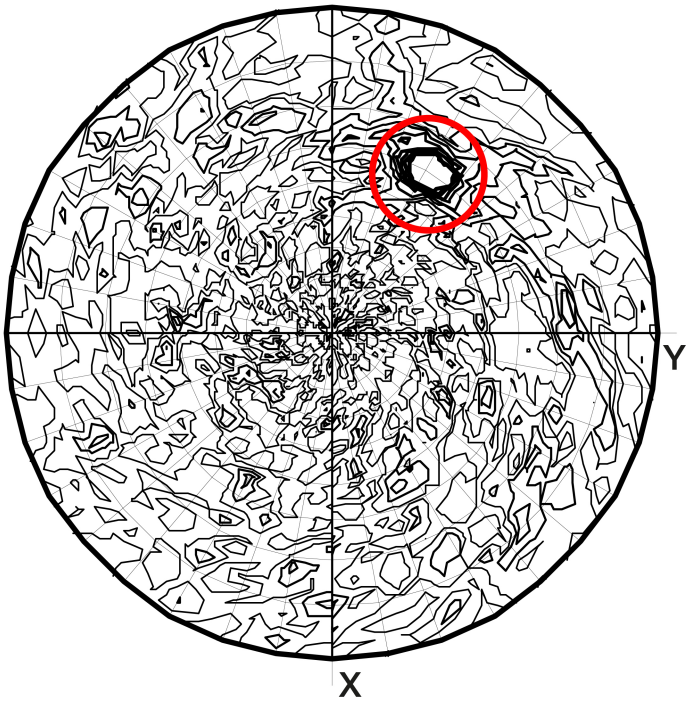

Figure 5.6 - Self-rotation function
Self-rotation function peaks calculated by MOLREP (Vagin and Teplyakov, 2010). A single peak above 30% of the origin is highlighted with a red circle.

θ	Φ	χ	P(i)/P(0)
0	0	0	1
119.21	**-31.46**	**180**	**0.53**

Table 5.4 - Top self-rotation function peaks
List of self-rotation function peaks above 30% of the origin. Related to Figure 5.6.

The N-terminal domain of TsdA had been identified as homologous to the C-terminal catalytic domain of SoxA. Specifically, the N-terminal domain of *Allochromatium vinosum* TsdA and the C-terminal domain of *Rhodovulum sulfidophilum* SoxA (PDBID: 1H32) have a sequence identity of 25%. This made it a candidate for an initial molecular replacement model for this domain. The remaining C-terminal sequence of TsdA was analysed using the PSI-BLAST server (EBI) identifying a region of *Pseudoalteromonas haloplanktis* nitrite reductase (PDBID: 2ZOO) as a potential molecular replacement model with 28% sequence identity.

Molecular replacement was carried out using MOLREP (Vagin and Teplyakov, 2010). Two copies of a search model derived from the nitrite reductase structure were placed with contrast scores of 7.90 and 3.63. Both contrast scores were above 2.5, which was indicative of a correct molecular replacement solution. With these two models placed, a search for two copies of a search model derived from SoxA was only able to place a single copy, albeit with a contrast score of 8.31 (Table 5.5).

Copy	TF/sig	Contrast
1st 2ZOO model	1.0	7.90
2nd 2ZOO model	3.50	3.63
1st 1H32 model	4.30	8.31

Table 5.5 - Initial molecular replacement scores

Although only one copy of the SoxA model was placed, this placement relative to a placed copy of a model for the C-terminal domain offered a potential search model for the whole protein. Having already placed one of these copies, a second copy of the whole protein was located using MOLREP (Vagin and Teplyakov, 2010) with a contrast score of 9.85. The 2-fold non-crystallographic axis that relates these two placed copies explains the non-origin peak in the self-rotation function (Figure 5.6 & Table 5.4). ARP/wARP (Langer et al., 2008) was successfully used to auto-build 463 of the 540 residues of the TsdA structure with this molecular replacement solution used as a starting model. Subsequent rounds of manual model building and refinement were carried out using COOT (Emsley et al., 2010) and REFMAC (Murshudov et al., 1997) utilising automatically generated local NCS restraints and included modelling two heme groups into the residual electron density. The structure was refined to an R_{work} of 16.95% and R_{free} of 19.51% (Table 5.6).

Refinement statistics	Native TsdA
Resolution range (Å)	17.22-1.60
Number of reflections (work / test)	56619 / 3010
Number of atoms	3902
Protein residues	466
Mean B-factors (protein, glycans, ligands, waters)	18.56
R_{work} (%)	16.95
R_{free} (%)	19.51
R.M.S.D. from ideal values (bonds, Å)	0.023
R.M.S.D. from ideal values (angles, °)	2.044
Ramachandran favoured (%)	96.58
Ramachandran outliers (%)	0
Molprobity score	1.33 (96th percentile)

Table 5.6 - Refinement statistics

5.5 Tetrathionate co-crystallisation of TsdA

Crystals grew in 2 weeks at 20°C in a condition containing 70% protein solution and 30% mother liquor (Figure 5.7). Protein solution contained TsdA protein at a concentration of 380 µM and 2 mM sodium tetrathionate. Mother liquor was composed of 100 mM ammonium acetate, pH 4.6, 7.5% PEG 10,000.

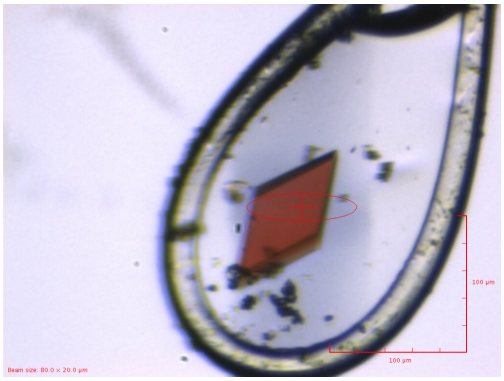

Figure 5.7 - Tetrationate soaked TsdA protein crystal
TsdA crystal mounted at Diamond Light Source's i03 beamline. Red oval shows beam dimensions and scale bars in the bottom right indicate a 100µm x 100µm square.

5.5.1 Data collection

Crystals were cyroprotected with 30% ethylene glycol and flash frozen in liquid nitrogen. A dataset was collected at Diamond Light Source's i03 beamline. Initial test images were taken to determine an appropriate strategy and subsequently 180

degrees of data were collected in case of incorrect determination of crystal symmetry at this point (Table 5.7).

Data Collection	
Diffraction source	Diamond Light Source, i03
Detector	Pilatus 6M
Energy (eV)	13477
Wavelength (Å)	0.92
Temperature (K)	100
Detector resolution (Å)	1.10
Exposure per image (s)	0.2
Ω start (°)	158
Ω end (°)	338
Ω osc (°)	0.15
Number of images	1200

Table 5.7 - Initial data collection

5.5.2 Data processing

Data were processed using XIA2 (Winter et al., 2013) running in 3daii mode, indexing a monoclinic cell (a = 79.28 Å, b = 70.42 Å, c = 57.91 Å, α = 90.00°, β = 129.34°, γ = 90.00 °) in the C 1 2 1 spacegroup with data to 1.29 Å (Table 5.8 and Figure 5.8).

Data Collection	**Tetrathionate**
Resolution range (Å)	44.52-1.29 (1.32-1.29)
Space group	C 1 2 1
Unit cell dimensions	
a, b, c (Å)	79.28, 70.42, 57.91
α, β, γ (°)	90.00, 129.34, 90.00
Total reflections	170161 (12379)
Unique reflections	60432 (4484)
Multiplicity	2.8 (2.8)
Completeness (%)	97.8 (98.0)
Mean I/sigma(I)	15.5 (2.3)
R_{merge} (%)	3.4 (47.8)
R_{pim} (%)	2.3 (33.5)
CC(1/2)	99.9 (86.0)

Table 5.8 - Data processing

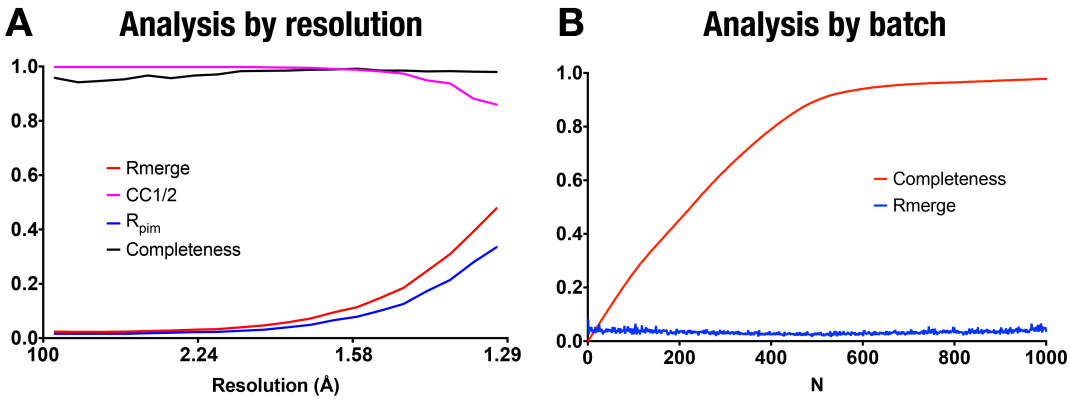

Figure 5.8 - TsdA data processing analysis
Panel A: Analysis by resolution. Panel B: Analysis by batch.

5.6 TsdA-tetrathionate structure determination, refinement & model validation

5.6.1 Structure determination

PHENIX.XTRIAGE (Adams et al., 2010) suggested no evidence of twinning or translational non-crystallographic symmetry. Cell content analysis by MATTHEWS (Kantardjieff and Rupp, 2003) implied the presence of a single copy of TsdA (Table 5.9).

Nmol / asym	Matthews coefficient	% solvent	P(1.29)	P(tot)
1	2.42	49.27	1.00	1.00

Table 5.9 - Cell content analysis

Analysis of the self-rotation function calculated with MOLREP (Vagin and Teplyakov, 2010) revealed no peaks consistent with a non-crystallographic symmetry axis, only a single non-origin peak with a 100% intensity representing the crystallographic 2-fold axis along b. PHASER (McCoy et al., 2007) was used to place a single copy of the refined native structure in the asymmetric unit of the new tetrathionate dataset (Table 5.10).

Copy	RFZ	TFZ	LLG
TsdA	62.6	42.0	11804

Table 5.10 - Molecular replacement with tetrathionate dataset

5.6.2 Model building & refinement

Limited rounds of model building and refinement were carried out using COOT (Emsley et al., 2010) and REFMAC (Murshudov et al., 1997), resulting in a final model with an R_{work} of 14.83% and an R_{free} of 16.40% (Table 5.11).

Refinement statistics	Tetrathionate crystal
Resolution range (Å)	44.52-1.29
Number of reflections (work / test)	53579 / 3003
Number of atoms (protein, ligand, water)	1784, 90 (heme) & 5 (thiosulfate), 202
Protein residues	234
Mean B-factors (protein, ligands, waters)	22.39, 14.94 (heme) & 21.83 (thiosulfate), 33.86
R_{work} (%)	14.82
R_{free} (%)	16.40
R.M.S.D. from ideal values (bonds, Å)	0.030
R.M.S.D. from ideal values (angles, °)	2.418
Ramachandran favoured (%)	97.90
Ramachandran outliers (%)	0.42
Molprobity score	1.22 (95th percentile)

Table 5.11 - Refinement statistics

5.7 Structural analysis

The higher resolution tetrathionate TsdA structure was used for the analysis of the overall protein structure because this model was refined against a higher resolution dataset and exhibited the same overall fold as the lower resolution structure (R.M.S.D. = 0.156 Å).

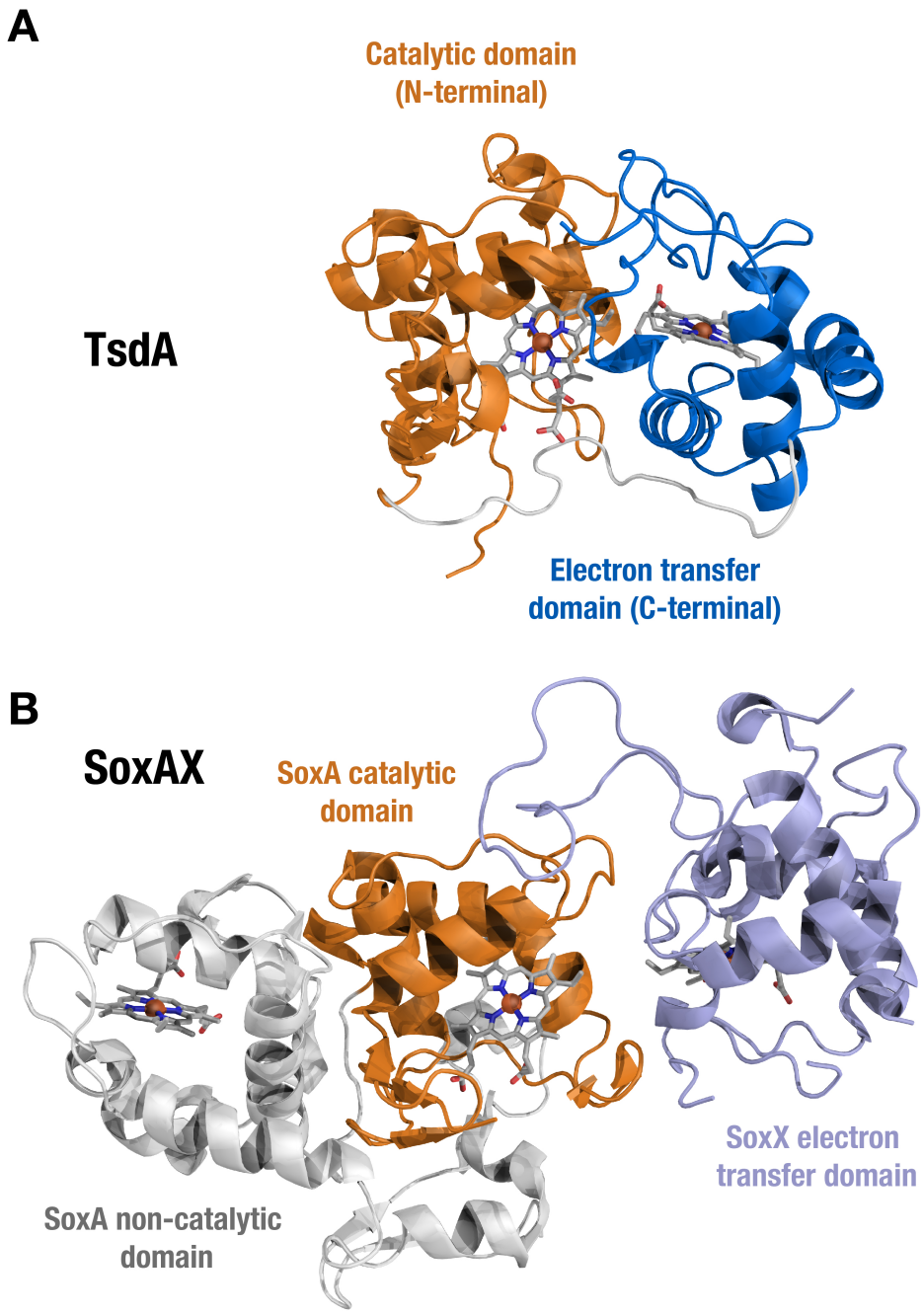

Figure 5.9 - Structure of TsdA and comparison with SoxAX
Protein molecules shown in cartoon representation with catalytic domains coloured in orange. Heme groups are shown in stick representation with carbon atoms coloured grey, nitrogen in blue, oxygen in red and iron atoms shown as orange spheres.

The structure of TsdA revealed a protein composed of two cytochrome c-like domains linked by an unstructured linker region (Figure 5.9A). The N-terminal catalytic domain of TsdA was homologous to the C-terminal domain (residues 177-282) of *Rhodovulum sulfidophilum* SoxA [PDBID: 1H32] (R.M.S.D. = 1.15 Å over 45 equivalent Cα atoms) and featured a heme group with Cys123 and His80 ligating the heme iron (Figure 5.14). The C-terminal domain of TsdA exhibited a standard mitochondrial cytochrome c fold with homology to *Pseudoalteromonas haploplantis* nitrite reductase [PDBID: 2Z00] (R.M.S.D. of 2.47 Å over 55 equivalent Cα atoms) and featured a heme group with the heme iron ligated by Met236 and His191. Significant heme ruffling was observed in the C-terminal heme, however the functional importance of this is unknown. Comparison of the TsdA and SoxAX structures showed that although both proteins pack the same faces of their catalytic and electron transfer domains against one another, the exact orientation is different, resulting in different distances between the two heme groups (Figure 5.9).

The organisation of the two cytochrome c-like domains of TsdA results in a protein structure that positions the two heme groups with a minimum edge-to-edge distance of 8.1 Å. This would allow extremely rapid transfer of electrons between heme molecules and may be important as each heme can only accept one electron, yet the forward thiosulfate oxidation reaction will release two electrons. This distance is 10.7 Å in the SoxAX structure, however the difference in distance is unlikely to be relevant to the enzyme mechanisms as electron transfer between heme groups will be much more rapid than the catalysis.

Analysis of the heme environments showed that both groups are accessible to bulk solvent (Figure 5.10). Electrostatic analysis of the protein surface revealed that there was a highly basic entrance and channel to the active site on the Cys123 side of the N-terminal heme, suggesting the use of electrostatic funneling to attract anionic substrate into the active site (Figure 5.10A). In contrast, the C-terminal heme iron was

completely buried, but one edge of the heme group was exposed through a cleft (Figure 5.10B). Sequence conservation analysis was carried out with CONSURF using 150 unique sequences selected from a CSI-BLAST search (Ashkenazy et al., 2010). This analysis revealed that the exposed edge of the C-terminal heme was in an area of increase sequence conservation suggesting a possible binding site for an electron acceptor protein.

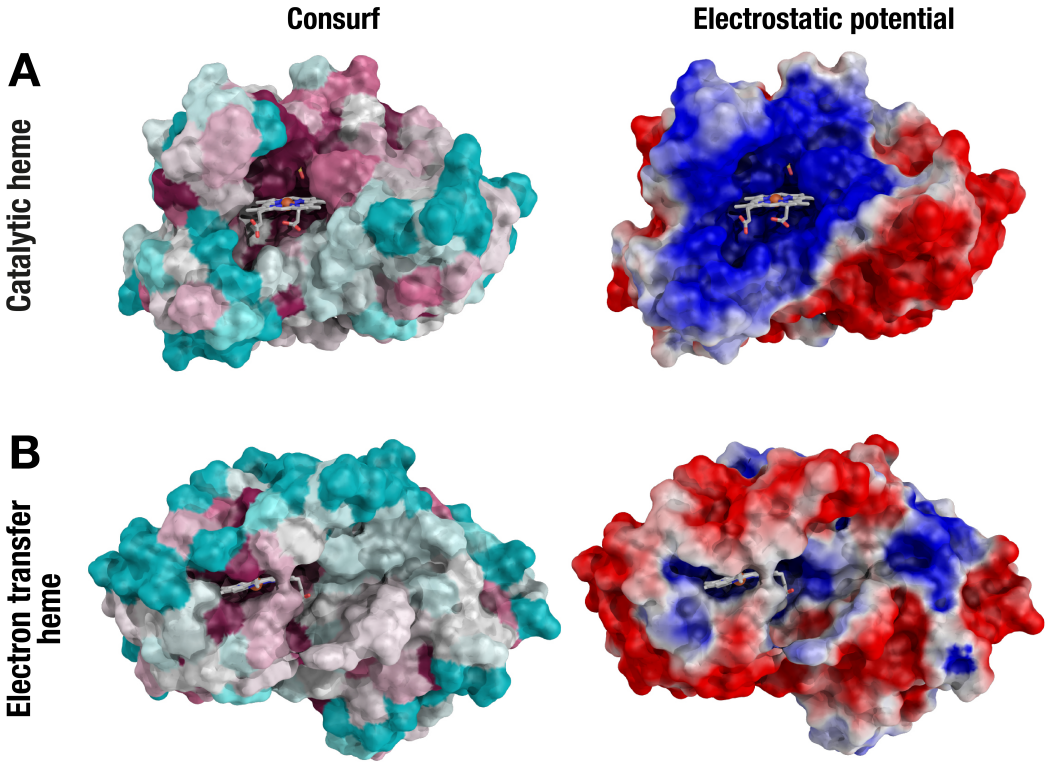

Figure 5.10 - Accessibility and environment of heme groups in TsdA
Sequence conservation analysis using CONSURF (Ashkenazy et al., 2010) shows sequence conservation on the surface of TsdA in the left column. Regions of high sequence conservation are coloured in purple and those of low sequence conservation in teal. Electrostatic analysis as calculated by ABPS (Dolinsky et al., 2004) in PyMOL (Schrödinger, LLC) is shown in the right column. Red indicates regions of negative charge and blue of positive charge. Proteins are shown as surfaces and heme groups as sticks.

This suggested the possibility that electrons released from thiosulfate oxidation by TsdA are initially accepted by the N-terminal heme, transferred to the C-terminal

heme and then to a possible electron acceptor protein that binds on the TsdA surface near to the exposed edge of the C-terminal heme.

5.7.1 TsdA active site

Inspection of the difference map peaks for the native structure revealed two problematic regions. Firstly, there was ambiguous electron density close to the catalytic heme and catalytically important Cys123 residue that could not be modelled with confidence (Figure 5.11). Secondly, there was poor quality electron density for residues 234-243, a region that contained the heme-coordinating residue Met-236. It could be argued that the accumulation of radiation damage over the course of collection 720 degrees of data may have influenced the uninterpretable electron density at the catalytic heme, however, when only processing the first 180 degrees of data and refining the model against this dataset, the observed electron density was almost identical in these regions. It was likely that the observed electron density at the catalytic heme was a result of either multiple different molecular species and conformational flexibility, or both.

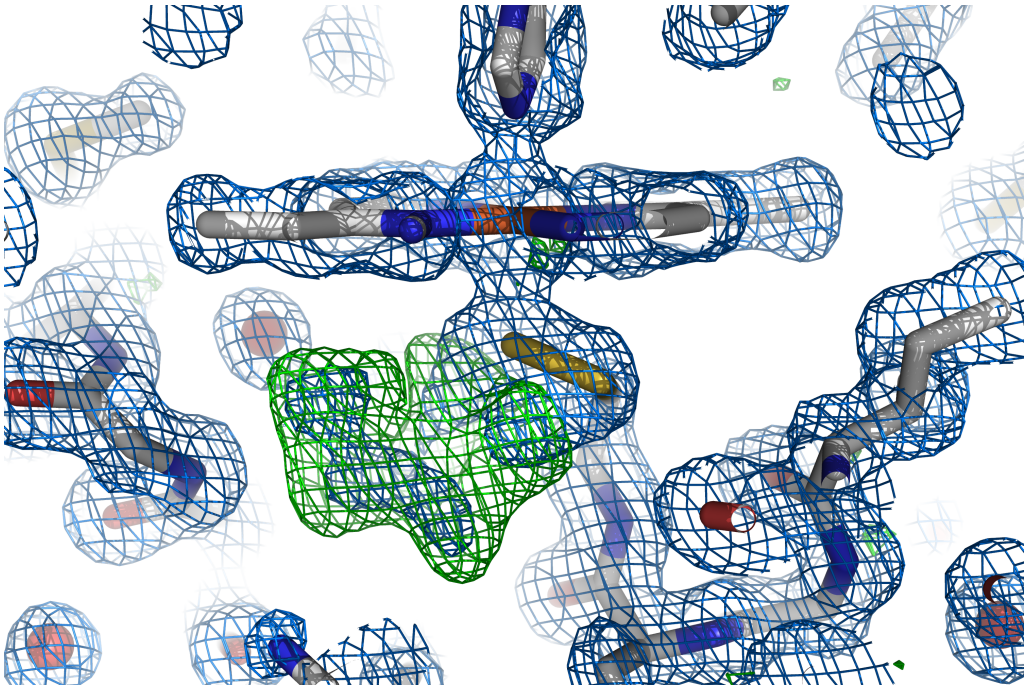

Figure 5.11 - Electron density at Cys123
Protein shown in stick representation with water molecules displayed as red spheres. 2Fo-Fc electron density map contoured at 1.4σ (blue). Fo-Fc map contoured at 3σ (green).

In contrast, observed electron density at the active site in the tetrathionate structure could be modelled as a S-sulfane modified cysteine residue with the terminal sulfur atom coordinating the iron of the heme group (Figure 5.12). There was additional density in the active site region that appeared to resemble a single molecule of thiosulfate. However upon modelling this, a negative residual map peak indicated that this molecule was not at 100% occupancy. Occupancy refinement was carried out in REFMAC (Murshudov et al., 1997) with the density best modelled as either a single thiosulfate molecule at 46% occupancy or a single water molecule at 54% (Figure 5.12).

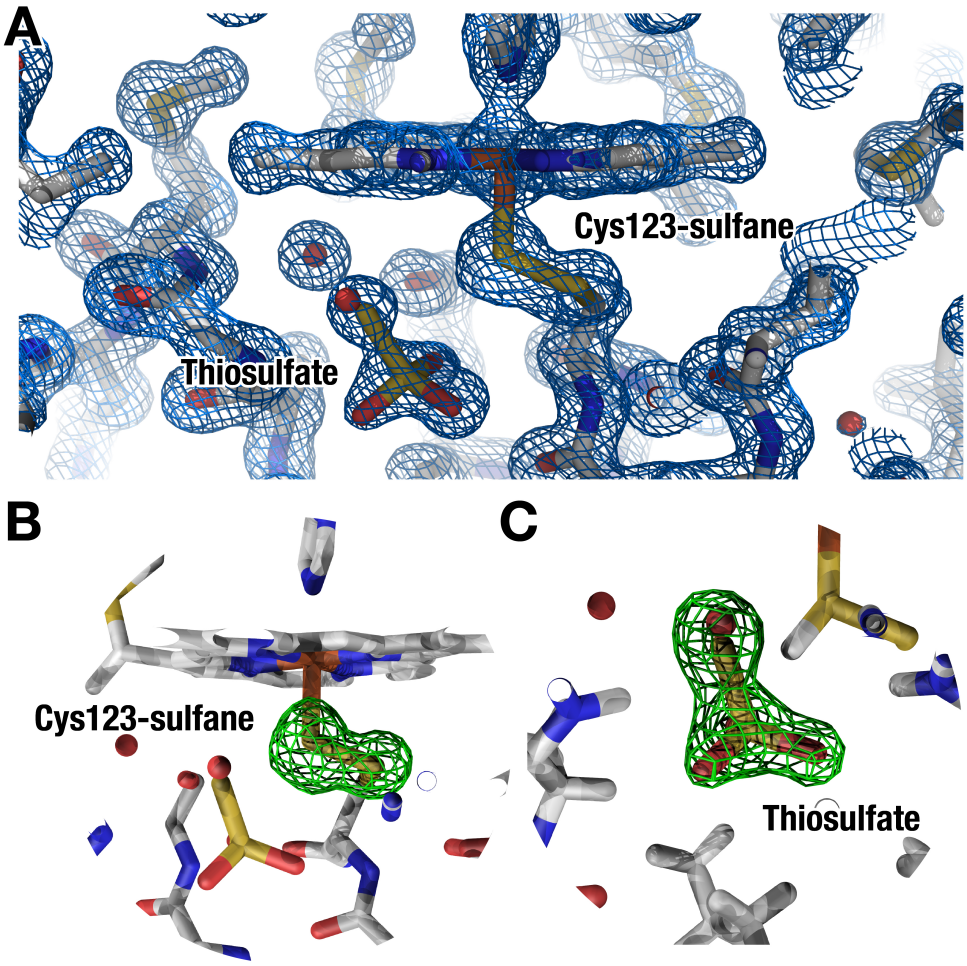

Figure 5.12 - TsdA-tetrathionate active site
Protein shown in stick representation with water molecules displayed as red spheres. Panel A: 2Fo-Fc map contoured at 1.4σ (blue). Panels B & C: Omit (Fo-Fc) maps contoured at 3σ (green) following removal of the two sulfur atoms in the Cys123 S-sulfane modified residue (B) and both the thiosulfate and water molecule (C).

To confirm the existence of Cys123 sulfonation and the presence of the bound thiosulfate and water molecule in the active site, two separate omit maps were calculated. Either the two sulfur atoms of the Cys123-sulfate residue or both the thiosulfate and water molecule were removed before the B-factors for all atoms in the model were reset back to the Wilson B factor. These structures were then re-refined in REFMAC resulting in the omit maps observed in Figure 5.12B & C (Murshudov et al., 1997).

It should noted that the tetrathionate structure also exhibited much more well-defined electron density for residues 234-243 compared to the native structure,

allowing the Met236 (that ligates the electron transfer heme) and surrounding residues to be modelled with greater confidence.

5.7.2 Anomalous TsdA tetrathionate dataset

An additional TsdA-tetrathionate dataset was collected at a longer wavelength (1.40 Å) in order to verify the interpretation of the electron density at the active site. This allowed the positions of anomalously scattering atoms to be located which could then be compared to the current model of the active site. In order to accurately measure the small differences in intensities between Friedel mates that occur in the presence of anomalous scatters, it was necessary to collect a dataset with a high multiplicity (Table 5.12).

Data Collection	
Diffraction source	Diamond Light Source, i03
Detector	Pilatus 6M
Energy (eV)	8856
Wavelength (Å)	1.40
Temperature (K)	100
Detector resolution (Å)	1.73
Exposure per image (s)	0.04
Ω start (º)	171
Ω end (º)	1371
Ω osc (º)	0.15
Number of images	8000

Table 5.12 - Long wavelength anomalous data collection

Data were processed using XIA2 (Winter et al., 2013) running in 3da mode, indexing a monoclinic cell (a = 79.37 Å, b = 70.37 Å, c = 57.95 Å, α = 90.00°, β = 129.39°, γ = 90.00 °) to a resolution of 1.66 Å in the C 1 2 1 space group (Table 5.13).

Data Collection	Tetrathionate
Resolution range (Å)	21.72-1.66 (7.42-1.66)
Space group	C 1 2 1
Unit cell dimensions	
a, b, c (Å)	79.37, 70.37, 57.95
α, β, γ (°)	90.00, 129.39, 90.00
Total reflections	609427 (31969)
Unique reflections	28404 (1914)
Multiplicity	21.5 (16.7)
Completeness (%)	97.5 (90.3)
Mean I/sigma(I)	32.8 (4.4)
R_{merge} (%)	6.8 (66.0)
R_{pim} (%)	2.1 (23.5)
CC(1/2)	1.00 93.8
Anomalous completeness (%)	96.8 (88.9)
Anomalous multiplicity	10.6 (8.0)
Anomalous correlation	0.64 (0.106)

Table 5.13 - Long wavelength anomalous data processing

Due to the slight differences in the cell dimensions compared to the original tetrathionate dataset, the current model was used in a rigid body refinement against the new data, before a small number of model building and refinement cycles were carried out using COOT (Emsley et al., 2010) and REFMAC (Murshudov et al., 1997). Phases from the final refinement in REFMAC (Murshudov et al., 1997) were converted from PHI/FOM values to Hendrickson Lattman coefficients using the CCP4 (Winn et al., 2011) program HLTOFOM, allowing a residual log likelihood gradient map to be calculated using SHARP (Vonrhein et al., 2007) after modelling the contribution of the heme iron atoms (Figure 5.13).

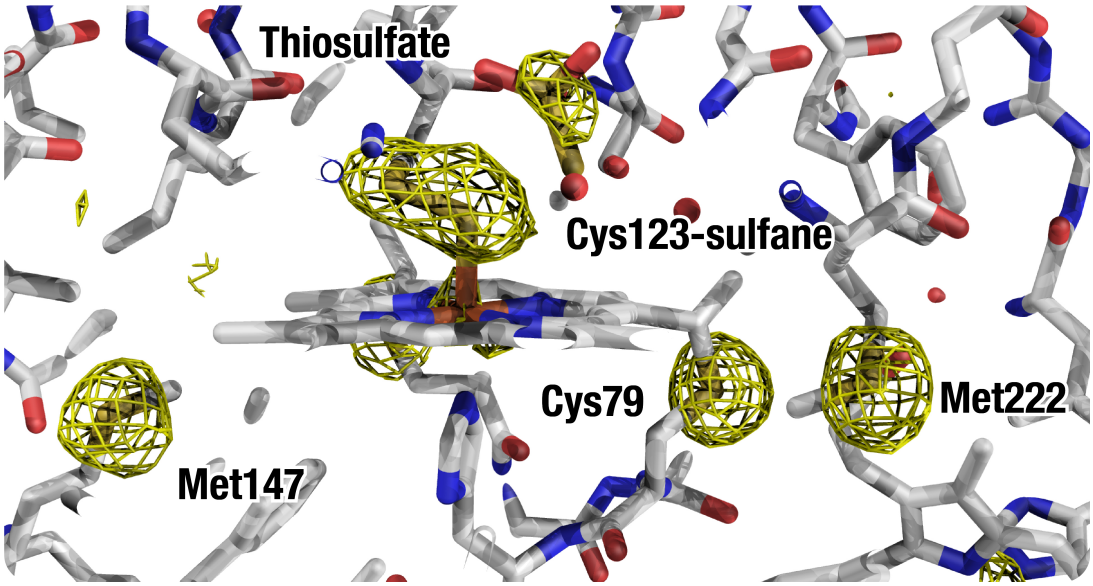

Figure 5.13 - Residual log likelihood gradient map
Residual log likelihood gradient maps in long wavelength (1.4Å) data after modelling contributions of the heme iron atoms. Maps are contoured at 3σ (yellow) and reveal density around sulfur atoms present in the model.

The generated residual log likelihood gradient maps showed defined density around the sulfur atoms, validating the interpretation of the cysteine S-sulfane modification and the presence of a bound sulfur-containing molecule, likely thiosulfate.

Comparison of the TsdA and SoxAX active site cavities showed that the TsdA cavity was significantly smaller than the SoxAX cavity, primarily due to the insertion of residues from the electron transfer domain (Figure 5.14). In contrast, the SoxAX active site was composed exclusively of residues from the catalytic domain and resulted in its comparatively larger size. This was consistent with the need for the SoxAX active site cavity to accommodate the SoxYZ carrier arm for substrate delivery.

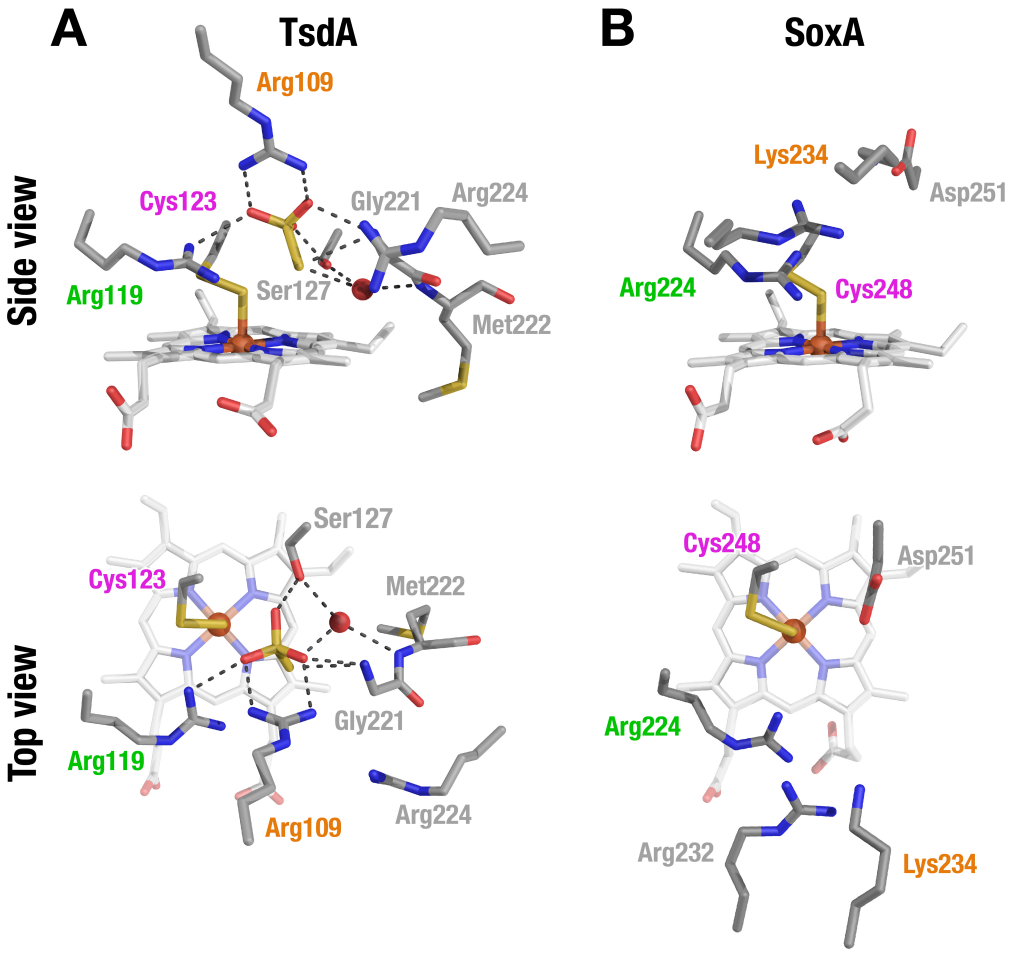

Figure 5.14 - TsdA and SoxA active site comparison
Protein residues, heme groups and thiosulfate are shown as sticks. Iron atoms of heme groups and water molecules are shown as spheres. Panel A shows the new TsdA structure, while panel B shows the SoxAX structure (PDBID: 1H32). Non-covalent interactions are indicated with black dotted lines. Structurally homologous residues are labelled with equivalent colours.

The TsdA structure revealed the presence of a bound molecule of thiosulfate in the TsdA active site cavity (on the Cys123 side of the N-terminal catalytic heme), offering insight into how TsdA binds both substrates and intermediates (Figures 5.12, 5.13 & 5.14). The presence of thiosulfate in the crystallisation environment is likely due to the thiosulfonation of Cys123 by tetrathionate, which produces it as a product. The thiosulfate molecule forms interactions with the side chains of residues Arg109, Arg119 and Ser127. Sequence analysis showed that Arg119 is highly conserved in both TsdA and SoxA sequences. Arg119 is conserved in TsdA sequences, being

replaced by lysine in some, whereas in SoxA sequences, this amino acid is replaced by either lysine or serine residues. Ser127 is well conserved in TsdA sequences, but is not present in SoxA sequences. Interestingly, although Arg224 of the TsdA electron transfer domain does not form interactions with the bound molecule of thiosulfate, it is in a position that could allow it to bind negatively charged substrates or intermediates. This residue is partially conserved in TsdA sequences with no equivalent in SoxA. Finally, SoxAX features an additional arginine residue between the Arg109 and Arg119 that would be well placed to coordinate the thiosulfate molecule if it binds in a similar location as in the TsdA structure.

To assess the functional importance of these residues identified in the active site, our collaborators carried out mutagenesis experiments. They mutated each residue to alanine and then assessed the impact on steady state reaction kinetics in the presence of a low (0.2 mM) or high (1 mM) concentration of ferricyanide (Table 5.14). There was no significant difference for the R224A mutant suggesting no significant role in catalysis. There were changes in enzyme activity for the other three mutants with R119A exhibiting the most dramatic change, consistent with it being the most well conserved of the chosen residues for mutation. Analysis of the visible spectra for each mutant in the presence and absence of thiosulfate suggested that there were changes in the amount of high spin heme relative to the wild type enzyme (illustrated by changes in the size of the 625 nm peak). R109A and R119A had increased amounts of high spin heme whereas S127 had reduced amounts (Figure 5.15).

Variant	0.2 mM $K_3Fe(CN)_6$			1 mM $K_3Fe(CN)_6$		
	K_m (mM $S_2O_3^{2-}$)	k_{cat} (s^{-1})	n	K_m (mM $S_2O_3^{2-}$)	k_{cat} (s^{-1})	n
Wild type	0.22 ± 0.03	920 ± 50	1.4 ± 0.25	2.2 ± 0.3	4280 ± 150	1.52 ± 0.25
S127A	4.4 ± 0.5	1530 ± 150	2.4 ± 0.7	10.7 ± 6.2	3060 ± 920	1.1 ± 0.4
R119A	17 ± 2	378 ± 20	1.2 ± 0.1	47 ± 7	640 ± 40	1.0 ± 0.05
R109A	1.11 ± 0.25	780 ± 95	1.6 ± 0.5	5.8 ± 0.8	3210 ± 310	1.7 ± 0.3
R224A	0.44 ± 0.05	870 ± 50	1.3 ± 0.2	2.5 ± 0.8	3670 ± 460	1.14 ± 0.26

Table 5.14 - Kinetic analysis of TsdA mutants
Measurements of activity were carried out at pH4.25 in the presence of 0.2 mM or 1 mM potassium ferricyanide acting as an electron acceptor. n = apparent Hill coefficient. Standard errors for curve fits with duplicate points are shown. Experiments carried out by collaborators in the Berks laboratory (Biochemistry Department, University of Oxford).

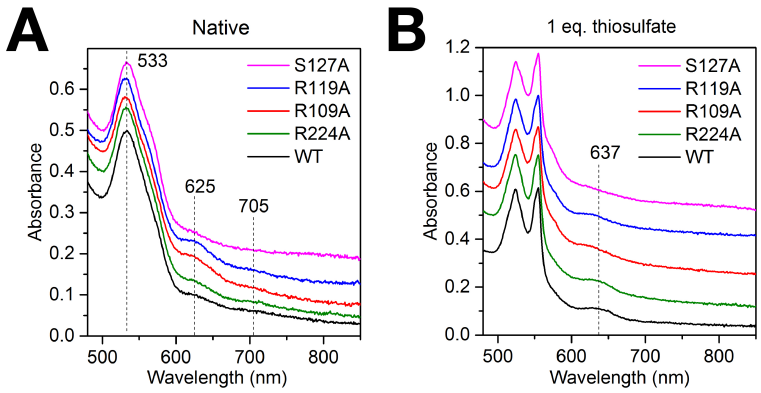

Figure 5.15 - Visible spectra of TsdA active site mutants
Proteins were at concentrations of 25 μM. Panel A shows visible spectra of untreated samples. Panel B shows visible spectra of proteins incubated for 3 minutes with equimolar thiosulfate. Experiments carried out by collaborators in the Berks laboratory (Biochemistry Department, University of Oxford).

To investigate further, our collaborators subjected each mutant to ESI-MS analysis including separate incubations with thiosulfate or tetrathionate (Figure 5.16). This revealed that the R109A and R119A mutants were present in multiple modified forms and given the increase in high spin heme content, suggested that these modifications were at the Cys123 catalytic residue. These two residues may help to suppress inappropriate reactivity at Cys123.

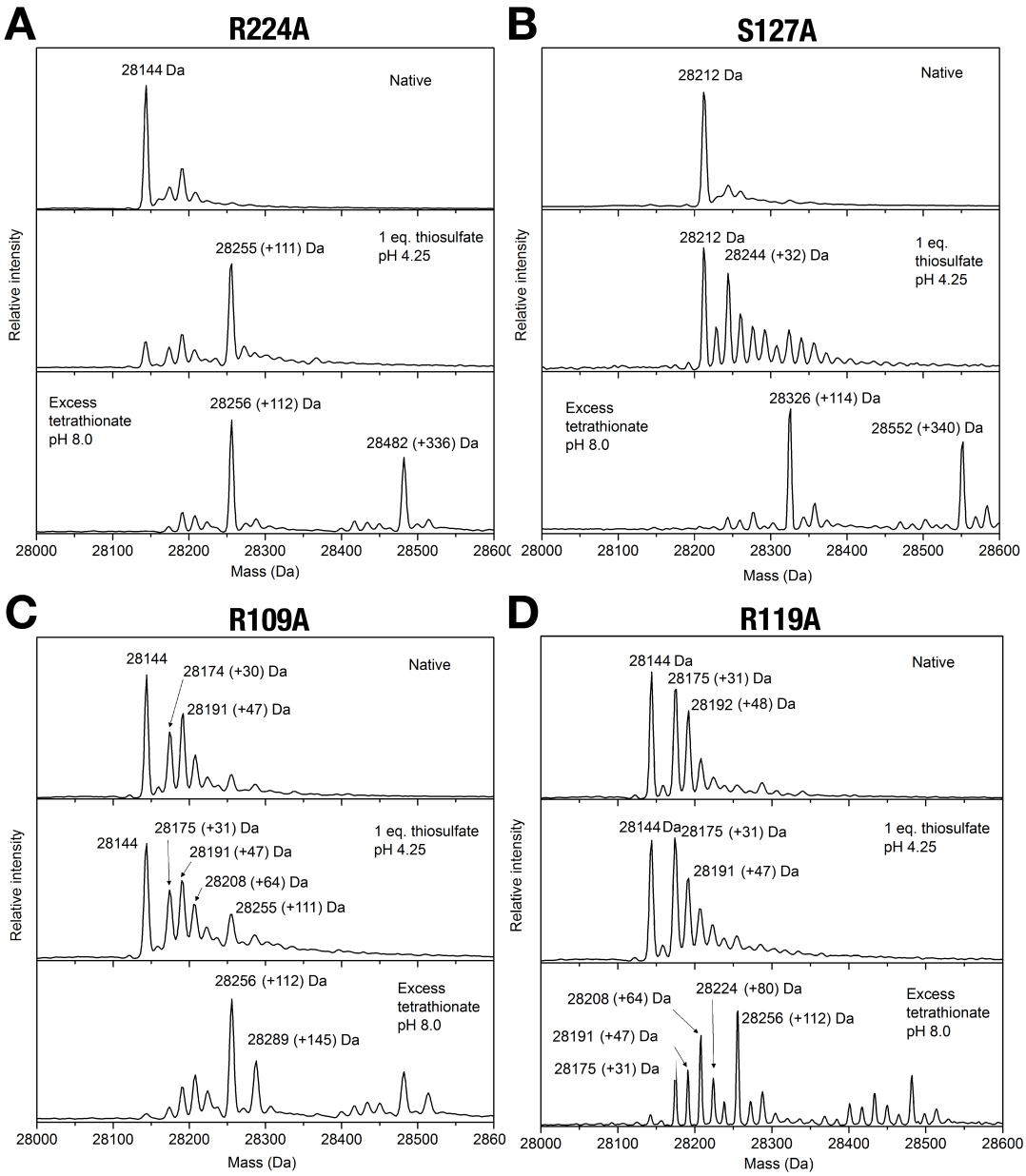

Figure 5.16 - ESI-MS of TsdA active site mutants
TsdA mutants were untreated, incubated with equimolar thiosulfate for 3 minutes or incubated with excess tetrathionate for 1 hour. Experiments carried out by collaborators in the Berks laboratory (Biochemistry Department, University of Oxford).

The visible spectra showed that all mutants were reduced by incubation with thiosulfate (Figure 5.15), however the three mutants that affected enzyme activity did not form the band at 637nm that indicated S-thiosulfonation at Cys123. Subsequent analysis by ESI-MS (Figure 5.16) showed that a significant proportion of the mutants remained unmodified, with only small peaks corresponding to S-thiosulfonation (+111 Da). Taken together, this suggests that the S-thiosulfonation adduct is formed, but it is

unstable due to the absence of key interactions that result from mutation. Indeed, incubation with tetrathionate at pH 8.0 resulted in the formation of this adduct.

5.8 Discussion

Prior to the crystallisation and structure solution of TsdA, collaborators in the Berks laboratory (Biochemistry Department, University of Oxford) carried out a number of biochemical experiments that provided evidence supporting the notion that S-thiosulfonation of the catalytic heme-ligating cysteine residue of TsdA (and likely other SoxA family enzymes) was a catalytic intermediate. They showed that this cysteine residue was chemically reactive and detachable from the heme iron, that it was indeed S-thiosulfonated when exposed to thiosulfate or tetrathionate, that enzymatic activity was not impeded by the modification and that the adduct was removed during enzyme turnover.

Solution of the native TsdA crystal structure revealed uninterpretable electron density at the active site. This was likely due to the presence of multiple modifications, conformations or both. In order to obtain a structure solution with interpretable density at the active site, the enzyme was co-crystallised with DTT, tetrathionate or thiosulfate. The only structure solution that exhibited interpretable density at the active site was the tetrathionate co-crystallisation. This model featured an S-sulfane modification of the catalytic Cys123 residue with the terminal sulfur atom coordinating the heme iron atom, as observed in multiple SoxAX structures. The model at the active site was verified using simulated annealing maps and exploiting the anomalous scattering of x-rays by sulfur atoms to generate a residual log likelihood gradient map identifying the positions of anomalously scattering sulfur atoms. Our collaborators' ESI-MS experiments support the notion that this species is formed as a result of decomposition from the S-thiosulfonated catalytic cysteine residue. They showed that

this residue was unmodified in the purified sample and that once it was thiosulfonated in vitro, the S-sulfane modified species began to appear over time. Although the terminal sulfur atom of the S-sulfane modified cysteine coordinates the heme iron in our structure, it is likely that, in its native state, the cysteine thiolate ligates the heme iron as suggested by previous EPR experiments.

The tetrathionate co-crystallised structure featured a molecule of thiosulfate in the TsdA active site bound by three active site residues. The importance of these residues for binding was confirmed by our collaborators' mutagenesis experiments. Multiple partial modifications were observed in ESI-MS analysis of the R109A and R119A mutants suggesting that these residues, as well as participating in catalysis, may help prevent inappropriate reactivity at Cys123 with other sulfur containing molecules that may be present in the cell.

The structure also revealed an electron transfer route, with electrons released in the oxidation of thiosulfate passed from the catalytic heme to the electron transfer heme and then to an external electron donor.

Taking both the biochemical evidence of our collaborators and the new structural data, a mechanism of catalysis is proposed (Figure 5.17). The two half reactions of TsdA involve disulfide bond formation with incoming thiosulfate molecules acting as nucleophiles. The location of the thiosulfate molecule seen in the crystal structure appears to occupy the position of the thiosulfate molecule involved in the first half reaction as only a small movement would be required to reach the expected geometry of the disulfide bond-forming transition state. It is likely that the S-thiosulfonate adduct remains bound to these conserved residues, stabilising the intermediate. At this stage, the reaction is paused until the heme groups are reoxidised. The location of a second thiosulfate molecule required for the second half reaction is proposed to be adjacent to Ser127, a residue that coordinates a water molecule in our structure. This residue is not present in SoxA sequences and the

active site of SoxA is enlarged in this region, consistent with it being the location of the SoxY carrier arm that acts as the substrate in the second half reaction of SoxA.

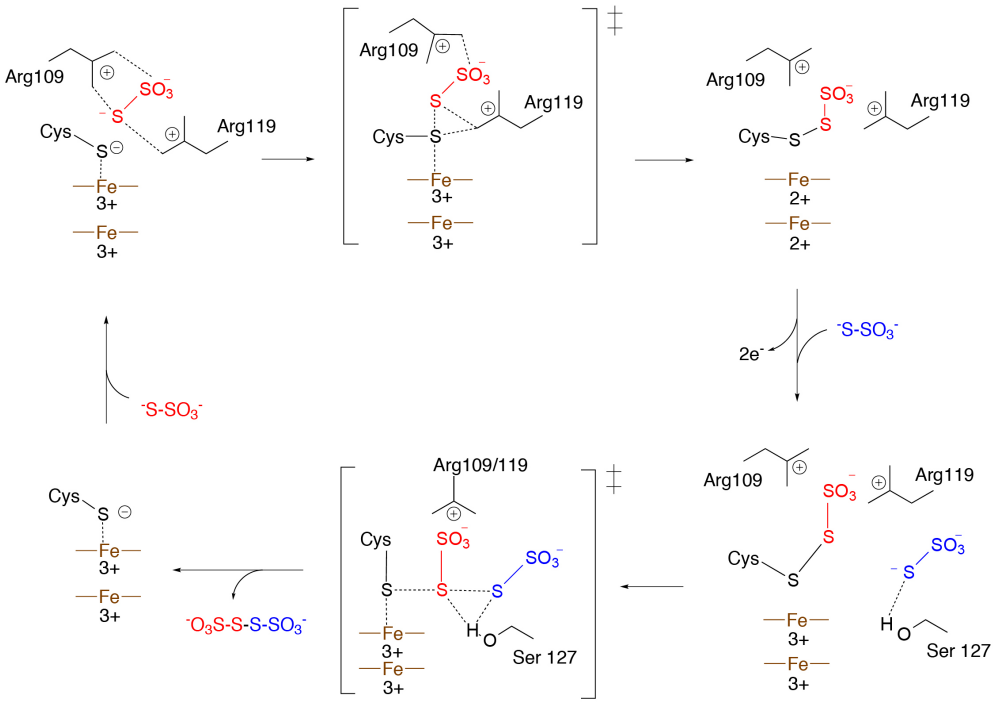

Figure 5.17 - Proposed reaction mechanism
Top heme is the N-terminal catalytic heme and the bottom heme is the C-terminal electron transfer heme. Thiosulfate molecules of first and second half reactions are coloured red and blue, respectively. Arg109 and Arg119 form interactions with the first thiosulfate molecule and following S-thiosulfonation of Cys123, remain bound to stabilize this intermediate. Heme reoxidation then occurs via an external electron acceptor before a second molecule of thiosulfate participates in a thiol-disulfide exchange reaction with the thiosulfonate group.

This work was published back-to-back with another paper that presented both biochemical experiments and a number of X-ray crystal structures of *Allochromatium vinosum* TsdA (Brito et al., 2015). The authors crystallised native TsdA along with a number of mutants and in the presence and absence of bisulfite, dithionite and tetrathionate. Their models revealed the same overall protein fold as presented in this chapter. Interestingly, the authors observed ligand switching at the electron transfer heme, with Lys235 coordination of the heme iron under oxidative conditions switching

to Met236 coordination under reducing conditions. Our two structures both show ligation of the electron transfer heme iron by Met236. The authors proposed that this was a mechanism by which TsdA could hinder the back-reaction occurring upon reduction of the electron transfer heme.

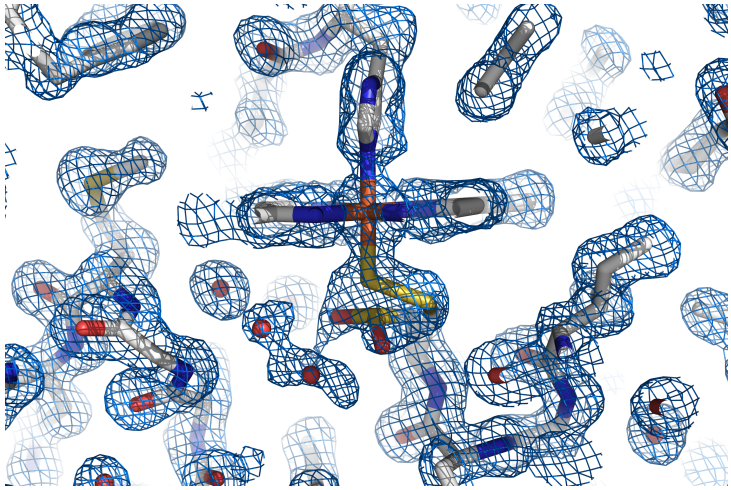

Figure 5.18 – Active site electron density observed by Brito et al.
PDBID 4WQ8 shown in stick representation with water molecules displayed as red spheres. 2Fo-Fc electron density map contoured at 1.4σ (blue) as calculated for the native structure presented in this chapter.

Inspection of the electron density around the catalytic heme of TsdA in their deposited structures revealed heterogeneous electron density, as observed in our native TsdA structure, into which the authors had modelled multiple different species and conformations of the catalytic cysteine residue and ligands. In some cases, they had modelled the Cys123 residue as a sulfane derivative (as modelled in this chapter) and in other cases they had modelled an unmodified cysteine residue in two conformations, either with the Sγ atom coordinating the heme iron or dissociated from it. The authors also observed a bound thiosulfate molecule in a very similar position to that observed in this chapter. When overlaying the various structures presented by these authors into the electron density observed in the native structure in this chapter, one particular model appeared to offer a possible molecular explanation for the electron density with the catalytic residue at position 123 modelled as either a cysteine

S-sulfonate, coordinating the heme iron, or partially oxidised cysteine sulfonate group (Figure 5.18).

A model for the ambiguous electron density observed at the TsdA active site for the native structure in this chapter was not presented because it was impossible to do this with sufficient confidence. The authors of this paper were considerably less conservative in modelling different species into the ambiguous electron density surrounding the Cys123 residue and in a location equivalent to where the thiosulfate molecule is observed in our tetrathionate structure. Presumably their models were just a small number of possible models that would explain the observed electron density. In order to confidently model the active site, it would be necessary to obtain crystals that contained protein molecules that did not feature a heterogeneous mix of species and conformations.

5.9 Future work

The homology of catalytic domains of SoxA and TsdA has offered an opportunity to explore the catalytic mechanism of SoxA family enzymes, something that has previously not been possible due to the intractability of SoxAX to biochemical experimentation. The new TsdA structure, in combination with our collaborators biochemical experiments, offers evidence consistent with our proposed reaction mechanism. In order to provide additional validation of the mechanism, further experimentation would be required. It would be highly advantageous to solve the crystal structures of the enzyme trapped in intermediate states, although this was not possible by soaking with DTT, tetrathionate or thiosulfate. Active site mutants or alternative substrate intermediate soaks may offer the solution.

Chapter 6: Appendix

6.1 X-ray crystallography

6.1.1 Protein crystals and diffraction

The sitting-drop vapour-diffusion technique involves drops containing protein mixed with mother liquor being sealed in a well containing a reservoir of mother liquor. The precipitant-containing mother liquor in the reservoir absorbs water from the protein-containing drop thereby increasing both the protein and precipitant concentration in the drop. Initially, protein and precipitant concentrations should be below the solubility curve, however, as both the concentration of protein and precipitant rise, they may reach conditions optimal for crystal nucleation (Figure 6.1). If crystal nucleation does occur, protein concentration in the drop will begin to decrease as the crystal grows and crystal growth will halt once the conditions fall back below the solubility curve. However, if the concentrations are too high, conditions may enter the precipitation zone and protein molecules will precipitate out of solution.

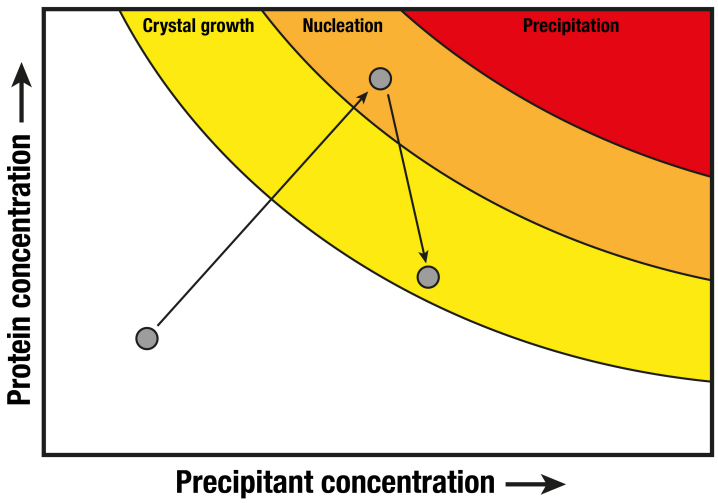

Figure 6.1 - Phase diagram showing crystal growth by vapour diffusion

Protein crystals are periodic assemblies of protein molecules held together by intermolecular interactions at so called crystal contacts. Protein crystals can be divided into translationally periodic repeating unit cells of identical content called the crystal lattice. There are seven different crystal lattices (triclinic, monoclinic, orthorhombic, trigonal, hexagonal, tetragonal or cubic) that, in combination with lattice centring, give 14 Bravais lattice types. The unit cell of a crystal form is defined by three distances (a, b and c) and three angles (α, β and γ) (Figure 6.2).

In the simplest case the unit cell possesses no internal symmetry and is itself the asymmetric unit (P1 space group). If the unit cell does contain internal symmetry, the contents of the asymmetric unit in combination with this space group symmetry can be used to generate the contents of the unit cell. The chiral nature of protein molecules means that protein crystals can only exist in one of 65 (of the 230) possible space groups.

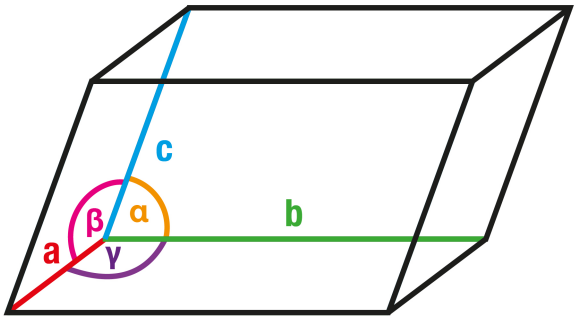

Figure 6.2 - Unit cell dimensions

When protein crystals are exposed to an X-ray beam, protein atoms absorb the energy and subsequently reemit it in all directions. Scattered X-rays from each lattice plane destructively interfere at most scattering angles, however, angles that satisfy Bragg's law result in constructive interference that gives rise to the characteristic spots on diffraction images. In order to satisfy Bragg's law and constructively interfere, the

differences in path length of the X-rays must be an integral number of wavelengths (Figure 6.3).

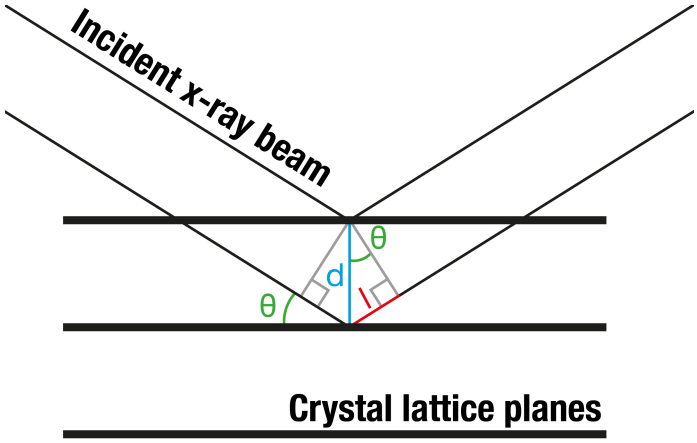

Figure 6.3 - Bragg's law
Diffraction maxima occur when Bragg's law is satisfied. $n\lambda = 2d\,sin\,\theta$ *, where n = integer, λ = wavelength, d = distance between diffracting lattice planes, θ = scattering angle*

The three dimensional diffraction pattern obtained from a protein crystal is the convolution of diffraction by the lattice and by the protein, representing a Fourier transform from real space into reciprocal space. Each reflection corresponds to a reciprocal lattice point that, in turn, corresponds to a set of lattice planes that run through the crystal. Each point on the reciprocal lattice is defined by Miller indices (h, k, l), with integer values indicating the number of times a lattice plane intercepts the unit cell axes (Figure 6.4). Miller indices with higher integer values indicate finer sampling of the unit cell and hence contain more information about the contents of the unit cell.

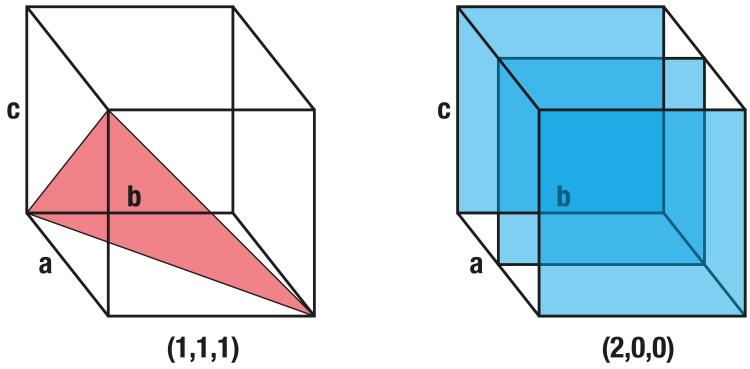

Figure 6.4 - Lattice planes and Miller indices

6.1.2 Data collection

In the case of using monochromatic X-rays from synchrotron sources, crystals are mounted on a goniometer via a nylon loop. Rotation of the crystal, after centring in the X-ray beam, brings the different lattice planes into conditions that satisfy Bragg's law. A 2-dimensional detector measures the intensities of the diffracted x-rays. Initially, a few diffraction images are collected so that a data collection strategy can be devised. These images are then indexed, a process in which the unit cell dimensions are determined by analysing the spatial relationship between strong reflections. This process is dependent on highly accurate detector parameters, such as the position of the incident beam and the distance between the crystal and detector. Crystal and detector parameters are then refined, resulting in a more accurate determination of reflection positions. Once the unit cell dimensions have been determined accurately and each reflection assigned Miller indices, it is then possible to predict where other reflections will be located on the 2 dimensional diffraction images as the crystal is rotated. Integration of these reflections and subsequent analysis of intensities can yield information about crystal symmetry. Using this information, an appropriate data collection strategy can be executed that allows the measurement of all possible reflections while limiting X-ray dose. An appropriate oscillation range per image can

also be determined to minimise overlap of reflections. With the replacement of older CCD detectors with much more sensitive hybrid photon counting detectors (Pilatus detectors), it is now possible to use much smaller oscillation ranges (fine slicing), further reducing the problem of overlapping reflections.

Once a full dataset has been collected and all reflections assigned Miller indices, the intensities of the reflections are determined by integration. Fitting three-dimensional Gaussian profiles in both the plane of the diffraction image and across images results in a more accurate determination of the reflection intensity compared to just two-dimensional integration in the plane of the image. By fitting three-dimensional profiles at this stage, partial reflections do not need to be merged later on, as they would have been in the case of two-dimensional integration of each individual image. During this integration stage, purely geometrical (Lorentz and polarisation) effects are corrected for.

After integration, the data is organised into a list of Miller indices and their associated intensities. Due to a number of factors, including X-ray beam fluctuation, crystal orientation and detector settings, equivalent reflections often exhibit variance in their intensities due to crystal symmetry or Fridel's law. Intensity differences between equivalent reflections are corrected for during the process of scaling, placing them all on a common scale. Merging is then carried out, combining the equivalent reflections. The completeness of a dataset is defined by the number of unique reflections observed divided by the total number of possible unique reflections (Equation 6.1) and the multiplicity by the number of observed reflections divided by the number of unique reflections observed (Equation 6.2).

$$Completeness = \frac{Number\ of\ unique\ reflections\ observed}{Total\ number\ of\ possible\ unique\ reflections}$$

Equation 6.1 - Data completeness

$$Multiplicity = \frac{Number\ of\ observed\ reflections}{Number\ of\ unique\ reflections\ observed}$$

Equation 6.2 - Data multiplicity

Multiple observations of equivalent reflections allow various statistics to be calculated that give indications of data quality and consistency. These are calculated for the dataset as a whole and also for sets of reflections grouped by resolution in so-called resolution bins. By analysing how these various statistics vary with resolution, an assessment of the overall dataset resolution can be made. Reflections should be included in a dataset only if they add meaningful information.

R_{merge} is one such statistic and describes the difference between intensities of individual reflections to the mean intensity of equivalent reflections, expressed as a fraction of the mean (Equation 6.3). This statistic is a measure of internal consistency of the data, however, because the errors are summed, an increase in multiplicity will also cause the R_{merge} value to rise. A similar statistic that takes into account the multiplicity of the data, R_{pim}, down-weights the errors associated with redundant reflections (Equation 6.4).

$$R_{merge} = \frac{\sum_{hkl} \sum_i |I_i - \langle I_{hkl} \rangle|}{\sum_{hkl} \sum_i \langle I_{hkl} \rangle}$$

Equation 6.3 - R_{merge} equation

$$R_{pim} = \frac{\sum_{hkl} \sqrt{\frac{1}{n-1}} \sum_i |I_i - \langle I_{hkl} \rangle|}{\sum_{hkl} \sum_i \langle I_{hkl} \rangle}$$

Equation 6.4 - R_{pim} equation

The correlation coefficient between two random half datasets is another statistic used to assess data quality and consistency. Unmerged reflections are split randomly into two half datasets with intensities then correlated and scored as a

fraction of perfect correlation. The mean signal-to-noise ratio (I/σI) is also an important factor to consider when assessing the quality and resolution of a dataset.

Processing of raw diffraction data into the scaled and merged datasets used in this thesis was predominantly carried out using XIA2 (Winter et al., 2013), an automated data processing pipeline. XIA2 allows the whole process of indexing, integration, space group determination, scaling, merging and conversion into structure factor amplitudes to be automated. It even truncates the resolution of the dataset to an appropriate level in a manner determined by the previously mentioned data quality and consistency statistics. Although it is possible to go from raw diffraction images to structure factor amplitudes in an automated manner, it is important to inspect the log files and to understand the decisions that have been made by XIA2. There are many reasons why it may be sometimes necessary to manually reprocess the data. One such example may involve the removal of a bad batch of images that would improve overall data quality. Another example may occur in the case of non-trivial space group determination where the data may need to be processed in an alternative space group.

6.1.3 Structure solution and generation of electron density maps

Each observed reflection in the diffraction pattern corresponds to a structure factor that describes the amplitude and phase of the scattered x-ray wave. Structure factors are related to the electron density in a relationship described in Equation 6.5 that represents a Fourier transformation from real to reciprocal space. They can also be described as the sum of the scattering by each atom present in the unit cell (Equation 6.6). A Fourier transform from reciprocal to real space as described in Equation 6.7, allows calculation electron density at each point in the unit cell.

$$F(h,k,l) = V\int_{x=0}^{1}\int_{y=0}^{1}\int_{z=0}^{1}\rho(x,y,z).\exp[2\pi i(hx+ky+lz)]\,.dxdydz$$

Equation 6.5 - Structure factor equation expressed in terms of electron density

$$F(h,k,l) = \sum_{j=1}^{atoms} f_j.\exp\left[2\pi i\left(hx_j+ky_j+lz_j\right)\right].\exp\left[-B_j\sin^2\theta/\lambda^2\right]$$

Equation 6.6 - Structure factor equation expressed in terms of contributions from each atom
f_j = atomic scattering factor, B_j = thermal factor

$$\rho(x,y,z) = \frac{1}{V}\sum_h\sum_k\sum_l|F(h,k,l)|.\exp\left[-2\pi i(hx+ky+lz) + i\alpha(h,k,l)\right]$$

Equation 6.7 - Electron density equation

In order to calculate an electron density map the amplitude and phase of each structure factor are required (Equation 6.8). The intensity of a given reflection measured in the diffraction pattern gives information about the amplitude of each structure factor (Equation 6.9), however it is not possible to measure the phase of such reflections and this results in the phase problem of crystallography.

$$F(h,k,l) = |F(h,k,l)|.\exp\left[i\alpha(h,k,l)\right]$$

Equation 6.8 - Structure factor described by its intensity and phase

$$I \propto |F(h,k,l)|^2$$

Equation 6.9 - Relationship between reflection intensity and structure factor amplitude

The phase problem of crystallography can be solved by either molecular replacement or experimental phasing techniques. It was possible to solve each

structure presented in this thesis by molecular replacement, so discussion will be limited to this technique.

Molecular replacement relies on the availability of previously solved structures with sufficient structural homology to the protein of interest. Patterson maps are calculated directly from the experimentally determined reflection intensities (Equation 6.10) with peaks describing interatomic vectors.

$$P(u,v,w) = \frac{1}{V}\sum_h \sum_k \sum_l |F(h,k,l)|^2 . \exp\left[-2\pi i(hx + ky + lz)\right]$$

Equation 6.10 - Patterson function

Molecular replacement involves fitting the Patterson map calculated using the experimental intensities to Patterson maps generated *in silico* for different orientations of the search model (Figure 6.5). This allows placement of the search model in the unit cell from which structure factor phases can be calculated. To reduce the computational requirements, the process is split into two steps, a rotation function search followed by a translation function search. In order to determine the correct rotation of the search model, only the vectors between atoms of individual protein molecules are taken into account and, as such, the rotation function search only includes Patterson peaks that fall within a radius (similar to that of the molecule) from the origin. The search model is rotated and a Patterson map calculated for each orientation (Figure 6.5C). These maps are then compared to the experimental Patterson map and scored for similarity. After the rotation has been determined, a translation function search is carried out to locate the position of the model in the unit cell. Patterson maps are calculated for each position of the model in the unit cell and then compared to the experimental map, this time taking into account Patterson peaks that represent intermolecular vectors (Figure 6.5D).

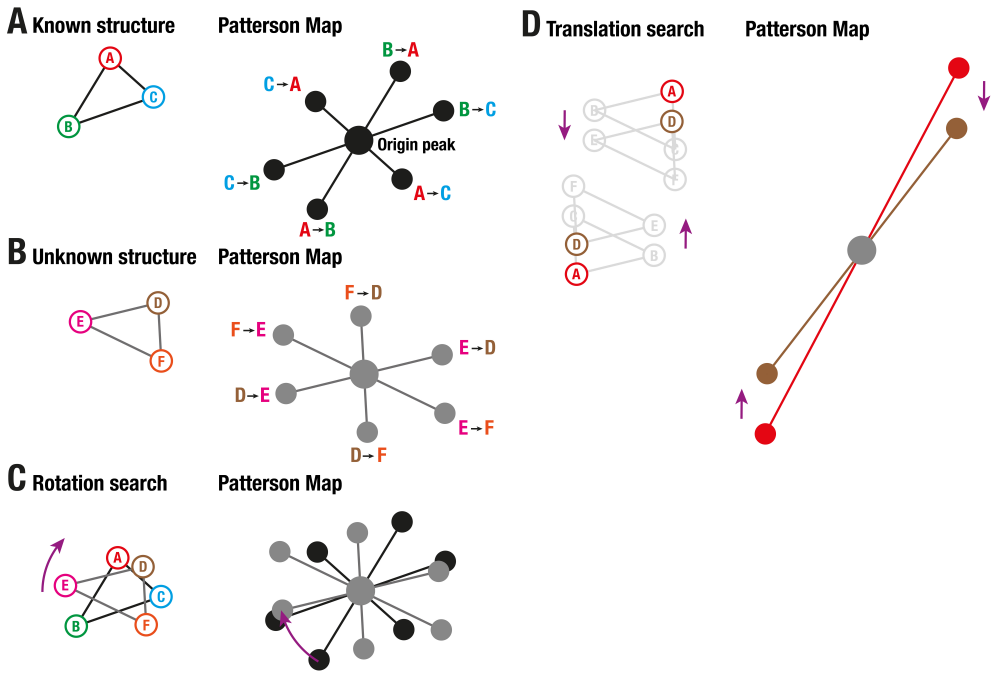

Figure 6.5 - Molecular replacement process
Panel A: Known molecular structure and calculated Patterson map. Panel B: Unknown molecular structure with Patterson map calculated using experimentally determined intensities. Panel C: Rotation of model of known structure. Panel D: Translational search. Patterson map simplified by only including peaks representing the vectors between atoms A or D of two identical molecules.

Molecular replacement carried out in this thesis used either MOLREP (Vagin and Teplyakov, 2010) or PHASER (McCoy et al., 2007). MOLREP (Vagin and Teplyakov, 2010) utilises Patterson-based rotation and translation searches in order to attempt placement of a search model into the unit cell. Potential molecular replacement solutions are scored based on the correlation coefficient between experimental Patterson map peak intensities and those generated by the potential solution and a packing function. This score is then converted in to a contrast value, which is the ratio of the top score to the mean score. The authors suggest a contrast score of over 2.5 represents a definite solution.

In contrast, PHASER (McCoy et al., 2007) uses algorithms based on maximum likelihood multivariate statistics, resulting in improvements over traditional Patterson-based methods in identifying correct molecular replacement solutions from noise. PHASER scores potential solutions using a log-likelihood-gain (LLG) function. This

function describes how much better the placement of the search model explains the experimental data compared to that of a random atom model. A Z-score is then calculated that gives a measure of how many standard deviations the LLG of a solution is above the LLG of randomly oriented search models. The authors suggest that a final translation function Z-score of over 8.5 represents a definite solution.

If the asymmetric unit contains more than one copy of a molecule, calculation of a self-rotation function (SRF) can reveal important information about the arrangement of the molecules with respect to one another. The Patterson map calculated using the experimentally observed intensities is rotated and superimposed onto itself around a particular axes described by spherical polar coordinates (Figure 6.6). The presence of crystallographic symmetry will result in a peak with an intensity equivalent to the origin peak, whereas a non-crystallographic symmetry axis will result in a peak less than 100% of the origin.

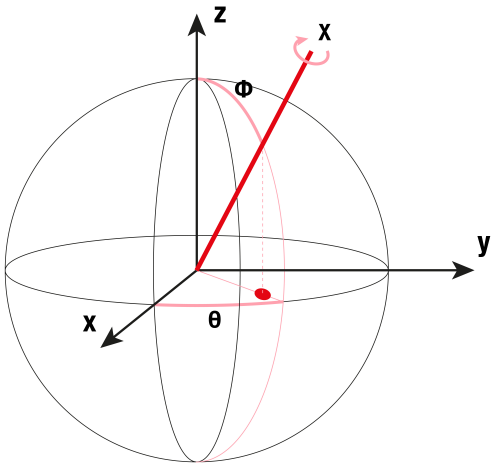

Figure 6.6 - Self-rotation function illustration

6.1.4 Refinement

The goal of refinement is to obtain a model that best describes the experimentally determined structure factor amplitudes. In order to achieve this, various model parameters (including positional coordinates, B-factors and occupancies) can be adjusted during refinement. Due to the inherently low observation-to-parameter

ratio, it is necessary to provide additional information in the form of restraints. Such restraints include bond lengths, dihedral angle and torsion angles, which are used to penalise the model if their values deviate too far from known mean values. In addition, non-crystallographic symmetry restraints can be applied if there is more than one copy of an identical molecule in the asymmetric unit.

Refining the model in this way leads to improved phases, which in turn lead to more detail in the calculated electron density maps. The model is then modified to more accurately fit this new electron density map and the process is repeated. Due to the iterative nature of the refinement process and the calculation of electron density maps using phases generated by the model, it is possible for model bias to occur.

The progress of a refinement is monitored by the R-factor (Equation 6.11), a statistical measure that compares observed structure factor amplitudes with those calculated by the model. The R-factor value decreases as the model better explains the observed data. To assess the extent of model bias, a small number of reflections (usually ~5%) are excluded from the refinement, allowing calculation of R_{free}. If the model is accurate and free from model bias, it will be able to predict the amplitudes of these missing reflections.

$$R = \frac{\sum_{hkl}\left||F_{obs}| - |F_{calc}|\right|}{\sum_{hkl}|F_{obs}|}$$

Equation 6.11 - R factor

Refinement carried out in this thesis used either REFMAC (Murshudov et al., 1997), AUTOBUSTER (Bricogne et al.) or PHENIX.REFINE (Adams et al., 2010).

References

Abidi, S.M.A., Saifullah, M.K., Zafiropulos, M.D., Kaput, C., Bowen, M.A., Cotton, C., and Singer, N.G. (2006). CD166 expression, characterization, and localization in salivary epithelium: implications for function during sialoadenitis. J. Clin. Immunol. *26*, 12–21.

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., et al. (2010). PHENIX : a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D Biol. Crystallogr. *66*, 213–221.

Aira, L.E., López-Requena, A., Fuentes, D., Sánchez, L., Pérez, T., Urquiza, A., Bautista, H., Falcón, L., Hernández, P., and Mazorra, Z. (2014). Immunological and histological evaluation of clinical samples from psoriasis patients treated with anti-CD6 itolizumab. mAbs *6*, 783–793.

Allen, G.S., and Stokes, D.L. (2013). Modeling, Docking, and Fitting of Atomic Structures to 3D Maps from Cryo-electron Microscopy. Methods Mol. Biol. Clifton NJ *955*, 229–241.

Alric, J., Tsukatani, Y., Yoshida, M., Matsuura, K., Shimada, K., Hienerwadel, R., Schoepp-Cothenet, B., Nitschke, W., Nagashima, K.V.P., and Verméglio, A. (2004). Structural and functional characterization of the unusual triheme cytochrome bound to the reaction center of Rhodovulum sulfidophilum. J. Biol. Chem. *279*, 26090–26097.

Amunts, A., Brown, A., Bai, X., Llácer, J.L., Hussain, T., Emsley, P., Long, F., Murshudov, G., Scheres, S.H.W., and Ramakrishnan, V. (2014). Structure of the Yeast Mitochondrial Large Ribosomal Subunit. Science *343*, 1485–1489.

Anderson, W.A., and Freeman, R. (1962). Influence of a Second Radiofrequency Field on High-Resolution Nuclear Magnetic Resonance Spectra. J. Chem. Phys. *37*, 85–103.

Anfinsen, C.B. (1973). Principles that govern the folding of protein chains. Science *181*, 223–230.

Appia-Ayme, C., Little, P.J., Matsumoto, Y., Leech, A.P., and Berks, B.C. (2001). Cytochrome complex essential for photosynthetic oxidation of both thiosulfate and sulfide in Rhodovulum sulfidophilum. J. Bacteriol. *183*, 6107–6118.

Arai, S., Maehara, N., Iwamura, Y., Honda, S., Nakashima, K., Kai, T., Ogishi, M., Morita, K., Kurokawa, J., Mori, M., et al. (2013). Obesity-associated autoantibody production requires AIM to retain the immunoglobulin M immune complex on follicular dendritic cells. Cell Rep. *3*, 1187–1198.

Aruffo, A., Bowen, M.A., Patel, D.D., Haynes, B.F., Starling, G.C., Gebe, J.A., and Bajorath, J. (1997). CD6-ligand interactions: a paradigm for SRCR domain function? Immunol. Today *18*, 498–504.

Ashkenazy, H., Erez, E., Martz, E., Pupko, T., and Ben-Tal, N. (2010). ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. Nucleic Acids Res. *38*, W529–W533.

Baker, L.A., and Rubinstein, J.L. (2010). Radiation damage in electron cryomicroscopy. Methods Enzymol. *481*, 371–388.

Bamford, V.A., Bruno, S., Rasmussen, T., Appia-Ayme, C., Cheesman, M.R., Berks, B.C., and Hemmings, A.M. (2002). Structural basis for the oxidation of thiosulfate by a sulfur cycle enzyme. EMBO J. *21*, 5599–5610.

Banchereau, J., and Steinman, R.M. (1998). Dendritic cells and the control of immunity. Nature *392*, 245–252.

Bardischewsky, F., Quentmeier, A., and Friedrich, C.G. (2006). The flavoprotein SoxF functions in chemotrophic thiosulfate oxidation of Paracoccus pantotrophus in vivo and in vitro. FEMS Microbiol. Lett. *258*, 121–126.

Beck-Sickinger, A.G., and Mörl, K. (2006). Posttranslational Modification of Proteins. Expanding Nature's Inventory. By Christopher T. Walsh. Angew. Chem. Int. Ed. *45*, 1020–1020.

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

Betz, S.F. (1993). Disulfide bonds and the stability of globular proteins. Protein Sci. Publ. Protein Soc. *2*, 1551–1558.

Bikker, F.J., Ligtenberg, A.J.M., End, C., Renner, M., Blaich, S., Lyer, S., Wittig, R., van't Hof, W., Veerman, E.C.I., Nazmi, K., et al. (2004). Bacteria binding by DMBT1/SAG/gp-340 is confined to the VEVLXXXXW motif in its scavenger receptor cysteine-rich domains. J. Biol. Chem. *279*, 47699–47703.

Birnbaum, M.E., Mendoza, J.L., Sethi, D.K., Dong, S., Glanville, J., Dobbins, J., Ozkan, E., Davis, M.M., Wucherpfennig, K.W., and Garcia, K.C. (2014a). Deconstructing the peptide-MHC specificity of T cell recognition. Cell *157*, 1073–1087.

Birnbaum, M.E., Berry, R., Hsiao, Y.-S., Chen, Z., Shingu-Vazquez, M.A., Yu, X., Waghray, D., Fischer, S., McCluskey, J., Rossjohn, J., et al. (2014b). Molecular architecture of the αβ T cell receptor-CD3 complex. Proc. Natl. Acad. Sci. U. S. A. *111*, 17576–17581.

Bjorkman, P.J., Saper, M.A., Samraoui, B., Bennett, W.S., Strominger, J.L., and Wiley, D.C. (1987a). Structure of the human class I histocompatibility antigen, HLA-A2. Nature *329*, 506–512.

Bjorkman, P.J., Saper, M.A., Samraoui, B., Bennett, W.S., Strominger, J.L., and Wiley, D.C. (1987b). The foreign antigen binding site and T cell recognition regions of class I histocompatibility antigens. Nature *329*, 512–518.

Blackwell, J.M., Jamieson, S.E., and Burgner, D. (2009). HLA and Infectious Diseases. Clin. Microbiol. Rev. *22*, 370–385.

Bluestone, J.A., Mackay, C.R., O'Shea, J.J., and Stockinger, B. (2009). The functional plasticity of T cell subsets. Nat. Rev. Immunol. *9*, 811–816.

Bodian, D.L., Skonier, J.E., Bowen, M.A., Neubauer, M., Siadak, A.W., Aruffo, A., and Bajorath, J. (1997). Identification of residues in CD6 which are critical for ligand binding. Biochemistry (Mosc.) *36*, 2637–2641.

Bordo, D., Deriu, D., Colnaghi, R., Carpen, A., Pagani, S., and Bolognesi, M. (2000). The crystal structure of a sulfurtransferase from Azotobacter vinelandii highlights the evolutionary relationship between the rhodanese and phosphatase enzyme families. J. Mol. Biol. *298*, 691–704.

Bott, C.M., Doshi, J.B., Morimoto, C., Romain, P.L., and Fox, D.A. (1993). Activation of human T cells through CD6: functional effects of a novel anti-CD6 monoclonal antibody and definition of four epitopes of the CD6 glycoprotein. Int. Immunol. *5*, 783–792.

Bowen, M.A., Patel, D.D., Li, X., Modrell, B., Malacko, A.R., Wang, W.C., Marquardt, H., Neubauer, M., Pesando, J.M., and Francke, U. (1995). Cloning, mapping, and characterization of activated leukocyte-cell adhesion molecule (ALCAM), a CD6 ligand. J. Exp. Med. *181*, 2213–2220.

Bowen, M.A., Bajorath, J., Siadak, A.W., Modrell, B., Malacko, A.R., Marquardt, H., Nadler, S.G., and Aruffo, A. (1996). The amino-terminal immunoglobulin-like domain of activated leukocyte cell adhesion molecule binds specifically to the membrane-proximal scavenger receptor cysteine-rich domain of CD6 with a 1:1 stoichiometry. J. Biol. Chem. *271*, 17390–17396.

Bowen, M.A., Bajorath, J., D'Egidio, M., Whitney, G.S., Palmer, D., Kobarg, J., Starling, G.C., Siadak, A.W., and Aruffo, A. (1997). Characterization of mouse ALCAM (CD166): the CD6-binding domain is conserved in different homologs and mediates cross-species binding. Eur. J. Immunol. *27*, 1469–1478.

Bricogne, G., Blanc, E., Brandl, M., Flensburg, C., Keller, P., Paciorek, W., Roversi, P., Sharff, A., Smart, O.S., Vonrhein, C., et al. BUSTER version 2.10.1. Glob. Phasing Ltd.

Brito, J.A., Denkmann, K., Pereira, I.A.C., Archer, M., and Dahl, C. (2015). Thiosulfate dehydrogenase (TsdA) from Allochromatium vinosum: structural and functional insights into thiosulfate oxidation. J. Biol. Chem. *290*, 9222–9238.

Brown, M.H., and Lacey, E. (2010). A ligand for CD5 is CD5. J. Immunol. Baltim. Md 1950 *185*, 6068–6074.

Brusic, V., Bajic, V.B., and Petrovsky, N. (2004). Computational methods for prediction of T-cell epitopes--a framework for modelling, testing, and applications. Methods San Diego Calif *34*, 436–443.

Budamakuntla, L., Madaiah, M., Sarvajnamurthy, S., and Kapanigowda, S. (2015). Itolizumab provides sustained remission in plaque psoriasis: a 5-year follow-up experience. Clin. Exp. Dermatol. *40*, 152–155.

Cabeen, M.T., and Jacobs-Wagner, C. (2010). The Bacterial Cytoskeleton. Annu. Rev. Genet. *44*, 365–392.

Call, M.E., and Wucherpfennig, K.W. (2005). THE T CELL RECEPTOR: Critical Role of the Membrane Environment in Receptor Assembly and Function. Annu. Rev. Immunol. *23*, 101–125.

Campbell, M.G., Cheng, A., Brilot, A.F., Moeller, A., Lyumkis, D., Veesler, D., Pan, J., Harrison, S.C., Potter, C.S., Carragher, B., et al. (2012). Movies of Ice-Embedded Particles Enhance Resolution in Electron Cryo-Microscopy. Structure *20*, 1823–1828.

Caramelo, J.J., Castro, O.A., de Prat-Gay, G., and Parodi, A.J. (2004). The endoplasmic reticulum glucosyltransferase recognizes nearly native glycoprotein folding intermediates. J. Biol. Chem. *279*, 46280–46285.

Carrington, M., Nelson, G.W., Martin, M.P., Kissner, T., Vlahov, D., Goedert, J.J., Kaslow, R., Buchbinder, S., Hoots, K., and O'Brien, S.J. (1999). HLA and HIV-1: heterozygote advantage and B*35-Cw*04 disadvantage. Science *283*, 1748–1752.

Cascio, P., Hilton, C., Kisselev, A.F., Rock, K.L., and Goldberg, A.L. (2001). 26S proteasomes and immunoproteasomes produce mainly N-extended versions of an antigenic peptide. EMBO J. *20*, 2357–2366.

Castro, M.A.A., Oliveira, M.I., Nunes, R.J., Fabre, S., Barbosa, R., Peixoto, A., Brown, M.H., Parnes, J.R., Bismuth, G., Moreira, A., et al. (2007). Extracellular isoforms of CD6 generated by alternative splicing regulate targeting of CD6 to the immunological synapse. J. Immunol. Baltim. Md 1950 *178*, 4351–4361.

Cayrol, R., Wosik, K., Berard, J.L., Dodelet-Devillers, A., Ifergan, I., Kebir, H., Haqqani, A.S., Kreymborg, K., Krug, S., Moumdjian, R., et al. (2008). Activated leukocyte cell adhesion molecule promotes leukocyte trafficking into the central nervous system. Nat. Immunol. *9*, 137–145.

Chayen, N.E., and Saridakis, E. (2008). Protein crystallization: from purified protein to diffraction-quality crystal. Nat. Methods *5*, 147–153.

Chen, G.Y., and Nuñez, G. (2010). Sterile inflammation: sensing and reacting to damage. Nat. Rev. Immunol. *10*, 826–837.

Chen, L., and Flies, D.B. (2013). Molecular mechanisms of T cell co-stimulation and co-inhibition. Nat. Rev. Immunol. *13*, 227–242.

Chen, V.B., Arendall, W.B., Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr. D Biol. Crystallogr. *66*, 12–21.

Chen, Z.W., Koh, M., Van Driessche, G., Van Beeumen, J.J., Bartsch, R.G., Meyer, T.E., Cusanovich, M.A., and Mathews, F.S. (1994). The structure of flavocytochrome c sulfide dehydrogenase from a purple phototrophic bacterium. Science *266*, 430–432.

Chitteti, B.R., Bethel, M., Kacena, M.A., and Srour, E.F. (2013). CD166 and regulation of hematopoiesis. Curr. Opin. Hematol. *20*, 273–280.

Ciofani, M., and Zúñiga-Pflücker, J.C. (2010). Determining γδ versus αβ T cell development. Nat. Rev. Immunol. *10*, 657–663.

Cohen, N.R., Garg, S., and Brenner, M.B. (2009). Antigen Presentation by CD1 Lipids, T Cells, and NKT Cells in Microbial Immunity. Adv. Immunol. *102*, 1–94.

Cowtan, K. (2006). The Buccaneer software for automated model building. 1. Tracing protein chains. Acta Crystallogr. D Biol. Crystallogr. *62*, 1002–1011.

Crick, F. (1970). Central dogma of molecular biology. Nature *227*, 561–563.

Crosby, J., and Crump, M.P. (2012). The structural role of the carrier protein--active controller or passive carrier. Nat. Prod. Rep. *29*, 1111–1137.

Crotty, S. (2015). A brief history of T cell help to B cells. Nat. Rev. Immunol. *15*, 185–189.

Dahl, C., Engels, S., Pott-Sperling, A.S., Schulte, A., Sander, J., Lübbe, Y., Deuster, O., and Brune, D.C. (2005). Novel genes of the dsr gene cluster and evidence for close interaction of Dsr proteins during sulfur oxidation in the phototrophic sulfur bacterium Allochromatium vinosum. J. Bacteriol. *187*, 1392–1404.

Dahl, C., Franz, B., Hensen, D., Kesselheim, A., and Zigann, R. (2013). Sulfite oxidation in the purple sulfur bacterium Allochromatium vinosum: identification of SoeABC as a major player and relevance of SoxYZ in the process. Microbiol. Read. Engl. *159*, 2626–2638.

Dambe, T., Quentmeier, A., Rother, D., Friedrich, C., and Scheidig, A.J. (2005). Structure of the cytochrome complex SoxXA of Paracoccus pantotrophus, a heme enzyme initiating chemotrophic sulfur oxidation. J. Struct. Biol. *152*, 229–234.

Danev, R., Buijsse, B., Khoshouei, M., Plitzko, J.M., and Baumeister, W. (2014). Volta potential phase plate for in-focus phase contrast transmission electron microscopy. Proc. Natl. Acad. Sci. *111*, 15635–15640.

Denkmann, K., Grein, F., Zigann, R., Siemen, A., Bergmann, J., van Helmont, S., Nicolai, A., Pereira, I.A.C., and Dahl, C. (2012). Thiosulfate dehydrogenase: a widespread unusual acidophilic c-type cytochrome. Environ. Microbiol. *14*, 2673–2688.

Denzin, L.K., and Cresswell, P. (1995). HLA-DM induces CLIP dissociation from MHC class II alpha beta dimers and facilitates peptide loading. Cell *82*, 155–165.

Derewenda, Z.S. (2004). The use of recombinant methods and molecular engineering in protein crystallization. Methods San Diego Calif *34*, 354–363.

Derewenda, Z.S., and Vekilov, P.G. (2006). Entropy and surface engineering in protein crystallization. Acta Crystallogr. D Biol. Crystallogr. *62*, 116–124.

Destro, R., Roversi, P., Barzaghi, M., and Marsh, R.E. (2000). Experimental Charge Density of α-Glycine at 23 K. J. Phys. Chem. A *104*, 1047–1054.

Dolinsky, T.J., Nielsen, J.E., McCammon, J.A., and Baker, N.A. (2004). PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. Nucleic Acids Res. *32*, W665–W667.

Dolinsky, T.J., Czodrowski, P., Li, H., Nielsen, J.E., Jensen, J.H., Klebe, G., and Baker, N.A. (2007). PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. Nucleic Acids Res. *35*, W522–W525.

Dong, G., Wearsch, P.A., Peaper, D.R., Cresswell, P., and Reinisch, K.M. (2009). Insights into MHC class I peptide loading from the structure of the tapasin-ERp57 thiol oxidoreductase heterodimer. Immunity *30*, 21–32.

Dushek, O., Goyette, J., and van der Merwe, P.A. (2012). Non-catalytic tyrosine-phosphorylated receptors. Immunol. Rev. *250*, 258–276.

Dyda, F., Hickman, A.B., Jenkins, T.M., Engelman, A., Craigie, R., and Davies, D.R. (1994). Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases. Science *266*, 1981–1986.

E. Eikenberry, C.B. (2002). Pilatus: a 2-dimensional X-ray detector for macromolecular crystallography. Acta Crystallogr. Sect. - ACTA CRYSTALLOGR A *58*.

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

Endicott, J.A., Noble, M.E.M., and Johnson, L.N. (2012). The Structural Basis for Control of Eukaryotic Protein Kinases. Annu. Rev. Biochem. *81*, 587–613.

Fan, S.W., George, R.A., Haworth, N.L., Feng, L.L., Liu, J.Y., and Wouters, M.A. (2009). Conformational changes in redox pairs of protein structures. Protein Sci. Publ. Protein Soc. *18*, 1745–1765.

Fass, D. (2012). Disulfide Bonding in Protein Biophysics. Annu. Rev. Biophys. *41*, 63–79.

Firebaugh, M. (1988). Artificial intelligence: a knowledge-based approach (Boston: Boyd & Fraser).

Fischer, E., and Krebs, E. (1955). Conversion of Phosphorylase-B to Phosphorylase-a in Muscle Extracts. J. Biol. Chem. *216*, 121–132.

Flory, P.J. (1956). Theory of Elastic Mechanisms in Fibrous Proteins. J. Am. Chem. Soc. *78*, 5222–5235.

Fremont, D.H., Matsumura, M., Stura, E.A., Peterson, P.A., and Wilson, I.A. (1992). Crystal structures of two viral peptides in complex with murine MHC class I H-2Kb. Science *257*, 919–927.

Friedrich, C.G., Rother, D., Bardischewsky, F., Quentmeier, A., and Fischer, J. (2001). Oxidation of reduced inorganic sulfur compounds by bacteria: emergence of a common mechanism? Appl. Environ. Microbiol. *67*, 2873–2882.

Frigaard, N.-U., and Dahl, C. (2008). Sulfur Metabolism in Phototrophic Sulfur Bacteria. In Advances in Microbial Physiology, R.K. Poole, ed. (Academic Press), pp. 103–200.

Frigaard, N.-U., and Dahl, C. (2009). Sulfur metabolism in phototrophic sulfur bacteria. Adv. Microb. Physiol. *54*, 103–200.

Frueh, D.P., Goodrich, A., Mishra, S., and Nichols, S. (2013). NMR methods for structural studies of large monomeric and multimeric proteins. Curr. Opin. Struct. Biol. *23*, 734–739.

Fugmann, S.D., Lee, A.I., Shockett, P.E., Villey, I.J., and Schatz, D.G. (2000). The RAG proteins and V(D)J recombination: complexes, ends, and transposition. Annu. Rev. Immunol. *18*, 495–527.

Fujii, S.-I., Shimizu, K., Smith, C., Bonifaz, L., and Steinman, R.M. (2003). Activation of natural killer T cells by alpha-galactosylceramide rapidly induces the full maturation of dendritic cells in vivo and thereby acts as an adjuvant for combined CD4 and CD8 T cell immunity to a coadministered protein. J. Exp. Med. *198*, 267–279.

Gabathuler, R., Reid, G., Kolaitis, G., Driscoll, J., and Jefferies, W.A. (1994). Comparison of cell lines deficient in antigen presentation reveals a functional role for TAP-1 alone in antigen processing. J. Exp. Med. *180*, 1415–1425.

Gangemi, R.M., Swack, J.A., Gaviria, D.M., and Romain, P.L. (1989). Anti-T12, an anti-CD6 monoclonal antibody, can activate human T lymphocytes. J. Immunol. Baltim. Md 1950 *143*, 2439–2447.

Ganz, T. (2003). Defensins: antimicrobial peptides of innate immunity. Nat. Rev. Immunol. *3*, 710–720.

Garbi, N., Tan, P., Diehl, A.D., Chambers, B.J., Ljunggren, H.G., Momburg, F., and Hämmerling, G.J. (2000). Impaired immune responses and altered peptide repertoire in tapasin-deficient mice. Nat. Immunol. *1*, 234–238.

Garcia, K.C., and Adams, E.J. (2005). How the T cell receptor sees antigen--a structural view. Cell *122*, 333–336.

Gardiner, C.M., and Mills, K.H.G. The cells that mediate innate immune memory and their functional significance in inflammatory and infectious diseases. Semin. Immunol.

Garman, E., and Owen, R.L. (2007). Cryocrystallography of macromolecules: practice and optimization. Methods Mol. Biol. Clifton NJ *364*, 1–18.

Garza-Garcia, A., Esposito, D., Rieping, W., Harris, R., Briggs, C., Brown, M.H., and Driscoll, P.C. (2008). Three-dimensional solution structure and conformational plasticity of the N-terminal scavenger receptor cysteine-rich domain of human CD5. J. Mol. Biol. *378*, 129–144.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D., and Bairoch, A. (2005). Protein Identification and Analysis Tools on the ExPASy Server (Humana Press).

Gaudet, R., and Wiley, D.C. (2001). Structure of the ABC ATPase domain of human TAP1, the transporter associated with antigen processing. EMBO J. *20*, 4964–4972.

Ghosh, P., Amaya, M., Mellins, E., and Wiley, D.C. (1995). The structure of an intermediate in class II MHC maturation: CLIP bound to HLA-DR3. Nature *378*, 457–462.

Gimferrer, I., Calvo, M., Mittelbrunn, M., Farnós, M., Sarrias, M.R., Enrich, C., Vives, J., Sánchez-Madrid, F., and Lozano, F. (2004). Relevance of CD6-mediated interactions in T cell activation and proliferation. J. Immunol. Baltim. Md 1950 *173*, 2262–2270.

Gliubich, F., Gazerro, M., Zanotti, G., Delbono, S., Bombieri, G., and Berni, R. (1996). Active site structural features for chemically modified forms of rhodanese. J. Biol. Chem. *271*, 21054–21061.

Goulder, P.J.R., and Walker, B.D. (2012). HIV and HLA class I: an evolving relationship. Immunity *37*, 426–440.

Grandea, A.G., Golovina, T.N., Hamilton, S.E., Sriram, V., Spies, T., Brutkiewicz, R.R., Harty, J.T., Eisenlohr, L.C., and Van Kaer, L. (2000). Impaired assembly yet normal trafficking of MHC class I molecules in Tapasin mutant mice. Immunity *13*, 213–222.

Grassucci, R.A., Taylor, D.J., and Frank, J. (2007). Preparation of macromolecular complexes for cryo-electron microscopy. Nat. Protoc. *2*, 3239–3246.

Gregersen, L.H., Bryant, D.A., and Frigaard, N.-U. (2011). Mechanisms and evolution of oxidative sulfur metabolism in green sulfur bacteria. Front. Microbiol. *2*, 116.

Grein, F., Venceslau, S.S., Schneider, L., Hildebrandt, P., Todorovic, S., Pereira, I.A.C., and Dahl, C. (2010). DsrJ, an essential part of the DsrMKJOP transmembrane complex in the purple sulfur bacterium Allochromatium vinosum, is an unusual triheme cytochrome c. Biochemistry (Mosc.) *49*, 8290–8299.

Guermonprez, P., Valladeau, J., Zitvogel, L., Théry, C., and Amigorena, S. (2002). Antigen presentation and T cell stimulation by dendritic cells. Annu. Rev. Immunol. *20*, 621–667.

Guillet, J.G., Lai, M.Z., Briner, T.J., Smith, J.A., and Gefter, M.L. (1986). Interaction of peptide antigens and class II major histocompatibility complex antigens. Nature *324*, 260–262.

Guo, H.C., Jardetzky, T.S., Garrett, T.P., Lane, W.S., Strominger, J.L., and Wiley, D.C. (1992). Different length peptides bind to HLA-Aw68 similarly at their ends but bulge out in the middle. Nature *360*, 364–366.

Hafler, D.A., Fallis, R.J., Dawson, D.M., Schlossman, S.F., Reinherz, E.L., and Weiner, H.L. (1986). Immunologic responses of progressive multiple sclerosis patients treated with an anti-T-cell monoclonal antibody, anti-T12. Neurology *36*, 777–784.

Hahn, E.L., and Maxwell, D.E. (1952). Spin Echo Measurements of Nuclear Spin Coupling in Molecules. Phys. Rev. *88*, 1070–1084.

Hammer, G.E., Gonzalez, F., James, E., Nolla, H., and Shastri, N. (2007). In the absence of aminopeptidase ERAAP, MHC class I molecules present many unstable and highly immunogenic peptides. Nat. Immunol. *8*, 101–108.

Harty, J.T., Tvinnereim, A.R., and White, D.W. (2000). CD8+ T Cell Effector Mechanisms in Resistance to Infection. Annu. Rev. Immunol. *18*, 275–308.

Hassan, N.J., Barclay, A.N., and Brown, M.H. (2004). Frontline: Optimal T cell activation requires the engagement of CD6 and CD166. Eur. J. Immunol. *34*, 930–940.

Hassan, N.J., Simmonds, S.J., Clarkson, N.G., Hanrahan, S., Puklavec, M.J., Bomb, M., Barclay, A.N., and Brown, M.H. (2006). CD6 regulates T-cell responses through

activation-dependent recruitment of the positive regulator SLP-76. Mol. Cell. Biol. *26*, 6727–6738.

Hebert, D.N., Garman, S.C., and Molinari, M. (2005). The glycan code of the endoplasmic reticulum: asparagine-linked carbohydrates as protein maturation and quality-control tags. Trends Cell Biol. *15*, 364–370.

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

Hensen, D., Sperling, D., Trüper, H.G., Brune, D.C., and Dahl, C. (2006). Thiosulphate oxidation in the phototrophic sulphur bacterium Allochromatium vinosum. Mol. Microbiol. *62*, 794–810.

Hermans, I.F., Silk, J.D., Gileadi, U., Salio, M., Mathew, B., Ritter, G., Schmidt, R., Harris, A.L., Old, L., and Cerundolo, V. (2003). NKT cells enhance CD4+ and CD8+ T cell responses to soluble antigen in vivo through direct interaction with dendritic cells. J. Immunol. Baltim. Md 1950 *171*, 5140–5147.

Herter, S., Piper, D.E., Aaron, W., Gabriele, T., Cutler, G., Cao, P., Bhatt, A.S., Choe, Y., Craik, C.S., Walker, N., et al. (2005). Hepatocyte growth factor is a preferred in vitro substrate for human hepsin, a membrane-anchored serine protease implicated in prostate and ovarian cancers. Biochem. J. *390*, 125–136.

Hewitt, E.W. (2003). The MHC class I antigen presentation pathway: strategies for viral immune evasion. Immunology *110*, 163–169.

Hill, A.V. (1998). The immunogenetics of human infectious diseases. Annu. Rev. Immunol. *16*, 593–617.

Hinz, A., and Tampé, R. (2012). ABC transporters and immunity: mechanism of self-defense. Biochemistry (Mosc.) *51*, 4981–4989.

Hohenester, E., Sasaki, T., and Timpl, R. (1999). Crystal structure of a scavenger receptor cysteine-rich domain sheds light on an ancient superfamily. Nat. Struct. Biol. *6*, 228–232.

Howarth, M., Williams, A., Tolstrup, A.B., and Elliott, T. (2004). Tapasin enhances MHC class I peptide presentation according to peptide half-life. Proc. Natl. Acad. Sci. U. S. A. *101*, 11737–11742.

Hsu, H., Chen, C., Nenninger, A., Holz, L., Baldwin, C.L., and Telfer, J.C. (2015). WC1 is a hybrid γδ TCR coreceptor and pattern recognition receptor for pathogenic bacteria. J. Immunol. Baltim. Md 1950 *194*, 2280–2288.

Huang, J., Brameshuber, M., Zeng, X., Xie, J., Li, Q., Chien, Y., Valitutti, S., and Davis, M.M. (2013). A single peptide-major histocompatibility complex ligand triggers digital cytokine secretion in CD4(+) T cells. Immunity *39*, 846–857.

Hübscher, U., Maga, G., and Spadari, S. (2002). Eukaryotic Dna Polymerases. Annu. Rev. Biochem. *71*, 133–163.

Hudson, B.P., Quispe, J., Lara-Gonzalez, S., Kim, Y., Berman, H.M., Arnold, E., Ebright, R.H., and Lawson, C.L. (2009). Three-dimensional EM structure of an intact

activator-dependent transcription initiation complex. Proc. Natl. Acad. Sci. *106*, 19830–19835.

Hui, R., and Edwards, A. (2003). High-throughput protein crystallization. J. Struct. Biol. *142*, 154–161.

Humphrey, S.J., James, D.E., and Mann, M. (2015). Protein Phosphorylation: A Major Switch Mechanism for Metabolic Regulation. Trends Endocrinol. Metab. *26*, 676–687.

Janeway, C.A., and Medzhitov, R. (2002). Innate immune recognition. Annu. Rev. Immunol. *20*, 197–216.

Jayaraman, K. (2013). Biocon's first-in-class anti-CD6 mAb reaches the market. Nat. Biotechnol. *31*, 1062–1063.

John L. Rubinstein, M.A.B. (2014). Alignment of cryo-EM movies of individual particles by global optimization of image translations.

Johnston, R.J., Poholek, A.C., DiToro, D., Yusuf, I., Eto, D., Barnett, B., Dent, A.L., Craft, J., and Crotty, S. (2009). Bcl6 and Blimp-1 are reciprocal and antagonistic regulators of T follicular helper cell differentiation. Science *325*, 1006–1010.

Jorritsma, A., Schumacher, T.N.M., and Haanen, J.B.A.G. (2009). Immunotherapeutic strategies: the melanoma example. Immunotherapy *1*, 679–690.

Kantardjieff, K.A., and Rupp, B. (2003). Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein–nucleic acid complex crystals. Protein Sci. *12*, 1865–1871.

Kappler, U., Bernhardt, P.V., Kilmartin, J., Riley, M.J., Teschner, J., McKenzie, K.J., and Hanson, G.R. (2008). SoxAX cytochromes, a new type of heme copper protein involved in bacterial energy generation from sulfur compounds. J. Biol. Chem. *283*, 22206–22214.

Katz, B.A., Luong, C., Ho, J.D., Somoza, J.R., Gjerstad, E., Tang, J., Williams, S.R., Verner, E., Mackman, R.L., Young, W.B., et al. (2004). Dissecting and designing inhibitor selectivity determinants at the S1 site using an artificial Ala190 protease (Ala190 uPA). J. Mol. Biol. *344*, 527–547.

Kaufman, J., and Salomonsen, J. (1997). The "minimal essential MHC" revisited: both peptide-binding and cell surface expression level of MHC molecules are polymorphisms selected by pathogens in chickens. Hereditas *127*, 67–73.

Kaufman, J., Völk, H., and Wallny, H.J. (1995). A "minimal essential Mhc" and an "unrecognized Mhc": two extremes in selection for polymorphism. Immunol. Rev. *143*, 63–88.

Kaufman, J., Milne, S., Göbel, T.W., Walker, B.A., Jacob, J.P., Auffray, C., Zoorob, R., and Beck, S. (1999a). The chicken B locus is a minimal essential major histocompatibility complex. Nature *401*, 923–925.

Kaufman, J., Jacob, J., Shaw, I., Walker, B., Milne, S., Beck, S., and Salomonsen, J. (1999b). Gene organisation determines evolution of function in the chicken MHC. Immunol. Rev. *167*, 101–117.

van Kempen, L.C.L.T., Meier, F., Egeblad, M., Kersten-Niessen, M.J.F., Garbe, C., Weidle, U.H., Van Muijen, G.N.P., Herlyn, M., Bloemers, H.P.J., and Swart, G.W.M. (2004). Truncation of activated leukocyte cell adhesion molecule: a gateway to melanoma metastasis. J. Invest. Dermatol. *122*, 1293–1301.

van Kempen, L.C., Nelissen, J.M., Degen, W.G., Torensma, R., Weidle, U.H., Bloemers, H.P., Figdor, C.G., and Swart, G.W. (2001). Molecular basis for the homophilic activated leukocyte cell adhesion molecule (ALCAM)-ALCAM interaction. J. Biol. Chem. *276*, 25783–25790.

Kendrew, J.C., Bodo, G., Dintzis, H.M., Parrish, R.G., Wyckoff, H., and Phillips, D.C. (1958). A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis. Nature *181*, 662–666.

Kijima, N., Hosen, N., Kagawa, N., Hashimoto, N., Nakano, A., Fujimoto, Y., Kinoshita, M., Sugiyama, H., and Yoshimine, T. (2012). CD166/activated leukocyte cell adhesion molecule is expressed on glioblastoma progenitor cells and involved in the regulation of tumor cell invasion. Neuro-Oncol. *14*, 1254–1264.

Kilmartin, J.R., Maher, M.J., Krusong, K., Noble, C.J., Hanson, G.R., Bernhardt, P.V., Riley, M.J., and Kappler, U. (2011). Insights into structure and function of the active site of SoxAX cytochromes. J. Biol. Chem. *286*, 24872–24881.

van Kilsdonk, J.W.J., Wilting, R.H., Bergers, M., van Muijen, G.N.P., Schalkwijk, J., van Kempen, L.C.L.T., and Swart, G.W.M. (2008). Attenuation of melanoma invasion by a secreted variant of activated leukocyte cell adhesion molecule. Cancer Res. *68*, 3671–3679.

Kim, S.T., Shin, Y., Brazin, K., Mallis, R.J., Sun, Z.-Y.J., Wagner, G., Lang, M.J., and Reinherz, E.L. (2012). TCR Mechanobiology: Torques and Tunable Structures Linked to Early T Cell Signaling. Front. Immunol. *3*, 76.

Koch, J., Guntrum, R., Heintke, S., Kyritsis, C., and Tampé, R. (2004). Functional dissection of the transmembrane domains of the transporter associated with antigen processing (TAP). J. Biol. Chem. *279*, 10142–10147.

Koch, M., Stronge, V.S., Shepherd, D., Gadola, S.D., Mathew, B., Ritter, G., Fersht, A.R., Besra, G.S., Schmidt, R.R., Jones, E.Y., et al. (2005). The crystal structure of human CD1d with and without α-galactosylceramide. Nat. Immunol. *6*, 819–826.

Koch, M., Camp, S., Collen, T., Avila, D., Salomonsen, J., Wallny, H.-J., van Hateren, A., Hunt, L., Jacob, J.P., Johnston, F., et al. (2007). Structures of an MHC class I molecule from B21 chickens illustrate promiscuous peptide binding. Immunity *27*, 885–899.

Koch, M., Chitayat, S., Dattilo, B.M., Schiefner, A., Diez, J., Chazin, W.J., and Fritz, G. (2010). Structural basis for ligand recognition and activation of RAGE. Struct. Lond. Engl. 1993 *18*, 1342–1352.

Kofler, D.M., Severson, C.A., Mousissian, N., De Jager, P.L., and Hafler, D.A. (2011). The CD6 multiple sclerosis susceptibility allele is associated with alterations in CD4+ T cell proliferation. J. Immunol. Baltim. Md 1950 *187*, 3286–3291.

Koopmann, J.O., Post, M., Neefjes, J.J., Hämmerling, G.J., and Momburg, F. (1996). Translocation of long peptides by transporters associated with antigen processing (TAP). Eur. J. Immunol. *26*, 1720–1728.

Košmrlj, A., Read, E.L., Qi, Y., Allen, T.M., Altfeld, M., Deeks, S.G., Pereyra, F., Carrington, M., Walker, B.D., and Chakraborty, A.K. (2010). Effects of thymic selection of the T-cell repertoire on HLA class I-associated control of HIV infection. Nature *465*, 350–354.

Krebs, E., and Fischer, E. (1956). The Phosphorylase-B to Phosphorylase-a Converting Enzyme of Rabbit Skeletal Muscle. Biochim. Biophys. Acta *20*, 150–157.

Krissinel, E., and Henrick, K. (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallogr. D Biol. Crystallogr. *60*, 2256–2268.

Krissinel, E., and Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. J. Mol. Biol. *372*, 774–797.

Kristiansen, M., Graversen, J.H., Jacobsen, C., Sonne, O., Hoffman, H.J., Law, S.K., and Moestrup, S.K. (2001). Identification of the haemoglobin scavenger receptor. Nature *409*, 198–201.

Krivov, G.G., Shapovalov, M.V., and Dunbrack, R.L. (2009). Improved prediction of protein side-chain conformations with SCWRL4. Proteins *77*, 778–795.

Kühlbrandt, W. (2014). Cryo-EM enters a new era. eLife *3*, e03678.

Kunisawa, J., and Shastri, N. (2003). The group II chaperonin TRiC protects proteolytic intermediates from degradation in the MHC class I antigen processing pathway. Mol. Cell *12*, 565–576.

Kunisawa, J., and Shastri, N. (2006). Hsp90alpha chaperones large C-terminally extended proteolytic intermediates in the MHC class I antigen processing pathway. Immunity *24*, 523–534.

Lampen, M.H., Verweij, M.C., Querido, B., van der Burg, S.H., Wiertz, E.J.H.J., and van Hall, T. (2010). CD8+ T cell responses against TAP-inhibited cells are readily detected in the human population. J. Immunol. Baltim. Md 1950 *185*, 6508–6517.

Langer, G., Cohen, S.X., Lamzin, V.S., and Perrakis, A. (2008). Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. Nat. Protoc. *3*, 1171–1179.

Lanier, L.L. (2008). Up on the tightrope: natural killer cell activation and inhibition. Nat. Immunol. *9*, 495–502.

Lapinaite, A., Simon, B., Skjaerven, L., Rakwalska-Bange, M., Gabel, F., and Carlomagno, T. (2013). The structure of the box C/D enzyme reveals regulation of RNA methylation. Nature *502*, 519–523.

Lea, S.M., and Johnson, S. (2012). Putting the structure into complement. Immunobiology *217*, 1117–1121.

Leibundgut, M., Maier, T., Jenni, S., and Ban, N. (2008). The multienzyme architecture of eukaryotic fatty acid synthases. Curr. Opin. Struct. Biol. *18*, 714–725.

Leonhardt, R.M., Keusekotten, K., Bekpen, C., and Knittler, M.R. (2005). Critical role for the tapasin-docking site of TAP2 in the functional integrity of the MHC class I-peptide-loading complex. J. Immunol. Baltim. Md 1950 *175*, 5104–5114.

Lepault, J., Booy, F.P., and Dubochet, J. (1983). Electron microscopy of frozen biological suspensions. J. Microsc. *129*, 89–102.

Levesque, M.C., Heinly, C.S., Whichard, L.P., and Patel, D.D. (1998). Cytokine-regulated expression of activated leukocyte cell adhesion molecule (CD166) on monocyte-lineage cells and in rheumatoid arthritis synovium. Arthritis Rheum. *41*, 2221–2229.

Li, X., Mooney, P., Zheng, S., Booth, C., Braunfeld, M.B., Gubbens, S., Agard, D.A., and Cheng, Y. (2013). Electron counting and beam-induced motion correction enable near atomic resolution single particle cryoEM. Nat. Methods *10*, 584–590.

Li, X.X., Xu, X.D., Dan, Y.Y., and Zhang, M.L. (2008). The factors during protein crystallization: A review. Crystallogr. Rep. *53*, 1261–1266.

Liu, Y.-W., Denkmann, K., Kosciow, K., Dahl, C., and Kelly, D.J. (2013). Tetrathionate stimulated growth of Campylobacter jejuni identifies a new type of bi-functional tetrathionate reductase (TsdA) that is widely distributed in bacteria. Mol. Microbiol. *88*, 173–188.

Longenecker, K.L., Garrard, S.M., Sheffield, P.J., and Derewenda, Z.S. (2001). Protein crystallization by rational mutagenesis of surface residues: Lys to Ala mutations promote crystallization of RhoGDI. Acta Crystallogr. D Biol. Crystallogr. *57*, 679–688.

Lu, P., Bai, X., Ma, D., Xie, T., Yan, C., Sun, L., Yang, G., Zhao, Y., Zhou, R., Scheres, S.H.W., et al. (2014). Three-dimensional structure of human γ-secretase. Nature *512*, 166–170.

Ma, Z., and Finkel, T.H. (2010). T cell receptor triggering by force. Trends Immunol. *31*, 1–6.

Macho-Fernandez, E., and Brigl, M. (2015). The Extended Family of CD1d-Restricted NKT Cells: Sifting through a Mixed Bag of TCRs, Antigens, and Functions. Front. Immunol. *6*, 362.

Mackay, C.R., Marston, W.L., and Dudler, L. (1990). Naive and memory T cells show distinct pathways of lymphocyte recirculation. J. Exp. Med. *171*, 801–817.

Madsen, J., Mollenhauer, J., and Holmskov, U. (2010). Review: Gp-340/DMBT1 in mucosal innate immunity. Innate Immun. *16*, 160–167.

Malissen, B., and Bongrand, P. (2015). Early T cell activation: integrating biochemical, structural, and biophysical cues. Annu. Rev. Immunol. *33*, 539–561.

Manz, B.N., Jackson, B.L., Petit, R.S., Dustin, M.L., and Groves, J. (2011). T-cell triggering thresholds are modulated by the number of antigen within individual T-cell receptor clusters. Proc. Natl. Acad. Sci. U. S. A. *108*, 9089–9094.

Marion, D. (2013). An introduction to biological NMR spectroscopy. Mol. Cell. Proteomics mcp.O113.030239.

Martin, M.P., Gao, X., Lee, J.-H., Nelson, G.W., Detels, R., Goedert, J.J., Buchbinder, S., Hoots, K., Vlahov, D., Trowsdale, J., et al. (2002). Epistatic interaction between KIR3DS1 and HLA-B delays the progression to AIDS. Nat. Genet. *31*, 429–434.

Mateja, A., Devedjiev, Y., Krowarsch, D., Longenecker, K., Dauter, Z., Otlewski, J., and Derewenda, Z.S. (2002). The impact of Glu→Ala and Glu→Asp mutations on the crystallization properties of RhoGDI: the structure of RhoGDI at 1.3 Å resolution. Acta Crystallogr. D Biol. Crystallogr. *58*, 1983–1991.

Matsumura, M., Fremont, D.H., Peterson, P.A., and Wilson, I.A. (1992). Emerging principles for the recognition of peptide antigens by MHC class I molecules. Science *257*, 927–934.

Matulis, D., Kranz, J.K., Salemme, F.R., and Todd, M.J. (2005). Thermodynamic Stability of Carbonic Anhydrase: Measurements of Binding Affinity and Stoichiometry Using ThermoFluor. Biochemistry (Mosc.) *44*, 5258–5266.

Maupin-Furlow, J. (2012). Proteasomes and protein conjugation across domains of life. Nat. Rev. Microbiol. *10*, 100–111.

Mazza, C., Auphan-Anezin, N., Gregoire, C., Guimezanes, A., Kellenberger, C., Roussel, A., Kearney, A., van der Merwe, P.A., Schmitt-Verhulst, A.-M., and Malissen, B. (2007). How much can a T-cell antigen receptor adapt to structurally distinct antigenic peptides? EMBO J. *26*, 1972–1983.

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

McMullan, G., Faruqi, A.R., Henderson, R., Guerrini, N., Turchetta, R., Jacobs, A., and van Hoften, G. (2009). Experimental observation of the improvement in MTF from backthinning a CMOS direct electron detector. Ultramicroscopy *109*, 1144–1147.

Meenan, N.A.G., Sharma, A., Fleishman, S.J., Macdonald, C.J., Morel, B., Boetzel, R., Moore, G.R., Baker, D., and Kleanthous, C. (2010). The structural and energetic basis for high selectivity in a high-affinity protein-protein interaction. Proc. Natl. Acad. Sci. U. S. A. *107*, 10080–10085.

van der Merwe, P.A., and Dushek, O. (2011). Mechanisms for T cell receptor triggering. Nat. Rev. Immunol. *11*, 47–55.

Metcalfe, C., Cresswell, P., Ciaccia, L., Thomas, B., and Barclay, A.N. (2011). Labile disulfide bonds are common at the leucocyte cell surface. Open Biol. *1*.

Meydan, C., Otu, H.H., and Sezerman, O.U. (2013). Prediction of peptides binding to MHC class I and II alleles by temporal motif mining. BMC Bioinformatics *14*, S13.

Mitra, A.K., Célia, H., Ren, G., Luz, J.G., Wilson, I.A., and Teyton, L. (2004). Supine orientation of a murine MHC class I molecule on the membrane bilayer. Curr. Biol. CB *14*, 718–724.

Moraes, A.M., Jorge, S.A.C., Astray, R.M., Suazo, C.A.T., Calderón Riquelme, C.E., Augusto, E.F.P., Tonso, A., Pamboukian, M.M., Piccoli, R.A.M., Barral, M.F., et al. (2012). Drosophila melanogaster S2 cells for expression of heterologous genes: From gene cloning to bioprocess development. Biotechnol. Adv. *30*, 613–628.

Moran, A.E., and Hogquist, K.A. (2012). T-cell receptor affinity in thymic development. Immunology *135*, 261–267.

Moremen, K.W., Tiemeyer, M., and Nairn, A.V. (2012). Vertebrate protein glycosylation: diversity, synthesis and function. Nat. Rev. Mol. Cell Biol. *13*, 448–462.

Morrow, F. (2004). Marek's disease: a worldwide problem. In In Marek's Disease: An Evolving Problem, (Elsevier Academic Press),.

Mosmann, T.R., Cherwinski, H., Bond, M.W., Giedlin, M.A., and Coffman, R.L. (1986). Two types of murine helper T cell clone. I. Definition according to profiles of lymphokine activities and secreted proteins. J. Immunol. Baltim. Md 1950 *136*, 2348–2357.

Mukhopadhyaya, P.N., Deb, C., Lahiri, C., and Roy, P. (2000). A soxA gene, encoding a diheme cytochrome c, and a sox locus, essential for sulfur oxidation in a new sulfur lithotrophic bacterium. J. Bacteriol. *182*, 4278–4287.

Murata, K., Liu, X., Danev, R., Jakana, J., Schmid, M.F., King, J., Nagayama, K., and Chiu, W. (2010). Zernike phase contrast cryo-electron microscopy and tomography for structure determination at nanometer and subnanometer resolutions. Struct. Lond. Engl. 1993 *18*, 903–912.

Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997). Refinement of Macromolecular Structures by the Maximum-Likelihood Method. Acta Crystallogr. D Biol. Crystallogr. *53*, 240–255.

Nakayama, T., and Yamashita, M. (2010). The TCR-mediated signaling pathways that control the direction of helper T cell differentiation. Semin. Immunol. *22*, 303–309.

Natarajan, K., Li, H., Mariuzza, R.A., and Margulies, D.H. (1999). MHC class I molecules, structure and function. Rev. Immunogenet. *1*, 32–46.

Nelissen, J.M.D.T., Peters, I.M., Grooth, B.G. de, Kooyk, Y. van, and Figdor, C.G. (2000). Dynamic Regulation of Activated Leukocyte Cell Adhesion Molecule–mediated Homotypic Cell Adhesion through the Actin Cytoskeleton. Mol. Biol. Cell *11*, 2057–2068.

Nguyen, C., Haushalter, R.W., Lee, D.J., Markwick, P.R.L., Bruegger, J., Caldara-Festin, G., Finzel, K., Jackson, D.R., Ishikawa, F., O'Dowd, B., et al. (2014). Trapping the dynamic acyl carrier protein in fatty acid biosynthesis. Nature *505*, 427–431.

Nguyen, T.T., Chang, S.-C., Evnouchidou, I., York, I.A., Zikos, C., Rock, K.L., Goldberg, A.L., Stratikos, E., and Stern, L.J. (2011). Structural basis for antigenic peptide precursor processing by the endoplasmic reticulum aminopeptidase ERAP1. Nat. Struct. Mol. Biol. *18*, 604–613.

Nielsen, M.J., Andersen, C.B.F., and Moestrup, S.K. (2013). CD163 binding to haptoglobin-hemoglobin complexes involves a dual-point electrostatic receptor-ligand pairing. J. Biol. Chem. *288*, 18834–18841.

Nilsen, T.W., and Graveley, B.R. (2010). Expansion of the eukaryotic proteome by alternative splicing. Nature *463*, 457–463.

Nothaft, H., and Szymanski, C.M. (2010). Protein glycosylation in bacteria: sweeter than ever. Nat. Rev. Microbiol. *8*, 765–778.

Nowak, M.A., Tarczy-Hornoch, K., and Austyn, J.M. (1992). The optimal number of major histocompatibility complex molecules in an individual. Proc. Natl. Acad. Sci. U. S. A. *89*, 10896–10899.

Nurieva, R.I., Chung, Y., Martinez, G.J., Yang, X.O., Tanaka, S., Matskevitch, T.D., Wang, Y.-H., and Dong, C. (2009). Bcl6 mediates the development of T follicular helper cells. Science *325*, 1001–1005.

Oancea, G., O'Mara, M.L., Bennett, W.F.D., Tieleman, D.P., Abele, R., and Tampé, R. (2009). Structural arrangement of the transmission interface in the antigen ABC transport complex TAP. Proc. Natl. Acad. Sci. U. S. A. *106*, 5551–5556.

O'Donoghue, G.P., Pielak, R.M., Smoligovets, A.A., Lin, J.J., and Groves, J.T. (2013). Direct single molecule measurement of TCR triggering by agonist pMHC in living primary T cells. eLife *2*, e00778.

Ojala, J.R.M., Pikkarainen, T., Tuuttila, A., Sandalova, T., and Tryggvason, K. (2007). Crystal structure of the cysteine-rich domain of scavenger receptor MARCO reveals the presence of a basic and an acidic cluster that both contribute to ligand recognition. J. Biol. Chem. *282*, 16654–16666.

Osorio, L.M., Garcia, C.A., Jondal, M., and Chow, S.C. (1994). The anti-CD6 mAb, IOR-T1, defined a new epitope on the human CD6 molecule that induces greater responsiveness in T cell receptor/CD3-mediated T cell proliferation. Cell. Immunol. *154*, 123–133.

Pancer, Z., and Cooper, M.D. (2006). The evolution of adaptive immunity. Annu. Rev. Immunol. *24*, 497–518.

Patel, D.D., Wee, S.F., Whichard, L.P., Bowen, M.A., Pesando, J.M., Aruffo, A., and Haynes, B.F. (1995). Identification and characterization of a 100-kD ligand for CD6 on human thymic epithelial cells. J. Exp. Med. *181*, 1563–1568.

Paul, S., Weiskopf, D., Angelo, M.A., Sidney, J., Peters, B., and Sette, A. (2013). HLA class I alleles are associated with peptide-binding repertoires of different size, affinity, and immunogenicity. J. Immunol. Baltim. Md 1950 *191*, 5831–5839.

Peaper, D.R., and Cresswell, P. (2008). The redox activity of ERp57 is not essential for its functions in MHC class I peptide loading. Proc. Natl. Acad. Sci. U. S. A. *105*, 10477–10482.

Peh, C.A., Laham, N., Burrows, S.R., Zhu, Y., and McCluskey, J. (2000). Distinct functions of tapasin revealed by polymorphism in MHC class I peptide loading. J. Immunol. Baltim. Md 1950 *164*, 292–299.

Perham, R.N. (2000). Swinging arms and swinging domains in multifunctional enzymes: catalytic machines for multistep reactions. Annu. Rev. Biochem. *69*, 961–1004.

Plachy, J., Pink, J.R., and Hála, K. (1992). Biology of the chicken MHC (B complex). Crit. Rev. Immunol. *12*, 47–79.

Pos, W., Sethi, D.K., Call, M.J., Schulze, M.-S.E.D., Anders, A.-K., Pyrdol, J., and Wucherpfennig, K.W. (2012). Crystal structure of the HLA-DM-HLA-DR1 complex defines mechanisms for rapid peptide selection. Cell *151*, 1557–1568.

Printz, R.L., Magnuson, M.A., and Granner, D.K. (1993). Mammalian Glucokinase. Annu. Rev. Nutr. *13*, 463–496.

Procko, E., Raghuraman, G., Wiley, D.C., Raghavan, M., and Gaudet, R. (2005). Identification of domain boundaries within the N-termini of TAP1 and TAP2 and their importance in tapasin binding and tapasin-mediated increase in peptide loading of MHC class I. Immunol. Cell Biol. *83*, 475–482.

Purcell, A.W., Gorman, J.J., Garcia-Peydró, M., Paradela, A., Burrows, S.R., Talbo, G.H., Laham, N., Peh, C.A., Reynolds, E.C., López De Castro, J.A., et al. (2001). Quantitative and qualitative influences of tapasin on the class I peptide repertoire. J. Immunol. Baltim. Md 1950 *166*, 1016–1027.

Quentmeier, A., and Friedrich, C.G. (2001). The cysteine residue of the SoxY protein as the active site of protein-bound sulfur oxidation of Paracoccus pantotrophus GB17. FEBS Lett. *503*, 168–172.

Rammensee, H.G., Friede, T., and Stevanoviíc, S. (1995). MHC ligands and peptide motifs: first listing. Immunogenetics *41*, 178–228.

Randolph, G.J., Angeli, V., and Swartz, M.A. (2005). Dendritic-cell trafficking to lymph nodes through lymphatic vessels. Nat. Rev. Immunol. *5*, 617–628.

Reddy, V.S., Natchiar, S.K., Stewart, P.L., and Nemerow, G.R. (2010). Crystal Structure of Human Adenovirus at 3.5 Å Resolution. Science *329*, 1071–1075.

Reinherz, E.L., Geha, R., Rappeport, J.M., Wilson, M., Penta, A.C., Hussey, R.E., Fitzgerald, K.A., Daley, J.F., Levine, H., Rosen, F.S., et al. (1982). Reconstitution after transplantation with T-lymphocyte-depleted HLA haplotype-mismatched bone marrow for severe combined immunodeficiency. Proc. Natl. Acad. Sci. U. S. A. *79*, 6047–6051.

Ritter, C., Quirin, K., Kowarik, M., and Helenius, A. (2005). Minor folding defects trigger local modification of glycoproteins by the ER folding sensor GT. EMBO J. *24*, 1730–1738.

Roberts, N.A., Martin, J.A., Kinchington, D., Broadhurst, A.V., Craig, J.C., Duncan, I.B., Galpin, S.A., Handa, B.K., Kay, J., and Kröhn, A. (1990). Rational design of peptide-based HIV proteinase inhibitors. Science *248*, 358–361.

Roche, P.A., and Cresswell, P. (1991). Proteolysis of the class II-associated invariant chain generates a peptide binding site in intracellular HLA-DR molecules. Proc. Natl. Acad. Sci. U. S. A. *88*, 3150–3154.

Rodamilans, B., Muñoz, I.G., Bragado-Nilsson, E., Sarrias, M.R., Padilla, O., Blanco, F.J., Lozano, F., and Montoya, G. (2007). Crystal structure of the third extracellular domain of CD5 reveals the fold of a group B scavenger cysteine-rich receptor domain. J. Biol. Chem. *282*, 12669–12677.

Rodriguez, P.C., Torres-Moya, R., Reyes, G., Molinero, C., Prada, D., Lopez, A.M., Hernandez, I.M., Hernandez, M.V., Martinez, J.P., Hernandez, X., et al. (2012). A clinical exploratory study with itolizumab, an anti-CD6 monoclonal antibody, in patients with rheumatoid arthritis. Results Immunol. *2*, 204–211.

Romagnoli, P., and Germain, R.N. (1994). The CLIP region of invariant chain plays a critical role in regulating major histocompatibility complex class II folding, transport, and peptide occupancy. J. Exp. Med. *180*, 1107–1113.

Roos, W.H., Ivanovska, I.L., Evilevitch, A., and Wuite, G.J.L. (2007). Viral capsids: Mechanical characteristics, genome packaging and delivery mechanisms. Cell. Mol. Life Sci. *64*, 1484–1497.

Rother, D., Henrich, H.J., Quentmeier, A., Bardischewsky, F., and Friedrich, C.G. (2001). Novel genes of the sox gene cluster, mutagenesis of the flavoprotein SoxF, and evidence for a general sulfur-oxidizing system in Paracoccus pantotrophus GB17. J. Bacteriol. *183*, 4499–4508.

Roversi, P., Johnson, S., Caesar, J.J.E., McLean, F., Leath, K.J., Tsiftsoglou, S.A., Morgan, B.P., Harris, C.L., Sim, R.B., and Lea, S.M. (2011). Structural basis for complement factor I control and its disease-associated sequence polymorphisms. Proc. Natl. Acad. Sci. U. S. A. *108*, 12839–12844.

Ruskin, R.S., Yu, Z., and Grigorieff, N. (2013). Quantitative characterization of electron detectors for transmission electron microscopy. J. Struct. Biol. *184*.

Russo, C.J., and Passmore, L.A. (2014a). Electron microscopy: Ultrastable gold substrates for electron cryomicroscopy. Science *346*, 1377–1380.

Russo, C.J., and Passmore, L.A. (2014b). Controlling protein adsorption on graphene for cryo-EM using low-energy hydrogen plasmas. Nat. Methods *11*, 649–652.

Ryan, S.O., and Cobb, B.A. (2012). Roles for major histocompatibility complex glycosylation in immune function. Semin. Immunopathol. *34*.

Rychlewski, L., Jaroszewski, L., Li, W., and Godzik, A. (2000). Comparison of sequence profiles. Strategies for structural predictions using sequence information. Protein Sci. Publ. Protein Soc. *9*, 232–241.

Sadasivan, B., Lehner, P.J., Ortmann, B., Spies, T., and Cresswell, P. (1996). Roles for calreticulin and a novel glycoprotein, tapasin, in the interaction of MHC class I molecules with TAP. Immunity *5*, 103–114.

Sallusto, F., Lenig, D., Förster, R., Lipp, M., and Lanzavecchia, A. (1999). Two subsets of memory T lymphocytes with distinct homing potentials and effector functions. Nature *401*, 708–712.

Sarafianos, S.G., Marchand, B., Das, K., Himmel, D.M., Parniak, M.A., Hughes, S.H., and Arnold, E. (2009). Structure and function of HIV-1 reverse transcriptase: molecular mechanisms of polymerization and inhibition. J. Mol. Biol. *385*, 693–713.

Saric, T., Chang, S.-C., Hattori, A., York, I.A., Markant, S., Rock, K.L., Tsujimoto, M., and Goldberg, A.L. (2002). An IFN-gamma-induced aminopeptidase in the ER, ERAP1, trims precursors to MHC class I-presented peptides. Nat. Immunol. *3*, 1169–1176.

Sarrias, M.R., Grønlund, J., Padilla, O., Madsen, J., Holmskov, U., and Lozano, F. (2004). The Scavenger Receptor Cysteine-Rich (SRCR) domain: an ancient and highly conserved protein module of the innate immune system. Crit. Rev. Immunol. *24*, 1–37.

Sarrias, M.-R., Farnós, M., Mota, R., Sánchez-Barbero, F., Ibáñez, A., Gimferrer, I., Vera, J., Fenutría, R., Casals, C., Yélamos, J., et al. (2007). CD6 binds to pathogen-associated molecular patterns and protects from LPS-induced septic shock. Proc. Natl. Acad. Sci. U. S. A. *104*, 11724–11729.

Sauvé, V., Bruno, S., Berks, B.C., and Hemmings, A.M. (2007). The SoxYZ complex carries sulfur cycle intermediates on a peptide swinging arm. J. Biol. Chem. *282*, 23194–23204.

Sauvé, V., Roversi, P., Leath, K.J., Garman, E.F., Antrobus, R., Lea, S.M., and Berks, B.C. (2009). Mechanism for the hydrolysis of a sulfur-sulfur bond based on the crystal structure of the thiosulfohydrolase SoxB. J. Biol. Chem. *284*, 21707–21718.

Saveanu, L., Carroll, O., Lindo, V., Del Val, M., Lopez, D., Lepelletier, Y., Greer, F., Schomburg, L., Fruci, D., Niedermann, G., et al. (2005). Concerted peptide trimming by human ERAP1 and ERAP2 aminopeptidase complexes in the endoplasmic reticulum. Nat. Immunol. *6*, 689–697.

Schalkoff, R., J. (1992). Pattern Recognition: Statistical, Structural and Neural Approaches (Wiley).

Scheres, S.H. (2014). Beam-induced motion correction for sub-megadalton cryo-EM particles. eLife *3*.

Scheres, S.H.W. (2012). A Bayesian View on Cryo-EM Structure Determination. J. Mol. Biol. *415*, 406–418.

Schreiber, G., and Keating, A.E. (2011). Protein binding specificity versus promiscuity. Curr. Opin. Struct. Biol. *21*, 50–61.

Schroder, K., Hertzog, P.J., Ravasi, T., and Hume, D.A. (2004). Interferon-γ: an overview of signals, mechanisms and functions. J. Leukoc. Biol. *75*, 163–189.

Schrödinger, LLC The PyMOL Molecular Graphics System, Version 1.7.4. Glob. Phasing Ltd.

Schubert, U., Antón, L.C., Gibbs, J., Norbury, C.C., Yewdell, J.W., and Bennink, J.R. (2000). Rapid degradation of a large fraction of newly synthesized proteins by proteasomes. Nature *404*, 770–774.

Schumacher, T.N., De Bruijn, M.L., Vernie, L.N., Kast, W.M., Melief, C.J., Neefjes, J.J., and Ploegh, H.L. (1991). Peptide selection by MHC class I molecules. Nature *350*, 703–706.

Schuwirth, B.S., Borovinskaya, M.A., Hau, C.W., Zhang, W., Vila-Sanjurjo, A., Holton, J.M., and Cate, J.H.D. (2005). Structures of the Bacterial Ribosome at 3.5 Å Resolution. Science *310*, 827–834.

Schwarz, F., and Aebi, M. (2011). Mechanisms and principles of N-linked protein glycosylation. Curr. Opin. Struct. Biol. *21*, 576–582.

Selmer, M., Dunham, C.M., Murphy, F.V., Weixlbaumer, A., Petry, S., Kelley, A.C., Weir, J.R., and Ramakrishnan, V. (2006). Structure of the 70S Ribosome Complexed with mRNA and tRNA. Science *313*, 1935–1942.

Serwold, T., Gonzalez, F., Kim, J., Jacob, R., and Shastri, N. (2002). ERAAP customizes peptides for MHC class I molecules in the endoplasmic reticulum. Nature *419*, 480–483.

Shaw, I., Powell, T.J., Marston, D.A., Baker, K., van Hateren, A., Riegert, P., Wiles, M.V., Milne, S., Beck, S., and Kaufman, J. (2007). Different evolutionary histories of the two classical class I genes BF1 and BF2 illustrate drift and selection within the stable MHC haplotypes of chickens. J. Immunol. Baltim. Md 1950 *178*, 5744–5752.

Shen, L., and Rock, K.L. (2006). Priming of T cells by exogenous antigen cross-presented on MHC class I molecules. Curr. Opin. Immunol. *18*, 85–91.

Shinkai, Y., Rathbun, G., Lam, K.P., Oltz, E.M., Stewart, V., Mendelsohn, M., Charron, J., Datta, M., Young, F., and Stall, A.M. (1992). RAG-2-deficient mice lack mature lymphocytes owing to inability to initiate V(D)J rearrangement. Cell *68*, 855–867.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. *7*, 539.

Skonier, J.E., Bowen, M.A., Emswiler, J., Aruffo, A., and Bajorath, J. (1996a). Mutational analysis of the CD6 binding site in activated leukocyte cell adhesion molecule. Biochemistry (Mosc.) *35*, 14743–14748.

Skonier, J.E., Bowen, M.A., Emswiler, J., Aruffo, A., and Bajorath, J. (1996b). Recognition of Diverse Proteins by Members of the Immunoglobulin Superfamily: Delineation of the Receptor Binding Site in the Human CD6 Ligand ALCAM. Biochemistry (Mosc.) *35*, 12287–12291.

Skonier, J.E., Bodian, D.L., Emswiler, J., Bowen, M.A., Aruffo, A., and Bajorath, J. (1997). Mutational analysis of the CD6 ligand binding domain. Protein Eng. *10*, 943–947.

Smart, O.S., Womack, T.O., Flensburg, C., Keller, P., Paciorek, W., Sharff, A., Vonrhein, C., and Bricogne, G. (2012). Exploiting structure similarity in refinement: automated NCS and target-structure restraints in *BUSTER*. Acta Crystallogr. D Biol. Crystallogr. *68*, 368–380.

Smith-Garvin, J.E., Koretzky, G.A., and Jordan, M.S. (2009). T cell activation. Annu. Rev. Immunol. *27*, 591–619.

Soldà, T., Galli, C., Kaufman, R.J., and Molinari, M. (2007). Substrate-specific requirements for UGT1-dependent release from calnexin. Mol. Cell *27*, 238–249.

Somoza, J.R., Ho, J.D., Luong, C., Ghate, M., Sprengeler, P.A., Mortara, K., Shrader, W.D., Sperandio, D., Chan, H., McGrath, M.E., et al. (2003). The structure of the extracellular region of human hepsin reveals a serine protease domain and a novel scavenger receptor cysteine-rich (SRCR) domain. Struct. Lond. Engl. 1993 *11*, 1123–1131.

Speir, J.A., Stevens, J., Joly, E., Butcher, G.W., and Wilson, I.A. (2001). Two different, highly exposed, bulged structures for an unusually long peptide bound to rat MHC class I RT1-Aa. Immunity *14*, 81–92.

Spiro, R.G. (2002). Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. Glycobiology *12*, 43R – 56R.

Stanley, P. (1989). Chinese hamster ovary cell mutants with multiple glycosylation defects for production of glycoproteins with minimal carbohydrate heterogeneity. Mol. Cell. Biol. *9*, 377–383.

Starr, T.K., Jameson, S.C., and Hogquist, K.A. (2003). Positive and negative selection of T cells. Annu. Rev. Immunol. *21*, 139–176.

Stein, N. (2008). CHAINSAW: a program for mutating pdb files used as templates in molecular replacement. J. Appl. Crystallogr. *41*, 641–643.

Suga, M., Lai, T.-L., Sugiura, M., Shen, J.-R., and Boussac, A. (2013). Crystal structure at 1.5Å resolution of the PsbV2 cytochrome from the cyanobacterium Thermosynechococcus elongatus. FEBS Lett. *587*, 3267–3272.

Sun, J.C., Ugolini, S., and Vivier, E. (2014). Immunological memory within the innate immune system. EMBO J. *33*, 1295–1303.

Sun, L., Zhao, L., Yang, G., Yan, C., Zhou, R., Zhou, X., Xie, T., Zhao, Y., Wu, S., Li, X., et al. (2015). Structural basis of human γ-secretase assembly. Proc. Natl. Acad. Sci. U. S. A. *112*, 6003–6008.

Swaminathan, B., Cuapio, A., Alloza, I., Matesanz, F., Alcina, A., García-Barcina, M., Fedetz, M., Fernández, O., Lucas, M., Orpez, T., et al. (2013). Fine mapping and functional analysis of the multiple sclerosis risk gene CD6. PloS One *8*, e62376.

Swart, G.W.M. (2002). Activated leukocyte cell adhesion molecule (CD166/ALCAM): developmental and mechanistic aspects of cell clustering and cell migration. Eur. J. Cell Biol. *81*, 313–321.

Takeuchi, O., and Akira, S. (2010). Pattern recognition receptors and inflammation. Cell *140*, 805–820.

Tan, P., Kropshofer, H., Mandelboim, O., Bulbuc, N., Hämmerling, G.J., and Momburg, F. (2002). Recruitment of MHC class I molecules by tapasin into the

transporter associated with antigen processing-associated complex is essential for optimal peptide loading. J. Immunol. Baltim. Md 1950 *168*, 1950–1960.

Tanovic, A., Samel, S.A., Essen, L.-O., and Marahiel, M.A. (2008). Crystal structure of the termination module of a nonribosomal peptide synthetase. Science *321*, 659–663.

Te Riet, J., Zimmerman, A.W., Cambi, A., Joosten, B., Speller, S., Torensma, R., van Leeuwen, F.N., Figdor, C.G., and de Lange, F. (2007). Distinct kinetic and mechanical properties govern ALCAM-mediated interactions as shown by single-molecule force spectroscopy. J. Cell Sci. *120*, 3965–3976.

Te Riet, J., Helenius, J., Strohmeyer, N., Cambi, A., Figdor, C.G., and Müller, D.J. (2014). Dynamic coupling of ALCAM to the actin cortex strengthens cell adhesion to CD6. J. Cell Sci. *127*, 1595–1606.

The International HIV Controllers Study (2010). The Major Genetic Determinants of HIV-1 Control Affect HLA Class I Peptide Presentation. Science *330*, 1551–1557.

The MHC sequencing Consortium (1999). Complete sequence and gene map of a human major histocompatibility complex. Nature *401*, 921–923.

Tiwari, J.L., and Terasaki, P.I. (1981). HLA-DR and disease associations. Prog. Clin. Biol. Res. *58*, 151–163.

Trost, B., Bickis, M., and Kusalik, A. (2007). Strength in numbers: achieving greater accuracy in MHC-I binding prediction by combining the results from multiple prediction tools. Immunome Res. *3*, 5.

Ukita, S., Fujii, T., Hira, D., Nishiyama, T., Kawase, T., Migita, C.T., and Furukawa, K. (2010). A heterodimeric cytochrome c complex with a very low redox potential from an anaerobic ammonium-oxidizing enrichment culture. FEMS Microbiol. Lett. *313*, 61–67.

Vagin, A., and Teplyakov, A. (2010). Molecular replacement with MOLREP. Acta Crystallogr. D Biol. Crystallogr. *66*, 22–25.

Vaidehi, N., Bhattacharya, S., and Larsen, A.B. (2014). Structure and dynamics of G-protein coupled receptors. Adv. Exp. Med. Biol. *796*, 37–54.

Van Driessche, G., Devreese, B., Fitch, J.C., Meyer, T.E., Cusanovich, M.A., and Van Beeumen, J.J. (2006). GHP, a new c-type green heme protein from Halochromatium salexigens and other proteobacteria. FEBS J. *273*, 2801–2811.

Vantourout, P., and Hayday, A. (2013). Six-of-the-best: unique contributions of γδ T cells to immunology. Nat. Rev. Immunol. *13*, 88–100.

Vidović, D., and Matzinger, P. (1988). Unresponsiveness to a foreign antigen can be caused by self-tolerance. Nature *336*, 222–225.

Vignali, D.A.A., Collison, L.W., and Workman, C.J. (2008). How regulatory T cells work. Nat. Rev. Immunol. *8*, 523–532.

Villanueva, M.S., Sijts, A.J., and Pamer, E.G. (1995). Listeriolysin is processed efficiently into an MHC class I-associated epitope in Listeria monocytogenes-infected cells. J. Immunol. Baltim. Md 1950 *155*, 5227–5233.

Vivier, E., Tomasello, E., Baratin, M., Walzer, T., and Ugolini, S. (2008). Functions of natural killer cells. Nat. Immunol. *9*, 503–510.

Vonrhein, C., Blanc, E., Roversi, P., and Bricogne, G. (2007). Automated structure solution with autoSHARP. Methods Mol. Biol. Clifton NJ *364*, 215–230.

Voorhees, R.M., Fernández, I.S., Scheres, S.H.W., and Hegde, R.S. (2014). Structure of the Mammalian Ribosome-Sec61 Complex to 3.4 Å Resolution. Cell *157*, 1632–1643.

Walker, B.A., Hunt, L.G., Sowa, A.K., Skjødt, K., Göbel, T.W., Lehner, P.J., and Kaufman, J. (2011). The dominantly expressed class I molecule of the chicken MHC is explained by coevolution with the polymorphic peptide transporter (TAP) genes. Proc. Natl. Acad. Sci. U. S. A. *108*, 8396–8401.

Wallny, H.-J., Avila, D., Hunt, L.G., Powell, T.J., Riegert, P., Salomonsen, J., Skjødt, K., Vainio, O., Vilbois, F., Wiles, M.V., et al. (2006). Peptide motifs of the single dominantly expressed class I molecule explain the striking MHC-determined response to Rous sarcoma virus in chickens. Proc. Natl. Acad. Sci. U. S. A. *103*, 1434–1439.

Walsh, C.T., Garneau-Tsodikova, S., and Gatto, G.J. (2005). Protein posttranslational modifications: the chemistry of proteome diversifications. Angew. Chem. Int. Ed Engl. *44*, 7342–7372.

Warne, T., Serrano-Vega, M.J., Baker, J.G., Moukhametzianov, R., Edwards, P.C., Henderson, R., Leslie, A.G.W., Tate, C.G., and Schertler, G.F.X. (2008). Structure of a beta1-adrenergic G-protein-coupled receptor. Nature *454*, 486–491.

Wearsch, P.A., and Cresswell, P. (2007). Selective loading of high-affinity peptides onto major histocompatibility complex class I molecules by the tapasin-ERp57 heterodimer. Nat. Immunol. *8*, 873–881.

Wearsch, P.A., Peaper, D.R., and Cresswell, P. (2011). Essential glycan-dependent interactions optimize MHC class I peptide loading. Proc. Natl. Acad. Sci. U. S. A. *108*, 4950–4955.

Weidle, U.H., Eggle, D., Klostermann, S., and Swart, G.W.M. (2010). ALCAM/CD166: cancer-related issues. Cancer Genomics Proteomics *7*, 231–243.

Welsh, R.M. (1984). Natural killer cells and interferon. Crit. Rev. Immunol. *5*, 55–93.

Welte, C., Hafner, S., Krätzer, C., Quentmeier, A., Friedrich, C.G., and Dahl, C. (2009). Interaction between Sox proteins of two physiologically distinct bacteria and a new protein involved in thiosulfate oxidation. FEBS Lett. *583*, 1281–1286.

West, A.P., Koblansky, A.A., and Ghosh, S. (2006). Recognition and signaling by toll-like receptors. Annu. Rev. Cell Dev. Biol. *22*, 409–437.

White, K.A., Enjuanes, L., and Berkhout, B. (2011). RNA virus replication, transcription and recombination. RNA Biol. *8*, 182–183.

Whitney, G.S., Starling, G.C., Bowen, M.A., Modrell, B., Siadak, A.W., and Aruffo, A. (1995). The membrane-proximal scavenger receptor cysteine-rich domain of CD6

contains the activated leukocyte cell adhesion molecule binding site. J. Biol. Chem. *270*, 18187–18190.

Wiegers, G.J., Kaufmann, M., Tischner, D., and Villunger, A. (2011). Shaping the T-cell repertoire: a matter of life and death. Immunol. Cell Biol. *89*, 33–39.

Williams, A.P., Peh, C.A., Purcell, A.W., McCluskey, J., and Elliott, T. (2002). Optimization of the MHC class I peptide cargo is dependent on tapasin. Immunity *16*, 509–520.

Williamson, M.P., Havel, T.F., and Wüthrich, K. (1985). Solution conformation of proteinase inhibitor IIA from bull seminal plasma by 1H nuclear magnetic resonance and distance geometry. J. Mol. Biol. *182*, 295–315.

Wilson, I.A., and Fremont, D.H. (1993). Structural analysis of MHC class I molecules with bound peptide antigens. Semin. Immunol. *5*, 75–80.

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., et al. (2011). Overview of the ıt CCP4 suite and current developments. Acta Crystallogr. Sect. D *67*, 235–242.

Winter, G., Lobley, C.M.C., and Prince, S.M. (2013). Decision making in xia2. Acta Crystallogr. D Biol. Crystallogr. *69*, 1260–1273.

Wodara, C., Bardischewsky, F., and Friedrich, C.G. (1997). Cloning and characterization of sulfite dehydrogenase, two c-type cytochromes, and a flavoprotein of Paracoccus denitrificans GB17: essential role of sulfite dehydrogenase in lithotrophic sulfur oxidation. J. Bacteriol. *179*, 5014–5023.

Wong, W., Bai, X., Brown, A., Fernandez, I.S., Hanssen, E., Condron, M., Tan, Y.H., Baum, J., and Scheres, S.H.W. (2014). Cryo-EM structure of the Plasmodium falciparum 80S ribosome bound to the anti-protozoan drug emetine. eLife *3*.

Wright, H.T. (1991). Nonenzymatic deamidation of asparaginyl and glutaminyl residues in proteins. Crit. Rev. Biochem. Mol. Biol. *26*, 1–52.

Wu, Y., Vendome, J., Shapiro, L., Ben-Shaul, A., and Honig, B. (2011). Transforming binding affinities from three dimensions to two with application to cadherin clustering. Nature *475*, 510–513.

Wüthrich, K., Wider, G., Wagner, G., and Braun, W. (1982). Sequential resonance assignments as a basis for determination of spatial protein structures by high resolution proton nuclear magnetic resonance. J. Mol. Biol. *155*, 311–319.

Xu, C., and Ng, D.T.W. (2015). Glycosylation-directed quality control of protein folding. Nat. Rev. Mol. Cell Biol. *16*, 742–752.

Yewdell, J.W. (2011). DRiPs solidify: progress in understanding endogenous MHC class I antigen processing. Trends Immunol. *32*, 548–558.

Yin, Y., Wang, X.X., and Mariuzza, R.A. (2012). Crystal structure of a complete ternary complex of T-cell receptor, peptide-MHC, and CD4. Proc. Natl. Acad. Sci. U. S. A. *109*, 5405–5410.

Yu, D., Rao, S., Tsai, L.M., Lee, S.K., He, Y., Sutcliffe, E.L., Srivastava, M., Linterman, M., Zheng, L., Simpson, N., et al. (2009). The transcriptional repressor Bcl-6 directs T follicular helper cell lineage commitment. Immunity *31*, 457–468.

Yusupov, M.M., Yusupova, G.Z., Baucom, A., Lieberman, K., Earnest, T.N., Cate, J.H.D., and Noller, H.F. (2001). Crystal Structure of the Ribosome at 5.5 Å Resolution. Science *292*, 883–896.

Zajonc, D.M., Cantu, C., Mattner, J., Zhou, D., Savage, P.B., Bendelac, A., Wilson, I.A., and Teyton, L. (2005). Structure and function of a potent agonist for the semi-invariant natural killer T cell receptor. Nat. Immunol. *6*, 810–818.

Zander, U., Faust, A., Klink, B.U., de Sanctis, D., Panjikar, S., Quentmeier, A., Bardischewsky, F., Friedrich, C.G., and Scheidig, A.J. (2011). Structural basis for the oxidation of protein-bound sulfur by the sulfur cycle molybdohemo-enzyme sulfane dehydrogenase SoxCD. J. Biol. Chem. *286*, 8349–8360.

Zhang, J., Chen, Y., Qi, J., Gao, F., Liu, Y., Liu, J., Zhou, X., Kaufman, J., Xia, C., and Gao, G.F. (2012). Narrow groove and restricted anchors of MHC class I molecule BF2*0401 plus peptide transporter restriction can explain disease susceptibility of B4 chickens. J. Immunol. Baltim. Md 1950 *189*, 4478–4487.

Zhang, W., Wearsch, P.A., Zhu, Y., Leonhardt, R.M., and Cresswell, P. (2011). A role for UDP-glucose glycoprotein glucosyltransferase in expression and quality control of MHC class I molecules. Proc. Natl. Acad. Sci. U. S. A. *108*, 4956–4961.