



Structural Studies on *Staphylococcus aureus* Quorum
Sensing Proteins AgrA and AgrB

Thomas Gwyn Warwick, MChem (Hons)

Thesis submitted in partial fulfilment of the degree of:

Doctor of Philosophy

1st Supervisor – Professor Jonas Emsley

2nd Supervisor – Professor Paul Williams

Internal Assessor – Doctor Ingrid Dreveny

The University of Nottingham

School of Pharmacy

February 2021



Abstract

Staphylococcus aureus is a human commensal pathogen notorious for the number and severity of infections it may cause and is endemic in hospitals worldwide. Virulence in *S. aureus* is partially controlled by a Quorum Sensing (QS) feedback circuit called *agr* (accessory gene regulator) which comprises four genes *agrB*, *agrD*, *agrC*, and *AgrA* that together regulate expression of major virulence factors. The activating switch of the QS loop requires the LytTR-type response regulator protein AgrA to bind to its cognate promoters to trigger gene expression. At present comprehension of the mode of AgrA dimerisation resulting from interaction with promoter DNA, and the asymmetry existing in the AgrA dimer, as a result of binding the two linear repeat sequences which comprise each promoter DNA, is absent.

The crystal structure of AgrA bound to the upstream linear repeat binding site of the P2 promoter has been solved at 4.8 Å from crystals containing 67 % solvent. The crystal structure exhibits an extremely hydrophobic patch that mediates a dimer that closely resembles physiological dimers of other response regulators such a ComE, FixJ, and LytR. Mutation of either of two proximal cysteine residue pairs that are conserved amongst all staphylococcal AgrA phosphate-receiving domains was shown to retard cell growth during *agr* dormancy. The molecular envelope of a beryllium trifluoride-activated AgrA dimer bound to the P3 promoter was constructed from small-angle X-ray scattering measurements and shows a bend of ~ 55 ° imposed on the DNA by AgrA.

AgrB is responsible for processing the linear precursor protein AgrD into the mature QS signal molecule, and is the archetypical member of the AgrB family of proteases for which there is no experimental structural data. Consequently, little is understood about the mechanism of AgrD processing and transport out of the cell.

A robust purification of AgrB solubilized into a number of detergents is presented. T7 western blotting failed to show any sample activity *in vitro*, however, circular dichroism measurements indicate that the sample is folded, primarily α -helical and exhibits a melting temperature of 63 °C. Crystallisation experiments were performed *in surfo* and attempted *in meso*, however, no crystals



were obtained. Preliminary negative staining electron micrographs and an atomic AgrB model are presented and discussed with respect to improving sample suitability for structural studies. The purification protocol developed was used to prepare AgrB for nanobody crystallisation chaperone generation at VIB nanobody core.

The results of this work provide structural evidence that *agr*-dependant virulence factor upregulation in *S. aureus* occurs through bending of promotor DNA by AgrA. A hitherto undiscovered cysteine pair in the crystal structure of full-length AgrA suggests that AgrA may encode an additional oxidative checkpoint to QS in *S. aureus*. Whether this cysteine pair constitutes a genuine oxidative checkpoint or is simply required for protein stability *in vitro* requires further investigation, however.



Acknowledgements

Thank you to the Medical Research Council for funding this PhD studentship and also to Jonas and Paul for accepting my candidacy. Especial thanks must be given to Dr. Philip Bardeland and Dr. Simon Caulton for the vast amount of light they have shed on the practical aspects of life sciences during the PhD process; as a graduate chemist this was much needed.

Thank you and good luck to all members of the University of Nottingham structural biology group. To those whom I have the privilege of calling friends thank you for the extra support, solidarity, weekends away and continued friendship.

Thank you to Ingrid for seeing to my annual progression and providing thoughtful feedback on my research. Thank you also to Dr. Ewan Murray for understanding that things do not always go according to plan and science, when all said and done, can be tough.

Particular thanks must be given to my mother, my girlfriend and my sister, who have always been there for me. I wouldn't have made it this far without your support and capacity to listen.



Table of Contents

Section	Title	Page
Chapter 1	Primary Introductions and Aims	
1.1.1	Infectious Diseases and the <i>Staphylococcus aureus agr</i> System	1
1.1.2	AgrD	7
1.1.3	AgrB	9
1.1.4	AIP	11
1.1.5	AgrC	12
1.1.6	AgrA	14
1.1.7	Genes Regulated by AgrA That Are Not Encoded by RNAIII	24
1.1.8	RNAIII	25
1.1.9	Additional Components of the <i>agr</i> System	26
1.1.10	The Potential of <i>agr</i> Components in Combating Antimicrobial Resistance	27
1.2.1	X-ray Crystallography	29
1.2.2	Introduction	29
1.2.3	The Basics of a Crystal and Diffraction Patterns	34
1.2.4	The Bragg Equation	36
1.2.5	The Fourier Transforms	40
1.2.6	The Phase Problem and Solutions to It	42
1.2.7	Model Building, Refinement and the R Factor	46
1.2.8	Model Bias and R_{free}	50
1.2.9	General Single Crystal X-ray Diffraction Experiment Workflow	51
1.3.1	PhD Aims	52
Chapter 2	AgrA – Expression, Purification and Crystallisation	
2.1	Preliminary Work on AgrA	53
2.2	Initial Attempts to Purify AgrA_C199S	53
2.3	Sample Confirmation and Activity Assaying of AgrA_C199S	61
2.4	AgrA_C199S Purifies as a Monomer in Solution	64
2.5	Initial Crystallisation Experiments ...AgrA_C199S with and without DNA	65
2.6	Control Crystallisation Experiments for AgrA_C199S/P2 _{fragment}	69
2.7.1	Optimisation of the AgrA_C199S/P2 _{fragment} Crystals	71
2.7.2	Chemical Grid Optimisation	71
2.7.3	Hanging Drop Optimisation	72
2.7.4.1	Seeding Optimisations	73



2.7.4.2	Streak Seeding into Hanging Drops	73
2.7.4.3	Streak Seeding into Sitting Drops	74
2.7.4.4	Seeding into Hanging Drops from Seed Stocks	74
2.7.4.5	Seeding into Sitting Drops from Seed Stocks	74
2.7.4.6	Discussion of the Seeding Results	74
2.7.5	Additive Screening	76
2.7.6	Macroseeding <i>In Situ</i>	77
2.7.7	Protein-to-DNA Ratio	80
2.7.8	Obtaining the Stoichiometric AgrA_C199S/P2 _{fragment} Complex	80
2.7.9	PEG Polymer Chain Length	81
2.7.10	Protein-to-Crystallisation Solution Ratio	81
2.7.11	Addition of D-sorbitol to the Crystallisation Solution	82
2.7.12	Greased Wells	82
2.7.13	Blunt-End DNA	82
2.7.14	Brominated DNA	83
2.7.15.1	Mimicking Phosphorylation ...in Pursuit of Better X-ray Diffraction	84
2.7.15.2	Preparation of the D59E Point Mutant	84
2.7.15.3	Beryllium Trifluoride as an Orthophosphate Mimetic	87
2.7.16	Streak Seeding into Iodine-Containing Drops	88
2.8	Crystallisation Optimisation Summary	88
2.9	Initial X-ray Diffraction Experiments on AgrA_C199S/P2 _{fragment} Crystals	92
2.10.1	Cryoprotection of the AgrA_C199S/P2 _{fragment} Crystals	96
2.11	Phasing the AgrA_C199S/P2 _{fragment} Crystals	98
Chapter 3	AgrA – Low-Resolution Structure Solution	
3.1	Crystal Structure of Full-Length AgrA_C199S in Complex with DNA	110
3.2	Major Fixes to the Initial Model	110
3.3	Overall Relationship Between the Two Domains	112
3.4	Overall Features of the N-terminal Domain	114
3.5	Overall Features of the C-terminal Domain	116
3.6	The Phosphate-Binding Region	118
3.7	Cysteines 55 and 123 Are Close to Each Other but Not to Cysteine 199	120
3.8	Helix 4 is Hydrophobic	123
3.9	An AgrA... Dimer Resembles Other Response Regulator Dimers	125
3.10	Further Analysis of AgrA_C199S Crystallographic Dimer	136
3.2.1	Final Structure Comments	139
3.2.2	On the Difficulty of Obtaining High-Resolution Crystal Data	139



3.2.3	Model Building at Low Resolution and Comments on Model Validity	141
Chapter 4	AgrA – Better Crystals and Structure-Activity Experiments	
4.1.1	Pursuit of Better AgrA Crystals via Surface Entropy Reduction	150
4.1.2	Surface Entropy Reduction on AgrA_C199S	150
4.1.3	SER Construct 1 – AgrA_K77A_C199S	154
4.1.4	SER Construct 2 – AgrA_Y100T_K101A_C199S	154
4.1.5	SER Construct 3 – AgrA_K101A_C199S	154
4.1.6	Comments on the Possible Role of Lys101 in AgrA Dimerisation	155
4.2.1	Pursuit of Better AgrA Crystals via Additional Crystallisation Experiments	161
4.2.2	AgrA_D59E_C199S/P2 _{fragment}	161
4.2.3	AgrA_C199S/P3 _{full}	161
4.2.4	AgrA_C199S/P3 _{full} and AgrA_C199S/P2 _{full} With and Without [BeF ₃] ⁻	162
4.2.5	AgrA N-terminal Domain Only	162
4.3.1	Structure-Activity Experiments on AgrA_C199S	166
4.3.2	SAXS... on the [BeF ₃] ⁻ -Activated AgrA Dimer/P3 _{full} Complex	166
4.3.3	Oxidising and Reducing EMSAs to Probe Cys55 and Cys123 <i>In Vitro</i>	172
4.3.4	<i>Lux</i> -based Reporter Assays Probe Cys55 and Cys123 <i>In Vivo</i>	176
4.3.5	Docking of IQS Analogues to AgrA	180
4.3.5.1	Preliminary Work on IQS Analogues	180
4.3.5.2	Co-crystallisation Attempts of IQS Analogues with AgrA	181
4.3.5.3	<i>In Silico</i> Docking of IQS Analogues to AgrA Using AutoDock Vina	182
4.3.5.4	Predicted Binding Site of the IQS Analogues in Context	196
Chapter 5	AgrB – Expression, Purification, and Characterisation	
5.1	Preliminary Work on AgrB2	198
5.2	Positive Identification of AgrB in the Recombinant Sample	199
5.3	Initial Attempts to Purify AgrB2	202
5.4	Generating a Monodisperse DDM_AgrB2 Sample	204
5.5	Assaying for Activity of DDM_AgrB2	209
5.6	Circular Dichroism Experiments on DDM_AgrB2	211
5.7	Detergent Removal and Retention of the DDM_AgrB2 Unit	213
5.8	Detergent Screening and the Oligomeric State of AgrB2	216
5.9.1	Structural Experiments on AgrB	227
5.9.2	Crystallisation Experiments <i>In Surfo</i>	227
5.9.3	Crystallisation Experiment Attempts <i>In Meso</i>	231
5.9.4	Negative Staining at the University of Warwick	239
5.10	Application to Instruct for Nanobody Preparation at Instruct-ERIC	247



5.11.1	Bioinformatical Interrogation of an AgrB1 Model	248
5.11.2	Basic Topology of the AgrB1 Model	248
5.11.3	Comments on Model Features With Respect to Sample Crystallisability	251
5.11.4	Literature Observations Pertinent to Model Interrogation	251
5.11.5.1	Features of the Model	252
5.11.5.2	Surface Features	252
5.11.5.3	Features in Relation to Residues That Drive AgrD Specificity	254
5.11.5.4	Features in Relation to Possible Arrangement of AgrBs in the AgrB Dimer	256
Chapter 6	Conclusions and Future Research	
6.1	AgrA – Conclusions	264
6.2	AgrB – Conclusions	264
6.3	AgrA – Future Research	265
6.4	AgrB – Future Research	268
6.5.1	Additional Work	273
6.5.2	Work Related to <i>agr</i>	273
6.5.3	Work Unrelated to <i>agr</i>	274
Chapter 7	Methods	
7.1	General Methods	275
7.2	SDS-PAGE Gel Preparation and Running Parameters	275
7.3	Western Blotting	276
7.4	Agarose Gel Preparation	277
6.5	Agarose Gel Electrophoresis	278
7.6.1	AgrA	279
7.6.2	Construct Generation – Wild-Type AgrA, _C199S and D59E_C199S	279
7.6.3	Construct Generation – SER Mutants	281
7.6.4	Construct Generation – Individual AgrA Domains...	284
7.6.5	SER Mutant Test Expression	285
7.6.6	Expression (for Full-Length, Wild-Type and All Mutants)	285
7.6.7	Purification (for Full-Length, Wild-Type and All Mutants)	286
7.6.8	AgrA ₁₋₁₄₁ Purification (N-terminal Domain Only)	287
7.6.9	AgrA ₁₃₅₋₂₃₈ _C199S Purification (C-terminal Domain Only)	288
7.6.10	SEC Peak Shift DNA Binding Assay	288
7.6.11	Oligoduplex Generation [P ₂ fragment and All Variants Thereof]	289
7.6.12	Sparse Matrix Crystallisation Screening	290
7.6.13	Control Crystallisation Experiments	290
7.6.14.1	Crystal Optimisation [Continues Through 7.6.14.14]	291



7.6.15	Cryocooling, Data Collection, Model Building and Refinement	297
7.6.16	Preparation of AgrA_C199S Dimer on Full Promoter Regions	299
7.6.17	Strategy for Introducing [BeF ₃] ⁻ to AgrA_C199S Without Precipitation	301
7.6.18	AgrA_C199S Dimer on P3 _{full} DNA Preparation for SAXS	301
7.6.19	AgrA_C199S Dimer on P3 _{full} DNA SAXS Data Collection	302
7.6.20	Oxidising and Reducing EMSAs on AgrA_C199S	303
7.6.21	EMSA Gel Preparation	306
7.6.22	IQS Analogue Docking to AgrA _C Using AutoDock Vina	306
7.6.23	<i>agrP3::lux</i> Bioreporter Assays and Growth Curves	307
7.7.1	AgrB	308
7.7.2	Construct Generation	308
7.6.3	Expression	309
7.7.4	Purification	309
7.7.5	Thermofluor Assay	311
7.7.6	T7-AgrD Cleavage Assay	313
7.7.7	Circular Dichroism Measurements on DDM_AgrB2	315
7.7.8	Negative Staining	315
Appendices		
8.1	Appendix I – SDS-PAGE Gel Marker	I
8.2	Appendix II – Column Calibration Curves	II
8.3	Appendix III – Preliminary PDB Validation Report	III
8.4	Appendix IV – IQS Analogue Docking Log Files for IQS1 – IQS15	XV
8.5	Appendix V – Summary of <i>In Surfo</i> AgrB Crystallisation Trials	XXXI
8.6	Appendix VI – Instruct Application Documents (Abstract)	XXXV
8.7	Appendix VII – Instruct Application Documents (Motivation Letter)	XXXVI
8.8	Appendix VIII – ITC Sample Preparation Schematic	XXXVII
8.9	Appendix IX – Preliminary... Thermograms for IQS3 ...AgrA_C199S	XXXVIII
8.10	Appendix X – Key Parameters for All Constructs Discussed	XXXIX
Bibliography		XL



List of Figures

Figure	Description	Page
Chapter 1	Primary Introductions and Aims	
1.1	Cartoon overview of the <i>agr</i> system	6
1.2	Sequence alignment of AgrD allelic variants	8
1.3	Crystal structure of the LytTR domain of AgrA in complex with DNA	16
1.4	Representations of response regulator dimer/promoter complexes	20
1.5	Skeletal formulae of select AgrA inhibitors	23
1.6	A representation of a sinusoidal wave	31
1.7	Derivation of the Bragg equation	37
Chapter 2	AgrA – Expression, Purification and Crystallisation	
2.1	Data for select failed AgrA_C199S purification attempts	55
2.2	Data for additional select failed AgrA_C199S purification attempts	57
2.3	Representative chromatographs for a successful AgrA_C199S purification	59
2.4	DNA binding assay results for AgrA_C199S	63
2.5	Initial hit crystals of AgrA in complex with P2 _{fragment} DNA	68
2.6	Control crystallisation experiments	70
2.7	Streak seeding results	75
2.8	Macroseeding <i>in situ</i> results	79
2.9	Comparison of large AgrA_C199S/P2 _{fragment} crystals	90
2.10	Initial X-ray diffraction data	93
2.11	Diffraction data from optimised AgrA_C199S/P2 _{fragment} crystals	97
2.12	Packing plots after molecular replacement of AgrA LytTR domain	99
2.13	Structural homologues of AgrA N-terminal domain for molecular replacement	102
2.14	Example of electron density maps after molecular replacement	104
2.15	Before and after images of reassigned asymmetric unit	107
2.16	Packing plots for rough initial atomic model after molecular replacement	108
Chapter 3	AgrA – Low-Resolution Structure Solution	
3.1	Before and after major fixes to the initial atomic model	111
3.2	Several views of AgrA_C199S/P2 _{fragment} asymmetric unit	113
3.3	The N-terminal domain of the AgrA_C199S/P2 _{fragment} crystals	115
3.4	Hoogsteen base pairing in the AgrA_C199S/P2 _{fragment} crystals	117
3.5	Acidic pocket in the AgrA_C199S N-terminal domain	119
3.6	Proximity of Cys55 and Cys123	122



3.7	Hydrophobic nature of helix 4 in AgrA_C199S N-terminal domain	124
3.8	AgrA SEC chromatograms showing entirely monomeric behaviour	126
3.9	The three crystallographically unique AgrA_C199S dimers in the unit cell	129
3.10	Comparison of AgrA_C199S dimer structure to homologous structures	130
3.11	Comparison of AgrA_C199S dimer interface to homologous structures	132
3.12	Superposition of an AgrA_C199S crystallographic dimer to docking result	138
3.13	MolProbity output for AgrA_C199S/P2 _{fragment} structure	142
3.14	Ramachandran plot for AgrA_C199S/P2 _{fragment} structure	143
3.15	Annotated B-factory putty diagram for the N-terminal domain of AgrA_C199S	145
3.16	Before and after electron density maps about helix 4	146
3.17	Sequence alignment of AgrA to other AgrA homologues	148
Chapter 4	AgrA – Better Crystals and Structure-Activity Experiments	
4.1	Targets for SER	152
4.2	Test expression result for SER mutants	153
4.3	Affinity chromatography data for SER mutants	157
4.4	Post-affinity chromatography gel data for SER mutants	158
4.5	SEC data for SER mutants	159
4.6	Post-SEC gel data for SER mutants	160
4.7	Purification data for AgrA ₁₋₁₄₁	163
4.8	Potential AgrA ₁₋₁₄₁ crystals	165
4.9	Sample quality control data for AgrA_C199S/P3 _{full} SAXS measurements	168
4.10	Key SAXS measurement charts	169
4.11	Molecular envelope of the AgrA_C199S/P3 _{full} sample	170
4.12	Oxidising and Reducing EMSAs for AgrA_C199S	174
4.13	Oxidising and Reducing EMSAs for AgrA_C199S show as heat maps	175
4.14	AgrA mutant EC ₅₀ curves	178
4.15	AgrA mutant growth curves	179
4.16	IQS2 docking result #1.	188
4.17	IQS3 docking result #1	188
4.18	IQS4 docking result #3	189
4.19	IQS5 docking result #1	189
4.20	IQS7 docking result #4	190
4.21	IQS8 docking result #3	190
4.22	IQS9 docking result #1	191
4.23	IQS10 docking result #4	191
4.24	IQS11 docking result #1	192



4.25	IQS12 docking result #5	192
4.26	IQS13 docking result #1	193
4.27	IQS14 docking result #3	193
4.28	IQS15 docking result #1	194
4.29	Spacefill representations of docking results for IQS3 and IQS11 on AgrA _C	195
Chapter 5 AgrB – Expression, Purification, and Characterisation		
5.1	AgrB sample confirmation by western blot and mass spectrometry	200
5.2	Thermofluor assay results for AgrB2	205
5.3	Before and after size exclusion chromatograms for monodisperse AgrB2	207
5.4	Representative gel data for successful AgrB2 purification	208
5.5	AgrD2 cleavage activity assay results	210
5.6	Circular dichroism measurements on AgrB2	212
5.7	Spin-concentrator retention assay for AgrB2	215
5.8	Skeletal formulae for key detergent molecules	219
5.9	SEC data for AgrB2 detergent screening experiments	220
5.10	Gel data for AgrB2 detergent screening experiments	221
5.11	Purification data for AgrB1 and sequence alignment to AgrB2	224
5.12	Gel result for 80 kDa contaminant peak common to AgrB SEC results	226
5.13	AgrB pseudo crystals	228
5.14	Diffraction data from AgrB pseudo crystals	229
5.15	Western blot result from AgrB pseudo crystals	230
5.16	Membrane protein crystal types	233
5.17	<i>In meso</i> crystallisation experiment stages	236
5.18	Crystals resulting from <i>in surfo</i> experiments on excess <i>in meso</i> sample	237
5.19	AgrB2_(His) ₁₀ size exclusion chromatogram comparison to AgrB1_(His) ₆	238
5.20	LMNG_AgrB2 EM Grid 1	241
5.21	LMNG_AgrB2 EM Grid 2	242
5.22	LMNG_AgrB2 EM Grid 3	243
5.23	DDM_AgrB2 EM Grid 1	244
4.24	DDM_AgrB2 EM Grid 2	245
5.25	Zoom-in on EM grid features	246
5.26	Schematic representation of the AgrB1 model	250
5.27	Cartoon, electrostatic surface and conservation of the AgrB1 model	253
5.28	AgrB1 and AgrB2 hydrophobicity plots	255
5.29	<i>In silico</i> predictions of the AgrB1 dimer	269
5.30	Schematic representation of an AgrB1 dimer predicted by ClusPro	260



5.31	Residues involved in the AgrB1 dimer predicted by ClusPro	261
5.32	Sequence alignment of AgrB1 to other staphylococcal AgrBs	262
Chapter 6	Conclusions and Future Research	
6.1	Potential curly arrow mechanism for AgrD processing by AgrB	271



List of tables

Table	Description	Page
Chapter 2	AgrA – Expression, Purification and Crystallisation	
2.1	Crystal data used to build the initial AgrA_C199S/P2 _{fragment} model	109
Chapter 3	AgrA – Low-Resolution Structure Solution	
3.1	Crystal data used to build the AgrA_C199S/P2 _{fragment} structure solution	147
Chapter 4	AgrA – Better Crystals and Structure-Activity Experiments	
4.1	Summary of SER constructs	153
4.2	Summary of conditions that generated prospective AgrA ₁₋₁₄₁ crystals	165
4.3	Skeletal structures of all IQS analogues docked to AgrA	185
4.4	Summary of predicted binding energies for IQS analogue docking results	187
Chapter 5	AgrB – Expression, Purification, and Characterisation	
5.1	Summary of key detergents and properties	218
Chapter 7	Methods	
7.1	SDS-PAGE gel recipes	275
7.2	TAE buffer recipe	277
7.3	PCR reaction summary for: wild-type AgrA, _C199S and _D59E_C199S	279
7.4	Construct sequences for: wild-type AgrA, _C199S and D59E_C199S	280
7.5	SER mutant primers	282
7.6	PCR reaction summary for SER mutants	282
7.7	SER construct sequences	283
7.8	Primers and construct sequences for AgrA ₁₋₁₄₁ and AgrA ₁₃₅₋₂₃₈	284
7.9	Oligonucleotide summary for all P2 _{fragment} variants	285
7.10	Oligonucleotide summary for P2 _{full} and P3 _{full}	300
7.11	EMSA sample buffer recipes	303
7.12	AgrA_C199S EMSA stock solution components	303
7.13	Oligonucleotide summary for EMSA probe	304
7.14	EMSA TBE running buffer recipe	304
7.15	Summary of EMSA reactions	306
7.16	EMSA gel recipe	306
7.17	Construct sequences for all AgrB1 and AgrB2 variants	308
7.18	Summary of thermofluor reactions	312
7.19	Summary of T7-AgrD cleavage assay reactions	314
7.20	T7-AgrD cleavage assay buffer recipe	314



List of Acronyms and Abbreviations

Shorthand	Full Description
ADP	Adenosine diphosphate
<i>agr</i>	Accessory gene regulator
AIDS	Acquired immunodeficiency syndrome
AIP	Auto-inducing peptide
AMR	Antimicrobial resistance
APS	Ammonium persulphate
ATP	Adenosine triphosphate
Au	Absorbance units
BME	Beta mercaptoethanol
bp	Base pair
BrDU	Bromouracil
Buff	Buffer
CD	Circular dichroism
CMC	Critical micelle concentration
CPM	7-Diethylamino-3-(4'-maleimidylphenyl)-4-methylcoumarin
Cryo-EM	Cryo-electron microscopy
Da	Dalton (equivalent to 1 g mol^{-1})
DDM	n-Dodecyl β -D-maltoside
DLS	Dynamic light scattering
DMNG	Decyl maltose neopentyl glycol
DMSO	Dimethyl sulfoxide
DNA	Deoxyribose nucleic acid
DOPG	1,2-Dioleoyl-sn-glycero-3-phospho-rac-(1-glycerol)
dsDNA	Double-stranded DNA
DTT	Dithiothreitol
ECL	Enhanced chemiluminescence
EDTA	Ethylenediaminetetraacetic acid
EM	Electron microscopy
EMBL	European molecular biology laboratory
EMSA	Electrophoretic mobility shift assay
ERIC	European research infrastructure consortium
FC-12	n-Dodecylphosphocholine



FDA	Food and drug administration
Fig	Figure
Frac	Fractions
FT	Flow through
g	Gravity, force of (equivalent to 9.81 m s^{-1})
GF	Gel filtration
GFP	Green fluorescent protein
GI	GenInfo identifier
HAART	Highly active antiretroviral treatment
HEPES	(4-(2-Hydroxyethyl)-1-piperazineethanesulfonic acid), buffer
HIV	Human immunodeficiency virus
HPLC	High-performance liquid chromatography
Hrs	Hours
HPK	Histidine protein kinase
IUPAC	International union of pure and applied chemistry
IPTG	Isopropyl β -D-1-thiogalactopyranoside
ITC	Isothermal titration calorimetry
LB	Lysogeny broth
LDAO	Lauryldimethylamine-N-oxide
LLG	Log likelihood gain
LMNG	Lauryl maltose neopentyl glycol
MAG	Monoacyl glycerol
MGL	Molecular Graphics Laboratory
Mins	Minutes
MOF	Metal-organic framework
MR	Molecular replacement
M_r	Relative mass (in g mol^{-1})
MS	Mass spectrometry
NEB	New England biolabs
NMR	Nuclear magnetic resonance
nmRC	Nanoscale and microscale research centre
No/ N°	Number
OD	Optical density
PCR	Polymerase chain reaction
PDB	Protein data bank
PEG	Polyethylene glycol



PhD	Doctor of Philosophy
PMSF	Phenylmethanesulfonyl fluoride
PPGBA	Poly(propylene glycol) bis(2-aminopropyl ether)
PSM	Phenol-soluble modulin
PVDF	Polyvinylidene difluoride
QS	Quorum sensing
RLU	Relative light units
RMSD	Root mean square deviation
RNA	Ribonucleic acid
RP-HPLC	Reverse-phase high-performance liquid chromatography
rpm	Revolutions per minute
SAD	Single-wavelength anomalous dispersion
SANS	Small-angle neutron scattering
SAXS	Small-angle X-ray scattering
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulphate – polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
Secs	Seconds
SEM	Scanning electron microscopy
SER	Surface entropy reduction
SMALP	Styrene-maleic acid lipid particle
SPR	Surface plasmon resonance
ssDNA	Single-stranded DNA
TB	Terrific broth
TBE	Tris-borate-ethylenediaminetetraacetic acid, buffer
TBS	Tris-buffered saline, buffer
TBS _t	Tris-buffered saline with tween® 20, buffer
TCEP-HCl	Tris(2-carboxyethyl)phosphine hydrochloride
TEM	Transmission electron microscopy
Term	Terminus, pertaining to protein domain or end
TEMED	N,N,N',N'-tetramethylethylenediamine
TLC	Thin-layer chromatography
UCSF	University of California San Francisco
VIB	Vlaams instituut voor biotechnologie
WB	Western blot
YFP	Yellow fluorescent protein



Chapter 1: Primary Introductions and Aims

1.1.1: Infectious Diseases and the *Staphylococcus aureus* agr System

The history of modern humanity is vividly coloured by those organisms with the ability to kill or debilitate us. The slow sophistication of medicine could be roughly charted as a timeline of epidemics followed by humanitarian response. For example: the bubonic plague outbreak in the 14th century; the spread of smallpox across the Americas and the major outbreak of the 18th century;^[1] cholera in 19th century London; polio and Human Immunodeficiency Virus (HIV) in the 20th century and several coronaviruses in the early 21st century. From bloodletting to antimicrobial pills and vaccination, pioneering work by celebrated figures such as Edward Jenner,^[2] Alexander Fleming^[3] and Jonas Salk^[4] has rendered most of these diseases anachronisms. More challenging illnesses, such as HIV for example, remain embedded in the human population^[5] and require treatments for which development and validation is complicated and expensive.^[6] And so it often is when treatments must discriminate between the host and the causative agent whilst overcoming issues associated with bioavailability.

Therapies for some of the diseases listed above are rightly well known, for example the Highly Active Antiretroviral Treatment (HAART) treatment of Acquired Immunodeficiency Syndrome^[7] (AIDS) and the small-molecule antibiotics penicillin and methicillin used to combat a variety of common bacterial infections.^[8] Inevitably some therapies do not acquire the same levels of widespread awareness as those listed above despite saving many lives. For example, the Ervebo® vaccine against the extremely deadly ebola virus recently gained Food and Drug Administration (FDA)-approval and was efficaciously prescribed to help control outbreaks of ebola prior to gaining this status.^[9] Despite its early success in preventing ebola transmission and the worldwide attention garnered by the 2013 West African outbreak however, its FDA-approval was a comparatively minor news story.

Severe illness are not restricted to viral infections and the growing problem of antimicrobial resistance^[10] (AMR) is arguably less tractable (to say nothing of antifungal resistance) than the problems associated with vaccination against debilitating viruses as the latter are not endemic in healthcare facilities, in which patients are usually already ill. Being acellular they also cannot



reproduce independently as bacteria do. This is not to imply, however, that common or rare viruses pose no threat when introduced to hospital wards and nursing homes.

The rapidity in which bacterial generations live and die, horizontal gene transfer between bacterial cells and over-prescription of antibiotic drugs are three of the reasons why approved antibiotic medicines are becoming less capable of curing infections against which they were historically effective. This has two major corollaries: A) the roster of effective antibiotics is shrinking and B) the range of antibiotic-resistant bacterial strains is increasing. These in turn lead to increased chances of serious patient illness or even death occurring from infection by resistant strains and increased chances of complications during routine medical operations. For those who are already infirm such risks may be dramatically increased. For example, estimates of the current annual death toll in the United Kingdom from sepsis is 40,000 – 50,000 and this figure is increasing due to the growing number of resistant strains.^[11] Furthermore, which bacterial strains may acquire resistance and when they might do so is unpredictable. Consequently the development of antibiotic treatments is always reactive, whilst proactive development and validation of new treatments requires a large investment of labour and financial capital.^[12]

To mitigate this rapid resistance to new drugs, and the humanitarian and scientific problems that accompany it, there is a growing interest amongst the research community in non-lethal, so-called second generation antibiotics. Unlike early antibiotics such as penicillin, these aim to be selective to the target pathogen. Although acquisition of antibiotic resistance is inevitable, and new pathogenic strains will always be evolving along with all life for as long as it continues to exist, it is hoped that this non-lethality will lessen the Darwinian selection pressure applied by bactericidal and bacteriostatic antibiotics so as the new medicines have a longer window of efficacy. Their selectivity meanwhile ought to lessen the negative side effects associated with some antibiotics such as allergies,^[13] intolerances^[14] and even mental health.^[15] Many of these detrimental side effects are only now being correlated with and tentatively attributed to the effects of current antibiotics and their inappropriate prescription, and the field of study is simultaneously fashionable and somewhat controversial.



Amongst the problematic bacterial species, *Staphylococcus aureus* is a Gram-positive human commensal bacteria which is notoriously rapid at acquiring resistance to new antibiotic drugs. Less than two years after the development of methicillin by the Beecham Group^[16] the first methicillin-resistant *S. aureus* (MRSA) strain was reported.^[17,18] Vancomycin, whose glycopeptide structure is radically different β -lactam-type antibiotics, was introduced in 1958^[19] and a resistant *S. aureus* strain was isolated from an infected patient in Michigan in 2002.^[20] This is in spite of the fact that vancomycin was never used as a first response drug due to its low bioavailability and acute detrimental side effects.^[21] *S. aureus* strains now carry an impressive arsenal of enzymes which enable it to resist most current antibiotics including β -lactamases, aminoglycoside-modifying enzymes and enzymes which cause efflux of the antibiotic out of the bacterium.^[22] This ability of *S. aureus* to rapidly acquire resistance to new antibiotics is compounded by the number and severity of infections it is capable of causing which include: meningitis,^[23] scalded skin syndrome,^[24] toxic shock syndrome,^[25] pneumonia^[26] and endocarditis^[27] amongst many others.

Crucially, production of virulence factors by *S. aureus* during its growth phase within the host have been observed to be somewhat dependant on one another.^[28,29] These studies observed mutants which suppressed production of a single virulence factor such as protein A,^[30] which binds the host's immunoglobulins to prevent phagocytosis, or haemolysin,^[31] which causes pitting and eventual death of red blood cells. These suppressions were observed with concomitant upregulation of other virulence factors which together provided evidence for some master chromosomal region controlling *S. aureus* virulence factor expression. This was eventually termed accessory gene regulator^[29] (*agr*).

Efforts to locate and sequence the *agr* locus on the *S. aureus* chromosome lead first to identification of the *agrBDCA* genes, expression of which is governed by the adjacent, downstream promoter, P2. The single transcription product, RNAII, encodes four protein gene products: AgrB, AgrD, AgrC and AgrA.^[32] The same group published a further study^[33] demonstrating that a second promoter, P3, is transcribed divergently to P2 and 186 bp (Base Pairs) further downstream of it. Its transcription product, RNAIII, was shown to be responsible for the production of a multitude of virulence factors and exotoxins. This was achieved by artificially inducing RNAIII production in host cells containing



RNAIII-inducible plasmids transformed into Δagr *S. aureus* cells and quantizing the subsequent gene expression profile. A third promoter, P1, was also discovered but it only drives transcription of AgrA and is not considered to be of great scientific interest.^[33] Sequencing of these promoter regions indicated that they are each a pair of linear repeats containing a 9 bp consensus sequence separated by 12 bp for the P2 promoter and 16 bp for the P3 promoter (originally determined as a 20 bp spacer assuming the AgrA binding sequence was 7 bp in length).^[34,35,36]

Another significant piece of the puzzle insofar as the nature of the *agr* locus was discovered when it was shown that the *agr* system could be activated independently of *S. aureus* cell growth by exposing cells to spent media^[37] indicating some extracellular signal was tying the entire *agr* system together. Mass Spectrometry (MS) on the cell growth media revealed a cyclic thiolactone to be the extracellular signal, called an Auto Inducing Peptide^[38] (AIP). This seminal study demonstrated that the *agr* system was part of an intercellular communication mechanism that enables the commensal bacteria to coordinate their combative activity and host immune response to better increase chances of establishing an infection and ultimately survive. This type of intercellular communication is known as Quorum Sensing (QS) and is observed in many bacterial pathogens.^[39,40,41] Crucially, the response coordination is dependent upon cell population density and the coordination cascade is triggered when a “quorum” is present.

A detailed introduction to the components that comprise the *S. aureus* QS system is presented on the following pages, however, it is worth first presenting a brief textual description of how the system functions. AgrD is the linear peptide precursor to AIP. The transmembrane protease AgrB cleaves the C-terminus from AgrD and also catalyses internal cyclisation of AgrD to a thiolactone intermediate. This cyclised intermediate is then somehow transported outside of the cell, where the surface peptidase SpsB cleaves the N-terminus to release the mature AIP signal molecule. At a critical concentration AIP binds to its cognate receptor, the histidine kinase AgrC. This triggers an allosteric signal through the cell membrane via AgrC to initiate phosphate-transfer from adenosine triphosphate (ATP) to the transcriptional regulator protein AgrA via a dedicated cytoplasmic phosphate-transfer domain on AgrC. Phosphorylation triggers dimerisation of AgrA, for which the P2 promoter has the



higher affinity. Phosphorylated AgrA subsequently binds to the P2 promoter as a dimer to drive transcription of the *agrBDCA* genes, positively reinforcing the QS feedback loop leading to increased concentrations of AIP and phosphorylated AgrA. Once the cytoplasmic pool of phosphorylated AgrA is sufficiently high in concentration it binds to the lower affinity P3 promoter to drive production of virulence factors and exotoxins, triggering a phenotypic switch to virulent lifestyle. This mode of operation is presented in figure 1.1.

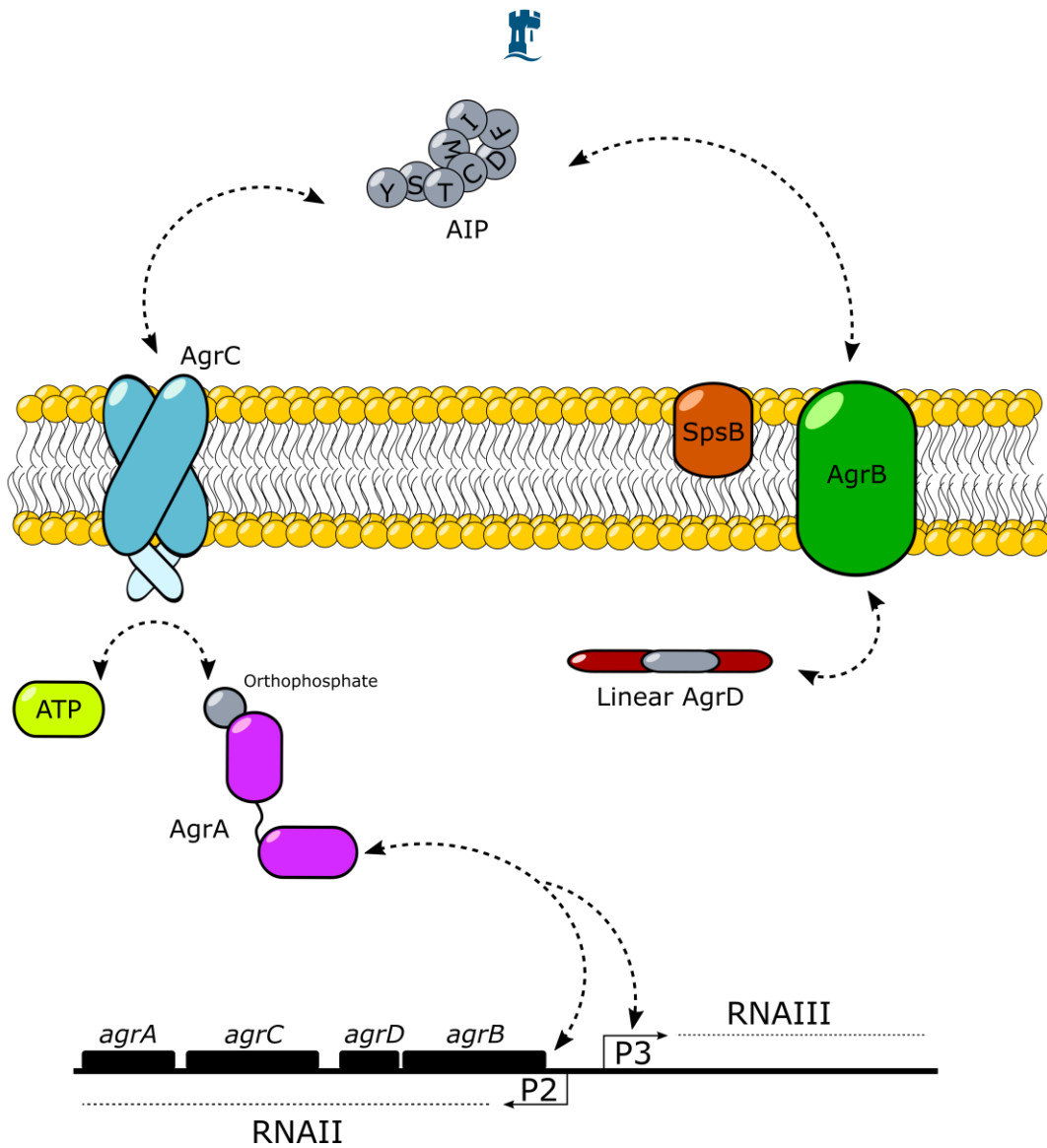


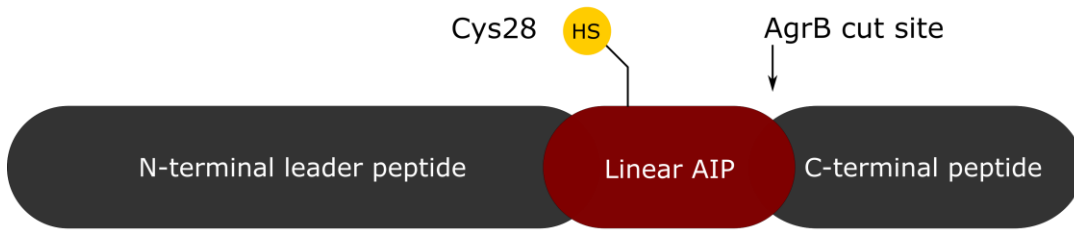
Figure 1.1. A cartoon representation of the *agr* QS circuit in *S. aureus*. The peptide AgrD (grey and red) undergoes C-terminal cleavage and cyclisation by AgrB (green) to produce a cyclised intermediate of the AIP unit covalently bound to the N-terminal fragment of AgrD. This intermediate undergoes N-terminal cleavage by SpsB (orange) to release the mature AIP molecule (grey). At a critical concentration, AIP binds to its cognate receptor, AgrC (blue) which catalyses the transfer of orthophosphate from ATP (lime) to the response regulator, AgrA (purple). Upon phosphorylation, AgrA dimerises and binds to pairs of linear repeat promoter sequences (black) to upregulate the *agr* system and a number of virulence factors. The oligomeric state of AgrB and the transport mechanism of the cyclised AIP outside of the cell are unknown.



Before introducing each of the components which comprise the *agr* system in more detail it must be noted that the entire *agr* system exhibits divergence across *S. aureus* strains. This divergence was first observed in AIP with four distinct sequences being reported.^[38] Eventually four allelic variants of the system were described: *agr-1*, *agr-2*, *agr-3* and *agr-4* each of which encodes an AIP unique to that particular allele. Unsurprisingly these four AIP variants are produced by four allelic variants of AgrB and are recognized by four allelic variants of AgrC. AgrA however, is conserved across the four *agr* groups.^[38] The sequence identities of AgrD2, 3, and 4 relative to AgrD1 are 47.8, 52.2, and 87.0 % respectively. The corresponding values for the four AIPs are 33.3, 60.0, and 87.5 % and the corresponding values for the four AgrBs are 79.1, 60.4, and 89.7 %.

1.1.2: AgrD

AgrD is the peptide precursor to the signal molecule AIP. AgrD1 contains 47 amino acids whilst AgrD2, 3 and 4 contain 46. Broadly speaking these sequences can be split into three regions.^[42] The N-terminal ~ 24 amino acids form an amphiphilic helix which is presumed to target the peptide to the inner membrane so as it may encounter AgrB. Evidence for this function comes from replacing these residues in AgrD1 with an artificial amphiphilic helix. Doing this still allows for its processing into AIP1. Deletion of the first 14 amino acids in AgrD1 however, abolishes production of AIP1.^[43] The central region of the AgrD amino acid sequence is the linear precursor to the mature AIP signal molecule which is formed via catalytic C-terminal cleavage of AgrD and cyclisation by AgrB.^[37] The C-terminal region of AgrD is also negatively charged and is the most highly conserved region, with many of these conserved amino acids being absolutely necessary for processing of AgrD to AIP.^[42] It has also been shown that the C-terminal AgrD cleavage product must be rapidly sequestered away from the reaction to drive the equilibrium position towards AgrD cleavage, and the reaction is significantly retarded if this is not the case.^[44] Finally AgrD is phenol-soluble modulins (PSM)-like and consequently toxic to host cells when overexpressed for recombinant studies.^[45,46] The four AgrD sequences corresponding to the four *agr* groups are given in figure 1.2.

**A****B**

CLUSTAL O(1.2.4) multiple sequence alignment

```

                                     AgrB cut site
                                     ↓
AgrD1      MNTLFNLFDFITGILKNIGNIAAYSTCDFIMDEVEVPKELTQLHE- 46
AgrD2      MNTLVNMFFDFIIKLAKAIGIVGGVNACSSLFDEPKVPAELTNLYDR 47
AgrD3      MKKLLNKVIELLVDFNSIGYRAAYINCDFLLDEAEVPKELTQLHE- 46
AgrD4      MNTLLNIFDFITGVLKNIGNVASYSTCYFIMDEVEVPKELTQLHE- 46
          *:.*. * .::: . : ** .. * ::** :** ***:***

```

Figure 1.2. Two representations of the AgrD peptide. (A). A simplified schematic view showing key features. (B). A Clustal Omega sequence alignment of the four allelic variants of AgrD^[38] annotated as follows: (*), completely conserved, (:), highly conserved, (.), weakly conserved. The amino acids corresponding to the mature AIP signal molecule are coloured red. The cysteine residue which displaces that of AgrB to form the cyclic thiolactone-bridged peptide are coloured green. The N-terminal residues before the AgrB cut site that are coloured black form an amphipathic helix. The C-terminal residues contain abundant negative charge, and the cleavage product must be rapidly sequestered away from the reaction site to drive AIP biosynthesis.



1.1.3: AgrB

Being an integral membrane protease AgrB is intrinsically difficult to study owing to the problems associated with overexpressing membrane proteins and obtaining stable and active forms of the sample in aqueous solutions.^[47,48,49] Furthermore invocation of detergent and lipid technologies^[50] to generate aqueous sample solutions often complicates and interferes with experimental design. Beyond its demonstrated role in C-terminal cleavage and cyclisation of AgrD to generate the thiolactone-containing AIP precursor,^[42,51] and the necessity of residues His77 and Cys84 for catalytic activity^[52] little is known about AgrB and its structure-activity relationships. A number of presumed structural homologues found in both of Gram-positive and -negative bacteria that exhibit sequence diversity (< 30 % identity relative to the primary sequence of AgrB1) are all presumed to belong to the “AgrB class” of proteases.^[53]

Published hydrophobicity plots suggest it contains 6 transmembrane helices and a possibly a 27 amino acid loop which extends across the outer membrane leaflet.^[51] Exact mechanistic insights on AgrB have remained elusive until the most recent study to-date^[44] unambiguously proved that AgrB is responsible for cyclisation of AgrD by reconstituting both AgrB and AgrD into nanodiscs and performing Reverse-Phase High-Performance Liquid Chromatography HPLC (RP-HPLC). Flying aliquots of the reaction mixture down the RP-HPLC column at different time intervals followed by MS on the separated species allowed for positive identification of the reaction products to be made whilst integration of the chromatographic peaks at each of the intervals allowed a detailed picture of the reaction kinetics to be established. The data showed cleavage of AgrD generated two products: the AgrD C-terminus fragment (AgrD_C) and the AgrD N-terminal helix attached to the cyclised AIP fragment (AgrD_{N-AIPcyclised}).

The relative concentration of AgrD decreased dramatically within the first 20 minutes in line with a sharp increase in AgrD_C as expected. However, AgrD_{N-AIPcyclised} reached a maximum concentration within the first 20 minutes before decreasing steadily with time over the entire 2 hour reaction. A steady increase in a third product, identified as the AgrD N-terminal helix attached to the linear AIP



fragment ($\text{AgrD}_{\text{N-AIPlinear}}$), was observed over the entire 2 hour reaction. Thus the data suggests two reactions are taking place. The first is a fast and reversible proteolytic cyclisation of AgrD to give $\text{AgrD}_{\text{N-AIPcyclised}}$ and the second is a slow, irreversible ring opening of $\text{AgrD}_{\text{N-AIPcyclised}}$ to give $\text{AgrD}_{\text{N-AIPlinear}}$.

The kinetic data demonstrated that a rapid removal of AgrD_{C} is required to drive the formation of the thiolactone intermediate, which is retarded if this fragment is allowed to accumulate, and so it has been postulated that the conserved residues in the cleaved C-terminus target it towards some degradation pathway. In addition to this the group were able to show that AgrB2 is dimeric by crosslinking AgrB2 with glutaraldehyde and employing native MS, and demonstrated that it is only active when embedded into lipid bilayers.^[44]

As yet there are no studies probing what role, if any, AgrB has in transferring the cyclised cleavage product to the extracellular region although mature AIP is detectable when only AgrD and AgrB are overexpressed in *E. coli*.^[42] Furthermore there are few studies examining how the N-terminus is cleaved from the cyclised AIP to give the mature signalling molecule. A generic surface peptidase, SpsB, has been implicated.^[54] This was achieved by demonstrating inhibition of mature AIP production in the presence of SpsB inhibitors.

Pharmacological inhibition of AgrB is of interest owing to the essential role AgrB plays in AIP biosynthesis. Blocking the biosynthetic pathway via AgrB inhibition should in theory mute the *agr* system and attenuate *S. aureus* virulence. In one seminal study it was demonstrated that the chemical ambuic acid acts as an inhibitor of cyclic peptide signal molecule biosynthesis in a variety of Gram-positive bacteria^[55] including: *S. aureus*, *Staphylococcus epidermidis*, *Staphylococcus saprophyticus*, *Staphylococcus lugdunensis*, *Listeria monocytogenes* and *Enterococcus faecalis*. Additionally, for *S. aureus* and *S. epidermidis*, *agr* groups 1, 2 and 3 were all tested. Quantitative MS allowed for determination of AIP concentrations after exposure of each of the strains to ambuic acid. The 50 % inhibitory concentration values (IC_{50}) were of the order 2 – 20 μM for most of the strains, however, *S. lugdunensis agr* type 1 and *S. epidermidis agr* groups 2 and 3 had much higher IC_{50} values. For



completion, at the time of the study discussed ambuic acid was already known to have antifungal properties,^[56] and the hypothesis that the AgrB family of proteins was the target for ambuic acid stemmed from the observed reduction in FsrD (the *E. faecalis* AgrD analogue) turnover by FsrB^[57] (the *E. faecalis* AgrB analogue).

1.1.4: AIP

AIP is the signalling molecule to which the *agr* system responds and is formed by displacement of the AgrB Cys84-AgrD linear thioester intermediate by the side chain of AgrD Cys28 to generate a thiolactone.^[37,57] Thiolactones are not the only class of signalling molecules employed in QS systems in the bacterial kingdom, however. *Staphylococcus intermedius* secrete regular lactones,^[59] *Streptococcus pneumoniae* secrete linear heptadecapeptides,^[60] whilst Gram-negative bacteria generally employ small-molecule signal molecules such as N-acyl-homoserinelactones.^[61] Formation of a thiolactone is theoretically less energetically favourable than formation of a regular lactone since a C-S bond (272 kJ mol^{-1}) is being formed and these are weaker than C-O bonds (358 kJ mol^{-1}).^[62] Thus $\sim 86 \text{ kJ mol}^{-1}$ less energy is stored as bond energy. Nor is cyclisation entropically favoured since the cyclised product has few conformations which it may assume. Presumably the bond energy released in cleaving the C-terminus of AgrD and the extra entropy afforded by having two discrete peptide fragments is sufficient to compensate for this apparent deficit. Proton nuclear magnetic resonance (NMR) studies on *S. aureus* AIPs have demonstrated the relevant protons to be solvent-shielded^[63] and therefore less susceptible to chemical attack. This probably helps to increase the functional lifetime of the AIP in the extracellular space, which is $\sim 3 \text{ hrs}$,^[64] by retarding hydrolysis of the thioester bond. This shielding may be related to the relative rigidity of the cyclic AIP molecule and may help binding to AgrC by presenting fewer possible conformations to it upon approach. The reported equilibrium dissociation constants, K_d , for AIPs binding to AgrC are in the region of $10 - 30 \text{ nM}$ ^[65] (from EC_{50} curves) indicating AgrC is extremely sensitive to intercellular AIP concentrations.

All of the four AIPs of the *S. aureus agr* system contain a thiolactone ring formed of five amino acids however, the number of “overhanging” N-terminal residues ranges from 2 – 4.^[66] There are a number



of studies which have explored these differences. AIP1 and AIP4 differ by a single D5Y mutant and appear to be functional identical, whilst AIP2 and AIP3 are distinct from one another and AIP1 and AIP4. The three subclasses appear to be involved in a cross-inhibitory system between *S. aureus* strains and each AIP can bind to, but not activate, a non-cognate AgrC. It is not unequivocally established however, that this apparent cross-inhibition is truly combative and may yet prove to be an additional method of cell-to-cell communication.^[66,67]

1.1.5: AgrC

AgrC is the receptor of AIP. At a threshold concentration AIP binds to the hosting site on the outer membrane region of AgrC whereupon it phosphorylates AgrA within the cell cytoplasm.^[68] As such the signal is allosteric. This mechanism of action is of the relatively well-understood Histidine Protein Kinase, (HPK)-type, and indeed AgrC is structurally homologous with other HPKs.^[69] Upon AIP binding, AgrC cleaves a phosphate anion from ATP by allowing ATP to enter its binding pocket on the kinase subdomain. The phosphate ion is transferred from ATP to a histidine residue (His239 on AgrC1) on the adjacent phosphorylation subdomain of AgrC, with the concomitant release of adenosine diphosphate (ADP). This phosphate ion is then transferred to the response regulator AgrA.

This mechanism of action has been exploited to demonstrate that AgrC is active as a dimer and that signal transduction proceeds by so-called trans-autophosphorylation.^[70] Without any prior knowledge of AgrC oligomerisation it would be reasonable to assume that once AIP has bound to AgrC to induce the allosteric signal, AgrC receives ATP at its kinase subdomain and the phosphate ion is transferred to the receiving histidine residue in the phosphorylation subdomain on the same AgrC protomer. If this were the case then mutating the phosphate-receiving histidine in AgrC ought to prevent its ability to phosphorylate AgrA in a β -lactamase reporter assay. Such mutants did not exhibit inhibition however, and β -lactamase activity was reported in spite of the mutation of this essential subdomain.^[70] This suggested that some other species was apparently compensating for the deliberately-introduced deficiency. The hypothesis that AgrC exists as a homodimer, with each protomer being able to compensate for deficiencies in its partner AgrC, was tested by co-expressing two AgrC molecules



each containing a unique set of point mutants. One had the kinase subdomain mutated to be inactive and the other had the phosphorylation subdomain mutated to be inactive. Furthermore, additional point mutations were introduced asymmetrically to the dimeric pair to ensure each was totally inactive when expressed alone. Any activity therefore must be due to the dimeric pair being proximal to one another and with each protomer able to compensate for the subunit deficiency on the partner protomer. Activity was observed from the β -lactamase reporter when the asymmetric mutants were co-expressed and the hypothesis of trans-autophosphorylation was proven.

There is an experimental structure available for the ATP-binding subdomain of AgrC^[71] (PDB: 4bxi) and this has been used to probe for AgrC...AgrA interactions by generating a homology fold for AgrA and docking it to the experimental AgrC structure. There is very little discussion of the resulting predicted interface in the corresponding paper, however.^[71]

It is common for therapeutic intervention on infections caused by microorganisms to mimic or extend what occurs naturally in an environment in which said microorganism is commonly found.^[72] In fact many antibiotic medicines are derived from compounds secreted by Gram-negative anaerobes and fungi found in soil which are in competition with one another for nutrients.^[73] The carbapenems^[74] and cephalosporins^[75] are important examples of this. The same holds true for the body of work concerning AgrC inhibition, of which the most expansive study demonstrated that simplified AIP derivatives could inhibit *agr* in the multi-drug resistant clinical *S. aureus* strain USA300.^[76] A relatively large library of 63 compounds was screened in a Yellow Fluorescent Protein (YFP) reporter assay in which all four *agr* groups were tested. The IC₅₀ values varied significantly but were generally in the 100 – 500 nM range. Three structural elements were reported to be necessary for potent AgrC inhibition and these were: A) aromatic residues at positions 3 and 4 in the macrocyclic ring, B) a thioester linkage in the macrocycle with amide derivatives exhibiting abolished AgrC inhibition and C) a septyl alkane chain, with octyl chains showing reduced potency. These observations clearly relate to structural features present in the parent AIP molecule which are necessary for selective AgrC binding and cross-communication/cross-inhibition.^[66,67]



The requirement for a thioester linkage in the aforementioned inhibitor study is contrast to a report in which amide-linked AIP analogues were shown to also inhibit AgrC and possess greater hydrolytic stability than their respective thioester analogues.^[77]

1.1.6: AgrA

AgrA is the response regulator of the *agr* system and belongs to the unusual LytTR class^[34] which account for only ~ 2.7 % of all prokaryotic response regulators.^[78,79] Other than AgrC it is the only *agr* QS protein with any experimental structural data^[80] (PDB: 3bs1). Only the residues 136 – 238 comprising the DNA-binding domain of AgrA were crystallised, however, this was in complex with an oligoduplex corresponding to the upstream AgrA binding site on the P2 promoter. The crystal structure, solved at 1.6 Å, shows AgrA bending the oligonucleotide through ~ 30 °. Incidentally the crystal structure of the DNA-binding domain of AgrA was the first experimental structure for a LytTR-type protein. The fold consists of 10 β-strands divided into 3 antiparallel β-sheets each containing 3, 5 and 2 strands. These sheets are anchored in place by two short helices whilst a third, longer helix sits across the face of and parallel to the central β-sheet. The overall appearance of the domain is reminiscent of a fist where the knuckles correspond to the loop regions that interdigitate with the DNA major and minor grooves. These loops exist between the β-sheets and contain two basic residues which form direct contacts with DNA nucleobases. These are His169 and Arg233, both of which nestle into adjacent major groves. The residue Asp201 nestles into the intervening minor grove and contacts a thymine nucleobase through a water-mediated H-bond. There are abundant other direct and water-mediated contacts between AgrA residues and the DNA sugar-phosphate backbone. The DNA-binding domain of AgrA in complex with DNA is shown in figure 1.3.

The strength of the interactions between AgrA and P2 and P3 regions of *S. aureus* DNA has been quantified via gel shift mobility assays.^[81] The K_d values for phosphorylated AgrA were reported as 0.16 nM for P2 and 1.7 nM for P3. These values are an order of magnitude lower (higher affinity) than the corresponding values reported for unphosphorylated AgrA. Unsurprisingly these values increase by two orders of magnitude (lower affinity) when those residues which make direct contacts with



DNA (fig. 1.3B) were mutated. Prior to these binding studies there was some doubt as to whether AgrA was a response regulator at all, owing to the hitherto paucity of data demonstrating a direct interaction between AgrA and DNA^[82] and the hitherto unclassified fold that it appeared to exhibit.

In the same study in which the K_d values were reported,^[81] DNase1 protection assays and purine sequencing allowed for the simultaneous exact quantisation of the P2 and P3 promoter element sequences by examining areas of DNA protected from the DNase1 enzyme as well as areas which appeared to exhibit hyper-sensitivity to cleavage. The latter are indicative of DNA curvature and were observed in these assays,^[83] demonstrating that at least part of the mode of P2 and P3 upregulation by AgrA was by the commonly-observed method of DNA bending, which increases its surface area relative the unbent form and results in RNA polymerase (RNAP) recruitment.^[84]

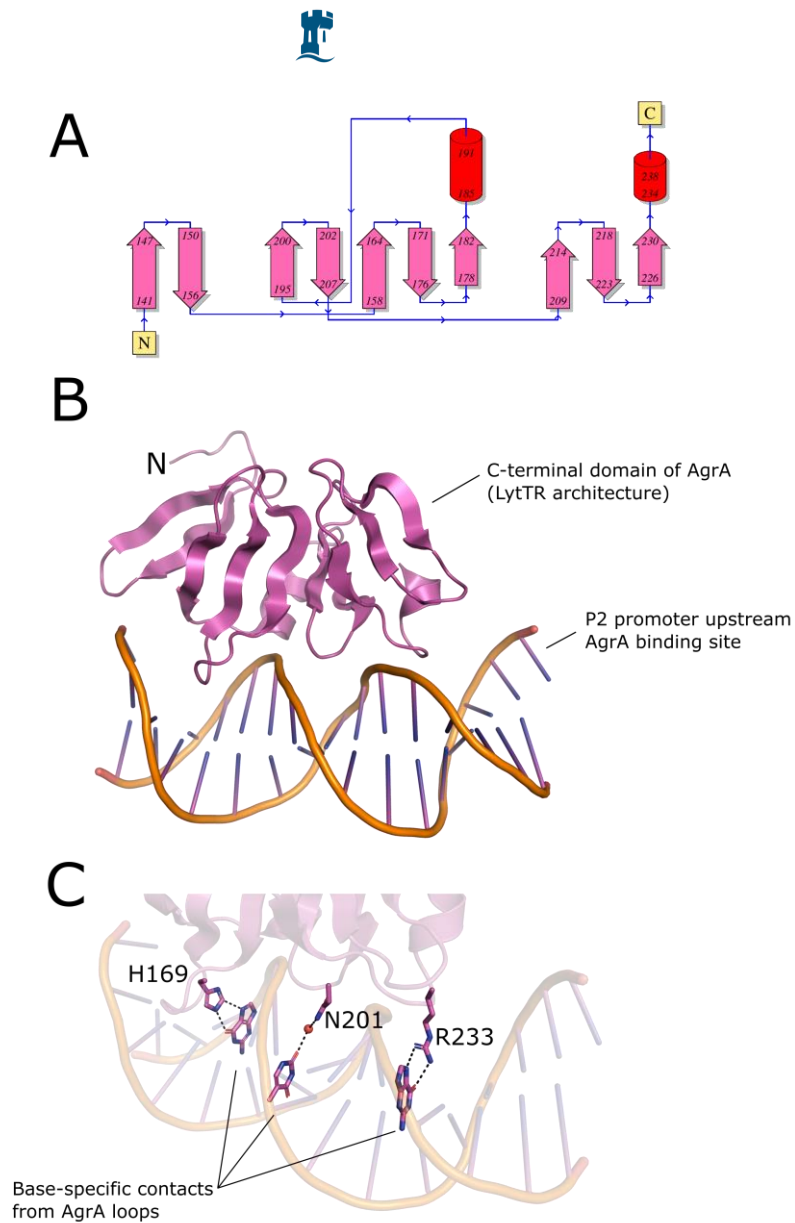


Figure 1.3. Three representations of the C-terminal domain of AgrA in complex with DNA corresponding to the upstream AgrA binding site on the P2 promoter region (PDB: 3bs1). (A). A topology diagram of the LytTR fold. (B). Viewed face-on with respect to the DNA helical axis highlighting the bend of $\sim 30^\circ$ imparted into the DNA by AgrA and the interdigitation of AgrA loop regions with DNA grooves. (C). Rotated slightly around the horizontal axis relative to (B) to highlight residues that make polar contacts (shown as black dashed lines) with the DNA molecule. His169 and Arg233 make base-specific contacts to guanine nucleobases, whilst Asp201 contacts the DNA molecule through a water-mediated H-bond to a thymine nucleobase. DNA deoxyribose rings, phosphate groups and all water molecules except those involved in the H-bond, are omitted for clarity.



Regions of DNA to which AgrA binds are, as already mentioned, imperfect linear repeats and mutating the P3 promoter such that it matches P2 (AC-to-CT) in sequence but not spacer length has been shown to enhance its affinity for AgrA beyond that of the P2 promoter to picomolar levels.^[81] Shortening the spacer region between P3 binding sites to match that of P2 has been shown to result in AgrA-independent P3 transcription.^[82] Given that the promoter regions are linear repeats, each of which can only bind to a single AgrA protein, AgrA must bind as a dimer to each of its promoters. Given what is known about the spacing between these binding sites and the structure of the AgrA DNA-binding domain, it has been inferred that the AgrA dimerisation interface must occur between the phosphate-receiving N-terminal domains^[80] as the DNA-binding domains would be too far apart ($\sim 10 \text{ \AA}$) to form protein...protein contacts in the physiological complex. As yet there is no structural data for the remainder of the AgrA protein, however, or any such data for AgrA in complex with a full promoter element.

The exact quantification of the pairs of linear repeats in each of the P2 and P3 promoter elements, coupled with the apparent ease of binding AgrA to these short oligoduplexes *in vitro*, has revealed some intricacies about *agr* regulation in relation to the subtle differences in binding modes of AgrA to each promoter. As expected from the larger spacer between AgrA binding sites on the P3 promoter compared to the P2 promoter, it is a less efficient binder of AgrA. This can be observed in relevant K_d values on the previous pages. In a later study^[85] Surface Plasmon Resonance (SPR) competition assays were performed on both phosphorylated and unphosphorylated AgrA pre-incubated with P1, (sequenced in the same study) P2 and P3. These samples were then flown over an SPR chip containing immobilised biotinylated P1, P2 and P3 promoters. For unphosphorylated AgrA the assay confirmed the K_d values determined previously and the binding affinity for the promoters was determined to be $P2 > P3 > P1$. For phosphorylated AgrA, however, the order of affinity changed to $P3 > P2 > P1$, suggesting that binding of AgrA to the P2 promoter is less drastically affected by AgrA phosphorylation/dimerisation than to the P3 promoter.

These results were corroborated by putting Green Fluorescent Protein (GFP) expression under the control of each of the promoters separately on a pRMC2M plasmid, on which AgrA was placed under



the tetO promoter and expression induced by the addition of anhydrotetracycline. The intracellular concentrations of both AgrA and GFP were measured via real-time PCR. A non-phosphorylatable AgrA mutant, D59A, was also used as a control and assumed to be constitutively inactive. This allowed for GFP expression to be plotted at known concentrations of AgrA and AgrA_D59A. The expression of GFP was found to be much higher when under the control of the P2 promoter compared to P1 and P3. Furthermore, the expression of GFP from the P2 promoter was the same for both AgrA and AgrA_D59A. Although the magnitude of GFP expression from the P3 promoter was found to be lower than that from P2, the change in expression levels of GFP from the P3 promoter was found to be substantially different for phosphorylated AgrA and AgrA_D59A. The same GFP expression pattern was also observed when GFP expression was measured by flow-cytometry. Put together these findings suggested that AgrA is capable of basal regulation of the *agr* system via P2 independent of phosphorylation, with this latter step seemingly being crucial for P3 expression and the switch to a virulent phenotype.

Given that the two AgrA binding sites on each of the promoter sequences are not identical^[81] it is perhaps unsurprising that AgrA appears to interact with each binding site differently, i.e. the dimer is not symmetric with respect to how it is positioned on, and interacts with, the promoter DNA. The extent to which this occurs is surprising however, with AgrA forming substantially more protein-DNA contacts with the linear repeat that is furthest away from the transcription initiation site, whilst that which is closer forms comparatively few contacts.^[85] This is somewhat counterintuitive as it implies that the AgrA protomer in the dimer that is closer to the RNAP molecule is less well bound to the DNA.

Here it is worth commenting on the unusual dimerisation mode of LytTR-type response regulators. This explanation is somewhat repeated in section 3.9 but is important to both sections. The DNA-binding (C-terminal) domains on AgrA must be arranged line astern to facilitate binding to a pair of linear repeats which have no inversion symmetry with respect to their nucleotide sequences. It has already been noted that these domains are placed too far apart on the promoter elements to form direct protein···protein contacts, so the dimerisation interface must presumably occur between the N-



terminal phosphate-receiving domains. These domains are, with the notable exception of CheY,^[86] connected to the C-terminal domains by an unstructured linker.^[87] This linker and the toxicity and/or insolubility that often accompanies overexpression full-length response regulators for recombinant studies^[80,88,89] means analysis of them is often relegated to either of the two domains alone. A general schematic representation for a response regulator dimer/promoter complex is shown schematically in figure 1.4A.

Crystal structures and Small-Angle X-ray Scattering (SAXS) studies on response regulator N-terminal domain dimers, where they have been coaxed to form *in vitro*, has revealed that they possess perfect, or near-perfect two-fold rotation symmetry.^[90] Consequently, the full-length response regulator dimer possesses rotation and tandem organisation of their N- and C-terminal domains respectively. There is an extreme paucity of full-length response regulator dimer structures in complex with promoter DNA, but those which have been solved clearly display this convoluted organisation. A good example is *Escherichia coli* KdpE^[91] (PDB: 4kny) (fig. 1.4B and C) which regulates expression of ATPases.

To return to the non-identical manner in which AgrA was found to bind to each linear repeat for any of the P2 or P3 promoters,^[85] there is a key feature of experimental design that revealed them that is worth noting. Only the DNA-binding domain of AgrA was crystallised separately with short oligonucleotides corresponding to each of the two AgrA binding sites for the P2 and P3 promoter regions. Thus they cannot include any steric perturbations imparted by the N-terminal domain dimer or imparted by the two binding sites being covalently connected through a piece of spacer β -form DNA. What is being witnessed then is the “intrinsic” binding fingerprint of the AgrA DNA-binding domain for each of the DNA fragments. The results of the study are undoubtedly meaningful, but it may yet be found that by having these additional physiologically-required modifications to the experimental design, greater or lesser differentiation between the two binding sites on each promoter is observed.

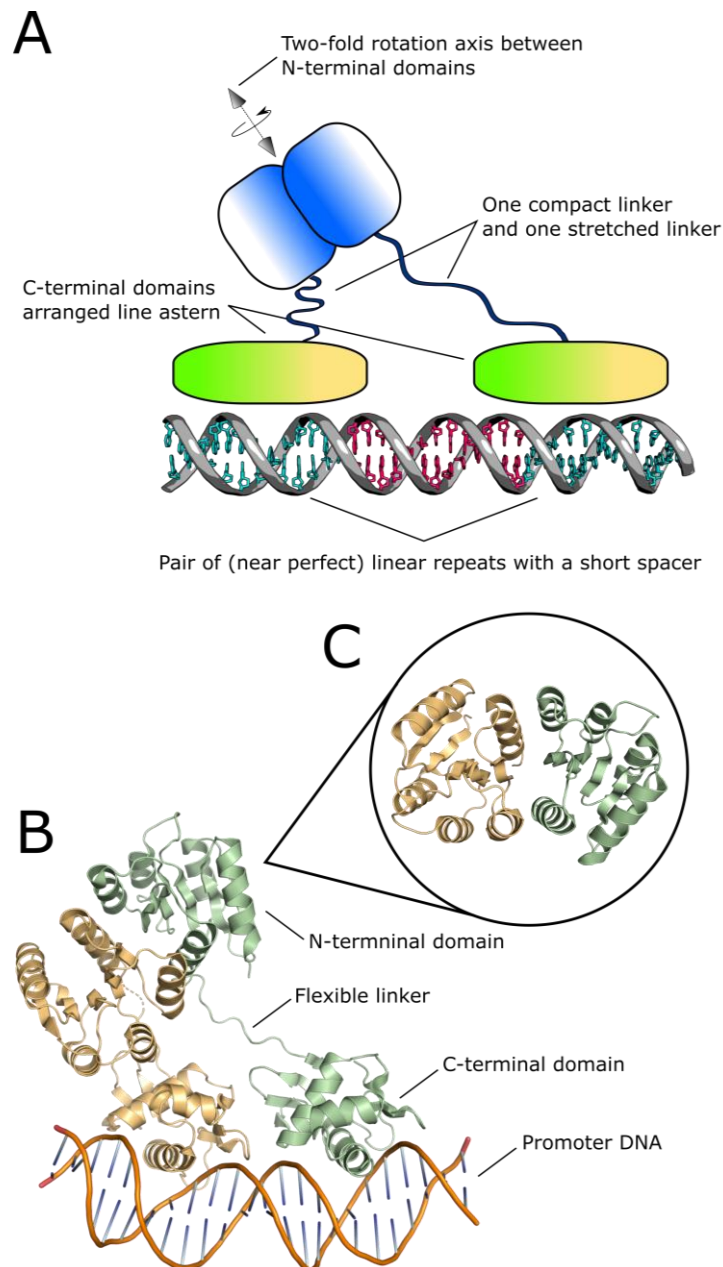


Figure 1.4. Representations of response regulator dimer/promoter complexes. (A). A schematic representation. The DNA-binding domains are arranged line astern (represented by the gradient fill) and the phosphorylated N-terminal domains are arranged with non-crystallographic two-fold symmetry. This organisation of domains must occur to drive gene transcription. (B). The *E. coli* KdpE dimer bound to an oligoduplex corresponding to its cognate promoter exhibits the arrangement shown in (A). The green protomer in exhibits an unstructured linker between the N- and C-terminal domains. The corresponding linker in the beige protomer has not been modelled. (C). The two N-terminal domains (inset) reside on a non-crystallographic two-fold rotation axis which lies normal to the page.



The asymmetric nature of AgrA binding to its promoters^[81,85] has proven to be crucial for chemical inhibition of transcription activation, which has mostly focussed inhibiting the direct interaction between AgrA and promoter DNA. This probably reflects the comparative ease of designing valid experiments to develop such inhibitors with only structural information for the DNA-binding domain available to researchers. Inhibitors which prevent phosphorylation and/or block dimerisation must at present rely on *in silico* modelling or solving the structure of the N-terminal domain to confirm where on AgrA they may bind. The former is generally considered weak evidence if it is the only data presented and the latter has not yet been forthcoming. Similarly, there are no reports of inhibitors which function through blocking interactions between AgrA and AgrC, or AgrA and RNAP. Both of these modes of AgrA inhibition are theoretically possible, although the former may prove to be indistinct from inhibiting phosphorylation and therefore dimerisation.

The first major study concerning AgrA inhibition involved *in silico* compound fragment screening for a library of 500 fragments.^[92] Five compounds were subsequently synthesised and titrated into a sample of the DNA-binding domain of AgrA in an NMR spectrometer. The changes in amide bond chemical shifts upon titration correspond to a highly conserved hydrophobic cleft at the extreme C-terminus of AgrA. This region is shown in figure 1.5B. Three of the five compounds exhibited disruption of DNA binding activity when analysed via Electrophoretic Mobility Shift Assays (EMSAs) and millimolar amounts of the compounds were required for this disruption, which is typical of the fragment screening approach of identifying inhibitory compounds.

Considering the two AgrA protomers bound to any of the P2 or P3 promoters, the protomer that is furthest away from the transcription initiation site engages the DNA through the two loops which contain Tyr189 and Tyr229 respectively.^[80,85] In the protomer that is closest to the transcription initiation site the contacts from residues in the Tyr229-containing loop are absent.^[85] Resonances corresponding to Tyr229 could not be assigned in the above-mentioned inhibitor study, but the following residues that are spatially-proximal to it: Ser231, Val232, Arg233, Asn234, Lys236, Lys237 and Ile238 (fig. 1.5B) all exhibited changes in their chemical shifts upon fragment titration.^[92] Since most of the subsequent studies concerning AgrA inhibition either do not attempt to demonstrate where



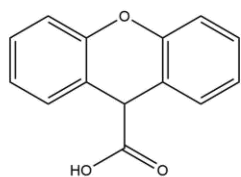
their respective low-potency inhibitors bind^[93] or shown them as modelled into this hydrophobic pocket^[94] (the only inhibitor binding site to-date with physical evidence to support it) it must be concluded at this stage that inhibition of gene transcription by AgrA proceeds through the protomer in the dimer that is furthest way from the transcription initiation site. That is the protomer which forms the most contacts with the DNA. To be certain of this, a series of the prospective inhibitors would need to be titrated into a sample of AgrA in the presence of DNA fragments containing to both binding sites of each promoter, or each binding site individually. This may also betray any differential inhibition of between the P2 and P3 promoters.

To-date, the most studied AgrA inhibitor is savarin^[95] (a portmanteau of *Staphylococcus aureus* virulence inhibitor). This compound lacks experimental evidence as to where on the AgrA surface it binds, however, docking studies predict a $\pi \cdots \pi$ stack with Tyr229 and polar contacts with Arg218^[95] (fig. 1.5B). These residues are extremely close to the experimentally-evidenced C-terminal pocket mentioned previously and clinical isolates containing the AgrA_R218H point mutant have been shown to have defective *agr* systems and reduced virulence,^[96] further highlighting the functional importance of this region of AgrA.

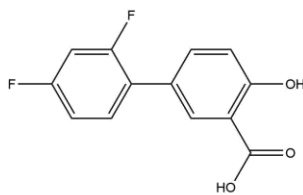
There is a study which takes an alternative route to targeting the above-mentioned C-terminal pocket. It instead discusses repurposing of the anti-inflammatory drug diflunisal for *agr* inhibition via binding to the N-terminal domain of AgrA.^[97] A comparatively poor EMSA does show evidence of DNA binding inhibition for full-length AgrA at 200 μM diflunisal. The authors also dock diflunisal to a homology fold of the AgrA N-terminal domain focusing on a 10 \AA^3 region centred on the phosphate binding site but seem to lack to confidence in their results, stating that diflunisal could bind anywhere on DNA before concluding “localization of the diflunisal-binding site on the surface of AgrA will have to await a cocrystal [sic] structure of AgrA and diflunisal.” Diflunisal is shown along with other known AgrA inhibitors in figure 1.5A.



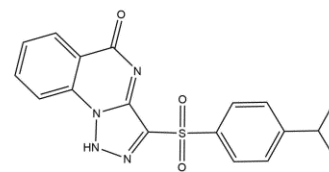
A



9H-Xanthene-9-carboxylic acid

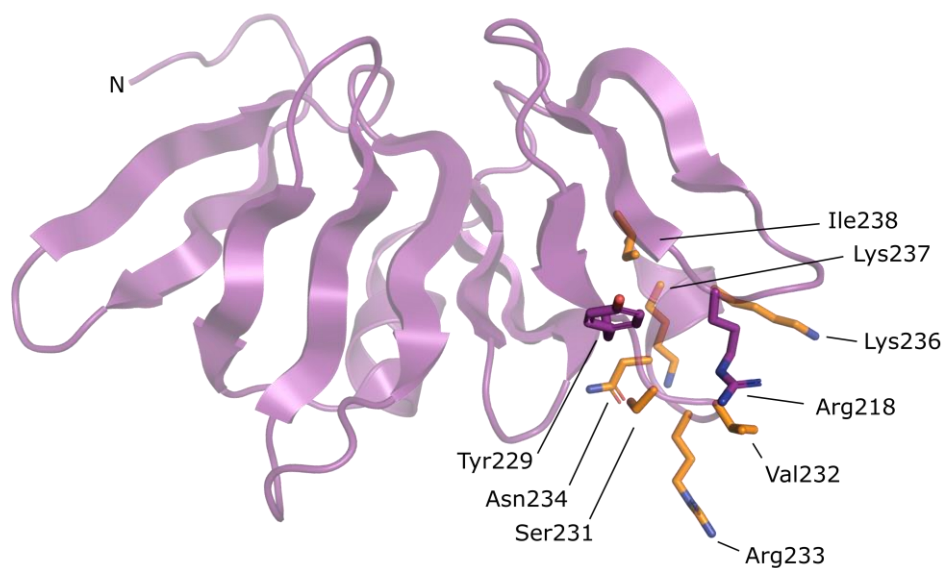


Diflunisal



Savarin

B



Key



-  Residues identified by *Leonard et. al.* to shift upon titration of 9H-Xanthene-9-carboxylic acid
-  Residues with which savarin is predicted to interact (from docking studies)

Figure 1.5. (A). Skeletal structures of salient AgrA inhibitors. 9H-Xanthene 9-carboxylic acid is the only inhibitor with accompanying experimental evidence as to where on the AgrA molecule it binds. (B). The C-terminal domain of AgrA. Residues in the hydrophobic pocket identified via NMR studies^[92] are shown in orange and those which are predicted to make contacts with savarin^[95] are shown in purple.



1.1.7: Genes Regulated by AgrA That Are Not Encoded by RNAIII

AgrA regulates transcription of genes that are not transcribed into RNAIII.^[98] Studies on this topic had previously been performed, however, they used the laboratory *S. aureus* strain RN6390 which contains a number of critical mutations to make it safer to work with.^[99] In the more recent study^[98] the clinical MRSA strain MW2 was used. Advances in the lower threshold used for genome annotation had also been made in the intervening time and so the study is considerably more relevant and illuminating.^[100]

Wilde type, Δagr and $\Delta RNAIII$ mutants were screened to triangulate those which were RNAIII-independent and the most surprising result was that AgrA downregulates many more genes than it upregulates with over 80 genes downregulated but only 6 upregulated.^[98] Of the upregulated genes, two encode the pro-inflammatory and leukocidal phenol-soluble modulins (PSMs) PSM β 1 and PSM β 2, one is the chaperone GroEL, one is a hypothetical membrane spanning protein and two are hypothetical proteins of unknown function MW0370 and MW0372. The significant upregulation of the two *psm* genes, corroborated by real time PCR experiments, lead to the hypothesis that AgrA binds directly to their promoters to initiate transcription. This hypothesis was proven by using the DNA sequences of the AgrA binding sites on the P2 and P3 promoters to search for similar binding sites on the *S. aureus* genome proximal to *psm* genes. Three putative binding sites were returned and two were eventually confirmed to bind to AgrA via EMSAs using *psm β 1* and *psm β 2* probes.^[98]

Of those genes downregulated, many are involved in amino acid and carbohydrate metabolism^[98] supporting the hypothesis that there is a metabolic burden accompanying the virulent phenotype of *S. aureus* cells.^[101] The large number of downregulated genes and the lack of any plausible consensus sequence between them was deemed sufficient to rule out direct regulation by AgrA binding to all of their promoters. It has been tentatively postulated that there may yet be an additional regulatory circuit with which *agr* may be intertwined,^[102] or that downregulation may occur through interactions between AgrA and RNAP and/or interactions between AgrA and the transcription factors which regulate transcription of the downregulated genes.^[98] Absent amongst the downregulated genes are



any surface-binding proteins. Prior to this study it was thought that the *agr* molecular circuitry downregulated such proteins as part of the progression to an infectious state^[58] but this was proven not to be case at least for the MW2 strain of *S. aureus*.

1.1.8: RNAIII

Expression of RNAIII is driven by the P3 promoter. It is a relatively large ~ 0.5 kb molecule and is the primary effector of the *agr* system.^[103] The modes in which it alters gene expression are varied and include: direct translation by RNAIII itself,^[103] binding to exogenous messenger RNA (mRNA) and shuttling it to RNase III thus shortening its half-life^[104] and inhibition of a global repressor of toxins^[105] (Rot) which belongs to the SarA family of repressors.^[106] Rot, as the name suggests, is responsible for suppressing expression of toxins, proteases and lipases and upregulating clumping factors and coagulases.^[107] There is considerable, but not total overlap between genes regulated by *agr* and genes regulated by Rot,^[107,108] indicating that *agr* function does not occur solely through the facile mode of Rot inhibition. Selectively augmenting Δ *agr* mutants can restore *agr* function in some instances (this is how Rot was discovered).^[106] To pick a salient example, in a Rabbit model of *S. aureus* endocarditis disease progression, a Δ *agr* strain was compared to the wild-type (WT) strain. The Δ *agr* mutant case exhibited fewer bacteria, lessened establishment of endocarditis and lessened progression of the cells to the liver, all in a dose-dependent manner.^[109] Deleting Rot, however, in the Δ *agr* mutant completely restored virulence to WT levels.^[110]

In *S. aureus* RNAIII contains the *hld* gene which encodes δ -toxin. Interestingly, not all Staphylococci embed this gene on RNAIII whilst others embed more than one copy.^[111,112] The δ -toxin molecule is a short amphipathic helix in aqueous solution.^[113] It oligomerises to form pores in red blood cells, a feature which also allows it to disrupt cell membranes at high enough concentrations.^[114] The fact that RNAIII does not influence PSM expression but does appear to influence PSM secretion^[98] suggests that δ -toxin may, through its pore-forming mechanism, aid in the yet unknown transport mechanism of PSMs outside of the cell.



Amongst the proteins downregulated by RNAIII is protein A,^[98] a surface-anchored immunoglobulin-binding protein that is encoded by the *spa* gene.^[30] The discovery of RNAIII-dependant *spa* inhibition involves a multitude of complicated experiments but simply put, it was already known that RNAIII was a translational regulator of *hla*^[115] (which encodes another pore-forming molecule α -toxin)^[116] and the *hla* sequence was observed to have some complementarity to *spa* mRNA.^[58,115] Eventually it was demonstrated that RNAIII secondary structure elements forms a stable complex with *spa* mRNA, containing an overlap of over 30 bp in which the ribosome-binding site and start codon of *spa* mRNA is obstructed.^[115]

1.1.9: Additional Components of the *agr* System

There are additional important features of the *agr* system worth introducing. There are dozen-or-so other transcriptional regulators that can activate the *agr* system in both a positive and negative manner.^[117,118,119,120] Of these *agr*-modulating transcription factors, the extent to which they regulate *agr* has shown to be remarkably strain-dependant. One important example is SarA for which putative binding sites were initially identified between the P2 and P3 regions.^[121] These regions were later confirmed to be protected by SarA in DNase1 protection assays.^[122,123,124] The crystal structure of a SarA/DNA complex indicates that the SarA mode of action is through over-winding of the DNA to encourage transcription factor binding.^[125] This is relevant to the fact that the P3 promoter has a longer spacer between AgrA binding sites than P2,^[81] which when slightly shortened leads to AgrA-independent RNAIII transcription.^[82,126] There is no confirmed link between these two observations, however.

Another significant feature concerns the nutrient-sensing protein CodY, which binds branched-chain amino acids to relay this nutrient-availability signal.^[127] Growing WT *S. aureus* cells in the absence of isoleucine leads to increased levels of RNAIII expression,^[128] which lends credence to the hypothesis that CodY downregulates *agr* during the exponential cell growth phase. Furthermore putative CodY binding sites have been identified in the *agrC* genetic region of RNAII^[129] and the *hld* region of



RNAIII.^[130,131] Pull-down assays of CodY indeed showed enrichment of genetic fragments which encode AgrC^[129] but the exact mechanism of *agr* interference by CodY remains unknown.

1.1.10 The Potential of *agr* Components in Combating Antimicrobial Resistance

It will be observed that AgrA has received significantly more discussion than the other *agr* QS factors. This represents the comparative abundance of studies upon it. Whilst it has proven difficult to express and purify in full-length form it is at least water-soluble unlike AgrB and AgrC, and not as toxic to the expression host as AgrD. Consequently it is the most amenable to recombinant studies. The fact that it was only as recently as 2004 proven to be a response regulator highlights^[81] that **it may be** difficult protein to study, and if it is the easiest of the *agr* factors to study then the others by extension must be challenging. It is for these reasons that most of the available *agr* inhibitors target AgrA, reportedly by interfering with DNA-binding activity.^[92,95] It must be admitted, however, that response regulators are not common therapeutic targets^[132] owing to the lack of deep substrate-binding pockets observed in proteases and transporters, and the difficulty in leveraging selectivity for one response regulator over another in the absence of these.

Thus ambuic acid appears to be a promising route to *agr* inhibition since it appears to be selective towards an integral membrane protease found in only a small number of problematic Gram-positive bacteria.^[57] Being an integral membrane protease, obtaining an experimental structure of AgrB to facilitate drug discovery projects is likely to be challenging. That being said structure-based drug design, although beneficial, is not an absolute requirement for drug development. Indeed this must be the case since ~ 70 % of FDA-approved drugs target membrane channels, transporters or receptors,^[133] yet the fraction of known structures which membrane proteins comprise is very small at ~ 2.5 % (according to <https://blanco.biomol.uci.edu/mpstruc/>^[134] which contains 4422 structures at the time of writing compared to 172,175 in the Protein Data Bank). This is not to imply an experimental AgrB structure would be a hindrance however, and being an entirely new class of transmembrane protease it would be of extremely high scientific interest and accelerate modifications to the ambuic acid molecule to increase its potency.



The same view could be taken for AgrC for which inhibition has started at the obvious point of derivatising AIP.^[77] For the AIP-derivatives and ambuic acid, efforts should be focused scaling up their syntheses to ensure they are economical whilst shifting focus to *in vivo* modes of validation. In the short term structural information that is less highly resolved than a full structure determination such as MS may pinpoint where exactly these inhibitors bind and so serve as a starting point for derivatisation to increase their potency.

Much has been said of recombinant and structural investigations into *agr* but little of *in vivo* microbiological work. The discovery of the CAAX-type transmembrane protease MroQ on the *agr* locus as recently as 2019^[135] and RNAIII-independent gene transcription by AgrA^[98] suggests there is a lot still to learn at the genetic level about QS in *S. aureus* and the *agr* system. It is hard to predict what benefits may come of measuring gene transcription in different *S. aureus* strains and rational augmentations thereof. If more could be learned of any molecular communication between *agr* and other regulatory systems^[136] however, then there may yet be discovered new opportunities to intervene and attempt to attenuate *S. aureus* virulence.

So there are ample routes to intervene and mute QS in *S. aureus* and therefore ample routes to develop lead compounds which may support, over considerable time and investment, the development of FDA-approved drugs. Since the *agr* system is not strictly required in the lifecycle of *S. aureus* cells^[137] any therapeutics which intervene upon it ought to be non-lethal and significantly reduce the selection pressure associated with current antibiotics.^[138] Although eventual resistance to these new drugs is inevitable, their non-lethality should at least reduce the rate at which resistance occurs and therefore the rate at which new drugs need to be developed with obvious humanitarian and economic benefits. The idea is simple enough to state but the scale of the issue is worth pausing to consider. This introduction is a brief summary of the most relevant literature of but one major problematic pathogen.



1.2.1: X-ray Crystallography

The subject of X-ray crystallography is centuries old, originating with speculative treatises upon the symmetric nature of naturally occurring objects such as snowflakes^[139] and emerged in its modern form as a fusion of mineralogy^[140] and pioneering work on X-rays during the late 19th and early 20th centuries.^[141,142] It is a highly interdisciplinary and complex subject which draws from all of the sciences.^[143] Physicists are required to provide the theoretical basis for new methods of data collection and treatment, engineers are required to turn these ideas into real experimental stations and hardware, computer scientists are required to write programs to allow for accurate and intuitive processing and conveyance of the data whilst chemists and biologists are required to give the subject context and currency. Consequently to treat the subject in full could result in a work that spanned volumes and is beyond the knowledge of the experimenter. What follows is an abbreviated textual overview of the very basics of the subject, followed by a deeper analysis of some of the fundamentals that are important to satisfactory structure solution via X-ray crystallography.

1.2.2: Introduction

With the exception of black holes all physical objects scatter electromagnetic light.^[144] That is to say photons striking an object from any particular direction will be sent veering off in different directions depending both on the nature of the incoming photons and the object doing the scattering.^[145,146] In this scattered form the photons are meaningless with respect to an observer trying to form an image of their surroundings.

If one looks at the wall or desk, incidental light is striking it from whatever light source is illuminating it and being scattered (via Mie scattering)^[147] by the minute blobs of various dyes and pigments which give it colour. These scattered photons are now travelling in all directions away from the object and some of these are traveling in the vague direction of one's eyes. In order for the object to be perceived the scattered photons must be refocussed onto a single point. It is towards this purpose that the eye has evolved.^[148]



Unfortunately for structural scientists electrons do not scatter visible light.^[149] Instead electrons scatter X-rays^[150] the wavelengths of which are several thousand times shorter^[151] and about the same length as a covalent bond.^[152] To try to refocus scattered X-rays using a physical lens would just result in further scattering of the X-rays by the bonds and atoms within the lens therefore. Thus individual molecules cannot be imaged using either visible light or lens technology.

Fortunately all waves, irrespective of their wavelength and amplitude, obey a well-understood mathematical transformation when scattered, called the Fourier transform.^[153] If the photons scattered from this page for example could somehow be accurately collected and quantised without refocussing them the resulting pattern would be the Fourier transform of the page. Remarkably, if such quanta are substituted back in to the Fourier transform equation raised to the power of minus one, with the necessary rearrangement of terms, an image of the original object as perceived by the eye would be formed.^[153]

If the X-rays scattered from a crystal can be accurately collected and quantised, then it becomes possible to perform the refocussing job of a lens mathematically instead of physically. There is, however, one major barrier to this goal: it is physically impossible to experimentally collect all the information needed to do this calculation.^[154] Before considering why this lack of information is a problem there are two things worth thinking about: A) what are the key properties of X-rays crystallographers are interested quantifying? And B) why must the matter under study be in the crystalline state to begin with?

Starting with A) the properties of X-rays of interest are their amplitude, A , and phase, Φ . The amplitude of a wave may be defined as its vertical displacement from the origin of a sinusoidal graph if modelled propagating as such. The phase of a wave is its horizontal displacement from the origin of a sinusoidal graph (fig. 1.6). Here an unavoidable complexity intrinsic to photophysics is encountered, but simply put, the amplitude of a single photon, being relativistic, has no definite value^[155,156] but strings of them propagating together as an electromagnetic wave do.^[157] The transition between these two behaviours is mathematically complex and probabilistic. In the simplistic case, however, where a



single photon is merely being represented graphically, it is safe to assume that amplitude, A , takes arbitrary units of displacement from the origin whilst phase, Φ , takes units of degrees ($^{\circ}$) or radians (rad).

Two adjacent waves of equal wavelength, with a phase offset of 180° relative to one another, would destructively interfere and cancel each other out. Two adjacent waves of equal wavelength with a phase offset of 360° would constructively interfere to give a single wave with the combined amplitude of the two constituent waves. Any other phase offset would give partially destructive interference with a waveform (the combined result of amplitude and phase) unique to the particular combination of waves which combined to create it.^[158]

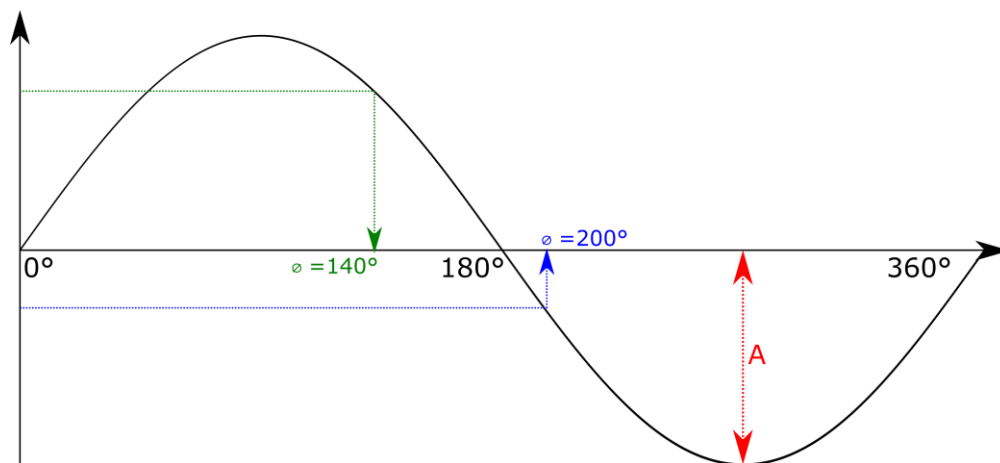


Figure 1.6. A generic representation of a sinusoidal wave. The phase of the wave at two arbitrary points as it propagates through space is demarked by the green and blue dashed arrows. The amplitude of the wave is demarked by the double-headed dashed red arrow. A wave of phase Φ will always be perfectly in phase with a wave of phase $\Phi + 360$, provided they both have the same wavelength, λ (not demarked on the diagram above but is defined as the spatial period in which the wave completes a single wave cycle. This would be the space between the points 0° and 360° on the diagram above).



As for B) why must matter be crystalline in order for its atomic structure to be studied is best understood by considering scattering from non-crystalline (amorphous) material. Since X-rays are scattered equally in three dimensions upon striking electrons,^[150] and since molecules in such samples are orientated in every conceivable direction their scattering pattern is diffuse and approximately equal in every direction.

To explain with an analogy, imagine shining a laser pen (the X-ray photons) at a mirror (the atoms in a sample) and seeing the laser light on wall or roof somewhere. Now imagine the same scenario but with the mirror in a new orientation (representing another orientation of molecules within the sample or another scattering event from the same molecule). A spot would be observed elsewhere on the wall or roof. Now repeat the thought experiment for every possible orientation of the mirror (representing all molecular positions within the amorphous sample). One would eventually be surrounded by a room of reflected laser light. In reality every atom scatters the incidental photons approximately spherically and interference effects are effectively averaged out in the scattering pattern because the distribution of molecules within the sample is random.

Returning to a real scenario, the number and therefore the intensity of scattered X-ray photons actually falls with higher scattering angles^[150] (scattering of photons by atomic matter is expressed probabilistically over given cross-sections)^[145] and photon detectors are two-dimensional. Thus the scattering observed from amorphous sample is a diffuse circle decreasing in intensity towards its perimeter. In other words the scattering is indiscreet. Such scattering must be described using continuous functions which preclude substitution back into the Fourier transform equation.

In a crystal, however, molecules are arranged symmetrically into small clusters which repeat almost perfectly.^[159] The primary consequence of this is that there are special orientations of the sample relative to the incident X-ray beam at which some of the scattered X-rays constructively interfere. These geometric laws (called Bragg conditions)^[142] are stringent since they must be at once satisfied in three orthogonal directions for constructive interference to occur. Consequently, destructive interference between adjacent scattered waves is the overriding result, with very few diffracted waves



satisfying the Bragg condition. At the X-ray detector this manifests as a mostly blank image with a few sharp maxima observed for a given crystal orientation or detector position.

These maxima (called Bragg peaks) encode two of the three pieces of information needed to perform the reverse Fourier transform on them to arrive back at the image of the crystalline sample. They have a discrete position, allowing them to be indexed with respect to some origin and they each have a measurable intensity. Since the measured intensity of an electromagnetic wave is proportional to the square of its amplitude, the latter can be derived.

The third piece of information required for the mathematical refocussing are the relative phase offsets of each of the waves causing all of the maxima. It is this information that cannot be collected experimentally. The diffraction pattern looks the way it does because of the phase offsets of all of the scattered waves, but the phases cannot be derived from it. This is called the phase problem.^[154]

Crystallographers do not so much as solve this problem as keep trying phase estimates until the mathematical refocussing of the scattered waves resembles something that is chemically and structurally reasonable.^[160] An accurate set of phase estimates substituted into the reverse Fourier transform along with experimental amplitudes generates an accurate electron density map i.e. an image of the object that originally scattered the X-rays (remember X-rays are scattered by electrons).^[150] This electron density map, although corresponding to a repeating portion of the crystal, contains the combined contributions from all of the atoms in the crystal and consequently all possible damage and disorder that may have been introduced. It is the job of the crystallographer to interpret this map and build a chemically and physically sensible structure into it. This structure must then be validated which is done by comparing how well it predicts data that was experimentally collected at the X-ray detector. These steps are called model building and refinement respectively.^[161]

This description, being simple, overlooks a large number of practical aspects that go into solving an experimental structure, primarily concerning how the quality of the crystal affects the data, and how the quality of data affects the electron density maps and their interpretation. Some of these will be addressed on the following pages.



1.2.3: The Basics of a Crystal and Diffraction Patterns

Crystals are repeating arrangements of singular, or clusters, of atoms or molecules. The portion of a crystal that repeats through translation only, i.e. with no symmetry between units of itself, is called the unit cell.^[159] Unit cells stack like blocks and from an atomic perspective there are essentially infinite unit cells per crystal except in very extreme cases where the unit cell is very large.^[162]

Although the dimensions of each unit cell vary from crystal-to-crystal, the requirement of perfect geometric tessellation of crystal subunits forbid many of the possible spatial arrangements and therefore many possible unit cell shapes.^[163] In the loosest description every crystal belongs to one of six crystal families: cubic, tetragonal, orthorhombic, monoclinic, triclinic, and hexagonal.

The unit cells are easy enough to denote by algebra and are as follows: cubic, a cube; tetragonal, a square-based prism; orthorhombic, a rectangular-based prism; monoclinic, a square-based prism that is leaning to one side, such that it has become a parallelogram with one (hence mono) non-right angle; triclinic, a parallelogram-based prism that is leaning in two orthogonal directions such that it has three (hence tri) non-right angles and all sides of unequal lengths; hexagonal, a hexagonal-based prism.^[161]

The hexagonal family is unusual in that it may be divided into two subsystems.^[164] One containing genuine hexagonal units with six-fold rotation symmetry, and the other, called the trigonal system, in which units with three-fold symmetry agglomerate into a hexagon. Imagine stacking equilateral triangles (which contain a three-fold rotation axis at their centre) together into a hexagon. Confusingly, the trigonal system is made up of rhombohedral unit cells. The rhombohedral unit cell is a rhombohedron for which the six faces are identical rhombi.^[164]

For ease of memory the cuboidal unit cells can be grouped with their non-cuboidal counterparts, Cubic and rhombohedral; tetragonal and monoclinic, orthorhombic and triclinic; and hexagonal is on its own. These are the seven crystal systems. The unit cell by definition contains all of the symmetry observed between molecules within the crystal. These are called symmetry operations. Examples of symmetry operations are: rotation axes, screw axes (a rotation followed by a translation parallel to the rotation axis), mirror planes, glide planes (where the mirror image is translated parallel to the mirror



plane) and inversion centres.^[165] Consequently, it is possible to divide the unit cell into a yet smaller region, the contents of which exhibit no crystallographic symmetry between them. This is called the asymmetric unit. Application of all of the symmetry operations to the asymmetric unit gives the unit cell. Translation of the unit cell gives the entire crystal. A diffraction pattern encodes information pertaining to all of the electron density within the unit cell. Since the asymmetric unit is the only atomistically unique portion of the crystal however, only electron density corresponding to a single asymmetric unit is modelled into during a routine structure solution.^[161]

Note that it is the minor imperfections in the arrangement of the unit cells that actually allows diffracted waves to be observed at the detector. If their arrangement were truly perfect then the diffracted waves would manifest as infinitesimally small mathematical points on the detector. Too much imperfection between unit cells tarnishes the diffraction pattern however, and worsens the quality of any calculated electron density maps.^[161]

There is a finite number of non-redundant ways in which symmetry operations can be grouped together and applied to an object to build a three-dimensional crystal.^[163] These are called space groups. Group theory mathematics allows determination of these space groups and it has been determined that there are 230 of them.^[165] This reduces to 219 when chiral space groups are not considered. Proteins are chiral molecules and so they cannot arrange themselves into lattices containing mirror or glide symmetry (and by extension inversion symmetry). The number of space groups possible for protein crystals consequently reduces further to 65.^[166] These contain rotation and screw symmetry operations only.

Owing to the mathematical laws of diffraction the size and geometry of the unit cell is inversely related to the spacing of the maxima on the diffraction pattern.^[142] It is therefore possible to infer some information about the symmetry of a crystal directly from the diffraction pattern such as crystal family and unit cell dimensions. Since a diffraction image is a two dimensional snapshot of a three dimensional pattern, and since all diffraction patterns possess inversion symmetry even if the molecules in the unit cell do not,^[167] some information (translational relationships) regarding the



arrangement of molecules in the crystal is not conveyed. Consequently it is not possible to infer the space group directly from the diffraction pattern. For example, screw axes, which contain a translation parallel to a rotation axis, appear the same as formal rotation axes until further processing of the diffraction images is undertaken. A formal categorisation of the space groups as they appear on a series of two dimensional diffraction images is needed therefore to unambiguously categorise diffraction patterns. These are called Laue classes and are directly obtainable from diffraction patterns.^[142,168]

Some other features of diffraction patterns are also worth mentioning. The relative intensities of each of the maxima relates to the positions of the atoms in the unit cell. The relative positions of the maxima encode information only pertaining to the arrangement of the asymmetric units and not the positions of the atoms within it.^[169] This is because the diffraction pattern is an interference pattern from which the intensity of a wave causing a given maximum is generated by interference from scattering from all of the atoms within the unit cell. Another way of phrasing this is that the diffraction pattern is the Fourier transform of the unit cell contents, with the Fourier transform of the lattice superimposed over the top of it.

1.2.4: The Bragg Equation

The Bragg equation^[142] is fundamental to X-ray crystallography as it relates features of a crystal (in “real” space) to features of a diffraction pattern (in “reciprocal” space) and is given below:

$$n\lambda = 2d\sin(\theta)$$

Its derivation in relation to figure 1.7 is as follows: any two adjacent X-ray photon waves, represented as black arrows, scattered by regions of a crystal, represented as fuchsia circles and separated by distance d , have a path difference of $x + x'$. In order for the waves to constructively interfere (a prerequisite to them being observed at the detector) the path difference between them must be equal to an integer number of wavelengths, $n\lambda$. Therefore, for an observed reflection:

$$x + x' = n\lambda$$

The distances x and x' may be expressed trigonometrically as:



$$x = d\sin(\theta)$$

$$x' = d\sin(\theta')$$

Therefore, for an observed reflection, substitution for x gives:

$$d\sin(\theta) + d\sin(\theta') = n\lambda$$

It is possible to construct a mathematical plane such that $\theta = \theta'$. Such a plane is called a Miller plane and the equation becomes:

$$d\sin(\theta) + d\sin(\theta) = n\lambda$$

Which is equal to:

$$n\lambda = 2d\sin(\theta)$$

QED.

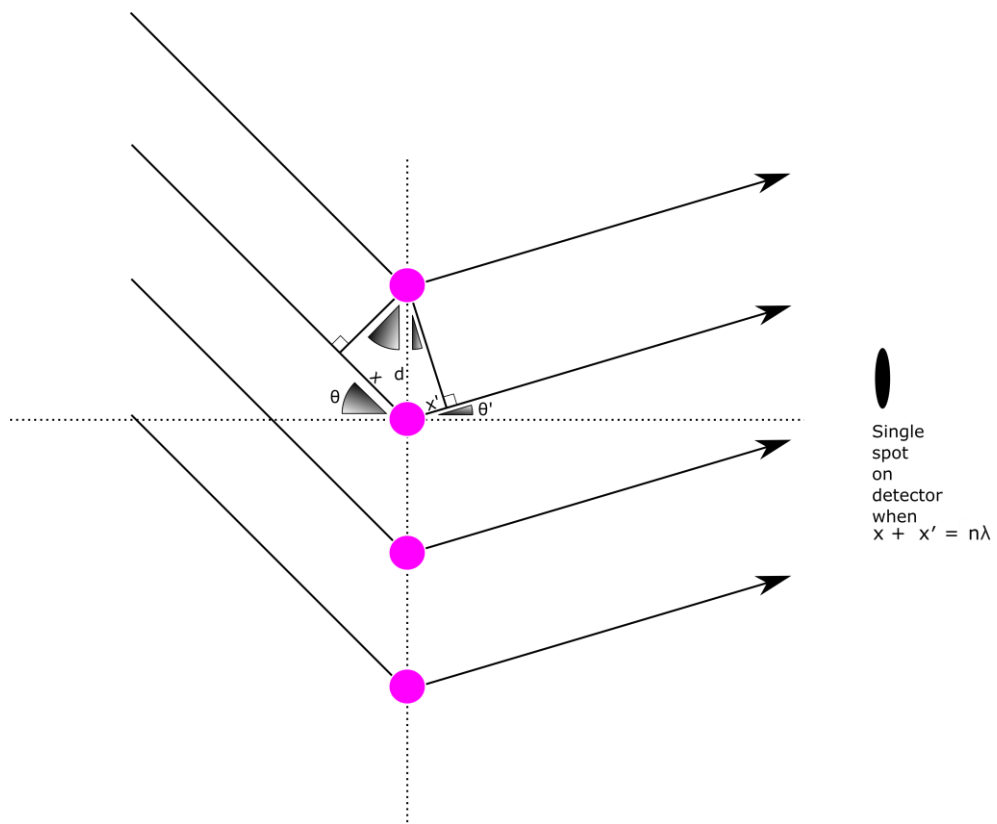


Figure 1.7. Derivation of the Bragg equation.



In the Bragg equation, n is any integer value, λ is the photon wavelength in Å, d is distance between adjacent Miller planes in Å, and θ is the angle of incidence/reflection of the X-ray beam from the Miller plane in °. The Miller plane is constructed such that that the angle of incidence and reflection of the X-ray beam from it are identical and such that this plane is normal to the plane of the scattering. It may be thought of as an imaginary mirror that sits where a real mirror would sit to cause a given Bragg peak from the X-ray source. One Miller plane may be constructed per Bragg peak and since they are constructed such that each scattered X-ray beam may be mathematically treated as a reflection, the maxima on diffraction patterns are sometimes referred to as reflections.

Each Miller plane, if extrapolated to extend throughout the crystal, will intersect the unit cell axes at fractional coordinates. These coordinates can be reciprocated and normalised to one in order convert them to integer indices which are called Miller indices. By convention the indices are denoted h , k , and l , which respectively intersect unit cell axes x , y and z .^[140] For example, the Miller plane that intersects the unit cell axes at $x = 1/2$, $y = 1/2$ and $z = 1/2$ is the 222 Miller plane.

In order to fully appreciate the Miller plane, one must imagine it continuing throughout the entire crystal. For example, imagine the 333 Miller plane running throughout a crystal. It will intersect the unit cell axes at $x = y = z = 1/3$. Since it continues infinitely through the crystal however, it will intersect the axes of an adjacent unit cell at $x = y = z = 2/3$, then the next adjacent one at $x = y = z = 3/3$ and then back to $1/3$. Plus every individual unit cell will have its own “originating” 333 plane, so every unit cell contains 3 333 planes intersecting the unit cell axes at $1/3$, $2/3$ and $3/3$. And so it is for all other Miller planes.

Since Miller indices are reciprocal coordinates, Miller planes with larger indices are more closely spaced than those with smaller indices. Taken to its extreme planes with large indices offer higher resolution information about the crystal since it is scattering from these that planes corresponds to scattering from electrons that are physically close together in the crystal. In the Bragg equation the spacing between Miller planes is given by d . Rearrangement of the Bragg equation for d gives:



$$d = \frac{n\lambda}{2\sin(\theta)}$$

And since $\sin(\theta)$ takes its maximum value of 1, at $\theta = 90^\circ$, the equation further simplifies to:

$$d = \frac{n\lambda}{2}$$

From this two major features of diffraction patterns may be inferred. Since $\sin(\theta)$ tends to 1 as θ approaches 90° , it becomes apparent that the high-resolution data (smallest value for d) is the high angle data. That is to say maxima observed at 40° relative to the incident beam originate from X-ray photons scattered by more closely-separated regions of the crystal than maxima observed at 20° .

In practise $n = 1$ in all cases since waves that are scattered from Miller planes such that their mutual path difference is (say) 2λ , then the distance d between the planes is half of the value than those planes for which the mutual path difference is λ . In other words when $n = 1$, $d = 1$, when $n = 2$, $d = 1/2$, when $n = 3$ $d = 1/3$ (relative to one another for a given crystal) and so on.

The second feature that may be observed from the Bragg equation in this form is the resolution limit of a diffraction experiment. Since $\sin(\theta) = 1$ at $\theta = 90^\circ$, the resolution limit is $\lambda/2$ (diffraction out to 90°). Given that synchrotron data collections for protein crystals are performed at $\lambda \sim 0.98 \text{ \AA}$, the maximum obtainable resolution is $\sim 0.5 \text{ \AA}$ although protein crystals seldom diffract to this limit.

The Bragg equation may further be rearranged to give:

$$\sin(\theta) = \frac{n\lambda}{2} \times \frac{1}{d}$$

In this form the origin and nature of the term “reciprocal space” become apparent since reflections measured at the X-ray detector are done so with their angle of reflection, θ , related to the real space Miller plane separation distance, d , raised to the power of -1, $(1/d)$, hence the term “reciprocal space”.

Because of the stringent requirements of satisfying the Bragg condition in three dimensions simultaneously, either the crystal or the detector need to be moved in a diffraction experiment in order



to bring as many possible reflections into the Bragg condition, such that enough of them are measured to be able to perform the reverse Fourier transform to yield interpretable electron density maps.

Observation of maxima on a diffraction pattern may also be explained by invoking a mathematical construction called the Ewald sphere,^[170] which is often illustrated as a sphere existing in real space surround a crystal. It is a mathematical sphere of radius $1/\lambda$, and thus has units of “per length” and so is purely a reciprocal space construct. It is constructed such that any point on its surface lies on the origin of the diffraction pattern (the origin of reciprocal space, or the 000 Miller plane) and with its diameter parallel to the incident X-ray beam. This construction implies that maxima only intersect the surface of the sphere when they in the Bragg condition. Its construction can then offer purely mathematical arguments as to why a portion of reflections may never be observed in a given crystal orientation and determination of unit cell parameters.^[171]

1.2.5: The Fourier Transforms

The forwards Fourier transform, mathematically describes what occurs when incident X-ray photons are scattered by electrons in a crystal.^[153] It is usually given as:

$$F(hkl) = \int_{cell} \rho(xyz). \exp[2\pi i(hx + ky + lz)] dV$$

$F(hkl)$ is, because of the inclusion of the i term, a complex number. It is called the “structure factor” and because it is a complex number it has a real and imaginary component. These real and imaginary components are the structure factor amplitudes, $|F(hkl)|$ and structure factor phases, $\Phi(hkl)$. Only the structure factor amplitudes are collected during a diffraction experiment. There is one structure factor per observed reflection hkl (“reflection” will be used instead of “maxima” here it is the conventional nomenclature). $\rho(xyz)$ is the electron density at any point xyz within the unit cell and the final term in squared brackets relates Miller indices hkl to their real space coordinates xyz .

To translate the equation approximately into words, a reflection, that is encoded by the number $F(hkl)$, is obtained by taking the electron density at any point in the unit cell $\rho(xyz)$ and multiplying it by the



complex number $\exp[2\pi i(hx+ky+lz)]$, and then integrating this function with respect to volume between the limits of the unit cell.

Integration of a continuous function such as the Fourier transform as it is given above is computationally expensive. This is not a problem for data collection since the Fourier transform is being carried out physically in real time. It becomes a problem however, when trying to validate the model that has been built into the electron density maps since this relies upon performing a forwards Fourier transform upon it *in silico* to see how well the model-derived $|F(hkl)|$ values agree with the experimentally-observed $|F(hkl)|$.

It is much simpler to perform these calculations if the Fourier transform can be expressed as a summation rather than an integral. This can be achieved by treating each atom as a point scatter of electrons. Whilst this treatment is valid, it is not enough to provide to satisfactory atomic models unless more dynamism is factored into the mathematics. This primarily concerns thermal vibration of the point-scattering atoms. The scattered intensity of X-rays becomes weaker the more an atom is thermally displaced from its origin,^[172] and provided this can be described mathematically and modelled for each atom then it may be introduced into the forwards Fourier transform to the right of the equals sign as an additional multiplication term. The resulting equation is complicated and is presented below:

$$F(hkl) = \sum_j f_j(\theta) \cdot \exp(-8\pi^2 U_j \sin^2(\lambda/2)) \cdot \rho(xyz) \cdot \exp[2\pi i(hx + ky + lz)]$$

The exact nature of everything in the equation is not of general importance. It is merely worth noting that the integral has become a summation, $f_j(\theta)$ is the so-called atomic scattering factor^[172] and these have been calculated and tabulated for every atom at every scattering angle for relevant radiation wavelengths^[173] and U_j is a thermal displacement parameter which describes the motion of the atom as a root mean square from an equilibrium position. Note that the $8\pi^2 U_j$ term is closely related to the so-called B-factor,^[174] or temperature factor, itself given by $8\pi^2 U_j^2$ and which is often used to convey two dimensional thermal disorder for atoms in protein crystals, in units of \AA^2 .^[175]



The reverse Fourier transform, which is the process performed when a suitable set of initial phase estimates has somehow been determined, is presented below:

$$\rho(xyz) = \frac{1}{V} \sum_{hkl} F(hkl) \cdot \exp[-2\pi i(hx + ky + lz)]$$

This too may be broken into simple observations for the purposes of explanation. Because it is performed for every reflection collected in the diffraction pattern it is a summation. The $\rho(xyz)$ and $F(hkl)$ (electron density and structure factor) terms have swapped places either side of the equals sign, thus electron density at xyz is being calculated from the summation structure factors $F(hkl)$. There is also a negative sign in the exponential and so the reverse Fourier transform is a reciprocal of the forwards Fourier transform. The $1/V$ (reciprocal volume) term is necessary to achieve appropriate units, and it must be remembered also that $F(hkl)$ contains amplitude and phase information, and is expressed fully as $|F(hkl)| \cdot \Phi(hkl)$. Remember also that no information about $\Phi(hkl)$ is collected during the diffraction experiment.

Perhaps the most fundamental property of both Fourier transforms is the non-discrete relationship between their inputs and outputs.^[153] Every point in the unit cell $\rho(xyz)$ makes a contribution to every reflection $F(hkl)$ and inversely every observed reflection $F(hkl)$ contributes something to the electron density at every point $\rho(xyz)$. Thus an unanticipated gap in the diffraction data adversely affects the quality of the entirety of the calculated electron density map. Similarly, accurate phase estimates for only a subset of reflections will improve the quality of the entire calculated electron density maps.

1.2.6: The Phase Problem and Solutions to It

The fact that phases are not directly obtainable from diffraction data, also known as the phase problem,^[154] has been the subject of much research over the decades.^[160] Consequently, although it is problematic for structural solution, it is not an impossible hurdle. Some facts are known *a priori* to aid in determining an initial set of phases. Firstly, any phase estimates which return large negative values for electron density cannot be correct since it is not physically possible to have fewer than zero electrons at a given point in space. This reduces the number of feasible phase estimates somewhat.



Secondly, largely through crystallographic studies, the shapes and conformations of common groups and moieties are known. Thus phase solutions that yield a hexagonal ring structure for example are more likely to be correct than those which do not. The same holds true for phase estimates that return electron density peaks that are too close together to be adequately explained by known bonding modes. Such features may be searched for in the calculated electron density maps (real space) or reciprocal space. These observations alone, when combined with modern computing power are enough to determine a set of phase estimates that produce an interpretable electron density map. The crystallographer is usually in possession of knowledge as to what ligands, reagents and solvents may be present in the crystal and so a satisfactory structure may be built. This approach is called direct methods^[176] and is common in chemical crystallography where structures have up to a few hundred unique atoms, and may be performed using the SHELXS program.^[177]

Another method to solve the phase problem that is common in chemical crystallography and underpins some methods in macromolecular crystallography is the Patterson method.^[178] In this method all phases are set to zero and the reverse Fourier transform is performed on the structure factor amplitudes $|F(hkl)|$ only. The resulting map is a convoluted electron density map in which each peak corresponds to a vector between a pair of atoms (so there is one peak per pair of atoms) and from which it is possible to determine the interatomic distances between some or all of the atoms in the unit cell.

The Patterson map, being a map of pair-wise interatomic vectors, includes self-vectors between atoms to themselves. Since these have a directional magnitude of zero they all superimpose at the centre of the Patterson map on top of one another to give an extremely intense peak. Importantly, the magnitude of any peak in a Patterson map is proportional to the product of the scattering power (in turn proportional to the number of electrons) of the atoms causing the Patterson peak. So the second largest peak after the central self-vector peak always corresponds to the two heaviest atoms in the unit cell. Thus it is possible to deduce the number of electrons and therefore the identity of the atoms causing the Patterson peak and, provided the space group is also known, their positions. Once one or two atoms are positively located it places a constriction on all subsequent phase estimates since their



substitution into the reverse Fourier transform must not drastically increase or decrease the calculated number of electrons at this site in real space. Since n atoms produces n^2 interatomic vectors including self-vectors, Patterson methods are not alone used for phasing macromolecular crystal diffraction data because the Patterson map is too complex.

Patterson methods are used in conjunction with other phasing methods to provide a set of initial phase estimates, however. In chemical crystallography it may be combined with direct methods resulting in something called dual-space methods,^[179,180] which essentially means contributions to phases are estimated in real and reciprocal space. In macromolecular crystallography Patterson methods are used in conjunction with Molecular Replacement^[181,182] (MR) and Single-wavelength Anomalous Dispersion (SAD) methods.^[183]

Before discussing MR and SAD methods, a comment must be made on the contribution of heavy atoms to diffraction patterns. Since X-rays interact with atoms in a manner proportional the number of electrons “orbiting” the nucleus,^[150] and because a diffraction pattern is an interference pattern arising from scattering from all of the atoms,^[142,143] crystallographically-ordered heavy atoms tend to dominate scattering and have a large, systematic effect on the intensities of all of the observed reflections. This is an important addition to the point regarding how locating one atom in a structure drastically reduces the number of subsequent phase estimates that stand a chance of being “real”.

The contribution of heavy atoms to a diffraction pattern and how it may be exploited to overcome the phase problem is best illustrated with a simple one dimensional example. Imagine 1-iododecane forms a crystal with two molecules in the unit cell related by a two-fold rotation axis with their decyl chains aligned pointing together. When summing up the waves as per the reverse Fourier transform to produce electron density maps, the phase estimates used must be such that extremely large peaks in the electron density maps are observed at opposite ends of the unit cell corresponding to the iodine atoms. What is observed is, if the choice of phases is such that these expected large peaks are observed in accurate positions then small intervening peaks at positions corresponding to the ten carbon atoms, and minima at positions corresponding to the carbon-carbon bonds, are simultaneously



observed. These small peaks alone do not account for the full scattering power of the carbon atoms, but enough subsequent accurate phase estimates make up this defect. Thus the number of feasible phase estimates is significantly reduced by locating just a few heavy atoms.

SAD phasing has a basis that is more abstract than the other phasing methods but simply put an X-ray wavelength may be chosen such that the photons are more likely to promote an electronic transition from an inner atomic orbital^[184] (usually a K or L shell) of a heavy atom. This electronic transition is transient and the promoted electron returns to the ground state and, in becoming closer to the nucleus, a small amount of energy is released as a photon (X-ray fluorescence).^[185] This process is not instantaneous however, and the resulting photons have phase delay of 90° or $\lambda/4$ relative to the instantaneously diffracted waves.^[183] The probability of an electronic transition occurring, however, is small and so only a fraction of the incident waves will exhibit this phase delay.^[161,183]

To appreciate why the anomalous scattering can be beneficial, it must be noted that the geometric conditions which allow a reflection hkl to be observed must apply in the inverse direction for $-h-k-l$. This is called Friedel's law and the reflections hkl and $-h-k-l$ are called Friedel pairs.^[167] In the absence of anomalous scattering both Friedel pairs will have the same structure factor amplitude and phase. Since the anomalous scattering is 90° out of phase for both Friedel opposites however, (not 90° for one and -90° for the other) and because not all of the atoms are scattering X-rays equally, there arises an in-equivalence in the intensities of the Friedel pairs.^[161,183]

Differences in the Friedel pair intensities imparted by the anomalous scattering are small, and consequently accurate, high-redundancy data needs to be collected, often by performing multiple scans of the crystal.^[183] This in turn mandates low X-ray doses to avoid damage.^[186] Once sufficient anomalous scattering has been collected however, it is possible to determine a narrow range of phase estimates that satisfy the observed differences in intensities for the Friedel pairs and accurately position the heavy atom(s) in the unit cell.^[187] Once the heavy atoms have been positioned, Patterson methods and direct methods may be used to "bootstrap" the rest of the phases to obtain an electron density map of suitable quality for model building.



MR is only possible if there is a structural homologue of the target molecule.^[182] It relies primarily on Patterson methods and systematically modifying phases of a solved homologous structure, or those obtained from a Fourier transform of a homology model, and applying them to the target structure. Patterson maps for large molecules are very complex it is generally not possible to deduce atomic positions from it^[178] as is done for chemical crystals. It is true, however, that Patterson maps for large molecules are not featureless, and this can be exploited to determine an initial set of phase estimates.

The homologue molecule is first rotated about three orthogonal axes and a Patterson map is calculated for each iterative rotation.^[188] The agreement between the new Patterson map of the homologue and that calculated for the target molecule is statistically compared and the rotation process continues until the best statistical agreement between the two Patterson maps is found.^[188] The same process is then repeated for iterative translations of the homologue within the asymmetric unit of the target molecule again until the best agreement between the two Patterson maps is found.

Since the phases of the homologous structure are modified in an estimable way during rotation and translation it is possible in a successful case to substitute the phases derived from it, in a position that gives the best match between the Patterson maps, into the reverse Fourier transform equation along with the experimentally-observed $|F(hkl)|$ values to obtain interpretable electron density maps.

The fact that the complex nature of macromolecular Patterson maps precludes high-resolution structural data being derivable directly from them, but that they can be used to systematically modify high-resolution information from a homologous structure, make MR a powerful phasing tool that does not itself require high-resolution experimental diffraction data to work. The quality of the final electron density maps will always be governed by the resolution of the experimental diffraction data, however.

1.2.7: Model Building, Refinement and the R Factor

Model building in X-ray crystallography is performed such that the best match between diffraction parameters calculated from the model and those observed in the experimental diffraction pattern are achieved whilst paying respect to what is chemically and physically sensible.^[161] The extent to which



a model can be parametrised, and therefore the ability to build it accurately, depends on the quality and quantity of diffraction data collected. This is the so-called data:parameter ratio.^[189]

It has already been mentioned that the contribution of an atom to scattering of X-rays can be modelled by treating each atom as point scatterer.^[172] This is not sufficient to build accurate models however, as atoms are thermally mobile and the contribution to scattering decreases with thermal mobility. Atoms may be treated therefore as having spherical deviation about an equilibrium position and such a treatment has four parameters per atom: three positional coordinates and a (root mean square) principle radius. This is called isotropic modelling.^[190] A more accurate treatment of thermal motion may be made by treating atoms as ellipsoids which move in a concerted manner with those atoms to which it is bound. In such a model there are nine parameters per atom: three positional coordinates, three orientational vectors and three principle radii. This is called anisotropic modelling.^[190]

Building and modelling an electron density map by identifying atoms and their positions, allowing them to vibrate etc., all subtly modify of the initial set of phase estimates used to generate the initial electron density map. Whether or not these modified phases are better or worse than the initial phase estimates or some other previous set of estimates is performed through a process called refinement.^[189]

A theoretical diffraction experiment is performed on the atomic model with all its known symmetry and modelled features, and a calculated diffraction pattern is obtained. Although no phase information is contained in the experimentally-observed diffraction pattern, a statistical comparison can be made between the structure factor amplitudes of the calculated and observed diffraction patterns (between $|F(hkl)|_{\text{calc}}$ and $|F(hkl)|_{\text{obs}}$). If the agreement between these two values is observed to increase the modifications made to the model are deemed to be valid, the latest phase estimates are accepted and model building continues. If the agreement becomes worse they are deemed to be incorrect and are usually rejected. Each time a cycle of refinement is performed on an improved model, the phase estimates become a slightly more accurate estimate of those which were present but uncollectable



during the diffraction experiment. Because of this, and because the amplitudes have been experimentally measured and are invariant, the quality of the entire electron density maps tends to improve with more cycles of accurate model building and refinement. This effect is much more pronounced for small chemicals however, for which high-resolution data has been collected.

The actual process of refinement is mathematical calculation performed *in silico* to produce the best fit between the model as built and the experimentally-observed diffraction data. Least squares regression fitting of the data is typically used for chemical structures,^[191] whilst maximum likelihood methods are used in macromolecular refinement^[192,193] as it puts an additional statistical weighting on reflection intensities (model alterations that produce better agreement with intense reflections are not penalised like-for-like if there is less-good agreement with weak reflections).

Refinement may be unstable when a lot of parameters are being fit to comparatively few data,^[194] and generally speaking macromolecular structure modelling relies on poorer data:parameter ratios than chemical structure modelling owing to the nature of macromolecular crystals. Consequently proteins are almost universally modelled as isotropic whilst small molecules are almost universally modelled as anisotropic.

There are ways in which the data:parameter ratio can be altered after the point of data collection and this represents a major divergence in how chemical and macromolecular structures are solved. The two modes of modifying this ratio are via restraints and constraints.^[195] A restraint may put an artificial limit on how much an atom can thermally vibrate and in which direction it does so by applying a penalty function that penalises larger movements in certain directions. Restraints may also be applied such that the mobility of an atom is forced to match that of an atom to which it is bound. Additionally, the geometry of a functional group may be restrained to match some ideal structure. These restraints effectively “add data”.

A constraint removes all degrees of freedom of vibration and geometry and affixes an atom and its thermal motion to a value set by the crystallographer. Furthermore these parameters are not subject to refinement and so constraints “remove parameters”. Constraints are more severe invocations than



restraints and their use in chemical crystallography must be judicious since the data is usually good enough to build a model that allows for more freedom of the atoms. Furthermore restraints and constraints are overtly scripted and specified by the crystallographer to apply to specific atoms, as are the extent to which the penalty function applies

In macromolecular crystallography the data is seldom good enough to allow such elaborate tailoring of restraints to every atom in model however,^[195] and since the exact shape and connectivity of amino acids and their side chains is already known at high resolution they are usually severely restrained all of the time to ideal geometry.^[196] Modelling of a polypeptide chain is more concerned with side chain position and orientation, and backbone tracing since it is these parameters that mostly govern protein fold and function. Secondary structure restraints are definable within model building packages also.^[197] Thermal vibration of amino acids is modelled using the B-factor which describes two dimensional isotropic movements of atoms and these may be restrained such that B-factors for covalently bound atoms are similar. Overt scripting of restraints is necessary however, when ligands are observed in the electron density maps since these must be allowed some flexibility to account for the corresponding electron density.^[198]

The quality of an atomic model is quantified using the R factor^[199] and its calculation is given as:

$$R = \frac{\sum |F(hkl)|_{obs} - \sum |F(hkl)|_{calc}}{\sum |F(hkl)|_{obs}}$$

Thus the R factor is therefore a measure of agreement between the structure factor amplitudes obtained from the atomic model and the sample crystal in the diffraction experiment. The R factor has become more sophisticated over time and it is common for programs to either apply a weighting function that applies more credence to reflections with lower standard uncertainties associated with them. For similar reasons refinement is sometimes performed on the square of the structure factor amplitudes also.

A typical R factor for a good chemical crystal structure solution is around 6 %^[200] although porous structures such as metal-organic frameworks^[201] (MOFs) and zeolites^[202] contain large voids which



harbour large amounts of heavily-disordered solvent which is not well handled by chemical diffraction data processing software, and so R factors of > 10 % are commonly observed. For protein structures R factors of ~ 20 % are common.^[203] Since the program suites are programmed to interpret and process poor data, however, significantly higher values are acceptable. It will be seen that there is no information about phases included in the R factor equation which leads into the next section.

1.2.8: Model Bias and R_{free}

Calculation of the R factor is not unbiased.^[190,204] One goal of building an atomic model is to lower the R factor by incrementally improving the phase estimates. However the data that is being used to measure the correctness of these phase estimates during each cycle of refinement ($|F(hkl)|_{obs}$) are also used in the refinement process itself.^[194] Exacerbating this issue is the fact that phases contribute more information to image (and therefore the atomic model) than do amplitudes, and so inaccurate phase estimates can cause electron density to appear that does not actually exist. Consequently, it is possible to build a model which with each refinement appears to lower the R factor but is in fact wrong. This is called model bias.

The best way to ameliorate this is to leave a subset of reflections (usually 5 %) distributed across a broad resolution range out of the model building and refinement processes altogether, and assess how well these are predicted by the atomic model during refinement.^[205] This agreement is denoted by R_{free} and assignment of these reflections is performed strictly before any model building. R_{free} is given by:

$$R_{free} = \frac{\sum |F(hkl)|_{freeobs} - \sum |F(hkl)|_{calc}}{\sum |F(hkl)|_{freeobs}}$$

So it is calculated the same as the biased R factor, and the same intensity-based weighting functions are sometimes applied.

One area where model bias can become particularly egregious is in the fitting of ligands which may not truly exist. This can be overcome by calculating something called an OMIT map wherein the ligand atoms are deleted from the model and a new set of maps are calculated using data corresponding only to the ligand-less model.^[206] If no electron density reappears then the ligand is a



phantom arising only due to model bias and it is good practise to provide OMIT maps for ligands even if the diffraction data are good.

R_{free} is an opportune way for unscrupulous crystallographers to misrepresent the quality of their atomic models. All data may be included in refinement and the R_{free} set only assigned at the very end. It is also possible to not submit the final round of model building for refinement, usually in anticipation that modifications made to the model to ameliorate some geometric or torsional issue may be undone. This again misrepresents the accuracy of the atomic model and sets an artificial standard for other crystallographers, although making refinement metadata a prerequisite for structure deposition does prevent this.

1.2.9: General Single Crystal X-ray Diffraction Experiment Workflow

A sample crystal is irradiated with an X-ray source and a diffraction pattern containing a large number of reflections is collected. The intensities of the reflections are estimated by integrating their detector profile. Absorption and other sample illumination effects, which may occur to different degrees along each of the crystal axes are then corrected for and all of the diffraction data are scaled to compensate for these effects. This is called scaling.

Miller indices are assigned to each of the reflections by first determining the origin of the diffraction pattern, aided by the so-called orientation matrix of the crystal, and then counting out from the origin along three orthogonal directions h , k and l . This process is called indexing and allows for the Laue class to be determined.

Analysis of reflections which are systematically absent and which reflections have equal intensities allows for estimates of the point group, and then space group to be made. Once a satisfactory assignment has been made, usually aided by other quality control metrics such as completeness, symmetry-equivalent reflections are averaged in a process called merging. A list of unique reflections hkl with their associated intensities and standard uncertainties is produced in a step called reducing. A set of phase estimates is somehow obtained. These are combined with the measured intensities and electron density maps produced into which an atomic model is built and refined.



1.3.1: PhD Aims

The aims of the PhD project covered by this thesis are as follows:

- Develop a purification scheme for, characterise and solve the structure of full-length AgrA in either its *apo* form or in complex with promoter DNA to elucidate any of the following: the presumed structural rearrangement upon phosphate-binding, its mode of dimerisation, its asymmetry with respect to DNA binding and the manner in which it distorts promoter DNA.
- Develop a purification scheme for, characterise and solve the structure of AgrB in either its *apo* form, in complex with ambuic acid or in complex with AgrD.
- Investigate any experimental structures that are forthcoming to learn something of their structure-activity relationships and add to the existing knowledge of *S. aureus agr* with the broad and long-term goal of attenuating *S. aureus* virulence by interfering with this QS molecular circuitry.



Chapter 2: AgrA – Expression, Purification and Crystallisation

2.1: Preliminary Work on AgrA

A pCOLD1 construct for overexpressing full-length AgrA had already been generated by a former PhD student within the University of Nottingham *S. aureus* research group **Dr. Yanin Jaiyen**. Briefly, the construct contains an N-terminal (His)₆ affinity tag followed by a factor Xa protease site. pCOLD1 vectors also contain a so-called translation enhancing element which has the amino acid sequence MNKVH.^[207] These genetic elements add an additional 19 amino acids to the N-terminus of the AgrA sequence and the resulting construct expresses a 30 kDa AgrA molecule.

Prior research into AgrA within the group did not extend to structural studies so recombinant AgrA samples were not purified beyond Ni²⁺-affinity chromatography. Nevertheless the extreme sensitivity of the sample to oxidation was apparently problematic, with the sample precipitating out of solution within minutes of eluting from the Ni²⁺-affinity column unless 10 mM dithiothreitol (DTT) was immediately added to the sample fractions. The cysteine pair Cys199 and Cys228 is reported in the literature to be sensitive to oxidation.^[208] Specifically, disulphide bond formation between these two residues completely abolishes binding of AgrA to the P2 and P3 promoters *in vitro*. Consequently, before any structural experiments on AgrA were undertaken, the C199S point mutant was introduced via standard inverse Polymerase Chain Reaction (PCR) methods.

2.2: Initial Attempts to Purify AgrA_C199S

A priori knowledge of AgrA allowed some decisions about the purification strategy to be made before any experiments were performed. Its function is to bind to DNA, and since it has nM affinity for its cognate promoter sequences it is reasonable to assume it may have some affinity for any β -form DNA, a large amount of which is released into the sample matrix upon cell sonication. It was decided therefore to wash the AgrA_C199S sample with 1.5 M NaCl whilst it was immobilised on the Ni²⁺-affinity column to remove any sample-bound DNA fragments.



Secondary structure predictions and homology modelling indicated the presence of a flexible linker existing between the two discrete AgrA domains. To attempt to reduce the conformational heterogeneity of the sample in solution, 10 % v/v glycerol was added to all sample buffers except the Size Exclusion Chromatography (SEC) running buffer, in which it was present at 5 % v/v. It was also hoped that by acting as a crowding agent the glycerol might mitigate sample oxidation by reducing the number of collisions between AgrA_C199S molecules. As the flexible linker was considered to be sensitive to proteases that were also present in the sample matrix, 1 mM phenylmethylsulfonyl fluoride (PMSF) was added to the sample lysis buffer immediately prior to cell sonication.

Sodium dodecyl sulphate Poly Acrylamide Gel Electrophoresis (SDS-PAGE) analysis of sample fractions after Ni²⁺-affinity chromatography exhibited sufficient yield (~ 5 mg mL⁻¹ per litre of expression culture) for downstream purification (fig. 2.1C. The same gel marker was used for all gels presented in this thesis, and the M_r values of all of the constituent bands are shown in appendix I). It is worth noting that although the C199S point mutant lessened the sensitivity of AgrA to oxidation it did not completely abolish it, suggesting that two of the three remaining cysteines are sensitive to oxidation. Thus if the addition of reducing agent to the sample fractions was not expeditious, sample precipitation was still observed. Furthermore it was hard to estimate exactly how long it took for the sample to precipitate since it was dependent upon yield, which ultimately dictated the concentration of AgrA_C199S in the 2.5 mL fraction tubes, and thus the likelihood of AgrA molecules encountering one another. In any case full-length AgrA_C199S still required some protection with reducing agents.

Another issue encountered was the almost complete loss of the AgrA_C199S sample during SEC with subsequent SDS-PAGE gels exhibiting a dramatic reduction sample recovery from SEC compared to those gels obtained immediately after Ni²⁺-affinity chromatography (fig. 2.1B and D). No significant void peaks were visible on the chromatograms however, suggesting: A) the sample loss was either due to oxidation and/or precipitation during spin-concentration, B) some other co-purified species was degrading the AgrA_C199S into small soluble fragments or C) the sample had some intrinsic propensity to degrade that was exacerbated by processing and purification.

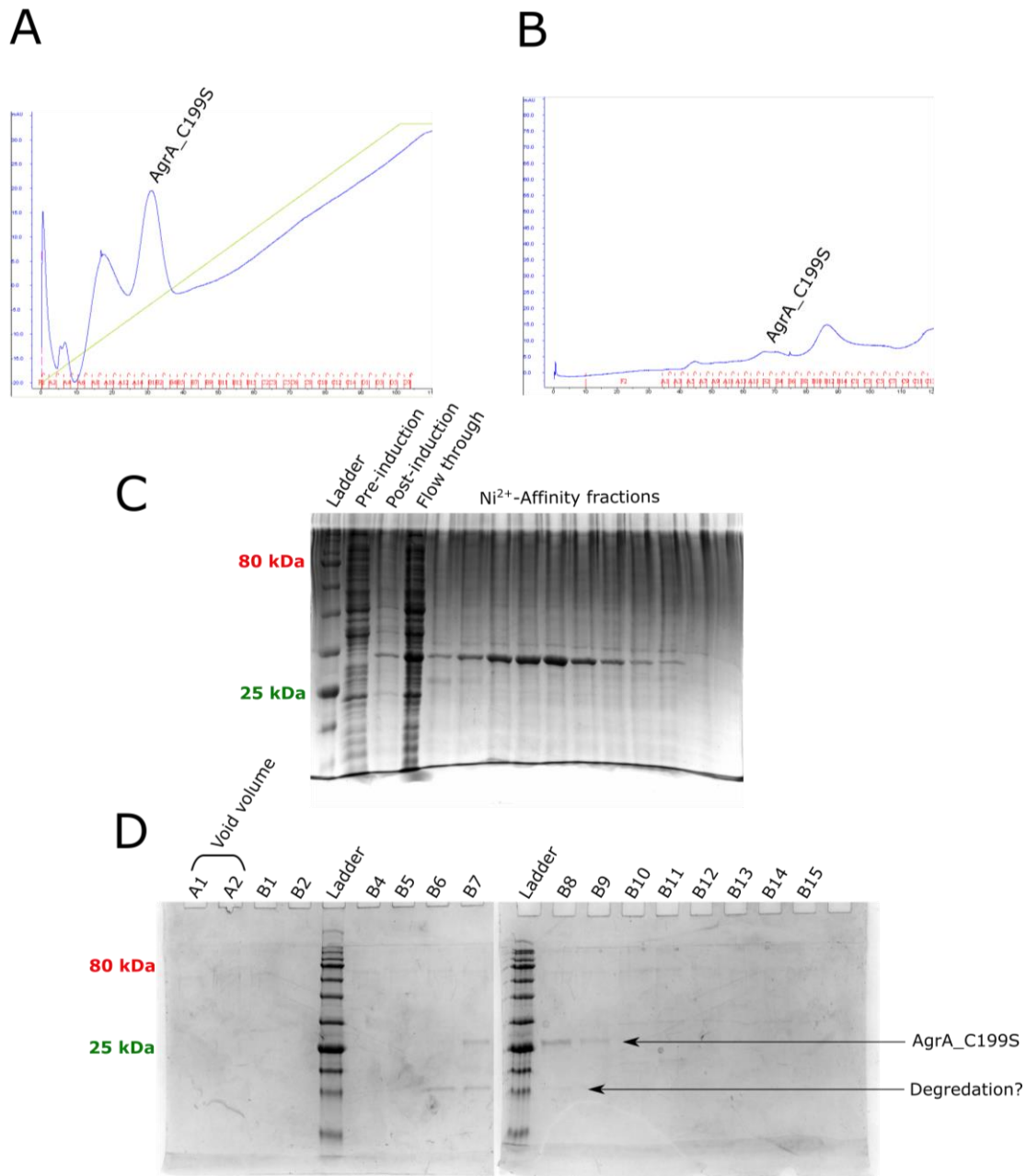


Figure 2.1. Initial attempts to purify AgrA_C199S. (A). The Ni²⁺-affinity chromatogram for AgrA_C199S. The first peak is a contaminant (gel of it not shown) and the second peak corresponds to AgrA_C199S. (B). A size exclusion chromatogram for AgrA_C199S ran on a Hi-Load® Superdex® 75 column. The running buffer is presented in the AgrA_C199S purification methods section 7.6.7. (C). The corresponding SDS-PAGE gel for (A) showing purified AgrA_C199S. (D). The corresponding SDS-PAGE gels for (C) showing almost total loss of the AgrA_C199S sample during SEC. Gels are 16 % polyacrylamide, reducing.



To try to protect the sample from degradation, the 1 mM PMSF (which inhibits only serine proteases) was substituted with cOmplete™ EDTA-free protease inhibitor tablets. These did not improve the sample recovery however, and since they are considerably more expensive to use than 1 mM PMSF on a “per purification” basis, the latter was reverted to.

Spin-concentration is often necessary prior to SEC since samples applied to the gel filtration column must occupy a volume that is only a small percentage of the column accessible volume in order to achieve effective sample separation. It can sometimes be detrimental to macromolecular samples however, due to the forces involved. To test whether spin-concentration was responsible for the sample loss, anion exchange chromatography was performed instead of SEC. To perform ion exchange chromatography samples must first be diluted until any salts present in the sample buffer are below a concentration at which they might interfere with binding of the sample to the column (< 20 – 50 mM) thus the sample is diluted rather than concentrated.

The Ni²⁺-affinity fractions were diluted into NaCl-less and imidazole-less buffer and loaded onto a 5 mL HiTrapQ column. Sample elution was performed exactly as per Ni²⁺-affinity chromatography only with so-called buffer *B* containing 1 M NaCl instead of 500 mM imidazole. A sample peak was observed on the resulting chromatogram and SDS-PAGE analysis revealed intact AgrA_C199S, however, it was clear that a large percentage of the sample was still being lost (fig. 2.2A and B).

Unless macromolecular targets for crystallisation overexpress extremely well, a spin-concentration step prior to establishing crystallisation experiments is unavoidable. As such the sample recovered from the anion exchange column had to be spin-concentrated anyway. Subsequent SDS-PAGE revealed further sample loss so anion exchange chromatography was abandoned in the pursuit of an AgrA_C199S sample suitable for structural studies.

To test whether some co-purifying species was degrading the AgrA_C199S sample, an ultracentrifugation step for 1 hr at 100,000 g was added between the cell lysis and Ni²⁺-affinity chromatography steps to try to remove contaminants. Subsequent SDS-PAGE analysis revealed that the ultracentrifugation was not helpful however (fig. 2.2C) and the idea was also abandoned.

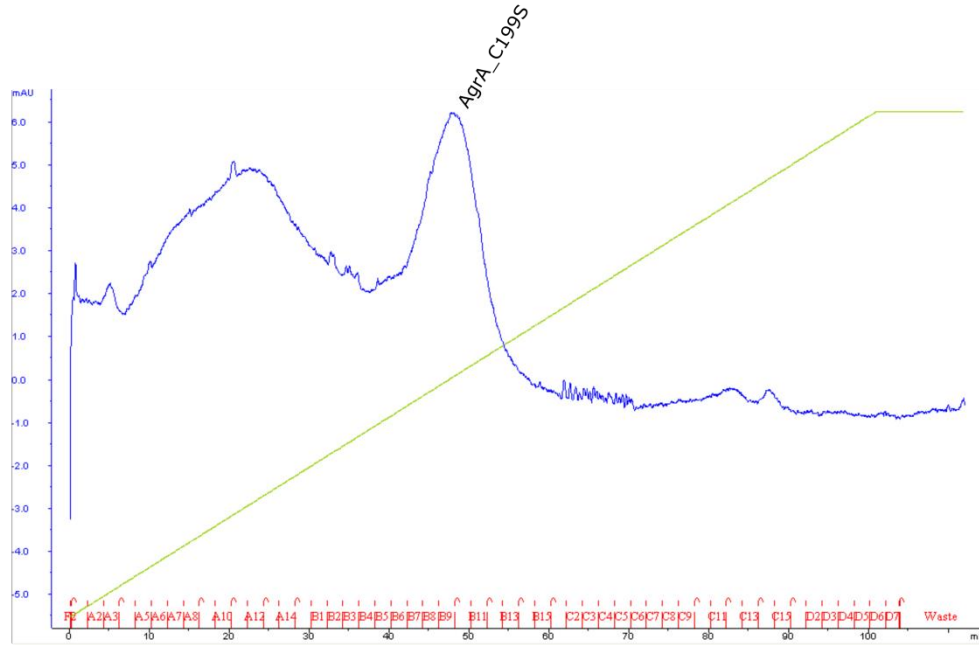
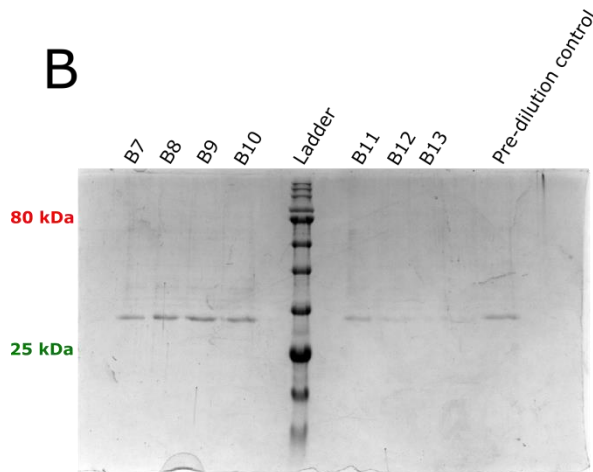
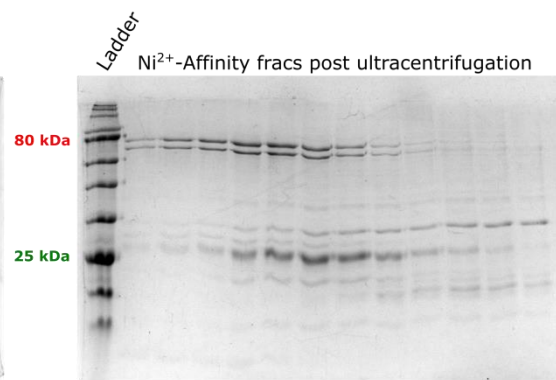
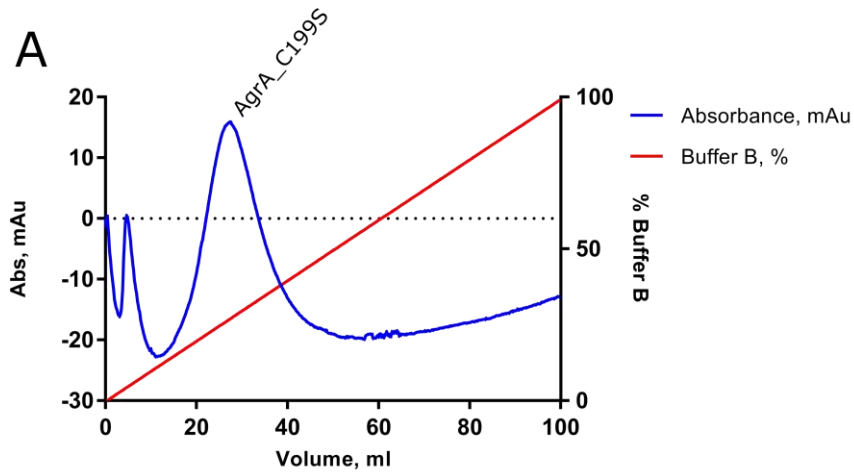
**A****B****C**

Figure 2.2. Failed AgrA_C199S purification modifications. (A). The anion exchange chromatogram for AgrA_C199S showing a distinct but small AgrA_C199S peak. (B). The SDS-PAGE gel corresponding to (A) showing a small amount pure and intact AgrA_C199S. Large amounts of sample was lost after Ni²⁺-affinity chromatography (gel not shown but essentially identical to that shown in figure 2.1B). The sample did not remain intact during spin-concentration prior to crystallisation experiments. (C). The SDS-PAGE gel resulting from ultracentrifugation of Ni²⁺-affinity purified AgrA_C199S in attempting to isolate it from potentially harmful species in the sample matrix. A multitude of species result, probably from the compressive forces in the ultracentrifuge. Neither strategy was used in subsequent AgrA_C199S purifications. Gels are 16 % polyacrylamide, reducing.

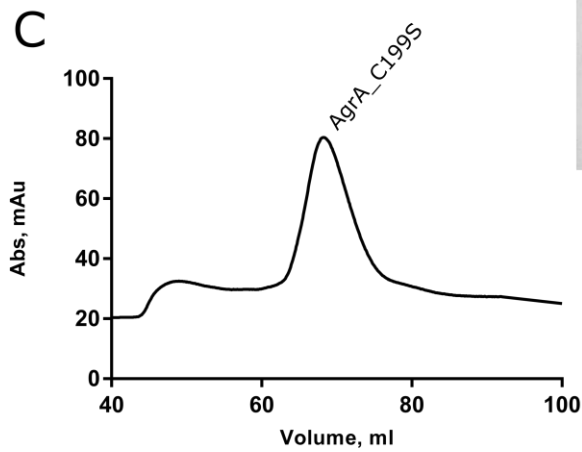
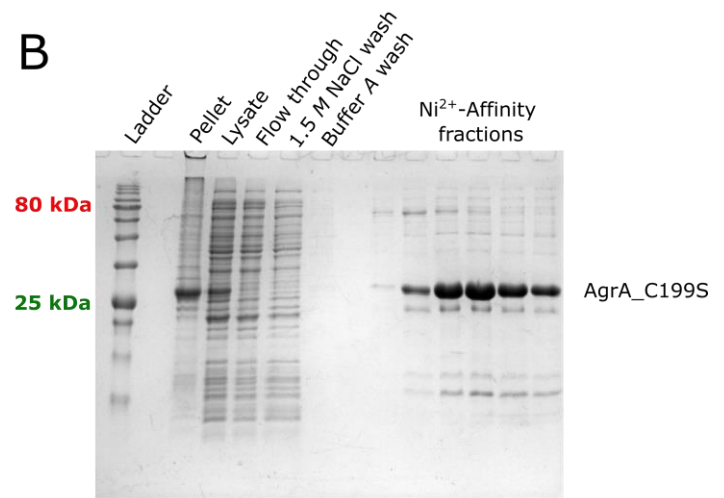


Owing to the lack of successful AgrA_C199S purification after overexpression and the relatively trivial Ni²⁺-affinity step, several changes were made to the purification procedure at once and a somewhat forceful approach was adopted. Hitherto the gel filtration attempts had been performed one day after Ni²⁺-affinity chromatography. Although the SDS-PAGE gels after SEC revealed sample degradation, they provided no insight upon when this degradation was occurring. It may have been shortly after running the first SDS-PAGE gel to assess the efficacy of Ni²⁺-affinity chromatography, overnight, during spin-concentration or on the column itself. It was decided therefore to perform Ni²⁺-affinity and SEC all in a single day by equilibrating the SEC column the day before the intended purification and keeping it at 4 °C overnight, removing it only when the sample was ready to be loaded.

Bearing the idea that some contaminant might be degrading the sample in mind, it was also decided that each cell pellet from 1 L of expression culture would be resuspended lysed into 50 mL of lysis buffer in order to dilute the crude sample matrix that exists after cell lysis and thereby (hopefully) reduce the susceptibility of AgrA_C199S to proteases. 1 mM EDTA was also included in the SEC running buffer to deactivate any co-purifying metalloproteases. The running buffer was kept on ice during the SEC run in order to keep the entire process as cold as possible. Gratifyingly, these changes resulted in two column purification which produced monodisperse AgrA_C199S in yields of 2 – 5 mg per litre of expression culture (fig. 2.3).



B



D

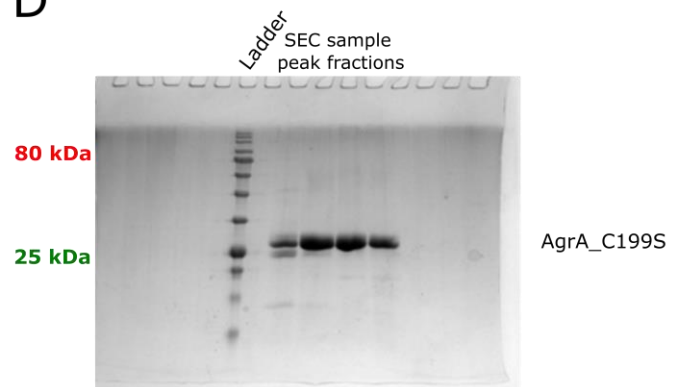




Figure 2.3 (previous page). **Chromatogram and SDS-PAGE gel data** for a successful AgrA_C199S purification after modifications to the initial procedure. (A). The Ni²⁺-affinity chromatogram with a single sample peak eluting at ~ 30 % buffer *B*. (B). Gel corresponding to the Ni²⁺-affinity chromatography step showing reasonably pure AgrA_C199S in modest yields (10 μL samples of unconcentrated 2.5 mL fractions per lane). (C). Size exclusion chromatogram for AgrA_C199S again showing a single sample peak eluting at ~ 68 mL. Some material elutes in the void volume and the absorbance never returns fully to baseline between the void and sample peaks. The best explanation for this is slow, irreversible aggregation of the AgrA_C199S molecules. This may also explain why the Ni²⁺-affinity purified sample did not remain intact when previously left overnight. (D). Gel corresponding to SEC showing highly pure AgrA_C199S. No higher molecular weight contaminants are observed but a few lower molecular weight contaminants are. These however are relatively low in abundance compared to the AgrA_C199S band. Gels are 16 % polyacrylamide, reducing.



Initial crystallisation experiments were performed on sample which was stored, usually overnight, at 4 °C. Once a crystallisation condition and appropriate sample concentration was discovered, however, the sample was spin-concentrated and flash frozen on N₂ (l) and later defrosted with no adverse effects as evidenced by a second SEC step on the defrosted material. This chromatogram (not shown) was identical to that performed during the initial purification. An additional 10 mM reducing agent (DTT or β-mercaptoethanol (BME)) was added to the defrosted sample to compensate for the instability of these reagents.

It is noteworthy that after the AgrA_C199S sample eluted from the SEC column it was stable, remaining in solution for several days at 4 °C. This seems in direct contradiction to the previous propensity of the sample to degrade. Which of the several modifications imparted this stability was not investigated further.

2.3: Sample Confirmation and Activity Assaying of AgrA_C199S

Although it was evident from the intensity and positioning of the SDS-PAGE bands corresponding to the Ni²⁺-affinity sample peak that the correct species was probably being purified, these observations did not constitute absolute proof of sample identity. It is always worth investing some time into confirming the correct sample before investing further time into structural investigations, in case it is discovered that what one thought was target sample is actually some other spurious protein masquerading as the desired sample. This is not such an infrequent issue as it may first seem and has led to summaries of various common contaminant affinities for different chromatographic technologies to be published.^[209,210]

Western blotting of AgrA_C199S was attempted but the results (now shown) were inconclusive. Although the anti-(His)₄ antibody detected an antigen to give chemiluminescence the gel images were smeary, preventing the positive identification of the correct AgrA_C199S band. The same was also true of the blot membranes when exposed to Ponceau stain.

Since the genetic sequences of the promoters to which AgrA binds are known and because AgrA has nanomolar affinity (K_d) for its cognate promoters,^[81] even in the absence of the phosphate ligand



required for *agr* upregulation *in vivo*,^[85] it was relatively trivial to devise a DNA binding assay to confirm the sample identity. This was initially performed on a preparative scale using SEC to measure shifts in retention volumes of AgrA_C199S before and after incubation with 2 molar equivalents of the 15 bp oligoduplex used to crystallise to the C-terminal domain of AgrA.^[80] This oligoduplex corresponds to the upstream AgrA binding site of the P2 promoter and the results of this peak shift assay were incontrovertible (fig. 2.4) ultimately leading to the crystallisation of full-length AgrA_C199S bound to this DNA fragment, herein denoted AgrA_C199S/P2_{fragment}. Once an initial structure solution had been built, EMSAs were used to detect DNA binding and probe structure-activity relationships as it requires smaller sample volumes (these are discussed later in section 4.3.3).

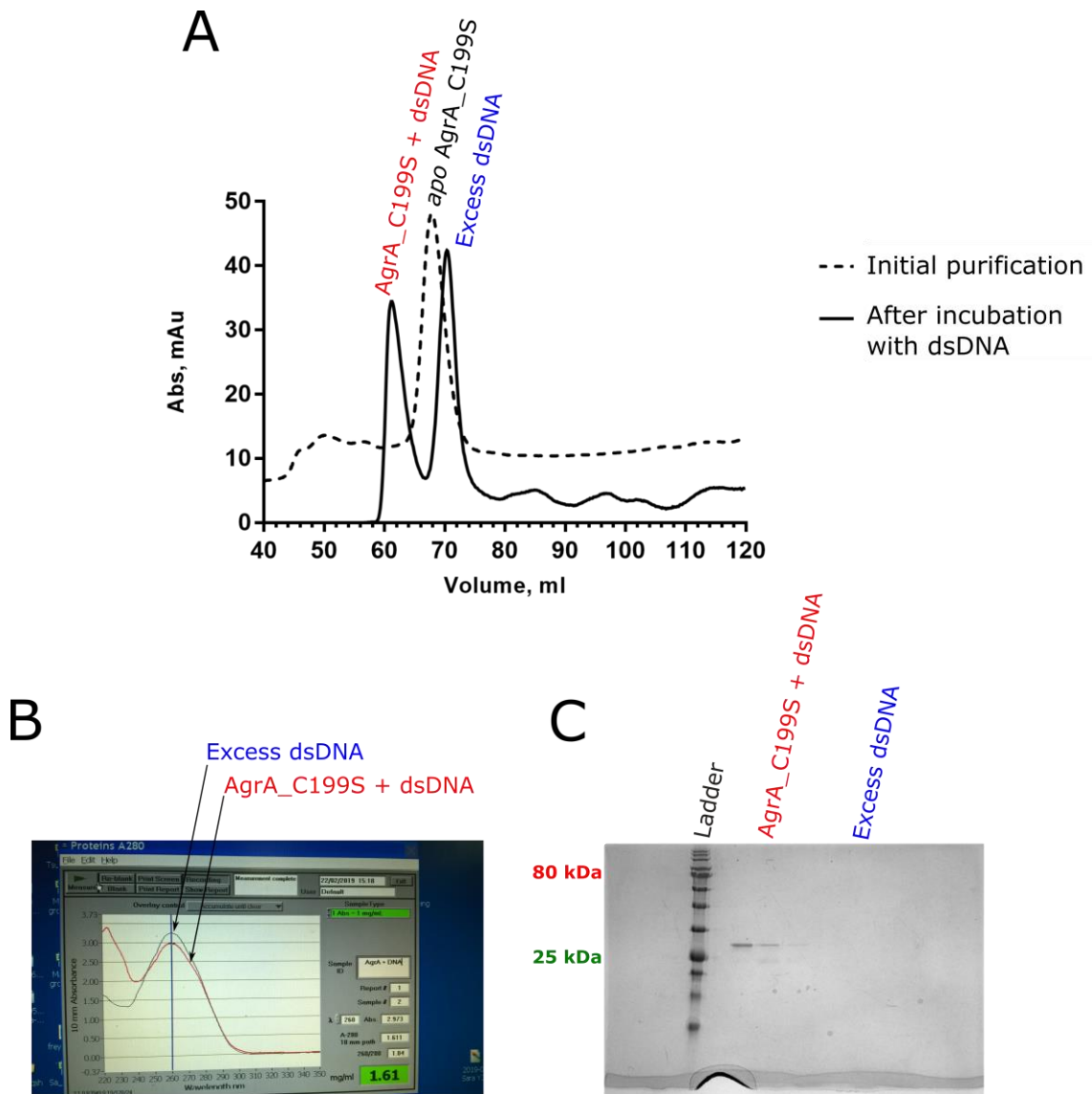


Figure 2.4. SEC peak shift assay for AgrA_C199S. (A). Size exclusion chromatogram for the *apo* AgrA_C199S sample (dashed line) over which is superimposed the size exclusion chromatogram resulting from incubation of a small amount of *apo* sample with 2 molar equivalents of a 15 bp oligoduplex (solid line). Two peaks are subsequently observed. The first to elute corresponds to the AgrA/DNA complex and the second to excess DNA. This was confirmed by Nanodrop™ 1000 absorbance measurements (B) which returned absorption maxima at 260 nm for both peaks, showing the presence of DNA, and SDS-PAGE analysis (C) of both peaks, of which only the earliest-eluting peak exhibits an AgrA_C199S band. dsDNA Stands for Double-Stranded DNA. The gel is 16 % polyacrylamide, reducing.



2.4: AgrA_C199S Purifies as a Monomer in Solution

AgrA_C199S purifies exclusively as a monomer according to SEC. This is in contrast to other examples of full-length response regulators which usually exhibit some dimerisation in solution even in the absence of a phosphate or phosphate-like ligand.^[89,90,211] Furthermore, attempts to mimic phosphorylation using either beryllium trifluoride, [BeF₃], as an orthophosphate mimetic,^[212] or by generating the D59E point mutant (both strategies discussed in more detail in section 3.9) also failed to induce any dimerisation of AgrA_C199S according to SEC. Both of these strategies have been employed to isolate response regulator dimers for structural studies^[211] and the D59E point mutant produces constitutively active AgrA *in vivo* according to *lux*-based reported assays in which the P3 promoter region is tethered to a luciferase reporter (Dr. Ewan Murray, unpublished data, not shown)

Binding of AgrA_C199S to the 15 bp oligoduplex also failed induce any dimerisation. Since each oligoduplex can only bind to one AgrA protomer, the observed lack of dimerisation implies that binding of AgrA to a single promoter binding site is not alone sufficient to induce dimerisation. This observation is important because the exact order of events: phosphorylation, dimerisation and DNA binding are not known in detail. Thus it conceivable that AgrA exists on the DNA as either a monomer or dimer and phosphorylation allows either recruitment of a second AgrA protomer or structurally modifies the dimer respectively, to upregulate the *agr* system.

The reasons as to why AgrA_C199S does not dimerise *in vitro* even when exposed to the necessary stimuli are many but may be as simple as the ionic strength of the sample buffer preventing the necessary AgrA···AgrA contacts from forming. Conversely the reason may relate to selective differentiation of AgrA between the promoters that are under its control. AgrA is known to augment *agr* upregulation in at least one redox-dependant manner, by upregulating the glutathione peroxidase gene *BsaA* at the expense of *agr* in response to oxidative stress.^[208] There are also the two *psm* genes discussed in the introduction that are directly up regulated by AgrA in addition to the large number of genes downregulated by it.^[98] Taking these into consideration it is possible that dimerisation of AgrA



and subsequent binding to its promoter regions may not be as simple as a phosphate-driven on/off switch that is readily detectable *in vitro*.

Techniques which are more sensitive to detecting protein oligomerisation such as Dynamic Light Scattering (DLS),^[213] SAXS^[214] or ultracentrifugation gradient separation^[215] may yet reveal some transient AgrA dimer in either its *apo* form or in response to the above-mentioned stimuli. These, however, were not performed in the pursuit of AgrA crystals due to limited time and the difficulties encountered when optimising said crystals.

2.5: Initial Crystallisation Experiments for AgrA_C199S With and Without DNA

The first set of crystallisation experiments for AgrA_C199S were performed on the *apo* protein and for AgrA_C199S in complex with the above-mentioned 15 bp oligoduplex in parallel. This short piece of DNA was crucial in obtaining crystals of the C-terminal domain of AgrA.^[80] In addition to the 15 bp dsDNA this oligoduplex contains unpaired 5' A/T “overhangs” enabling it to form H-bonds to complimentary regions on adjacent DNA molecules. These complimentary overhangs form critical lattice contacts in the crystals of the C-terminus/DNA complex via reverse Hoogsteen base pairing between adjacent DNA molecules, to give rise to a pseudo-continuous double helix (with breaks in the sugar-phosphate backbone every 15 nucleotides) along a 4_1 screw axis.^[80]

A detailed explanation of how the AgrA_C199S/P2_{fragment} complex was generated can be found in the methods section 7.6.11. Briefly, AgrA_C199S was concentrated to 250 μ M, and 2 molar equivalents of the oligoduplex stock solution were added. The oligoduplex stock was generated through annealing of single stranded oligonucleotides 5'-TTTAACAGTTAAGTAT and 5'-AATACTTAACTGTAA in Milli-Q™ H₂O to reduce the number of unique chemical species present during the crystallisation screening. The DNA was allowed to bind to AgrA_C199S for 30 mins on ice after which two two-fold serial dilutions were performed into SEC running buffer to achieve protein concentrations of 250, 125 and 62.5 μ M.

Identical serial dilutions were also performed on the *apo* sample to achieve two parallel sets of samples. The decision to start crystallisation experiments at these sample concentrations was not



determined by any empirical method such as the pre-crystallisation test [Hampton Research] but the trade-off between adequate sample concentration and large enough sample volume that inevitably occurs when dealing with low-to-moderately expressed proteins such as full-length AgrA_C199S.

Crystallisation experiments were established using the commercial screens: Morpheus®, PACT Premier™, MIDASPlus™ [Molecular Dimensions] and Nucleix [Qiagen]. Initially 250 and 125 μM samples were used in conjunction with the Morpheus® screen, however, inspection of the experimental drops immediately after sealing of the experiments revealed unsatisfactory levels of precipitate for almost all of the conditions, indicating that the sample concentration was too high. Subsequently 125 and 62.5 μM samples were thenceforth used with the four aforementioned screens and the experiments were incubated at 10 °C.

Crystals were observed in PACT Premier™ wells: F1, F6 and F7 for 62.5 μM AgrA_C199S/P2_{fragment} and: F1, F2, F3, F6 and F7 for 125 μM AgrA_C199S/P2_{fragment}. No crystals were observed for *apo* AgrA_C199S at any concentration. The PACT Premier™ screen is a systematic screen and all conditions in row F contain: 100 mM bis-tris propane pH 6.5, 20 % w/v polyethylene glycol (PEG) 3350. Every condition contains a unique anion however, which is maintained at 200 mM. Although all of the crystals were very small (estimated to be around 10 – 20 μm for the longest axis) those grown in the presence of 200 mM NaF at 62.5 μM AgrA_C199S/P2_{fragment} corresponding to well F1 (fig. 2.5A) were the largest.

125 μM AgrA_C199S/P2_{fragment} in condition F1 (fig. 2.5B) contained the greatest number of crystals, however, these were noticeably smaller than those grown at 62.5 μM . Wells F2 and F3, corresponding to 200 mM NaBr and NaI respectively, only produced crystals at 125 μM AgrA_C199S/P2_{fragment}. Furthermore these crystals were noticeably narrower than those grown in F1. Crystals grown wells F6 and F7, corresponding to 200 mM HCOONa and CH₃COONa respectively, appeared at both concentrations of AgrA_C199S/P2_{fragment} but were both narrower than those grown in F1 and significantly less abundant.



Given these observations the chemical composition of condition F1 (200 mM NaF, 100 mM bis-tris propane pH 6.5, 20 % w/v PEG 3350) was chosen as the starting point for crystal optimisation. It appeared as if the crystals grew in one of two habits with some growing as elongated rectangular-based prisms and others growing to produce a cuboidal shape. It was difficult to be certain of this however, due to the small crystal size. Another observation made at this stage was that generally the *apo* AgrA_C199S sample produced much more precipitate than AgrA_C199S/P2_{fragment} sample across all of the four commercial screens used. Furthermore the precipitate that was observed from the AgrA_C199S/P2_{fragment} sample was generally of much finer texture and more evenly distributed. This qualitative observation supports the idea that addition of DNA stabilises the AgrA sample *in vitro*, however, this was not tested quantitatively.

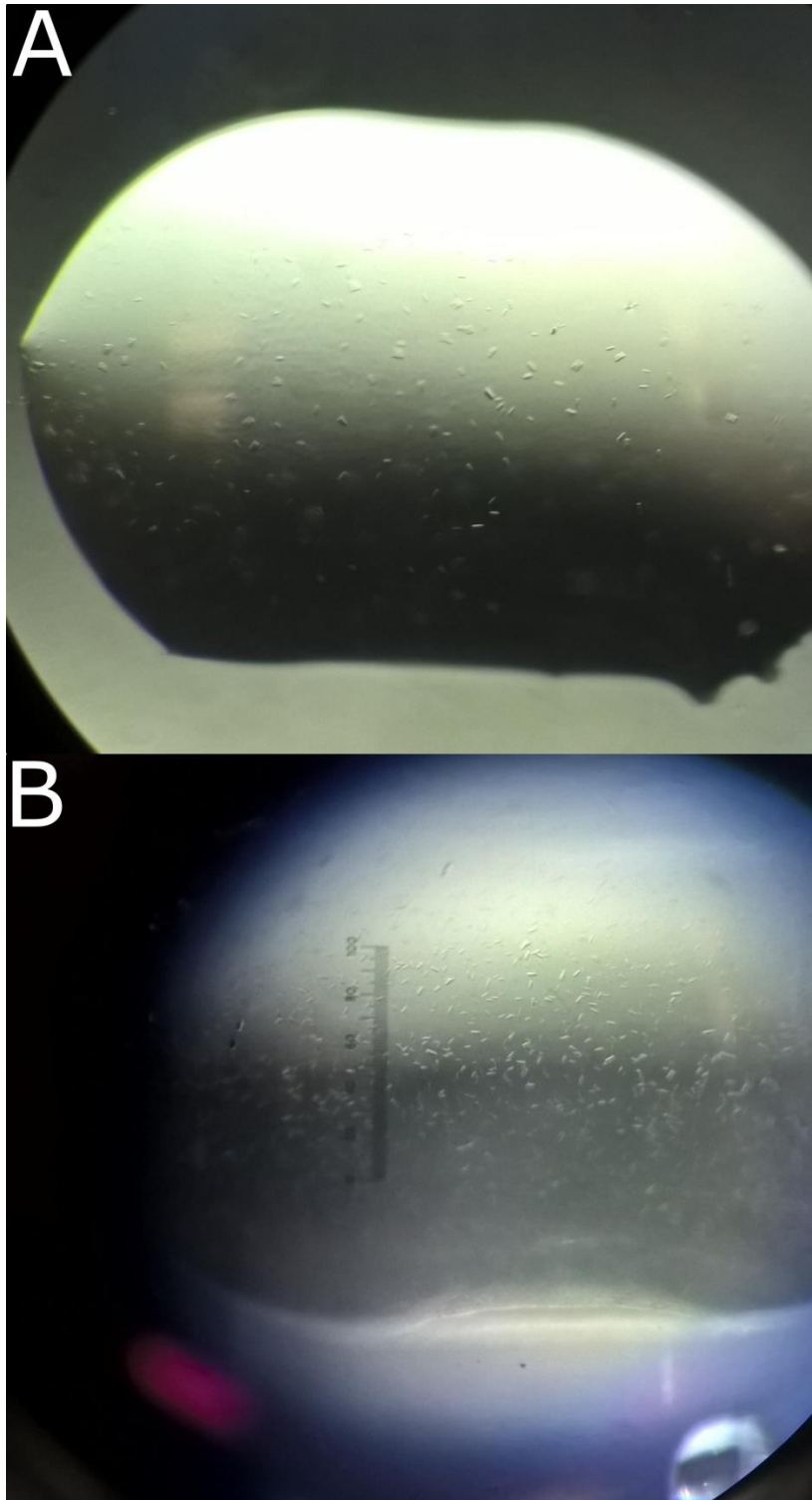


Figure 2.5. Initial AgrA_C199S/P2_{fragment} crystals grown in PACT Premier™. (A). Those grown at a sample concentration of 62.5 μ M. (B). Those grown at a sample concentration of 125 μ M.



2.6: Control Crystallisation Experiments for AgrA_C199S/P2_{fragment}

Before probing the successful crystallisation condition it was considered prudent to assess the crystals to determine whether they corresponded to AgrA_C199S/P2_{fragment} or to a false positive. Often enough the latter is the case for protein crystallisation experiments which is not surprising given the chemical diversity of the crystallisation mother liquors that are typically used.

In the particular case of the AgrA_C199S/P2_{fragment} complex, crystallising the excess DNA was a concern since crystals of it would probably be hard to distinguish from genuine protein/DNA complex crystals. The simplest approach was deemed to be a control crystallisation experiment in which sample buffer, containing DNA but not AgrA_C199S, was mixed with the crystallisation solution in exactly the same fashion as the experimental case and the drops monitored for precipitate and/or crystals.

Controls corresponding to DNA concentrations of 125, 250 and 500 μM in sample buffer were prepared in triplicate. It is worth highlighting that whilst undoubtedly this is a sound strategy to test for false positive crystals, DNA (and proteins) may act as precipitants for chemicals just as PEGs etc. act as precipitants for macromolecules. That is to say simply deducting a species from the crystallisation experiment can never constitute a perfect control since the fundamental physical and chemical environment of the drop will have changed.

Gratifyingly however, these control crystallisation experiments remained permanently clear of precipitate and crystals during the several months for which they were monitored (fig. 2.6). Even though other methods of probing crystals were available^[216] such as crushing the crystals to check for hardness, fishing and washing the crystals and running them on an SDS-PAGE gel, staining the crystals with a small dye or checking for fluorescence from tryptophan residues and so on, each test is often inconclusive and so they were not performed.

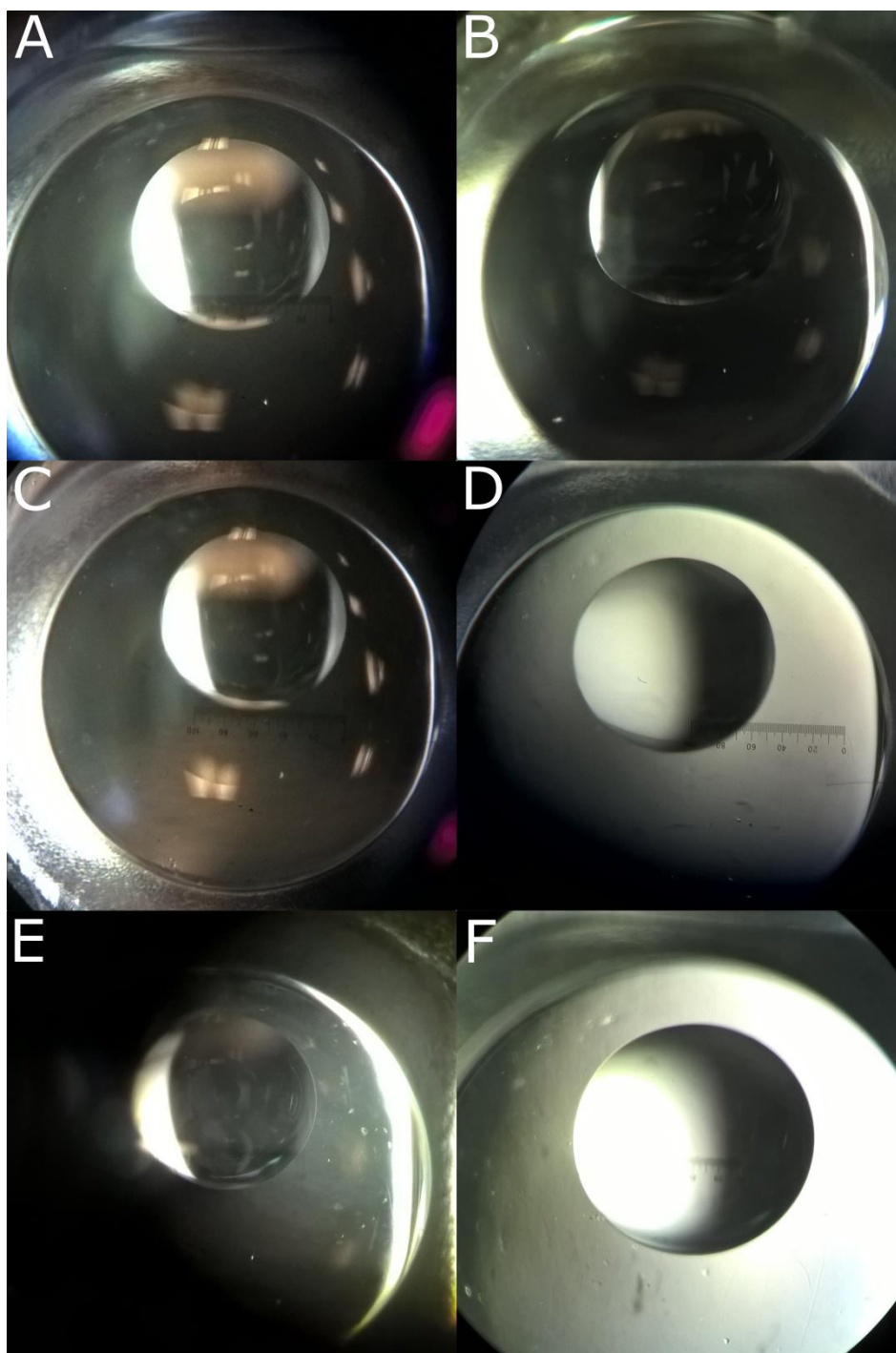


Figure 2.6. Control experiments for the AgrA_C199S/P2_{fragment} crystals in which SEC running buffer and DNA were mixed with a solution matching the PACT Premier™ condition F1, in ratios identical to the crystallisation experiments. No crystals or precipitate was observed after: (A) one day, (B) two days, (C) one week, (D) two weeks, (E) one month or (F) two months. Note that these time intervals are approximate but the observation that no crystals or precipitate was observed stands.



2.7.1: Optimisation of the AgrA_C199S/P2_{fragment} Crystals

The crystal optimisation and X-ray diffraction screening processes were iterative, and many trips to Diamond Light Source were required throughout the optimisation process to judge the quality of crystals resulting from all of the various crystal optimisation strategies employed. For the sake of simplicity however, the optimisation and diffraction screening processes will be discussed separately with cross referencing as needed.

2.7.2: Chemical Grid Optimisation

The chemical space around the PACT Premier™ F1 condition was explored by varying the NaF concentration from 150 to 225 mM, in 25 mM increments, and by varying the PEG 3350 concentration from 13 to 23 % w/v in 2 % increments. The bis-tris propane was maintained at 100 mM at pH 6.5 in all conditions. A separate but otherwise identical grid screen was prepared using NaBr instead of NaF with a view to performing experimental phasing of the crystallographic data using the K edge of the bromine atoms should any crystals form. Note that due to the **low solubility of NaF** in H₂O, the maximum concentration of the NaF stock achievable was 0.5 M. This meant that it was impossible to prepare conditions corresponding to: 19, 21 and 23 % w/v PEG 3350 at 225 mM NaF; 21 and 23 % w/v PEG 3350 at 200 mM NaF; and 23 % w/v PEG 3350 at 175 mM NaF owing to insufficient space in the designated final volume. This was not an issue for the NaBr grid screen however, since NaBr can be readily prepared to 1 M in H₂O.

AgrA_C199S/P2_{fragment} was prepared as described and concentrated to 62.5 μM, and incubated with 1 molar excess of DNA (DNA concentration would later become an optimisation variable). Experimental drops were set up by mixing 1 μL of sample with 1 μL of optimisation condition in a 48 well sitting drop crystallisation plate and incubating this against an 80 μL reservoir. Each grid screen NaF and NaBr was prepared in duplicate, each one of which was incubated at 10 °C and 20 °C, thus four grids were prepared in total.

Considering that the original crystals grew at 20 % w/v PEG 3350 and that the optimisation grids only extended to 21 % w/v PEG 3350, further optimisations were prepared wherein the PEG 3350



concentration was increased from 21 to 27 % w/v in 2 % increments, with as much NaF as was physically possible to dilute into the remaining volume from the 0.5 M NaF stock solution. Again, duplicate grids were prepared for NaF and NaBr and these were incubated at 10 and 20 °C.

Interestingly, although the original crystals grew at 10 °C, none of the optimisation grids reproduced crystals at this temperature. This is probably due to a discrepancy between the pH of the bis-tris propane stock solution prepared by the experimenter and that used by the manufacturer of the original screen. pH is affected by temperature, especially for tris and tris-based buffers.^[217] Therefore the true crystallisation pH of the original experiment may not have been achieved by the optimisation grids which were incubated at 10 °C once any discrepancy is taken into account. This argument is supported by the fact that crystals failed to grow in any of the conditions in rows E or G of the original PACT Premier™ screen. These rows are chemically identical to those in row F except that conditions in row E are unbuffered whilst conditions in rows G and H are maintained at pH 7.5 and 8.5 respectively. Taken together these observations suggest a strong pH dependency for AgrA_C199S/P2_{fragment} crystal formation. Furthermore, mildly acidic crystallisation pHs are common for protein/DNA complex crystals owing to the poor acid-solubility of DNA.

Crystals regrew in many of the NaF-based conditions which were incubated at 20 °C. The largest crystals were observed in 175 mM NaF, 100 mM bis-tris propane pH 6.5, 21 % w/v PEG 3350 thus this condition was used for downstream optimisation experiments.

2.7.3: Hanging Drop Optimisation

Growing crystals in a drop suspended over the vapour diffusion reservoir has been known to increase the size of crystals and is the default crystallisation strategy in some laboratories.^[218] With this in mind 5 µL hanging drops containing 1:1 mixtures of AgrA_C199S/P2_{fragment} and crystallisation solution were prepared on siliconised glass cover slides and incubated over 500 µL reservoirs. Identical experiments were established in parallel in a single 24 well hanging drop plate which was maintained at 20 °C and periodically inspected for crystal formation.



No crystals formed in any of these experiments. At this stage in time however, some of the original hit crystals had been shipped to Diamond Light Source and, from the crystal fishing process, it was apparent that the crystals were all either stuck to the bottom of the plate or were in a thick skin that covered the meniscus of the drop. No crystals existed in the body of the drop. This complicated the fishing process but taken with the entirely negative hanging drop results lead to the conclusion that the AgrA_C199S/P2_{fragment} crystals needed some physical support in order to form.

This awkward property manifested throughout the remainder of the project and no way to ameliorate it was discovered. Nonetheless it did present an opportunity separate the nucleation and crystal growth processes via combining the crystal seeding approach^[219] with hanging drop method in which no spontaneous nucleation of AgrA_C199S/P2_{fragment} occurred.

2.7.4.1: Seeding Optimisations

All seeding experiments were performed at AgrA_C199S/P2_{fragment} concentrations at 100, 80 and 50 % of the concentration used that produced the best crystals during sparse matrix screening. There reasons for this were: A) delivering seeds to an experimental drop bypasses the requirement for nucleation which occurs at higher concentrations of sample than does crystal growth and B) it was thought that lowering the sample concentration might prevent spontaneous nucleation and so encourage crystals to grow only from the seeds. For the sake of the following discussion an abbreviated method for each the seeding experiments is presented below.

2.7.4.2: Streak Seeding into Hanging Drops

A hair belonging to the experimenter was briefly wiped through some of the original hit crystals and then immediately wiped through a 5 μ L experimental drop which had been equilibrating against reservoir for \sim 7 hrs at 20 °C. The experiment was resealed and inspected for crystals. Care was taken not to leave the drop exposed to air for too long as this was presumed to introduce oxidative and evaporative effects detrimental to crystal formation and growth.



2.7.4.3: Streak Seeding into Sitting Drops

A hair was wiped through the original crystals and then wiped through a 2 μL sitting drop which had also been equilibrating for ~ 7 hrs at 20 $^{\circ}\text{C}$. The same precautions were taken as the above.

2.7.4.4: Seeding into Hanging Drops from Seed Stocks

Seed stocks were prepared using seed beads [Hampton Research] using the supplied protocol. Briefly, some of the crystals which grew during initial chemical grid optimisation were broken up using acupuncture tools and scraped from the bottom of the plate using a fine acupuncture needle. 2 μL of reservoir was added to the drop to retard any evaporation. All of the crystal debris was transferred to a chilled Eppendorf® tube containing a seed bead and a further 78 μL of cold reservoir was added. This was then vortexed for short bursts with intermittent incubations on ice to prevent dissolution of the crystal fragments. 5 ten-fold serial dilutions were then prepared via dilution into cold reservoir to achieve a batch of seed stocks ranging from [seed stock] = x down to [seed stock] = x^{-5} . These seed stocks were then titrated into experimental drops as described in the methods section 7.6.7.4.

2.7.4.5: Seeding into Sitting Drops from Seed Stocks

The same seed stocks as described above were titrated into sitting drops as described in the methods section 7.6.7.4.

2.7.4.6: Discussion of the Seeding Results

The most successful seeding strategy, as judged by those which produced the highest number of geometric single crystals, was streak seeding into hanging drops. Generally speaking a continuous streak of crystals grew along the streak line, all of which were firmly stuck to the glass cover slip. In a few instances however, a few of these appeared to become dislodged at some stage and fall into the bulk of the drop and grow to an appreciable size. The diversity of results obtained from streak seeding into hanging drops is presented in figure 2.7.

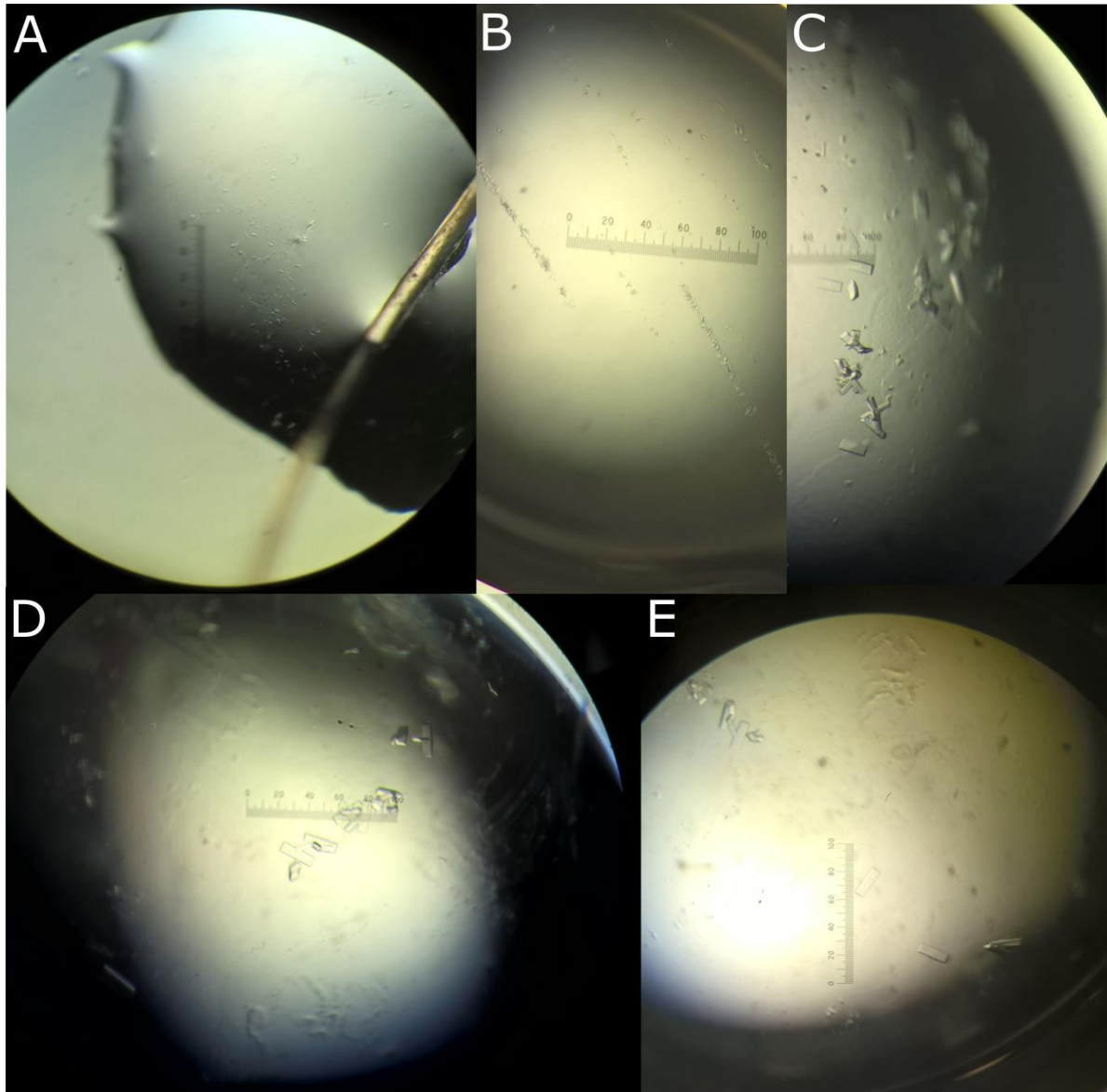


Figure 2.7. Streak seeding into hanging drop experiment results. (A). The microseeds were delivered by a fibre, in this case a hair of the experimenter, into fresh hanging drops (B) through (E). In a many of experiments no crystals grew after streak seeding and so there was nothing to photograph. Often enough the resulting crystals were tiny and clustered together along the streak line(s) as in (B). These were of little use for diffraction experiments. Occasionally however, delivery of the seeds was such that only a few microseeds were delivered to the hanging drops, resulting in a loose cluster of crystals all in a line with a few break-away single crystals as in (C) through (E). Such isolated crystals are the desired result but took repeated attempts to successfully grow.



The seed stocks only produced crystals down to $[\text{seed stock}] = x^{-2}$. When titrated into the hanging drops the crystals grew generally to a larger size than those resulting from the streak seed approach, however, there was also higher degree of crystal twinning and the crystals were always stuck to the bottom of the well. Both streak seeding and delivering seed stocks into sitting drops was less efficacious than for hanging drops, although crystals did grow in both of those cases.

It is worth stating explicitly that there was a high failure rate in reproducing crystals, with about half of the drops simply failing to regrow crystals at all. Of the half that did, only a small percentage went on to produce large single crystals suitable for X-ray diffraction experiments, with the rest all being “much of a muchness” and producing crystals of unsuitable size and/or quality. This necessitated a high degree of redundancy in crystallisation experiments. Nevertheless it was clear that streak seeding into a large hanging drop could produce geometric single crystals of AgrA_C199S/P2_{fragment} that were much larger than the original hit crystals.

2.7.5: Additive Screening

Although relatively large crystals were grown from the seeding experiments, preliminary X-ray diffraction analysis revealed only poor diffraction from the crystals even by macromolecular standards. Reflections were only observed out to approximately 5 – 6 Å even at high beam transmission and relatively long exposure times. Thus it was becoming apparent the previously-observed weak diffracting power of the crystals was not strictly related to their small size but was related to intrinsic properties of the bulk crystal.

Additive screening, in which reagents that are known to have a stabilising effect on macromolecules are titrated into the crystallisation experiment, can be an effective way increasing crystal size and improving their internal order.^[220] Thus it was employed to try to improve the quality of the AgrA_C199S/P2_{fragment} crystals. The rationale is much the same as the initial crystallisation experiments, with a large number of chemically diverse reagents being titrated into the original crystallisation condition in parallel and the resulting drops incubated and periodically inspected.



Improvements in the crystal may be visually obvious and confer some morphological or dimensional improvements, or purely internal and invisible to the naked eye.

For the AgrA_C199S/P2_{fragment} crystals both the Angstrom Additive Screen™ [Molecular Dimensions] and the Additive Screen [Hampton Research] were tried in parallel. Both screens exist as an assortment of 10 x additive stock solutions. Generally it is wiser to dilute the additives into the reservoir for, if the reagent is volatile, it will simply evaporate entirely from the experimental drop thus this approach was used for additive screening of the AgrA_C199S/P2_{fragment} crystals. It was initially performed using the sitting drop method as these are easier to perform and parallelise. Any reagent that conferred a benefit to the crystals could then be used in conjunction with the seeding and hanging drop methods. The optimised crystallisation condition of: 175 mM NaF, 100 mM bis-tris propane pH 6.5, 21 % w/v PEG 3350 was prepared at 1.1 x concentration for the additive screening experiments to compensate for the dilution incurred by addition of the additive stock solution.

The efficacy of each of the additives was judged by inspection of the resulting crystals for improvements in size over crystals grown under control conditions. The most striking trend amongst the efficacious additives was that they were almost exclusively simple sugars or other short chain polyols. Here it is worth mentioning that the addition 1 M D-sorbitol is absolutely necessary to overexpress soluble full-length AgrA under osmotic shock conditions.^[221,222] Whether the two observations are related and simple sugars actively stabilise the AgrA protein during overexpression or whether the two are entirely coincidental is impossible to say.

Of all of the additives which led to an increase in crystal size, L-rhamnose gave the largest overall improvement relative to the control crystals. The final L-rhamnose concentration in the experimental drops was 1.2 % w/v and so was included at this concentration in all further optimisation experiments.

2.7.6: Macroseeding *In Situ*

Both streak seeding and seed stocks belong to a category of seeding called microseeding as only fragments of the starting crystals are being used as the seeds. Entire crystals may be used as a starting point for additional crystal growth however, and this is called macroseeding.^[219] Usually the entire



macroseed crystal is fished, washed and transferred to fresh mother liquor for continued crystal growth. Because the AgrA_C199S/P2_{fragment} crystals were extremely soft and were either in a skin or stuck to a surface of the experimental plate however, it was deemed inadvisable to disturb them to ensure as little damage as possible occurred between growing the crystals and delivery to the synchrotron beam. The traditional method of macroseeding was not therefore viable.

To overcome this, a strategy was devised to deliver fresh mother liquor to the original crystal and macroseed it *in situ*. Experimental drops containing good candidate crystals which had grown to maximum size were opened and the “depleted” mother liquor removed 1 μL at a time until it was impossible to remove any more. Fresh mother liquor was then carefully added onto the resulting crystal-containing puddle of remaining liquid and the experiment resealed. This was only ever tried on hanging drops but there is no obvious reason why it would not work for sitting drops also. The position and orientation of the candidate crystals was noted and the pipetting manoeuvres were then performed as far away from them as possible. Since the AgrA_C199S/P2_{fragment} crystals were generally stuck in place there was little risk of accidentally aspirating them, however, a few crystals inevitably succumbed to the bludgeoning effects of the pipette tip.

These macroseeding experiments were critical to obtaining structure-grade crystals of the AgrA_C199S/P2_{fragment} complex (fig. 2.8). Several rounds of microseeding were performed on crystals until they were very large. Comparison of the crystals before and after microseeding, and to the original hit crystals are striking with the latter being orders of magnitude larger. One noteworthy downside to the method was the propensity of some crystals to sprout small crystalline appendages after the second or third round. This could theoretically be mitigated by opening the drop to halt the crystal growing process but this was never performed.

In situ macroseeding of macromolecular crystals does not appear to be a very common approach to increasing their size and it was an original idea of the experimenter rather than a report in the scientific literature that prompted its trial here. Since it requires no skill other than a steady hand it may be generally applicable.

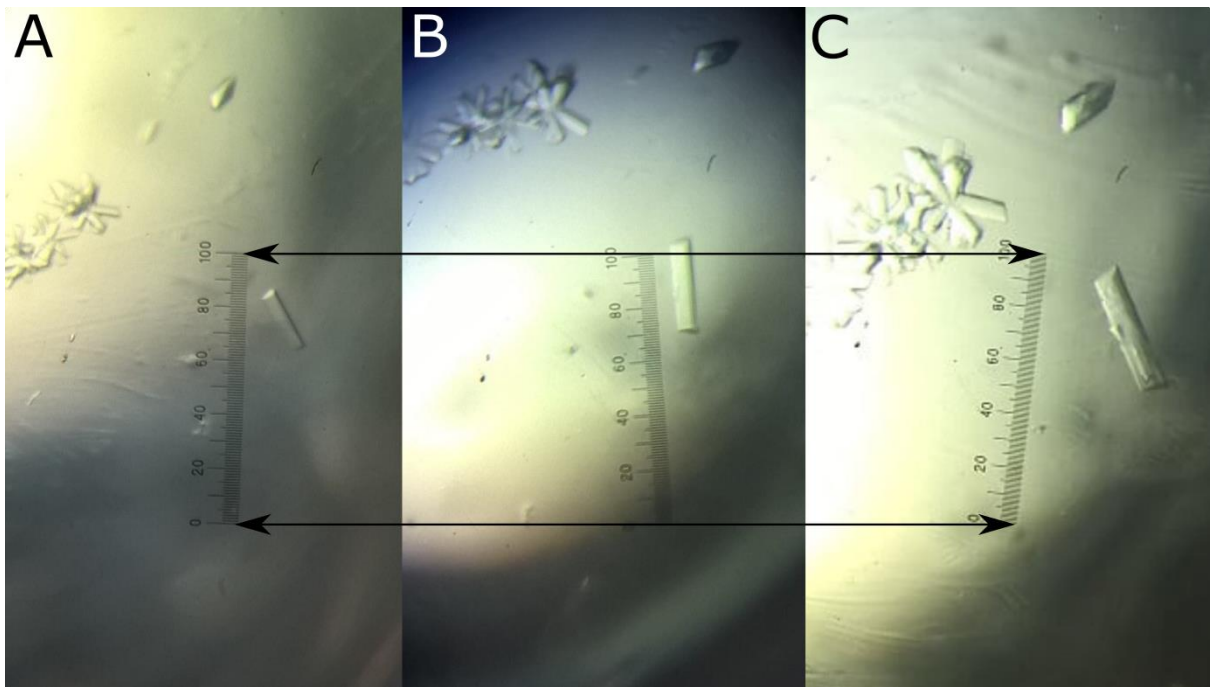


Figure 2.8. Results for *in situ* macroseeding of the AgrA_C199S/P2_{fragment} crystals. (A). The crystals at maximum size after streak seeding into a hanging drop. The streak line is visible on the left along with a well-defined single crystal lying off to one side. (B). The same crystal at maximum size after treatment with fresh crystallisation mother liquor. Enlargement along all axes is visible. (C). A second round of *in situ* macroseeding results in further enlargement and some twinning (not always the case). Significant enlargement is seen when (C) is compared directly to (A).



2.7.7: Protein-to-DNA Ratio

Having too much excess DNA present in the crystallisation solution was considered a possible reason as to why the crystals did not grow to large sizes. DNA molecules with no AgrA bound to them still had the capacity to H-bond to those that were bound, via their 5' A/T overhangs and so it was plausible that excess DNA was being included into the crystals and prematurely terminating their growth. To test this, the AgrA_C199S/P2_{fragment} complex was generated in parallel reactions containing AgrA_C199S and DNA at ratios of: 1.0:2.0, 1.0:1.5, 1.0:1.2 and 1.0:1.0. These were then used to set up 5 μ L hanging drop crystallisation experiments containing 2.5 μ L of both the AgrA_C199S/P2_{fragment} complex and crystallisation solution. Six crystallisation drops were prepared per ratio and these were then streak seeded into and monitored for crystal growth.

All AgrA_C199S:DNA ratios produced crystals after several days of incubation at 20 °C. Those resulting from AgrA_C199S/P2_{fragment} prepared at 1.0:1.2 AgrA_C199S:DNA produced slightly larger crystals and so this ratio was adopted for all subsequent optimisations.

2.7.8: Obtaining the Stoichiometric AgrA_C199S/P2_{fragment} Complex

This optimisation strategy follows the same arguments as that just discussed. One of the central dogmas of macromolecular crystallisation is that the mother liquor needs to be as homogenous as possible. Thus when attempting to co-crystallise a host macromolecule bound to a guest ligand, SEC is often used to remove excess ligand to achieve the aforementioned homogeneity since it will inevitably yield a more monodisperse sample mixture than if the researcher simply pipetted sufficient ligand into a solution of the host to get them to bind.

This does overlook certain scenarios, however. The K_d of a ligand is a ratio between K_{on} and K_{off} and unless the ligand is extremely non-labile some excess will be required to shift the equilibrium position to favour K_{on} . If a stoichiometric complex is prepared via SEC and then some of the ligand dissociates but the complex crystallises anyway then it was probably not worth performing SEC. If an excess of ligand is required to crystallise the sample however, then there is no point performing the SEC assuming it has already performed when initially purifying the host macromolecule.



The stoichiometric AgrA_C199S/P2_{fragment} complex was prepared by performing a second SEC step on the defrosted AgrA_C199S sample after incubation with 2 molar equivalents of the DNA. The sample was reconcentrated to 62.5 μ M and crystallisation experiments were performed as already described. No crystals were ever observed from the experiments involving the stoichiometric AgrA_C199S/P2_{fragment} complex prepared from SEC and so the strategy of simply titrating in 1.2 molar equivalents of DNA shortly before performing crystallisation experiments was maintained for all subsequent optimisations.

2.7.9: PEG Polymer Chain Length

Crystallisation experiments were set up in the 5 μ L hanging drop format using the optimised crystallisation containing: 175 mM NaF, mM bis-tris propane, 1.2 % w/v L-rhamnose and 21 % w/v PEG 3350. In addition PEG chain lengths 1500, 3500 and 4000 were also used in parallel to the original to test whether this had any bearing on crystal size. The experiments were steak seeded into and incubated at 20 °C. All PEG chain lengths produced crystals after several days. PEGs 3350, 3500 and 4000 produced crystals which were all equally large. PEGs 3500 and 4000 also produced a lot of large, robust crystals. These exhibited a diffraction pattern (not shown) consistent with those typically obtained for chemical crystals however, and so PEG 3350 was maintained for all subsequent optimisations.

2.7.10: Protein-to-Crystallisation Solution Ratio

Varying the ratio of crystallisation sample-to-crystallisation solution is a common approach to crystal optimisation. For the AgrA_C199S/P2_{fragment} complex 5 μ L hanging drop crystallisation experiments were prepared containing either: 2.0 and 3.0, 2.5 and 2.5 or 3.0 and 2.0 μ L of AgrA_C199S/P2_{fragment} sample and crystallisation solution respectively. Experiments were incubated at 20 °C and monitored for crystal growth. The ratio was not found to noticeably affect crystal growth since all crystals which grew in these experimental drops were approximately the same size and so the 1.0:1.0 ratio was maintained for all subsequent optimisations.



2.7.11: Addition of D-sorbitol to the Crystallisation Solution

It has already been mentioned that simple sugars and polyols significantly increased the size of crystals of AgrA_C199S/P2_{fragment} grown *de novo* (without seeds) and that this may be related to requirement of D-sorbitol in the bacterial growth media to produce soluble full-length AgrA. Neither of the two additive screens tested contained D-sorbitol itself as an additive however, so crystallisation experiments were established containing 3 % w/v D-sorbitol. Parallel experiments were performed containing: 3 % w/v sucrose, 3 % w/v trehalose, 3 % w/v glucose and a 1.5 % w/v L-rhamnose positive control, and negative control in which no sugar additives was also included.

These experiments were prepared in the 2 μ L sitting drop format and no seeds were administered. Once again L-rhamnose outperformed all of the other sugars although D-sorbitol did produce AgrA_C199S/P2_{fragment} crystals which were larger than those grown in the absence of sugar additives.

2.7.12: Greased Wells

To try to overcome the sticky nature of the crystals, a very small amount of silicone grease [Dow Corning] was rubbed into the wells of a 48 well sitting drop crystallisation plate using a Kimwipe™ [Cole-Parmer]. Crystallisation experiments were then set up as described. No crystals grew in wells which had been greased.

2.7.13: Blunt-End DNA

To check whether the complimentary 5' A/T overhangs in the DNA which afforded the AgrA_C199S/P2_{fragment} crystals was crucial for crystal formation, or rather amounted to unnecessary entropy in the sample, a “blunt-end” version of the oligoduplex was prepared from single stranded oligonucleotides that did not contain the 5' A/T overhangs. Crystallisation experiments were performed identically to what has been described already. Sitting drop experiments without the invocation of seeds were used to check for *de novo* crystal formation whilst streak seeding into hanging drops used to try to see if the blunt-end material would produce crystals from “sticky-end” seeds. All experiments failed absolutely to produce any crystals. As already mentioned, the high



failure rate in producing large single crystals mandated all crystallisation experiments to be performed with a high degree of experimental redundancy. Thus it was probably not due to random error that the blunt-end DNA-bound material failed to crystallise but rather the necessity of the 5' A/T overhangs.

2.7.14: Brominated DNA

Given that the protein/DNA complex in the AgrA_C199S/P2_{fragment} crystals contained 136 amino acids not present in the crystal structure of the C-terminus/DNA complex, phase information using the latter as search model for MR was unlikely to provide sufficient phase information alone to generate electron density maps of sufficient quality to build these remaining residues. Either a second search model that was structurally homologous to the remaining part of the structure or some *ab initio* phase information was presumed at the time to be required.

Conveniently, heavy atoms may be readily included into oligonucleotides during synthesis. This makes introduction of heavy atoms into protein/DNA complex crystals for SAD phasing experiments relatively trivial. Moreover, in order to obtain phase information from heavy atoms within the crystal, these atoms must be crystallographically ordered^[183] which they inevitably will be if they are covalently bound to oligonucleotides which are in turn bound to the sample protein.

This approach to obtaining *ab initio* phases was used to solve the structure of the C-terminus/DNA complex crystals.^[80] A thymine on each of the single stranded oligonucleotides was replaced with 5-bromouracil (5-BrdU) to introduce two bromine atoms per C-terminal domain. This was sufficient to generate enough anomalous signal to overcome the phase problem.

These two oligonucleotides: 5'-TTTAACAGTTAAG[5-BrdU]AT-3' and 5'-AA[5-BrdU]ACTTAACTGTTAA-3' were purchased from Sigma Aldrich and annealed to prepare a stock solution of brominated DNA. The brominated DNA had no observable effect upon crystallisation of full-length AgrA_C199S.

Diffraction from the resulting AgrA_C199S/P2_{fragment} crystals was never sufficient (< 6 Å) for any *ab initio* phase estimates to be obtained. The crystals required a high dose of X-rays to observe



reflections out to even comparatively poor resolutions. This often resulted in severe X-ray damage which further complicated data collection. Nonetheless it remains a promising and elegant strategy to return to if strongly-diffracting crystals of AgrA_C199S/P2_{fragment} crystals are grown in the future.

2.7.15.1: Mimicking Phosphorylation of AgrA_C199S in Pursuit of Better X-ray Diffraction

Bacterial response regulators undergo some structural modification after becoming phosphorylated which increases their binding affinity towards their cognate receptors.^[87,211] Solution-based NMR studies have demonstrated that response regulators can spontaneously sample the phosphorylated, or “on” state even in the absence of a phosphate ligand.^[223,224]

This phenomenon creates a less homogenous sample than if the response regulators adopted a single conformation in solution, and was considered to be amongst the possible reasons as to why the AgrA_C199S/P2_{fragment} crystals tended to be small and exhibit only poor X-ray diffraction.

Attempts were therefore made to lock AgrA_C199S in an “on” conformation in the pursuit of better quality crystals. Phosphoryl donors, such as lithium potassium acetyl phosphate and pyrophosphate, are intrinsically unstable in aqueous solutions owing to the very same chemistry that provides energy during cellular respiration. This lability of phosphorylating agents is an issue for crystallisation experiments since they usually take place over a long time frame. Thus invocation of some other, more stable orthophosphate mimetic was necessary in order to pursue this line of investigation.

2.7.15.2: Preparation of the D59E Point Mutant

For response regulators driven by phosphorylation of aspartic acid residues, it has been shown in many studies that mutating this residue to a glutamic acid effectively achieves permanent phosphorylation of the target by promoting the same structural rearrangement.^[90,225,226,227] In simple terms this mutation pushes the carboxylate group away from the surface of the protein by incorporating a single additional methylene group into the amino acid side chain. Since carboxylate groups and orthophosphate groups are not directly structurally analogous, this implies that a variety of negatively charged species could activate response regulators *in vitro* provided they are small, anionic



and present at sufficient concentrations. Although some dedicated study would be required in order to draw any definite conclusions upon this, it is worth bearing mind during the later discussion of the AgrA_C199S crystal structure since the crystallisation drop is comparatively rich in small anions (F^- and $H_2PO_4^{1-}$) which may bind to and activate AgrA_C199S. This in turn would mean that mimicking phosphorylation at an earlier stage in the purification was a redundant exercise.

The D59E point mutant was introduced to the pCOLD_AgrA_C199S construct via inverse PCR. The C199S_D59E double mutant expressed and purified identically to the C199S mutant inclusive of the observation that the double mutant was also monomeric on SEC meaning that, despite being forced into an “on” conformation, no dimerisation was observed *in vitro*. Here again it is worth noting that the D59E mutant is constitutively active *in vivo*. Crystals of the double mutant in complex with the P2_{fragment} DNA were grown by seeding AgrA_C199S/P2_{fragment} complex crystals into hanging drops. The resulting crystals diffracted no better than the single mutant crystals (between 4 – 5 Å) at which resolution even quite significant structural alterations would be impossible to distinguish anyway.

2.7.15.3: Beryllium Trifluoride as an Orthophosphate Mimetic

Beryllium trifluoride, $[BeF_3]^-$, is a small trigonal pyramidal anion which forms spontaneously in aqueous mixtures of Be^{2+} and F^- .^[228] It has been used successfully as an orthophosphate mimetic in numerous structural and kinetic studies owing to the stability of the macromolecular derivatives it produces, which are ionic rather than covalent.^[212,229,230] Thus it was used as another possible route to mimicking phosphorylation of AgrA_C199S for structural studies and for subsequent EMSA and SAXS experiments too.

There are numerous ways in which $[BeF_3]^-$ may be administered to a macromolecule during a crystallisation experiment, and even more so if the macromolecule has increased stability upon binding DNA. For example it is possible to: A) generate the $[BeF_3]^-$, add it to the macromolecule and then add the DNA. B) Add DNA to the macromolecule, generate the $[BeF_3]^-$ and then add it to the macromolecule/DNA complex (the reverse of what occurs *in vivo*). C) Generate $[BeF_3]^-$ and add it to the crystallisation solution. D) Add the F^- anions to the either the macromolecular sample or the



crystallisation solution then add the Be^{2+} cations to the solution which did not receive F^- and so allow $[\text{BeF}_3]^-$ to form during the crystallisation experiments. E) Soak the $[\text{BeF}_3]^-$ into existing crystals or F) add $[\text{BeF}_3]^-$ to the SEC running buffer.

For the AgrA_C199S/P2_{fragment} complex crystals all of these experimental designs were tried but option F) since it was anticipated that the $[\text{BeF}_3]^-$ would do more harm than good to the sample in a lot of scenarios, and gel filtration columns are expensive and communal. Option E) whilst it did not damage the crystals or affect their visual properties in any noticeable way did not yield any improvement in X-ray diffraction. Option D) produced excessive precipitate in the drop within a short time of establishing the crystallisation experiment. The rate of $[\text{BeF}_3]^-$ formation is probably slower in presence of high concentrations of extremely large PEG molecules than in one in which PEG is absent. Thus the Be^{2+} cations probably denature the protein before any beneficial chemistry takes place. Option C) still produced crystals, but these were typically very small so that benefits of the optimisations deduced hitherto was undone by the presence of $[\text{BeF}_3]^-$. Option A) immediately precipitated the sample and so it was impossible to proceed with crystallisation experiments.

Option B) stood out in that it allowed the $[\text{BeF}_3]^-$ to be administered without precipitation of the sample. It differs from option A) only in as much as, in the case of AgrA_C199S, the latter is already bound to and perhaps protected by the short piece of DNA. Given that without the DNA being first administered immediate sample precipitation upon administration of $[\text{BeF}_3]^-$ was immediately observed, it is postulated the $[\text{BeF}_3]^-$ anion does indeed bind to AgrA and that the DNA shields egregiously susceptible regions of the protein such that it is not denatured.

This is further evidenced by the fact that if the $[\text{BeF}_3]^-$ was added to the sample immediately after the addition of DNA, precipitation was observed. If ~ 20 mins was allowed to pass however before the addition of the $[\text{BeF}_3]^-$ no precipitation was observed, suggesting that sample denaturation is a function of DNA binding. Administration of $[\text{BeF}_3]^-$ to AgrA_C199S was at first performed in one hundred-fold molar excess, and later dropped to ten-fold excess for SAXS studies. Although crystals



did grow from beryllium trifluoridated sample, the corresponding X-ray diffraction was once again no better than when it had not been used.

The inability of $[\text{BeF}_3]^-$ to improve the diffraction limits of the crystals was disappointing. The exploration of this strategy did at least enable a method of administering $[\text{BeF}_3]^-$ to AgrA to be discovered. This may inevitably prove useful for downstream experiments in which phosphorylation of AgrA is a variable of the study, such as crystallisation of the N-terminal domain truncations of AgrA, NMR experiments, EMSAs, Isothermal Titration Calorimetry (ITC) and so on.

2.7.16: Streak Seeding into Iodine-Containing Drops

During the initial chemical grid screening, *de novo* growth was screened for in both NaF and NaBr. In the end the NaF-based condition was deemed best as it produced the largest crystals. In the meantime a number of strategies had been discovered to both bypass nucleation and grow relatively large if weakly-diffracting crystals of AgrA_C199S/P2_{fragment}. The benefit of including heavy atoms into a crystal in order to solve the phase problem has already been discussed.^[183,187] Another key benefit conferred by the inclusion of heavy atoms however is the fact that, since the degree of scattering of X-ray photons by an atom is proportional to the number of electrons “orbiting” the nucleus, and since a diffraction pattern is an interference pattern, increased scattering from one or more heavy atoms in a crystal lattice leads to increased reflection intensities for all of the observed reflections. In simple terms, the entire map is “brightened” by the inclusion of enough heavy atoms.^[231]

With these arguments in mind AgrA_C199S/P2_{fragment} crystals were generated via streak seeding into hanging drops containing 175 mM NaI instead of NaF whilst all other reagents in the crystallisation condition were unchanged. Interestingly the resulting crystals were **larger** than those which typically grew in NaF (fig. 2.9H). Some of the diffraction properties of these crystals were also different to those crystals grown in NaF. For example they were less sensitive to X-ray damage (as evidenced by more consistent numbers of reflections per diffraction frame after long exposures), had higher signal-to-noise ratios ($I/\sigma I$) and better innershell completeness. The resolution limit was not increased, however. All subsequent crystallisation experiments were therefore performed in the presence of NaI.



2.8: Crystallisation Optimisation Summary

Evidently a lot of different optimisation strategies were tested in the pursuit of well-diffracting crystals of full-length AgrA_C199S. For the benefit of the reader and people following on from this body of work, a brief summary of how large crystals of AgrA_C199S/P2_{fragment} crystals may be grown is presented below.

A batch of small crystals for microseeding can be prepared by incubating AgrA_C199S, purified as described in the methods section 7.6.7, at 62.5 μM with 75 μM of the aforementioned DNA for ~ 20 mins. If the sample has been defrosted 10 mM fresh BME or DTT, or 1 mM tris(2-carboxyethyl)phosphine hydrochloride (TCEP-HCl) should be added to ensure the sample remains reduced. The sample should then be mixed 1:1 with a solution of 100 mM bis-tris propane pH 6.5, 175 mM NaF, 21 % w/v PEG 3350 and 1.2 % w/v L-rhamnose to achieve a 2 μL experimental drop which should be incubated against an 80 μL reservoir in the sitting drop format at 20 °C until crystals appear. As many drops as is reasonable should be prepared. An entire 48 well plate should be sufficient to overcome the hit-and-miss nature of the crystallisation.

These crystals can then be microseeded via streak seeding into a 5 μL hanging experimental drop consisting of 2.5 μL of sample worked up as just described and 2.5 μL of a solution of 100 mM bis-tris propane pH 6.5, 175 mM NaI, 21 % w/v PEG 3350 and 1.2 % w/v L-rhamnose incubated against a 500 μL reservoir at 20 °C. Once again as many drops as possible should be prepared since the streak seeding adds yet another variable to the system.

Drops should be inspected for large single crystals which grew well away from the steak line or drops which contain a comparatively small number of crystals. These can then be macroseeded *in situ* by carefully removing 2-3 μL of mother liquor and replacing it with fresh and resealing the experiment. The crystals should carry on growing after a day or two and reach their new maximum after four or five days. A few rounds of macroseeding may be required. The crystals will eventually tend to sprout small crystalline appendages. In practise these have not had any adverse effects upon the diffraction



data but in any case can be avoided by opening and resealing the drop periodically until the X-ray diffraction experiment.

If the initial crystals fail to grow removal of the 1.2 % L-rhamnose may help since the experimenter has noted it suppresses nucleation in sitting drops resulting in fewer crystals. Sometimes this can be too effective and no crystals grow and this phenomenon varies drop-to-drop. If the bis-tris propane buffer stock has been purchased at pH 7.5 crystallisation experiments should be incubated at 10 °C. An assortment of large AgrA_C199S/P2_{fragment} crystals grown via these methods is compared to some of the initial hit crystals in figures 2.9A through H.

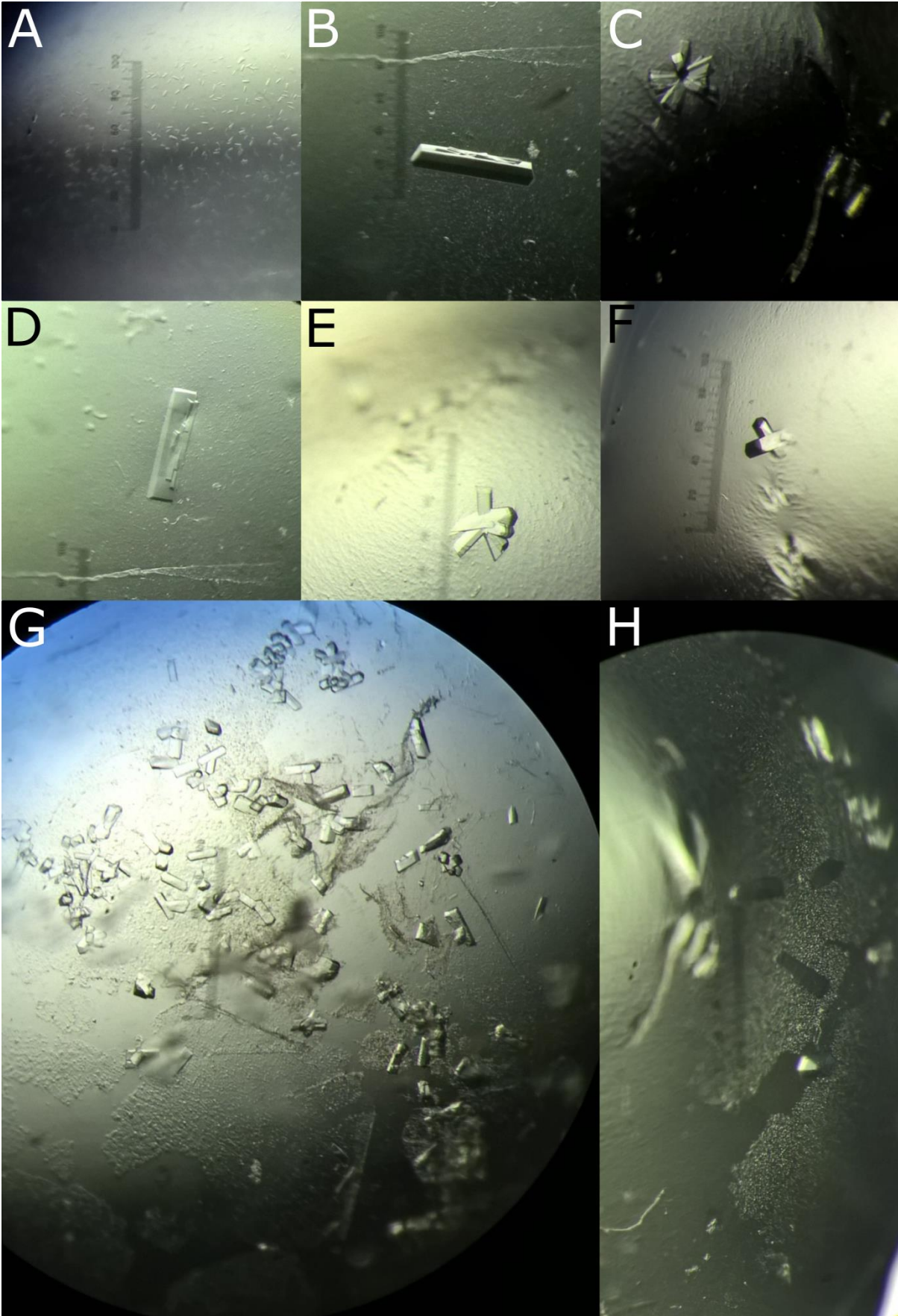




Figure 2.9 (previous page). Assorted optimised AgrA_C199S/P2_{fragment} crystals. (A). The initial hit crystals obtained in well F1 of the PACT Premier™ screen which are very small. (B). A large single crystal grown via the *in situ* macroseeding procedure described, that is orders of magnitude larger than those in (A). Note the non-blunt ends of the crystal. (C). A rod cluster also grown after delivering seed stocks to a hanging drop followed by *in situ* macroseeding. This clustering tended to occur when seed stocks were used in place of streak seeding as the way of delivering the microseeds to the hanging drop. (D). A similar crystal in proportions to (B) but it has blunt ends. (E). Similar to (C) but the point of origin of the cluster is hard to determine and it may have started life as a single crystal with twins sprouting from it later into the growth process. (F). A two-component twin with non-blunt ends and satisfying proportions. (G). AgrA_C199S/P2_{fragment} crystals grown by delivering seed stocks into a hanging drop followed by a round of *in situ* macroseeding. The NaF in the mother liquor has been replaced by NaBr. The islets of skin in which the crystals reside are visible owing to the necessary disturbance of the drop in order to perform the macroseeding steps. All of the crystals move together in these skins and are extremely difficult to manipulate. (H). The same as (G) but with NaI instead of NaF. The crystals are hard to distinguish from the experimental drop and appear as gaps in the precipitate. Despite the evidently successful optimisation procedures explored diffraction from all of these crystals remained comparatively poor.



2.9: Initial X-ray Diffraction Experiments on AgrA_C199S/P2_{fragment} Crystals

Crystals from the first successful crystallisation experiment were shipped to Diamond Light Source. Test exposures lead to the observation of Bragg peaks with intensity and spacing characteristic of macromolecular crystals (fig. 2.10), however, it was not possible to collect sufficient data to index the crystals owing both to their small size and weak diffraction.

Crystals grown from the various seeding methods were substantially larger allowing for larger wedges of data to be collected. The crystals indexed reliably in space group 23, I222, but failed to diffract beyond 5.5 Å. Furthermore the completeness of the diffraction at this resolution was typically around 70 % which is low.

Occasionally the crystals indexed in space group 5 C121, however, data from the same crystals were also successfully indexed in I222 by other data processing software pipelines. The completeness in C121 was substantially lower than in I222, at around 40 %. The space group C121 belongs to the monoclinic crystal system and is lower symmetry than I222 which could explain the lower completeness value. It is also possible that a calculated deviation of one of the lattice angles from 90 ° lead to the assignment of this space group in some instances. In any case, the diffraction data that indexed reliably in I222 were used for phasing and eventual structure solution without any further issues being encountered. A typical crystal had unit cell dimensions 60.28 x 98.34 x 198.88 Å which gives an estimated unit cell volume of 1,179,000 Å³.

The Mathew's coefficient is an estimate of the volume of space taken up by crystal contents (protein or solvent) in Å³ per unit relative mass (M_r) in g mol⁻¹.^[232,233] It will therefore take different values depending upon the packing density of a crystal. It may thus be used to probe the estimated number of copies of macromolecules within the unit cell and henceforth the asymmetric unit. These values may be compared to entries in the Protein Data Bank from which it is possible to infer the probability of $N = x$ where N is the probability of x copies of protein in the asymmetric unit. Mathew's coefficients can be readily calculated provided the space group and unit cell dimensions are known, along with the M_r of the sample.

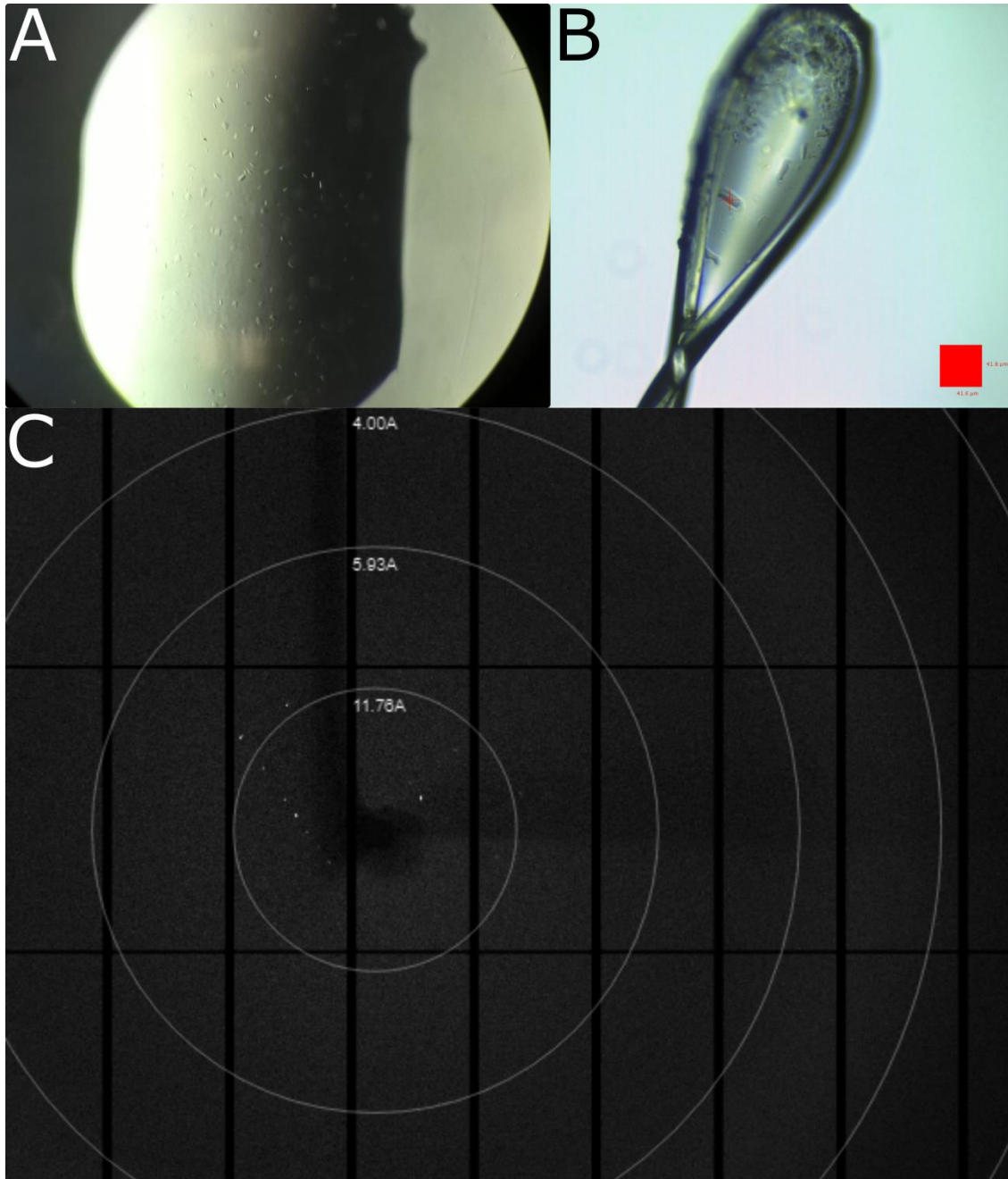


Figure 2.10. Diffraction data for the initial hit crystals. (A). The crystals which were nominated for diffraction screening. (B). A fished and cryoprotected crystal mounted on the goniometer at Diamond Light Source. The red box is for scale and has dimensions $41.6 \mu\text{m}^2$ thus the crystals appear to be $\sim 10 \times 20 \text{ \AA}$. (C). A diffraction image negative for the mounted crystal showing low-intensity, low-resolution Bragg peaks characteristic of a macromolecular crystal. Resolution rings are shown also.



For the AgrA_C199S/P2_{fragment} crystals the combined M_r of protein and nucleic acid is 39,300 g mol⁻¹. The unit cell volume has already been specified. Accordingly, the probabilities of one or two copies of the protein/DNA complex molecules residing in the asymmetric unit are 38.7 or 61.3 % respectively. The estimated solvent volume for these two scenarios is 67.17 or 34.34 % respectively.

The solvent volume for a typical macromolecular crystal is between 40 and 50 %.^[233] Judging from the relatively poor X-ray diffraction of the AgrA_C199S/P2_{fragment} crystals it was inferred at this stage that they probably did not contain two molecules per asymmetric unit since this would have required the crystals to be relatively densely packed with comparatively little solvent. Generally speaking crystals which contain less bulk solvent diffract better than those which contain more, and so the poor diffraction observed from the crystals in question would have been even harder to explain.^[234]

The most likely explanations for the poor diffraction at this stage were that the crystals contained a lot of disordered bulk solvent, were very small, and that the putative flexible linker between AgrA domains may have imparted some structural heterogeneity into the crystal lattice.

It was not until the *in situ* macroseeding and the addition of the L-rhamnose further improvements to the diffracting power of the crystals was observed and some traction was gained in solving the structure. For the first batch of crystals which were macroseeded *in situ* and subsequently sent to Diamond Light Source, L-rhamnose was added only to the “top up” mother liquor and was not present in the original crystallisation drop. This was simply because the crystals predated the discovery of the benefits of including L-rhamnose. This is an important point because it was obvious from the diffraction images that these crystals diffracted better from either of the long ends than in the middle. This suggests that inclusion of the L-rhamnose was only improving the packing of the crystalline material that formed after the *in situ* macroseeding. It is tempting to conclude from this that despite the fact that L-rhamnose can inevitably diffuse throughout the entire crystal it is only beneficial when it is present during the crystal growth phase.

The X-ray diffraction from these crystals was significantly better than those previously tested, extending out to 3.8 – 5 Å with acceptable signal-to-noise ratios and data redundancy. There remained



a major data pathology, however, concerning the innershell completeness. Even the lowest angle data had a completeness of only 60 – 70 %, dropping to 40 % for the highest angle data. This is unusual and unfortunate as the quality of every region of the electron density map is affected by the absence of any reflections (the maps and the diffraction pattern are two mathematical expressions of the crystal contents). It is beneficial therefore to be in possession of all of the reflection data that it is theoretically possible to observe.

The reason for the low completeness of the innershell data is hard to know for certain. Incorrect indexing of the data could cause it but there were instances where the diffraction data was > 90 % complete in I222 (albeit only extending out to ~ 6 Å). Too high doses of X-rays may cause adjacent reflections to assimilate into a single larger reflection on the detector, which may be treated as a single reflection by data processing software. Similarly a high degree of mosaicity^[235] or other deviations from ideal crystallinity can spread reflection intensities out over a greater area of the detector than would be the case for reflections from an ideal crystal. This too can cause adjacent reflections to overlap and be treated as a single reflection by data processing software.

It may be that a portion of the innershell reflections were simply too weak to observe. This is unlikely given that the crystals were receiving about 70 % of the total beam transmission available with exposures of 0.1 – 0.2 secs, oscillations per exposure of about 0.2 – 0.5 ° for 90 – 180 ° of data. These parameters represent the maximum dose of X-rays possible whilst still collecting a broad enough wedge of data to yield an adequate number of reflections for downstream processing since the crystals often deteriorated over the course of the experiment. So if it was the case that the missed innershell reflections were too weak to observe it was not feasible to compensate for this by relying on higher X-ray doses. It is more probable that the alignment of the crystal with respect to the X-ray beam rendered a number of reflections uncollectable. Further to this issue of low innershell completeness, inclusion of 175 mM NaI instead of 175 nM NaF in the crystallisation mother liquor did appear to alleviate the issue somewhat and give innershell completeness values closer to 100 %. An I anion has 54 electrons compared to 10 electrons in a F anion so it will diffract X-rays much more strongly.^[173] Whether this was the cause of the observed increase in innershell completeness is hard to say without



either a much higher resolution dataset or sufficient crystals grown from NaI and NaF to be tested for a statistically significant comparison to be made. Crystals grown in 175 mM NaI did not, however, diffract to a higher resolution than those grown in 175 mM NaF as already stated.

2.10.1: Cryoprotection of the AgrA_C199S/P2_{fragment} Crystals

Most diffraction data are collected at 100 K to mitigate damage to the crystals caused by the high energy X-ray photons^[186] and to reduce dynamic disorder.^[236] Ice is crystalline water and will diffract X-ray photons strongly, however, resulting in unusable diffraction images.^[237] Consequently some anti-freezing agent needs to be introduced to the crystal before it can be cryocooled.^[238] It has been well documented that the specific choice of anti-freeze agent can lead to marginal gains in diffraction resolution.^[239,240] This fact was not overlooked during the crystal optimisation process.

Here **the low solubility of NaF** posed another problem, as it was impossible to dilute all the components into one another with final concentrations that matched the crystallisation condition whilst leaving room for an anti-freeze agent. The latter typically takes up 20 – 40 % v/v of the total cryoprotectant solution. To overcome this, most of the reagents were diluted into one another to achieve final concentrations which were half of the concentrations at which they existed in the PACT Premier™ screen. This was because upon establishing a crystallisation experiment in which sample is mixed with crystallisation solutions at 1:1 ratios the reagents are diluted two-fold anyway. This strategy worked and the following cryoprotectant solution was used: 87.5 mM NaF, 50 mM bis-tris propane pH 6.5, 17.5 % w/v PEG 3350, 0.6 and w/v L-rhamnose and 40 % v/v anti-freezing agent.

Several anti-freezing agents were tried including: glycerol, ethylene glycol, D-trehalose and PEG 400. If crystals were grown in different halide salts then they replaced NaF in the cryoprotectant solution. The only anti-freezing agent which totally abolished ice formation was glycerol. The propensity of ice formation in the crystals is probably related to their high solvent content. A comparison of the original hit crystals and an optimised crystal mounted on the goniometer at Diamond Light Source is presented in figure 2.11. A representative diffraction frame from the optimised crystal is also presented.

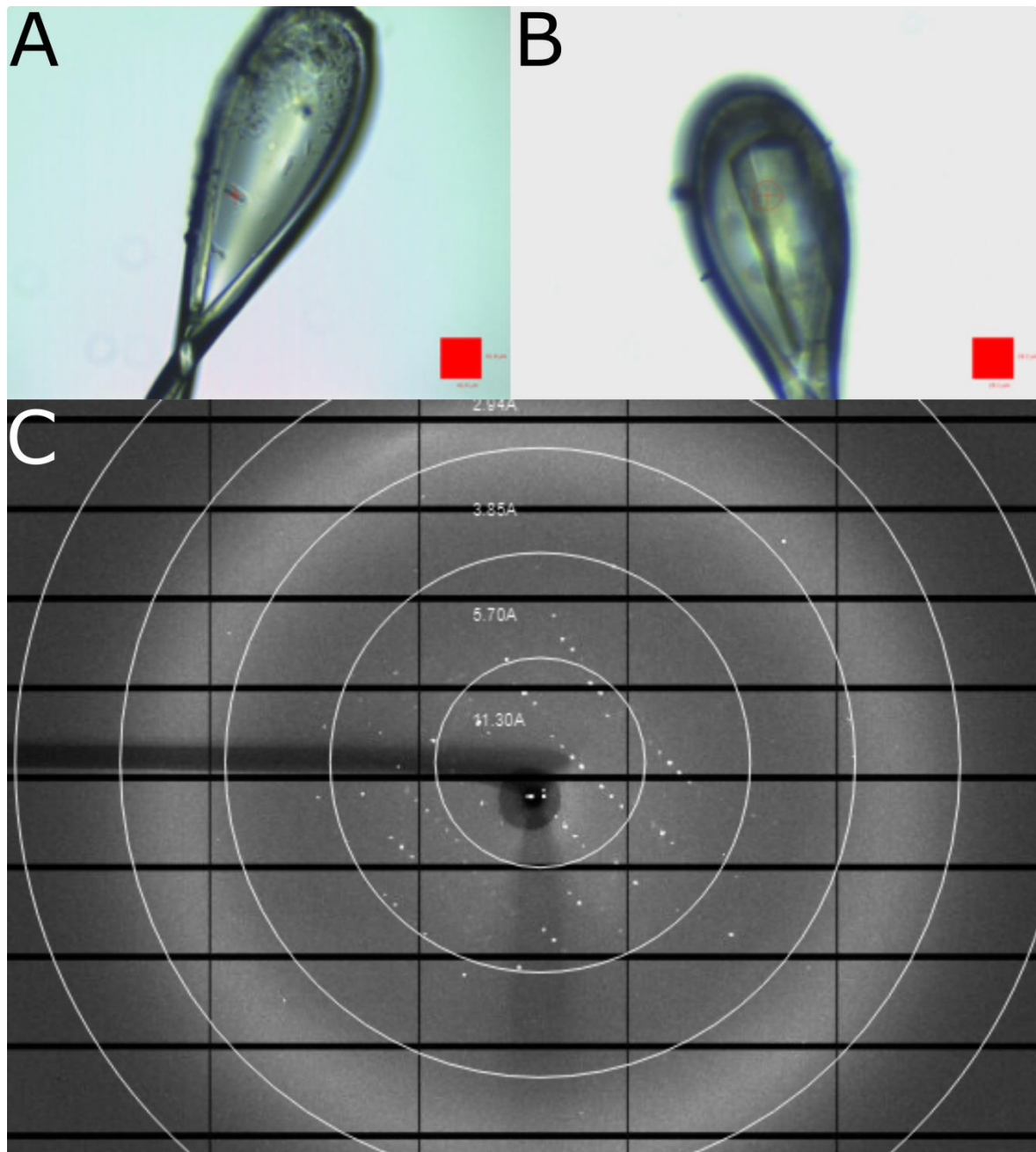


Figure 2.11. Comparison of the original (A) and optimised (B) AgrA_C199S/P2_{fragment} crystals mounted on the goniometer at Diamond Light Source. The red box is for scale and has dimensions 41.6 μm^2 . (C). A diffraction frame corresponding to crystal (B) exhibiting macromolecular diffraction with approximately circular and discrete Bragg peaks. The lack of ice damage indicates the efficacy of the cryoprotectant solution also. Note how Bragg peaks are seldom observed beyond 3.85 Å.



2.11: Phasing the AgrA_C199S/P2_{fragment} Crystals

Initial phases for the AgrA_C199S/P2_{fragment} crystals were obtained via iterative MR steps using PHASER.^[182] (vers. 2.7) [CCP4]. The first MR step used the C-terminal domain in complex with DNA (PDB: 3bs1) as the search model.^[80] Water molecules, Mg²⁺ cations and five amino acids corresponding to the putative flexible linker, KDNSV, were removed. These five residues correspond to residues 136 – 140 in the AgrA sequence and are located between the N and C-terminal domains.

Placement of this domain was possible even with the data which was only complete to ~ 60 %. A unique MR solution with one copy of the search model per asymmetric unit was found in I222. Whenever the search model was used to obtain phase estimates for poorer datasets the same solution was also found. The so-called Log Likelihood Gain (LLG), which is a comparative measure of how well the experimental diffraction data can be predicted from an atomic model over a random distribution of those same atoms,^[241] was approximately 60. This is indicative of a probable solution^[241] notwithstanding the fact that the search model only contained roughly two thirds of the total number of atoms in the protein/DNA complex.

Manual inspection of the molecular replacement solution reaffirmed its probability as the correct solution since there were no egregious overlaps between adjacent C-terminal domain or DNA molecules. Furthermore these were spaced such that they were not entirely close packed with ample space for the remaining N-terminal domain units to form the remaining crystal contacts. Unmodelled electron density betrayed the location of the N-terminal domains, however, it was of too poor quality to perform any model building *de novo*. Packing plots for this partial molecular replacement solution are shown in figure 2.12.

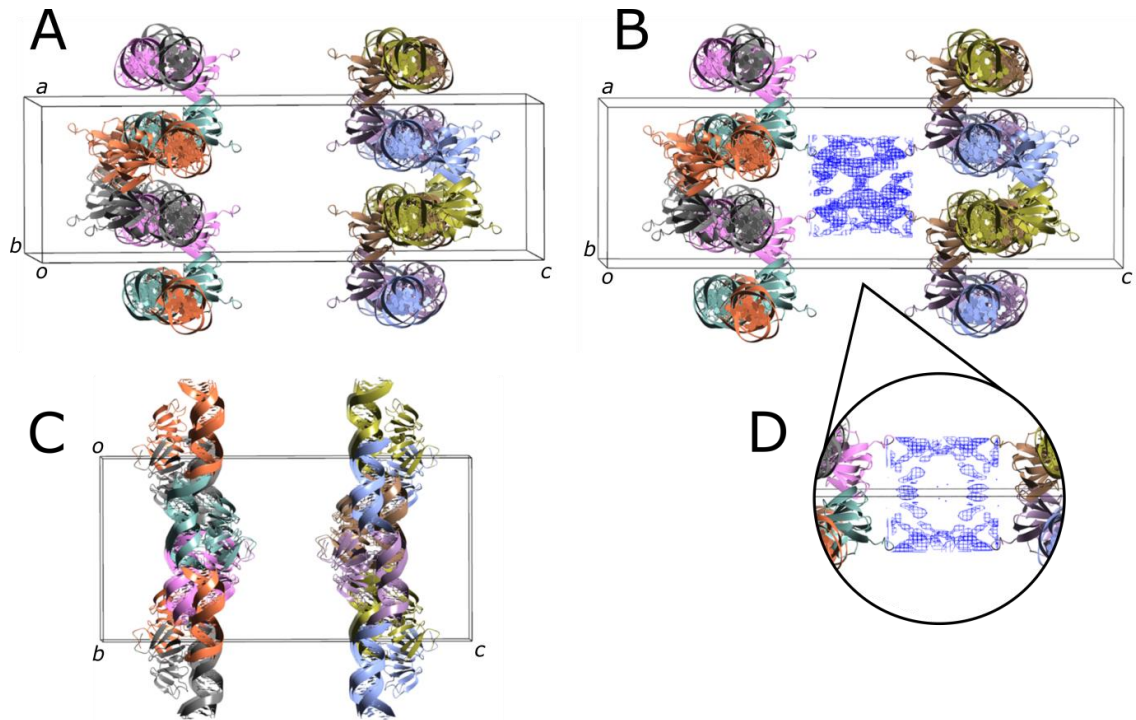


Figure 2.12. Various packing plots of the AgrA_C199S/P2_{fragment} crystals after molecular replacement using the C-terminal domain (PDB: 3bs1). (A) and (B). Looking along the crystallographic *b* axis. The C-terminal domains bound to the DNA molecules are aligned either pointing towards the unit cell vertexes or its geometric centre as per I222 symmetry. The unmodelled electron density in (B) is conspicuous and betrays the location of the unmodelled N-terminal domain. (C). Looking along the crystallographic *a* axis. Adjacent DNA molecules form a Hoogsteen base pair (not drawn) via their respective 5' A/T overhangs to form a pseudo-continuous double helix running parallel to the crystallographic *b* axis. (D). A zoom-in showing what appeared at this stage to be a large solvent cavity adjacent to the unmodelled electron density. Electron density is contoured at 1.5 σ . The box corresponds to a single unit cell. The linker residues KDNSV and Mg²⁺ cations were left in the search model here to help convey the symmetry, but were deleted before further phasing attempts. Only molecules which have at least one atom intersecting the unit cell are drawn hence the absence of some C-terminal domains. In the crystal structure all of the DNA molecules have a corresponding AgrA_C199S C-terminal domain bound to them.



In order to attempt to build the remaining portion of AgrA_C199S, a homology model for residues 1 – 135 (AgrA₁₋₁₃₅) was generated using the Phyre2 server.^[242] The linker residues KDNSV were left out owing to their putative unstructured nature. The position of these residues was probably not going to be accurately encapsulated in the homology model and so likely reduce its ability to be positioned accurately during MR. In addition to the homology model it was considered that some extra coordinate data from solved AgrA homologues might help, or even be necessary to correctly build the remaining unit.

To obtain possible search models to aid solving the AgrA₁₋₁₃₅ portion of the structure, amino acid sequences similar to that of AgrA₁₋₁₃₅ were searched for in the PDB using the BLAST server.^[243] Four results were returned: the phosphate-receiving domain of ComE_D58E^[90] (PDB: 4mld) and ComE_D58A^[90] (PDB: 4ml3), full-length ComE^[90] (PDB: 4cbv), all from *Streptococcus pneumoniae*, and the phosphate-receiving domain of LytR^[88] (PDB: 6m8o) from *S. aureus*. These had sequence identities of 33.33, 33.33, 47.27 and 33.64 % to AgrA₁₋₁₃₅ respectively. These homologues, the homology fold for AgrA₁₋₁₃₅ and a superposition of AgrA₁₋₁₃₅ with the phosphate-receiving domain of LytR are shown in figure 2.13.

Since the C-terminal domain of AgrA_C199S in complex with DNA had already been positioned, it seemed sensible to first preform a second MR step using the coordinates from the first MR step as the starting point for the second one. It is worth stating explicitly that positioning of the C-terminal domain in complex with DNA before attempting to position the N-terminal domain homologue was observed to be critical as to whether or not the outcome was successful.

When the linker residues KDNSV were left on the C-terminal domain during MR, positioning of any of the candidate search models of the N-terminal domain failed as evidenced by PHASER “thrashing” and producing dozens of low LLG score (0 – 30) results. Positioning of the homology fold of AgrA₁₋₁₃₅ also failed. Hoping that it was the linker residues that were responsible for the failure of the second MR step, they were removed and the process was repeated. All failed apart from LytR, which was positioned to return an LLG of approximately 120. Interestingly the relative position of the



homologues was nearly always the same irrespective of whether the process was successful or a failure as judged by the LLG values, however, the relative orientation of the domains was different. This is perhaps because these phosphate-receiving domains are all globular in shape and lack any significant protrusions.

It should also be noted that the MR steps hitherto discussed were performed on several datasets collected at Diamond Light Source but the steps subsequently discussed were performed using the single best dataset available at the time. This had a completeness of approximately 60 % and a maximum resolution of 3.6 Å. (Note that resolution is taken here to mean the literal maximum angle out to which Bragg peaks were observed after some cut-off criteria was applied; in reality the resolution of the electron density maps would be significantly lower owing to the low completeness). Other data metrics for this dataset are shown in table 2.1 at the end of this chapter. Note that this is not the data that was eventually used to build and refine the structure.

Successful placement of the LytR domain lead to an initial chimeric structure solution containing the C-terminal domain of AgrA_C199S bound to P2_{fragment} DNA, and the N-terminal domain of LytR. The problem thus became one of trying to “convert” the LytR N-terminal domain sequence to that of the N-terminal domain of AgrA_C199S whilst maintaining the apparent validity of this initial solution.

The homology fold AgrA₁₋₁₃₅ was superposed over the N-terminal domain of LytR to check their similarity at the atomic level. The Root Mean Square Deviation (RMSD) between all of the atoms in the two domains found to be 2.0 Å which implies the two units have a very similar structure, not forgetting of course that the homology model of AgrA₁₋₁₃₅ is indirectly based on experimental data from structural homologues to begin with. This apparent similarity inspired a strategy wherein the homology model of AgrA₁₋₁₃₅ was superposed over the LytR unit positioned in the second MR step followed by subsequent deletion of this LytR “template” molecule to leave just amino acids corresponding only to full-length AgrA_C199S, minus the five amino acid linker, in (hopefully) accurate enough positions to serve as a starting point for further model building and refinement.

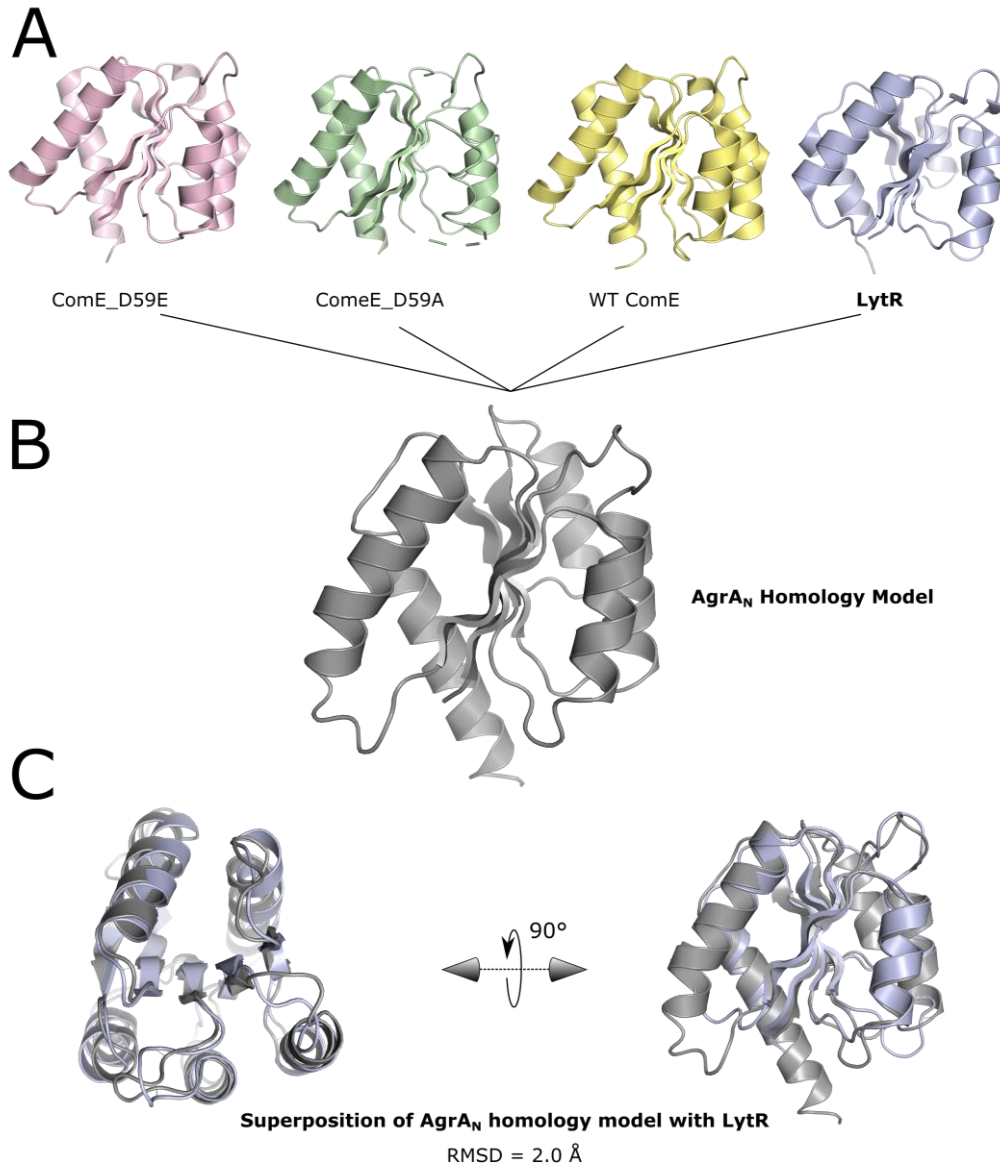


Figure 2.13. Homologues of, and the homology model generated for, AgrA₁₋₁₃₅. (A). Left-to-right. The N-terminal domain of ComE_D58E (pink; PDB: 4mld), the N-terminal domain of ComE_D58A (green; PDB: 4ml3), N-terminal domain of ComE abstracted from the full-length structure solution (yellow; PDB: 4cbv) and the N-terminal domain of LytR (blue; PDB: 6m8o). (B). The homology fold for AgrA₁₋₁₃₅ generated using the Phyre2 server (grey). (C). Superposition of the N-terminal domain of LytR and the AgrA₁₋₁₃₅ homology model viewed from two different angles. Despite the clear structural homology the sequence identity is low, at 30 – 40 %.



These steps were carried out and another MR step was performed using the resulting coordinates on the merged reflection data to see if PHASER could position this LytR-derived, full-length AgrA_C199S/DNA complex unit. When the superposition was performed using PyMOL™^[244] (vers. 2.2.0) [Schrödinger Inc]^[244] or COOT^[196] (vers. 0.8.9) [CCP4] the MR step failed, however, if the superposition was performed using Chimera^[245] (vers. 1.12.2rc) [UCSF] the MR step succeeded, to yield an LLG of 257. Ten cycles of refinement with REFMAC^[193] resulted in an R and R_{free} of 28 and 35 % respectively. This divergence between R and R_{free} topped out at 7 % and did not increase with further refinement cycles implying that model bias was not detrimentally egregious at this stage and that the refinement was stable despite poor experimental diffraction data.

Inspection of the electron density maps suggested that this initial solution was probably correct since it concurred with key features of atomic model although there were of course quite large peaks and holes visible in the difference density maps. Samples of these maps are shown in figure 2.14.

The observed dependence upon which program was used to perform the superposition of AgrA₁₋₁₃₅ over LytR probably reflects the accuracy of the atomic coordinates needed for it to be an efficacious search model for MR and reaffirms the point that every option should be explored for difficult macromolecular crystallographic cases.

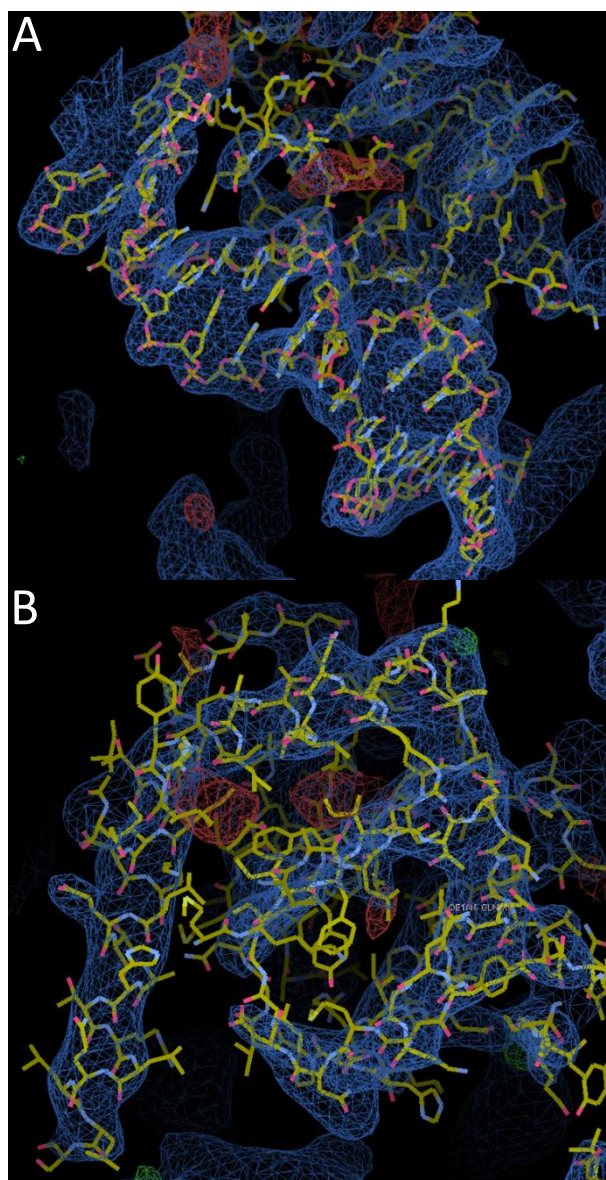


Figure 2.14. Electron density maps after the LytR unit had been displaced by the homology fold of AgrA₁₋₁₃₅, the MR step repeated and 10 cycles of refinement in REFMAC performed. (A). The electron density around the DNA, although very low in resolution, is relatively clear along the sugar-phosphate backbone. The electron density around the DNA-binding domain of AgrA_C199S, which is mostly β form, is substantially poorer. (B). Electron density around the N-terminal domain, which is mostly α form and consequently a lot clearer. Again the reader is reminded that the data is only $\sim 60\%$ complete and directed to table 2.1 at the end of this chapter. Electron density is shown as a blue mesh and is contoured at 1.5σ . Screenshots are taken from COOT.



Inspection of this initial structure solution revealed that the asymmetric unit contained N- and C-terminal domains from two different protomers in the crystal. The evidence for this was the distance which would have to be spanned by the linker residues KDNSV if the two domains in the asymmetric unit belonged to a single protomer. This distance was 32.5 Å if measured between the C α atoms of residues 135 and 141, and is too far for five amino acids to span.^[246] Moreover the respective breaks in the polypeptide chain were aligned pointing away one another, implying that these five residues would have had to reverse their direction and looped back across the face the N-terminal domain which is highly implausible. Inspection of the packing of this initial structure solution revealed a symmetry copy of the N-terminal domain close by with its C-terminal break pointing towards the N-terminal break of the C-terminal domain. The distance between the respective C α atoms for these two domains was 12.5 Å which is a much more plausible distance for five amino acids to span.^[246]

Although modelling of crystallographic structures is performed on their asymmetric unit, for large structures such as macromolecules it is intuitive to model units that correspond to entire discrete polypeptide chains. This sometimes requires reassignment of which portions of which symmetry copies of the protomers in the crystal comprise the asymmetric unit. The total number and nature of the symmetrically-unique portion of atoms cannot change, however.

This reassignment was performed manually in PyMOL™ by generating atoms for the N-terminal domain at the desired symmetrically-equivalent site and deleting those corresponding to the original copy of the N-terminal domain. These coordinates were saved and MR step performed, again using the same merged reflection data. A comparison of the asymmetric units before and after this reassignment is shown in figure 2.15. Reassignment of asymmetric unit symmetry contents lead to improvement in LLG from the MR step, returning a value of 430. This is noteworthy since from a technical standpoint the asymmetric unit contents have not changed and the two units that have been swapped are crystallographically identical. Also noteworthy is the slight drop in R_{free} to 33 %, once this solution was subject to ten rounds of refinement in REFMAC which implies that it is also a better atomic model for the experimental data at this early stage.



In any case the resulting asymmetric unit contained a single N-terminal domain comprised of residues 1 – 135, a single C-terminal domain comprised of residues 141 – 238 with the C199S point mutant, and a single 15 bp oligoduplex with 5' A/T overhangs as shown in figure 2.15D. This model was by no means complete and contained substantial errors including: a significant packing clash between adjacent N-terminal domains, a broken/discontinuous sugar-phosphate DNA backbone, 5-bromouracil instead of thymine at certain DNA bases and generally poor polypeptide and side chain geometry. At the time of experimenting it was decided to pursue better quality crystals and crystal data before investing significant time into building a satisfactory model, however. This was because some useful insights could be gleaned from the structure even in its current state and higher quality crystal data generally speaking makes the entire structure solution through deposition and publication processes easier. It was also borne in mind that the initial model presented here may be used as starting point if better data was forthcoming. A packing plot for this initial solution is presented in figure 2.16A.

Several LLG values have been presented in this section to demonstrate the efficacy of the convoluted phasing process. It is not valid to compare LLG values from MR solutions performed using different datasets.^[182,241] Nor is it correct to use LLG values to judge the quality of a model which has been built into electron density by some other method after the initial MR step. This is because, in the former case, resulting models are being judged by the accuracy with which they predict the experimental data, which will differ between data collections. That is to say LLG is an internal metric with respect to each dataset. In the latter case, atoms manually built, correctly or not, into electron density will always give a better-than-random fit to the experimental data when placed there by PHASER, and so LLGs will tend to increase even if the structure solution is wrong. That is to say PHASER cannot independently judge the accuracy of phase estimates obtained by refinement. To this end it is worth saying that the LLG values quoted in this section were all obtained from positioning of whole clusters of atoms, the coordinates for which were not obtained from refinement against the experimental data. Furthermore all LLG values quoted were all obtained from the same dataset. Refinements that were performed were done so simply to obtain R and R_{free} values from parallel MR steps to enable cursory comparisons to be made.

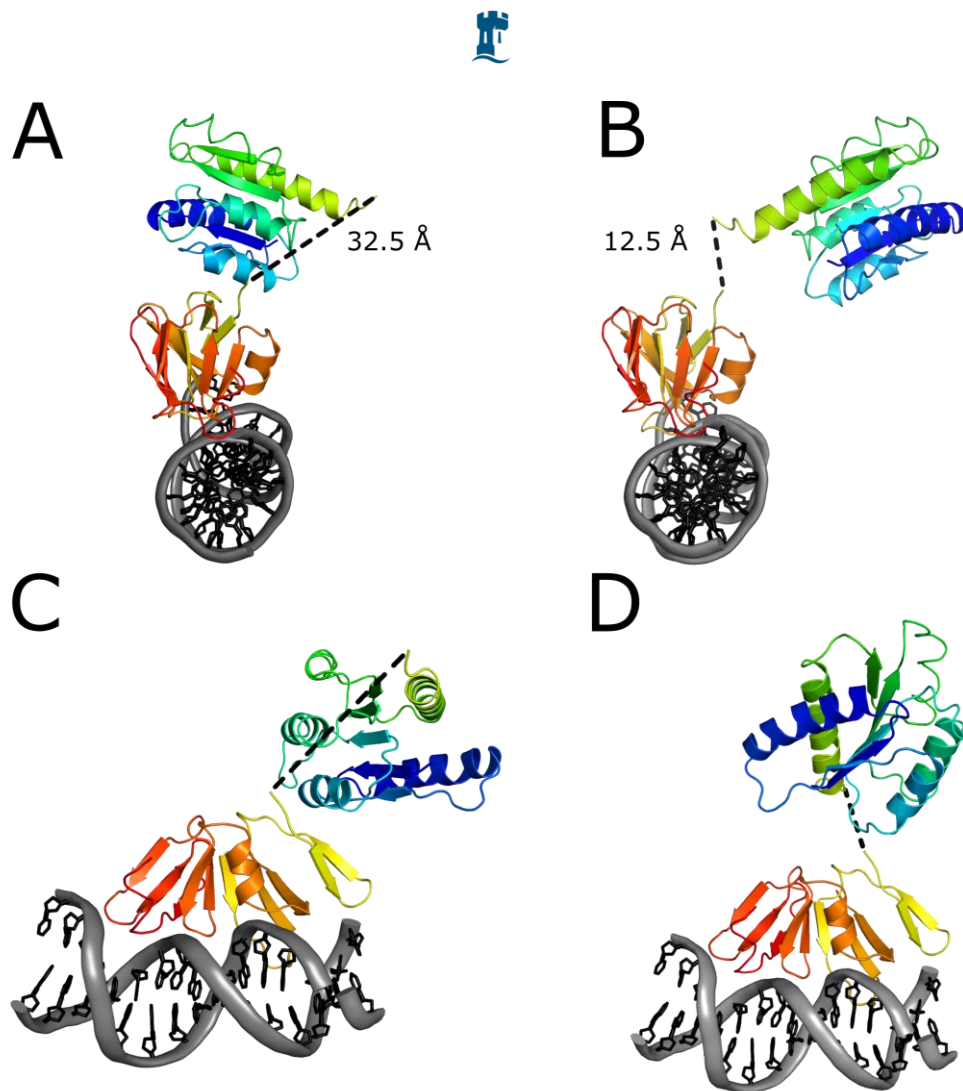


Figure 2.15. The AgrA_C199S/P2_{fragment} crystal asymmetric unit contents before and after reassignment. The asymmetric contains a single AgrA_C199S molecule and a single P2_{fragment} DNA molecule. (A). Original asymmetric unit viewed along the DNA helical axis. (B). the reassigned asymmetric unit viewed along the DNA helical axis. (C). The original asymmetric unit viewed face-on with respect to the DNA helical axis. (D). The reassigned asymmetric unit contents viewed face-on with respect to the DNA helical axis. Essentially, the N-terminal domain has been rotated 180 ° about a crystallographic two-fold axis. The protein is coloured as a rainbow with the N-terminus shown as blue and the C-terminus shown as red. The break in the polypeptide chain corresponding to residues KDNSV is shown as a black dashed line and has Cα···Cα distances of 32.5 and 12.5 Å for the original ((A) and (C)) and reassigned ((B) and (D)) asymmetric units respectively.

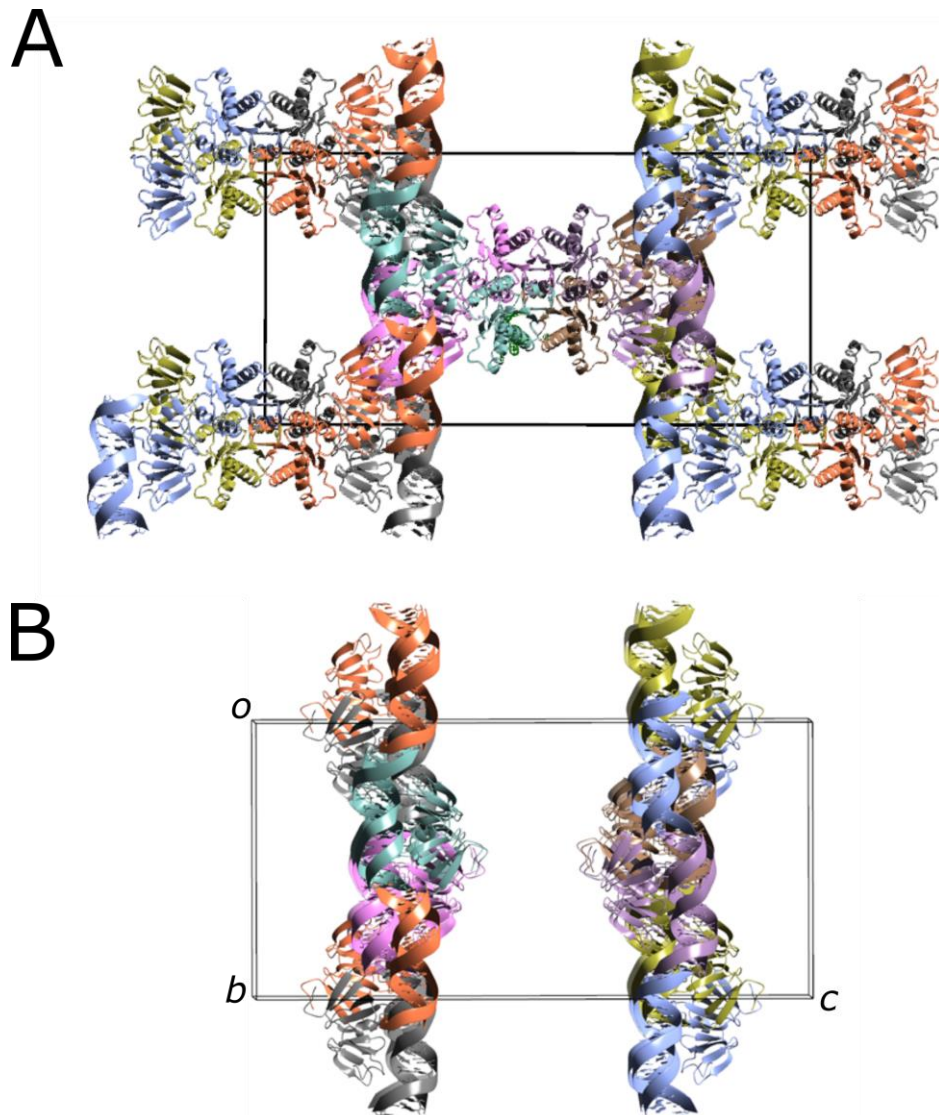


Figure 2.16. Packing plot of the initial AgrA_C199S/P2_{fragment} model after all steps discussed in this section were performed. (A). The I222 cell is, apart from solvent molecules etc., fully populated with proteinaceous and nucleic acid material. Large solvent cavities, somewhat exaggerated by the cartoon display, are observed above and below each lattice point. These are located at the cell vertexes and its geometric centre. The view is along the crystallographic *a* axis. (B). The previously-shown packing plot for the C-terminal domain only in the same orientation as (A) for reference. The same colour scheme is retained for both images.



Parameter	Value
Data Collection	
Wavelength, (Å)	0.9687
Space group	I 2 2 2
Cell dimensions, (Å)	
a, b ,c (Å)	59.83, 97.21, 195.25
α , β , γ (°)	90.00, 90.00, 90.00
Beam size, (μm)	50 x 50
Transmission, (%)	69.62
Exposure, (s)	0.2
Oscillation, (°)	0.2
Number of Images	450
Max. Resolution, (Å)	3.59
Total no. reflections	10574
No. unique reflections	2629
Number of reflections	10574
Redundancy	4.0 (3.17)
Spherical completeness, (%)	60.3 (47)
Mean I/ σ (I)	6.23 (1.41)
R _{merge}	0.088 (1.08)
R _{pim}	0.050 (0.656)
CC _{1/2}	0.993 (0.249)

Table 2.1. Summary of key diffraction data used to build the crude initial model of the AgrA_C199S/P2_{fragment} complex discussed in section 2.11. Values in parenthesis are for the highest resolution shell. Data was collected at Diamond Light Source beamline I04 and processed with autoPROC+STARANISO.^[247] Note the low spherical completeness. Generally completeness tapers off gradually with scattering angle and most of the low-resolution reflections are collected to 100 % completeness, however, this was not so for the AgrA_C199S/P2_{fragment} crystals.



Chapter 3: AgrA – Low-Resolution Structure Solution

3.1: Crystal Structure of Full-Length AgrA_C199S in Complex with DNA

The method for cryocooling the crystals was described in section 2.10.1. Crystal data was collected at Diamond Light Source beamline I04. Crystal data was integrated using DIALS^[248] (vers. 1.14.2), indexed using POINTLESS^[249] (vers. 1.11.19) and scaled and merged using AIMLESS^[199] (0.7.4) as part of the CCP4 suite^[250] (vers. 7.0.072). An anisotropic resolution cut-off was applied to the data using the STARANISO server^[251] (vers. 3.315) with the following cut-off criteria: $R_{\text{pim}} \leq 0.6$, $I/\sigma(I) \geq 2.0$ or $CC_{1/2} \geq 0.3$. Initial crystallographic phases were obtained via MR using PHASER^[182] with the search model described in section 2.11. Manual building was performed in COOT^[196] (vers. 0.8.9) and reciprocal space refinement was performed using PHENIX^[197] (vers. 1.17.1) with secondary structure restraints for the N-terminal domain, group B-factors and optimised X-ray/stereochemistry weights enabled. The atomic model refined to an R and R_{free} 30 and 36 % with acceptable model geometry given the quality of the crystal data and the high solvent content of 67 %. The Mathew's coefficient was $3.69 \text{ \AA}^3 \text{ Da}^{-1}$. A sequence alignment of *S. aureus* AgrA to a variety of AgrA homologues is presented in figure 3.17 along with diffraction data in table 3.1 at the end of this chapter and a provisional PDB validation report is presented in appendix III.

3.2: Major Fixes to the Initial Model

After obtaining the initial model there remained significant errors within it. These included: A) a broken sugar-phosphate backbone in the DNA molecule, B) two 5-bromouracil nucleobase derivatives occupying the positions of two thymines in the DNA molecule and C) complete catenation of the polypeptide backbone between adjacent symmetry copies of the N-terminal domain in the crystal. This was not only incorrect in itself but led to numerous severe packing clashes. Without high resolution electron density maps available it was difficult to build experimentally-justified fixes for many of these issues, particularly the peptide catenation. Fixes were built nonetheless with due attention paid to plausibility and ideality as well as the usual metrics which are used to judge model quality. Before and after images of these fixes are presented in figure 3.1.

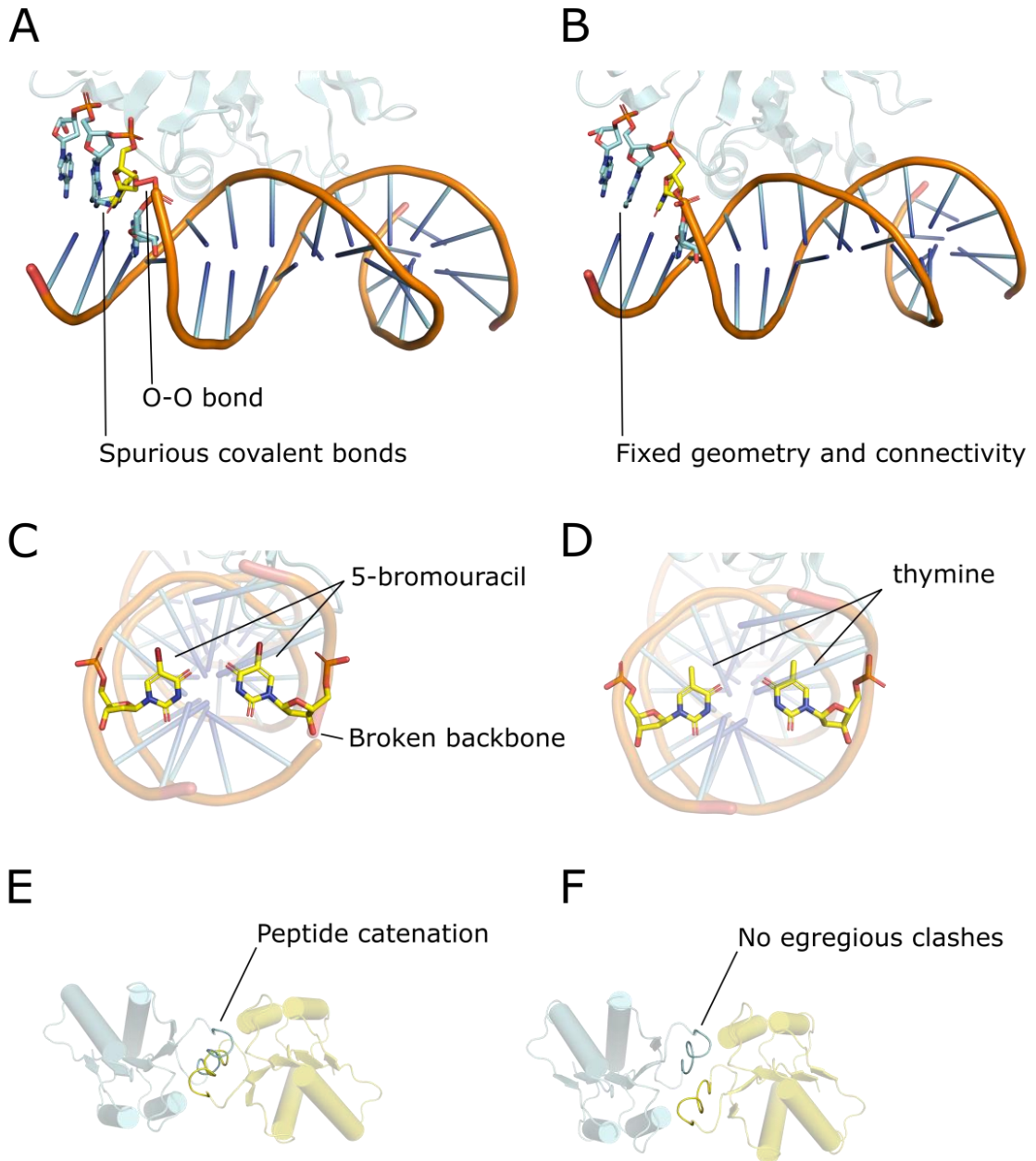


Figure 3.1. Major fixes to the initial model. (A). The broken sugar-phosphate backbone is displayed. A spurious O-O bond is visible along with other incorrect covalent bonds. (B). The sugar-phosphate backbone and corresponding nucleobases are fixed and have appropriate bonding and geometry. (C). The 5-bromouracil derivative which is not present in the oligoduplex used to crystallise full-length AgrA_C199S. The broken sugar-phosphate backbone is also visible. (D). The 5-bromouracils have been replaced by thymines in the atomic model. (E). Catenation of the polypeptide backbone in the region of $\alpha 4$ in the initial atomic model. Two symmetry copies of the N-terminal domain of AgrA are coloured in cyan and yellow. (F). The modified atomic model exhibiting no catenation.



3.3: Overall Relationship Between the Two Domains

AgrA_C199S folds into two discrete domains as expected. Residues 1 – 135 comprise the N-terminal phosphate-receiving domain whilst residues 141 – 238 comprise the DNA-binding domain. Residues 136 – 140, which have the amino acid sequence KDNSV, form a putatively flexible linker for which only poor electron density was observed and so were not modelled. Additionally, the 19 non-AgrA_C199S residues at the very N-terminus of the expression construct which comprise the (His)₆ affinity tag, factor Xa protease site and the pCOLD translation enhancing element were unobserved in the electron density maps and were also not modelled. There is a 95 ° bend between the two domains, imparting distinct J-shape to the protein as a whole when viewed along the helical axis of the DNA molecules. The two domains are not packed against one another with the distances of closest approach being 12.9 Å between the peptide N atoms of Ser140 and Tyr156, and 12.5 Å between the carbonyl O atom of Asp52 and the peptide N atom of Lys192. All of these amino acids reside in loop regions. Several views of the asymmetric unit contents are presented overleaf in figure 3.2.

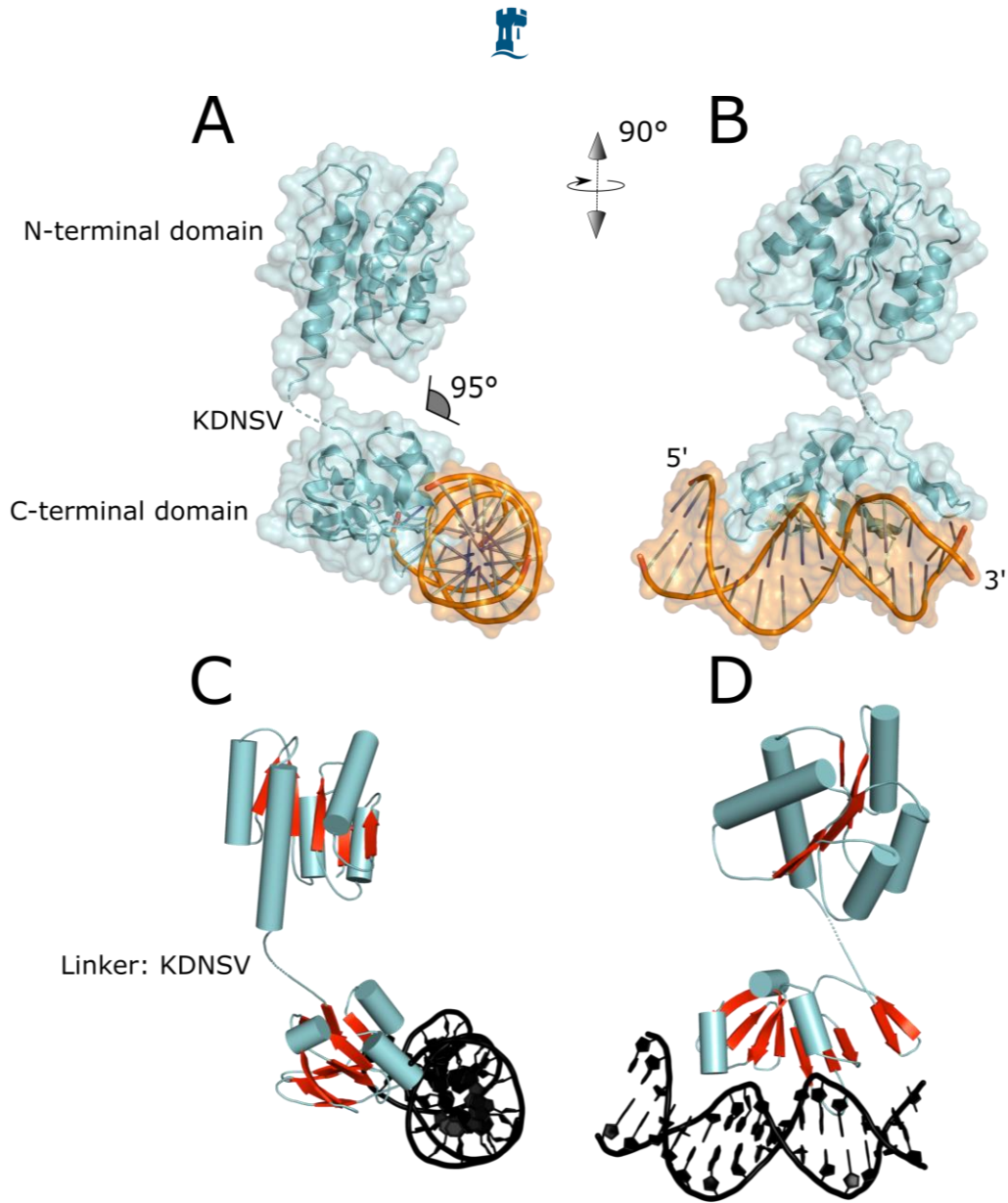


Figure 3.2. The asymmetric unit. (A). A single AgrA_C199S protomer in complex with DNA, which comprise a the asymmetric unit, is shown as a transparent surface. The protein is colored cyan and the DNA is colored orange with the domains labelled accordingly. (B). The same as (A) but rotated vertically through 90° so as the DNA helical axis is viewed face-on. (C). A simplified cartoon view of the asymmetric unit contents in the same orientation as in (A). Helices and loops are colored cyan, β -strands are coloured red and DNA is coloured black. (D). The same colour scheme as (C) but in the same orientation as (B). The missing residues KDNSV are represented as a dashed cyan line.



3.4: Overall Features of the N-terminal Domain

The N-terminal domain adopts the flavodoxin fold, in which a five-stranded parallel β -sheet is surrounded by five α -helices. Amino acids constituting the β -strands are as follows: 2 – 4 (β 1), 32 – 36 (β 2), 55 – 59 (β 3), 85 – 88 (β 4) and 106 – 109 (β 5). Amino acids constituting the α -helices are: 10 – 24 (α 1), 41 – 48 (α 2), 68 – 77 (α 3), 93 – 100 (α 4) and 115 – 132 (α 5). The β -sheet is slightly curved towards α 1 and α 5 which both lie to one side of it whilst α 2, α 3 and α 4 lie on the other side resulting in an α - β - α sandwich. The amino acids 63 – 66 form a β -turn in the loop that connects β 3 to α 3. The analogous region in CheY, that regulates cell motility in *E. coli* and has a sequence identity of 18.4 % to the N-terminal domain of AgrA, forms a tighter γ -turn.^[86] This reversal of the peptide backbone forms a “flap” near the active site and is a common feature of some response regulators.^[90,252] Several views of the N-terminal domain are presented in figure 3.3. Helix five extends for approximately two turns below the plane of the domain when viewed side-on. This imparts a “mushroom” shape to the protein surface. This feature is not present in the structural homologues LytR,^[88] that regulates cell autolysis in *S. aureus*, or CheY^[86] owing to all of their helices being of more similar length. Helix 4 appears to be distorted away from ideal helical geometry despite being directly involved in crystal packing.

Attempts to build it with ideal helical geometry resulted in large negative peaks in proximal areas of the difference density maps, in addition to returning a higher R_{free} value. It was the α 4 region which was completely catenated with a symmetry copy of itself after MR. Consequently, no secondary structure restraints were imposed upon it during refinement and the region was built such that no egregious packing clashes were introduced whilst also lowering R_{free} . Although the deviation away from an ideal α -helix could be entirely down to the poor quality of the crystal data, a bent α 4 is observed in the structure of full-length ComE, ComE_D58E and ComE_D58A (PDB: 4cbv, 4mld and 4ml3 respectively) which are all reported to be representative of the phosphorylated or “on” form.^[90]

This protein regulates cell competence in *Streptococcus pneumoniae* and has a sequence identity of 30 % relative to AgrA. A full discussion on the difficulty in modelling α 4 is presented in section 3.13.3.

For sake of simplicity however, the region is referred to as α 4 for the rest of this discussion.

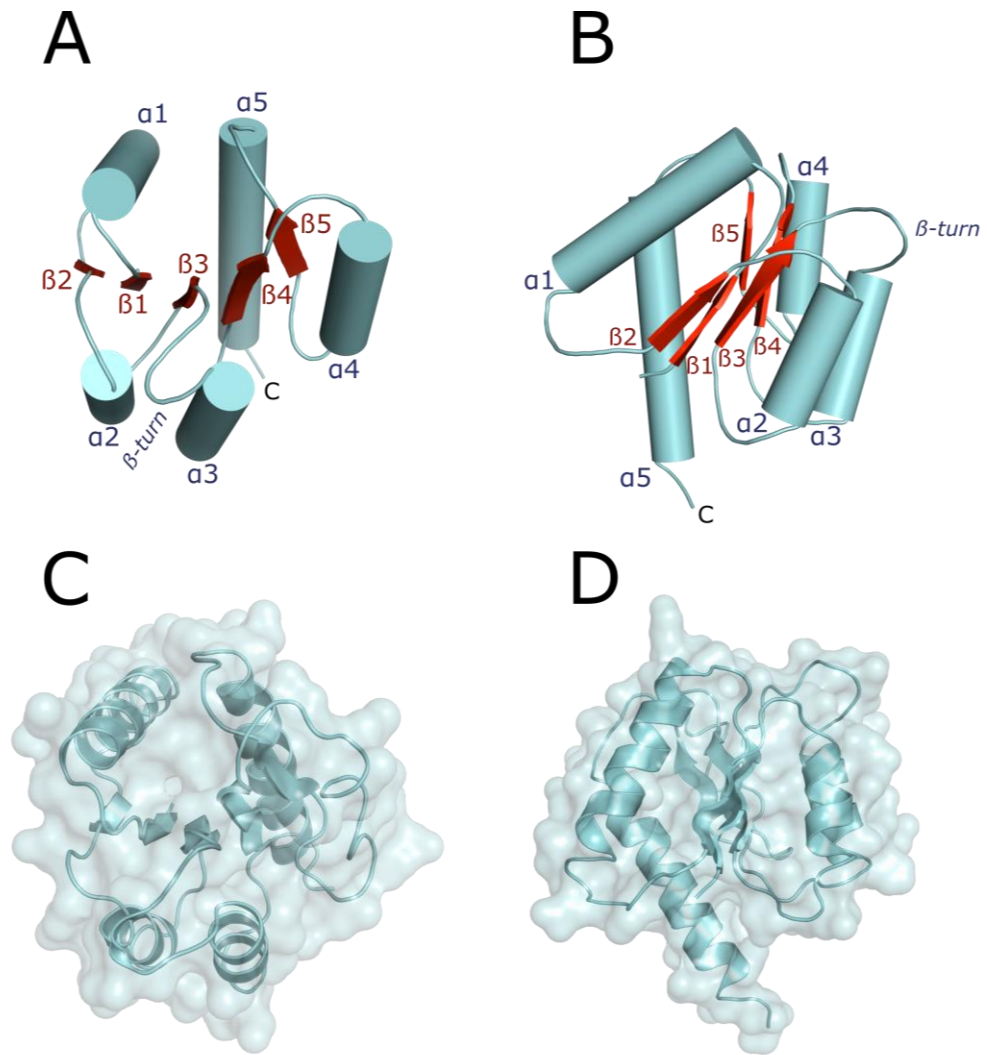


Figure 3.3. The N-terminal domain of AgrA_C199S. (A). The N-terminal domain of AgrA_C199S (residues 1 – 135) is shown as a simplified cartoon with helices and loops coloured cyan, and β -strands coloured red. The view is top-down with respect to the central β -sheet. (B). The same as (A) but viewed rotated horizontally through 90° so as to be viewed edge-on with respect to the central β -sheet. (C) and (D). The same two perspectives as (A) and (B) respectively but displayed as transparent surfaces with additional details of the polypeptide path shown.



3.5: Overall Features of the C-terminal Domain

High-resolution crystal structures of the C-terminal of domain of AgrA, **spanning residues 136 – 238**, in complex with DNA are available^[80,85] and so a detailed breakdown of that region of the structure is not duplicated here. No broad differences are observed between it and the full-length structure solution presented in this work. This is unsurprising given that the same oligoduplex was used in the pursuit of crystallogensis^[80] and any finer differences would be impossible to infer anyway given the low resolution of the diffraction data.

It is remarkable, however, that a crystal packing motif is common to both structures^[80] given that the structure presented here contains an additional 135 AgrA amino acids and a further 19 unbuilt amino acids imparted by the expression construct. Specifically, the 5' A/T overhangs on neighbouring DNA molecules in the crystal form a Hoogsteen base pair between them. Hoogsteen base pairing occurs when a purine nucleobase is rotated through 180 ° about the glycosidic bond to present the N atom at position 7 as the H-bond acceptor.^[253] In this arrangement the two nucleotides are arranged to one another, relative to Watson-Crick base pairing.

In both structures this motif effectively mitigates the curvature introduced into the DNA by the C-terminal domain of AgrA by reversing the direction of the curvature. This allows adjacent DNA molecules to run as a pseudo-continuous double helix parallel to the crystallographic *b* axis rather than curve back on itself. The Hoogsteen base pair involved in this packing motif is presented in figure 3.4.

In as much as it is possible to infer from a low-resolution structure the Hoogsteen base pair in the full-length structure solution appears slightly more strained than in the high-resolution structure of just the C-terminal domain in complex with DNA.^[80] This is inferred from the fact that preceding nucleotides do not overlap exactly when the two structures are superposed. This may be due to the necessity of accommodating the additional N-terminal domain in the full-length structure whilst retaining the Hoogsteen base pair packing motif. It may also be due to refinement against low-resolution data however, and in any case the Hoogsteen base pair is not of functional significance.

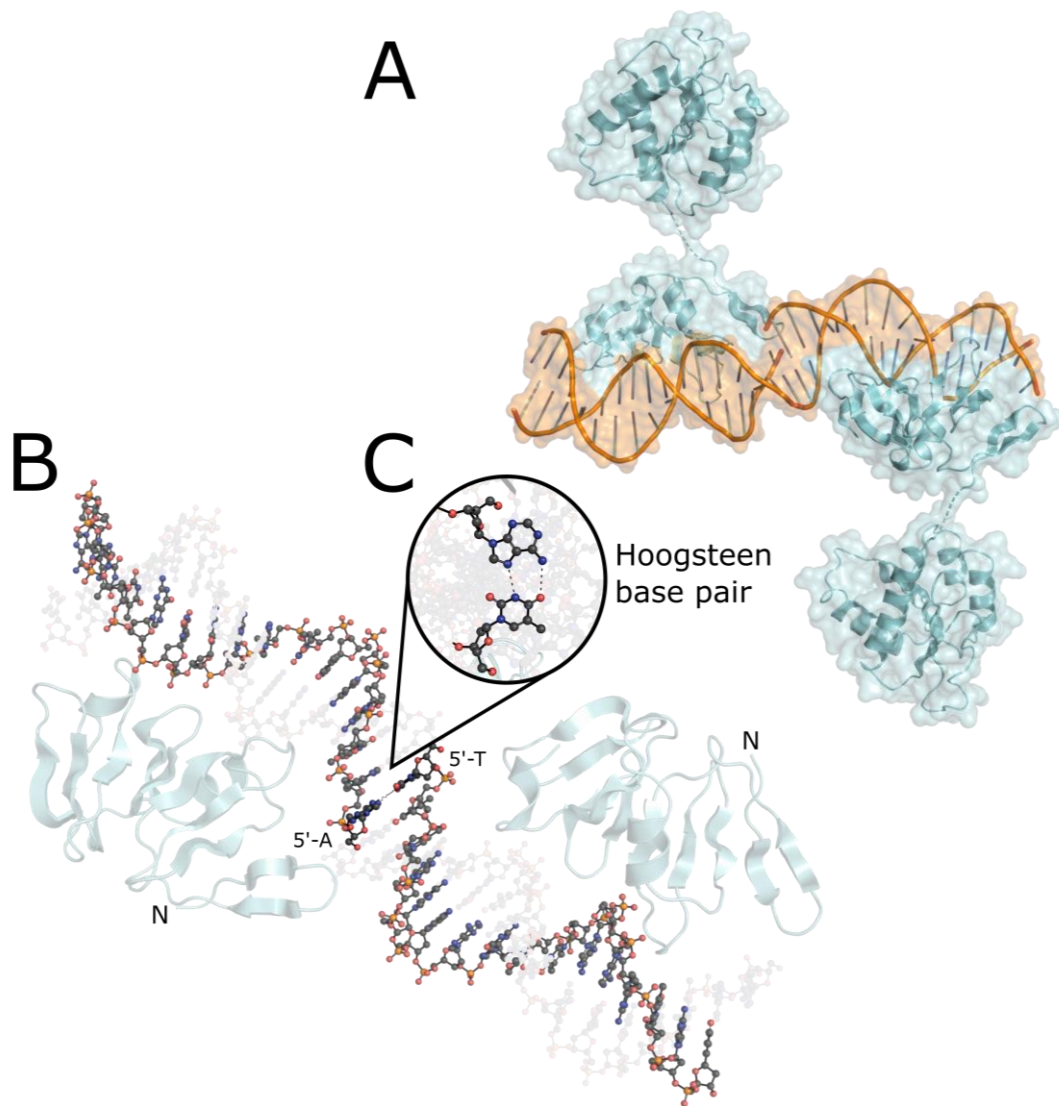


Figure 3.4. Representations of the Hoogsteen base pair that contributes to the crystal lattice of the AgrA_C199S/P2_{fragment} crystals. (A). A transparent surface of two adjacent fully-populated asymmetric units with the protein coloured cyan and the DNA orange. (B). Two asymmetric units again, but with the N-terminal domains removed for clarity. The DNA is shown as balls and sticks in this view. The two strands that form the Hoogsteen base pair are opaque and the complimentary strands are greyed out. (C). A zoom-in of the Hoogsteen pair showing the necessary 180 ° rotation around the adenine glycosidic bond to present N7 to the opposing thymine. The reversal of the DNA curvature is apparent and the motif repeats essentially infinitely, parallel to the crystallographic *b* axis.



3.6: The Phosphate-Binding Region

The functionally-requisite residue Asp59, which becomes phosphorylated by AgrC, is located at the top of $\beta 3$ and forms an acidic pocket along with Glu7, Asp8 and Asp9 which lie in the loop region between $\beta 1$ and $\alpha 1$. The residues Gln61 and Lys110, which all lie in the loop regions between $\beta 3$ and $\alpha 3$, and $\beta 5$ and $\alpha 5$ respectively, are also proximal. A diagram of this acidic pocket is presented in figure 3.5. The acidic region surrounding the phosphate-receiving aspartic acid is highly conserved amongst bacterial response regulators (see fig. 3.17) despite the presumed specificity required to ensure phosphorylation is performed by the cognate histidine kinase(s).^[87,211]

The proximity of Lys110 to the active site is not happenstance and the putative mechanism of response regulator phosphorylation invokes the amine functional group on the lysine side chain to electrostatically balance the abundant negative charge in the phosphorylation site.^[254] It has also been proposed that it may act as a proton donor to the phosphate anion during dephosphorylation also.^[254] In the high-resolution crystal structure of CheY (PDB: 3chy) a *cis* peptide bond occurs between the analogous lysine, Lys109, and Pro110 in order to present the Lys109 side chain to the active site.^[254] *Cis* peptide bonds are much rarer than *trans* peptide bonds owing to their higher energy.^[255] *S. aureus* appears to have overcome this energetic requirement to presenting the side chain of Lys110 to the phosphorylation site of AgrA by through the evolution of two intervening aspartic acid residues, Asp111 and Asp112, between Lys110 and Pro113 thus allowing the polypeptide backbone to twist in a more gradual fashion.

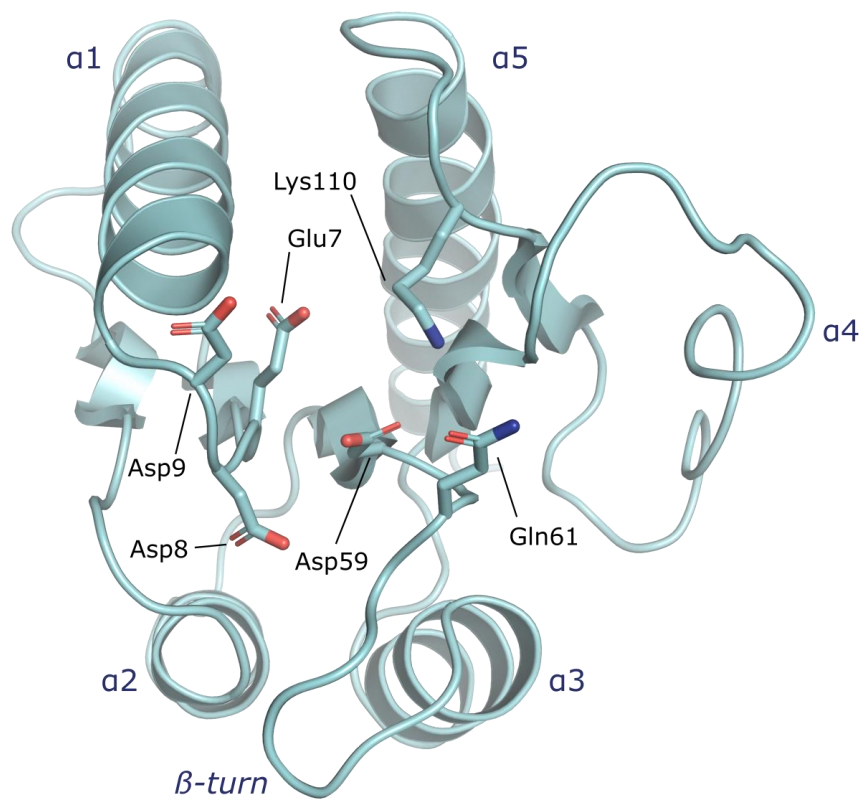


Figure 3.5. The phosphate-binding region of the N-terminal domain is displayed with relevant residues represented as sticks and labelled in black. The residue that becomes phosphorylated is Asp59 which resides at the centre of an acid pocket. The five α -helices and the β -turn are labelled in blue and α 4 is distorted away from ideal geometry.



3.7: Cysteines 55 and 123 Are Close to Each Other but Not to Cysteine 199

AgrA is known to regulate genes which are not on the *agr* operon.^[98] One salient example is the *BsaA* gene which encodes glutathione peroxidase that catalyses the reduction of H₂O₂ in order to mitigate oxidative stress.^[208] It has been demonstrated that when *S. aureus* cells are placed under oxidative stress, transcription of the *agr* operon, specifically transcription of RNAlII, is downregulated but transcription of *BsaA* is upregulated by AgrA.^[208] This oxidative checkpoint to quorum sensing has been linked to the cysteine pair Cys199 and Cys228 in the DNA-binding domain of AgrA. In the same study it was shown that the C199S point mutant cannot upregulate *BsaA* and so renders the cells much more susceptible to oxidation.

At the time of publication no experimental structure of full-length AgrA was available to the authors,^[208] with only the C-terminal domain in complex with DNA having been solved.^[80] In order to address the question of whether there were other cysteine pairs present in AgrA that may also play a role in redox-dependant QS regulation, the authors could only rely on a homology model for the remaining N-terminal domain of AgrA. The model presented suggests^[208] that the three cysteines in the N-terminal domain: Cys6, Cys55 and Cys123 are too far apart to form a disulphide bond. The experimental structure presented here contradicts this, however.

The residues Cys55 and Cys123 are spatially-proximal, with the distance of closest approach being 5.9 Å, occurring between the S atom on Cys55 and the Cβ atom of Cys123. These residues have been modelled as fully reduced with the Cys123 thiol pointing away from that of Cys55. The cysteine···cysteine distance was measured between the S and Cβ atoms because it is invariant with side chain rotation of any one of the cysteines, and the precise location of side chains cannot be accurately determined in low-resolution electron density maps. The S···S distance would be shorter than 5.9 Å if both thiols were pointing towards each other, however. **The spatial arrangement of these cysteine residues along with the Cys199 and Cys228 pair is shown in figure 3.6.**

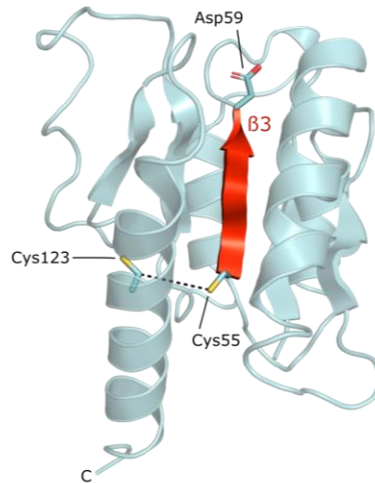
Both Asp59 and Cys55 reside on the same secondary structural unit. In particular, Asp59 is located at the carboxyl end of β3 whilst Cys55 is located at the amino end. Given that disulphide bond formation



or abolition usually confers some functional perturbation, either by subtly altering protein conformation or by completely destroying a given fold or interaction^[256] it is intriguing that Cys55 and Cys123 (and Asp59) are close together. This is redoubled given the known role of AgrA in redox-dependant modulation of QS.^[208] Furthermore both Cys55 and Cys123 are completely conserved in staphylococcal AgrA homologues (fig. 3.17). Mutations are observed in AgrA homologues belonging to more distant bacterial relatives, however. No partner cysteine is observed for Cys6 in the crystal structure which is striking given that it is even more highly conserved than Cys55 and Cys123, being present in the sequences of non-staphylococcal AgrA homologues also. The residues immediately subsequent to Cys6, Glu7 and Asp8, are equally well conserved as Cys6 itself.



A



B

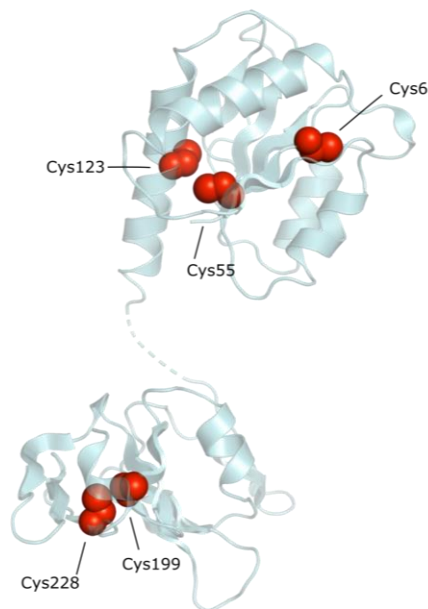


Figure 3.6. Relative locations of the five cysteine residues in AgrA. (A). The proximity of Cys55 and Cys123 is displayed with the distance between respective the S and C β atoms, of length 5.9 Å, shown as a black dashed line. The central β 3 strand connects Cys55 to the functionally-requisite Asp59 residue, and is coloured in red. The side chains of Cys55, Asp59 and Cys123 are also shown. (B). All five cysteines are shown as red spheres (Cys199 is mutated to Ser in the crystal structure) to highlight the two cysteine pairs, Cys199 and Cys228, and Cys55 and Cys123. Two pairs reside on separate domains (N-terminal and C-terminal respectively) and are mutually distant from the unpaired Cys6.



3.8: Helix 4 is Hydrophobic

The accepted paradigm in bacterial response regulator dimerisation is that the phosphate-receiving domains form a symmetric dimer with two-fold rotational symmetry whilst the DNA-binding domains arrange line astern in order to bind to the pair of linear DNA repeats.^[87,211] The flexible linker must therefore adopt a compact conformation in one of the protomers and a relatively extended conformation in the other to facilitate this asymmetry. A crystal structure of the full-length response regulator KdpE, that regulates expression of ATPases and has a sequence identity of 25 % relative to AgrA, bound to its cognate receptor as a dimer^[91] has already been shown (fig. 1.4) and exhibits this arrangement of domains. Furthermore many crystal structures of phosphorylated response regulator phosphate-receiving (usually N-terminal) domains are available.^[90,225,226,227] Through this abundance of structural data it has become apparent that the two-fold symmetric dimers that they form upon phosphorylation all invoke $\alpha 4$ as part of the structural unit through which they dimerise.^[211] There is diversity even amongst these $\alpha 4$ -mediated dimers however, with different secondary structure elements being invoked to form the complete dimer interface, in addition $\alpha 4$. It is significant therefore that $\alpha 4$ in AgrA_C199S is extremely hydrophobic.

Helix 4 has the amino acid sequence LTYLTFVY and presents very little charge to the protein surface. Although $\alpha 4$ lies to one side of the central β -sheet it is slightly removed from $\alpha 2$ and $\alpha 3$ and consequently more solvent exposed. Several views of $\alpha 4$ are presented in figure 3.7. This hydrophobic region is highly conserved amongst AgrA homologues with residues Thr97, Phe98 and Tyr100 being completely conserved whilst Leu93, Tyr95 and Leu96 exhibit conservative mutations to other hydrophobic residues (fig. 3.17). The charged and relatively entropic residues Glu92 and Lys101 reside at either end of $\alpha 4$ and are also completely conserved in AgrA homologues. Thus the role of Glu92 and Lys101 may be to prevent spontaneous aggregation of AgrA *in vivo* via $\alpha 4 \cdots \alpha 4$ association, and this inference is somewhat supported by purification data presented later in section 4.1.6. Structural evidence of this hydrophobic patch on the N-terminal domain of AgrA, combined with what has previously been reported for bacterial response regulators suggests, that $\alpha 4$ probably forms a significant part of the physiological dimerisation interface of AgrA.

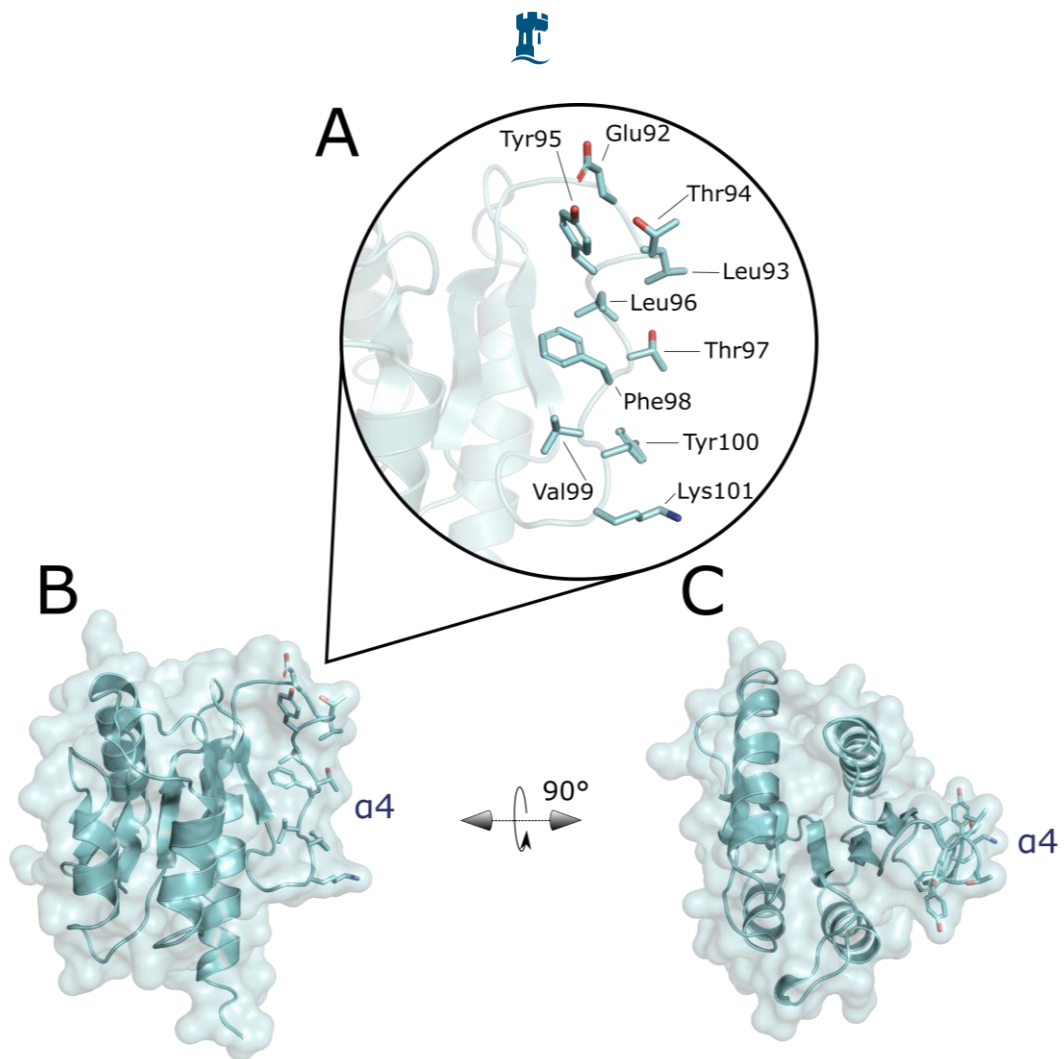


Figure 3.7. Different views of $\alpha 4$ are shown. (A). A zoom-in of $\alpha 4$ with side chains shown as sticks and labelled in black. (B). A transparent surface representation showing the path $\alpha 4$. (C). The same as (B) but rotated horizontally through 90° so as to be viewed top-down with respect to $\alpha 4$.



3.9: An AgrA_C199S Crystallographic Dimer Resembles Other Response Regulator Dimers

At the resolution at which the full-length AgrA_C199S structure solution presented here has been solved it is impossible to deduce whether or not the phosphate-binding site is occupied. Consequently it is impossible to deduce conclusively from the crystal structure alone whether AgrA_C199S is in an “on” or “off” state with respect to phosphorylation. Even if high-resolution crystal data were forthcoming, a comparative study would need to be undertaken of both phosphorylated and unphosphorylated crystal forms of the N-terminal domain of AgrA, coupled with an *in vitro* or *in vivo* assay which could demonstrate AgrA self-association upon phosphorylation.

Further frustrating the goal of commenting on any physiological relevance of AgrA_C199S dimerisation in the crystal structure is the inability of the experimenter to form AgrA_C199S dimers *in vitro*, which has already been detailed. Mutating the phosphate-receiving aspartic acid residue to glutamic acid mimics phosphorylation and induces dimerisation for variety response regulators *in vitro*.^[255,257] Exposing recombinant response regulator samples to phosphate donors such as acetyl phosphate, or phosphate mimetics such as beryllium trifluoride has also been shown to induce response regulator dimerisation *in vitro*.^[87,257,258] In attempting to study the AgrA_C199S dimerisation interface, the AgrA_D59E_C199S double mutant was prepared and AgrA_C199S was separately treated with beryllium trifluoride. It may also be the case that the C199S point mutant prevents the AgrA sample from dimerising *in vitro*, however, this would be hard to verify given the sensitivity of WT AgrA to oxidation.

Neither of these strategies induced dimerisation of AgrA_C199S as evidenced by size exclusion chromatography, with the sample always existing entirely as a monomer. This resistance to dimerisation presented a barrier to solution-based studies of the physiological dimerisation interface to enrich the crystal data. Comparative SEC chromatograms for AgrA_C199S, AgrA_D59E_C199S and [BeF₃]⁻-treated AgrA_C199S are presented in figure 3.8.

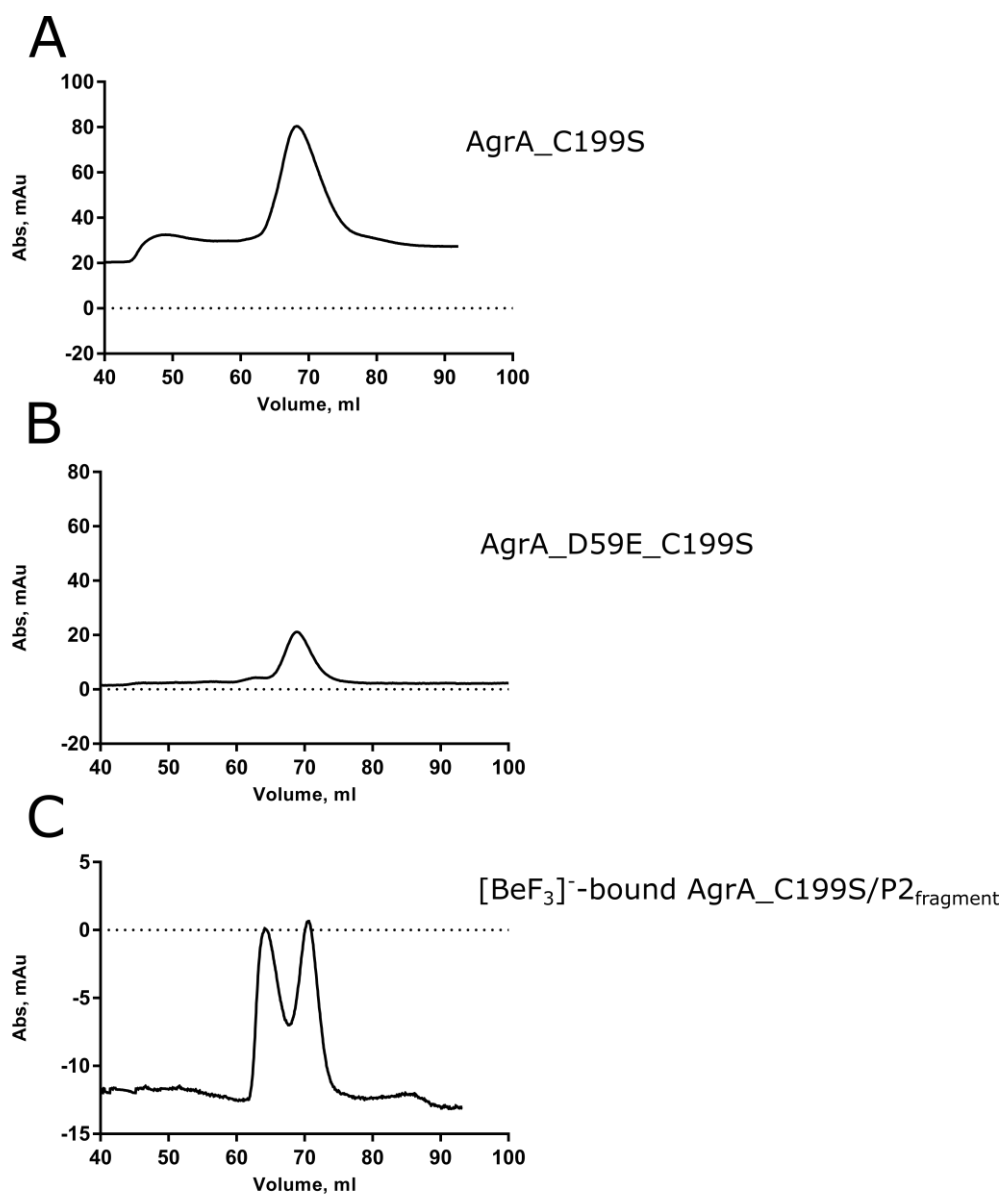


Figure 3.8. Comparison of size exclusion chromatograms for two experiments designed to encourage dimerisation of AgrA_C199S, to a control AgrA_C199S purification chromatogram. (A). AgrA_C199S monomer as it appears during the sample purification and before any crystallisation “work-up” with DNA or [BeF₃]⁻ etc. (B). The AgrA_D59E_C199S double mutant also purifies as a monomer. (C). AgrA_C199S treated with 10 molar equivalents of [BeF₃]⁻ and 2 molar equivalents of P2_{fragment} DNA used for crystallisation. The left-most peak corresponds to AgrA_C199S monomer bound to DNA whilst the right-most peak corresponds to the excess DNA.



Not all response regulators exhibit the same recalcitrance to dimerisation as AgrA_C199S however, particularly when the sample studied is the N-terminal domain only. There are abundant structures of response regulator N-terminal domains available in the literature and brief survey of these indicated that most of these samples purify as dimers or readily formed dimers with the appropriate chemical stimulus. Authors of these studies could therefore perform such experiments as SAXS, NMR or MS to infer which crystallographic dimer corresponded to the one which was prepared *in vitro*.^[90,223,224]

Consequently there is a well-evidence paradigm for response regulator dimerisation interfaces, which invariably occur through $\alpha 4$ as has already been alluded to.^[211] Another consequence of this is that there are many structures of physiologically representative response regulator dimers deposited in the Protein Data Bank against which to compare the AgrA_C199S structure in lieu of direct studies upon AgrA_C199S itself.

Also worth considering when making inferences about physiological relevance of crystallographic AgrA_C199S dimers are the basic steric and chemical requirements of AgrA binding to its cognate promoters. AgrA binds as a dimer to a pair of 9 bp direct repeats separated by an additional 12 bp of β -form DNA. The DNA used to crystallise both the C-terminal domain of AgrA and full-length AgrA_C199S consists of a 15 bp oligoduplex. This short oligoduplex itself contains the 9 bp sequence corresponding to the upstream AgrA recognition site on the P2 promoter^[81] flanked on either side by an additional 3 bp found on the *agr* locus required to ensure AgrA binding *in vitro*, plus the unpaired 5' A/T overhangs.

If one were to imagine two copies of the “crystallisation” DNA side-by-side, six additional base pairs would need to be built, including the compliments to the unpaired overhangs, to achieve the proper 12 bp spacer. The displacement along the helical axis, or rise axis, is $\sim 3.4 \text{ \AA}$ per bp for β -form DNA.^[259] Therefore six extra bp of spacer DNA corresponds to a rise of approximately 20.4 \AA . Over this distance the DNA cannot perform any major conformation feats such as looping back on itself as is observed for the 66 bp *Pseudomonas aeruginosa* LytTR-type transcriptional regulator AlgR promoter.^[260] It also places the DNA-binding domains too far apart to for the physiological dimer



interface to occur between them.^[80] Consequently the N-terminal domain of AgrA must form the physiological interface, and given what is observed for AgrA homologues, probably have perfect or nearly-perfect two-fold rotation symmetry and be roughly co-aligned.

There are three unique dimeric pairs formed between AgrA_C199S protomers in the I222 unit cell which exhibit contacts between N-terminal domains (fig. 3.9). One has the protomers arranged perfectly antiparallel (fig. 3.9C). The second has additional intermolecular contacts between N- and C-terminal domains of the two protomers (fig. 3.9B). This extra set of contacts aligns the C-terminal domains in an antiparallel fashion and would prevent their rotation in solution to achieve the proper line astern arrangement required to bind to a pair of linear repeats. The third (fig. 3.9A) co-aligns the N-terminal domains and bears striking similarities to other structures of phosphorylated response regulators. A comparison of this AgrA_C199S dimer to three homologues exhibiting crystal structures representative of their physiological dimers^[88,90,261] is presented in figure 3.10. A schematic of each of their dimer interfaces prepared using the DimPlot tool^[262] (included in LigPlus vers. 2.2) [EMBL] is presented in figure 3.11.

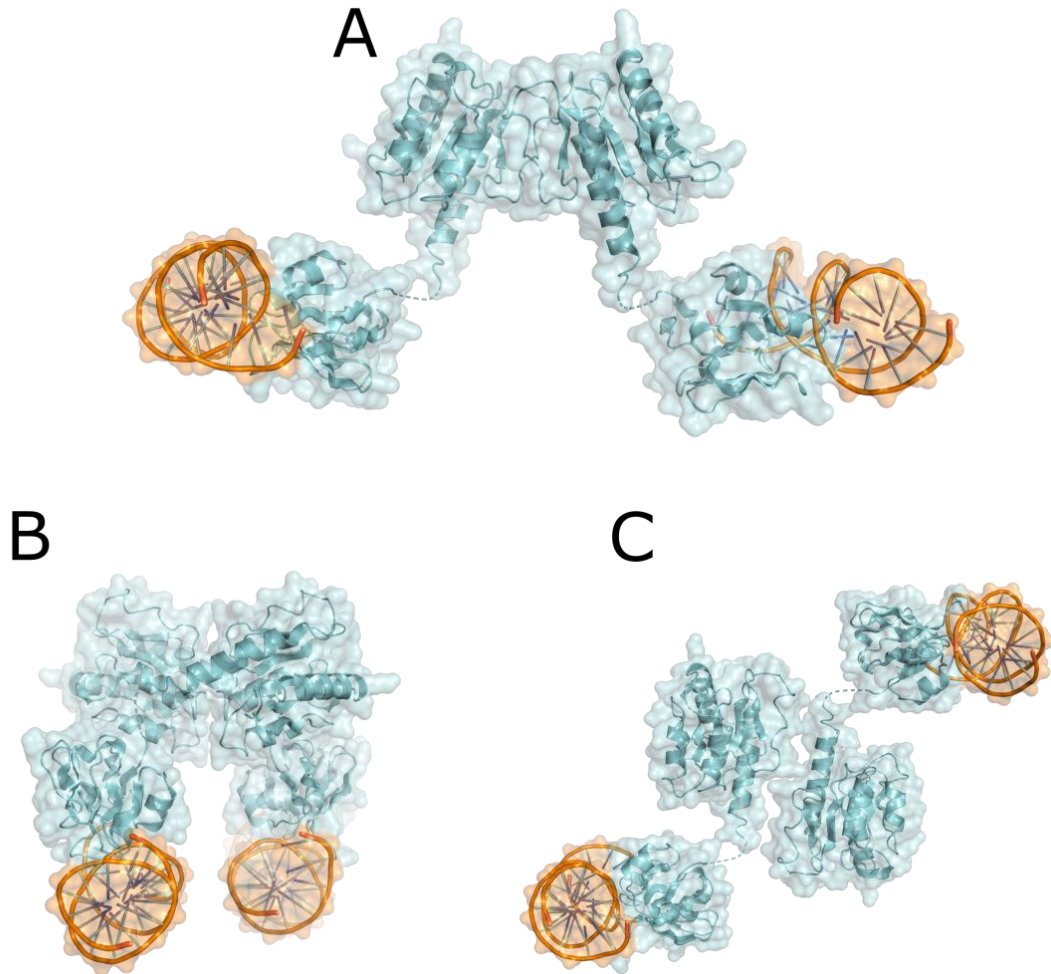
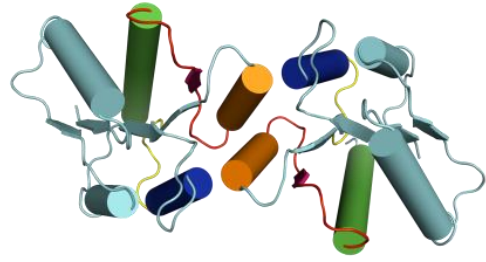
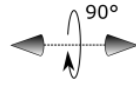
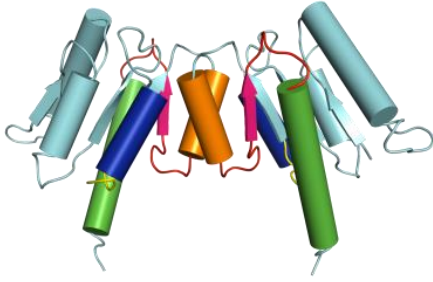


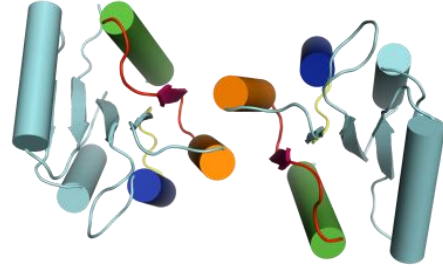
Figure 3.9. The three crystallographically unique dimers of AgrA_C199S as in I222 the unit cell. (A). The two protomers are approximately co-aligned with no intramolecular contacts. This would allow for rotation of the C-terminal domains in the solution state to achieve the line astern arrangement required to bind to a pair of linear repeats. (B). Additional intermolecular contacts are observed between N-and C-terminal domains of the two protomers in the dimer. This would prevent the aforementioned rotation. (C). The two protomers are arranged antiparallel. Whilst the C-terminal domains would be free to rotate in solution to achieve the necessary arrangement, it would mandate that the promoter DNA loop across the face of the two N-terminal domains. Whilst this is feasible, it would leave the two highly hydrophobic regions partially solvent exposed. Also exhibited in (C) is a possible reason as to why the AgrA_C199S/P2_{fragment} crystals diffracted poorly, with the two domains not close packed and exhibiting a clear ridge running between them consequential of Lys101.



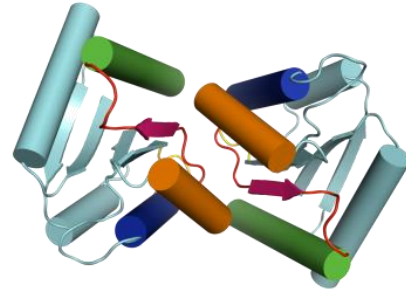
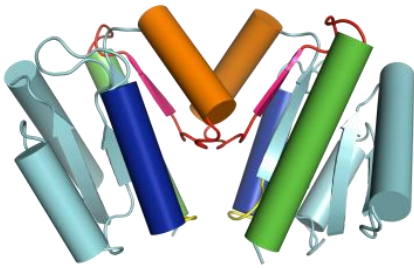
AgrA



FixJ



ComE



LytR

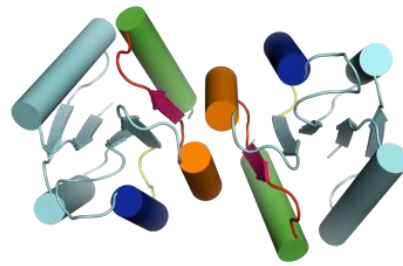
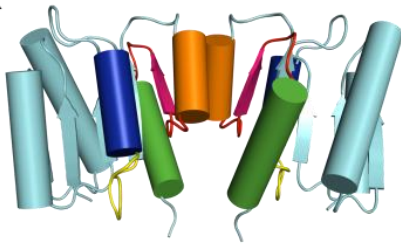
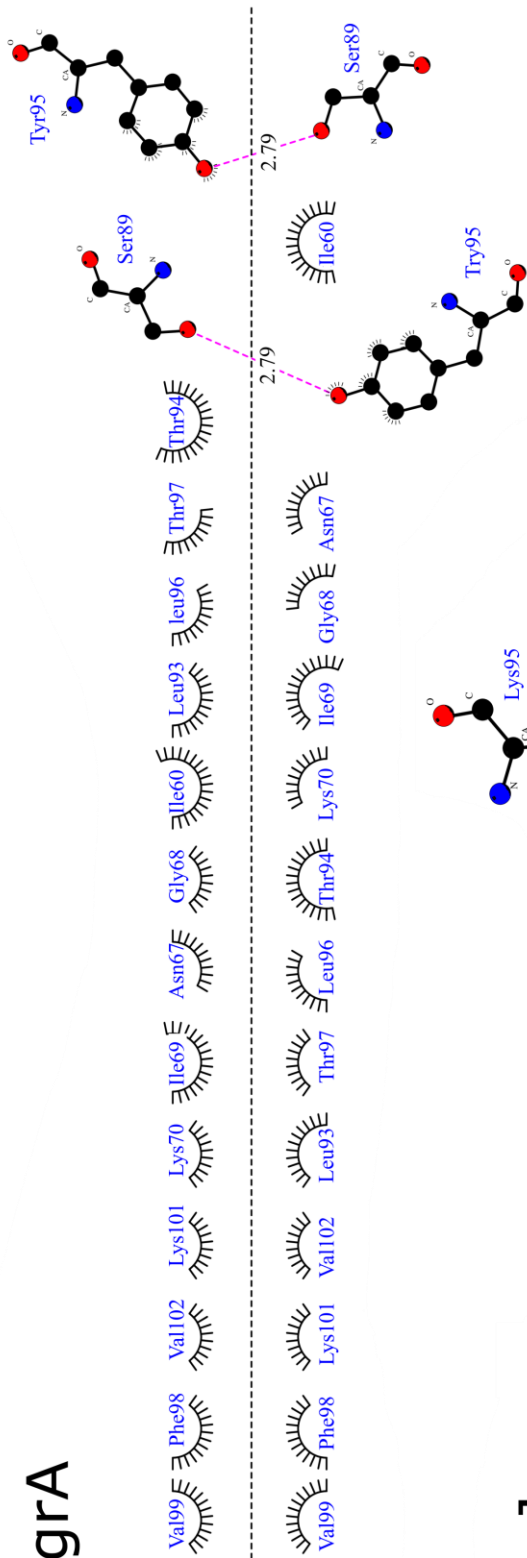


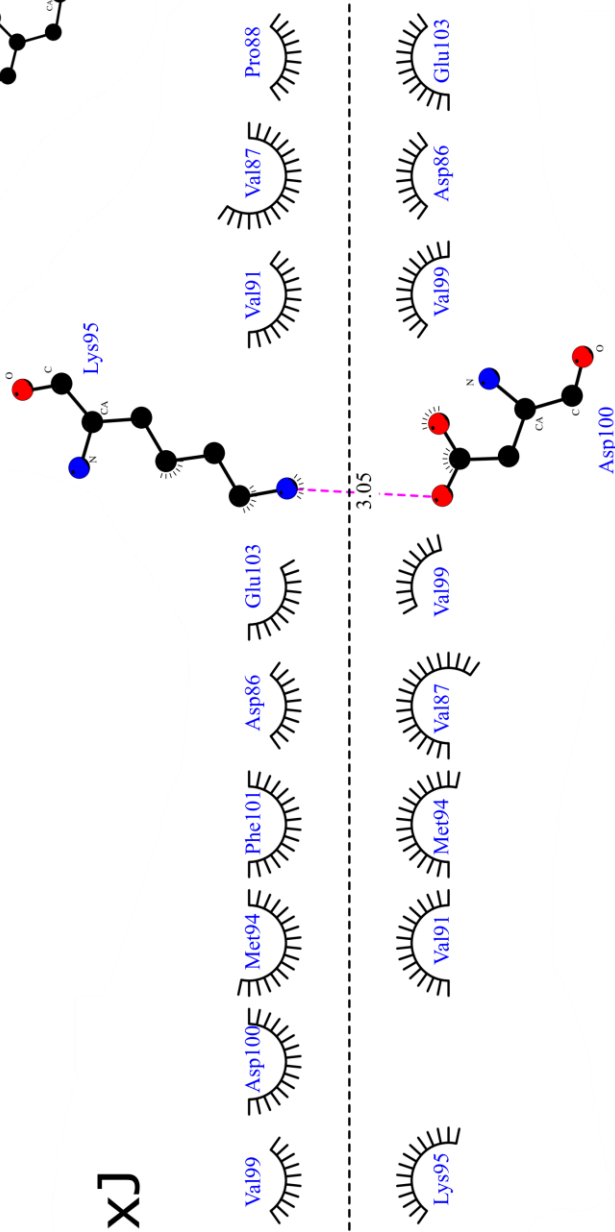


Figure 3.10 (previous page). Topographical views of a crystallographic AgrA_C199S N-terminal domain dimer and the physiological N-terminal domain dimers of: phosphorylated FixJ^[261] (PDB: 1d5w), ComE_D58E^[90] (PDB: 4mld) and LytR^[80] (PDB: 6m8o). FixJ regulates nitrogen fixation in *Sinorhizobium meliloti*, ComE regulates cell competence in *S. pneumoniae* and LytR regulates cell wall metabolism in *S. aureus*. The left-hand column shows a side-on view whilst the right-hand column shows a top-down view. Secondary structure elements involved in the four dimerisation interfaces are coloured as follows: $\alpha 3$ is coloured blue, $\alpha 4$ is coloured orange, $\alpha 5$ is coloured green, $\beta 5$ is coloured magenta, the loops either side of $\beta 5$ are coloured red and the loop between $\alpha 3$ and $\beta 4$ is coloured yellow. The remainder of the structures is coloured cyan. The same colour scheme is retained for all four proteins

AgrA

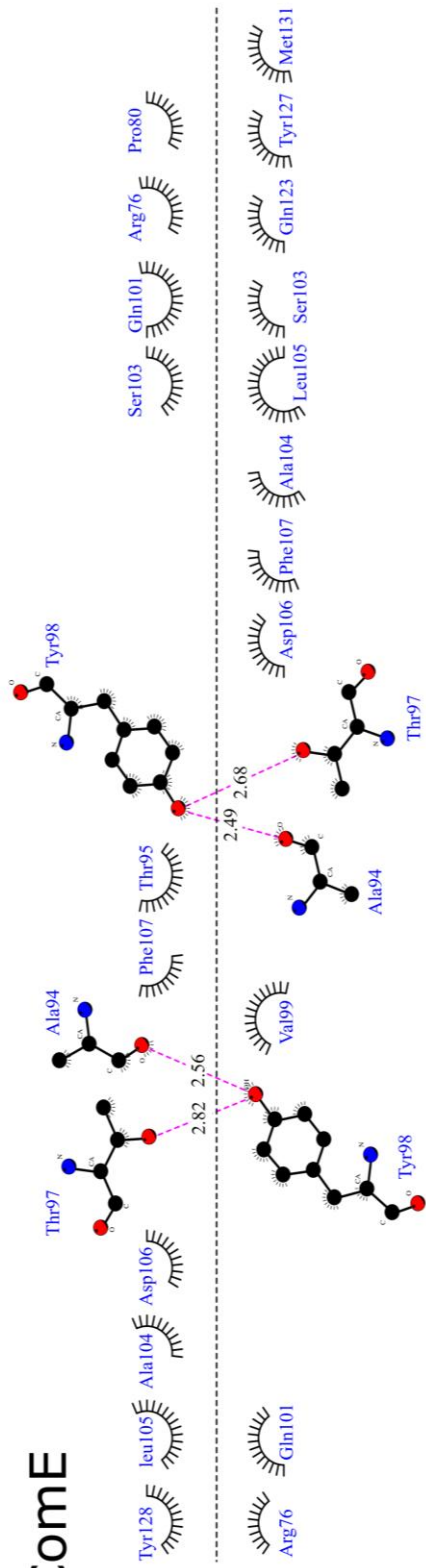


FixJ





ComE



LytR

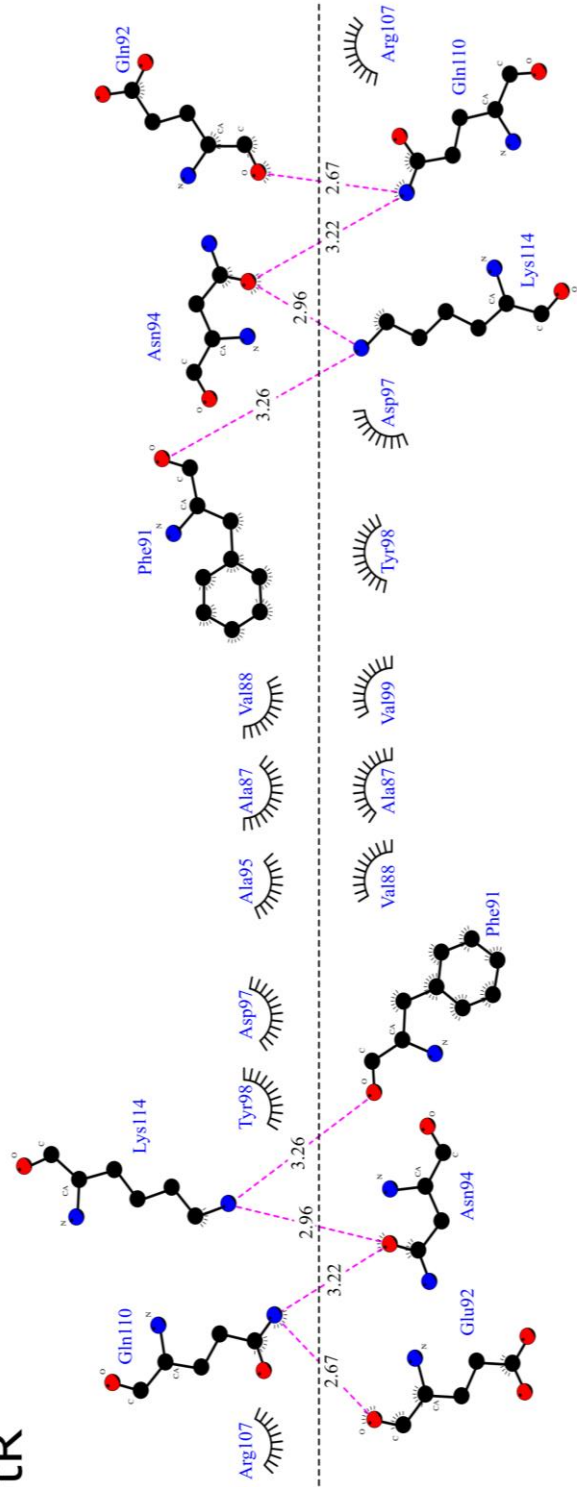




Figure 3.11 (previous two pages). Schematic views of the dimerisation interfaces for: AgrA_C199S, phosphorylated FixJ,^[261] ComE_D58E^[90] and LyrR.^[88] The interface for each structure is shown as a dashed horizontal line with each protomer in the dimer lying on either side of it. Residues involved in H-bonding are explicitly drawn and the H-bonds are displayed as dashed pink lines with their lengths displayed in Å. Carbonyl O and C, and C α and peptide N atoms are labelled in black for explicitly drawn residues. Hydrophobic interactions are shown as black eyelashes. All residues are labelled in blue.



The feature shared by all of the dimers presented in figure 3.10 is that an $\alpha 4 \cdots \alpha 4$ interaction forms the backbone of the dimer interface (orange-to-orange). Other secondary structure units are invoked in each structure however, to impart subtle differences. The crystal structure of AgrA_C199S bears the closest resemblance to the phosphorylated FixJ dimer, being comprised primarily of hydrophobic interactions between $\alpha 4$ s with only one polar contact formed at the dimer interface. In FixJ this contact is between Lys95, which resides at the carboxyl end of $\alpha 4$, and Asp100, which resides at the amino end of $\beta 5$. FixJ differs from the other three structures in lacking a residue with an aromatic side chain in $\alpha 4$. The polar contact at the AgrA_C199S dimerisation interface is between the hydroxyl groups on the Ser89 and Tyr95 side chains. These residues are located in the loop between $\beta 4$ and $\alpha 4$, and $\alpha 4$ respectively. The contact appears to hold the tyrosine ring such that it approaches its symmetry-equivalent face-on, facilitating $\pi \cdots \pi$ stack of 4.2 Å.

The ComE and LytR dimers are similar to each other insofar as they invoke a larger number of polar contacts at the dimer interface in addition to the abundance of hydrophobic contacts. These occur between a wider variety of secondary structure units also, and serve to bring the carboxyl ends each of the $\alpha 5$ s in the dimeric pair together (green-to-green) in a V-shape. LytR retains a $\pi \cdots \pi$ stack in spite of these extra contacts. It occurs between symmetry-equivalents of Phe91 which are located in $\alpha 4$ and has a length of 3.8 Å.

In ComE however, a $\pi \cdots \pi$ stack between aromatic side chains of residues in $\alpha 4$ is abjured due to the extra polar contacts present. The two Tyr98 residues in the $\alpha 4$ s are packed face-to-face, but there is significant shear perpendicular to the $\pi \cdots \pi$ axis which places the aromatic centroids at 6.3 Å apart which is too far for $\pi \cdots \pi$ stacking. In place of this is a C-H $\cdots\pi$ interaction of length 3.5 Å, occurring between the side chains of both Tyr81 residues in the dimeric pair. These Tyr81s are not crystallographically symmetrical despite the remainder of the two protomers in the dimer exhibiting excellent two-fold rotation symmetry, and reside in the loop between $\alpha 3$ and $\beta 4$ (yellow-to-yellow). Of the structures compared ComE is the only example to invoke this loop at the dimerisation interface.



A second C-H $\cdots\pi$ interaction is observed in the ComE structure. It is shorter than the other one with a distance of 2.7 Å and occurs between the methyl H atom on the side chain of Leu105 and the π cloud of Tyr127 (red-to-green). The Tyr127 residue is located on $\alpha 5$, and of the structures compared only ComE and LytR use $\alpha 5$ -located residues at the dimer interface. Instead of C-H $\cdots\pi$ interactions, LytR exhibits a small network of H-bonds between Gln110 and Lys114, which reside in $\alpha 5$; and Phe9, Glu92 and Asn94, which form the very carboxyl end of $\alpha 4$ and the following loop (green-to-orange).

The AgrA_C199S dimer is the only one of the compared structures to invoke residues located in $\alpha 3$ as part of the dimer interface (blue-to-any). Figure 3.10 demonstrates that for FixJ, ComE and LytR, each protomer in the dimer pair mutually presents its $\alpha 4$ into the cleft formed between $\alpha 4$ and $\alpha 5$ on its neighbour (orange goes between orange and green). Each AgrA_C199S protomer mutually presents its $\alpha 4$ to the gap formed between $\alpha 3$ and $\alpha 4$ on its neighbour, however (orange goes between orange and blue). Another way of visualising this is that, as viewed in figure 3.10, the $\alpha 4$ belonging to the left-most protomer in the AgrA dimer is at the back when viewed face-on, whilst for the others it is in front. This packing of $\alpha 3$ against $\alpha 4$ is mediated by the two symmetry-equivalent Ile60s, the residue immediately following the functionally-requisite Asp59, being buried in the clefts formed either side of the central $\alpha 4\cdots\alpha 4$ unit. In consequence to this the phosphorylation sites on the two protomers in the dimer are slightly tilted towards each other, whilst the carboxyl ends of the $\alpha 5$ are slightly tilted away from each other. This is in direct contrast to ComE and LytR in which the phosphorylation sites are tilted away from each other and the carboxyl ends of $\alpha 5$ are tilted towards each other.

3.10: Further Analysis of AgrA_C199S Crystallographic Dimer

The AgrA_C199S dimerisation interface was analysed using the PDBePISA server^[263] (vers 1.52) [EMBL] and was found to bury an area of 755 Å² per protomer, which accounts for 5 % of the total solvent-accessible area of AgrA. The P-value of an interaction is a statistical measure of its specificity and is defined by comparing the solvation energy gained through forming the observed interaction, with the solvation energy gained by forming an interaction through a randomly-chosen, but equally-



sized spot on the protein surface.^[264] The P-value represents the probability of this randomly-chosen “patch” yielding a lower gain in solvation energy (hence more stable) than the observed interface. For the AgrA_C199S dimer discussed, the P-value was found to be 0.029, with a solvation energy gain of $-7.0 \text{ kcal mol}^{-1}$. This compares to an average of $-1.6 \text{ kcal mol}^{-1}$ for randomly-chosen patches of 5 % of the surface. This is to say an AgrA_C199S dimer formed through a randomly-chosen patch comprising 5 % of the protein surface is 97 % less likely to exhibit a stronger interaction than the observed interface. This P-value was over an order of magnitude lower than observed for all other protein···protein interfaces found in the crystal and is remarkably low given the size the interface. For comparison, the P-values of the physiological dimers of ComE_D58E and phosphorylated FixJ are 0.197 and 0.173 respectively.

To assess whether the AgrA_C199S dimer being discussed can occur independently of the spatial arrangement of the C-terminal domains and DNA found in the crystal, *in silico* protein···protein docking was performed. The null hypothesis being: if the observed AgrA_C199S dimer is dependent upon the pseudo-continuous DNA helices holding the N-terminal domains of AgrA_C199S in such an orientation that the $\alpha 4 \cdots \alpha 4$ pack is the most thermodynamically favoured out of only a few possibilities, and more likely a crystallographic artefact therefore, then the interaction should not be readily detected by *in silico* docking of the N-terminal domain to itself. This is because the docking process samples all mutual 360° of space in which in which there is a high probability of detecting additional N-terminal···N-terminal interactions which are more stable than that highlighted but forbade by the restrictions imposed by the packing of DNA molecules and C-terminal domains.

The crystal structure of AgrA_199S in complex with DNA was truncated in PyMOL™ (vers. 2.2.0) to leave just the first 135 amino acids. This truncated model was docked to itself using the ClusPro server^[265,266,267] (vers. 2.0). A protein···protein interface almost identical to the one discussed was identified by the ClusPro server amongst the top 10 results. The docking result was superposed to the crystallographic model using PyMOL™ and two entities excellent similarity with an RMSD of 2.11 \AA between all of the atoms. This superposition is presented in figure 3.12.

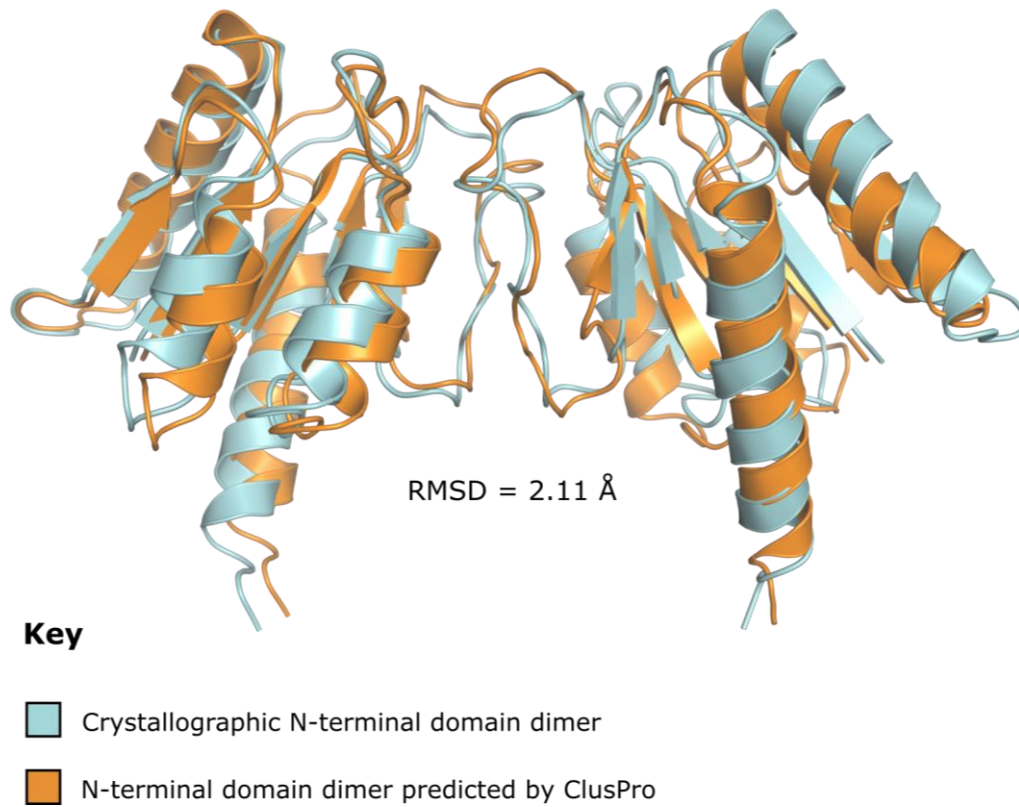


Figure 3.12. Superposition of the AgrA_C199S N-terminal dimer observed in the crystal structure (cyan) and ClusPro docking result (orange). The all-atom RMSD is 2.11 Å and was calculated in PyMOL™. The polar contact between Ser89 and Tyr95 in the crystal structure is observed in the docking result also.



The similarity of this docking result to the observed AgrA_C199S crystal structure implies that the $\alpha 4 \cdots \alpha 4$ packing interface may occur independently of the organisation of the remaining C-terminal domains and DNA molecules found in the crystal structure. This is perhaps unsurprising given that the solvation energy gained in forming the Hoogsteen base pair was calculated at $-1.2 \text{ kcal mol}^{-1}$ by the PDBePISA server, which is significantly less than that calculated for the dimerisation interface.

One final literature observation that is relevant to the potential physiological relevance of the AgrA_C199S dimer interface discussed is that, in trying to obtain crystal structures for both the constitutively active and inactive forms of ComE, both constitutively inactive (D58A) and active (D58E) mutants were crystallised individually. It was observed that both mutants exhibited identical crystal structures (corresponding to the active form according to the SAXS data) despite producing inactive and active phenotypes *in vivo* and purifying as a monomer and a dimer respectively.^[90] The authors conclude that the relatively high sample concentrations of ComE during crystallisation selected for only the phosphorylated form during crystal growth. The same phenomenon has been observed in the solution state.^[224] Although the low-resolution electron density maps for AgrA_C199S preclude any direct comments about occupancy of the phosphorylation site, it is reasonable to speculate the same phenomenon might have occurred. Thus the crystal structure combined with the subsequent analysis provides some evidence for the physiological AgrA dimerisation interface. This dimerisation interface is further probed via SAXS measurements, discussed later in section 4.3.2.

3.2.1: Final Structure Comments

3.2.2: On the Difficulty of Obtaining High-Resolution Crystal Data

Obtaining high-resolution crystal data was complicated by several factors. The high solvent content of the crystals of 67 % did not help. Generally speaking high solvent contents result in weaker diffraction from crystals because most of the solvent is not crystallographically ordered. This leads higher background scattering.^[234]

Another complication is fundamentally connected to the structure of AgrA and involves the flexible linker between the two domains. Having a multi-domain protein in which the domains do not interact



directly with each other but are held together by a flexible polypeptide tether are, again generally speaking, harder to crystallise. This is fundamentally related to the physics of crystallisation, which involves a loss of entropy from the sample molecules. In order for the Gibbs free energy of crystallisation, ΔG_{cryst} , to be negative, some energy must be liberated upon forming crystal contacts to compensate for the entropy loss, which disfavours spontaneity. Additionally, there is an entropy “pay back” from the release of solvent molecules (predominantly water) during crystallisation. This payback comes from ordered or partially-ordered, protein-associated solvent molecules being released from the sample solvation shell to facilitate the crystal contacts.^[268,269]

If a sample is very high in entropy, these other factors may not compensate sufficiently to make crystallisation a spontaneous process. If domains in a multi-domain protein do not associate into a single rigid globule that rotates and moves as a single entity in solution, there is a higher entropy barrier to crystallisation. Although this entropy barrier does not wholly influence the quality of a sample if it can be coaxed to crystallise, it does offer an explanation as to why the crystals might be small, with ΔG_{cryst} becoming positive quite early on in the crystallisation experiment. Smaller crystals generally exhibit weaker diffraction than larger ones and the AgrA_C199S/P2_{fragment} crystals tended to be very small. This effect would exacerbate the problems associated with high solvent content.

The AgrA_C199S/P2_{fragment} crystals nucleated along the bottom of the crystallisation well as discussed in a section 2.5. This made them very difficult to fish. The AgrA_C199S/P2_{fragment} crystals were extremely soft and their stickiness made avoiding damage during crystal fishing challenging. Even poor crystals have exquisite order relative to the universe at large, and disturbing them with a nylon loop can only be harmful even in the best of cases.

Another observation, again probably related to the high solvent content, was that the crystals did not remain intact during exposures to the X-ray beam. This was inferred from the reduction in the number of reflections peaks observed during each exposure, and a reduction in resolution of those which were observed during the diffraction experiment. This presented something of a paradox to collecting data from weakly-diffracting crystals since a higher dose of X-rays is required generate diffraction data,



but the higher doses diminish the quality of the diffraction data as it is collected.^[238] Those AgrA_C199S/P2_{fragment} crystals that were grown in the presence of iodide were less radiation sensitive however, perhaps because the iodide ions absorbed a significant number of otherwise damaging photons or sequestered some of the free radicals generated by the high energy X-ray photons.

3.2.3: Model Building at Low Resolution and Comments on Model Validity

The Structure validation was performed using MolProbity^[270] (vers 4.2) and the PDB validate server (vers. 4.4). All major validation issues detected by the PDB validate server were resolved. A detailed analysis of model quality can be found in the preliminary PDB validation report presented in appendix III. Some key metrics are discussed here, however.

As mentioned already the structure refined to have R and R_{free} values of 30 and 36 % respectively. These values are high, however, nineteen residues, corresponding to the (His)₆ tag, factor Xa protease site and the pCOLD-I translation enhancing element were not modelled. The five linker residues between the two domains were also not built. So there are 24 amino acids from which the scattering has not been accounted for. This will lead to higher R factors but leaving them unmodelled was the scientifically correct thing to do given that there is no experimental evidence for their exact position. Similarly, ordered solvent molecules could not be built which will have a small effect on R factors.

The geometry of the model and the fit to the electron density are all within acceptable tolerances. As a rule of thumb the MolProbity score of a model should be lower than the resolution of the data. The MolProbity score was determined to be 2.28 with a clashscore of 12.^[270] Some validation metrics output by MolProbity are shown in figure 3.13. Of the seven Ramachandran outliers^[271] (3.1 %) two are prolines, Pro30 and Pro40, and one is a glycine, Gly148, and these are found in loops. Prolines and glycines are common Ramachandran outliers owing to the cyclic side on the former and lack of side chain on the latter. They are frequently found as secondary structure disruptors.^[272] The outlier Tyr41 immediately follows Pro40. Residues Ser90 and Val102 are also located in loops and Ser63 is involved in the β -turn. A Ramachandran plot is given in figure 3.14.

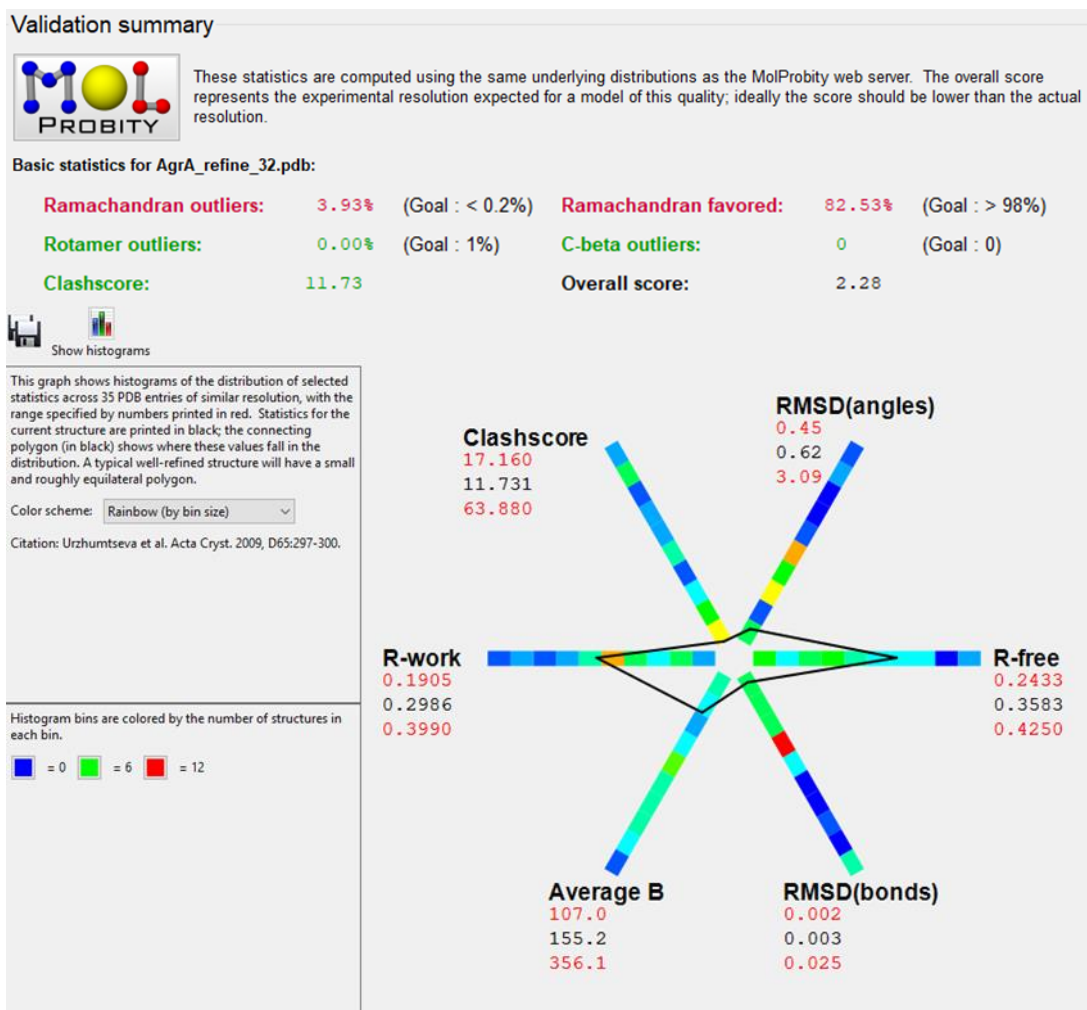


Figure 3.13. Various MolProbity quality metrics showing how the AgrA_C199S/P2_{fragment} structure solution compares to structures deposited in the Protein Data Bank at a similar data resolutions. The AgrA_C199S/P2_{fragment} structure compares favourably to its resolution neighbours. Note that the Ramachandran outliers were subsequently reduced to 3.1 % but the most up-to-date figure is not available due to the Covid-19 outbreak and this validation check represents the penultimate round of refinement. Again the reader is directed to the validation report presented appendix III, which corresponds to the atomic model after the final round of refinement was performed.



Ramachandran Plot for All Non-Pro/Gly Residues

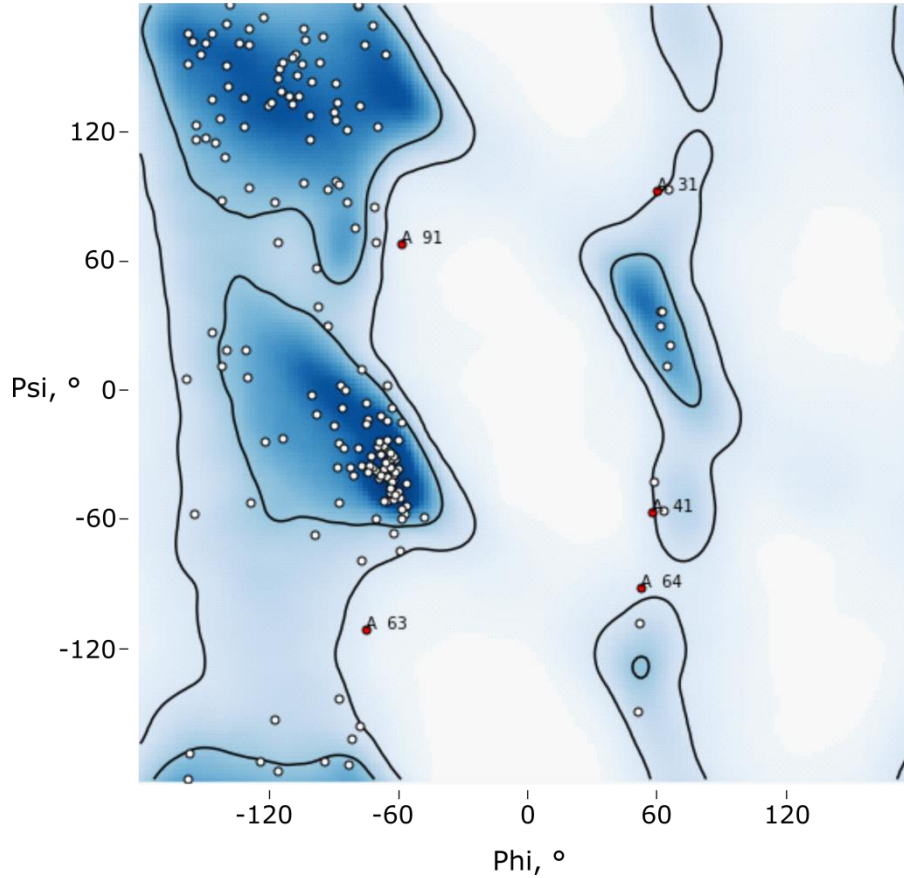


Figure 3.14. A Ramachandran plot of residues in the AgrA_C199S/P2_{fragment} atomic model. Note that like the MolProbity analysis it corresponds to the structure in the penultimate round of refinement. Residue 64, a threonine involved in the β -turn, was subsequently resolved to leave four non-Pro/Gly outliers in the structure.



Electron density for α -helices is usually quite well defined even at low resolution.^[273,274] This was not the case for $\alpha 4$ in the full-length AgrA_C199S structure, however. The complete catenation of these secondary structure units was likely a consequence of the LytR search model used for MR not being accurate for the mutual locations of the corresponding $\alpha 4$ s in AgrA_C199S. Even though the catenation was evidently incorrect, the difference electron density which ought to reveal its true location in the crystal structure was less well-defined than for the other α -helices. Furthermore imposition of ideal α -helical restraints returned higher R_{free} values.

Consequently it was difficult to build this region of the structure accurately. Secondary structure distortions are sometimes observed structures of response regulator N-terminal domains. For example in the structure of constitutively active PhoB_D10A_C53E (PBB: 2jb9), solved at 1.70 Å, $\alpha 3$ continues around a corner whilst the structure in general appears to be considerably strained,^[275] and in full-length ComE (PDB: 4cbv) $\alpha 4$ is bent away from the dimerisation interface.^[90]

The best model was judged to be the one which lowered R_{free} and made chemical sense (no egregious clashes and incorrect valences) and without respect to the expected α -helical structure for these residues therefore, and this resulted in a deviation away from ideal α -helical geometry for $\alpha 4$. In one respect this is valid since it is the solution which provided the lowest R_{free} . On the other hand deviation away ideal α -helical geometry involves breaking the hydrogen bonds which hold it together, and so requires good evidence that the energy released through packing the $\alpha 4$ s together is enough to compensate for the distortion. Many valid but unprovable arguments could be presented both for and against the model as it is built. It is sufficient to say it does not contain any chemical faux pas and provided the best fit to the data. A B-factor putty plot is presented in figure 3.15 and comparative difference density maps before and after the fixing of $\alpha 4$ are presented in figure 3.16.

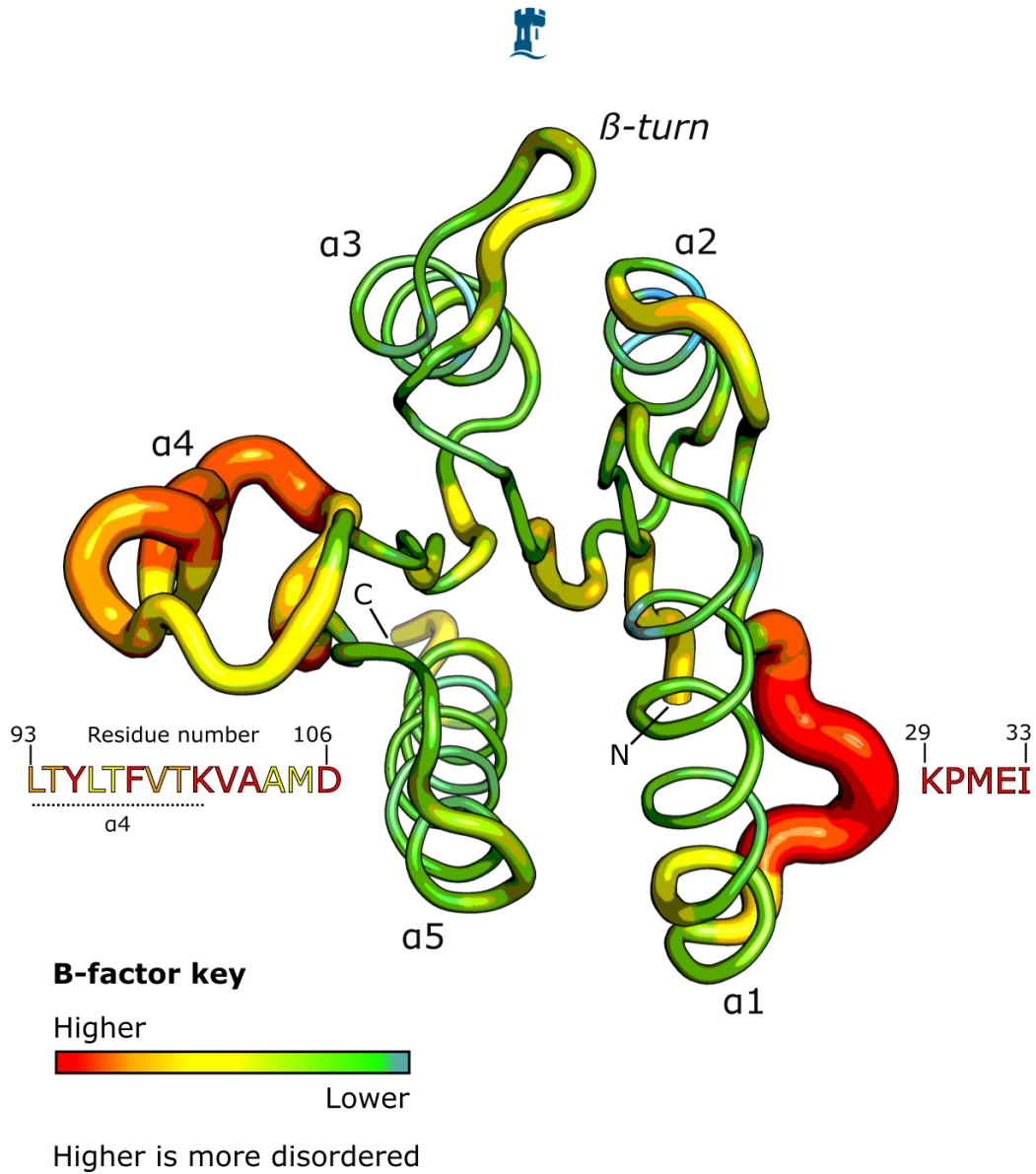


Figure 3.15. Annotated B-factor putty diagram for the N-terminal domain of AgrA_C199S. Higher B-factors (which take units \AA^2) are indicative of a greater magnitude deviation away from an equilibrium position for any given atom. The residues in $\alpha 4$ and in the loop between $\alpha 1$ and $\alpha 2$ (indicated by red putty and with amino acid sequences displayed) are particularly disordered regions of the structure solution.

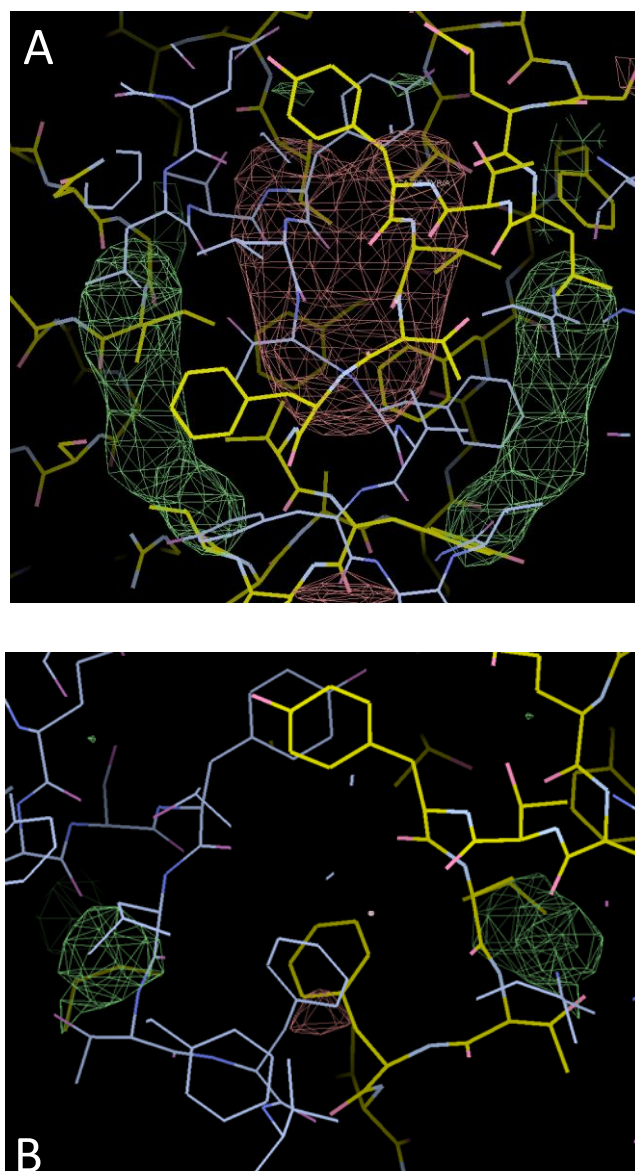


Figure 3.16. Difference density maps about $\alpha 4$. (A). Difference density maps for the catenated symmetry-equivalent $\alpha 4$ s. The large red region indicates too much scattering power has been built into the atomic model at these coordinates whilst the green regions indicate insufficient scattering power has been built at these coordinates. The blue bonds represent one AgrA_C199S molecule and yellow bonds represent the symmetry-equivalent molecule. (B). Difference density maps for the refined structure after the rebuilding of the area around $\alpha 4$ s. According to the difference density maps and the R_{free} value it is the best fit to the data but in order to achieve it $\alpha 4$ has forgone ideal helical geometry. The small dashes and artefacts for the bottom figure are pieces of the structure which lie just outside of the front clipping plane. Screenshots are taken from COOT.



Parameter	Value
Data collection	
Wavelength (Å)	0.9688
Space group	I 2 2 2
Cell dimensions	
a, b ,c (Å)	61.11, 96.95, 195.70
α , β , γ (°)	90.00, 90.00, 90.00
Resolution (Å)	54.12 (4.76)
Total no. reflections	13434 (1312)
No. unique reflections	2359 (236)
Redundancy	5.7 (5.6)
Spherical completeness (%)	75.2 (36.9)
Ellipsoidal completeness (%)	86.8 (71.3)
Mean I/ σ (I)	3.1 (1.0)
CC _{1/2}	0.92 (0.23)
R _{merge}	0.291 (2.39)
R _{pim}	0.138 (1.12)
R _{meas}	0.323 (2.65)
Refinement	
R _{work} /R _{free} (%)	29.7/35.7
No. of atoms	
Protein	1925
DNA	650
Ligand	0
Water	0
Average all-atom B-factor (Å ²)	155.0
Wilson B-factor (Å ²)	134.9
RMS deviations ^a	
Bond lengths (Å)	0.003
Bond angles (°)	0.62
MolProbity statistics	
Overall score	2.28
All atom clashscore	11.73
Ramachandran favoured (%)	82.5
Ramachandran outliers (%)	3.1
Solvent Content (%)	67

Table 3.1. Crystal data for the AgrA_C199S/P2_{fragment} crystals. ^aRMS stands for Root Mean Square.

^bValues in parenthesis are for the highest resolution shell.



CLUSTAL O(1.2.4) multiple sequence alignment

C6

```

C.cellulovorans -----MNIVICEDRV-QLEKTKEIENYAMMEDNGIKVVLAC6TANP-HDVLC6SFLQE-- 48
L.grayi -----MDMLSVFICEDSSMORKRVC6TGYINQYIKSQRLDMQVKLSTGNP-FEVIDELKAAE 54
E.faecalis -----MISIIICEDDWRQC6RQMLETYVKNYIMMESLDMELVFSC6TGNP-LEVLEFVK-SQ 51
B.cereus -----MLEIFIC6CEDEIEQRERLKKYIENYIMC6ENLDMKIVVCTGEP-QEVIDYLK-EN 51
S.silvestris MKSGELSVLSIFVCEDNKEQC6REC6RITKAIENIVLI-ELDMKC6LALATEKP-DDIIDFIK-EN 57
S.aureus -----MKIFVC6CEDDQC6REHMVSIC6IKNYIMIEEKC6PMELTLATNDC6POYEVLEQSK-EL 51
S.saprophyticus -----MKIYIC6CEDDVC6KQRENC6MVEIC6INNYIMIE-KPMEIALAC6TDNP-YEILKMSK-EN 49
S.epidermidis -----MKIFIC6CEDEC6REHMVSIC6INNYIMIEEKC6PMIEIC6EVAC6TADE-YDILERSK-NL 50
S.lugdunensis -----MKIFIC6CEDDC6PKQRENC6MVTIC6IKNYIMIEEKC6PMIEIALAC6TDNE-YEVLEQAK-NM 50

```

C55 D59

```

C.cellulovorans -DKADCYFLDILK--DEITGITLC55GSKIREEDPIASLVYITD59THAEMSFLTFYIKLA-LDFI 104
L.grayi TNEMGLYFLDIDLQC55PEMNGFQLAD59QKIREHDPC55RGFIID59FITC55THAELSYMD59TFYKVEALDYI 114
E.faecalis SRIIGLYFLDVC55DLQ-HEMSGLTD59LAAKIRC55YDNLGKIVFD59VTC55THGELSD59YLTFYKVEALDYI 110
B.cereus -NSTGLYFLDVC55DLQ-ADKSGIALGAEIRD59HYDTRGSIVFC55ITD59THSELTYLTFYKVEAMD59DI 109
S.silvestris -NGVGLYFLDIDLK-ATINGIELATKIREYDC55PRGFIVD59FVTC55THAEMSD59YLTFYKVE-MDYI 114
S.aureus -NDIGCYFLDIQLE-ADMNGIKLASEIRKD59HPVGNIIC55FVTD59SHSELTYLTFVYKVAAMD59DFI 109
S.saprophyticus -EDVGCYFLDIQLE-ADINGIKLGSEIRKD59YDPIGNIC55-FVTD59SHSELTYLTFVYKVSAMD59DFI 106
S.epidermidis -NDIGCYFLDIQLE-ADINGIKLASEIRKD59HPVGNIIC55FVTD59SHSELTYLTFVYKVAAMD59DFI 108
S.lugdunensis -NDIGCYFLDIQLS-TDINGIKLGSEIRKD59HPVGNIIC55FVTD59SHSELTYLTFVYKVAAMD59DFI 108

```

C123

```

C.cellulovorans IKDNPDSLKEKLLSTLKEAHRRYLKLGC123EQDNIQKLQIKTTGRTHNIDFQNIYFFEASPDC123S 164
L.grayi IKD-FNEMEKRVYDCLAKAEC123ERVNVKPEAGAVFTLQISD-KKVIREREDDILFFETSPTA 172
E.faecalis IKDNKEHLQC123QKICENIQLAHERITSERG-DTKKLFKFDGDTIRSVDC123MATIIFFESSVS 169
B.cereus TKDEFTDIQKRVIDCIDTANKRYITNRH-GNKKIFQTKLGDKVIS-NYDEILFFETSC123SQL 167
S.silvestris IKDNYRNIKERVHQCILDANKKYSTKST-DLQKIFSC123IKVDDRIINIEYSKILFFETSC123STI 173
S.aureus FKDDPSELKMRIIDCLETAHC123TRLKLLSKESNVDTIELKRGNSVYVQYDDIMFFESSTKS 169
S.saprophyticus FKDDPAELKTRIIDCLETAHC123TRLKLLSKERTVETIELKRGNSVYVQYDEVC123VMFFESSTNS 166
S.epidermidis FKDDPDELKSRIIDCLETSEC123SRLKLLSKESSVETIELKRGNSIYVQYDDVMFFESSTKS 168
S.lugdunensis FKDDPAELRTRIIDCLETAHC123TRLQLLSKDNSVETIELKRGNSVYVQYDDIMFFESSTKS 168

```

C199

```

C.cellulovorans HKIILHLENEHIEYFGRC199LKNYEG---HPDFRCH-SFIVNKSC199KIK-NIDSKERSINLDNC199GE 219
L.grayi HKVILHGKNRQIEFYGC199LKNIEKMLDDTFYRCHRSFIVNKNC199NIH-EMDSKN-GVLMINGE 230
E.faecalis HKIVLHLENGEIEFYGC199SLKEIE-EQSNDFYRCHKSYLIRKC199HISK--VIKSERIVEMSNC199GE 226
B.cereus HKIILHGMRQVEFYGC199LKEIV-ELDSRFYRCHNSYVC199VNKDNIA-EIDMKKREVMVC199NGE 225
S.silvestris HKVIVHAMDRQIEFYGC199KMKEIEAKLDSGYFRCHKSYIVC199NKENIK-EIDLNNRC199CIHMVC199NGE 232
S.saprophyticus HRLIAHLDNRQIEFYGNLC199KQIS-QIDDRFFRCHNSFVC199VNRHNIS-SINSKERVVYFKNC199GE 224
S.epidermidis HRLIAHLDNRQIEFYGNLC199KELD-QLDERFFRCHNSFVC199VNRHNIDSSIDSC199KSRVVYFKNC199E 227
S.aureus HRLIAHLDNRQIEFYGNLC199KELA-QLDERFFRCHNSFVC199INRHNIE-SIDSKERIVYFKNC199E 227
S.lugdunensis HRLIAHLDNRQIEFYGNLC199KELS-QLDDRFFRCHNSFVC199VNRHNIE-SIDSKERIVYFKNC199E 226

```

C228

```

C.cellulovorans LCYASARLIKGLLKETTFT 238
L.grayi NCYVSTKQIKSLSF----- 244
E.faecalis RCLVSVRAMKNL----- 238
B.cereus VCYASSRFLKGLKNIVQR- 243
S.silvestris RCLISTRMLKGLIE----- 246
S.aureus NCFASVRNVKKI----- 239
S.saprophyticus HCYASVRNVKKI----- 236
S.epidermidis HCYASVRNVKKI----- 239
S.lugdunensis HCYASVRNVKKI----- 238

```



Figure 3.17 (previous page). Clustal Omega sequence alignment^[276] of *S. aureus* AgrA [GI: 2422418] to homologues from bacterial genii: *Bacillus cereus*, strain NVH0597-99 [GI: 1960413]; *Solibacillus silvestris*, strain StLB046 [GI: 3028770]; *Listeria grayi*, strain DSM_20601 [GI: 2998208]; *Enterococcus faecalis*, strain 62 [GI: 3234790] and *Clostridium cellulovorans*, strain 743B [GI: 3274416], and staphylococcal species: *epidermidis*, strain W23144 [GI: 7366214]; *lugdunensis*, strain HKU09-01 [GI: 1512221]; *saprophyticus*, strain ATCC_15305 [GI: 2895503]. The alignment is annotated as follows: (*), completely conserved, (:), highly conserved, (.), weakly conserved. Completely-conserved residues have also been highlighted in red boxes for additional clarity. Pertinent cysteine residues are highlighted in green and are labelled with residue numbers as they occur in the *S. aureus* AgrA sequence. Note the conservation of Cys55 and Cys123 in all staphylococcal AgrA and the complete conservation of Cys6, Cys199 and Cys228 across all homologues.



Chapter 4: AgrA – Better Crystals and Structure-Activity Expts.

4.1.1: Pursuit of Better AgrA Crystals via Surface Entropy Reduction

The notion of particularly flexible macromolecules having an entropic barrier to crystallisation was discussed briefly in the previous chapter in the context of the poor diffraction data from the AgrA_C199S/P2_{fragment} crystals. In the most extreme cases it may prevent crystallisation from occurring at all.^[277] In less extreme cases it may reduce the quality of the resulting crystals. Losing entropy during any process impedes its spontaneity.^[278] Packing of amino acids with side chains that can adopt a high number of discrete side chain conformations represents a larger loss of entropy than packing of amino acids possessing shorter side chains with fewer possible conformations.

Glutamic acid, aspartic acid, arginine and lysine all have comparatively flexible side chains^[279] and so burying of these residues during crystallisation is entropically disfavoured, and this phenomenon is evident in the natural world. For example, lysine residues on average have the largest contribution to protein solvent-accessible surfaces^[280,281] but are the least abundant residue at oligomerisation interfaces,^[282] crystal lattice contacts,^[283] and at the interfaces of multi-component functional complexes.^[284,285] It has been argued that this negative correlation is evolutionary^[286] and consequent of the need to prevent spontaneous protein aggregation *in vivo*, which may lead to diseased states,^[287] e.g. sickle cell anaemia.^[288]

This phenomenon may be reversed in macromolecular crystallography in order to promote controlled aggregation of the target molecule^[289] and it is now relatively common to find examples of published structures with clusters of point mutations in which flexible amino acids, usually lysine, glutamic acid and arginine, have been mutated to shorter ones such as alanine or serine.^[290,291,292] This process is known as Surface Entropy Reduction (SER).

4.1.2: Surface Entropy Reduction on AgrA_C199S

AgrA is particularly rich in glutamic acid residues (8.2 %) which along with lysine residues, may be taken to betray areas which lie on the protein surface in the absence of reliable structural data.^[293]



From the AgrA_C199S structure solution however, it was possible to gain an understanding of the packing interactions in the crystal lattice and deduce whether or not they could be rationally modified to improve the quality of the symmcystals. It should be noted that this SER work was underway immediately prior to the Covid-19 outbreak. Consequently there are many follow-up experiments which remain to be performed.

The I222 unit cell contains a lattice point at its geometric centre. Inspection of this point for the AgrA_C199S structure revealed four symmetry-equivalent lysines, Lys101s, to all be converging upon each other. Further inspection of the structure suggested that Lys77 appeared to be preventing close contacts occurring between neighbouring N-terminal domains. These two areas were chosen as targets for SER and are shown in figure 4.1. Inspection of the sequence alignment given in figure 3.17 demonstrates that these target lysines are completely conserved in staphylococcal AgrA, however. So the choices were: A) mutate another patch on the surface and hope for a new crystal form which effectively amounts to starting again, B) abandon SER altogether or C) mutate the conserved lysines.

Option C) was chosen and a small library of point mutations was introduced to AgrA_C199S. These are shown in table 4.1. Eventually all of the point mutants were successfully introduced, transformed and had their expression tested. The resulting test expression gel is shown in figure 4.2. Immediately it looked as if K77A was going to be a problematic mutant, with the test cultures showing negligible expression. All of the other mutants appeared to express as well as the AgrA_C199S control.

Of the five constructs presented in table 4.1, three were purified and of these two were used to set up crystallisation trials. The crystallisation trials were established using identical strategies reagent concentrations etc. as were employed for the AgrA_C199S/P2_{fragment} crystals. The purification of these constructs and the properties of the purified sample are discussed in turn. Ni²⁺-affinity chromatography data is presented in figure 4.3. SDS-PAGE gels post Ni²⁺-affinity chromatography are presented in figure 4.4. Size exclusion chromatograms are presented in figure 4.5 and SDS-PAGE gels post SEC are presented in figure 4.6.

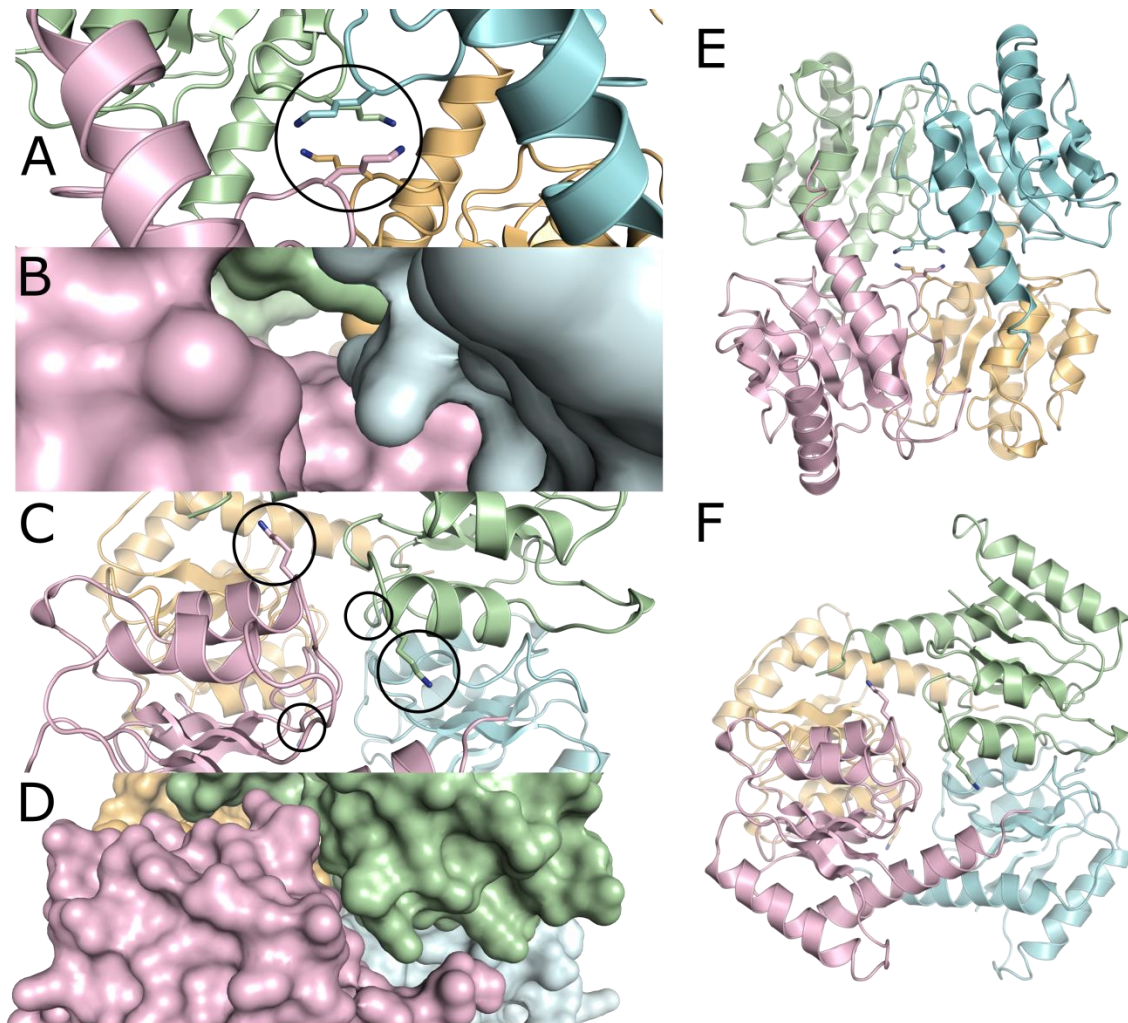


Figure 4.1. Summary of the rationale behind the choice of SER point mutants introduced to the AgrA_C199S construct. (A). A cartoon diagram of four symmetry-equivalent N-terminal domains of AgrA_C199S within the crystal, showing their respective Lys101s (highlighted in the circles) converging. (B). A surface representation of the same area showing a distinct cavity formed in the area of the Lys101s. (C). A cartoon diagram of four symmetry-equivalent N-terminal domains of AgrA_C199S showing the location of Lys77s (highlighted in the circles). (D). A surface representation showing the lack of close packing between the pink and green protomers. There is a deep chasm the interface between them. (E) and (F). Zoomed-out cartoon diagrams showing the spatial organisation of the entire four copies of the N-terminal domains and the locations of Lys101s (E) and Lys77s (F).



Construct	Purification observations
AgrA_K101A_C199S	Multiple Ni ²⁺ -affinity species.
AgrA_K101T_C199S	Not purified due to Covid-19.
AgrA_Y100T_K101A_C199S	As AgrA_C199S.
AgrA_Y100T_K101T_C199S	Not purified due to Covid-19.
AgrA_K77A_C199S	Aggregation/lead to cell death.

Table 4.1. Summary of AgrA_C199S mutants generated for SER studies. Salient observations made during purification are also presented.

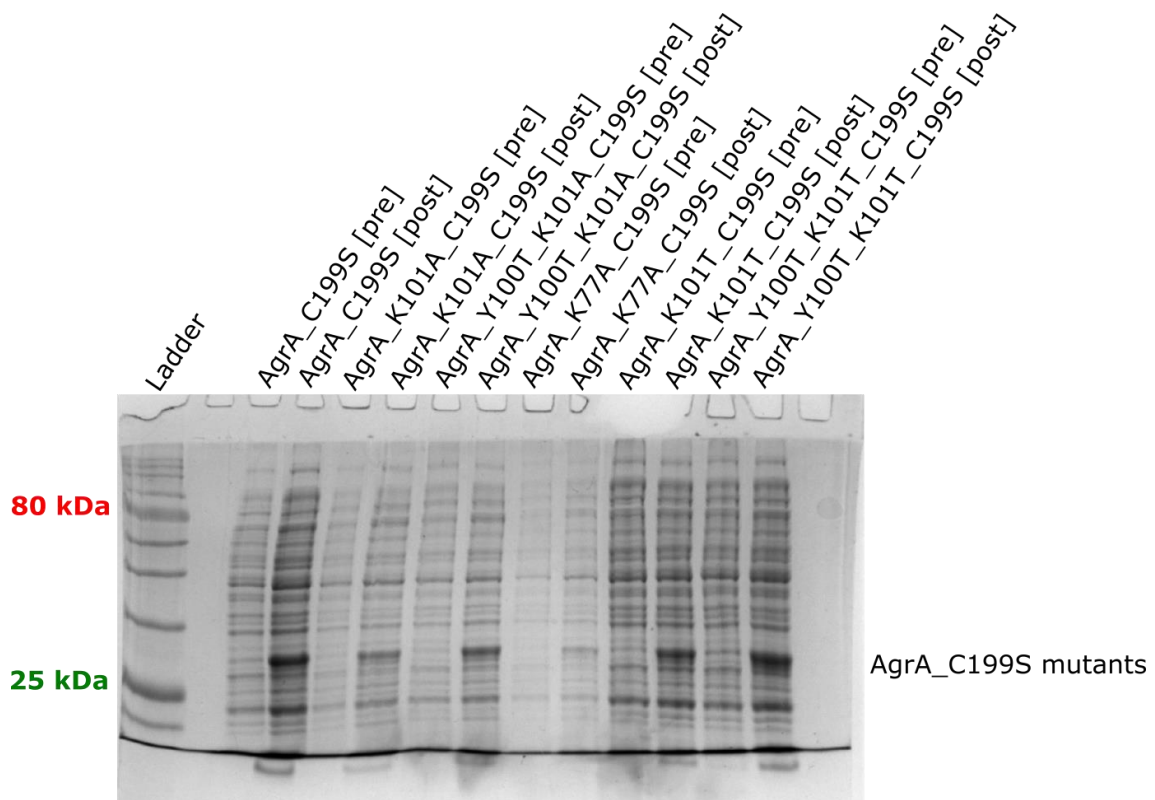


Figure 4.2. Test expression gel of the five AgrA_C199S SER mutant constructs and the AgrA_C199S control. The lanes correspond to cytoplasmic fractions of test cultures before treatment with isopropyl β -D-1-thiogalactopyranoside (IPTG), [pre]; and after overnight expression upon addition following addition of IPTG, [post]. The two left-most lanes following the ladder correspond to the control expression of AgrA_C199S. The gel is 16 % polyacrylamide, reducing.



4.1.3: SER Construct 1 – AgrA_K77A_C199S

Lys77 is completely conserved among staphylococcal AgrA homologues. Mutation to alanine resulted in a noticeable retardation of cell growth during expression. This can be observed on the test expression gel, which exhibits less proteomic matter in the corresponding lanes compared to the controls and the other SER constructs. Inspection of figure 4.1D shows a deep reverse S-shaped cleft between the pink and green protomers in the crystal structure. This appears to be caused by the Lys77 residues which protrude out of otherwise complimentary areas on the AgrA_C199S surface. It is possible that this unfortunate property of the crystal lattice is absolutely necessary *in vivo* in order to prevent an aggregation event that is encouraged by the K77A mutant. Changing the mutant residue to one that is small yet charged might yield soluble AgrA_C199S suitable for crystallisation experiments. Similarly, the region which opposes Lys77 in the crystal structure on the adjacent protomer could be mutated to achieve a double SER mutant (triple mutant including C199S) with amino acids which have complimentary chemical properties and that are electrostatically attracted to one another at pH 6.5. These strategies were not pursued in the time that was available, however.

4.1.4: SER Construct 2 – AgrA_Y100T_K101A_C199S

These two extra mutated residues are located at the carboxyl end of $\alpha 4$ which, as previously discussed, is hydrophobic. This construct behaved identically to AgrA_C199S during purification, giving a monomeric peak during SEC and exhibiting excellent purity. Consequently it was possible to establish crystallisation experiments with it. Sitting drop experiments were established using the optimised AgrA_C199S crystallisation strategy and condition to check for *de novo* growth of crystals of this SER mutant. Additionally, microseeds of the AgrA_C199S/P2_{fragment} crystals were delivered via streak seeding into to hanging drops containing sample corresponding to AgrA_Y100T_K101A_C199S. No crystals were observed in any of the experimental wells.

4.1.5: SER Construct 3 – AgrA_K101A_C199S

This construct retains the native Tyr100 immediately preceding the K101A mutant. The Ni²⁺-affinity chromatogram for AgrA_K101A_C199S exhibits two major peaks, the first of which has a sharp



right-hand shoulder. It was common to observe two peaks on the Ni²⁺-affinity chromatograms during AgrA_C199S purifications. The first peak always corresponded to a contaminant eluting at 10 – 15 % buffer *B*, and the second peak corresponded to the AgrA_C199S sample eluting ~ 40 % buffer *B*. The respective buffer *B* retention percentages for the AgrA_K101A_C199S Ni²⁺-affinity chromatogram do not match these values, however. The first peak, as well as being shouldered, elutes at ~ 25 % buffer *B* whilst the second peak elutes at ~ 50 % buffer *B*. Inspection of the corresponding SDS-PAGE gel shows that the second peak corresponds to AgrA_K101A_C199S, which is to be expected given the percentage buffer *B* retention. This sample was further purified and collected but crystallisation experiments using it were never established.

The first peak meanwhile exhibits a band that runs between the 58 and 80 kDa markers on SDS-PAGE. There is a similarly sized protein band in lanes corresponding to the first Ni²⁺-affinity peak for the AgrA_Y100T_K101A_C199S sample as well, however, here is significantly more of the AgrA_K101A_C199S SDS-PAGE gel; enough to suggest that it is overexpressed.

4.1.6: Comments on the Possible Role of Lys101 in AgrA Dimerisation

Although it is generally the case that exposure to sodium dodecyl sulphate and DTT in SDS-PAGE sample loading buffers, combined with heating to ~ 98 °C, breaks apart all non-covalent interactions, this is not always the case. For example, boiling of most membrane protein samples leads to excessive laddering of the sample on SDS-PAGE gels due to non-covalent hydrophobic interactions.^[294] It is feasible therefore that the band which runs between the 58 and 80 kDa markers on the SDS-PAGE gel corresponding to AgrA_K101A_C199S sample might be the hitherto illusive AgrA dimer, given the proximity of this lysine mutant to an extremely hydrophobic region on the surface of the AgrA and its high degree of conservation.

A band on an SDS-PAGE gel is very information-poor, however. In the instance discussed the band may correspond to a contaminant that sticks to the AgrA_K101A_C199S mutant (although this argument must also acknowledge that non-covalent interactions can withstand the heat and denaturants). Alternatively it may correspond to a contaminant that is present in very high abundance



and exhibits some affinity for the Ni²⁺-affinity column. The latter property could impart the former by collection on the column and followed by elution into a comparatively small volume.

These are all plausible scenarios and short of MS data or some other data which is capable of conclusively identifying denatured protein samples from an SDS-PAGE gel it is not possible to state anything for certain. It was not possible to collect these data due to the Covid-19 outbreak, however.

Both Ni²⁺-affinity chromatogram peaks for the AgrA_K101A_C199S mutant were subject to SEC. The peak which eluted at the highest percent buffer *B* % corresponds to AgrA_K101A_C199S monomer. The peak retention volume and profile is exactly the same as for both AgrA_C199S and AgrA_Y100T_K101A_C199S SEC peaks (fig. 4.5A and B). The peak which elutes at the lower-percent buffer *B* % gives a different chromatogram (fig. 4.5C).

Most of the sample elutes in the column void volume at ~ 45 mL and the corresponding SDS-PAGE sample band runs between the 58 and 80 kDa markers as it did after Ni²⁺-affinity chromatography. Two much smaller peaks are visible also, and have retention volumes of ~ 58 and 68 mL. SDS-PAGE analysis of these peaks indicate that they are probably the same species as they run at the same size at 25 kDa (fig. 4.6C) and may both correspond to degraded AgrA.

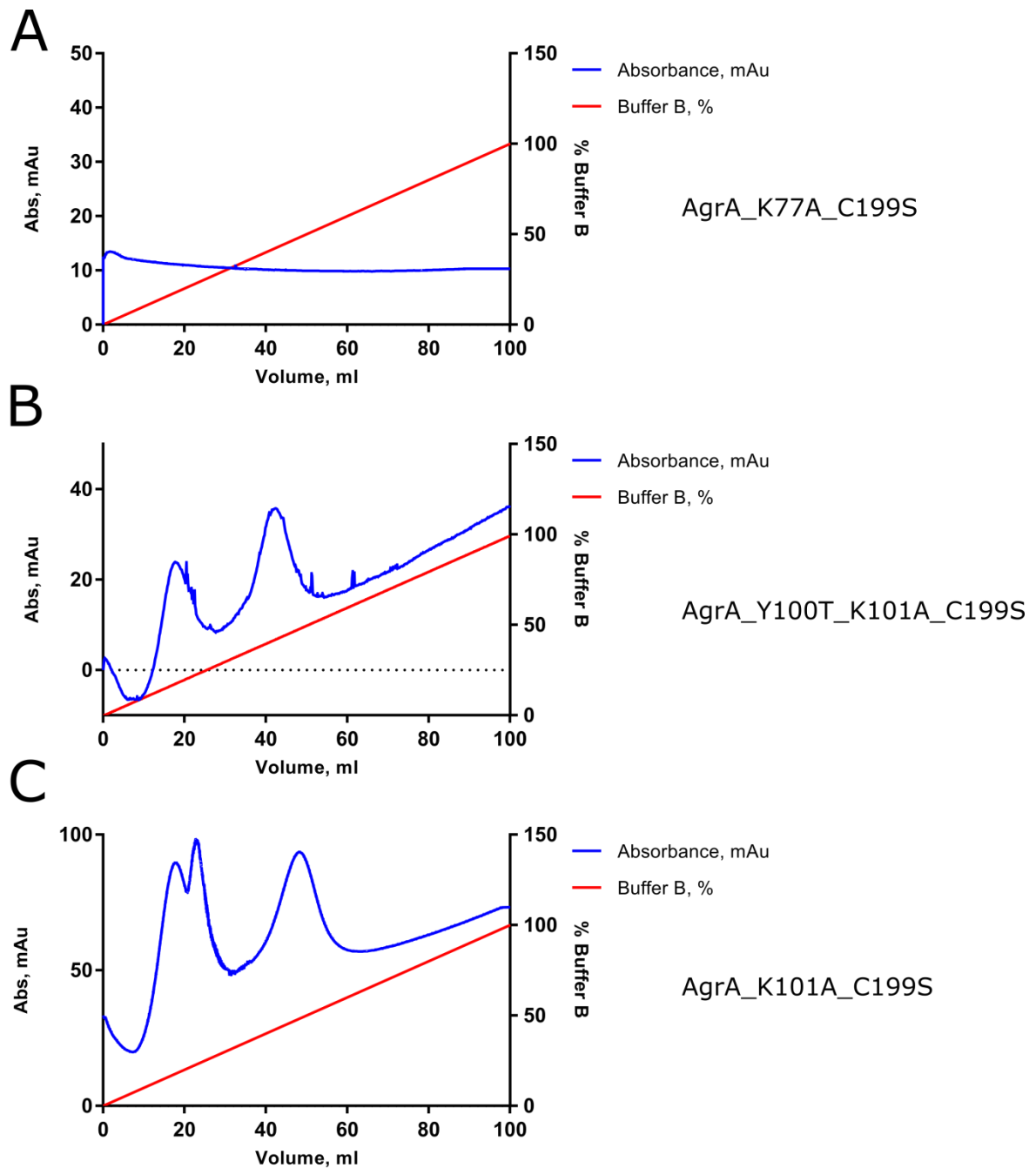


Figure 4.3. Ni²⁺-affinity chromatograms for AgrA_C199S SER mutants. (A). K77A. (B). Y100T_K101A. (C). K101A. The K77A mutant did not express and so the purification was retired at this step. The Y100T_K101A mutant behaved like the AgrA_C199S sample. The K101A mutant exhibits an atypical Ni²⁺-affinity chromatogram containing two peaks both eluting at percentages of buffer B. All of constructs contain the C199S mutation in addition those introduced for SER.

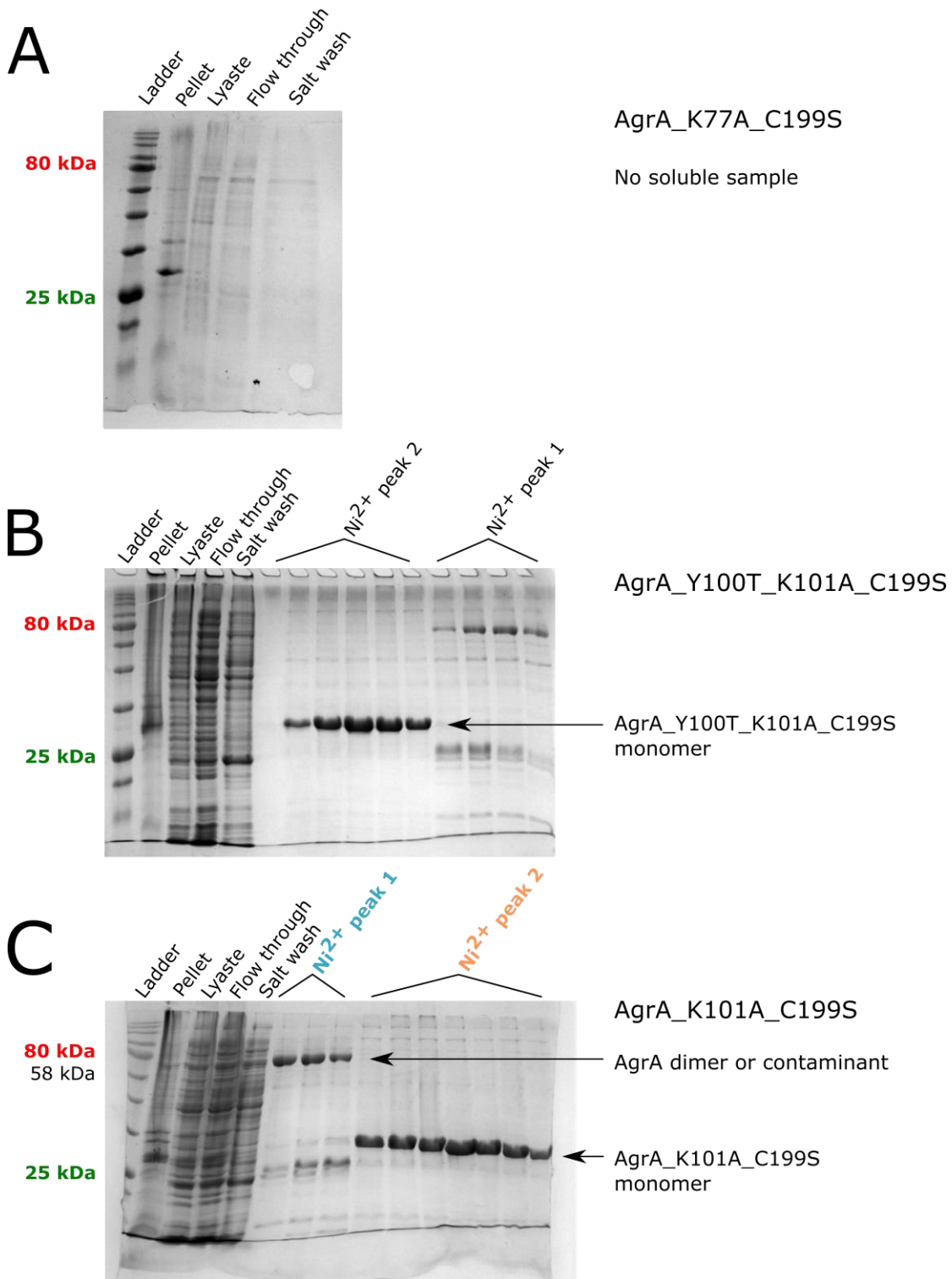


Figure 4.4. SDS-PAGE gels for AgrA_C199S SER mutants. (A). K77A (B). Y100T_K101A. (C). K101A. A small amount of AgrA_K77A is observed in the lane corresponding to the cell pellet in (A). Over expressed AgrA_Y100T_K101A and AgrA_K101A both run as monomers. The AgrA dimer or a contaminant is visible, particularly in (C). Gels are 16 % polyacrylamide, reducing.

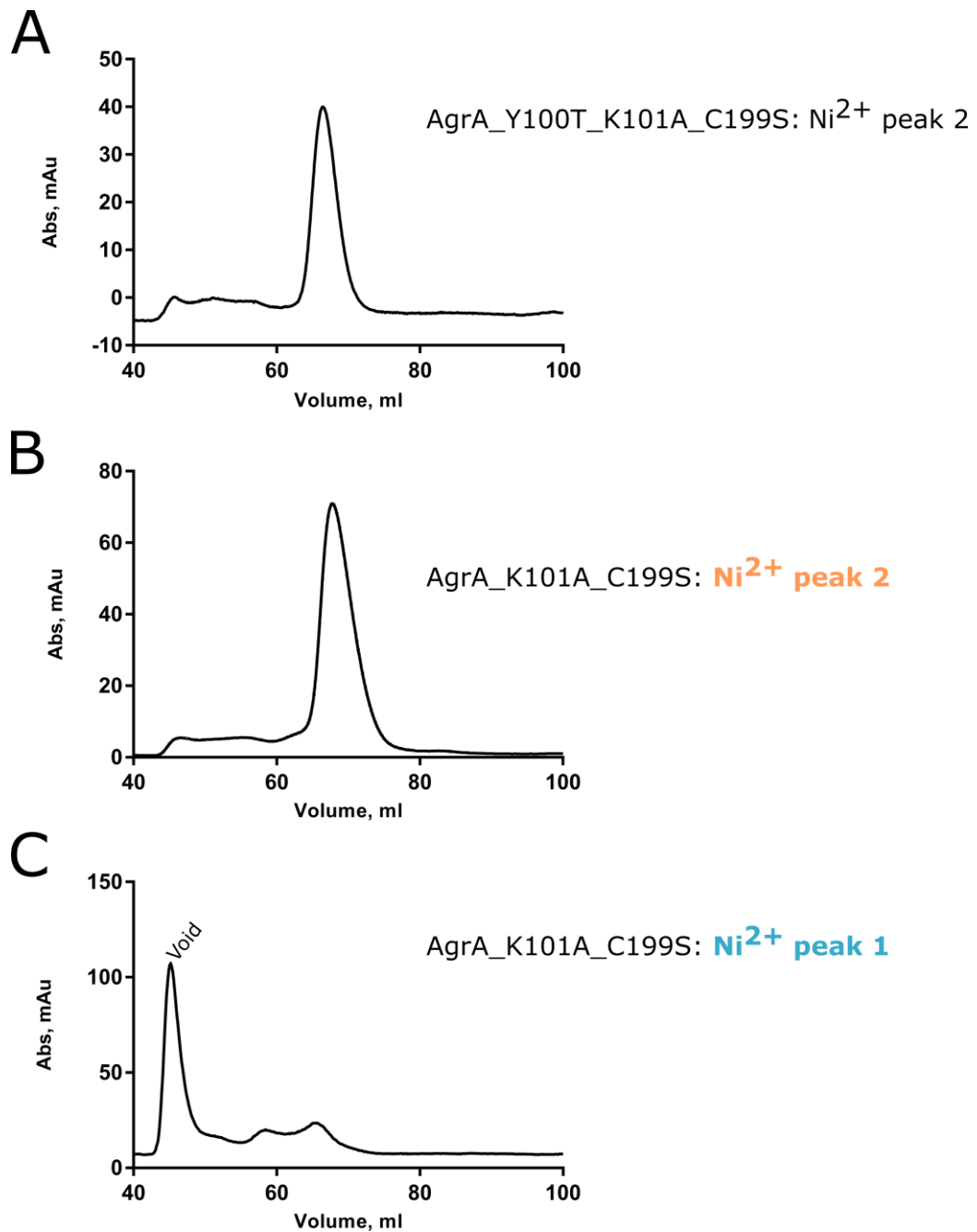


Figure 4.5. Size exclusion chromatograms for AgrA SER mutants. (A). Y100T_K101A. (B). The higher-percentage buffer *B* Ni²⁺-affinity chromatography peak from K101A. (C). lower-percentage buffer *B* Ni²⁺-affinity chromatography peak from K101. Y100T_K101A exhibits a peak corresponding to a monomer as does the K101A mutant obtained from the higher-percentage buffer *B* peak. Material from the lower-percentage buffer *B* peak elutes mostly in the void volume but yields two additional small peaks at ~ 58 and 68 mL which might correspond to degraded AgrA. The same running buffer and Hiload® Superdex® 75 16/600 column is used for all purifications.

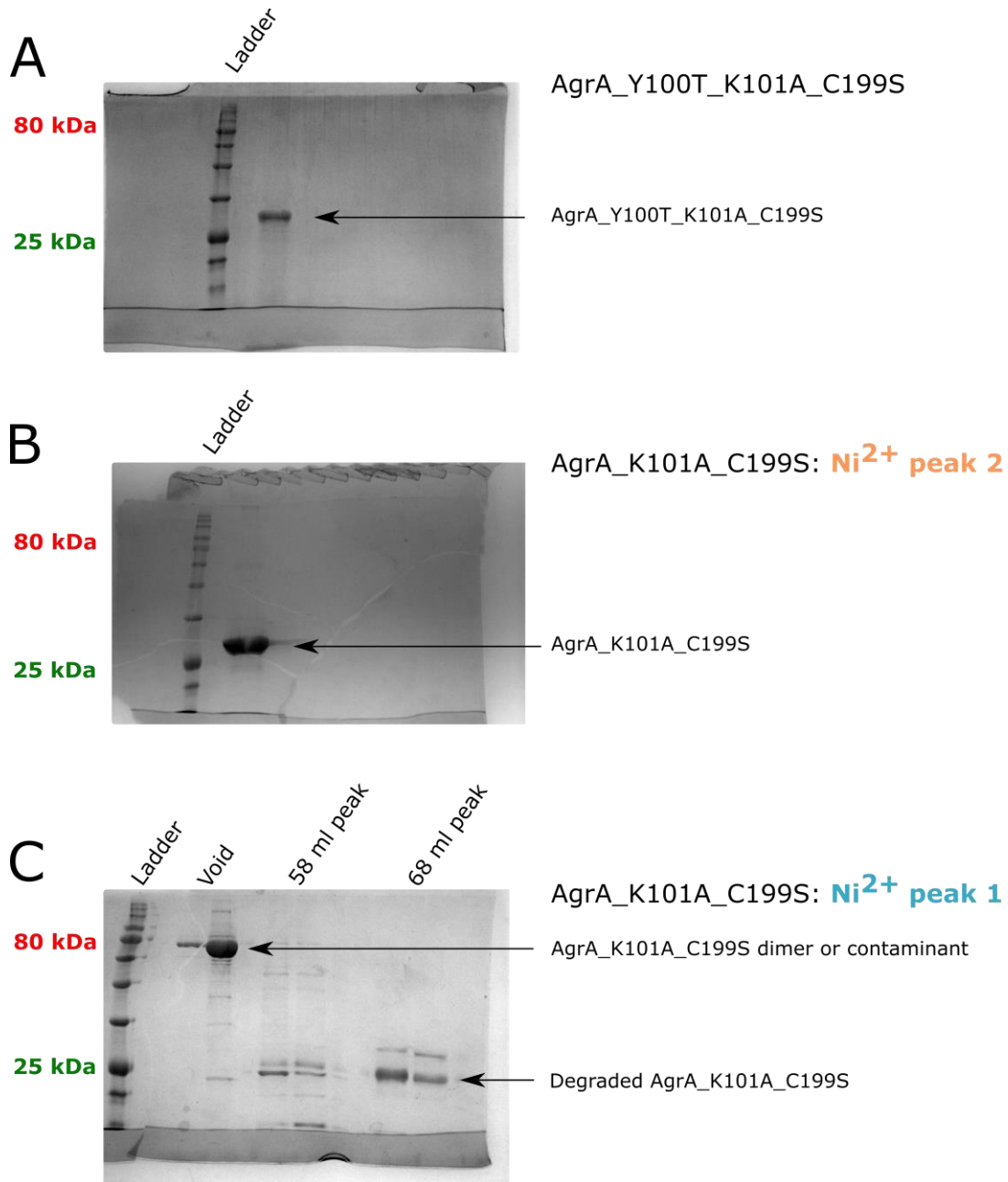


Figure 4.6. SDS-PAGE Gels for AgrA_C199S SER mutants. (A). Y100T_K101. (B). The higher-percentage buffer *B* Ni²⁺-affinity chromatography peak from K101A. (C). The lower-percentage buffer *B* Ni²⁺-affinity chromatography peak from K101A. Y100T_K101 exhibits excellent purity, as does the monomeric K101A. The bottom gel shows that the SEC void peak that corresponds to the band which runs between 58 and 80 kDa. The pairs of lanes correspond to the peaks at 58 and 68 mL respectively and seem to be degraded forms of a single species which runs at 25 kDa on SDS-PAGE. Gels are 16 % polyacrylamide, reducing.



4.2.1: Pursuit of Better AgrA Crystals via Additional Crystallisation Experiments

Additional crystallisation experiments were designed to provide either higher resolution data for the N-terminal domain to compliment the structure already presented, and to try to crystallise an entire AgrA dimer/promoter complex. The former would facilitate comparisons to be made between AgrA in phosphorylated and non-phosphorylated states, and perhaps mechanistic conclusions drawn which are impossible to infer from low-resolution electron density maps. The latter should provide a wealth of structural information regarding the physiological dimer/promoter complex including the exact nature of the dimerisation interface and the asymmetry between the two AgrA molecules.

4.2.2: AgrA_D59E_C199S/P2_{fragment}

Crystallisation experiments of this (constitutively active) sample were established in Morpheus®, PACT Premier™, MIDASPlus™ [Molecular Dimensions] and Nucleix [Qiagen] using the standard sitting drop vapour diffusion method. The concentration of the sample was 65.5 and 125 μ M with 2 molar equivalents of the DNA. The sample was worked-up identically to the AgrA_C199S/P2_{fragment} sample and the experiments were incubated at 20 °C. No crystals were observed from these experiments.

4.2.3 AgrA_C199S/P3_{full}

Complimentary single-stranded DNA (ssDNA) corresponding to the forwards and reverse sequences of the P3 promoter region, with complimentary unpaid 5' A/T overhangs were annealed to achieve an oligoduplex stock corresponding to the full P3 promoter region. This was incubated with purified AgrA_C199S in a 1.0:0.7 molar ratio to achieve a 0.2 molar excess of DNA after each DNA molecule had bound two AgrA_C199S molecules. The exact strategy for generating the oligoduplex stock and its subsequent incubation with AgrA_C199S can be found in the methods section 7.6.12. This sample was used to establish crystal screens in the following screens at Agr_C199S concentrations of 62.5 and 125 μ M: Morpheus®, Morpheus® II, Morpheus® III, PACT Premier™, MIDASPlus™, Stura Footprint Screens, JCSG Plus™, ProPlex™ [Molecular Dimensions], Nucleix, The PEGs Suite and The ComPAS Suite [Qiagen]. Two promising candidate crystals were observed in conditions



corresponding to Morpheus® well A12 and MIDASPlus™ well A11. Although later X-ray diffraction screening was to show that these were chemical crystals (data not shown) this was not known at the time of experimentation and so the conditions were used to establish the subsequent crystallisation experiments detailed below.

4.2.4: AgrA_C199S/P3_{full} and AgrA_C199S/P2_{full} With and Without [BeF₃]

Complimentary ssDNA corresponding to the forwards and reverse sequences of the P2 promoter region, with complimentary unpaired 5' A/T overhangs were annealed to achieve an oligoduplex stock corresponding to the full P2 promoter region. This stock solution was used to prepare a sample of AgrA_C199S/P2_{full} as already described above at AgrA_C199S concentrations of 62.5 and 125 μ M. Identical samples of AgrA_C199S/P3_{full} were also prepared in parallel. All four of these samples were divided in half, and one set was treated with 10 molar equivalents of [BeF₃]. Crystallisation experiments were then established for all eight samples in conditions corresponding to Morpheus® well A12 and MIDASPlus™ well A11 in 48 well sitting drop crystallisation experiments with 2 μ L experimental drops and 80 μ L reservoirs. These were incubated at 20 °C.

4.2.5: AgrA N-terminal Domain Only

A construct containing the first 141 amino acids of AgrA, AgrA₁₋₁₄₁, was prepared via inverse PCR using the construct used to purify full-length AgrA_C199S for crystallisation experiments, as the template DNA. This work was performed by Sara Zandomenighi, a PhD researcher within the *S. aureus* research group at the University of Nottingham. The Ni²⁺-affinity purified sample was provided to the experimenter with a view to dividing up the sample after SEC. The sample purified as a monomer with a retention volume of ~ 78 mL. This compares to ~ 70 mL for full-length AgrA_C199S when purified on the same column. Key purification data are presented overleaf in figure 4.7.

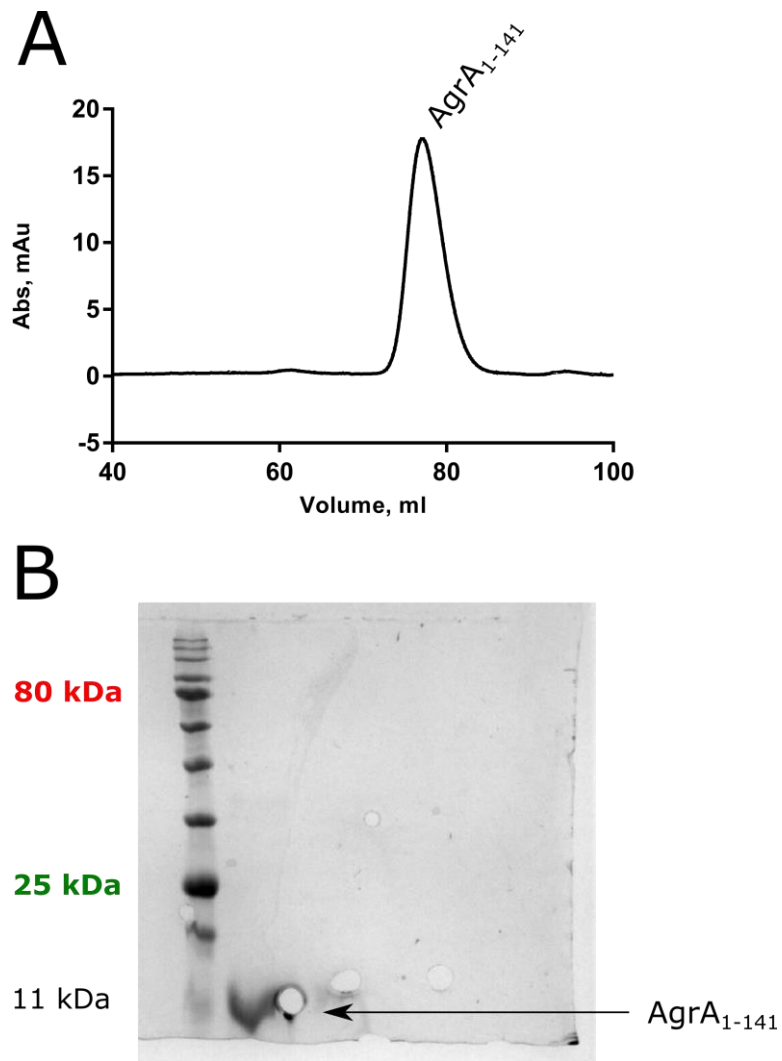


Figure 4.7. Purification data for AgrA₁₋₁₄₁. (A). Gel filtration chromatogram for AgrA₁₋₁₄₁ showing a retention volume of 77 mL, which corresponds to the monomeric state of the sample. The column used was a HiLoad® Superdex® 75 16/600 identical to the one used to purify full-length AgrA_C199S. (B). Purified AgrA₁₋₁₄₁ runs as a ~ 11 kDa species on the SDS-PAGE gel, and is somewhat obscured by a blister on the gel itself. The gel is 18 % polyacrylamide, reducing.



The sample was spin-concentrated 173 to μM , divided into two equal portions, one of which was diluted two-fold into SEC running buffer. Crystallisation experiments were established at both sample concentrations in the following screens: Morpheus®, PACT Premier™, MIDASPlus™, JCSG Plus™, ProPlex™, The LMB Screen™, MemGold™, and The Structure Screen [Molecular Dimensions] using standard sitting drop vapour diffusion methods. Crystals were observed in the following conditions: ProPlex™ well G3, MIDASPlus™ well E1 and JCSG Plus™ well E2. The compositions of these conditions are presented in table 4.2.

Two of the conditions are very similar in composition, containing 2 *M* ammonium sulphate at and near neutral pH. Of these two conditions, one contains the salt of cacodylic acid which has an arsenic atom at its centre and is therefore toxic. The condition corresponding to ProPlex™ well G3 was nominated for crystal optimisation therefore. A buffer control experiment was established using SEC running buffer only as the crystallisation sample. This did not generate any crystals or precipitate and so optimisation was continued. The concentration of ammonium sulphate in the crystallisation experiments was varied from 1.0 – 3.0 *M* in the 48 well sitting drop format with 2 μL experimental drops. All of these failed to reproduce the crystals however, even the control experiments which matched the composition of the original condition exactly.

The original hit crystals were cryocooled and screened for X-ray diffraction at Diamond Light Source. No X-ray diffraction was observed from the crystals. This indicates they are probably formed of protein and not simple salt or chemical crystals which would diffract synchrotron X-rays strongly. More time is required to reproduce and improve the crystals to see if any X-ray diffraction can be obtained from them. Pictures of the original hit crystals are presented in figure 4.8 overleaf.



Condition ID	Buffer	Salt	Precipitant
ProPlex™ G3	HEPES, 0.1 M, pH 7.0	None	Ammonium sulphate, 2 M
MIDASPlus™ E1	Tris, 0.1 M, pH 8.5	Li-citrate·4H ₂ O, 0.1 M	PPGBA* 400, 15 % v/v
JCSG Plus™ E2	Na-cacodylate, 0.1 M, pH 6.5	None	Ammonium sulphate, 2 M

Table 4.2. Summary of the conditions which afforded promising crystals resulting from sparse matrix crystallisation screening of AgrA₁₋₁₄₁. *PPGBA 400 is poly(propylene glycol) bis(2-aminopropyl ether).

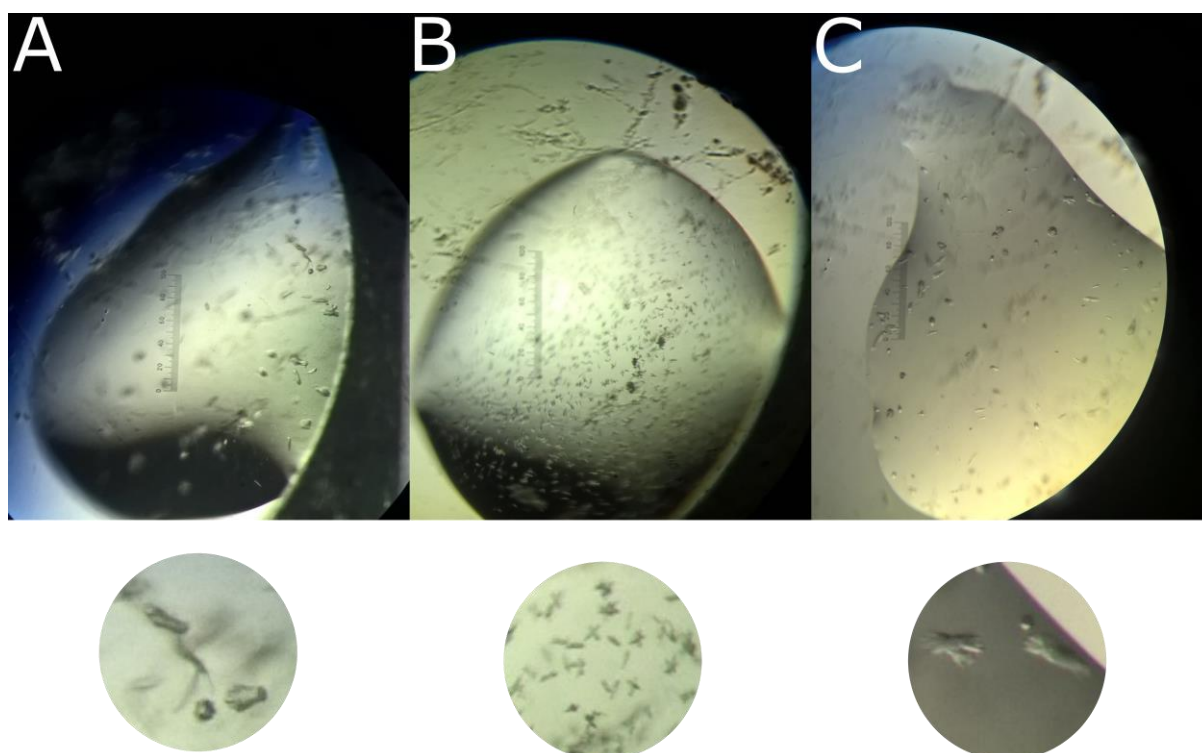


Figure 4.8. Pictures of the hit crystals obtained from sparse matrix crystallisation screening of AgrA₁₋₁₄₁ in the conditions indicated. (A) ProPlex™ G3. (B). MIDASPlus™ E1. (C). JCSG Plus™ E2. Respective zoom-ins are shown beneath. The leftmost crystals were fished and screened for X-ray diffraction but exhibited completely blank diffraction images.



4.3.1: Structure-Activity Experiments on AgrA_C199S

To try to validate the correctness of the AgrA_C199S/P2_{fragment} structure solution built at low resolution in the absence of higher quality diffraction data, and to try to give it some physiological context, several experiments were designed to probe the structure-activity relationships of AgrA. These delineated into three discrete threads: A) does the hydrophobic patch on surface play a role in AgrA dimerisation? B) What role, if any, does the proximal cysteine pair Cys55 and Cys123 play in a physiological setting? And C) can the available in-house AgrA inhibitor data be advanced upon?

In order to address these questions, four experiments were conceived: A) SAXS analysis to provide further information upon the nature of the AgrA dimerisation interface and the AgrA dimer/promoter complex. B) EMSAs under oxidative and reducing conditions to assess whether the aforementioned cysteine pair directly affects binding of AgrA to its promoters *in vitro*. C) *Lux*-based reporter assays to address whether mutating these cysteine residues affects turnover of the *agr* system *in vivo* and D) *in silico* docking of AgrA inhibitors developed in-house to generate hypotheses upon their mode of inhibition.

4.3.2: SAXS Experiments on the [BeF₃]-Activated AgrA Dimer/P3_{full} Complex

Some evidence for a hydrophobic patch on the surface of the N-terminal domain playing a role in AgrA dimerisation has already been presented. To briefly recap: the hydrophobic patch in itself is limited evidence since it will be higher in solvation energy than the rest of the surface; a crystallographic dimer that buries this patch against a copy of itself strongly resembles physiological dimers of structural AgrA homologues, and invokes the same secondary structure elements at the dimerisation interface; *in silico* docking of the AgrA N-terminal domains picks out a near-identical interface and *in silico* analysis of the crystallographic interface suggests that it is in or above the 97th percentile of energetically favourable dimers that the N-terminal domains can form.

AgrA is a strong binder of DNA even in the absence of phosphorylation.^[85] Although there are differences in the affinities of AgrA for the P2 and P3 promoter regions, and these affinities become higher upon phosphorylation, the increase is from low nanomolar to high picomolar (K_d). Preparing



AgrA/DNA complexes is trivial then provided the recombinant AgrA sample is folded, since the sample concentration is invariably orders of magnitude above the K_d during purification.

Isolating an AgrA dimer was eventually achieved by incubating AgrA with an oligoduplex corresponding to the full P3 promoter region, followed by incubation with excess $[\text{BeF}_3]^-$. Formation of this complex was confirmed by analytical SEC and presence of both protein and DNA in the sample peak was confirmed by SDS-PAGE and checking that the absorbance maxima occurred at a wavelength of 260 nm.

Such a sample has already been discussed as being used for crystallisation studies (section 4.2.4). The same sample work-up procedure and analysis was carried out for the crystallisation sample as for the SAXS sample. As the crystallisation experiments on this complex were unsuccessful, however, the sample quality control data is presented here in figure 4.9 with the SAXS data but applies to both samples. During a test purification, before submitting the sample for SAXS, $[\text{BeF}_3]^-$ was applied to the AgrA_C199S sample in one hundred-fold molar excess with 0.6 molar equivalents of DNA. These conditions were found to be too austere however, resulting in precipitation of $\sim 30 - 40\%$ of the sample. This was circumvented in the actual SAXS sample by applying $[\text{BeF}_3]^-$ in ten-fold molar excess and 1 molar equivalent DNA (exactly twice as much DNA needed to bind of all of the AgrA_C199S molecules given that the reaction has a 2:1 ratio of AgrA_C199S:P3 DNA).

The SAXS data were collected at Diamond Light Source beamline B21 by Dr. Alice Goode. Key charts are presented in figure 4.10. The Guinier plot^[295] is linear which is indicative of scattering from a monodisperse sample.^[296] The radius of gyration determined from the Guinier plot and the pair-distance distribution, P_r , plot agree and are 51.3 and 51.6 Å respectively, which is indicative of good quality data.^[297] The maximum dimension of the complex, r_{max} , was determined as 176 Å. The scattering data provided readily-interpretable molecular envelopes (fig. 4.11) when processed with DAMMIF^[298] and DAMMAVER,^[299] which are available as part of the ATSAS suite^[300] (vers. 3.0.1) [EMBL].

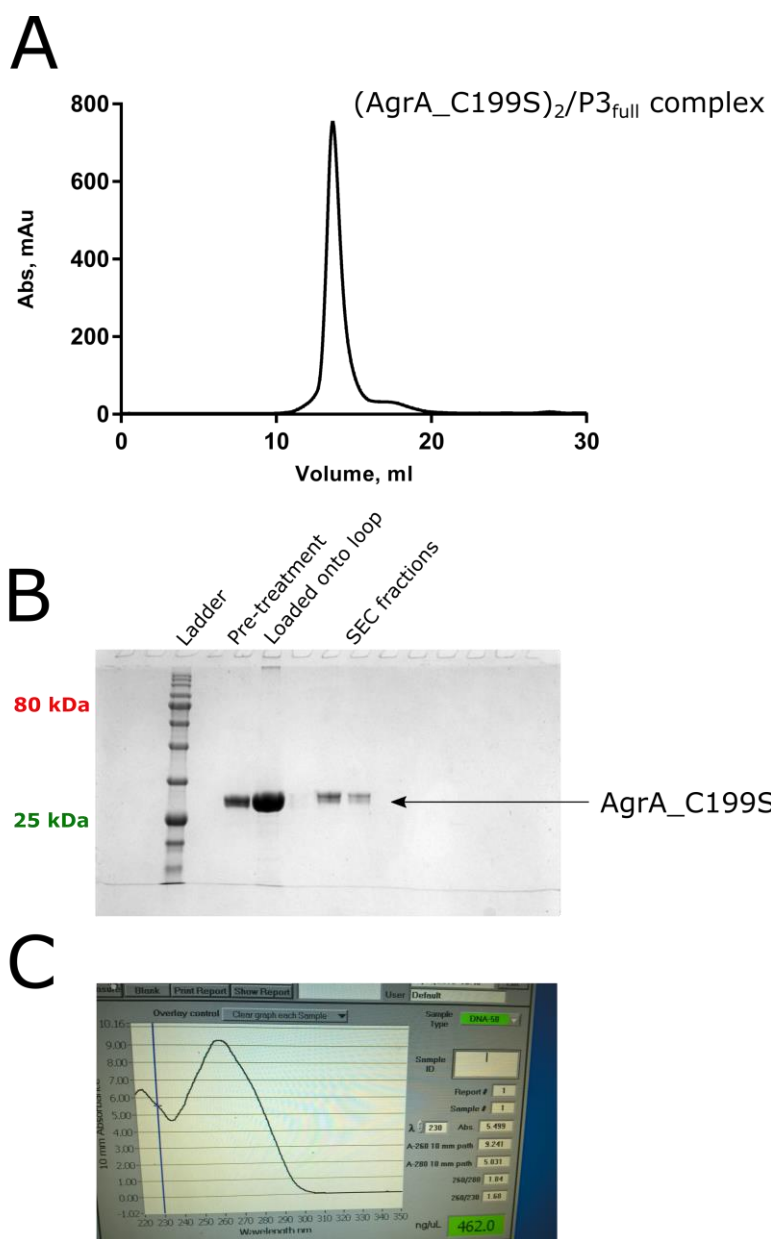


Figure 4.9. Purification data for the (AgrA_C199S)₂/P3_{full} complex. (A). the size exclusion chromatogram. The column used was a Superdex® 200 Increase 10/300 GL column and the running buffer is the same as for the AgrA_C199S sample used for crystallisation experiments. (B). SDS-PAGE analysis of the sample before any treatment, after treatment with DNA, [BeF₃]⁻ and spin-concentration, and after the analytical SEC step. (C). Nanodrop 1000 readout for the sample after analytical gel filtration showing an absorbance maximum at 260nm which betrays the presence of DNA in the sample.

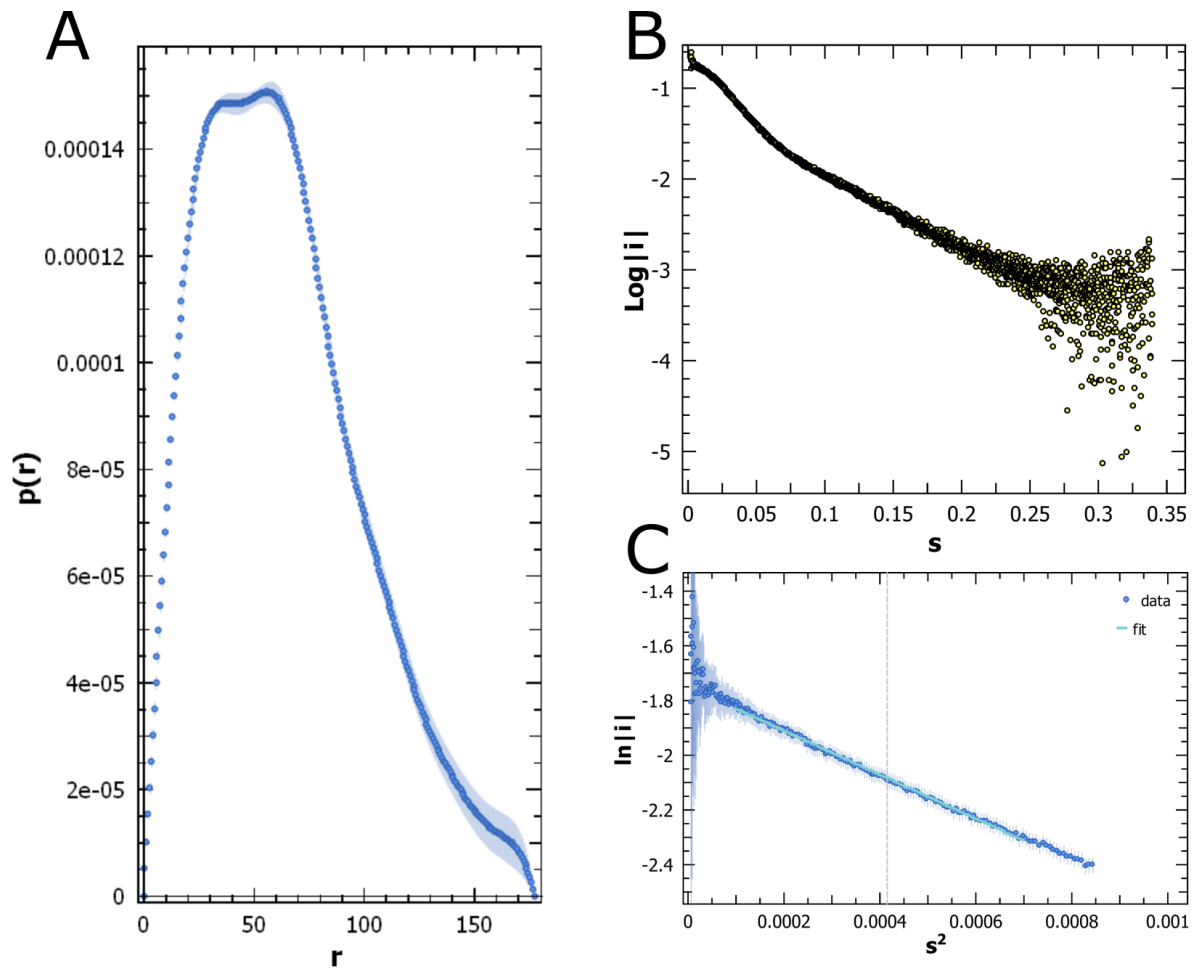


Figure 4.10. Key charts obtained from the buffer-subtracted SAXS data. (A). The pair distance distribution plot of P_r against r . The P_r function is approximately Gaussian but is slightly left-biased which is indicative of a rod-shaped particle.^[301] P_r has no units and r has units \AA . (B). A plot of the log of the scattering intensities, $\log|I|$, against the scattering vector, s . Logarithms have no units and s has units \AA^{-1} . (C). A Guinier plot of the natural logs of the scattering intensities, $\ln|I|$, against the square of the scattering vector, s^2 . The parameter s^2 has units \AA^{-2} . The plot is linear which is indicative of scattering from a monodisperse sample.

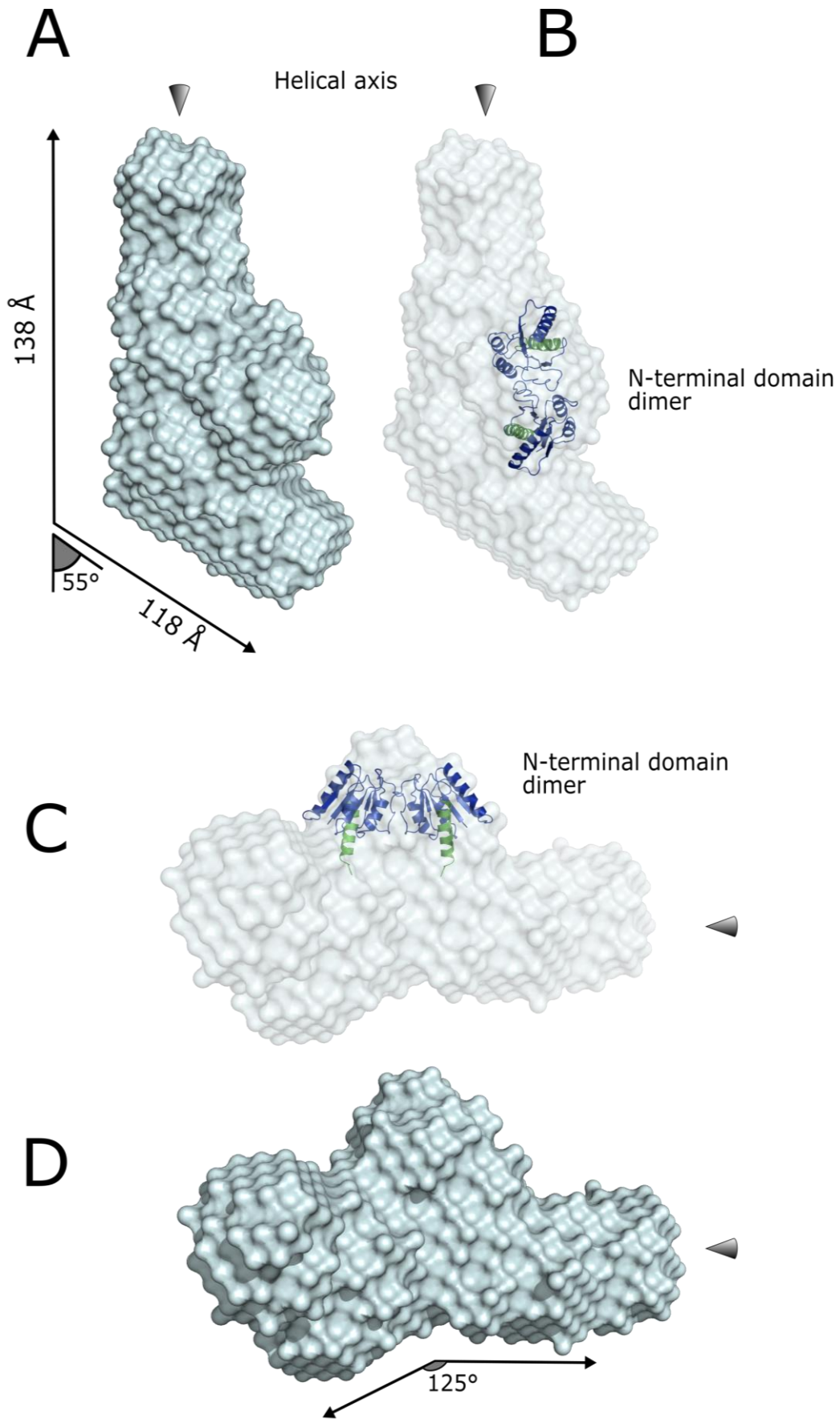




Figure 4.11 (previous page). SAXS-constructed molecular envelope of (AgrA_C199S)₂/P3_{full} prepared using DAMMIF^[298] and DAMMAVER,^[299] which are available as part of the ATSAS suite^[300] (vers. 3.0.1) [EMBL]. (A). Top-down view of the molecular envelope showing a sharp 55 ° bend and a protrusion, corresponding to the AgrA_C199S N-terminal dimer, lying above the inside apex of this bend (that is, pointing directly up out from the page towards the reader). (B). The same as (A) but with the envelope shown as a transparent surface and with the crystallographic AgrA_C199S N-terminal dimer discussed in section 3.9 modelled into the protrusion. The protein is shown as a dark blue cartoon with the two α5s, which immediately precede the C-terminal domains, coloured green for clarity. (C) and (D). The same two representations (A) and (B) respectively but viewed from the inside of the bend introduced to the DNA by AgrA_C199S. The angle and distance measurements were made in PyMOL™. The corresponding maximum particle dimension would be: $118\cos(55) + 138 = 205 \text{ \AA}$. This is within 20 % of the r_{\max} calculated directly from the P_r plot.



The path of the DNA helix is readily traceable within the molecular envelope. The AgrA_C199S dimer introduces a sharp 55 ° bend into the DNA molecule, indicating that the DNA-binding domains bend the DNA in a cooperative manner. DNase I protection assays have identified hypersensitive bases on the P3 promoter once bound to AgrA^[81] which is indicative of DNA curvature^[83] and the envelope presented here agrees with these findings. The DNA-binding domains themselves are not clearly visible within the envelope of the complex. There is, however, a distinct protrusion pointing out from the DNA, demonstrating unequivocally that the dimerisation interface does indeed occur between the N-terminal domains of AgrA. The crystallographic dimer discussed in section 3.9 was modelled into the protrusion using PyMOL™ and exhibits an excellent fit to the molecular envelope.

Another feature which can be observed in the envelope is that both AgrA_C199S molecules lie to one side of the DNA, and the bend they impart is perpendicular to the plane of the AgrA_C199S N-terminal domains. This is in marked contrast the crystal structure of the KdpE dimer bound to its promoter region^[91] (PDB: 4kny) in which the entirety of the KdpE dimer and the bend imparted into the DNA all occupy a single plane. The architecture of the KdpE DNA-binding domains is winged helix-turn-helix however, not LytTR-type.

The non-co-planarity observed in the envelope implies that the 95 ° angle between N- and C-terminal domains of the AgrA_C199S molecule observed in crystal asymmetric unit is present in the physiological dimer also. This feature may be characteristic of the way the LytTR family of proteins bend DNA, or a unique feature of the AgrA dimer/P3 promoter complex. Additional structural information regarding LytTR-type response regulators in complex with DNA would need to become available to see if any trends became apparent. This SAXS data and the full-length AgrA_C199S structure solution are the first structural data for any full-length LytTR-type DNA-binding protein in complex with DNA however, so there is nothing against which to make direct comparisons.

4.3.3: Oxidising and Reducing EMSAs to Probe Cys55 and Cys123 *In Vitro*

The strategy for the EMSAs involved incubating low μM quantities of acetyl phosphate-activated AgrA_C199S with 10 nM of 5' cyanine5.5-labeled DNA probe corresponding to the P3 promoter



region and a running a native PAGE gel in tris/borate/EDTA (TBE) buffer at low currents at 4 °C. The samples were split into two test populations which were treated separately with excess DTT or H₂O₂ prior to titrating in the DNA probe. The C199S point mutant in the sample acted as a pre-made control for the proximal cysteine pair in the C-terminal domain of AgrA (Cys199 and Cys228).

In the initial experiments, the oxidised and reduced samples were run on the same gel, but this was subsequently changed such that each test case was ran on a separate gel so as to avoid diffusion of the redox reagents, which could tarnish the results. The resulting gels were clear but all of the experiments were complicated by the failure of the DNA probe to fully anneal, and consequently running as a doublet of two single strands of DNA on the gel rather than a single band of dsDNA.

The ssDNA is too flexible for AgrA_C199S to bind and so only faint bands were observed for the gel shifts since only a small fraction of the DNA probe existed as dsDNA. This was not a problem with respect to AgrA_C199S activity which had been proven several times via SEC with different oligoduplexes for crystallisation and SAXS, but it did prevent conclusive results being drawn from the EMSAs. The probe was reordered with a view to repeating the experiments but it did not arrive prior to the outbreak of Covid-19 and so they have not yet been performed. Two representations of the gels resulting from the same EMSA experiment are shown in figures 4.12 and 4.13 **overleaf**.

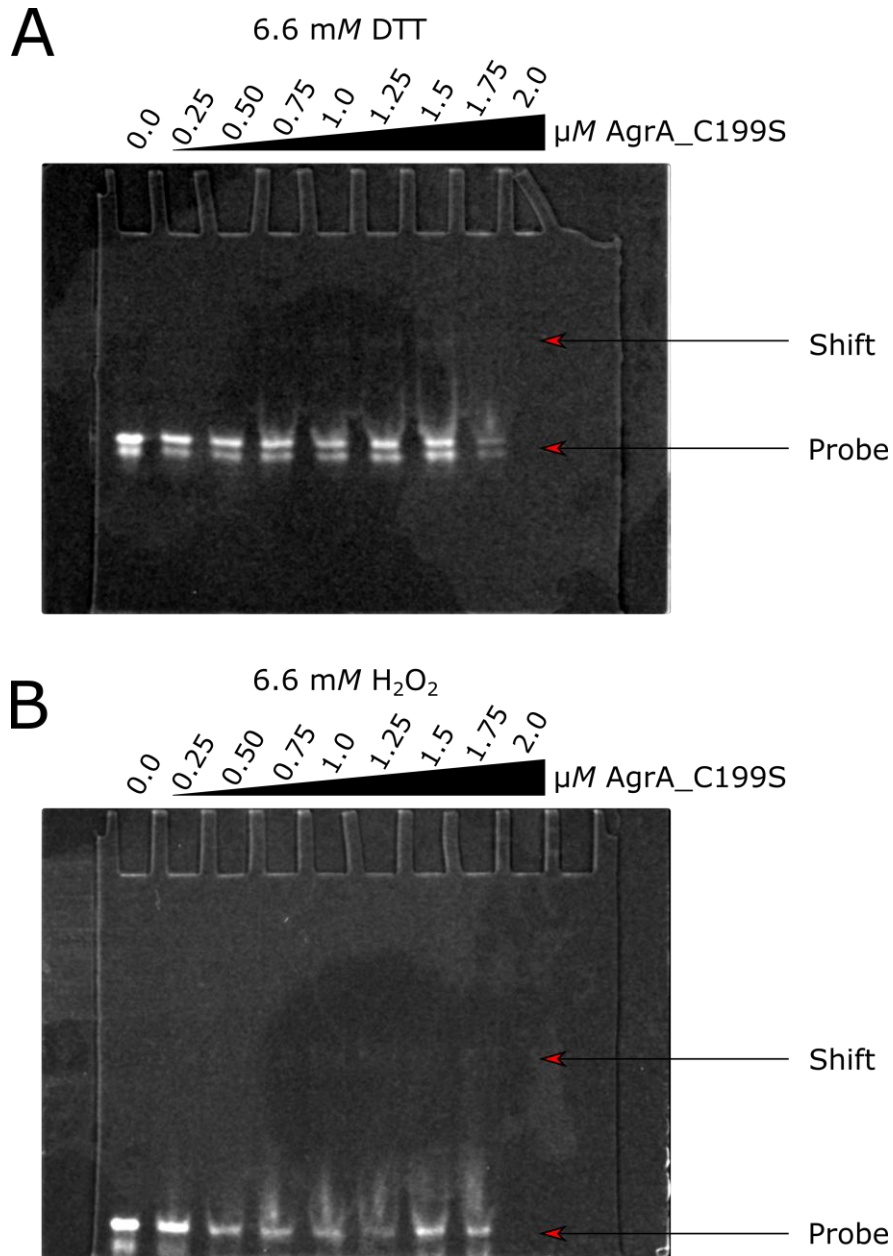


Figure 4.12. EMSA results. (A). Reducing EMSA. (B). Oxidising EMSA. Both results are visualised as negatives with identical exposures and no digital manipulation. The probe runs as a doublet for both experiments limiting the intensities of the bands corresponding to the gel shifts, which are visible but faint. Gels are 6 % acrylamide (prepared from a 30 % solution of 19:1 acrylamide:bis-acrylamide, not 37.5:1 which is used to prepare SDS-PAGE gels). Note that the shifts may not have printed well.

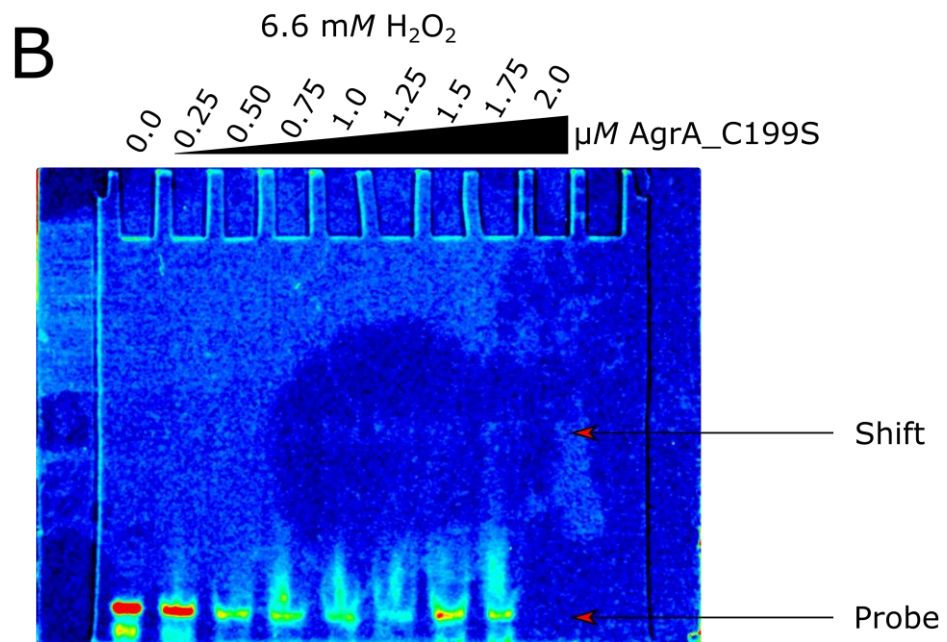
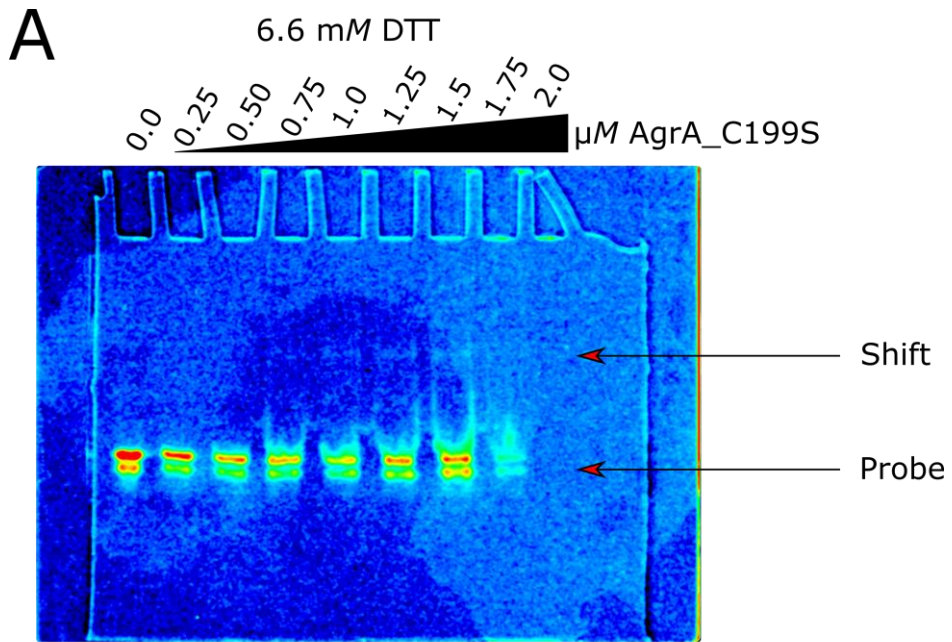


Figure 4.13. EMSA results. Gels are the same as the previous page but viewed as heat maps. Again no digital modification was made to the images. Note again that the gels may not have printed well.



4.3.4: *Lux*-based Reporter Assays to Probe Cys55 and Cys123 *In Vivo*

A robust *lux*-based reporter assay is available to the *S. aureus* research group.^[302] In the assay the *agr* P3 promoter is tethered to a luciferase gene and the *agr* system may be activated in a AIP dose-dependant manner. Bioluminescence can be trivially recorded and normalised against cell Optical Density at $\lambda = 600$ nm (OD_{600nm}) to get an accurate determination of perturbations in *agr* regulation. Measuring light output is arbitrary and so every measurement must be compared to a control in which the *agr* system, apart from the gene which it has been forced to upregulate (*lux*) is WT. Cell density was simultaneously quantified in order to normalise the light output to the abundance of cells in each of the test conditions and assess any perturbations in cell growth resulting from modifications to the *agr* genetic code.

The crystal structure determined in this work informed a number of point mutations that may perturb *agr* function when introduced into AgrA in the bioreporter assay. The mutants generated were: AgrA_C55S, AgrA_C123S, AgrA_C199S, AgrA_C228S, AgrA_D59N (constitutively inactive AgrA), AgrA_D59N_C123S and AgrA_D59N_C199S. At the outset it should be duly noted that all of bioreporter work was performed by Dr. Ewan Murray whose research primarily involved *in vivo* studies of *agr*, and not recombinant protein work or biophysics etc. The data presented here are his and are presented with permission.

The mutations involving Cys55 and Cys123 were generated to test the redox dependency of the cysteine pair observed in the N-terminal domain of AgrA. The C199S mutant meanwhile is known to enhance AgrA DNA-binding activity,^[208] however, it has not been reported whether this superactivity is dependant on prior phosphorylation of AgrA in order to occur, or whether the C199S mutant is truly constitutively (super)active. It was therefore interesting in itself to study and a good benchmark against which to compare the Cys55 and Cys123 mutations to.

It was found to be impossible to introduce mutants that simultaneously disrupted both disulphide bridges, suggesting that at least one is required to prevent toxicity to the host cells. No mutations were made for Cys6 which lies well away from the other cysteines in the N-terminal domain (the CB...CB



distances are 11.6 Å and 16.3 Å for Cys6 and Cys55, and Cys6 and Cys123, respectively). Similarly, the AgrA_D59E mutant (constitutively active) was also found impossible to generate in *S. aureus* cells and toxic when transformed into *E. coli* cells. Placing it under the control of a silent promoter ameliorated these effects but was of limited experimental value.

There is no dramatic effect upon activation of the P3 promoter when the potential N-terminal disulphide bond in the N-terminal domain is disrupted (C123S). The respective half-maximal effective concentration (EC₅₀) curves match that of WT AgrA. This is unlike the C199S mutant, which displays increased levels of activation and ~ five-fold lower EC₅₀ values when compared to WT AgrA (fig. 4.14). When in combination with D59N (constitutively inactive) both C123S and C199S mutants display significant growth retardation however, whilst all of the single mutants grow at the same rate as WT AgrA.

This data suggests that AgrA plays a role in the *S. aureus* cell lifecycle during *agr* dormancy, and that both of the disulphide bonds present in AgrA are important to this role. Furthermore the growth defect imparted by mutating them occurs independently of AIP concentration (figure 4.15). However, it remains unclear how the N-terminal domain disulphide bond is important, i.e. whether its mutation abrogates DNA-binding function or simply affects the stability of AgrA. These preliminary findings also corroborate the full-length AgrA model as it is built, by demonstrating that the Cys55 and Cys123 pair must be close so as to form a disulphide bridge that is disruptable in such a manner that produces a measurable effect (retarded cell growth). The phosphorylation-dependency of the C199S mutant superactivity compared to WT AgrA is also a novel finding.

Inevitably there are more mutant combinations which could be tested which may yet elucidate more layers of complexity in the *agr* system imparted by AgrA. This thread of research was also disrupted by the Covid-19 outbreak and is ongoing at the time of writing.

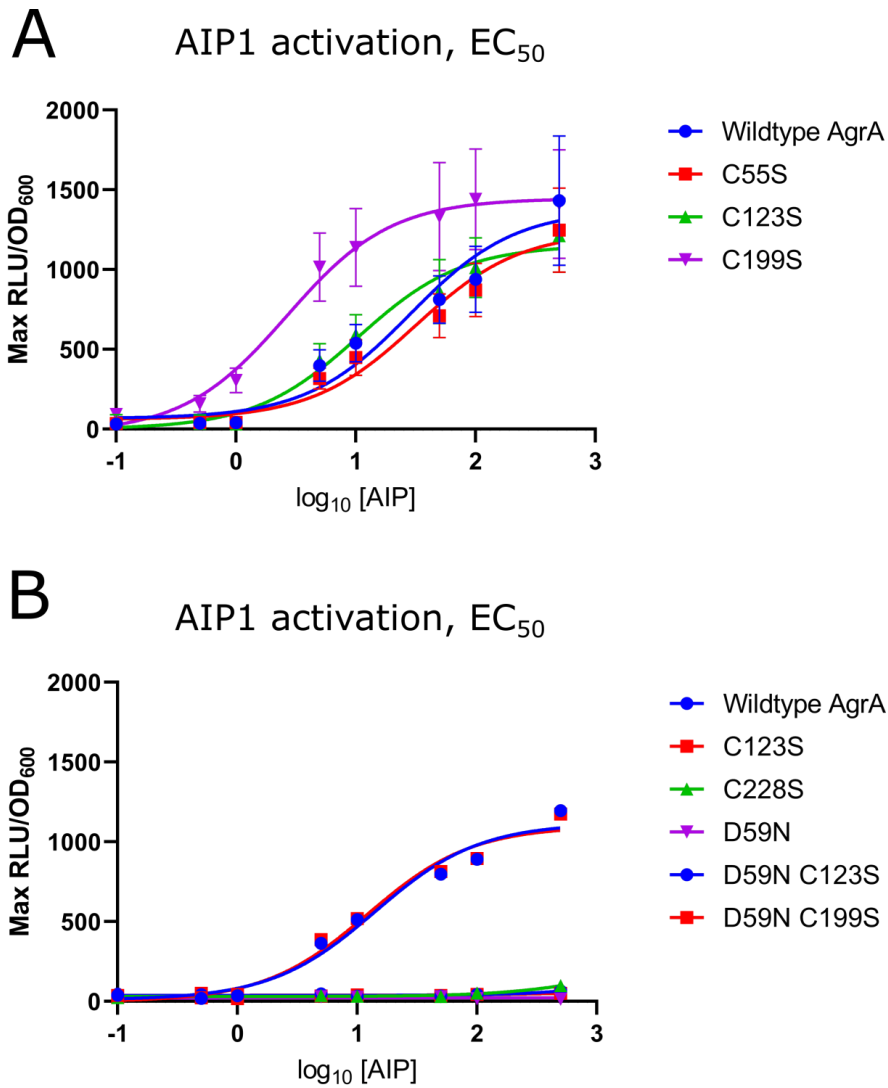


Figure 4.14. EC_{50} Curves for series of AgrA mutants showing normalised light output as a function of the log of the dose of AIP (which have no units). (A). Single mutants and WT only. (B). Single and double mutants plotted together. Work upon the double mutant EC_{50} curves is unfinished due to Covid-19 (hence the lack of error bars). All of the D59N mutants show abolished *agr* activity as does the C228S mutant. The C199S mutant is superactive whilst still requiring some AIP to initiate activity. RLU stands for Relative Light Units (which have arbitrary units).

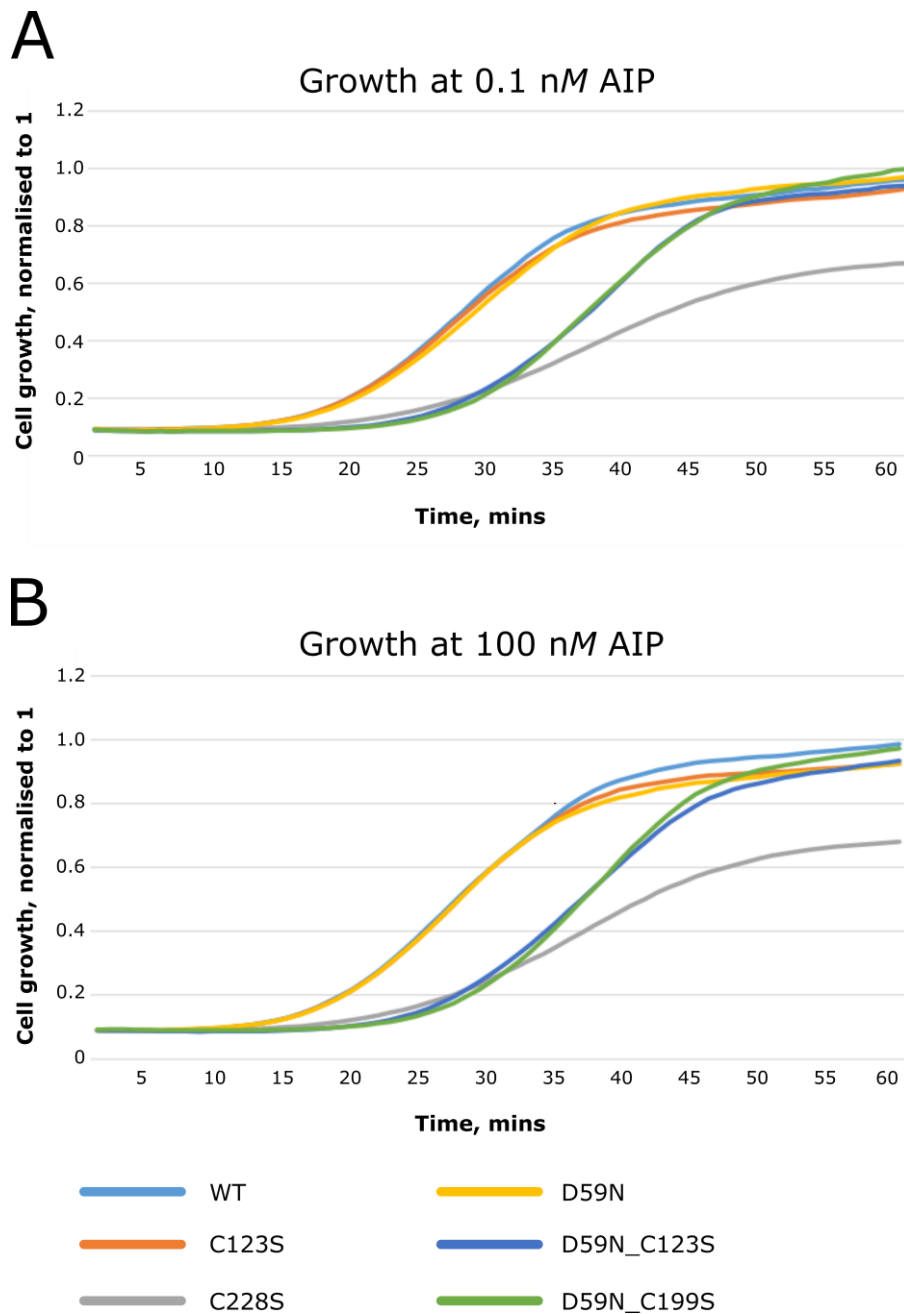


Figure 4.15. Growth curves for a series of AgrA mutants showing cell growth normalised to 1 for the WT AgrA population as a function of time (in mins). Top. (A). Cell growth at 0.1 nM AIP. (B). Cell growth at 100 nM AIP. The C228S mutant exhibits the most drastic growth retardation and does not appear to reach the same final cell population number as the other mutants. The D59N_C123S and D59N_C199S double mutants show less drastic growth retardation and appear to reach the same final population number as WT AgrA. All effects appear to be invariant with respect to AIP concentration.



4.3.5: Docking of IQS Analogues to AgrA

4.3.5.1: Preliminary Work on IQS Analogues

Both being commensal and opportunistic pathogens, *S. aureus* and *P. aeruginosa* compete with each other for the nutrients required to proliferate and infect hosts that are suffering from an immunocompromising illness, usually one that affects the respiratory system such as cystic fibrosis.^[303]

P. aeruginosa secretes a variety of anti-staphylococcal compounds such as 4-hydroxy-2-heptylquinoline-*N*-oxide^[304] which alters the metabolic profile of *S. aureus*, to produce the so-called small-colony variant of the pathogen. This variant displays reduced alpha-toxin production and slower growth than the regular-colony form.^[305] The QS signal molecule secreted by *P. aeruginosa*, *N*-(3-oxododecanoyl)-*L*-homoserine lactone, is an antagonist of the *S. aureus agr* system and leads to reduced virulence factor production.^[306] Pyochelin meanwhile is a secondary metabolite and iron-chelating agent produced by *P. aeruginosa* that allows the cells to ingest iron by transporting the chelated ions to the outer membrane transporter protein FptA.^[307] Pyochelin biosynthesis produces 2-(2-hydroxyphenyl)-thiazole-4-carbaldehyde (also known as aeruginaldehyde and abbreviated to IQS) as a by-product and there is some evidence that it has antimicrobial properties against some phytopathogens.^[308]

Research undertaken by previous PhD student Dr. Yanin Jaiyen demonstrated that IQS also exhibited anti-staphylococcal properties. Specifically, it was shown that IQS could reduce α -haemolysin expression in *S. aureus* in a highly strain-specific manner. In attempting to broaden the range of *S. aureus* strains against which IQS was efficacious and/or increase their potency as inhibitors, chemical synthesis of fourteen IQS analogues was performed. Of these analogues, IQS3 and IQS11 were more potent than IQS with respect to α -haemolysin expression in: group I *S. aureus* strain KH1187A, group II *S. aureus* strains TS2 and TS5 and group IV *S. aureus* strain TS3. The target of IQS was eventually determined to be AgrA, as 1 – 100 μ M IQS3 was shown to reduce light output from an AgrA bioreporter strain that could be induced independently of AgrC.



That AgrA was the target of the IQS analogues was further validated *in vitro* via EMSA experiments using recombinant AgrA and a DNA probe corresponding to the *agr* P2 promoter. Total abolition of the AgrA/P2 promoter complex was observed when 1 mM IQS3 in 10 % DMSO was titrated into a mixture of 600 nM AgrA and 10 nM P2 probe. Other IQS analogues required even higher concentrations to achieve abolition of the AgrA/P2 promoter complex.

That a by-product of secondary metabolite synthesis in *P. aeruginosa* was shown to inhibit *S. aureus* AgrA has important corollaries within the field of staphylococcal AMR, however, the data is limited by the low binding affinity of the IQS derivatives for AgrA and the lack of any data evidencing their binding site, which in turn precludes hypotheses upon their mechanism of inhibition. To overcome this, efforts were made (this work) to determine the binding site of these IQS analogues.

4.3.5.2: Co-crystallisation Attempts of IQS Analogues with AgrA

Efforts were made to experimentally determine the binding site of the IQS analogues via co-crystallisation studies with full-length AgrA_C199S. This approach was fundamentally limited, however, by the aforementioned low binding affinity of the IQS analogues, and compounded by the relatively high concentrations of AgrA_C199S required in the crystallisation experiments (62.5 μ M). The limited solubility of the IQS analogues at elevated concentrations and the requirement of 10 % DMSO to solubilise them prevented them from being titrated into the crystallisation experiments in a degree of excess comparable to those in EMSA experiments performed by Dr. Jaiyen (approximately fifteen hundred-fold). The presence dehydrating polymers and other polyols in the crystallisation experiments would have further decreased their power to solvate the IQS analogues, and taken together these reasons explain why all of the experiments resulted in immediate heavy precipitate being formed. Soaking of the IQS analogues into the AgrA_C199S/P2_{fragment} crystals was deemed an invalid strategy since the analogues function as DNA-binding inhibitors of AgrA and these crystals contain DNA bound to AgrA_C199S.

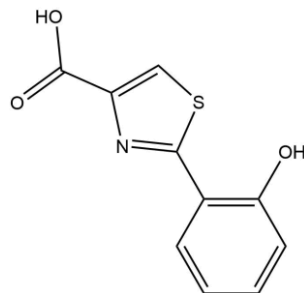
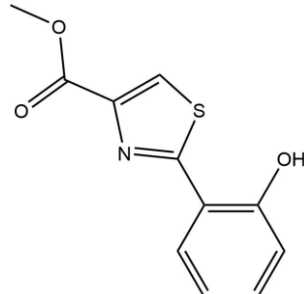
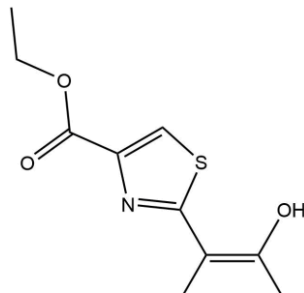
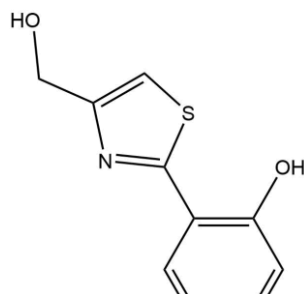
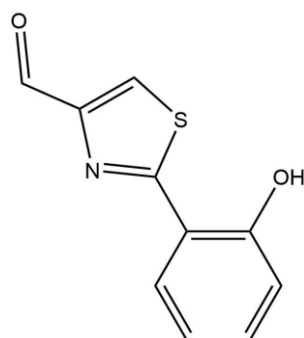


4.3.5.3: In Silico Docking of IQS Analogues to AgrA Using AutoDock Vina

In the absence of experimental data, an *in silico* docking approach was nominated as an alternative method to provide some information regarding the binding site of the IQS analogues. Whilst data generated *in silico* cannot alone be considered evidence for a given binding site, it does allow the formation of hypothesis that can be tested in the laboratory.

The docking experiments were performed using AutoDock Vina^[309] (vers. 1.1.2) in conjunction with MGL Tools^[310] (vers. 1.5.6) [both developed by the MGL lab, Scripps Institute] using the *apo* structure of the C-terminal domain of AgrA^[92] (AgrA_C) (PDB: 4g4k) as the receptor molecule. All fourteen IQS analogues and IQS itself were docked, and the entire of surface of the receptor was included in the search space. Each docking experiment returned the nine results, ranked by their corresponding predicted binding energies, thus generating in a total of 135 AgrA_C/IQS analogue complex “structures”. The skeletal structures of the fourteen IQS analogues and IQS itself can be found in table 4.3 and the log files, containing a full list of the predicted binding energies for each of the fifteen docking experiments, can be found in appendix IV.



IQS#	Skeletal structure	IUPAC Name
IQS1		2-(2-hydroxyphenyl)thiazole-4-carboxylic acid
IQS2		methyl 2-(2-hydroxyphenyl)thiazole-4-carboxylate
IQS3		ethyl 2-(2-hydroxyphenyl)thiazole-4-carboxylate
IQS4		2-(4-(hydroxymethyl)thiazol-2-yl)phenol
IQS5 (IQS)		2-(2-hydroxyphenyl)thiazole-4-carbaldehyde



IQS#	Skeletal structure	IUPAC Name
IQS6		(S)-2-(2-hydroxyphenyl)-4,5-dihydrothiazole-4-carboxylic acid
IQS7		methyl (S)-2-(2-hydroxyphenyl)-4,5-dihydrothiazole-4-carboxylate
IQS8		ethyl (S)-2-(2-hydroxyphenyl)-4,5-dihydrothiazole-4-carboxylate
IQS9		2-phenylthiazole-4-carbaldehyde
IQS10		2-(2-hydroxyphenyl)thiazole-4-carboxamide



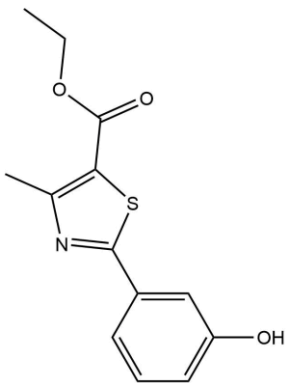
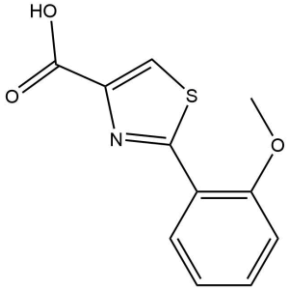
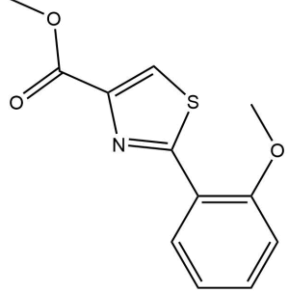
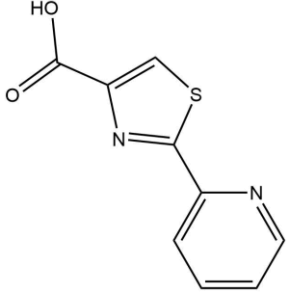
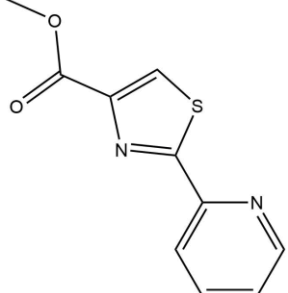
IQS#	Skeletal structure	IUPAC Name
IQS11		ethyl 2-(3-hydroxyphenyl)-4-methylthiazole-5-carboxylate
IQS12		2-(2-methoxyphenyl)thiazole-4-carboxylic acid
IQS13		methyl 2-(2-methoxyphenyl)thiazole-4-carboxylate
IQS14		2-(pyridin-2-yl)thiazole-4-carboxylic acid
IQS15		methyl 2-(pyridin-2-yl)thiazole-4-carboxylate

Table 4.3. Skeletal structures of all IQS analogues docked to AgrAc.



Only a small difference in the predicted binding energies was observed between the 135 docking results ($-4.1 \text{ kcal mol}^{-1}$ for the weakest binder and $-5.4 \text{ kcal mol}^{-1}$ for the strongest). The magnitude of the predicted binding energies is in agreement with the millimolar affinities for AgrA reported in Dr. Jaiyen's PhD thesis. The resulting predicted structures do not demonstrate a single site on the surface of AgrA_C to which the IQS analogues preferentially binds, with each IQS analogue occupying one of several sites on the AgrA_C surface, with similar predicted binding energies for each. Therefore there is no clear consensus as to where IQS analogues bind on surface of AgrA_C based upon the predicted binding energies alone.

Despite this narrow range of predicted binding energies however, a common interaction is observed between AgrA_C and all but two of the fifteen compounds docked, excepting IQS1 and IQS6. Furthermore this interaction occurs in the same hydrophobic pocket identified via NMR titration studies as the binding site to which a small series of AgrA inhibitors bind.^[92] It involves $\pi \cdots \pi$ stack between the phenyl/hydroxyphenyl/methoxyphenyl/pyridyl ring on the IQS analogue ligand and Tyr229 on AgrA_C. A small number of additional polar contacts to residues in the vicinity of Tyr229, primarily to Glu217 and Arg218, are also observed. That IQS1 and IQS6 are not predicted to bind to this area on AgrA_C is pertinent, for in testing whether the inhibitory properties of the IQS analogues is due to an ability to chelate iron, Dr. Jaiyen determined that only IQS1 and IQS6 chelated iron as well as the positive control iron-chelating agent, desferrioxamine.

Since all of the predicted binding energies for each of the nine binding modes for each of the fifteen compounds are very similar, they must be treated as equally plausible. There is however, no available evidence at present to suggest any functional relevance with respect to DNA-binding of areas on the AgrA_C surface that are not either A) on the DNA-binding face of AgrA_C or B) in the pocket previously mentioned.^[92] For this reason, and the fact that is not feasible to discuss all 135 unique results, only the results with the highest predicted binding energy that also exhibit a $\pi \cdots \pi$ stack with Tyr229 will be discussed. Schematic images of the predicted binding modes for all such results can be found in figures 4.16 – 4.28 and their corresponding predicted binding energies are listed in table 4.4.



Spacefill representations of IQS3 and IQS11 bound to AgrA_C in their predicted modes are presented in figure 4.29.

Of the residues with which the IQS analogues form contacts in addition to the $\pi\cdots\pi$ stack, Arg218 also exhibits contacts of one type or another with all thirteen of the analogues. The most abundant type of contact with Arg218 is a bifurcated H-bond, occurring between a single O atom acceptor in the alkylcarboxylate (IQS2, IQS3 and IQS7); carbaldehyde (IQS5 and IQS9); and carboxamide (IQS10) moieties, to N atom donors in the guanidinium head group on Agr218. An additional H-bond, between the N atom acceptor in the IQS thiazole ring to an N atom donor in guanidinium head group, is also observed (IQS2, IQS3, IQS5, IQS7 and IQS9). A bifurcated H-bond also occurs between Agr218 and IQS11, however, the donor atom is the phenoxy O atom, not the ethylcarboxylate O atom. In contrast to the above, the entire IQS11 molecule has its ethylcarboxylate group pointing away from Arg218. In this orientation the ether O atom of the ethylcarboxylate moiety acts as an H-bond acceptor for the side chain hydroxyl group of Ser202. The three IQS analogues that are not predicted to form polar contacts with Arg218 are: IQS8, IQS12 and IQS13. In these three cases the 3-carbon aliphatic portion of the Arg218 side chain is predicted to form non-specific hydrophobic interactions with the 6-membered ring of the IQS analogue.

IQS analogue	Best docking result # with $\pi\cdots\pi$ stack to Tyr229	Predicted binding energy, kcal mol⁻¹	Δ predicted binding energy to result #1, kcal mol⁻¹
IQS2	1	-5.4	NA
IQS3	1	-5.3	NA
IQS4	3	-4.6	0.1
IQS5	1	-4.7	NA
IQS7	4	-4.7	0.2
IQS8	3	-4.8	0.1
IQS9	1	-4.8	NA
IQS10	4	-4.9	0.1
IQS11	2	-5.0	0.1
IQS12	5	-4.6	0.3
IQS13	1	-4.8	NA
IQS14	3	-4.6	0.1
IQS15	1	-4.8	0.2

Table 4.4. Summary of predicted binding energies for IQS analogue docking results which exhibit a $\pi\cdots\pi$ stack to Tyr229 on AgrA_C.

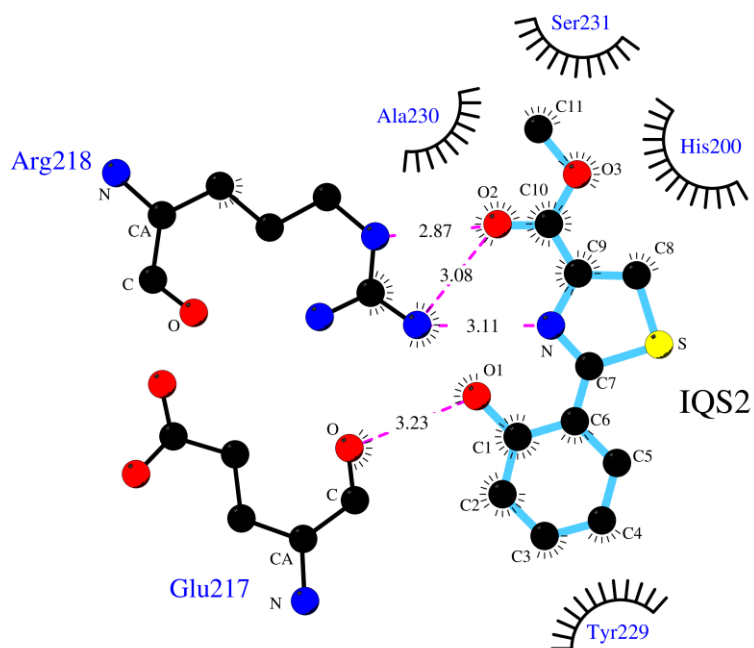


Figure 4.16. IQS2 docking result #1.

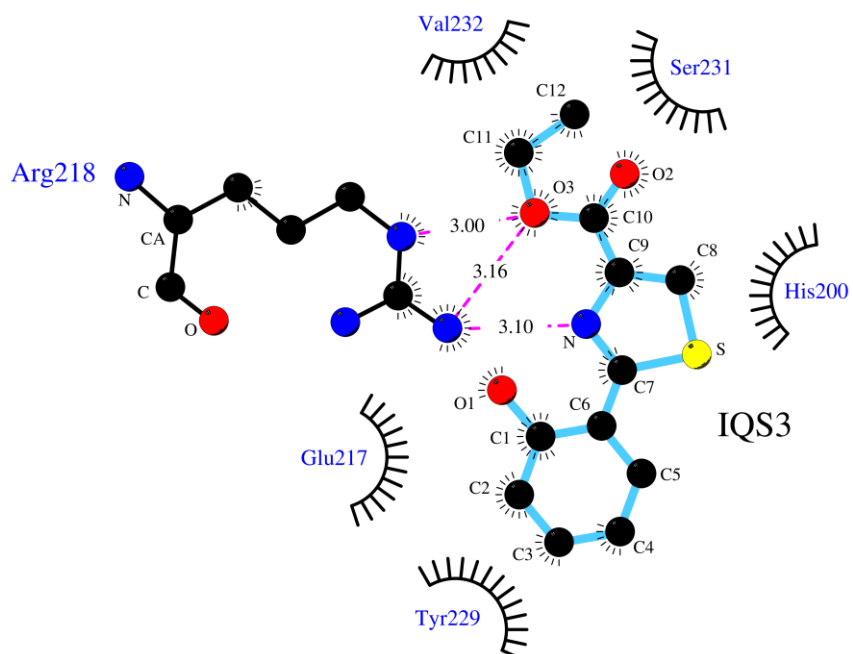


Figure 4.17. IQS3 docking result #1.

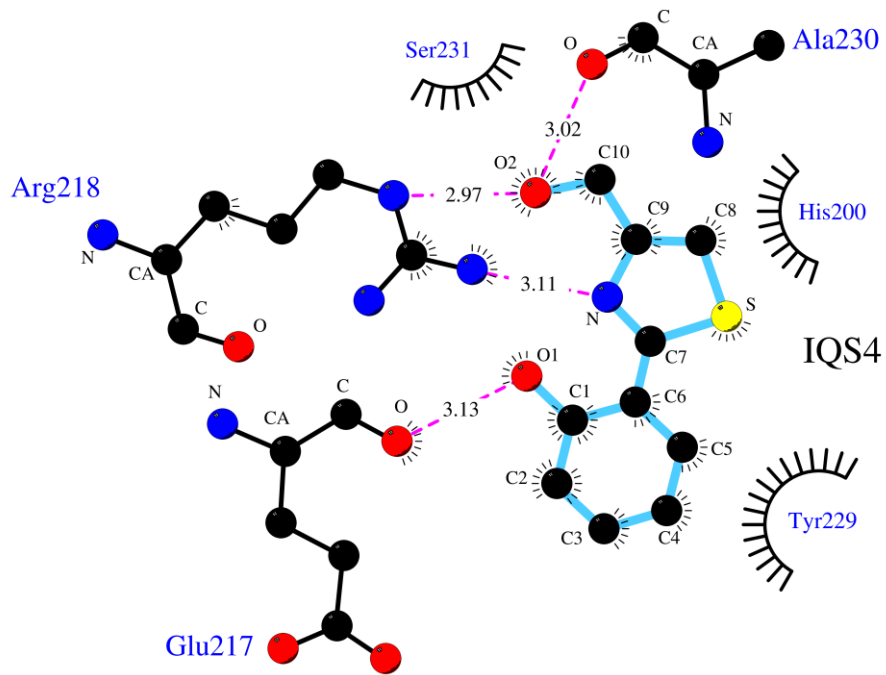


Figure 4.18. IQS4 docking result #3.

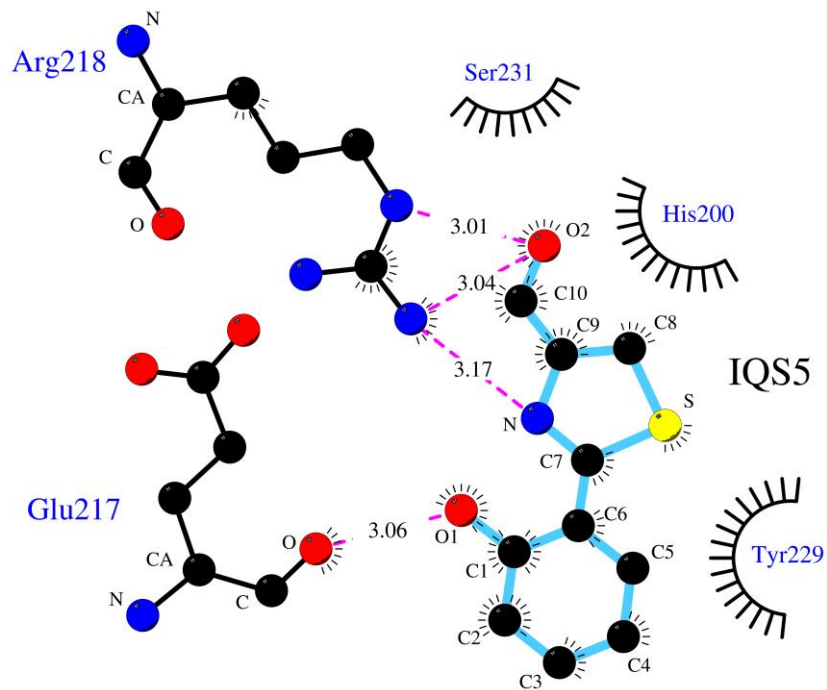


Figure 4.19. IQS5 docking result #1.

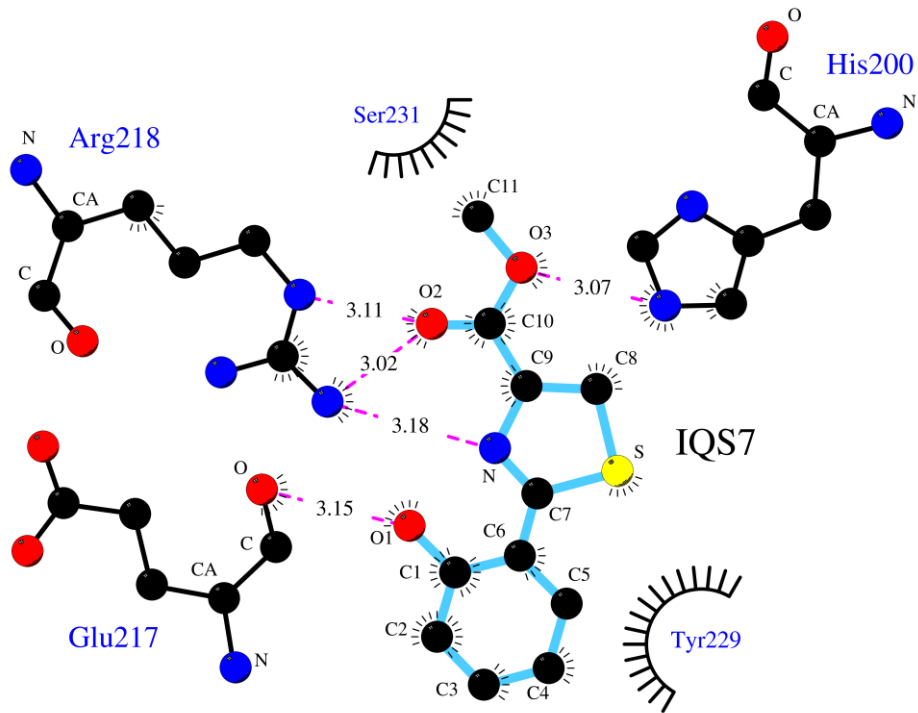


Figure 4.20. IQS7 docking result #4.

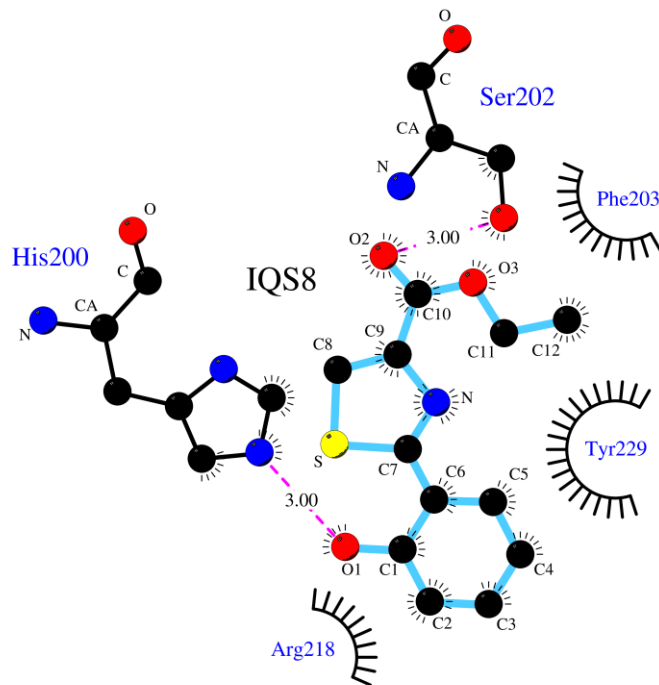


Figure 4.21. IQS8 docking result #3.

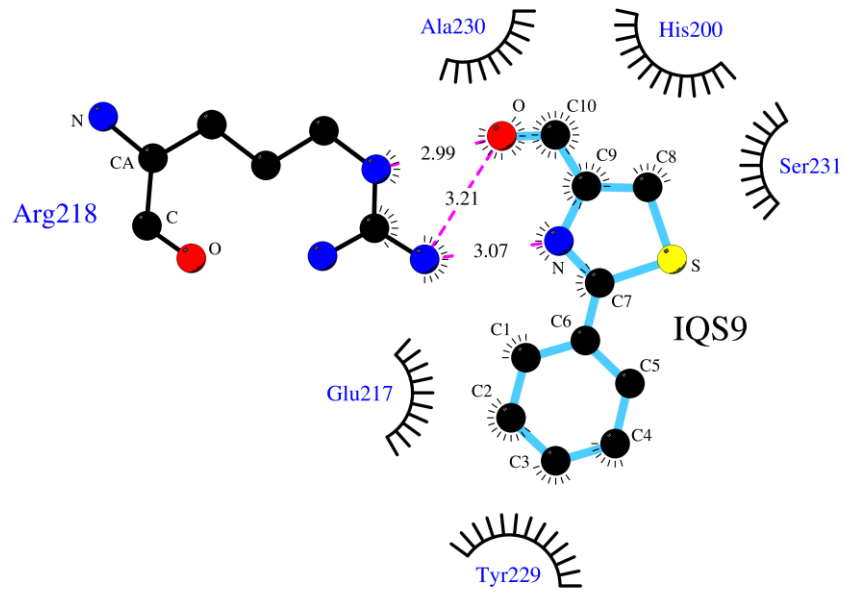


Figure 4.22. IQS9 docking result #1.

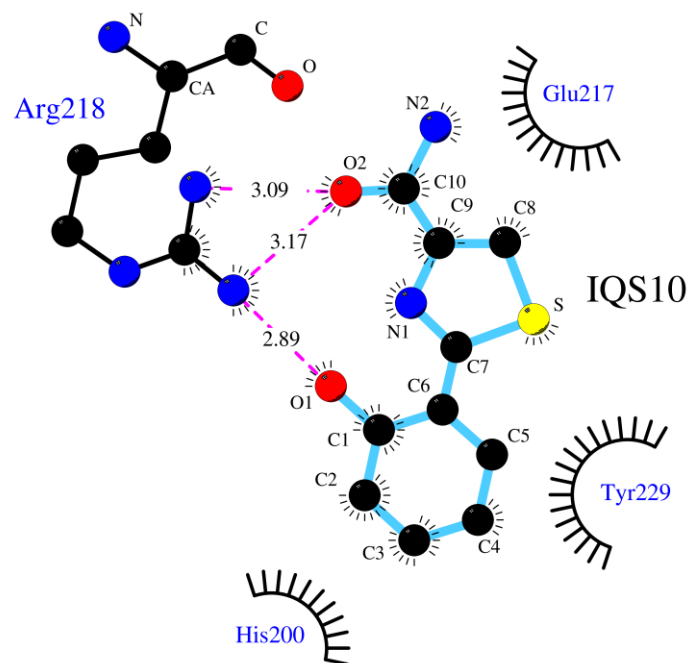


Figure 4.23 IQS10 docking result #4.

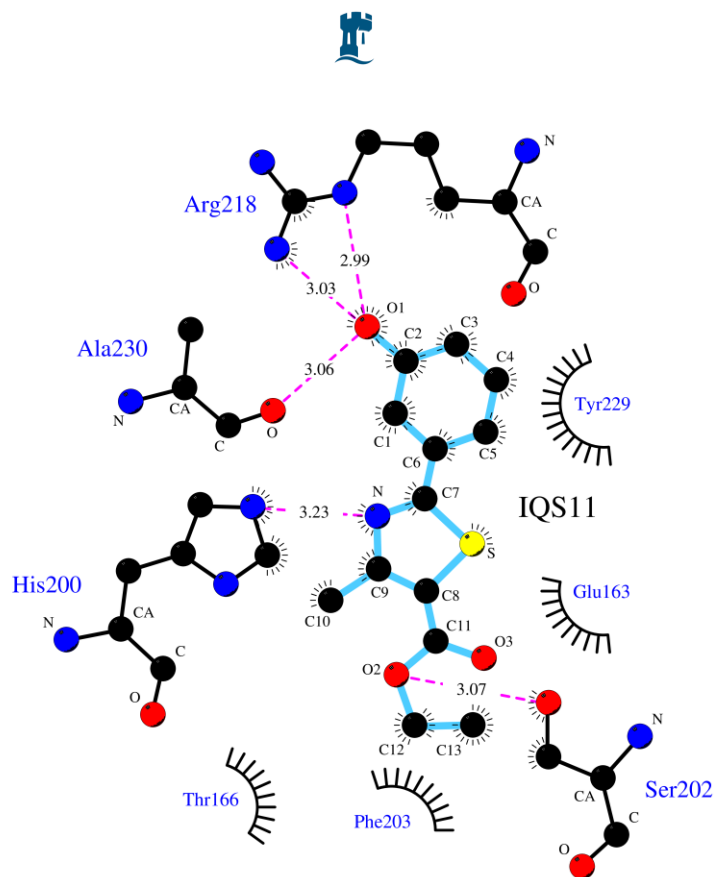


Figure 4.24. IQS11 docking result #1.

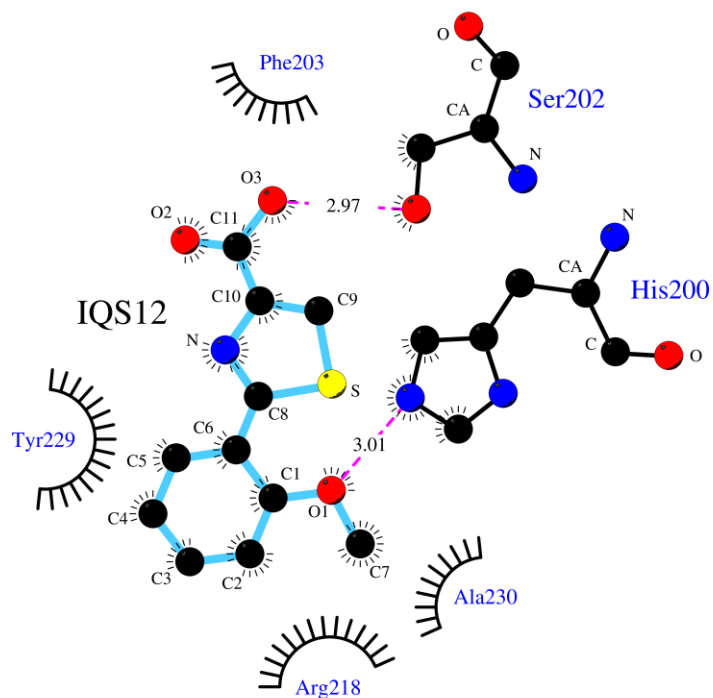


Figure 4.25. IQS12 docking result #5.

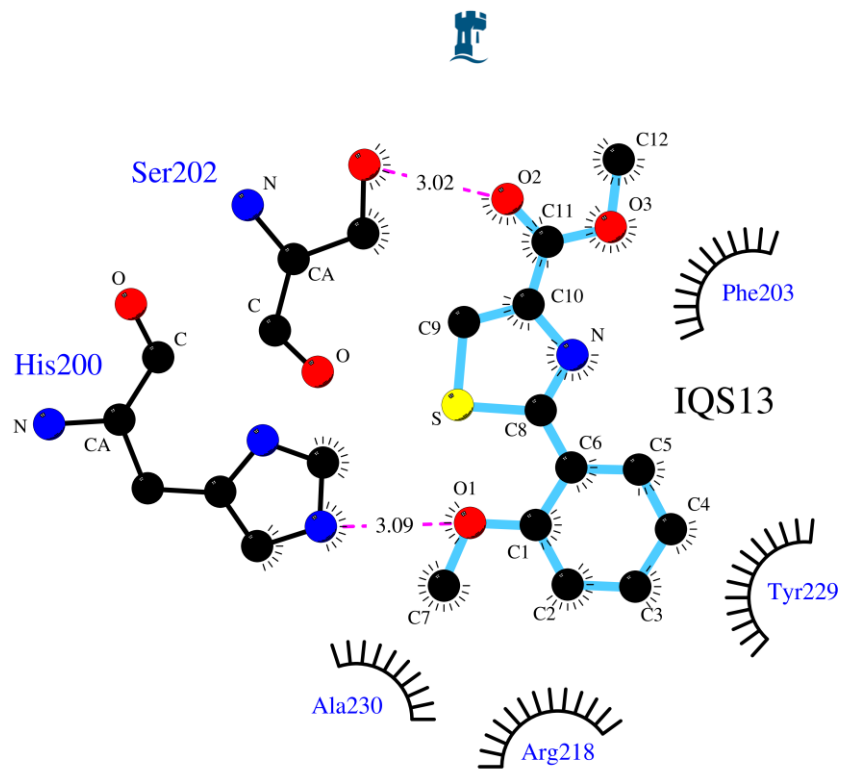


Figure 4.26. IQS13 docking result #1.

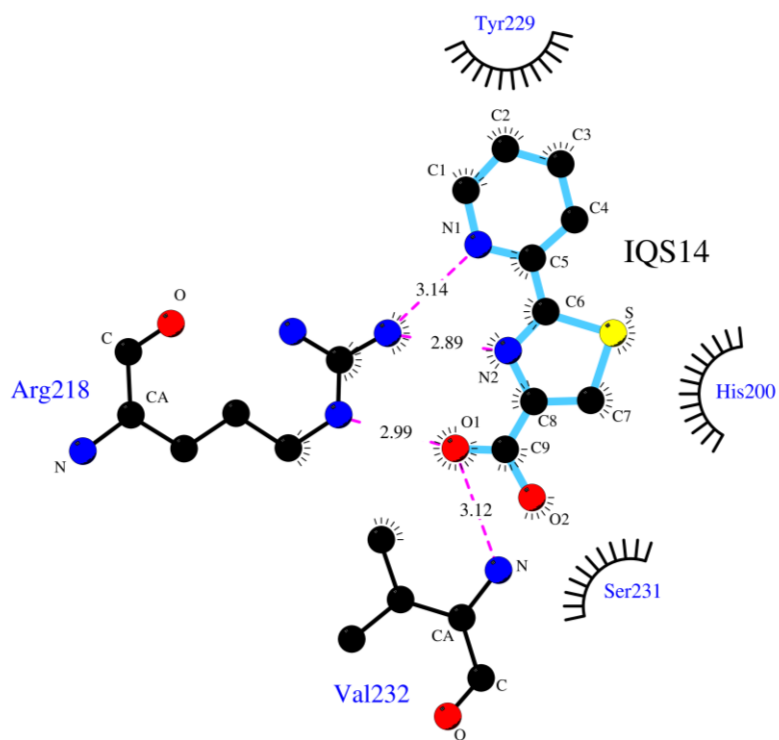


Figure 4.27. IQS14 docking result #3.

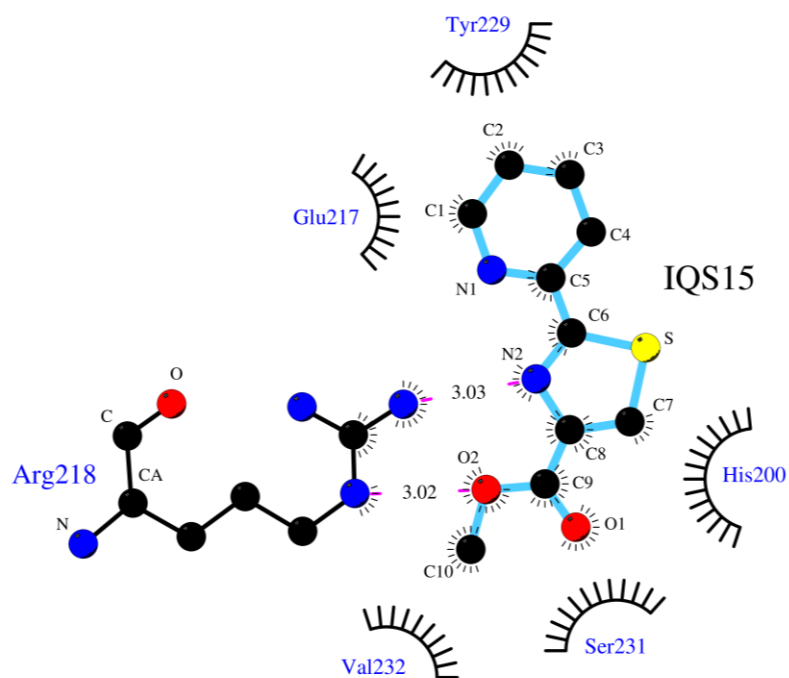


Figure 4.28. IQS15 docking result #1. For all such images: residues involved in H-bonding are explicitly drawn and the H-bonds are displayed as dashed pink lines with their lengths displayed in Å. Carbonyl O and C, and C α and peptide N atoms are labelled in black for explicitly drawn residues. Hydrophobic interactions are shown as black eyelashes. All residues are labelled in blue.

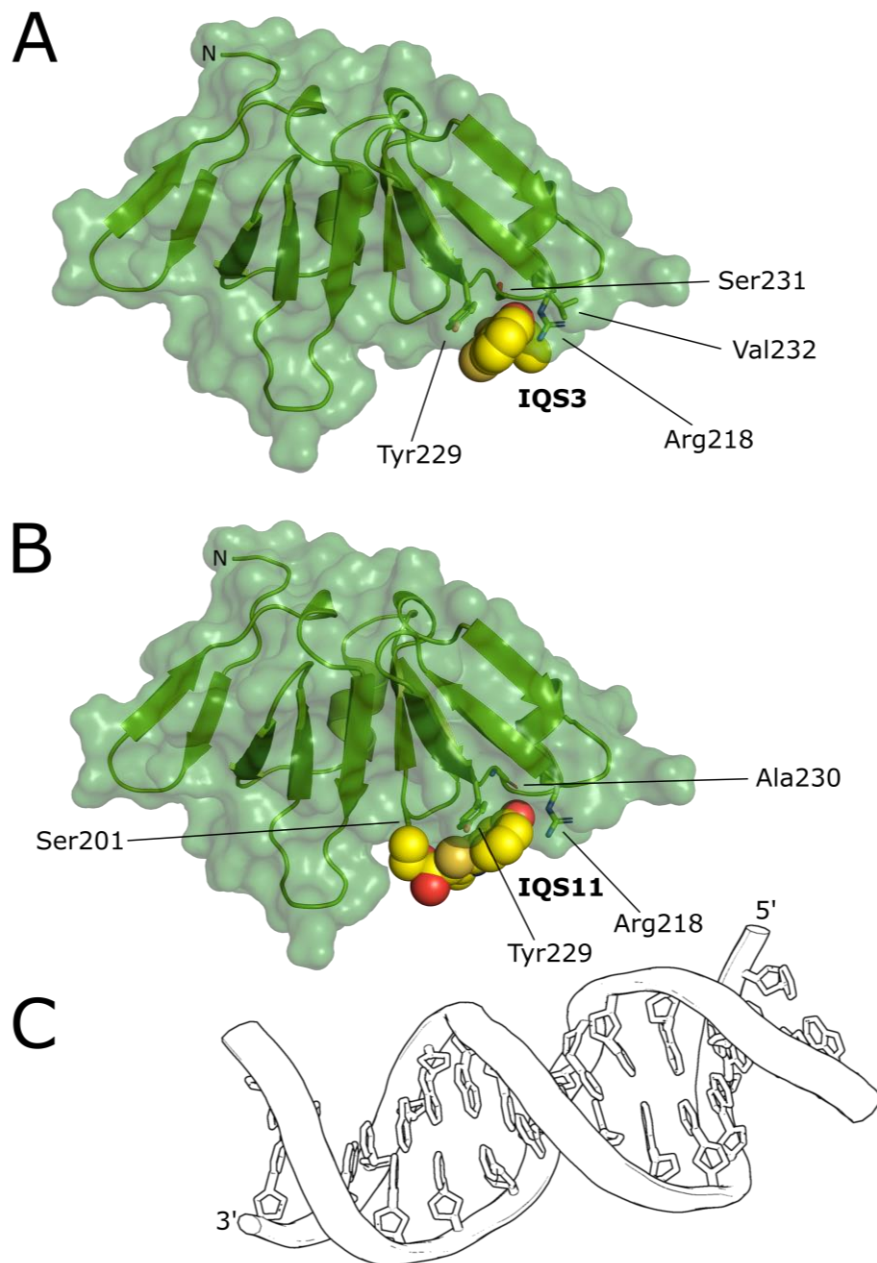


Figure 4.29. Spacefill representations of the results for IQS3 and IQS11 docked to AgrA_C. (A and B). IQS3 and IQS11 are represented as space-filling diagrams and the AgrA_C molecule are shown in green as a transparent surface. Residues that make key contacts with each of the IQS analogues are drawn. (C). P2 promoter DNA^[80] (PDB: 3bs1) is shown on the DNA-binding face of the AgrA_C molecule shown in (B) to demonstrate how the IQS analogues may sterically interfere with DNA binding.



4.3.5.4: Predicted Binding Site of the IQS Analogues in Context

That the IQS analogues are predicted to form polar contacts with Agr218 is important for two reasons.

A) Arg218 forms a water-mediated contact with the sugar-phosphate backbone of the DNA molecule in the in the crystal structure of the AgrA_C /DNA complex^[85] (PDB: 3bs1) and B) R218H mutants in AgrA_C found in *S. aureus* strains isolated from hospital patients have been shown to be *agr*-defective.^[96] It is reasonable to infer therefore that sequestering of Arg218 away from its physiological contact by the IQS analogues would abrogate AgrA DNA-binding function.

Residues in the AgrA_C loop containing Tyr229 form direct and water-mediated polar contacts to the DNA sugar-phosphate backbone, and directly to an adenine nucleobase through the spatially-proximal Arg233, in the upstream AgrA binding site on the P2 (PDB: 4xyq) and P3 (PDB: 4xqj) promoters but not in the downstream AgrA binding sites on the P2 (PDB: 4xxe) and P3 (PDB: 4xqq) promoters.^[80,85] Studies have shown that alanine mutations of residues that are spatially-proximal to Tyr229, including Leu171, Glu181 and His200, significantly impair the ability of AgrA to bind DNA whilst the alanine mutation of Tyr229 itself does not.^[126]

It has also been shown that the Tyr229A AgrA mutant does not alter the ability of AgrA to bend DNA (to initiate RNAP recruitment).^[126] Alanine scanning experiments coupled with GFP expression reporter assays have demonstrated however, that the Y229A and Y229G mutants significantly impair the ability of AgrA to initiate transcription from the P3 promoter, whilst Y229I and Y229F mutants do not.^[126]

Given that the Tyr229 is highly conserved amongst staphylococcal AgrA, and the observations stated above, authors conclude from this data that a large hydrophobic residue at position 229 is required to maintain maximum transcription activation from AgrA, and that given the proximity of the transcription initiation site to the AgrA binding sites, Tyr229 might stabilise a direct interaction between AgrA and RNAP.^[126] It is also worth recalling that Tyr229 is immediately adjacent to Cys228, which is the partner to Cys19, formation of a disulphide bond between which completely



abolishes DNA binding activity *in vitro* and reduces P3 expression *in vivo*.^[208]

A published docking study is available for the AgrA inhibitor savarin (fig. 1.5).^[95] Comparison of the predicted binding mode of the IQS analogues to that of savarin reveals the two to be essentially identical, with the savarin molecule also binding into a groove between Tyr229 and Arg218. The isopropylphenyl ring on the savarin molecule exhibits a $\pi \cdots \pi$ stack with Tyr229 whilst the triazole ring forms polar contacts with Arg218. The predicted binding energy of savarin to this area on AgrA_C is reported to be $-6.1 \text{ kcal mol}^{-1}$ and is comparable to those determined for the IQS analogues.

It is pertinent therefore that the IQS analogues, which have been shown by Dr. Jaiyen to reduce α -haemolysin expression from the P3 promoter in a small number of *S. aureus* strains, and that bind directly to AgrA, are predicted to have even a low affinity for this functionally-important area of AgrA. Although the IQS analogues have a similar predicted affinity for other areas on the AgrA_C surface there is no clear consensus between them, nor is there is there experimental evidence available to indicate any functional relevance of these areas.

It must be borne in mind that the Tyr229 region is located on the DNA-binding face of AgrA_C, therefore any appreciably-sized compound binding to this site would probably sterically interfere in with binding of DNA. Whether or not the mode of inhibition of AgrA DNA binding by the IQS analogues occurs through steric interference of DNA binding, through interfering with the function of Arg218 or through interfering with the role of Tyr229 in gene transcription, is unclear at this time. It may be that they are indistinguishable given the proximity of Arg218 and Tyr229 to DNA in the available crystal structures of AgrA_C/DNA complexes.



Chapter 5: AgrB – Expression, Purification, and Characterisation

5.1: Preliminary Work on AgrB2

Prior to arrival of the experimenter in the *S. aureus* research group considerable research had been undertaken on AgrB in-house. Consequently there were several constructs available to begin purification optimisation for structural studies. This is not to imply that a lot was known about AgrB, however. Basic characteristics such as the oligomeric state, orientation in the membrane, and the location of the N- and C-termini were all unknown. Secondary structure predictions suggest that AgrB is primarily alpha helical.^[51] An AgrD cleavage assay had been developed by a former researcher in the group, Dr. Victoria Steele but it was unreliable to perform, involving detection of cleaved T7-tagged synthetic AgrD via western blot over a two day period using photographic film to detect the chemiluminescence. The best results had exhibited approximately 3 % turnover of AgrD after several hours of incubation. Only AgrB1 activity had been detected, but even this was not specific to the concomitant AgrD1 as it also cleaved AgrD2, a phenomenon that does not occur *in vivo*.^[67] Furthermore AgrB2 activity had never been demonstrated.

The difficulty in addressing the unknowns outlined above arises from two primary causes: A) AgrB is an integral membrane protease and B) The gene products of the allelic variants of AgrB all behave differently *in vitro*. It will become apparent that an observation made for AgrB2 does not necessarily hold for AgrB1. Consequently, an experimental design suitable for AgrB1 may not be valid for AgrB2 and vice versa. With respect to structural studies, it was at least known that AgrB1 formed ladders on SDS-PAGE, whilst AgrB2 ran (predominantly) as a monomer. This was enough to choose AgrB2 as the starting point for structural experiments. Building upon the work of Dr. Victoria Steele, fellow *S. aureus* group researcher Dr. Philip Bardelang had demonstrated that C-terminal (His)₆-tagged AgrB2, when expressed in pET24(b)+, could be obtained in crude form via Co²⁺-affinity chromatography provided that the Co²⁺-chelated resin was loose, the incubation of the sample with the resin was sufficiently long (> 3 hours) and that the detergent concentration was not too high (< 2 % w/v).

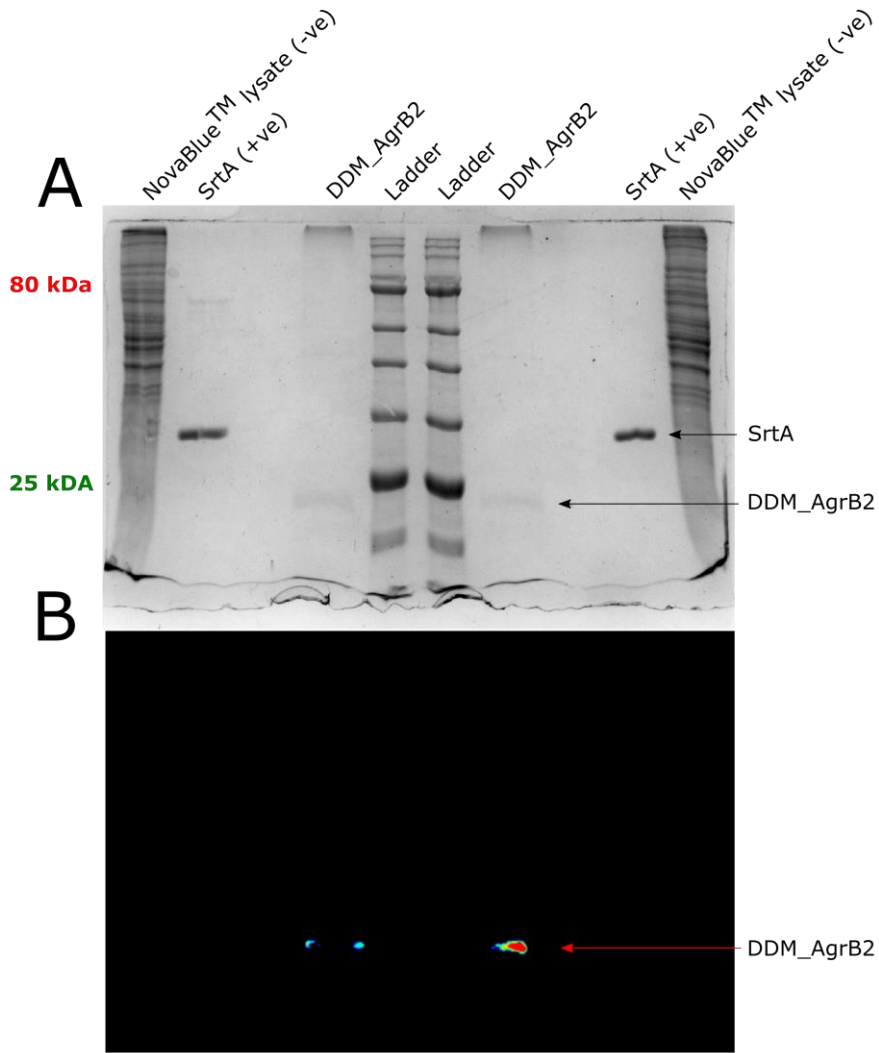


Lastly, it was also known that AgrB possess at least two catalytically-requisite residues, His77 and Cys84.^[52] Alanine point mutations of both of these in AgrB1 and AgrB2 had been generated by Dr. Philip Bardelang.

5.2: Positive Identification of AgrB in the Recombinant Sample

Because membrane proteins are usually poorly overexpressed,^[311,312,313] they are not always readily identifiable on SDS-PAGE since they may not stand out significantly against the expression host proteome. Even after elution from a metal affinity column there may still be enough background contamination to make a positive identification impossible. This was generally the case for AgrB2, so it was subject to an anti-(His)₄ western blot (fig. 5.1A) and to MS at the University of York (fig. 5.1C). The western blot positively identified AgrB2, however, the SrtA (*S. aureus* Sortase A) positive control failed to generate chemiluminescence. Note that after transferring the samples to the blot membrane, which was done symmetrically, the blot membrane was cut in half down the middle. One of the halves was exposed to an epitope-enhancing buffer in accordance with a published protocol^[314] to aid exposure of the epitope.

The MS analysis (denaturing Matrix-Assisted Laser Desorption/Ionisation-Time of Flight, or MALDI-TOF) was performed on a chymotrypsin digest of an SDS-PAGE band which prospectively corresponded to AgrB2. The MS experiment returned a single peptide fragment which matched the AgrB2 primary sequence that was supplied by the experimenter along with the recombinant sample. One peptide match to a supplied sequence is not sufficient to confirm the presence of AgrB2 in the recombinant sample. However, taken along with the western blot result, it is possible to conclude that the recombinant sample contains AgrB2.



C

Select Summary Report

Format As: [Help](#)

Significance threshold $p < 0.05$ Max. number of hits:

Standard scoring: MudPIT scoring Display non-significant matches Show sub-sets:

Show pop-ups: Suppress pop-ups Require bold red:

Preferred taxonomy:

Re-Search: All queries Unassigned Below homology threshold Below identity threshold

1. [C623_Warwick_Sequence1](#) Mass: 23037 Score: 57 Matches: 1(1) Sequences: 1(1)
 Query Observed Mr(expt) Mr(calc) ppm Miss Score Expect Rank Unique Peptide
 1 1214.5706 1213.5633 1213.5615 1.46 0 57 5.1e-05 1 U **F.DNKIDQFATY.L**

Proteins matching the same set of peptides:

[C623_Warwick_Sequence2](#) Mass: 22788 Score: 57 **Matches: 1(1) Sequences: 1(1)**

Protein sequence coverage: 5%

Matched peptides shown in **bold red**.

1 MNY**F****D****N****K****I****D****Q** **F****A****T****Y**LQKRNN LDHIQFLQVR LGMQVLAKNI GKLIVMYTIA
 51 YILNIFLFTL ITNLTFLVLR RHAHGAHAPS SFWCYVESII LFILLPLVIV
 101 NFHINFLIMI ILTVISLGI SVYAPAATKK KPFPVRLIKR KKYAIIVSL
 151 TLFIIITLIK EPFAQFIQLG IIIIEAITLLP IFFIKEDLKL EHHHHH

Unformatted sequence string: [197 residues](#) (for pasting into other applications).

Sort by residue number increasing mass decreasing mass

Query	Start - End	Observed	Mr(expt)	Mr(calc)	ppm	M	Score	Expect	Rank	U	Peptide
1	5 - 14	1214.5706	1213.5633	1213.5615	1.46	0	57	5.1e-05	1	U	F.DNKIDQFATY.L



Figure 5.1 (previous page). Confirmation of AgrB2 sample identity. (A). An SDS-PAGE gel containing the AgrB2 sample, a SrtA positive control and NovaBlue™ cell lysate negative control. The gel is 16 % polyacrylamide, reducing. (B). Blot membrane showing a signal for the AgrB2 sample but not for the SrtA control. The blot membrane was cut in half and the right hand side was treated with a stripping buffer to aid exposure of the epitope. (C). MS data for AgrB2, showing detection of single peptide fragment that matched the supplied primary AgrB2 sequence, entitled “C623_Warwick_Squence1.



5.3: Initial Attempts to Purify AgrB2

Although it was apparent that a crude N-dodecyl- β -D-maltoside (DDM) AgrB2 extract (herein denoted as DDM_AgrB2, with other detergent abbreviations preceding the underscore as appropriate) could be obtained via Co^{2+} -affinity chromatography, this sample was not suitable for crystallisation experiments owing to both its lack of purity and its low yield.

The issue of purity was overcome by adding more chromatographic steps to the purification, although this was to lead to more difficulties, including polydispersity and inactivity (both somewhat overcome and discussed later on). The issue of yield was harder to overcome owing to the kit available to research group. Simply increasing the expression culture volume, whilst certainly providing more AgrB2, was not such a simple route to increasing yields as it first seemed. This is because, after cell lysis, the membranous fraction must be separated via ultracentrifugation from the total cell lysate.

Centrifuges that operate at 100 – 200,000 g are somewhat specialist pieces of equipment and not common in all laboratories. The one to which the *S. aureus* research group had access had a total capacity of 8 x 6 mL which is sufficient for lysate corresponding to biomass obtained from ~ 1 L of expression culture. As many cells as desired could be resuspended into 48 mL and lysed, but efficient lysis would take some time to achieve and it would result in high concentrations of sample-harming enzymes in the lysate. Consequently it was one-at-a-time through the ultracentrifuge for each batch of cells lysed with each spin lasting 1 hour.

Initially, four sequential ultracentrifugation steps were performed in series. This was deemed unsustainable however, and was later cut to two to make the workflow easier. It was also found that the amount of material obtained from 4 L of expression culture exceeded its solubility limit when concentrated to fit onto the 5 mL SEC sample loop. This latter problem was also true of the amount of material prepared from 2 L of overexpression culture but was overcome by spin-concentration/dilution steps to incrementally remove the 0.5 M imidazole, in which Co^{2+} -affinity purified DDM_AgrB2 is only soluble up to ~ 2.0 – 3.0 mg mL⁻¹ (remembering that much of the absorbance used to calculate this figure comes from contaminants at this stage). Performing two



sequential gel filtrations was also attempted to mitigate this insolubility, however, this was also deemed unsustainable in a lab sharing communal instruments.

A full description of the purification of DDM_AgrB2 can be found in methods section 7.7.4, but in summary an acceptable an acceptable yield of 0.25 – 0.5 mg 1 L of expression culture was obtained by isolating membranes from biomass corresponding to 2 L of expression culture by processing each biomass from 1 L in series to yield 2 x 6 mL membrane extracts which were combined. Extraction of AgrB2 into DDM micelles was allowed to proceed overnight after which the extract was diluted two-fold into DDM-less buffer to halve the DDM concentration.

This ~ 24 mL sample, once the depleted membrane debris was removed, was added to an Econo-Column® containing 2.5 mL of Co²⁺-coordinated TALON™ Superflow® resin, which was mixed into the DDM_AgrB2 extract and allowed to bind the sample for several hours.

Once the sample had been detected on SDS-PAGE, the sample was spin-concentrated two-fold, made up to the original volume in imidazole-less buffer, then spin-concentrated ten-fold, made up to the original volume again before finally being spin-concentrated to fit onto the SEC loop. This mitigated the insolubility of DDM_AgrB2 in high concentrations of imidazole that re required to perform the step elution from the Co²⁺-affinity column.

The purification process was condensed into three full days. The first was spent isolating cell membranes and extracting AgrB2 into DDM overnight. The second was spent performing the Co²⁺-affinity chromatography and analysis, and equilibrating the SEC column (it was deemed better practise to wait until purified DDM_AgrB2 had been observed before equilibrating the SEC column given the expensive nature of DDM, which might otherwise be wasted). The third was spent performing the SEC and crystallisation experiments. The purification was robust, seldom failing except for when glycerol stocks of the expression constructs degraded, and after which fresh transformant colonies were prepared.



5.4: Generating a Monodisperse DDM_AgrB2 Sample

Although relatively pure DDM_AgrB2 could be obtained from Co^{2+} -affinity chromatography coupled with SEC, the sample was polydisperse, which is a barrier to crystallisation. A common, high-throughput method of identifying sample buffers which having a stabilising effect on the sample is the thermofluor assay.^[315,316] An increase in sample melting temperature (T_m) is taken as a surrogate for increased sample stability imparted by the sample buffer. The T_m of a sample may be determined by gradual heating of it in a real-time PCR machine. The sample matrix contains miniscule amounts of the dye SYPRO™ orange which exhibits reduced fluorescence quenching upon association with the unfolded hydrophobic interior of the sample protein. The ideal result is a sigmoidal increase in fluorescence maxima, at a wavelength of 570 nm, at the sample T_m . Since the PCR machine is capable of holding 96 reactions at once, and each reaction can be up to 100 μL in volume, many melting curves are obtained in one experiment for a variety of sample buffers. Although the actual magnitude of the T_m increase that is tantamount to a *bona fide* stabilising effect is often dubiously assigned after the experiment (usually 0.5 – 1.0 °C), it is a good if information-poor first approximation of what reagents may impart sample stability, and hopefully therefore, monodispersity.

This experiment was attempted with three replicate Co^{2+} -affinity purified DDM_AgrB2 samples to test the efficacy of the experiment before varying buffer composition, and returned a negative result (fig. 5.2). The exact composition of the samples and controls can be found in the relevant methods section 7.7.5. The negative results are probably due to the excess detergent present in the sample binding the SYPRO™ orange dye. This would explain the large initial fluorescence counts and the similarity in the fluorescence signal between the test and control (buffer only) samples. It was later learned that there are other dyes more suited to the hydrophobic nature of membrane proteins, however this was not known at the time of experimentation, and so this method of determining maximal sample stability was retired. Fortuitously, a separate sample quality control experiment performed in the absence of a reliable activity assay provided a serendipitous discovery of a sample buffer which yielded monodisperse AgrB2.

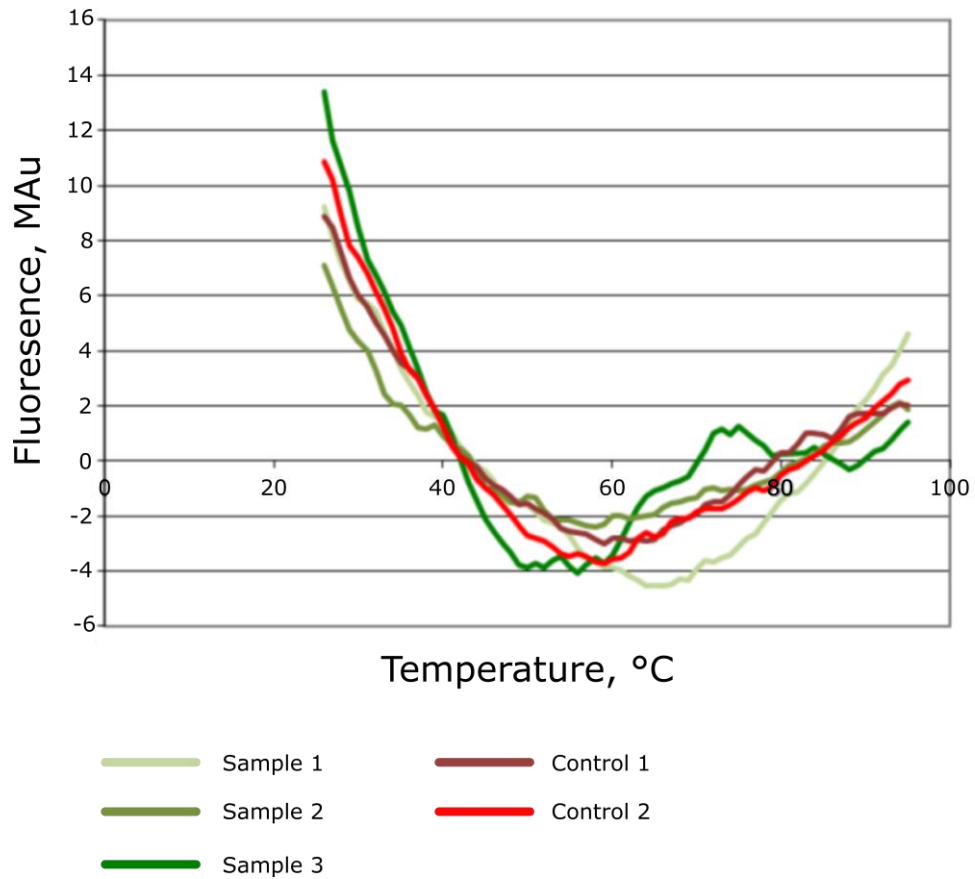


Figure 5.2. Thermofluor data for three replicate Ni²⁺-affinity purified DDM_AgrB2 samples and two buffer controls. The third buffer control was an outlier, exhibiting fluorescence that was not on the same scale as the other five conditions, and so was omitted. The decrease in fluorescence signal with increasing temperature is counter to what is expected for SYPRO™ orange and is probably due to the dye associating with detergent molecules.



To check the DDM_AgrB2 sample integrity with respect to secondary structure it was subject to Circular Dichroism^[317] (CD) and this experiment is discussed in fullness later in section 5.6. Here it is sufficient to say that high concentrations of Cl⁻ ions are known to interfere with CD measurements and must consequently be kept at low concentrations (≤ 50 mM) in samples whose intended use is for CD measurements.^[317]

Hitherto the DMM_AgrB2 sample had been transferred into a buffer containing: tris, 20 mM, pH 7.0, NaCl, 100 mM, DTT, 1 mM and DDM, 0.03 % w/v during SEC. In order to keep [Cl⁻] ≤ 50 mM, on the advice of biophysics PhD Dr. Alex Slater, the SEC running buffer was changed to: K₂H/KH₂PO₄, 20 mM, pH 7.5, NaCl, 50 mM, DTT, 1 mM and DDM 0.03 % w/v. This had the remarkable of “condensing” all of the shouldered DDM_AgrB2-containing peaks previously observed during SEC into a single Gaussian peak as shown in figure 5.3. Typical SDS-PAGE gels for a successful DDM_AgrB2 purification are shown in figure 5.4.

Henceforth the latter running buffer was used for all AgrB2 purifications with only the detergent type, and its concentration, being altered as the downstream experiments required. This serendipitous finding implies that the oligomeric state of AgrB2 *in vitro* is highly sensitive to the ionic strength of the sample buffer, and that low ionic strengths promote agglomeration into a single species (insofar as it is possible to deduce this from SEC coupled with ultraviolet light detection).

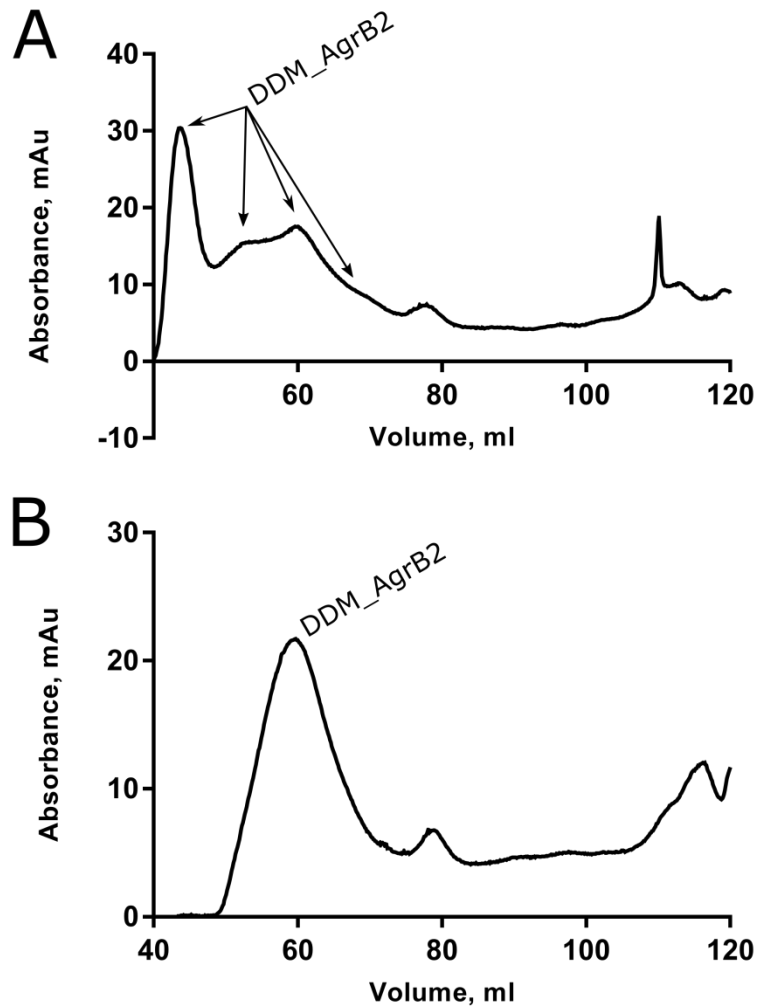


Figure 5.3. Size exclusion data for two DDM_AgrB2 samples using different running buffers. (A). The running buffer is: tris, 20 mM, pH 7.0, NaCl, 100 mM, DTT, 1 mM and DDM, 0.03 % w/v. Most of the DDM_AgrB2 elutes in the void volume (the sharp peak eluting at ~ 45 mL) whilst that which does not elute in the void volume exists as several unresolved species. (B). The running buffer is: K_2H/KH_2PO_4 , 20 mM, pH 7.5, NaCl, 50 mM, DTT, 1 mM and DDM 0.03 % w/v. All of the unresolved DDM_AgrB2 species appear to have agglomerated into a single species that elutes at ~ 62 mL. The running buffer corresponding to (B) was used for all subsequent AgrB2 purifications.

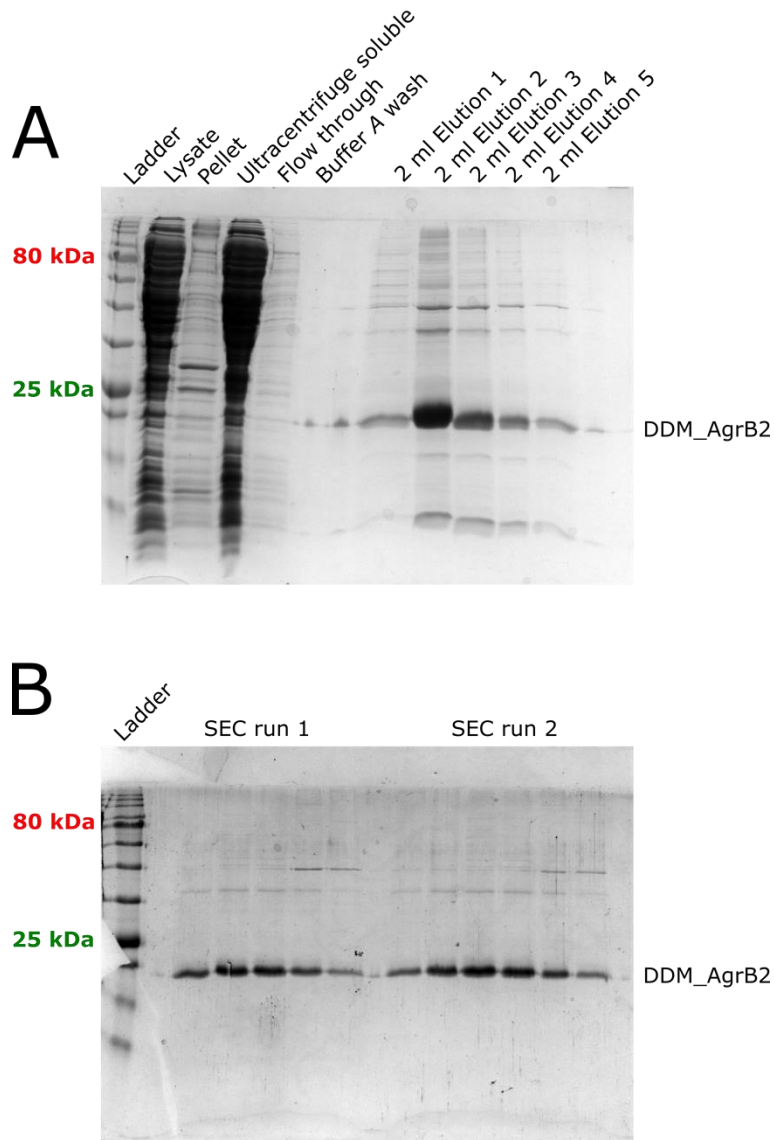


Figure 5.4. Representative SDS-PAGE gel data for a successful DDM_AgrB2 purification. (A). The gel corresponding to the Co^{2+} -affinity step. DDM_AgrB2 is eluted via 5 x 2 mL additions of buffer *B* and abundant sample has been expressed and crudely purified. (B). The gel corresponding to two SEC runs. The sample is of good purity and yield for a recombinant integral membrane protease. Gels are 16 % polyacrylamide, reducing.



5.5: Assaying for Activity of DDM_AgrB2

Only one attempt to assay DDM_AgrB2 sample for AgrD cleavage was made. A read through of the relevant experimental methods section 7.7.6 will highlight many of the experimental challenges involved. Chemists were employed on the same research grant as the experimenter to develop a more reliable assay using a fluorogenic AgrD substrate mimetic, however, this did not come to fruition within the time frame of the PhD process owing to difficulties involved in AgrD polypeptide synthesis. No evidence for turnover of AgrD2 by AgrB2 was demonstrated (fig. 5.5).

There are a number of reasons why folded membrane sample may not show activity once removed from the membrane. For example the chemical and physical environment of a detergent micelle is very different to a cell membrane, not least because many of the native lipids^[318] that may co-localise with the sample in the cell are abjured, as may be other co-factors that are necessary to for efficient AgrD turnover. The lateral pressure imparted by the membrane on the sample is also lost in a detergent micelle.^[319,320,321] Furthermore the extreme N-terminal region of AgrD, being hydrophobic, is presumed to target AgrD to the cell membrane where it can manoeuvre in two dimensions until it encounters AgrB.^[42,43] In this assay there is no cell membrane.

There are lipidic technologies that more closely mimic the host cell membrane, but these are ill generally suited to sample crystallogenesis as they have high relative masses, and in the case of Styrene-Maleic Acid Lipid Particles^[322] (SMALPs) are not stable at acidic pHs, which make redundant a broad swathe of potential crystallisation conditions.

It would be feasible to feed the DDM_AgrB2 to a SMALP or nanodisc to see if activity is re-instated and undoubtedly this is a good line of experimentation,^[44] but it would have ultimately taken time away from trying to crystallise DDM_AgrB2, an objective itself which had many innate barriers to overcome. Consequently no further effort was made to demonstrate proteolytic activity of the DDM_AgrB2 sample.

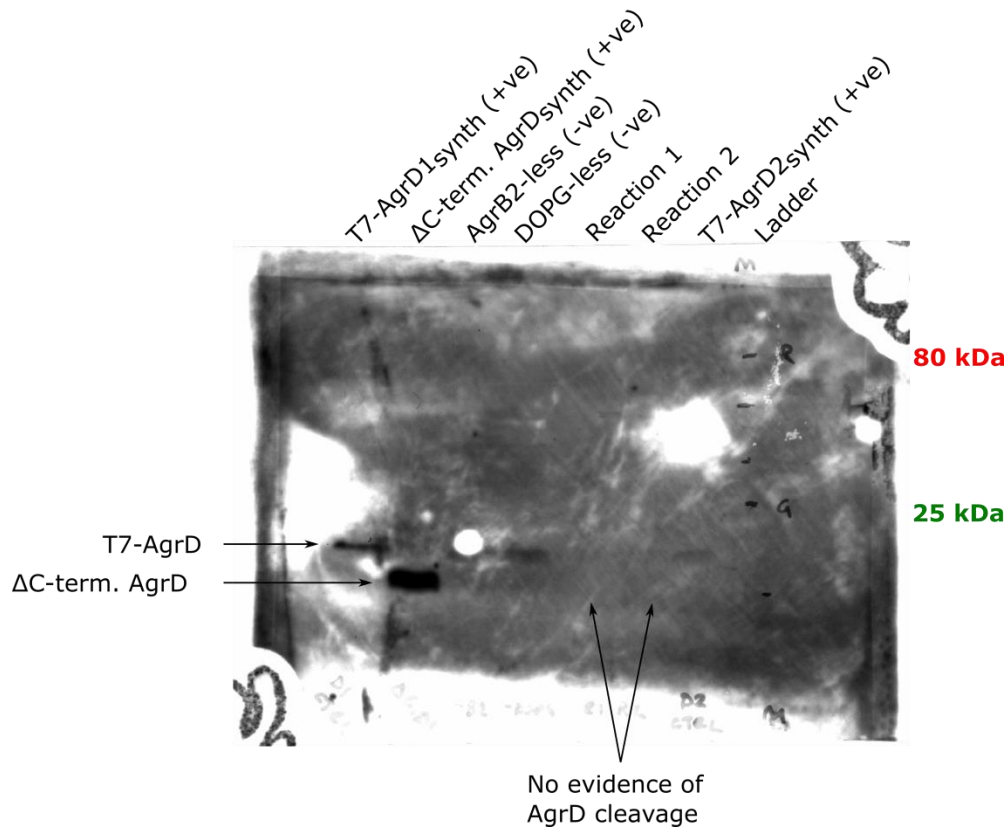


Figure 5.5. T7-Tag western blot for AgrD2 cleavage by DDM_AgrB2. The compositions of the lanes are as follows (left-to-right): an intact T7-tagged synthetic AgrD1 peptide control; a C-terminally truncated T7-tagged AgrD1 control; a full reaction minus DDM_AgrB2 negative control; a full reaction minus 1,2-dioleoyl-sn-glycero-3-phospho-(1'-rac-glycerol) (DOPG) negative control; two replicate AgrD cleavage reactions containing: reaction buffer, DDM_AgrB2, T7-tagged synthetic AgrD2 and DOPG (see relevant methods section 7.7.6 for exact composition); an intact T7-tagged synthetic AgrD2 control; ladder. No evidence for AgrD2 turnover by DDM_AgrB2 is observed. The (phosphorescent) skull and hand in the top right and bottom left corners are to aid orientation of photographic film, exposure of which is performed in a dark room. The gel used before transfer to the blot membrane was a 16 % tris-tricine gradient gel.



5.6: Circular Dichroism Experiments on DDM_AgrB2

The lack of proteolytic activity for the DDM_AgrB2 sample intended for structural studies lead to doubt over the structural integrity of it. In the absence of a reliable activity assay and with other assays in development, CD experiments were employed to detect what, if any, secondary structure elements the DDM_AgrB2 sample possessed, and whether the sample was folded therefore.

A CD experiment is a one-dimensional fingerprint technique which measures the non-linearity of the resultant of left- and right-hand circularly polarised light as they both pass through a solution-state sample. If the left- and right-hand circularly polarised photons propagate at the same speed through a sample, and are both in phase, then their resultant waveform is linear. However, if either of the hands is refracted (or absorbed) more strongly than the other one, as they would be if they were propagating through a chiral sample such as a protein, then their resultant is an ellipse. The degree of ellipticity of the resultant is directly related to the conformation, and physico-chemical environment of the chiral centres in the sample. If these extraneous parameters can be controlled, then the observed ellipticity of the resultant wave is diagnostic of protein sample secondary structure.^[317]

AgrB2 exhibited a CD spectrum characteristic of a protein which is folded and comprised almost entirely of α -helices^[323] (fig. 5.6A) as suggested by secondary structure predictions available in the literature.^[51] Additional spectra were collected over a temperature range of 10 – 90 °C (fig. 5.6B) which allowed for an estimate of the T_m of the DDM_AgrB2 sample to be made. This was found to be 63 °C (fig. 5.6C).

These results clearly show that despite the lack of demonstrable activity the purified DDM_AgrB2 sample had intact secondary structure. Furthermore, a T_m of 63 °C is indicative of excellent thermostability. This observation should be borne in mind when experimental design might benefit from gentle heating of the sample, as may be the case for extraction of AgrB2 from the cell membrane or when optimising *in vitro* activity assays.

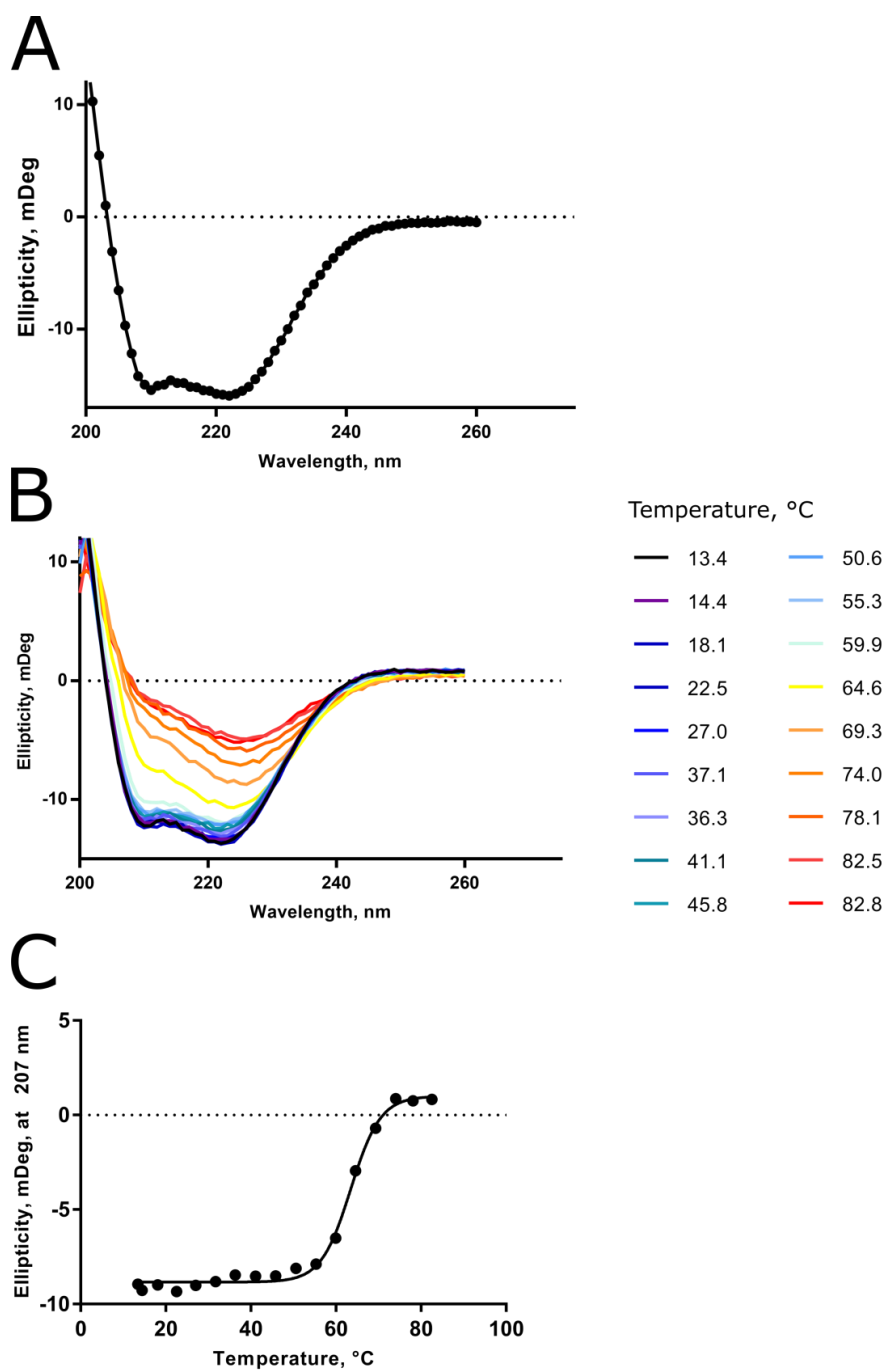


Figure 5.6. CD Spectra and a melting curve for DDM_AgrB2. (A). A CD spectrum collected at room temperature that exhibits the “double dip” at 208 and 222 nm that is classical of structures that are entirely, or almost entirely α -helical in composition.^[323] (B). CD spectra collected at temperatures of ~10 through 90 °C. The loss of the double minimum is indicative of sample denaturation and occurs at temperatures above 60 °C. (C) Ellipticity at 207 nm from (B) plotted against temperature. A sigmoidal fit of the data gives a sample T_m of 63 °C.



5.7: Detergent Removal and Retention of the DDM_AgrB2 Unit

One of the difficulties in crystallising a membrane protein is the requirement of having the sample solubilised in detergent to ensure sample solubility in aqueous buffers. The detergent has to be maintained above its Critical Micelle Concentration^[324] (CMC) in order for it to engulf the membrane protein sample, but be as low as possible to present the smallest steric barrier to crystallisation. This is simple enough except that efficient extraction of samples from cell membranes usually requires detergent concentrations that are many orders of magnitude above the CMC. Consequently, the problem becomes one of detergent removal.^[325]

Dialysis is not normally feasible for two reasons. A) The detergent/sample complex is usually not massively dissimilar in size to the free micelles, so no filter can discriminate effectively between the two. Therefore passage of both or none of these entities through the dialysis membrane occurs. B) Any dialysis that does occur is on the individual detergent molecules, which are in dynamic equilibrium with the micellar phase, and/or the lipid monolayer and/or any other phases known to occur at a isotherm for a given detergent.^[326] The net result of these is that efficacious dialysis is usually quite slow to perform.

Another strategy to remove excess detergent simply use spin-concentrators with the largest cut-off filter possible to spin-concentrate detergent/sample complexes without sample loss, and assume that this allows passage of a significant amount of the excess detergent through the spin-concentrator filter. The amount of detergent left in the sample may be estimated using Thin-Layer Chromatography (TLC) of the detergent/sample complex against several detergent standards, followed by iodine vapour sublimation staining in a vacuum desiccator.^[327] This, however, was not tried.

For DDM_AgrB2, the large cut-off spin-concentrator route was used to attempt to remove excess detergent. To determine whether a 100 kDa cut-off spin-concentrator could be used to elevate sample concentration prior to crystallisation experiments whilst hopefully permitting some of the DDM micelles through, a simple assay was devised to measure DDM_AgrB2 retention in such a filter. Purified DDM_AgrB2, 200 μ L was, applied to a 100 kDa cut-off spin-concentrator and spun to



reduce the volume to 100 μL and then spun again to 50 μL . Samples were taken after each spin and the experiment was performed in duplicate. The filtrate and retentate were analysed via SDS-PAGE against an un-concentrated DDM_AgrB2 control. To ensure low levels of DDM_AgrB2 were not passing through the filter, which would be a problem during long spin-concentration steps, the filtrates were pooled and spin-concentrated again, in a 10 kDa cut-off spin-concentrator which should reveal any lost DDM_AgrB2 that had passed through the 100 kDa cut-off spin-concentrator when analysed via SDS-PAGE.

The results (fig. 5.7) demonstrate absolute retention of DDM_AgrB2 by a 100 kDa cut-off filter. This permitted 100 kDa cut-offs to be used in future purifications. Since the DDM micelle is $\sim 72 \text{ kDa}$ ^[328] they should in all in theory pass through the filter. The design of the spin-concentrator filters means separation is not perfectly binary, and the manufacturer website recommends that filters with cut-off values that are at most half of the M_r of the sample are used (see the product page: <https://www.vivaproducts.com/vivaspin-centrifugal-concentrators/vivaspin-20.html>).

Similarly, micellar size is taken as the mean of a range of micellar sizes, which follow the normal distribution.^[329] Nonetheless using a large filter is convenient and should facilitate removal of some of the DDM. It should be stated explicitly that these results are valid only for DDM_AgrB2 in the SEC running buffer described previously as generating a monodisperse sample. It has already been shown that the oligomeric state of AgrB2 appears to depend on ionic strength of the sample buffer, and the sample used in this simple experiment was free of imidazole. Consequently, it may not be valid to use a 100 kDa cut-off spin-concentrator to reduce sample volume after Co^{2+} -affinity chromatography.

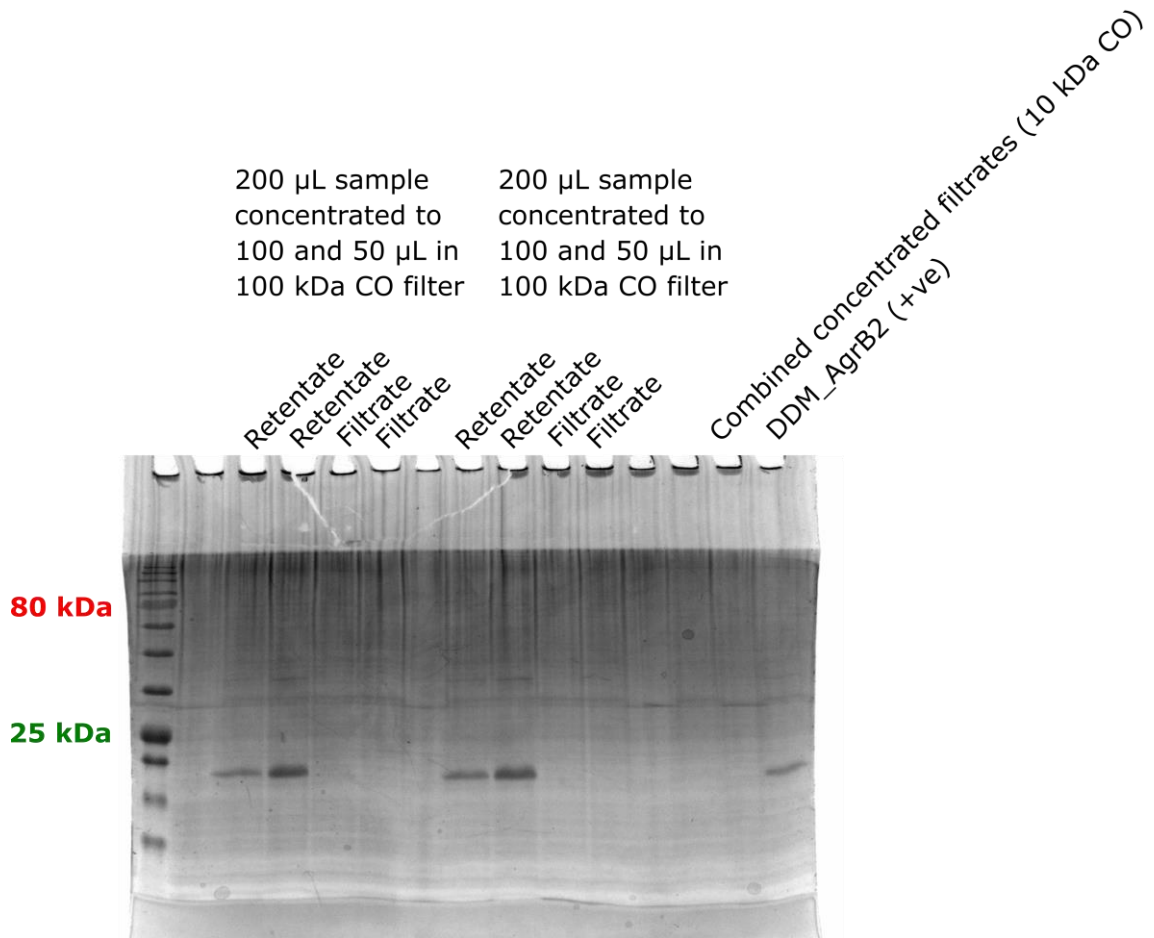


Figure 5.7. Gel result for a simple spin-concentration assay for DDM_AgrB2. The gel shows 100 % retention of AgrB2, when solubilised in DDM, by a 100 kDa cut-off (CO) spin-concentrator. This in turn suggests that the mass of the protein detergent complex is greater than 100 kDa, and that large cut-off filters may be used to aid detergent removal prior to crystallisation experiments. Note that the results may not be valid for all buffer systems. The gel is 16 % polyacrylamide, reducing.



5.8: Detergent Screening and the Oligomeric State of AgrB2

The detergent used to extract a protein from the cell membrane is often not the one best suited to crystallisation experiments. This is because, in trying to retain nativity of the sample and its environment, long-chain, non-ionic (so-called milder) detergents are chosen.^[50] Nor is it wise to limit crystallisation experiments *in surfo* to just one detergent if it is possible to try several. Longer detergent chains lead to larger micelles however, which are effective crowding agents that are not easy to remove.

Larger detergent molecules also more effectively shield the protein surface (which is after all the entire purpose of having the detergent there in the first instance) preventing the formation of protein...protein contacts during crystallisation experiments. Non-ionic detergents are not in themselves detrimental for crystallisation experiments but electrostatic attraction between adjacent point charges may aid formation of a crystal lattice. Since short-chain, ionic detergents are usually poor at embedding themselves into the cell membrane and extracting the intact protein sample, it is often necessary to extract the sample with a non-ionic, long-chain detergent and subsequently reconstitute the sample into a shorter-chain detergent later during the purification procedure.^[50]

There is some experimental freedom as to how the detergent is swapped. It may be performed during SEC provided the CMC is not so high that detergent usage becomes unaffordable. For high-CMC detergents, an additional metal affinity step may be introduced after SEC in which a step elution into a small volume is performed on the bench to reduce the buffer volume, and therefore mass of the high-CMC detergent required. Imidazole may be subsequently removed by spin-concentration/dilution steps into imidazole-less buffer.

Two additional detergents were nominated for crystallisation experiments for AgrB2 *in surfo* along with DDM. Consideration was also given to potentially using cryo-electron microscopy (cryo-EM) as a means to structure determination at a later date if a crystallographic approach was unsuccessful. In any case these detergents chosen were lauryldimethylamine-N-oxide (LDAO) and lauryl maltose neopentyl glycol (LMNG). A summary of key properties for detergents used in this section, plus N-



dodecylphosphocholine (FC-12) is given in table 5.1 and skeletal formulae of the detergent molecules are given in figure 5.8.

LDAO is remarkably hydrophilic despite its dodecyl chain, and has a CMC of ~ 1.5 mM.^[330] It also forms ~ 17 kDa micelles^[330] which are comparatively light and the detergent molecules have zwitterionic amine oxide head groups. It was considered to be particularly useful for crystallisation therefore. LMNG has remarkable properties that make it useful for crystallography and cryo-EM. Foremost it has a slower rate of dissociation from protein samples than DDM, and protein/detergent complexes may be reconstituted into detergent-less buffer,^[331] at least for a short time. This is obviously beneficial for generating interpretable micrographs. The micelles LMNG forms are ~ 400 kDa and appear to exhibit a broad size range.^[332] Such bulk is also useful for cryo-EM, which has a lower limit upon the size of the molecules for which it can be used efficaciously (52 kDa at the time of writing).^[333] The LMNG molecule is essentially two DDM molecules fused at the C2 carbon atom of the lauryl chain. The resulting quaternary carbon imparts some rigidity to the molecule which is desirable for crystallisation experiments. The large M_r of the LMNG micelle is probably counter-productive to crystallisation however and that is why it not the only additional detergent chosen for AgrB2 crystallisation experiments *in surfo*.

Before comparing the behaviour of AgrB2 solubilised in these detergents it should be stated that some crystallisation experiments of AgrB2 solubilised in FC-12 were performed as well. These samples were provided from other *S. aureus* group research experiments, and had not been subject to SEC. Typically the sample volumes were small and so the crystallisation experiments were not extensive but merely opportunistic. It is regrettable that there are no comparative size exclusion chromatograms and SDS-PAGE gels for these samples for comparison to those gathered by the experimenter. Similarly, a 1 g pot of DMNG, the decyl chain analogue of LMNG, was purchased but never used. Size exclusion chromatograms for DDM_, LMNG_, and LDAO_AgrB2 are presented in figure 5.9 and the corresponding SDS-PAGE gels are presented in figure 5.10.

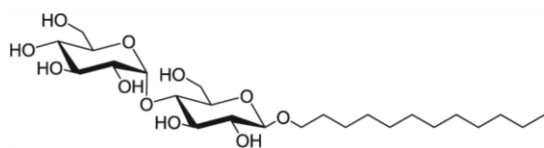


Detergent	M_r , g mol ⁻¹	CMC, % w/v	Aggregation N ^o	Micellar M_r , kDa	a Radius, Å	b Radius, Å
DDM	511	0.0087 ^a	78 – 149	37 – 76 ^c	13.8 – 14.3 ^e	28.0 – 29.5 ^e
LMNG	1005	0.001 ^a	391 ^b	235 – 622 ^d	NA	NA
LDAO	229	0.012 ^a	76	17 ^c	30.6 ^f	19.4 ^f
FC-12	351	0.047 ^a	54	19 ^c	20.7 – 21.2 ^e	13.4 – 13.6 ^e

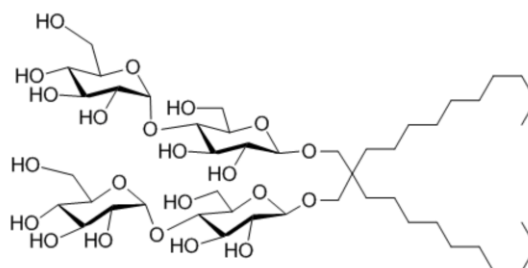
Table 5.1. Summary of key data for detergents discussed in this section. ^aMeasured in H₂O and available from Anatrace. ^bHas not been experimentally measured but may be inferred from the corresponding micellar mass which has been experimentally measured. ^cCalculated as the product of the M_r and aggregation number, where the latter is available from Anatrace. ^dCalculated via SEC. Values represent the peak limits. The retention volume of the peak was reported as 393 kDa^[332] and the explanation for the broad elution profile of LMNG micelles was attributed to it potentially forming rod-shaped micelles. It may also be due to adsorption of the detergent molecules to the column stationary phase, however. ^eMeasured in 20 mM “phosphate” buffer (counter ion not specified), pH 6.2 and 150 mM NaCl at 25 °C. These radii are derived from an ellipsoidal fitting of SAXS data^[334] and correspond to the alkyl chain length only. The dimension of the head groups were determined to be 6.3 Å for DDM and 3.0 Å for FC-12. The head group dimensions were determined to be directionally invariant. ^fMeasured in H₂O at 22 °C. The radii are derived from an ellipsoidal fitting Small-Angle Neutron (SANS) data.^[335] Measurements in higher ionic strength buffers showed expansion of the LDAO micelle along the a radius.^[335]



DDM: N-Dodecyl- β -D-maltoside



LMNG: Lauryl maltose neopentyl glycol



LDAO: Lauryldimethylamide-N-oxide



FC-12: N-Dodecylphosphocholine

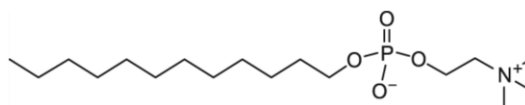


Figure 5.8. Skeletal structures for detergents discussed in this section. DDM and LMNG both have maltose head groups and are both β anomers. LMNG is essentially two DDM molecules fused at the C2 carbon atom of the dodecyl chain, which is therefore quaternary. LDAO and FC-12 are zwitterionic, containing both positive and negative charges in their respective head groups.

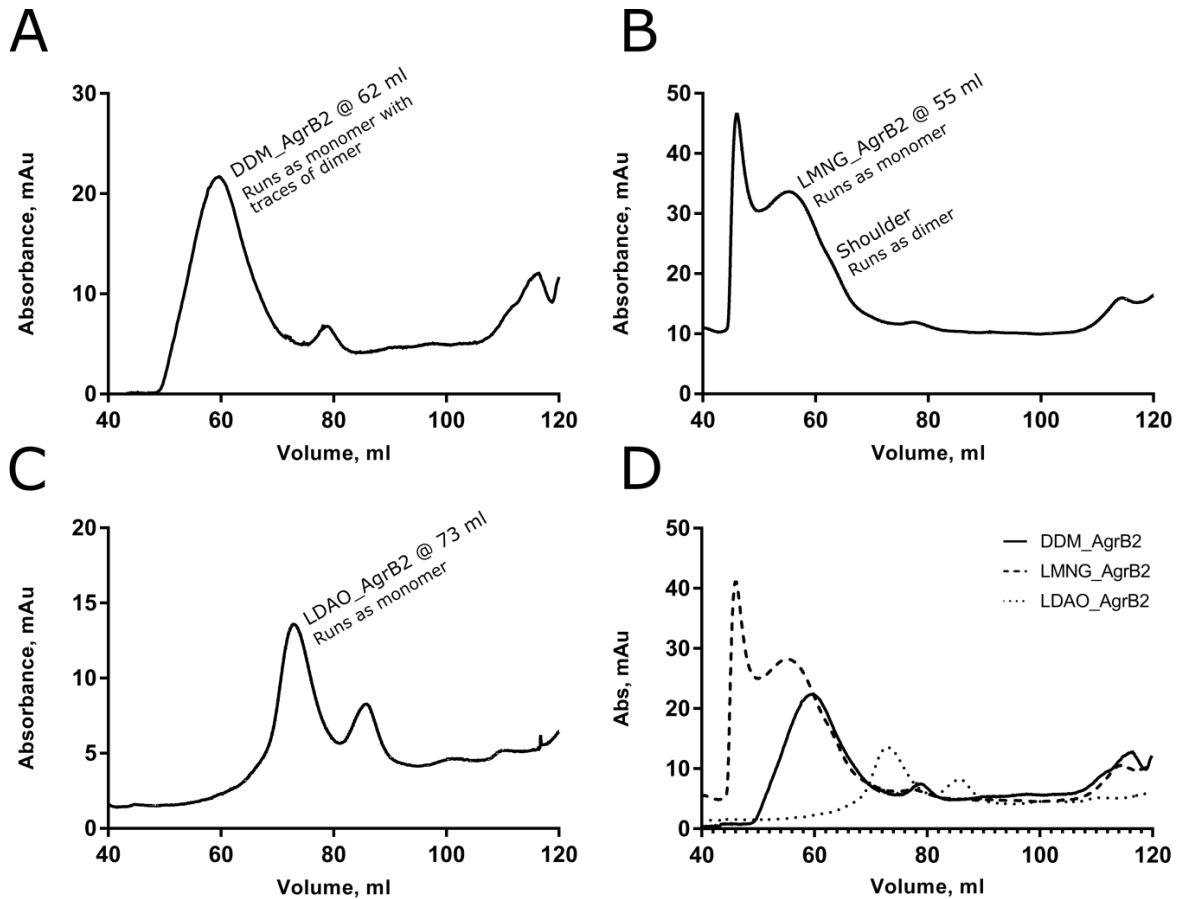


Figure 5.9. Comparative size exclusion chromatograms for AgrB2 reconstituted into different detergents. The primary sample peak is labelled with the retention volume and with a brief note as to how said sample runs on SDS-PAGE. (A). DDM_AgrB2. (B). LMNG_AgrB2. (C). LDAO_AgrB2. (D). A superposition of (A), (B), and (C) on the same scale. Generally speaking, the sample elutes with a retention volume corresponding to the detergent micellar mass. That a portion of LMNG_AgrB2 elutes in the void in (B) may not be due to sample aggregation but simply because a portion of the LMNG_AgrB2 sample is too large to enter the accessible volume. An analytical gel filtration was performed on the void peak to analyse this speculation but the sample precipitated upon defrosting. Although the supernatant was run the chromatogram (not shown) was featureless. A large increase in sample retention volume is observed for LDAO_AgrB2. A small peak that elutes later than all of the AgrB2 samples is observed and addressed in the ensuing text. The same running buffer (already given) and column, a HiLoad® Superdex® 200 16/600, was used for all purifications.

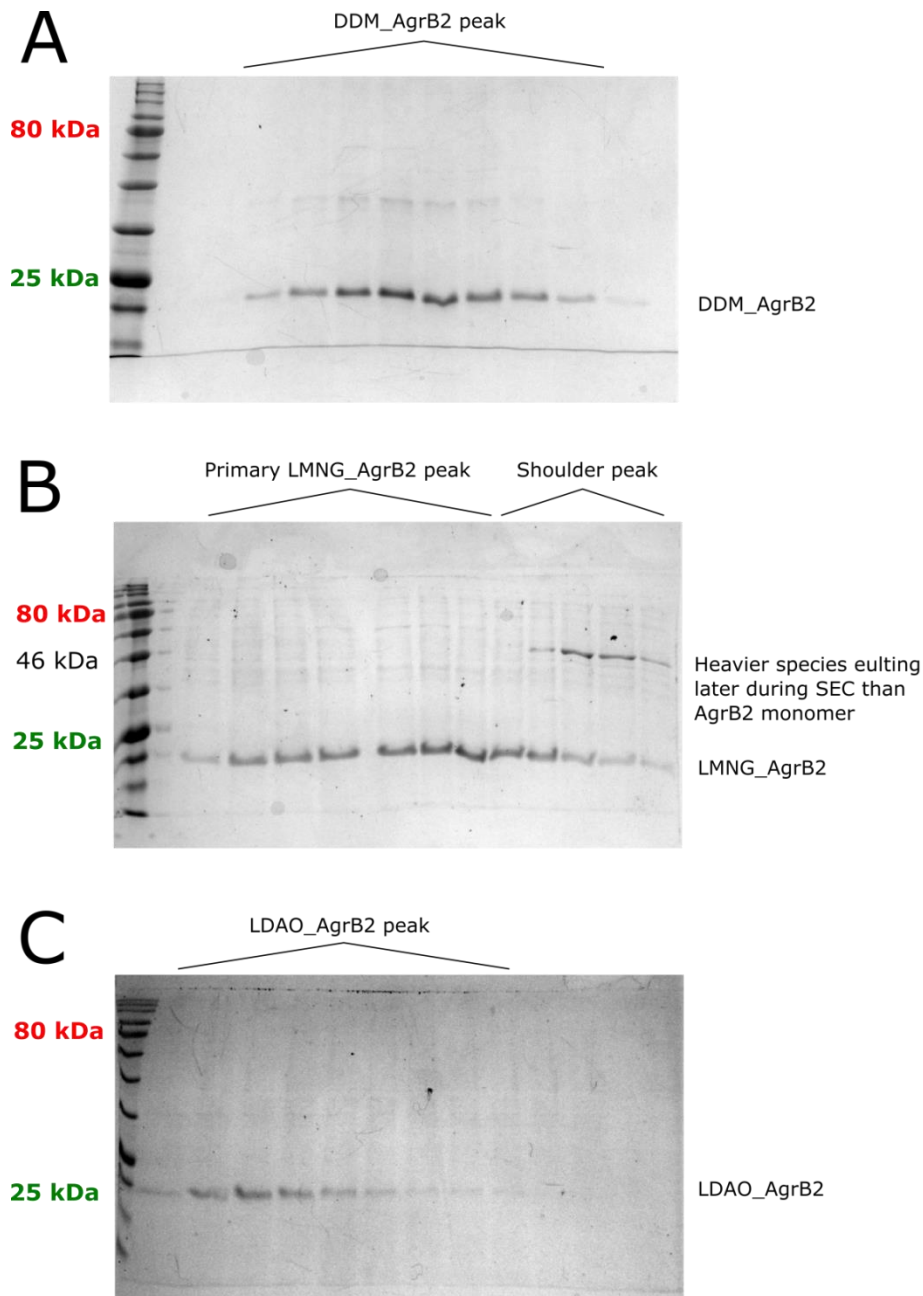


Figure 5.10. SDS-PAGE gel data for AgrB2 reconstituted into different detergents. (A). DDM_AgrB2. (B). LMNG_AgrB2. (C). LDAO_AgrB2. LDAO_AgrB2 runs entirely as a monomer. DDM_AgrB2 exhibits a faint band with an M_r corresponding to that expected for an AgrB2 dimer, that was common on many SDS-PAGE gels. It appears to co-elute with the portion of sample that runs as a monomer, however. For LMNG_AgrB2 this apparent dimer elutes later than the portion of sample that elutes as a monomer, which is the opposite way around to how size exclusion chromatography separates molecules. Gels are 16 % polyacrylamide, reducing.



The DDM_AgrB2 sample elutes at ~ 62 mL on a HiLoad® Superdex® 200 16/600 column. This corresponds to an apparent M_r of 250 kDa according the column calibration curve (given in appendix II). The M_r of a DDM micelle is ~ 72 kDa and the M_r of AgrB2 is 23 kDa. Clearly there is a discrepancy then between the actual M_r of the DDM_AgrB2 unit and that predicted by SEC, not forgetting of course that in analysing size exclusion chromatograms M_r is taken as being directly proportional to hydrodynamic radius. This discrepancy may be explained by either A) AgrB2 existing as a large oligomer or B) The DDM molecules adding such size to the AgrB2 monomer that it has a hydrodynamic radius equivalent to that of a globular particle of approximately ten times its detergent-less M_r .

Detergent micelles do not have a gap at their centre, so their size cannot be disproportional to their M_r due to that. Whilst micelles are not perfectly spherical, DDM micelles are more spherical than all of the following: FC-10, FC-12, OG, N-nonyl- β -D-glucoside (NG) and, 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS).^[334] To be precise DDM micelles are oblate spheroids approximately half as tall as they are wide, having a height of ~ 14 Å and a width of ~ 29 Å, plus a directionally-invariant head group layer ~ 6 Å deep.^[334] Nor are DDM micelles particularly big with their longest principle radius being 29 Å. Compare this to OG (42 Å), NG (60 Å), FC-12, (28 Å) and CHAPS (31 Å).^[334] Thus when there is talk of “swapping sample X into small micelles” clearly it is implied “micelles with lighter/charged head groups that break apart quaternary structures” since all of the micelles in the previous list, except for CHAPS, have detergent monomers with lower formula weights than that of DDM.

All of this seems to support, but not prove, option A). Looking at the SDS-PAGE gels, those which correspond to DDM_AgrB2 and LMNG_AgrB2 appear to shown an AgrB2 dimer band (this band recurred on all DDM_AgrB2 SDS-PAGE gels obtained after SEC). What is unusual for the LMNG_AgrB2 SEC chromatogram is that this apparent SDS-PAGE dimer elutes later during SEC than the monomer (the SDS-PAGE band corresponding to the heavier species elutes into later fractions than the band corresponding to the lighter one). This is the opposite way round to how SEC functions, in which larger molecules elute earlier than lighter ones. This dimer band elutes as a small right-hand

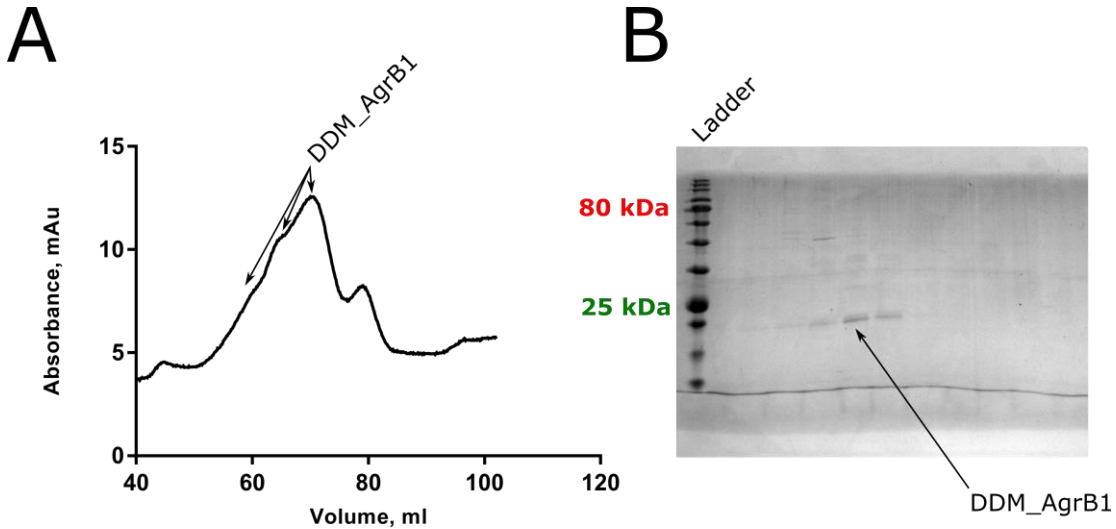


shoulder to the primary LMNG_AgrB2 peak. Furthermore AgrB2 is tryptophan-less, and the shoulder does not exhibit UV absorbance that is greatly different in magnitude to the LMNG_AgrB2 peak.

Returning to the apparent size discrepancy of the calculated and SEC-derived M_r of DDM_AgrB2, if the aforementioned band is not an AgrB2 dimer but actually a contaminant that gets co-solubilised with AgrB2, it must be similarly weakly absorbing. Furthermore, if it is not an AgrB2 dimer but contaminant that gets co-solubilised in LMNG micelles then this strongly suggests A) since it is more massive than AgrB2 (according to SDS-PAGE), solubilised in the same detergent as AgrB2, but elutes later on SEC than AgrB2. Therefore the entire protein/detergent complex must be smaller than that of the AgrB2/detergent complex. Presumably both species would possess a similar amount of detergent-added size and this would, when added to monomeric AgrB2, cause it to elute earlier and not later than the dimer/contaminant species. If this is not the case then the AgrB2 monomer must have a disproportionate amount of detergent bound to it, sufficient to make the protein/detergent complex elute even later during SEC than a protein of double the M_r solubilised in the same detergent. Therefore it is highly probable that DDM_AgrB2 is not a monomer under these purification conditions but a large oligomer of unknown size.

Examining the LDAO_AgrB2 size exclusion chromatogram, the LDAO_AgrB2 complex elutes at 73 mL. This corresponds to an estimated M_r of 93 kDa. The M_r of an LDAO micelle is 17 – 20 kDa, so there is still a size discrepancy but it is less than is observed for DDM_AgrB2. Furthermore there is no shoulder peak and no dimer/contaminant band on the corresponding SDS-PAGE gel. So if it is indeed a contaminant then it did not co-purify in this instance even though extraction of AgrB2 was always performed into DDM. There are no additional peaks on the size exclusion chromatogram to betray its presence also.

For completion, a size exclusion chromatogram and SDS-PAGE gel for DDM_AgrB1 is presented in figures 5.11A and B respectively. They clearly show multiple oligomeric species of AgrB1. A sequence alignment of AgrB1 and AgrB2 is also presented in figure 5.11C.



C
 CLUSTAL O(1.2.4) multiple sequence alignment

```

AgrB1      MNYFDNKIDQFATYLQKRNNLDHIQFLQVRLGMQVLAKNIGKLIVMYTTIAYILNIFLFTL 60
AgrB2      MNYFDNKIDQFATYLQKRNNLDHIQFLQVRLGMQIIVGNFFKILVTYSISIFLSVFLFTL 60
*****::: *:: *::* *::* :*::*****

AgrB1      ITNLTFFYLIRRHAGAHAPSSFWCYVESIILFILLPLVIVNFHINFLIMIILTVISLGVI 120
AgrB2      VTHLSYMLIRYNAHGAHAKSSILCYIQSILTFVFPYFLINIDINFTYLLALSIIGLISV 120
:::>::: *** :***** **:: **::**:: *:::* :*::*::*** : : *::*.* :

AgrB1      SVYAPAATKKKPIPVRLIKRKKYYAIIIVSLTLFIITLIIKEPFAQFIQLGIIIEAITLLP 180
AgrB2      VIYAPAATKKQPIPIKLVKRKKYLSIIMYLLVLILSLIIHPFYAQFMLLGILVESITLLP 180
:*****::***::*::***** :**:: * :*::*::***: :***: ***::*::*****

AgrB1      IFFIKEDLKLEHHHHHH 197
AgrB2      IFFPKEDLEHHHHHH-- 195
*** *****: .****

```

Figure 5.11. Purification data for DDM_AgrB1 and a sequence alignment of AgrB1 and AgrB2. (A). A size exclusion chromatogram for DDM_AgrB1, for which the running buffer was identical to that used for all other AgrB purifications hitherto presented. Multiple unresolved peaks are observed. (B). The SDS-PAGE gel corresponding to (A) showing monomeric AgrB1 regardless of retention volume. The gel is 16 % polyacrylamide, reducing. (C). A Clustal Omega sequence alignment of AgrB1 and AgrB2 sequences annotated as follows: (*), completely conserved, (:), highly conserved, (.), weakly conserved.



What may be definitively concluded from these observations is that SEC coupled with UV detection is not a good way to study the native oligomeric state of detergent-solubilised membrane protein samples. Light scattering detection^[213] would be much more informative. The conclusion upon AgrB2 oligomerisation that best fits the data presented is as follows:

AgrB2 exists as a large oligomer, probably a tetramer or hexamer and possibly a pentamer. This is comprised of AgrB2 dimeric subunits whose dimeric interface is more stable than the interface between these dimer subunits, as they sit in the oligomer. Consequently, a fraction of this dimer is partially preserved during SDS-PAGE when milder detergents are used to solubilise it (DDM and LMNG). This may be aided by the fact that the samples are not boiled prior to SDS-PAGE analysis. Reconstituting the DDM-solubilised oligomer into a harsh detergent (LDAO) effectively collapses this large oligomer into the monomeric form such that no trace of the oligomer is detectable on SEC or SDS-PAGE.

Here it is worth recalling that overexpression of AgrB and AgrD in *E. coli* cells is sufficient to detect the mature AIP signal molecule in the cell growth media.^[42] Transport of the cyclic AIP intermediate outside of the cell must therefore take advantage of a general bacterial peptide transport mechanism, occur spontaneously directly through the membrane once there is sufficient accumulation at the inner membrane to drive to this process, or occur via AgrB. If the latter is true then it is likely that AgrB would exist as a toroidal oligomer with a central cavity.

Lastly, it will be observed in all chromatograms the presence of an additional peak eluting at ~ 80 mL after the primary AgrB2 peak. This feature was common to all AgrB2 size exclusion chromatograms. The peak was collected, concentrated subject to SDS-PAGE, upon which it ran as a single 80 kDa species (fig. 5.12). This is in agreement with the M_r derived from its retention volume without invocation of any micellar mass. Given this, and that it exhibits a relatively strong absorbance signal during SEC but requires extensive concentration to be detectable on SDS-PAGE, it may correspond to contaminant that absorbs UV light strongly.

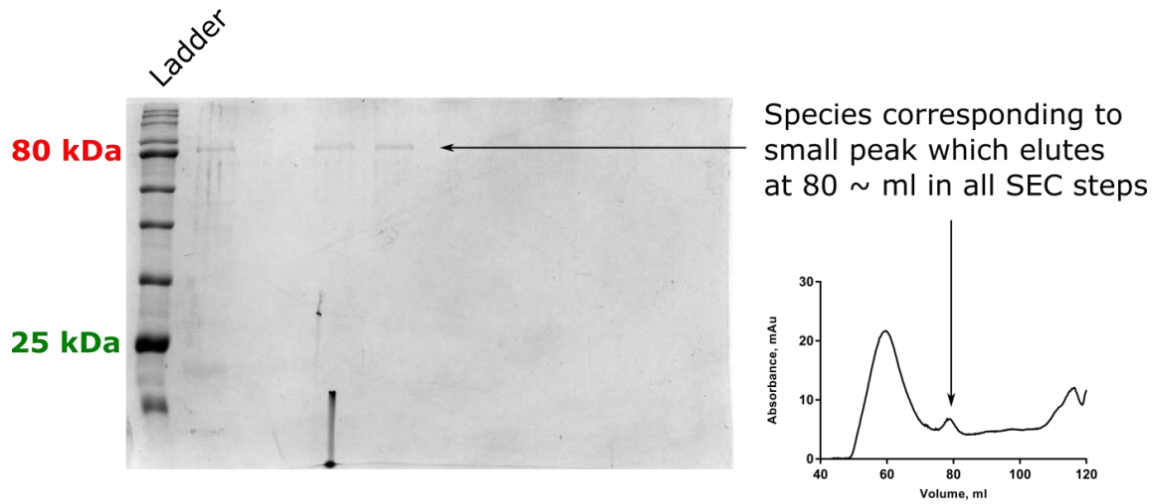


Figure 5.12. SDS-PAGE analysis of a peak common to all AgrB2 chromatograms. The peak fractions had to be extensively spin-concentrated in order to detect a band. The species runs at 80 kDa on the gel which corresponds to the mass derived from the peak retention volume column calibration curve. A representative DDM_AgrB2 chromatogram is inset to the left. The gel is 16 % polyacrylamide, reducing.



5.9.1: Structural Experiments on AgrB

5.9.2: Crystallisation Experiments *In Surfo*

A total of fifty-one 96 well sitting drop vapour diffusion experiments were established for AgrB1 and AgrB2 in DDM, LMNG, LDAO and FC-12. The sample concentration range for these experiments was 0.5 – 10.0 mg mL⁻¹. An exhaustive summary of these experiments can be found in the tables presented in appendix V.

No structure-grade crystals were generated by any of these experiments. There were entities which looked exactly like a crystal, with straight edges and regular faces (fig. 5.13). These occurred in many conditions and significant time was spent in trying to determine what they were and grow them larger. X-ray diffraction screening at Diamond Light Source failed to exhibit macromolecular X-ray diffraction, although they did exhibit a diffraction pattern that suggested a C₆ axis, which is in agreement with their shape (fig. 5.14B and C). They were also fished, ran on an SDS-PAGE gel and subsequently western blotted but revealed no significant protein content (fig. 5.15).

The question regarding what these entities were made of was somewhat solved by contacting the Structural Genomics Consortium via the Collaborative Crystallography Project 4 (CCP4) bulletin board. Reportedly, these entities are akin to liquid crystals of the protein/detergent complex and, in their experience, were not optimisable to structure-grade single crystals. Other members of the bulletin board have reportedly observed the same false positives. This lack of success in crystallising AgrB reflects the challenging nature of crystallising integral membrane proteins solubilised in detergent micelles. For the rest of the ensuing discussion these false positives are referred to as “pseudo crystals” despite not knowing for certain whether they fit the formal mathematical definition of a pseudo crystal.

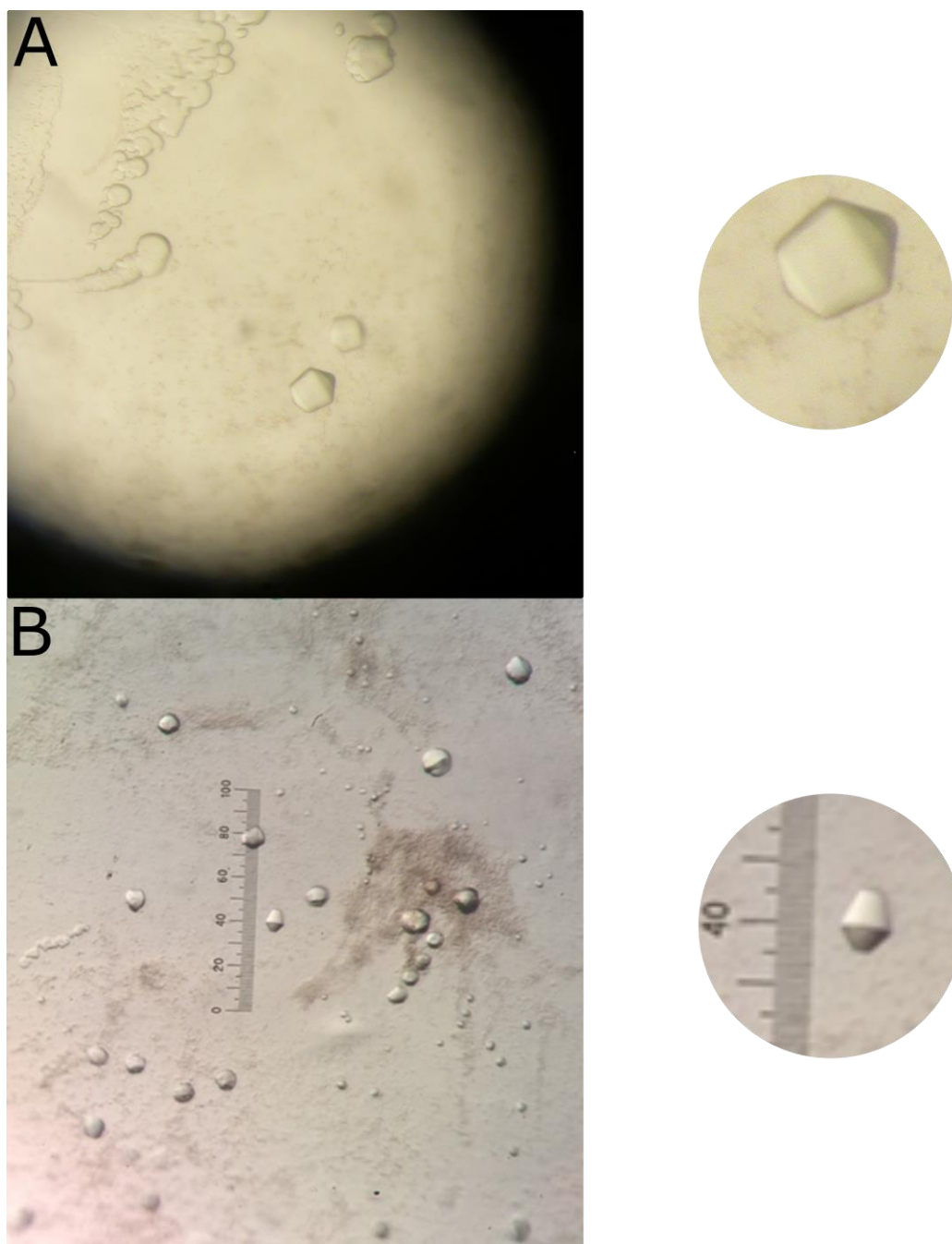


Figure 5.13. (A) and (B). Representative images of the apparent DDM_AgrB2 crystals shown with zoom-ins on the right highlighting the striking regularity of these entities. Western blotting, X-ray diffracting screening and communication with the Oxford Structural Genomics Consortium was later to inform that these were extremely deceptive false positives.

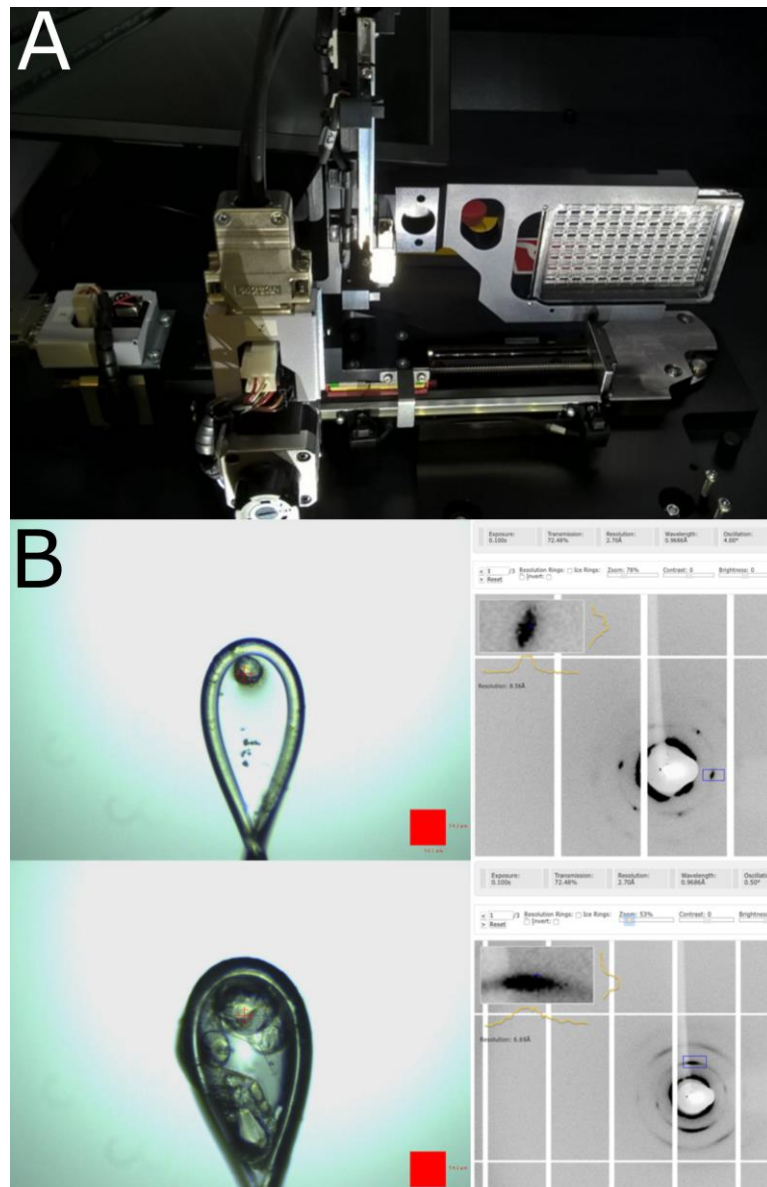


Figure 5.14. X-ray Diffraction screening experiments for the apparent DDM_AgrB2 pseudo crystals. (A). Owing to their extreme softness, attempts were made to shoot the crystals *in situ* on a home source Bruker D8 diffractometer equipped with a $\text{CuK}\alpha$ source. The source was too weak to yield any appreciable diffraction however, resulting only in blank diffraction images. (B). The pseudo crystals mounted on the goniometer at Diamond Light Source. For the topmost example the diffraction image exhibits six hexagonally-arranged, intense maxima. The bottommost example exhibits a pattern that is akin to a weak powder X-ray diffraction pattern.

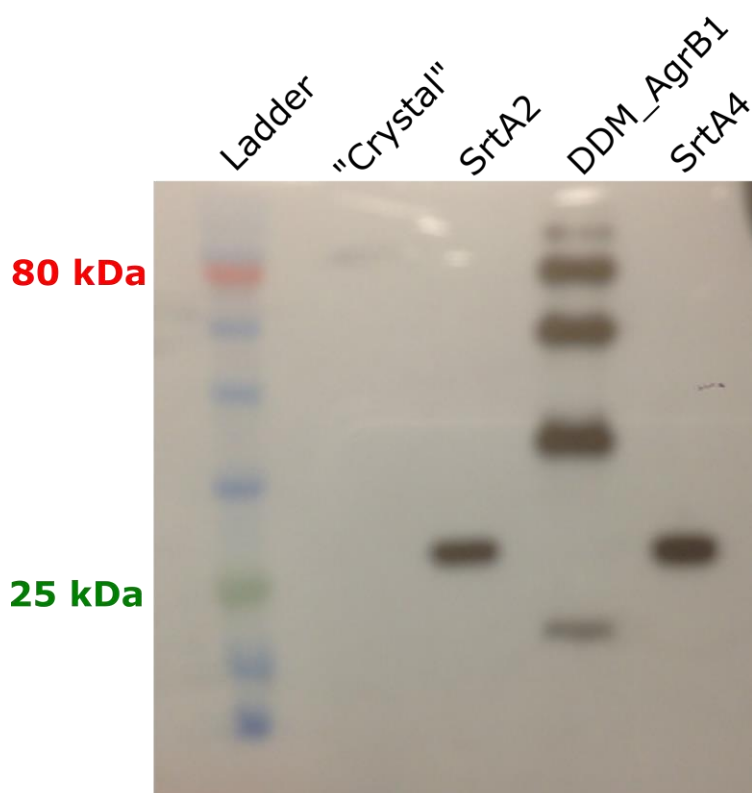


Figure 5.15. Anti-(His)₄ western blot for the DDM_AgrB2 pseudo crystals. A number of these were fished, washed in crystallisation solution and transferred to 3 x SDS-PAGE loading buffer. The controls are Sortase A from groups 2 and 4 *S. aureus*, and a solution-based sample of DDM_AgrB1. Incidentally the “laddering” of DDM_AgrB1 was a common observation for other researcher’s in the *S. aureus* group. An extremely faint band is observed at ~ 80 kDa in the lane corresponding to the pseudo crystals but insofar providing as evidence for these apparent crystals being “of” AgrB2, it is not very convincing. Whatever these entities are, their diffraction pattern and softness suggest a high liquid content, and probably some AgrB2 molecules have become included into them during formation or are stuck to its surface.



5.9.3 Crystallisation Experiment Attempts *In Meso*

Crystallisation of membrane protein in samples *in surfo*, when successful, generally results in so-called type II membrane protein crystals^[336] in which the crystal contacts are formed between the exposed hydrophilic regions that are not smothered in detergent molecules. The protein/detergent complex crystal subunits are usually packed at 90 °, or with a slightly obtuse angle to encompass the non crystallographically-ordered detergent molecules around the sample's hydrophobic girth. Such crystals are typically fragile and diffract weakly owing to their high solvent content (like the AgrA_C199S/P2_{fragment} crystals).

The popularisation of crystallisation experiments performed *in mesophase*,^[337,338] or simply *in meso*, can be attributed to Martin Caffrey and Vadim Cherezov.^[339] The method takes advantage of the complex phase behaviour of lipids in aqueous solutions to create a cell membrane-like lipid bilayer, into which the protein/detergent complex is reconstituted. Steady dehydration of this matrix promotes so-called type I^[336] crystals which, due to the orientational constraints enforced upon the protein molecules by the lipid bilayer, pack with the hydrophobic girth of the proteins forming the lattice contacts. An illustration of common membrane protein crystal types is given in figure 5.16.

The most common lipid used for crystallisation experiments *in meso*. It is a monoacylated glycerol (MAG) molecule in which the R group of the acylating agent is 18 carbon atoms in length and monounsaturated, containing a *cis* alkene with 9 carbon atoms lying to either side of the double bond (oleic acid).^[340] Different length R groups may be used and are often denoted as A:B monoacyl glycerol where A and B denote the number of carbons about the double bond. (monoolein is therefore 9:9 MAG).

The essence of the method is that, when monoolein is mixed with H₂O (or a sample solution) at 60:40 monoolein:sample at 20 °C, a liquid crystalline phase (a colloquialism of the term mesophase) of monoolein is formed with Pn3m symmetry.^[341,342] This phase may be thought of as a contiguous ordered array of fused monoolein vesicles, into which the protein diffuses whilst simultaneously shedding the detergent molecules hitherto used to solubilise it. It is assumed that the sample will



retain its native fold and activity. This viscous mesophase is then overlaid with a crystallisation condition which must not dissolve it or react with it. The exact process of how crystallisation then occurs is unknown.^[341] It has been suggested that components in the crystallisation condition may destabilise the mesophase, triggering various phase transitions, in one of which crystallisation occurs. Alternatively, the 20 °C isotherm of the monoolein:water phase diagram indicates that upon sufficient dehydration a lamellar liquid crystalline phase of monoolein bilayer sheets stacked on top of one another will form.^[341,342] Given the arrangement of molecules observed in type I crystals it is possible that crystallisation occurs in such a lamellar phase, but the ambiguity seems to be in how such a phase arises.

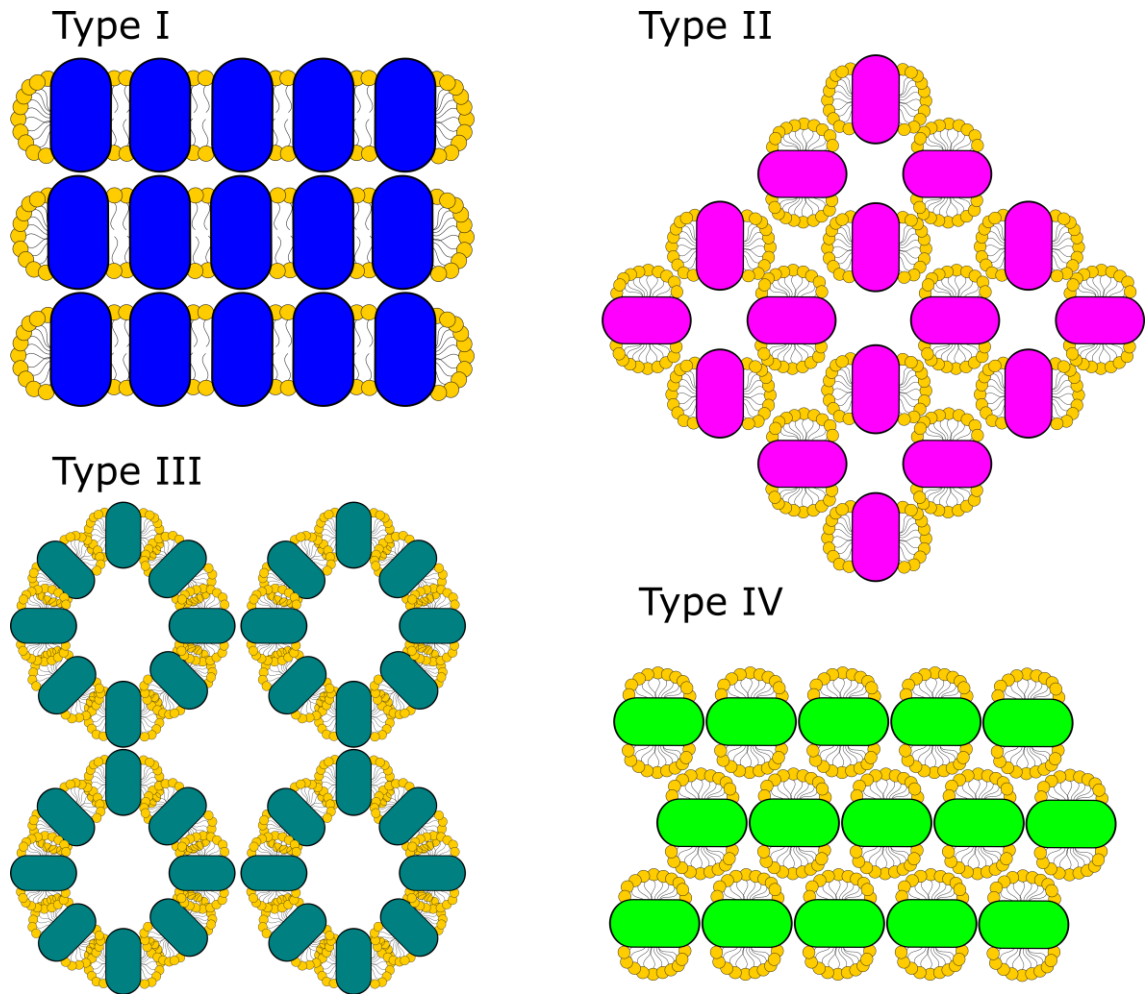


Figure 5.16. Simplified two dimensional representations of common types of membrane protein crystals.^[336] Types II, III and IV generally result from crystallisation *in surfo*. It will be observed that types II and III contain large solvent cavities, which is part of the reason such crystals tend to exhibit weak X-ray diffraction and are fragile. Type I crystals result from crystallisation in a lamellar phase as in *in meso* crystallisation. Unlike types II, III and IV, type I crystals form from contacts between the hydrophobic regions of the membrane protein sample.



Crystallisation *in meso* offers several advantages to crystallisation *in surfo*. The primary benefit is that the detergent molecules used to extract the sample from the cell membrane are forgone, along with all of the detrimental steric barriers they present to crystallisation. Additionally, the mesophase is assumed to be more hospitable to the protein sample than the detergent micelle.

Just because a method is beneficial it does not follow that it is easy to perform. To quote the Nature Protocols paper that outlines the method^[343] "...The cubic phase is extremely viscous (akin to a thick toothpaste) and sticky... many have tried to use the method, but have abandoned it in frustration because the material at the heart of the method, the cubic phase, was found to be difficult to prepare and to dispense." There are available robots which automate the process of setting up crystallisation trials *in meso*, however, these are extremely rare.^[339] The alternative, manual route to establishing the experiment proceeds thusly:

Monoolein, 60 mg, is transferred into one 250 μL Hamilton® syringe whilst sample, 40 μL , is transferred to another. The syringes are then coupled together and the two components manually mixed until they become translucent. The resulting mixture (the mesophase) is transferred into one of the syringes and the empty one is removed. A 10 μL Hamilton® syringe is then attached, into which the mesophase is transferred. The filled 10 μL Hamilton® is then decoupled and attached to a ratchet dispenser which when pressed will dispense ~ 200 nL, of the mesophase onto a glass plate. The commercial crystallisation kit solutions are then added on top of the mesophase drops and the experiment is sealed with another glass plate.

The kit necessary to perform this experiment upon AgrB was purchased but the experiment itself was not successful. Here it is worth highlighting some of the challenges of manually establishing crystallisation experiments *in meso*. The monoolein was supplied in 100 mgs in a blown glass ampoule. The monoolein has the same properties at room temperature as candlewax except that it is static. Consequently the scrapings are repelled by a spatula. Fortunately the monoolein can be melted and aspirated slowly and then allowed to re-cool. The air must be then squashed out of the monoolein



without ejecting it over the bench, to get an accurate measure of how much is in the syringe so the appropriate amount of sample can be added.

All of these steps proceeded successfully (fig. 5.17) but the terminal difficulty arose when transferring the mesophase to the 10 μ L Hamilton® syringe. Although the seal afforded by the barrel coupler was good for the two 250 μ L syringes, it was not mesophase-tight once the 10 μ L syringe was attached. Instead of being transferred into the smaller syringe, it was instead squeezed out of the sides at which point the experiment was terminated. The high cost of monoolein was not conducive to multiple attempts.

Incidentally, a (His)₁₀-tagged AgrB2 expression construct was used in conjunction with these crystallisation experiments. This purified identically to (His)₆-AgrB1 and not like (His)₆-AgrB2, and it was polydisperse (fig. 5.18A) further highlighting the subtle differences between the two allelic forms of AgrB.

The small amount of excess sample left over from the failed *in meso* crystallisation experiments was used to establish standard vapour diffusion *in surfo* crystallisation trials. In some of the drops it was noticeable that a few of the DDM pseudo crystals mentioned in section 5.8.2 seemed to have genuine needle-shaped crystals growing from them (fig. 5.19). These dissolved when cryoprotectant solution was added to them prior to X-ray diffraction screening.

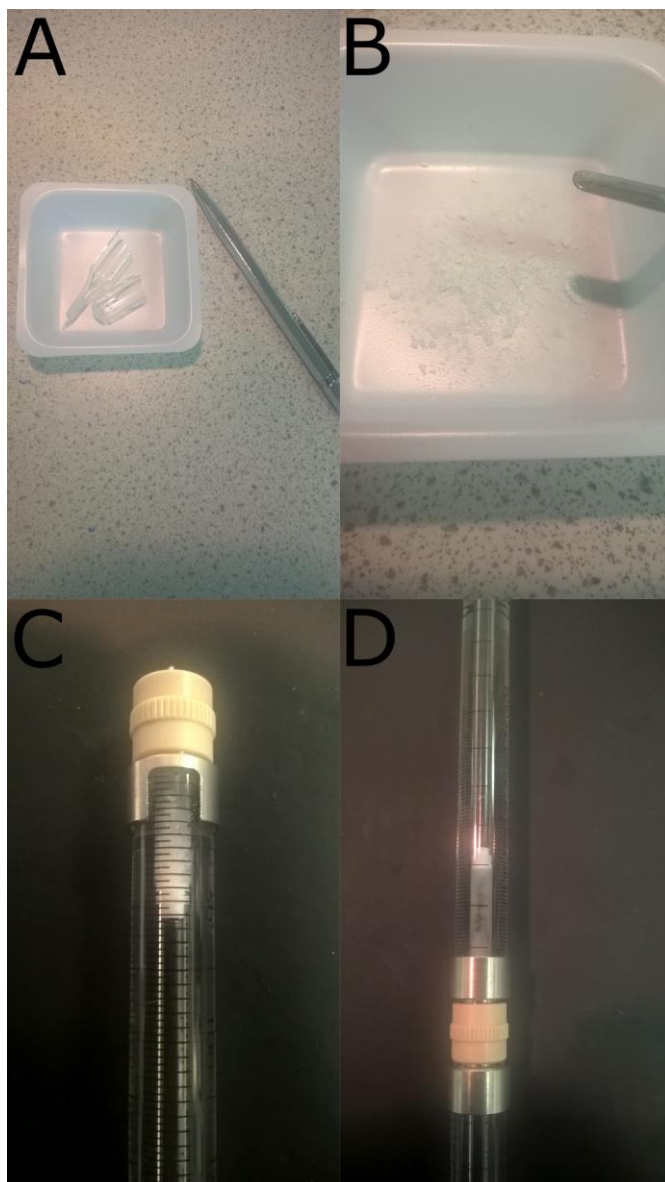


Figure 5.17. Pictures taken during the *in meso* crystallisation experiment for C-terminally (His)₁₀-tagged AgrB2. (A). The monoolein lipid as it arrived in the glass ampoule. (B). The powdered monoolein ready for melting. (C). The monoolein lipid successfully transferred to the 250 μ L Hamilton® syringe and starting to reset. (D). The monoolein lipid mixed with the purified DDM_AgrB2 sample via coupling of the two Hamilton® syringes and rapid mixing of the two samples. The mixture is starting to turn translucent presumably, as a result of formation of the cubic phase.

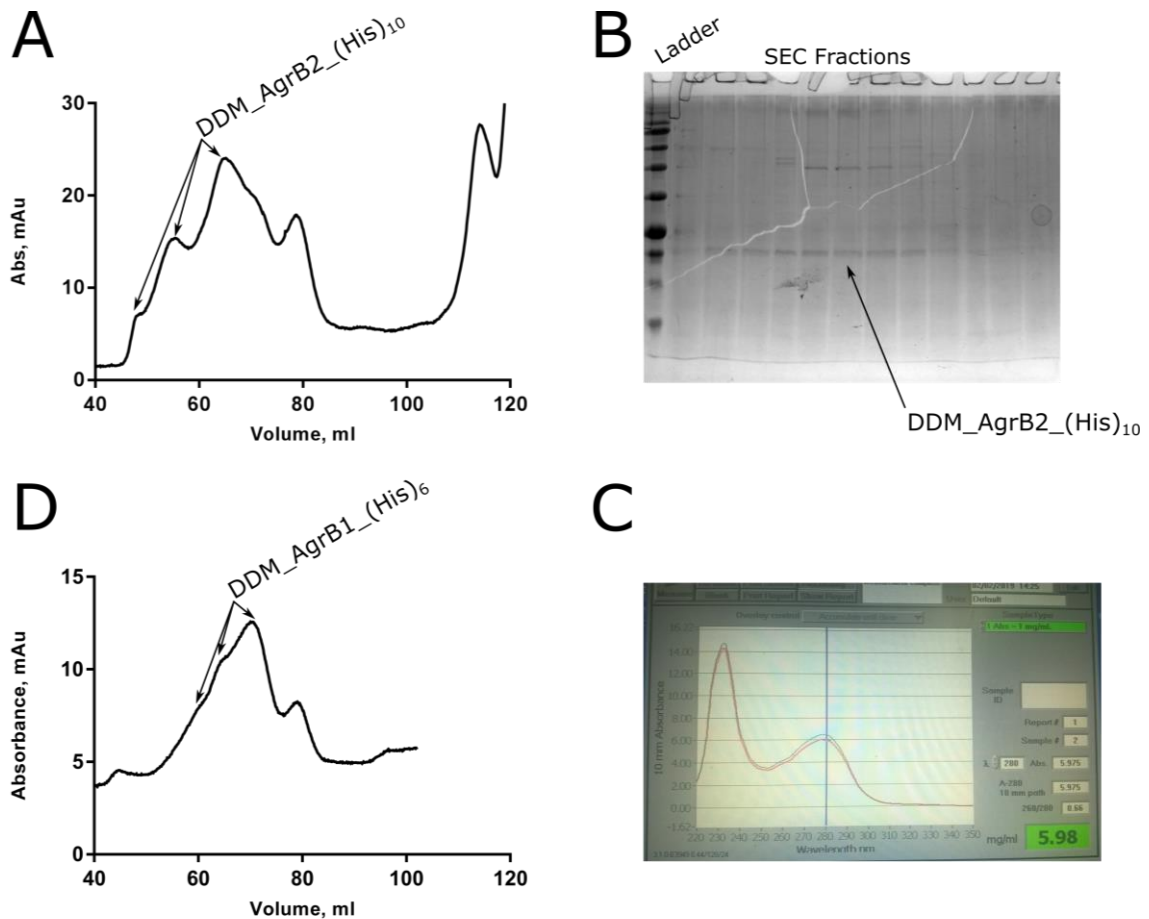


Figure 5.18. Purification data for (His)₁₀-tagged DDM_AgrB2. (A). A size exclusion chromatogram showing polydispersity. This is in contrast to the corresponding (His)₆-tagged sample which was shown in section 5.3 to give a Gaussian peak during SEC provided the appropriate running buffer is used. Said buffer was used here but clearly does not impart the same effect. (B). The corresponding SDS-PAGE gel showing a band corresponding to the purified sample. The gel is 16 % polyacrylamide, reducing. (C). The Nanodrop™ 1000 reading giving a raw absorbance reading of 5.98 absorbance units. This corresponds to DDM_AgrB2 concentration of 7.79 mg mL⁻¹, or 335 μM, once the theoretical extinction coefficient and appropriate M_r are substituted in to the Beer-Lambert equation. (D). The chromatogram corresponding to C-terminally (His)₆-tagged DDM_AgrB1. It bears resemblance to that shown in (A). Once again the same running buffer was used.

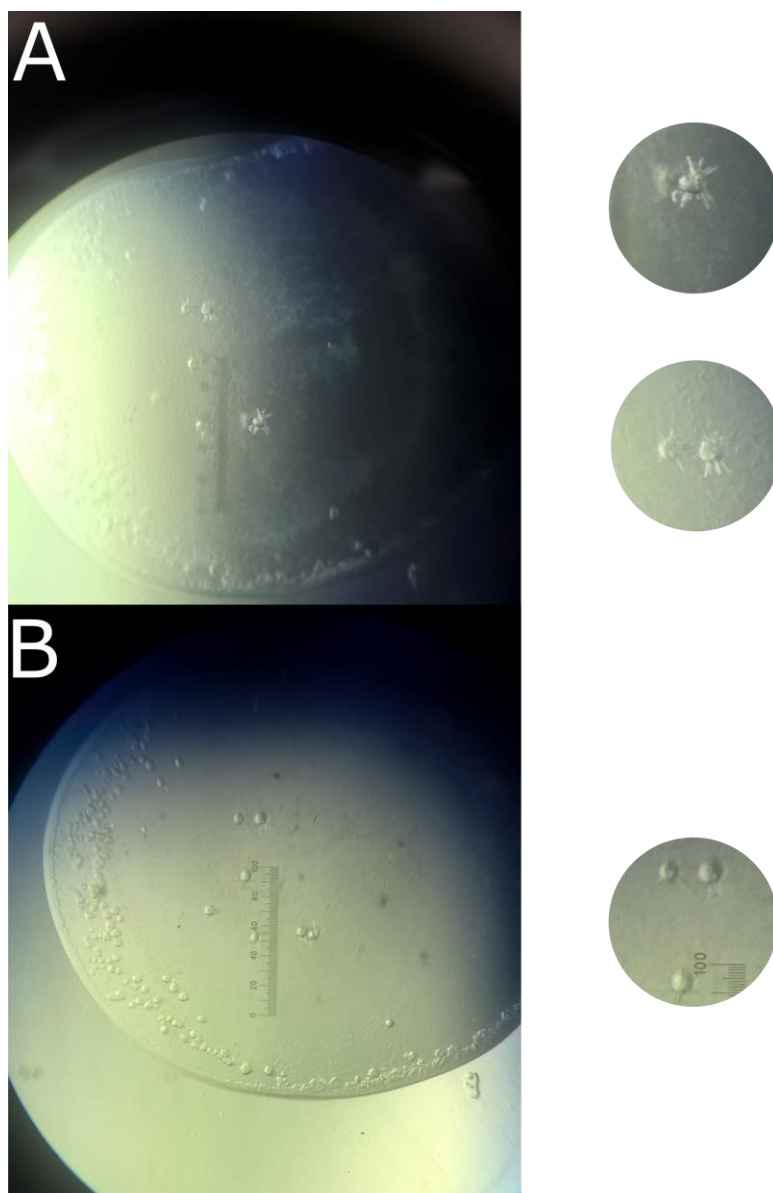


Figure 5.19. *In situ* crystallisation experiments for the (His)₁₀-tagged DDM_AgrB2 using sample that was left over from the failed *in meso* crystallisation experiment. Zoom-ins are shown to the right. Some hexagonal entities, which are the same as those pseudo crystals discussed in section 5.8.2, are observed. These have small crystalline appendages growing from them however, which dissolved once the cryoprotectant solution was added prior to X-ray diffraction screening.



5.9.4: Negative Staining at the University of Warwick

The awarding of the Nobel Prize in Chemistry to Jacques Dubochet, Joachim Frank and Richard Henderson for cryo-EM sparked an interest in pursuing this method in tandem to X-ray crystallography. DDM_AgrB2 and LMNG_AgrB2 were purified and subject to negative staining at the University of Warwick with the help of Dr. Saskia Bakker. Before discussing the micrographs obtained, a brief explanation of the process and goal of negative staining is presented below.

A small amount of purified sample solution is exposed to a small ($\sim 5 \text{ mm}^2$) carbon grid. It is hoped that the sample molecules settle onto the grid such that they are discrete, non-overlapping and ideally adopt a finite number of orientations into which the particles may be grouped. This latter point is crucial for full cryo-EM data collection. Achieving these criteria and producing a sample that yields discernible entities on the micrographs requires considerable optimisation, however. It is for this reason extensive negative staining of a sample is performed before proceeding to a full cryo-EM data collection. The process is similar to preparing grids for cryo-EM but the sample is not vitrified on the grids in liquid ethane and the microscopes used are standard Transmission Electron Microscopes (TEM) common in microscopy laboratories.

Once the sample has been exposed to the grids it is washed by adding a small drop of water onto the grid which is subsequently wicked away, followed by addition of an optically opaque stain such as uranyl acetate. When phosphate-based sample buffers are used, the washing step is essential due to its tendency to mineralise with heavy cations to produce, for example, uranyl phosphate. This was observed for AgrB2 negative staining attempts performed at the nanoscale and microscale Research Centre (nmRC) at the University of Nottingham, and manifested as large geometric silhouettes on the micrographs. After exposure to the stain, all excess fluid is wicked away and the grids are transferred to the microscope for visualisation. Ideally, the protein molecules appear as discrete light patches against a dark background, and possess some level of identifiable quaternary structure.

Micrographs for dilutions of DMM_ and LMNG_AgrB2 prepared from a 0.5 mg mL^{-1} stock are presented in figures 5.20 through 5.24. The granular patternation observed on all of the DDM_AgrB2



grids (figs. 5.23 and 5.24) is a consequence of too much protein/detergent complex sticking to the grids. An AgrB2 concentration of $0.0005 \text{ mg mL}^{-1}$ corresponds to $\sim 20 \text{ nM}$.

The LMNG_AgrB2 grids presented in figures 5.20 – 5.23 exhibit some interesting features. The larger grey ovals on the micrographs correspond to holes in the carbon on the grids and can be distinguished by their fine and even texture (more so than the rest of the micrograph). Despite too much protein sticking to the grids, there are entities which are approximately disc-shaped. Their size varies somewhat although none are orders of magnitude larger than others. This would agree with the breadth of the LMNG_AgrB2 size exclusion chromatogram and the reported broad size range of LMNG micelles^[332] assuming they correspond to such.

The choice of what constitutes an interesting or uninteresting region is somewhat arbitrary and the imagination is good at spotting patterns where none exist. For the sake of presentation however, the LMNG_AgrB2 micrographs have been delineated into three regions: holes in the grid (green circles), potential LMNG micelles (magenta circles) and potential protein/detergent complex molecules (burgundy circles). Some of the burgundy circles have been blown up and are presented in figure 5.25. Not all corresponding regions have been highlighted, just enough to give a representative idea.

Theoretically there should be no excess LMNG in the sample as the gel filtration running buffer did not contain any LMNG in accordance with the reported slow K_{off} rate.^[331] What is likely is that a portion of the “free micelles” co-elute with the LMNG_AgrB2 sample and so still pervade. This could potentially be overcome by taking the SEC-purified sample and re-binding it to a metal affinity column followed by a step elution into detergent-less buffer. This would guarantee as little contaminating LMNG as possible. In any case it appears that some of these entities visible on the micrographs exhibit a C_3 rotation axis. Micrographs with less background protein/detergent adhering to the grid and consequently better contrast would be needed to confidently draw conclusions.

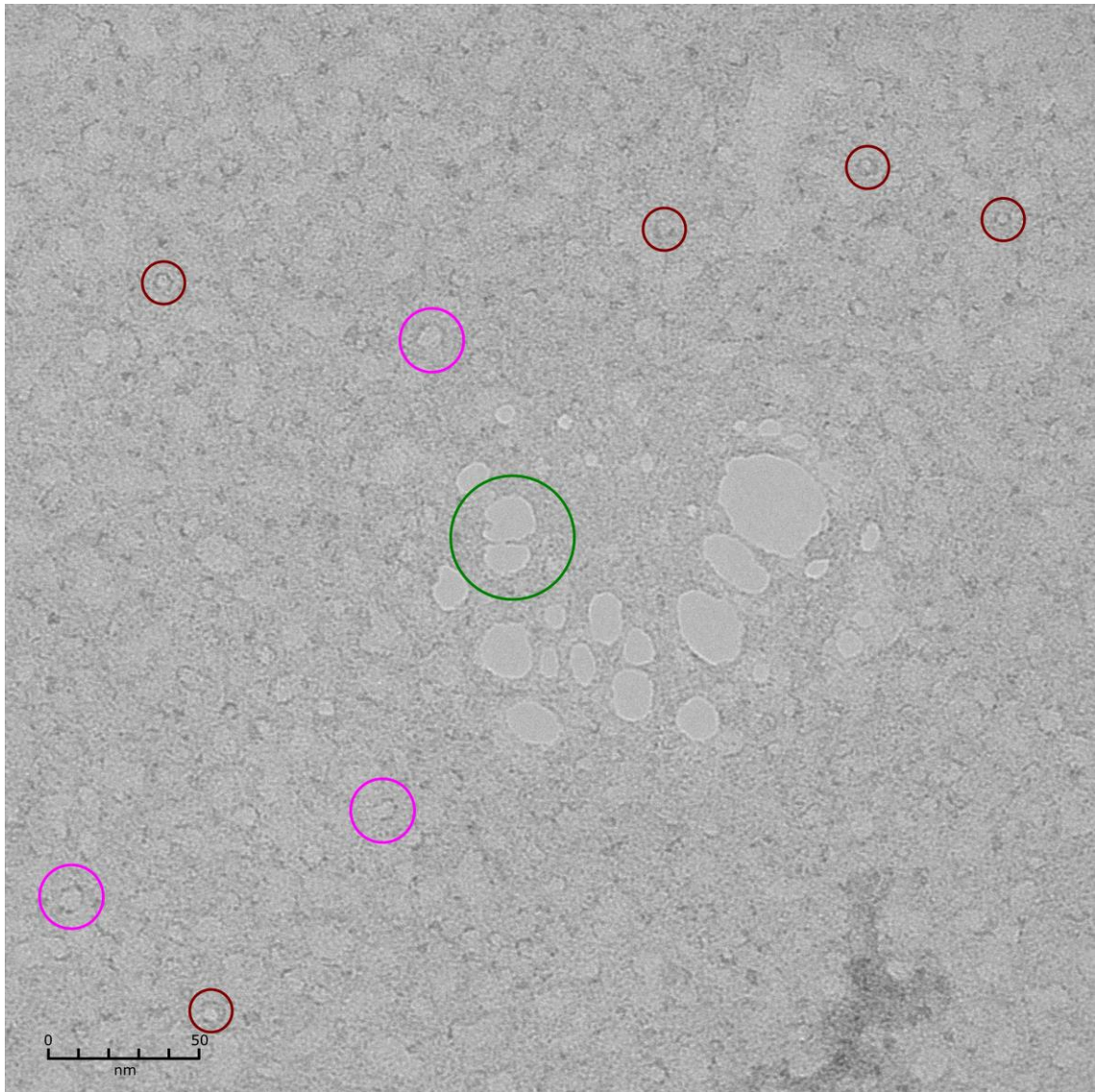


Figure 5.20. An LMMG_AgrB2 micrograph with features circled as previously indicated in the text.

The AgrB2 concentration is $0.0005 \text{ mg mL}^{-1}$, or 20 nM .

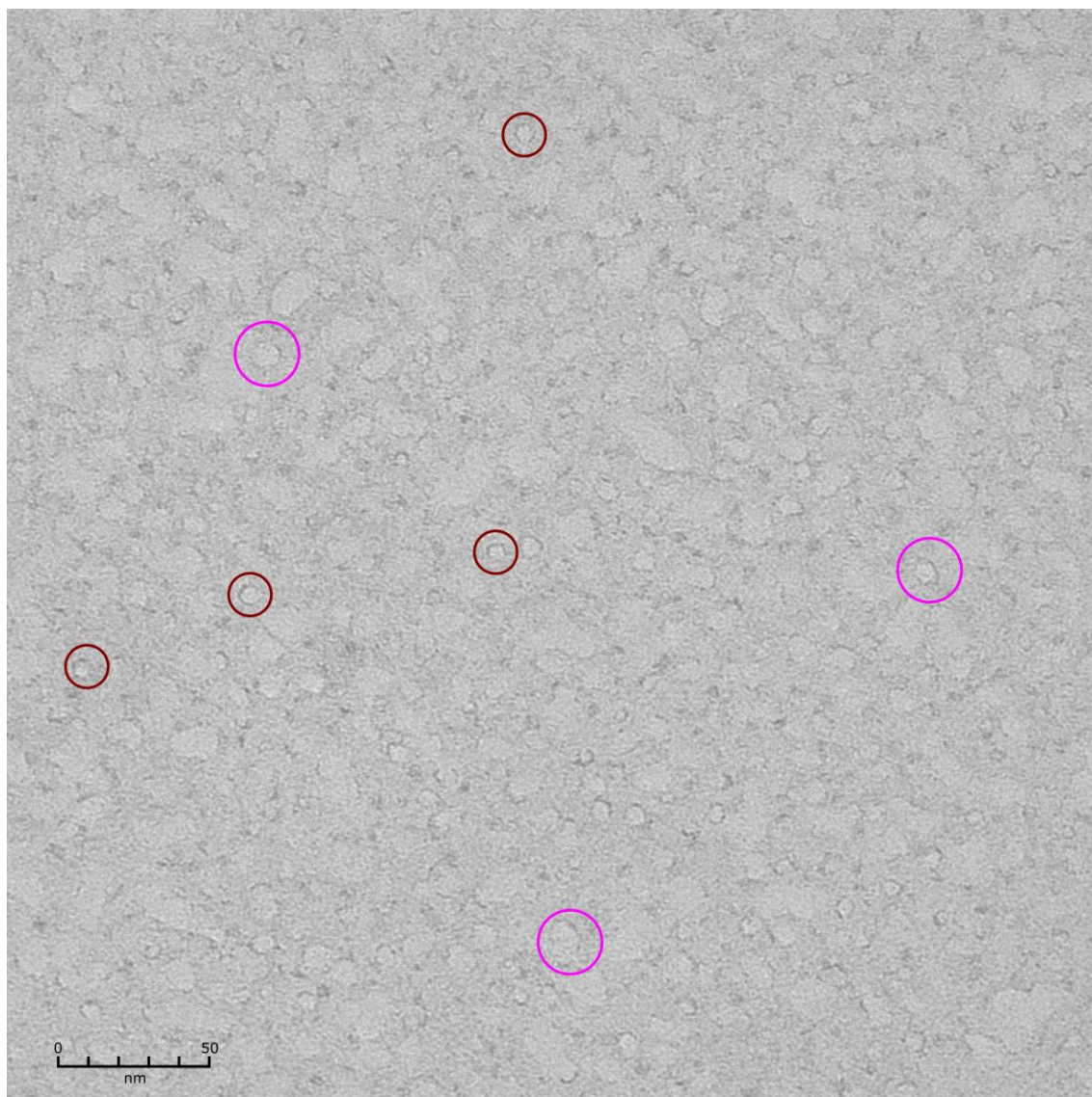


Figure 5.21. A second LMMG_AgrB2 micrograph with features circled as previously indicated. The AgrB2 concentration is $0.0005 \text{ mg mL}^{-1}$, or 20 nM .

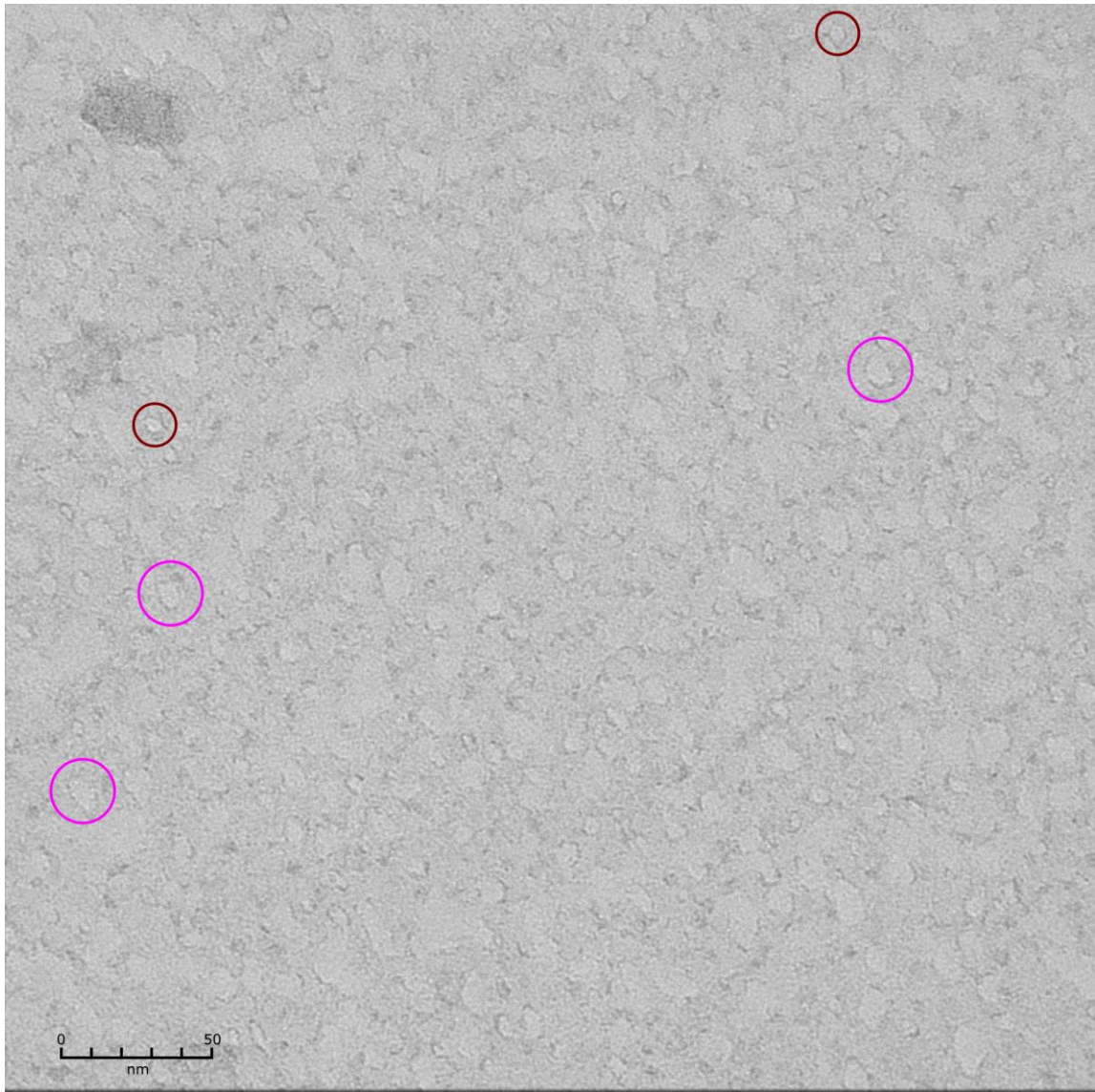


Figure 5.22. A third LMMG_AgrB2 micrograph with features circled as previously indicated. The AgrB2 concentration is $0.0005 \text{ mg mL}^{-1}$, or 20 nM .

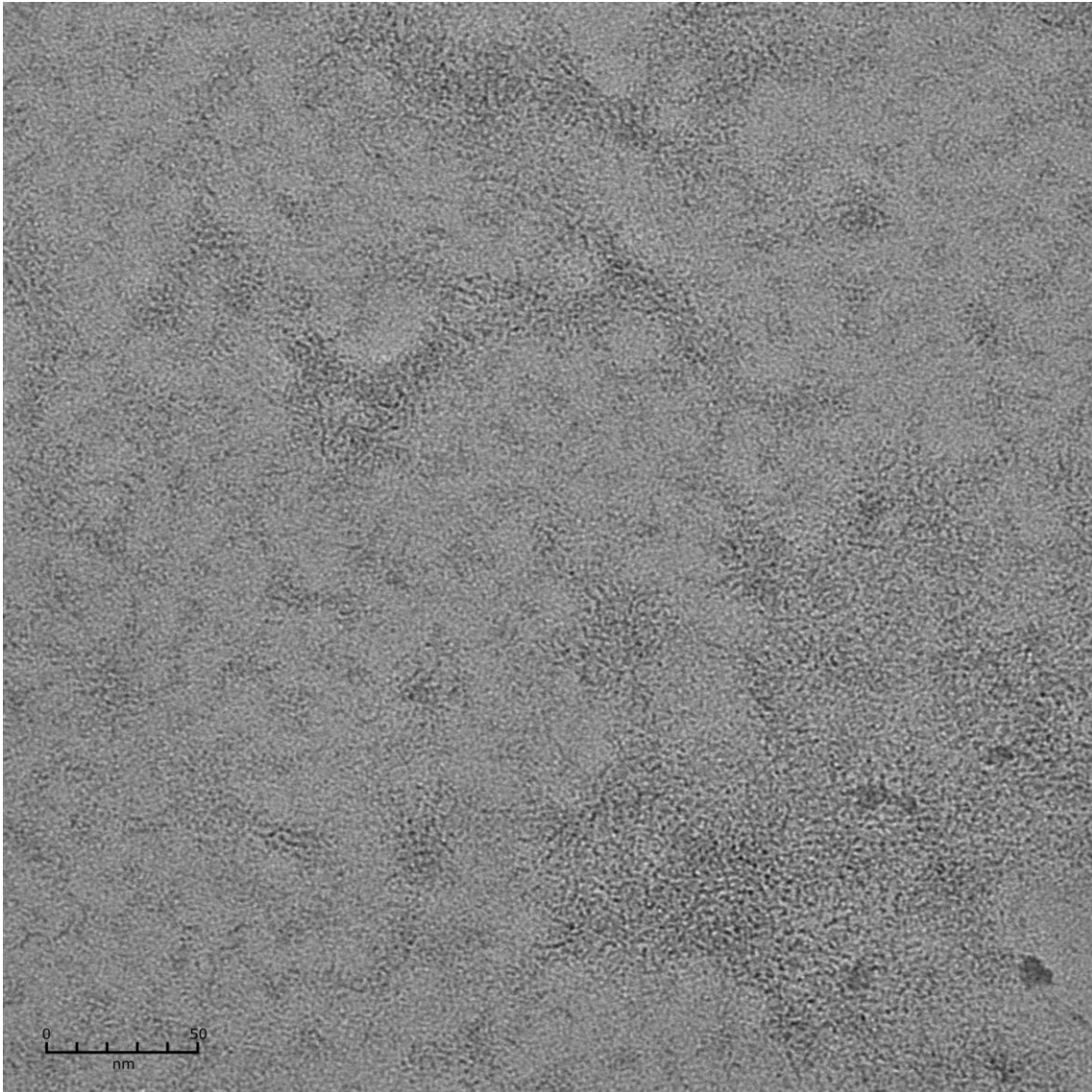


Figure 5.23. A DDM_AgrB2 micrograph. The granular patternation is due to too much protein/detergent complex sticking to the grid after wicking away the excess sample, or 20 nM.

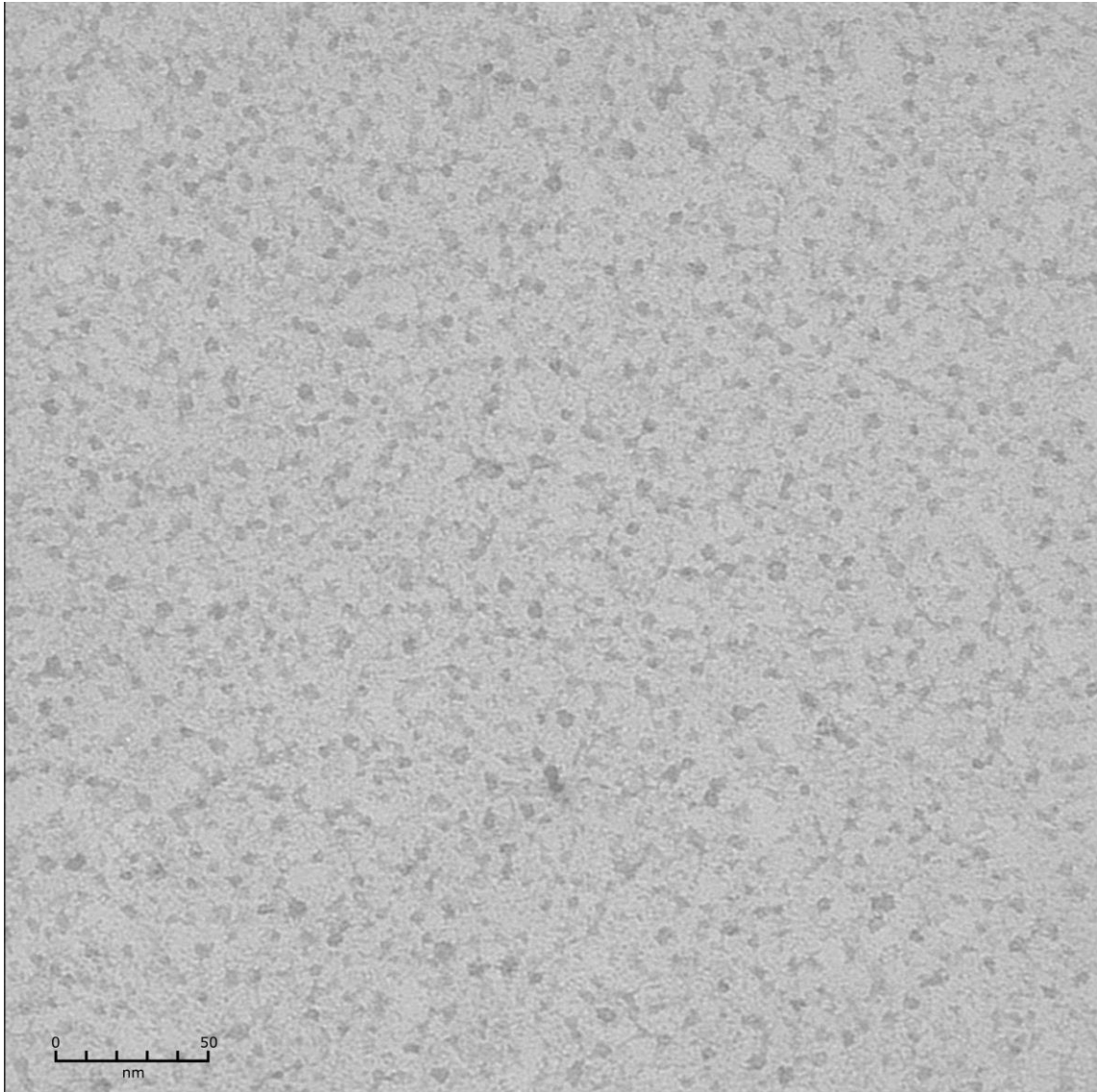


Figure 5.24. A second DDM_AgrB2 micrograph with the same pathology as the first. Again the sample concentration was $0.0005 \text{ mg mL}^{-1}$, or 20 nM . As this was already a one thousand-fold dilution of the 0.5 mg mL^{-1} stock, it was decided not to dilute the sample further.

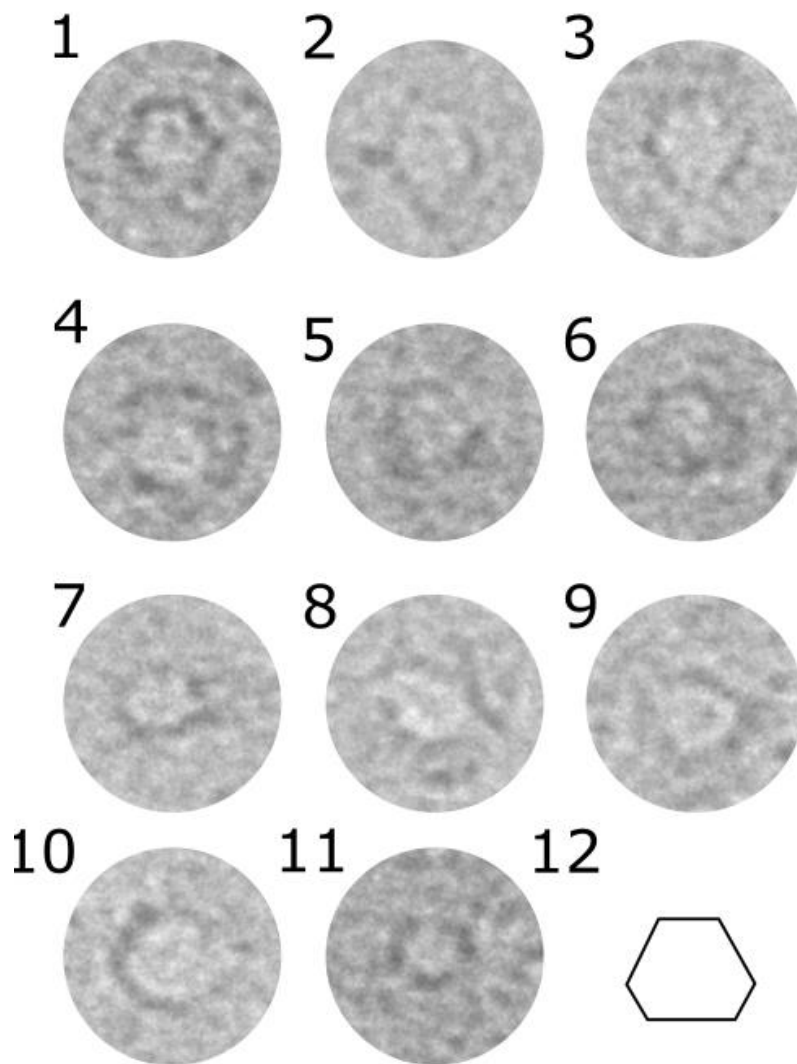


Figure 5.25. Some of the regions highlighted in the burgundy circles on the LMNG_AgrB2 micrographs blown up but still to scale. 1, 2, 3 and 4 bear the strongest resemblance to one another. 5 and 6 look similar to each other but appear to be distorted, especially at their lower left hand side. 12 is an estimated consensus shape of 1 – 11.



The *E. coli* formate transporter protein FocA (PDB: 3kcu and 3kcv) is a 31 kDa, 6 transmembrane helix integral membrane protein which forms a cyclic homopentamer, with a diameter of $\sim 80 \text{ \AA}$.^[344] This compares well to the observed diameter of $\sim 90 \text{ \AA}$ of the units highlighted in the burgundy circles on the LMNG_AgrB2 the micrographs. AgrB has an M_r of 23 kDa and secondary structure predictions also suggest 6 transmembrane helices are present.^[51] Consequently, if it were to form a cyclic oligomer the size may be anticipated to be in this size region.

Better micrographs may be obtained by: A) taking the peak SEC fraction and applying it the grids without any spin-concentration, which subsequent experience of negative staining has demonstrated to do more harm than good in most cases, and B) experimenting with “holey carbon” grids or graphene oxide grids. The former do not contain a back plate and the molecules are held in place by surface tension of the sample buffer, whilst the latter are more polarised than non-oxidised carbon-based grids. These may ameliorate the overcrowding of the grids observed for the micrographs presented. LMNG appears to be a more appropriate choice of detergent than DDM and holds promise for any future microscopy experiments undertaken upon AgrB.

5.10: Application to Instruct for Nanobody Preparation at Instruct-ERIC

AgrB2 has very little predicted extra-membranous structure, which is a problem for crystallisation *in surfo* since there is essentially nothing exposed to facilitate building of a crystal lattice. Nanobodies are light single-domain antibodies (12 – 15 kDa) that are generated during the natural immune response of camelids (camels, llamas, alpacas, vicuñas and guanacos). They have been employed successfully as crystallisation chaperones in structural determinations of difficult crystallisation targets. During the PhD project there was a call for applications to attend a course at the Belgium-based Instruct-ERIC site to learn about and develop nanobodies for samples provided by attendees. An application was made but owing to the competitive nature of the process and the limited number of places it was unsuccessful. The application documents can be found in appendices VI and VII. A nanobody generation and cocrystallisation project for AgrB is ongoing at the time of writing, details of which are provided in chapter 6 section 6.4.



5.11.1: Bioinformatical Interrogation of an AgrB1 Model

To gain insight into the structure of AgrB in the absence of experimental structural data a model, based upon the primary sequence of AgrB1, was generated in-house by the University of Nottingham biomembrane NMR research group. To interrogate the model and assess its validity, the AgrB1 sequence from group I *S. aureus* strain NCTC_8325 (used to generate the model) was aligned to AgrB1 sequences from three other *S. aureus* strains (group II strain MW2, group III strain N315 and group IV strain 6850), and fourteen other species in the staphylococcal genus (*saprophyticus*, *lugdenensis*, *schleiferi*, *simulans*, *hyicus*, *chromogenes*, *argentus*, *schweitzeri*, *warneri*, *hominis*, *haemolyticus*, *epidermidis* and *intermedius*) using the Clustal Omega server [EMBL].^[276] The sequence alignment is presented in figure 5.32. Sequence conservation was visualized in three-dimensions by mapping it onto the atomic surface of the AgrB1 model using the ProtSkin tool [McGill NMR Lab].^[345] An electrostatic surface of the model was also generated using the Adaptive Poisson-Boltzmann Solver (APBS) software suite (vers. 3.0).^[346,347]

5.11.2: Basic Topology of the AgrB1 Model

As inferred from the primary amino acid sequence, and in agreement with the CD data presented already, the model predicts AgrB to be almost entirely α -helical in secondary structure. An annotated schematic representation of the model is given in figure 5.26. There are six hydrophobic transmembrane (TM) α -helices all arranged approximately antiparallel to one another, and which presumably span the cell membrane in a multi-pass fashion. No re-entrant helices are predicted in the model. The N- and C-termini are located on the same face, in support HiBit assays performed by Dr. Ewan Murray (data not shown) which corroborates this and shows them both to be located on the cytoplasmic face. The two catalytic residues His77 and Cys84 are located at the N-terminal end, and mid-way through, TM α -helix 3 (TM3 and so on for the rest of this discussion) respectively. There are also two additional single-turn non-TM hydrophilic helices. One is located at the extreme N-terminus and the other in the loop region between TMs 4 and 5. All other regions between these helices and the TM helices are unstructured loops. The cytoplasmic face is lysine-rich, with ten such residues all



residing in close proximity and one instance of three lysines occurring consecutively. An additional three lysines are found on the extracellular face. Other basic residues such as arginine and histidine are also prevalent occurring with a frequency of six and five respectively. This abundance of solvent-exposed basic residues agrees with the high theoretical isoelectric point of 9.9.

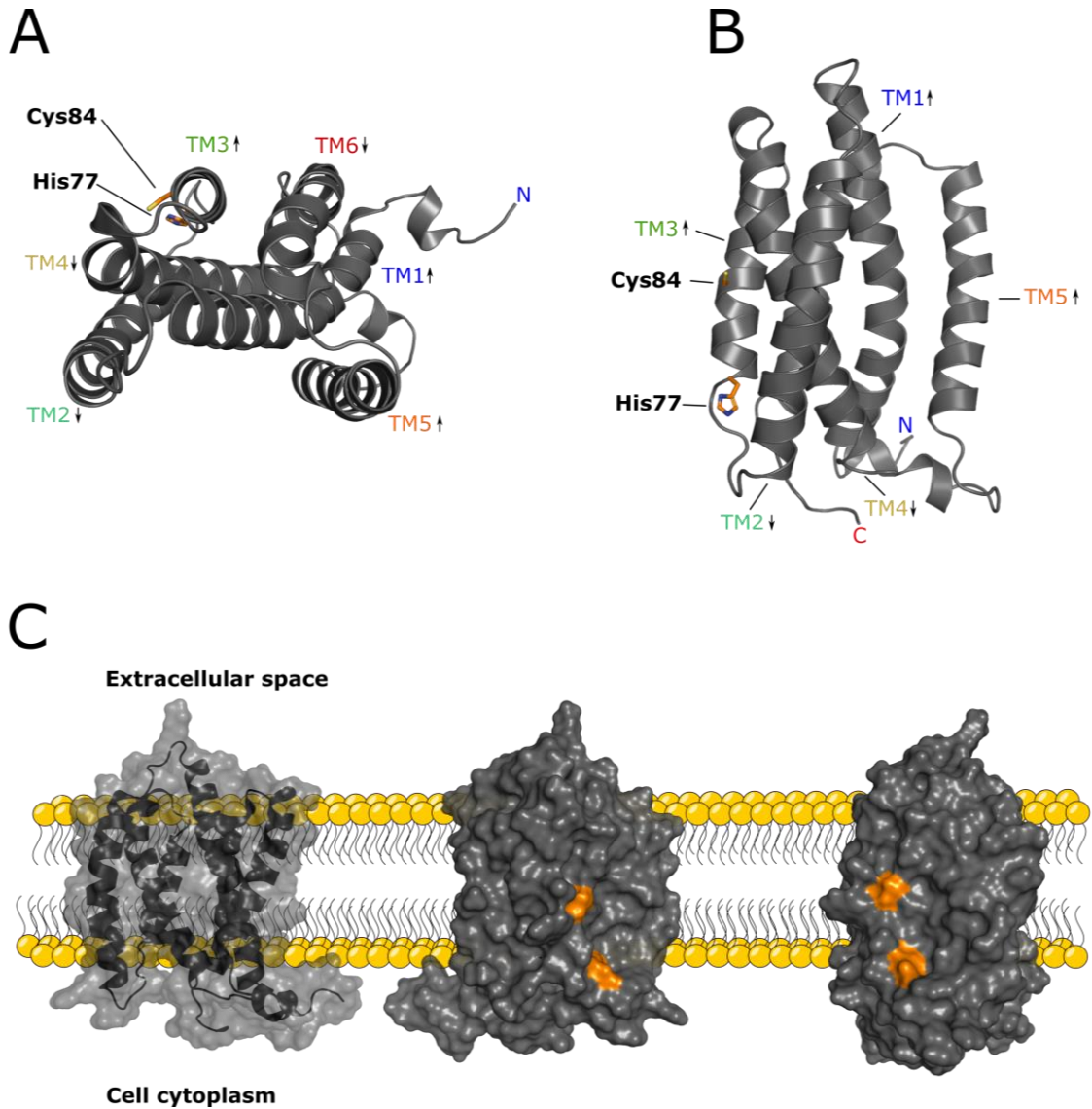


Figure 5.26. Various representations of the AgrB1 model. (A). Viewed top-down with respect to the TM helices. TM helix labels are coloured as a spectrum from blue-red according to standard N-to-C-termini rainbow convention. The arrows indicate the direction of the polypeptide chain relative to the cell cytoplasm (up is towards the extracellular face). (B). The same as (A) but viewed side-on with respect to the TM helices. The location of His77 and Cys84 on TM3 are shown. (C). Representations of how the AgrB molecule might sit in the cell membrane. In the surface representations the catalytic His77 and Cys84 residues are coloured orange.



5.11.3: Comments on Model Features With Respect to Sample Crystallisability

That the cytoplasmic face of AgrB1 is so lysine-rich (AgrB2 has eleven lysines) has important ramifications for crystallisation experiments of AgrB *in surfo*. In such experiments the hydrophobic regions of the protein are engulfed by a belt of detergent molecules, leaving only the hydrophilic loop regions exposed to form lattice contacts. For *S. aureus* AgrB, these corresponding regions are clearly dominated by entropic lysine residues which do not favour formation of lattice contacts owing to the loss of entropy associated with the solution-solid phase transition.^[279] This is clearly manifested in the complete failure of AgrB to crystallise *in surfo* and compounds the already-challenging task of crystallising an integral membrane protein. Inspection of the Clustal Omega sequence alignment (fig. 5.32) reveals only two of the thirteen lysines in AgrB1 to be completely conserved across all of the species aligned. This suggests that an AgrB homologue from a species other than *S. aureus* might yield more profitable crystallisation experiments.

5.11.4: Literature Observations Pertinent to Model Interrogation

From the outset it is difficult to interpret features the model because A) the role, if any, of AgrB in transport of the cyclised AIP intermediate outside of the cell is unknown. This would surely have important bearings upon AgrB quaternary structure and probably the tertiary structure as well and B) AgrB is the archetypical member of the “AgrB class” of integral membrane proteases, for which there is no representative structure available so it is not clear whether proteolysis occurs in the cell cytoplasm or within the cell membrane itself.^[51,52,53] There are some key studies however, which enable comments to be made on the AgrB model.

AgrB has been observed to immunoprecipitate as an oligomer containing an unknown number of monomer subunits and is active as at least a dimer *in vitro*.^[44,348] The size exclusion data for DDM_AgrB2 purifications presented in this thesis suggests that a large AgrB2 oligomer is being purified despite the extra detergent mass carried by the DDM_AgrB2 complex, however it runs as a monomer on SDS-PAGE. AgrB1 meanwhile runs predominantly as a dimer on SDS-PAGE but larger oligomers are clearly present.



Some authors have hypothesised in the absence of a clear transport mechanism that AgrB oligomerises to form a pore in the cell membrane through which the cyclised AIP intermediate is transferred outside of the cell,^[42,348] implying AgrB may have dual-functionality as both a protease and a transporter. This seems to be borne out by the observation that overexpression of AgrB and AgrD in *E. coli* is sufficient to detect mature AIP in the cell growth media.^[42] The sequence of AgrB has no ATP-binding cassette however, meaning it cannot take advantage ATP to drive this transportation, nor has there been discovered any gene on, or near to, the *agr* locus which may perform this transport role. It may be said with certainty that there is a strong bias towards symmetric oligomers amongst membrane proteins (~ 65 % of TM proteins are oligomers)^[349] owing to the lateral constraints imparted by the cell membrane.^[350]

5.11.5.1: Features of the Model

5.11.5.2: Surface Features

When viewed looking down onto the extracellular face there is one clear concave face, bound by TMs 1, 2, 4 and 5, and one convex surface bound by TMs 1, 3 and 6. The inside of this concave face is predominantly positively charged and highly conserved whilst the convex face is predominantly hydrophobic and is less well conserved. There is also highly conserved and positively charged “dogleg” formed by the non-TM α -helix at the very N-terminus. There is also a solvent-exposed patch that exhibits significant sequence diversity, and this is located at the C-terminal end of TM4 which corresponds to the extreme cytoplasmic side of the concave surface. Cartoon, electrostatic surface and conservation surface representations of the AgrB model are presented in figure 5.27.

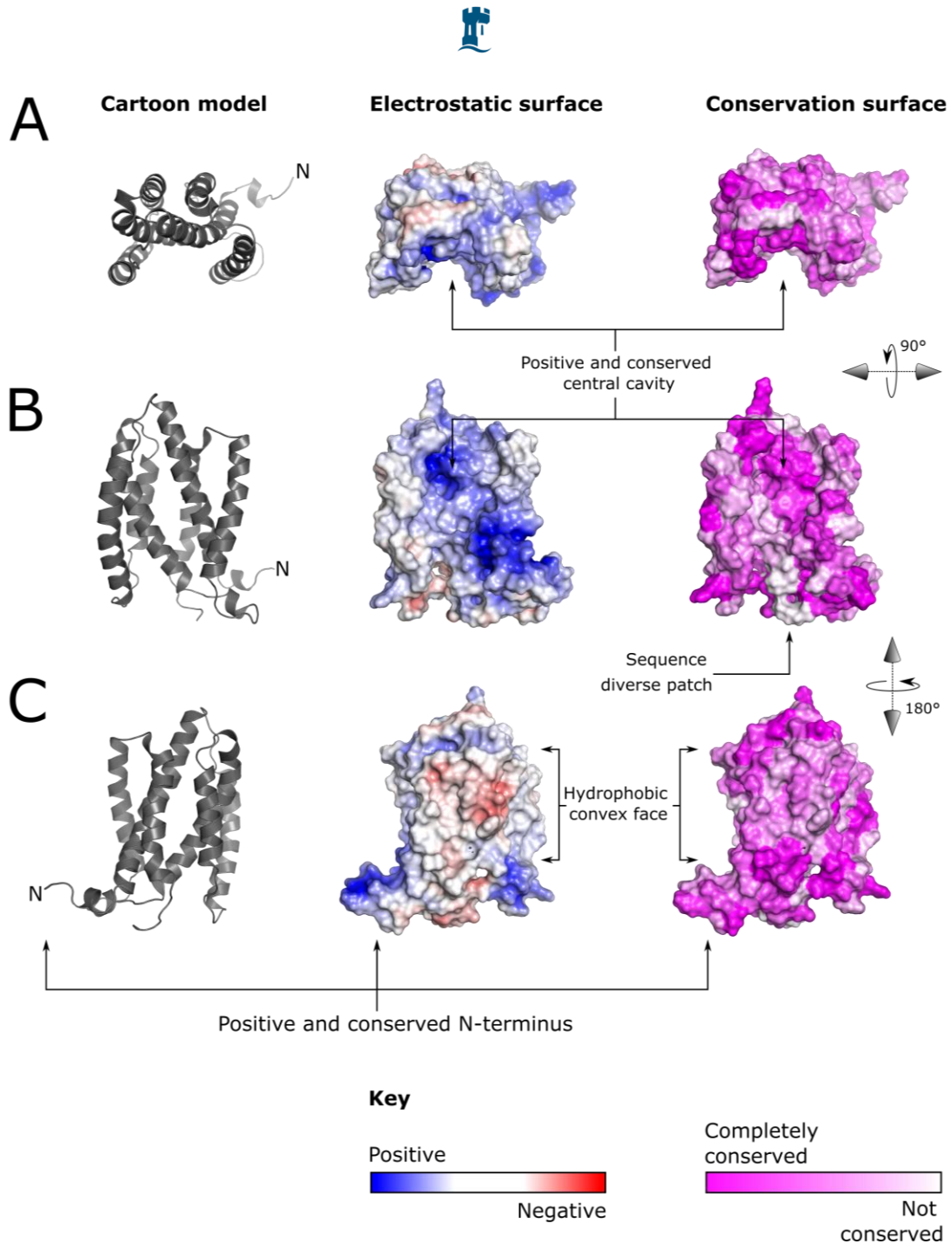


Figure 5.27. Three representations of the of the AgrB model. These are (left-to-right): a cartoon representation, an electrostatic surface calculated at pH 7.0 using APBS and a map of surface conservation created using the Clustal Omega output and ProtSkin. Each representation is displayed in three orientations: (A) top-down, (B) face-on and (C) rotated 180 ° vertically from (B). For the electrostatic surface representations, blue indicates regions of positive charge, red indicates regions of negative charge, and white indicates hydrophobic regions. For the conservation representations, darker pink represents more conserved regions and white represents unconserved regions.



5.11.5.3: Features in Relation to Residues That Drive AgrD Specificity

Chimeric studies in which AgrB1 and AgrB2 amino acid sequences have been progressively interchanged have revealed specificity towards AgrD1 and AgrD2^[351] (there is no cross-processing by WT AgrBs *in vivo*)^[67] is driven by different amino acids for each allelic variant. Specifically, the amino acids at positions 42 – 65 drive specificity for AgrD1 whilst amino acids at positions 67 – 75 and 126 – 141 drive specificity for AgrD2.^[351] Remarkably, inspection of the model reveals these two sets of amino acids to be located on spatially-distal areas of the atomic model, and are organised mutually orthogonal to one another (fig. 5.28C and D). The AgrD1 specificity residues are located entirely on TM2 whilst the AgrD2 specificity residues located exclusively at the cytoplasmic “perimeter” of the AgrB molecule, largely away from the TM helices.

It is tempting to conclude from this that the model is either incorrect or at least not representative of both AgrB1 and AgrB2, however, hydrophobicity plots (fig. 5.28A and B) generated from the primary amino acid sequences of AgrB1 and AgrB2 are almost identical to one another, and presumably hydrophobicity is the dominant effector of transmembrane topology. Cross referencing these specificity residues to the electrostatic surface and hydrophobicity plots reveals those residues that drive specificity towards AgrD1 to be largely hydrophobic in character, as is necessary for residues which form a TM helix, whilst those that drive specificity towards AgrD2 are generally positively charged. The hydrophobicity plots also demonstrate that the catalytic residue Cys84 is not located at a hydrophobicity minimum.

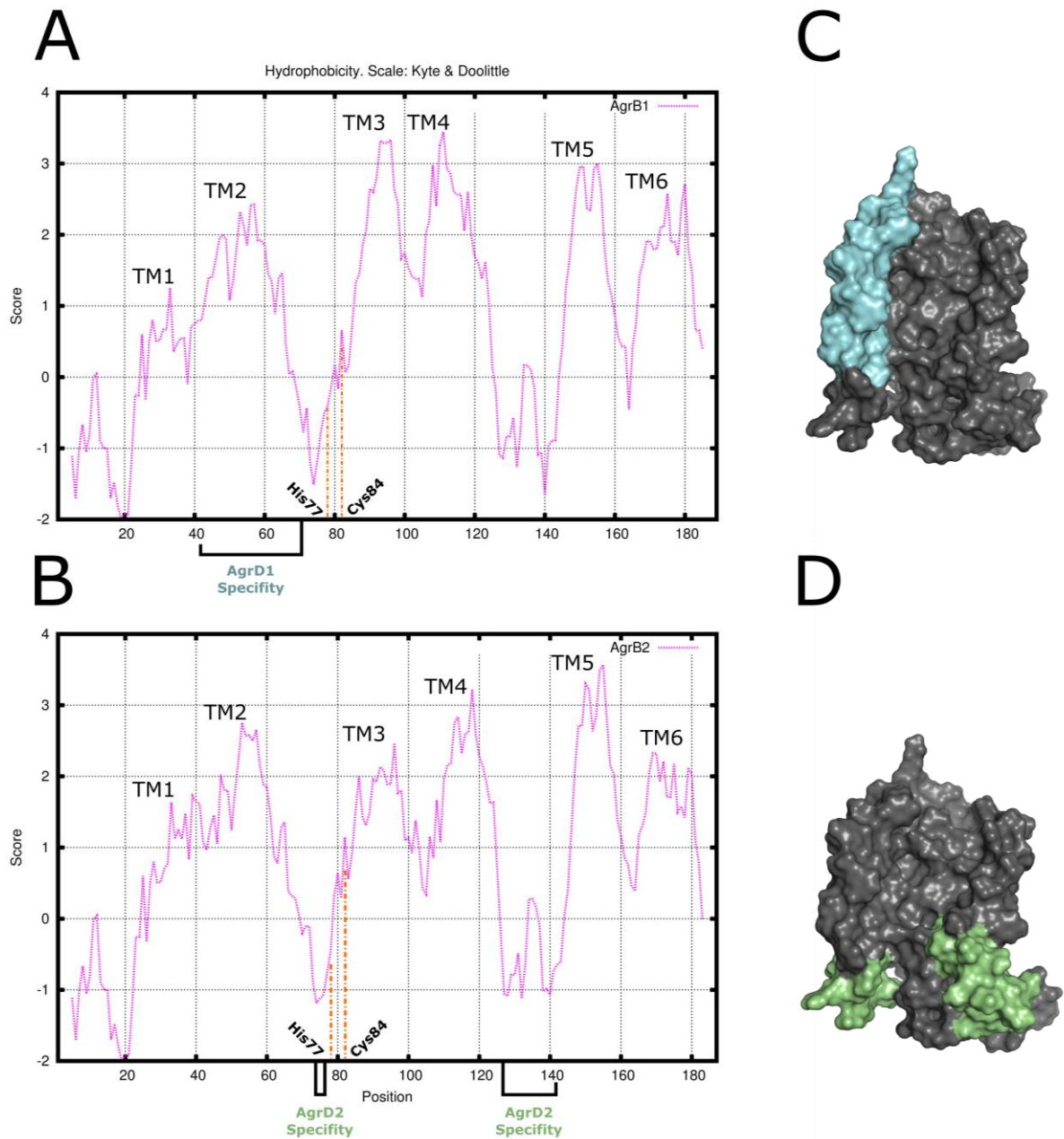


Figure 5.28. Annotated hydrophobicity plots for AgrB1 and AgrB2 and areas of the AgrB surface which drive AgrD1 and AgrD2 specificity. (A) and (B). Hydrophobicity plots for AgrB1 and AgrB2 respectively. The hydrophobicity maxima are suggestive of 6 TM helices. Catalytic residues are indicated by orange dashed lines and AgrD specificity residues are indicated. (C) and (D). Surface representations of AgrD1 (cyan) and AgrD2 (green) specificity areas on the surface of the model.



5.11.5.4: Features in Relation to Possible Arrangement of AgrBs in the AgrB Dimer

Some region or regions on the surface of AgrB must comprise the dimerisation interface. Presumably such a region would exhibit a high degree of conservation since dimer-abolishing random mutations would prevent AIP biosynthesis and place the cells at a competitive disadvantage. The same arguments also apply to those areas of the AgrB surface, if any, that are involved in transport of the cyclised AIP intermediate out of the cell.

It is possible to envisage an AgrB dimer in which the conserved central cavities on each protomer are placed together to create a pore that lies normal to the cell membrane and through which the cyclised AIP intermediate could pass (fig. 5.29A and B). There are no known examples of pore-forming dimers however, and such an arrangement would be extraordinary given that the buried surface area of such a dimer interface is likely to be comparatively small.

Mapping of the sequence alignment onto the surface of the AgrB model using the ProtSkin tool^[345] reveals the central AgrB cavity to be highly conserved. It is possible therefore that this region may itself form part of the AgrB dimerisation interface. From a purely steric perspective such a cavity running continuously along one plane of the AgrB molecule presents an opportune area into which structural elements on the second AgrB protomer could bind. Such an arrangement would have a larger buried surface area than a completely flat interface and is further made plausible by the orientational constraints imparted onto the AgrB molecules by the cell membrane. To explore this hypothesis further, two copies of the AgrB model were docked together using the ClusPro server (vers. 2.0).^[265,266,267]

Of the top ten docking results, five have the AgrB molecules arranged antiparallel and may be dismissed on those grounds alone. The “best” remaining result (which incidentally was the “best” of all ten) predicts the AgrB molecules to be arranged such that TM2 on each protomer nestles into the aforementioned concave face, and this arrangement is mutual such that each protomer “presents” and “receives” the same secondary structure elements (fig. 5.29C and D). To gain further insight into this predicted interface, a schematic diagram of the residues that form AgrB···AgrB contacts was prepared



using the DimPlot tool^[262] (included in LigPlus vers. 2.2) [EMBL]. The result is shown in figure 5.30. Residue labels were coloured according to conservation, using the same colour scheme as the ProtSkin results. The results indicate that over half of the residues that form AgrB···AgrB contacts, specific or otherwise, are completely or highly conserved (* or : according to the Clustal Omega sequence alignment). Furthermore the residues involved are exclusively hydrophobic or basic in character with a bias towards tyrosine and phenylalanine.

To assess how well the surface distribution of completely and highly-conserved residues predicted to form AgrB···AgrB contacts agree with the surface conservation indicated by the ProtSkin result, and to what *extent* the completely and highly conserved-residues located on the TM1-2-4-5 (concave) face of AgrB as indicated by ProtSkin also form contacts at the predicted dimer interface, a simple comparison was made between the ProtSkin result and the location of said residues on the surface of the AgrB model. The result is shown in figure 5.31. The completely and highly conserved residues predicted to form AgrB···AgrB contacts were again coloured according to conservation and displayed as both sticks and a solvent-accessible surface. Comparison of the resulting images with the ProtSkin result show good agreement between residues which are conserved across the *S. aureus* strains and staphylococcal species studied and those involved in AgrB···AgrB contacts at the predicted dimer interface.

Another feature of the dimer predicted by the ClusPro server is that two AgrB protomers are tilted inwards slightly at the extracellular side of the molecules and overlap one-another. This has the combined effect of bringing the catalytic residues His77 and Cys84, and AgrD1 and AgrD2 specificity residues all in close proximity to one another and lowering them slightly towards the cell cytoplasm (fig. 5.29A and B) relative to the instance where both AgrB protomers are aligned perfectly normal to the cell membrane (fig. 5.29C and D).

In the study in which the oligomeric AgrB immunoprecipitate was reported^[348] the following non-proteolytic (to prevent AgrD processing) mutants were prepared to see if they also interfered with AgrB oligomerisation: N39I, N39Y, M46K, R70G, P125H, T128I and K131E. All mutants failed to



disrupt oligomerisation in the immunoprecipitate. Although it may be these single site mutants are insufficient to disrupt the AgrB1 dimer, it is pertinent that none of these residues form contacts at the predicted dimerisation interface in the dimer predicted by ClusPro. According to the model presented in this work, all of these residues excepting Met46 are located in loop regions between the TM helices at the “top” or “bottom” of the AgrB molecule (fig. 5.28A and B). Met46 is located at the carboxyl end of TM2 (which is the extreme extracellular end in the model) and is amongst the residues required for AgrD1 specificity.

Taken together these bioinformatical observations suggest a plausible interface for the AgrB dimer is formed by conserved, hydrophobic and basic residues located primarily at the carboxyl-end of TM1, throughout TM2 and on the intervening loop. The spatial arrangement of catalytic and substrate-specifying residues on the AgrB dimer generated *in silico* that is mediated by this interface, would be conducive to AgrD cleavage and cyclisation in the cell cytoplasm, close to the inner membrane leaflet. Whether or not such dimer subunits associate to form a toroidal AgrB oligomer to enable transportation of the cyclised AIP intermediate out of the cell remains, open to speculation.

Being derived entirely from an *in silico* interrogation however, these observations only constitute a hypothesis. They would ultimately need testing in a laboratory to be proven correct or incorrect. The speculative dimer interface may be probed experimentally via alanine scanning^[352] of the conserved residues implicated to see if a non-oligomeric phenotype could be generated. The availability of Cys84 to the cell cytoplasm could be probed by derivatising it with a suitable biochemical reagent such as iodoacetamide and then probing for fluorescence from a thiol-specific dye, such as, 7-diethylamino-3-(4'-maleimidylphenyl)-4-methylcoumarin^[353] (CPM) relative to a control species.

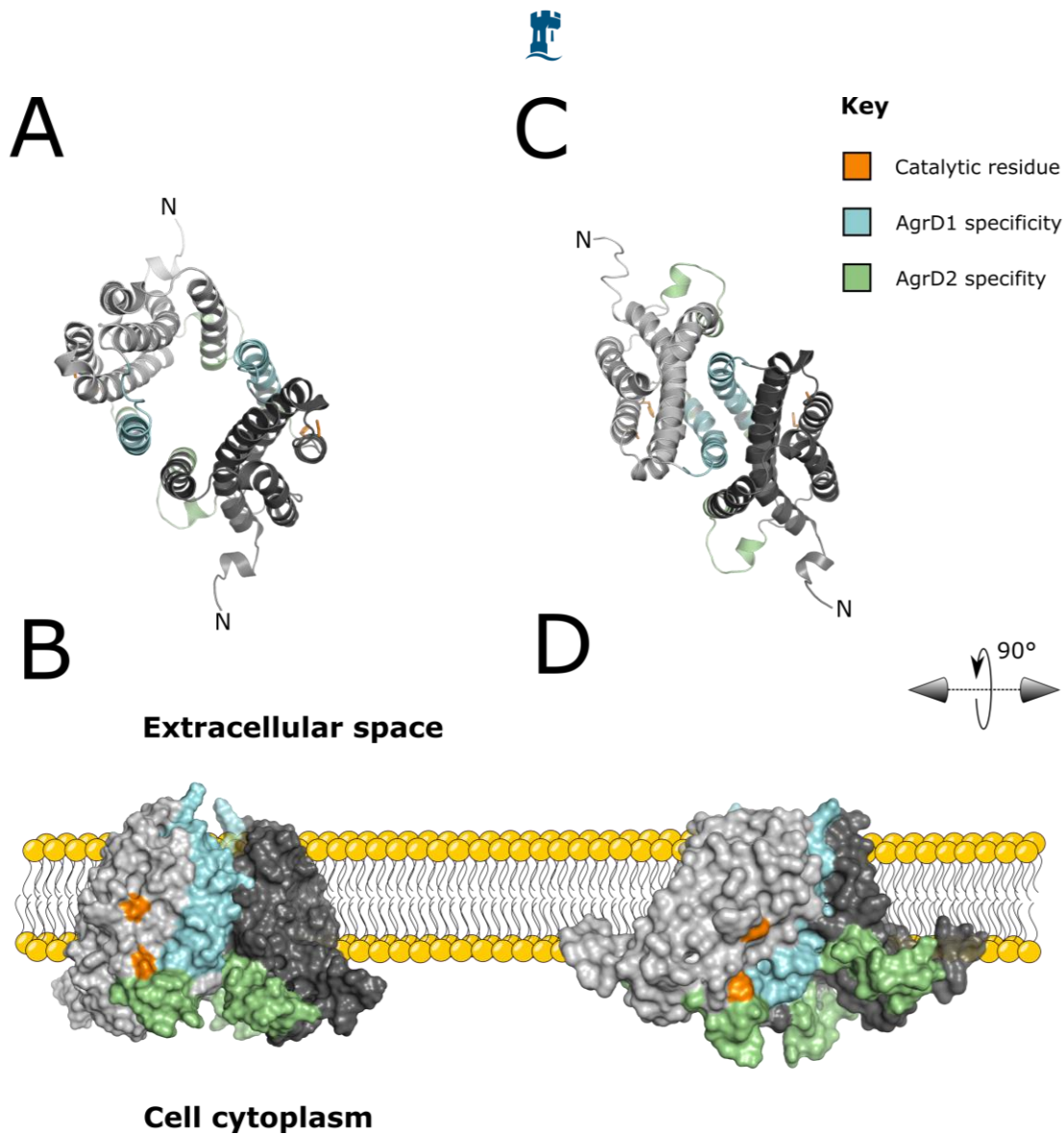


Figure 5.29. Two theoretical AgrB dimers proposed from the AgrB model, displayed in two different orientations. (A) and (B). An AgrB dimer prepared by manual alignment of the AgrB molecules in PyMOL™ such that the concave face of each AgrB molecule forms a pore normal to the cell membrane. (C) and (D). The “best” docking result output by ClusPro that exhibits interlocking of the concave faces. AgrD1 specificity residues (cyan), AgrD2 specificity residues (green) and the catalytic residues His77 and Cys84 (orange) are highlighted to demonstrate their proximity in each dimer in each case. The area enclosed by the substrate specificity and catalytic residues is smaller and closer to the cell cytoplasm in the dimer predicted by ClusPro, due to the inward tilting of the AgrB protomers.

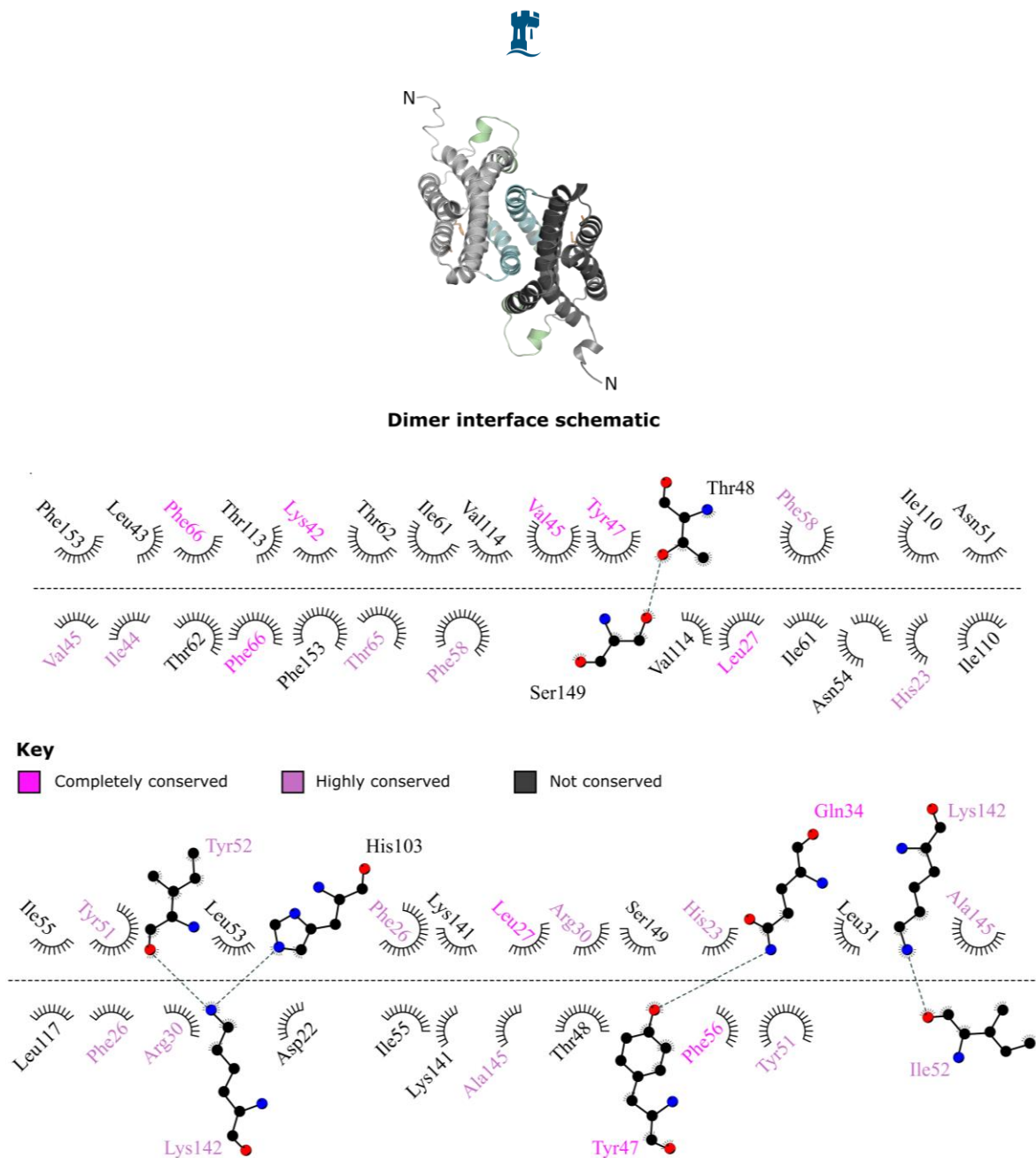


Figure 5.30. A schematic plot of the dimer interface for the dimer predicted by ClusPro. The interface is shown as a dashed horizontal line with each protomer in the dimer lying on either side of it. H-bonds are displayed as dashed green lines. Hydrophobic interactions are shown as black eyelashes. Residues are labelled according to sequence conservation indicated by the Clustal Omega alignment. Labels are coloured bright pink for completely conserved (*) residues and purple for highly conserved (:) residues.

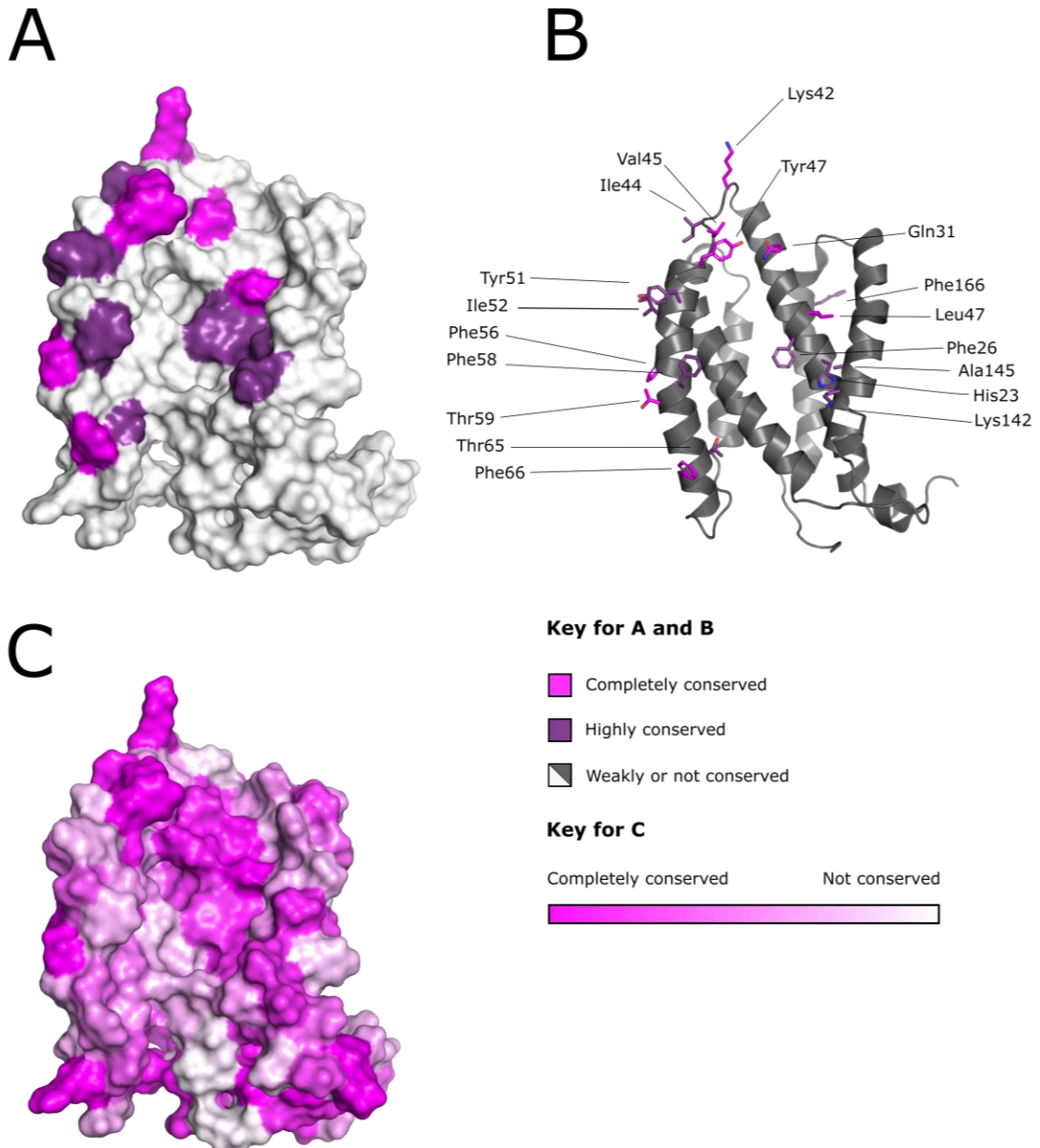


Figure 5.31. Comparison of sequence conservation of residues that are predicted to form AgrB...AgrB contacts at the dimer interface in the dimer predicted by ClusPro, to the ProtSkin result for all *S. aureus* strains and all staphylococcal species discussed. (A) and (B). Completely conserved (*) residues and highly conserved (:) residues that are predicted to form AgrB...AgrB contacts are shown as a solvent-exposed surface (A) and sticks (B) and are coloured bright pink and purple according to degree of conservation. (C). The ProtSkin surface conservation result viewed at the same orientation as (A) and (B). For the conservation representations, darker pink represents more conserved regions and white represents unconserved regions.



CLUSTAL O(1.2.4) multiple sequence alignment

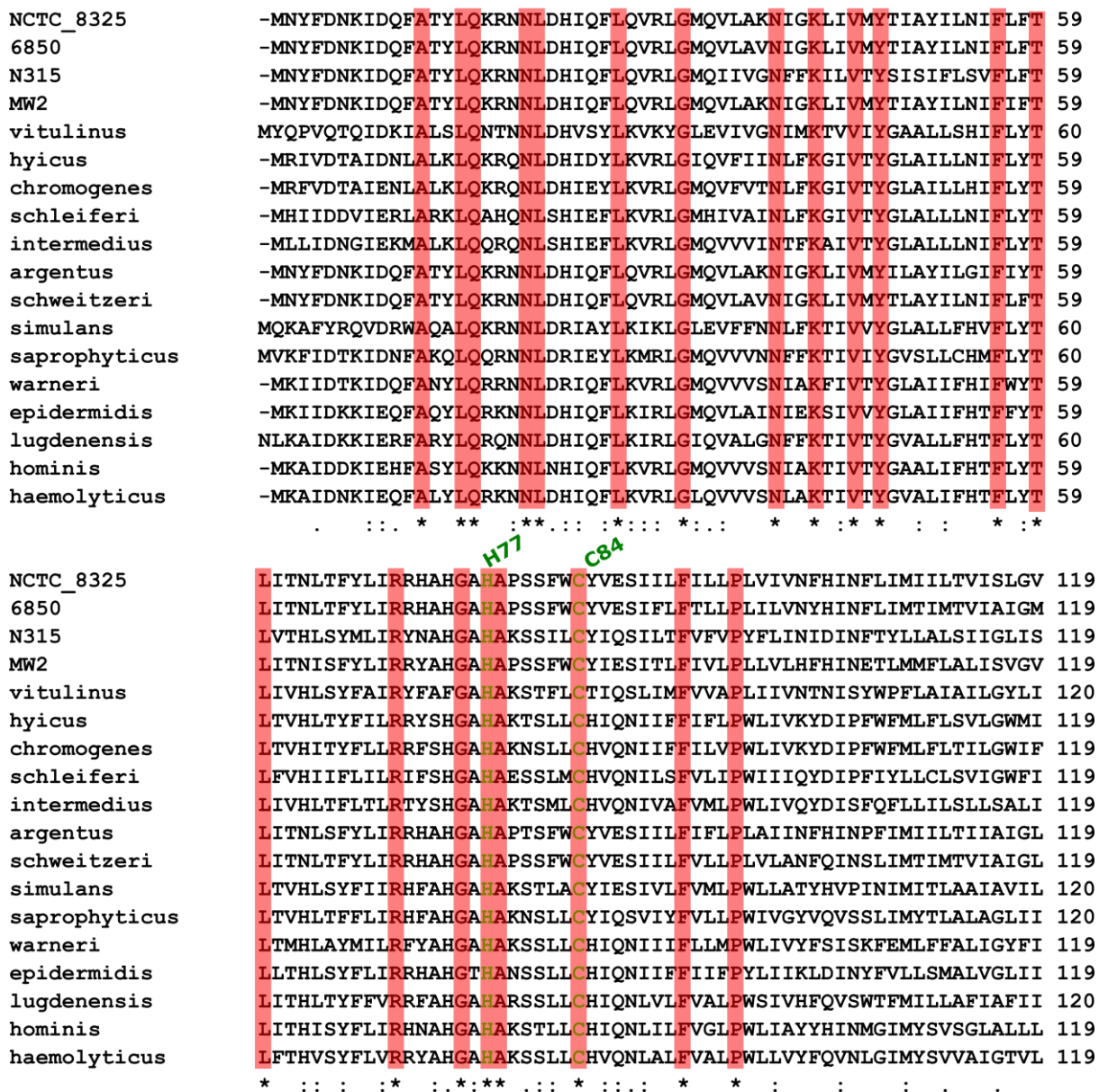


Figure 5.32. Clustal Omega sequence alignment of *Staphylococcus aureus* strains: NCTC_8325 (group I), 6850 (group IV), N315 (group III) and MW3 (group II), and staphylococcal species: *vitulinus*, strain 62, horse isolate; *hyicus*, pig nose isolate; *chromogenes*, pig nose isolate; *schleiferi*, strain 30743, dog isolate; *intermedius*, strain NCTC_11048; *argentus*, strain 60; *schweitzeri*, strain 59; *simulans*, pig nose isolate; *saprophyticus*, human nose isolate; *warneri*, strain 5811483, human nose isolate; *epidermidis*, strain ATCC_12228; *lugdenensis*, strain A1; *hominis*, strain 9525 and *haemolyticus*, strain 1312, dog isolate. Information regarding these strains is published here.^[354] Continues overleaf.



NCTC_8325	ISVYA PAAT TKKKPI PVRLIKRKKYYAIIVSLTLFIITLIIKEPFAQFIQLGIIIEAITLL	179
6850	IIRYA PAAT TKKKPI PVRLIKRKRNYAIIVSLIFFIITLIIKEPFAQFIQLGIIIEAITLL	179
N315	VVIYA PAAT TKKQPI PIKLVKRKKYLSIIMYLLVLILSLIIHPFYAQFMLLGIIIVESITLL	179
MW2	VIKYA PAAT TKKKPI PARLVKQKRYFSIIISTILFIITLTFVKEPYTQFIQLGIIQAITLL	179
vitulinus	IIKYA PMQ TKKHPI ILGKWRKGLKLRSAIATILIIISLFIKEPYQQLVCLGLVFEAVSLL	180
hyicus	VFKYA PAAT TRKQPI IKYSRKRALKIKSLSLMTLYFILFLVPEPFNHLIAYGAFLOSTTLL	179
chromogenes	VFYYA PAAT TRKQPI IKHSRKRALKIKSLSLMTLYFILFLVPEPFNHLIAYGAFLOSTTLL	179
schleiferi	VVIFA PAAT QKKPI IKQSKIKGLKIKSIITATILLVLSLFFPSPFDKLITYAVTLQSTTLL	179
intermedius	VIKYA PAAT TKKRPI APKKVKGLKIKSIIIVFVLLMTIACIVPPPNRFVVGVLQSFSTLL	179
argentus	IIKYA PAAT TKKKPI PVRLIKRKKYYATIVSLVFFIITLIIKEPFAQFIQLGIIIEAITLL	179
schweitzeri	IIRYA PAAT TKKKPI PVRLIKRKRNYAIIVSIIFFIITLIIKEPFAQFIQLGIIIEAITLL	179
simulans	LCIYA PAAT TKKQPI PTRLIKKKKITAIILATLMFAISFLVPNAYQOMILLGITLQGGSQL	180
saprophyticus	ISIYA PSAT TKKQPI PERLRRGKKKAIKICLTLIFLLISLFLNEPYQQLMLLGIVIIISILQF	180
warneri	VIAFA PAAT TKKQPI PKRLVKKKRILSIVVYMLLVLSFIIEKPYSQLILFGIIVESITLL	179
epidermidis	TILYA PAAT TKKQPI PRRLVKKRKKILSIFLYCTIVVISLVTKPEVKNLILFGVILESLTLL	179
lugdenensis	IICYA PAAT TKKQPI LPHLRKKKKRMAILSICFLVLMFLVSEPYMQLIALGMCLEAITLL	180
hominis	IIIIYA PAAT TKKQPI PERLKPRKKIKSIIIVVLLIIISIIIVPEPKQLILLGMILESTTLL	179
haemolyticus	IIYYA PSAT TKKQPI PSHLKMKKKLLSIIITMVLLIISFLAPEPFKQLILLGITLESITLL	179
	:** * *:** : : . : . : . . . :	
NCTC_8325	PIFFIKEDLK -----	189
6850	PIFFVRRT -----	187
N315	PIFFPKED -----	187
MW2	PIYYSKED -----	187
vitulinus	PIFYKEEETIS ----	191
hyicus	PIFFSKEE -----	187
chromogenes	PVFFPKEE -----	187
schleiferi	PIFNNKEEI -----	188
intermedius	PIFSIKEEV -----	188
argentus	PIFFIRRN -----	187
schweitzeri	PIFFVRRT -----	187
simulans	PIFFPKNRKDV L---	192
saprophyticus	PIFFPKEDY -----	189
warneri	PIFFPKED -----	187
epidermidis	PIFFPKEDINHGKH F	194
lugdenensis	PIFFSKEET -----	189
hominis	PFFFPKEDY -----	188
haemolyticus	PIFFPREDY -----	188
	*.: ..	

Fig. 4.32. Clustal Omega sequence alignment continued from the previous page. The alignment is annotated as follows: (*), completely conserved, (:), highly conserved, (.), weakly conserved. Completely-conserved residues have also been highlighted in red boxes for additional clarity. The catalytic residues His77 and Cys84 are highlighted in green and are labelled with residue numbers as they occur in the *S. aureus* strain NCTC_8325 sequence.



Chapter 6: Conclusions and Future Research

6.1: AgrA – Conclusions

A structural understanding of the *S. aureus* QS circuit has not been forthcoming to the research community. This is largely due to the difficulty in producing suitable soluble samples of the key proteins involved. The full-length AgrA_C199S structure solution, even at low resolution, represents a significant advance in this area, with only the ATP-binding subdomain of AgrC, and the C-terminal domain of AgrA, having been previously solved.

A hydrophobic helix within the N-terminal domain of AgrA appears to form a major part of the physiological AgrA dimerisation interface. Evidence for this comes from the similarity of a crystallographic AgrA_C199S dimer, to the structures of the physiological dimers of a number of AgrA homologues, and from agreement between said dimer with the SAXS molecular envelope of the phosphorylated AgrA_C199S/P_{3full} promoter DNA complex. Furthermore, this envelope indicates that AgrA_C199S places a ~ 55 ° bend into the DNA, occurring approximately perpendicular to the mean plane of the AgrA_C199S molecules. This structural data is the first for any full-length LytTR-type DNA-binding protein in complex with DNA.

Residues Cys55 and Cys123 form a cysteine pair, mutation of either of which to serine, retards cell growth during *agr* dormancy. *In silico* docking results for a series of in-house AgrA inhibitors suggests that they bind into a functionally-important groove located between Arg218 and Tyr229. A simple but atypical approach to growing large crystals has been outlined and used efficaciously for AgrA_C199S, and may be of general use to macromolecular crystallographers.

6.2: AgrB – Conclusions

Results on AgrB are abbreviated compared to those for AgrA owing to the challenges involved in studying recombinant membrane protein samples. The goal of solving the experimental structure of AgrB was not met, however, some foundational milestones have been achieved. Namely, a robust purification that yields pure, folded, and monodisperse DDM_AgrB2 has been outlined. The sample



exhibits a high thermostability of 63 °C and the purification is flexible. This allows for facile swapping of the sample into several detergents, upon which varied oligomerisation behaviour was observed. Preliminary negative stain micrographs suggest that LMNG is an appropriate choice of detergent, however, efforts need to be made to reduce the amount of sample settling onto the grids.

6.3: AgrA – Future Research

Research focussing on the AgrA_C199S_K101A double mutant would return informative results in a short time frame. Submitting the SDS-PAGE band corresponding to the sample which elutes in the void volume during SEC for MS may directly implicate Lys101 in being involved in mediating AgrA dimerisation, if the returned sequences matched that of AgrA. The K101A mutant could then be subsequently placed into *in vivo* reporter and growth assays to see if and how the system is perturbed.

Another experiment that ought to produce meaningful data would be to prepare the [BeF₃]⁻-activated AgrA_C199S dimer bound to the P2 promoter, and to other promoters also, to obtain comparative SAXS data for them all. Construction of the molecular envelopes might show differences in how AgrA binds to and distorts each of them, provided that these differences are fairly drastic (e.g. due the length of spacer DNA between the two AgrA binding sites). If all the envelopes are approximately identical, the differences in their respective affinities for AgrA must lay within the finer structural detail only available through high-resolution structural studies.

A crystal structure of the AgrA_C199S dimer bound to one of its full promoters would be of high scientific value. It would serve the same purpose as the SAXS experiments described above but be significantly richer in data. It will be observed that very little has been discussed of the DNA-binding domain of AgrA despite all of the known chemical inhibitors targeting this region. This is because the full-length structure presented in this work adds nothing new to this field of study. A crystal structure of an AgrA dimer/promoter complex should reveal in full the subtle asymmetry between the two AgrA protomers bound to the promoter DNA.^[81,85] This asymmetry is crucial to the mode of action of the current AgrA inhibitors and a high-resolution structure may facilitate further improvements in their potency.



The P3 promoter is probably the more scientifically interesting to pursue, as it is the one which upregulates virulence factors associated with infection. Practically speaking, however, it probably makes little difference which promoter is studied. The sequence corresponding to the nominated promoter may be incrementally truncated until binding of AgrA is abolished. This abolition should be readily detectable via ITC since binding of DNA is usually considerably exothermic, and EMSAs since AgrA is a high-affinity binder of DNA. Once the shortest sequence that exhibits binding to AgrA has been determined, crystallisation experiments should then be conducted. It may be wise to start using commercial screens of similar composition to the PACT Premier™ screen which afforded the AgrA_C199S/P2_{fragment} crystals. Similarly, it would be worth screening blunt-end DNA and DNA with complimentary overhangs (already tried) to increase the chances of success, since these parameters can make the difference between obtaining crystals or not. Phasing of any crystallographic data could probably be achieved experimentally using DNA containing 5-bromouracil substitutions.

The *apo* structure of full-length AgrA would also be of considerable interest as it would open up possibilities for AgrA/inhibitor co-crystals for sufficiently potent AgrA inhibitors. A condition which affords crystals of AgrA in complex with DNA is of very little use if the role of the inhibitor is to abolish DNA binding. These experiments could start using the SER mutants discussed in this work, in combination with sparse matrix screening to search for new crystallisation conditions. Owing to the high quality of the crystals obtained of the DNA-binding domain in complex with DNA^[80] and the poor quality of the crystals presented in this work, a significant amount of time may need to be dedicated to crystal optimisation.

Another thread of research that may require a significant investment of time but ultimately provide high-impact results if successful, would be to try to obtain co-crystals of the N-terminal domain of AgrA in complex with the phosphate-transfer subdomain of AgrC. Both can be readily expressed and purified so there is no barrier in that regard. A non-hydrolysable ATP analogue could be used to try to capture the complex at some point during the phosphate transfer process. A good starting point may be to simply mix the aforementioned domains in equimolar amounts, or with a slight excess of the AgrA N-terminal domain, add 10 – 100 molar equivalents of the non-hydrolysable ATP analogue and



proceed immediately to vapour diffusion crystallisation experiments. Whilst waiting for the experiments to proceed, SEC may be used to see if there is any association between the two domains *in vitro*, with and without the ATP analogue. If no association is observed, it may be found that under the dehydrating conditions of the crystallisation drop, association is eventually encouraged in much the same way that constitutively inactive ComE was forced into a phosphorylated conformation.^[90] The primary reason this thread of research is likely to be time consuming is that it may be found that one of the domains preferentially crystallises over the other, leading to false positives which would need indexed diffraction data to estimate their content.

A high-resolution structure of the N-terminal domain on its own would accompany the low-resolution full-length structure presented in this work. It may corroborate or disprove areas of the full-length solution accordingly, and it would be interesting to see if the same $\alpha 4 \cdots \alpha 4$ interface is present in the crystal lattice also. There are prospective crystals presented in this work that may be a good starting point for this line of experimentation.

Although AgrA_C199S was eventually coaxed to dimerise on DNA corresponding to the P3 promoter, a strategy for inducing AgrA dimerisation in the absence of DNA would enable a facile study into which mutants, informed by degree of conservation and location in the AgrA_C199S/P2_{fragment} crystal structure, could abolish AgrA dimer formation. It has been seen that considerable effort has been put into realising this idea already. Although it would be experimentally simpler to have a system which required no DNA to be present for AgrA dimer formation, it has not yet been confirmed whether the AgrA_K101A_C199S double mutant can dimerise to produce a complex identical to that used for the SAXS measurements presented earlier. That is to say, if the same sample preparation was conducted, but used the aforementioned double mutant, and the analytical SEC returned two peaks, corresponding to an AgrA_C199S_K101 monomer bound to P3 DNA, and unbound AgrA_C199S_K101, then the work presented here could be expanded into an assay for AgrA dimerisation *in vitro*. The complication here is that in order to test the system, a mutant that is known to abolish AgrA dimerisation is required. Given what has been presented already the experimenter suggests that the AgrA_C199S_K101A double mutant is be a good place to start.

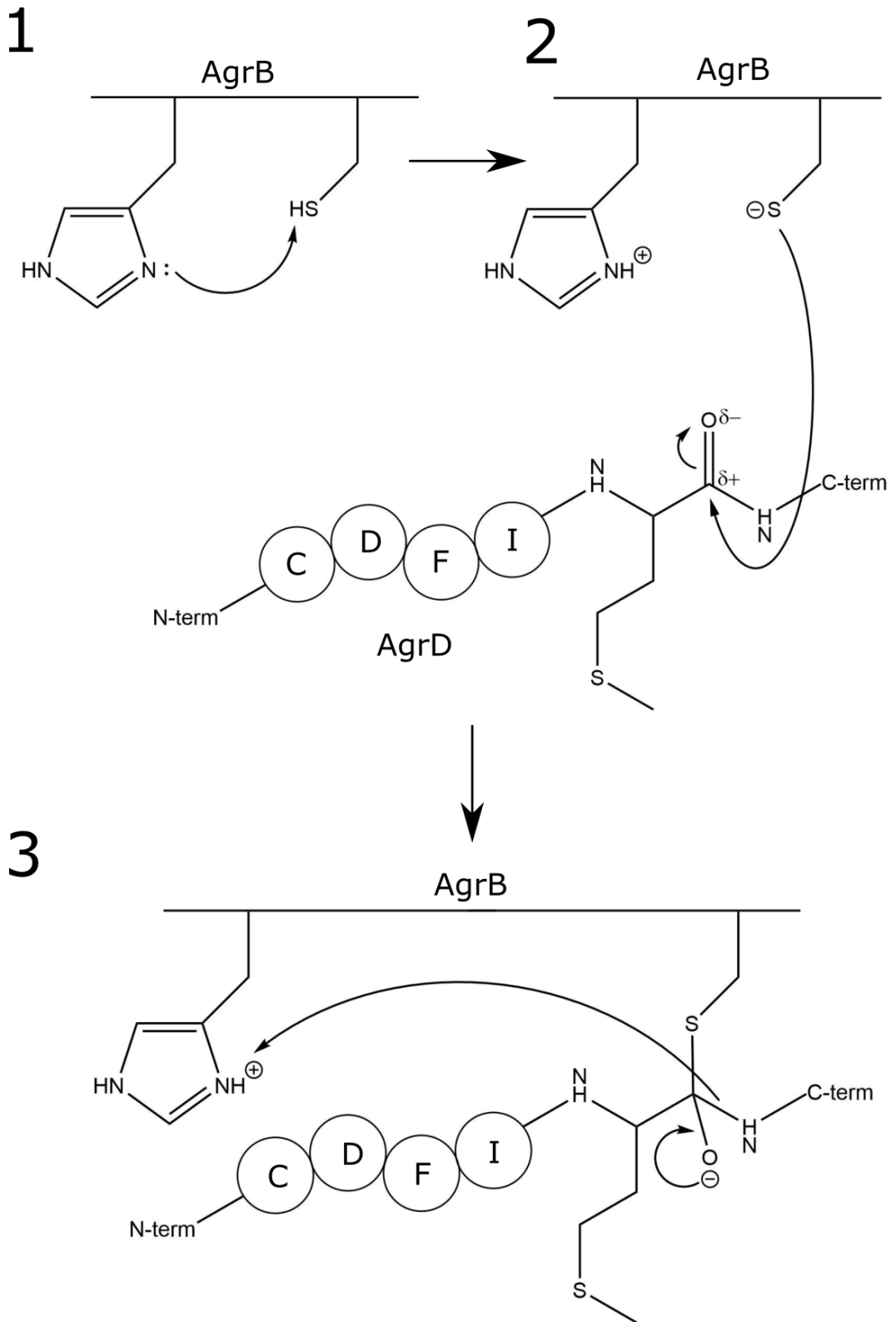


6.4: AgrB – Future Research

Radical modifications to the expression construct are needed to yield successful structural experiments. These could include: A) a loop insertion tag such as lysozyme to facilitate crystallisation, B) a nanobody/antibody crystallisation chaperone to provide some water-soluble, rigid domain to facilitate crystallisation in a similar fashion to a loop insertion tag or C) an oligomerisation tag (which could also be a loop insertion) that imparts some readily-identifiable symmetry to make electron micrographs easier to interpret.

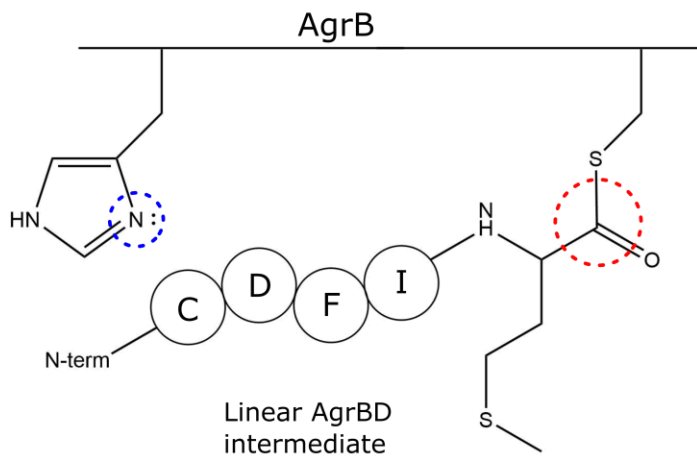
Regarding option B) purified DDM_AgrB1 and DDM_AgrB2 samples have been provided to VIB nanobody core by Dr. Philip Bardelang, Dr. Mohamad Saleem and the experimenter for nanobody generation and, at the time of writing, a library of genes encoding AgrB-binding nanobodies is available. These are being screened for AgrB-binding activity *in vitro* by Dr. Philip Bardelang in order to deduce the nanobody that is most suited to crystallisation experiments. The purification and crystallisation experiments presented in this work could be easily replicated for the AgrB···nanobody complex once the optimal nanobody for structural studies has been deduced.

One strategy which might be conducive to producing structure-grade AgrB crystals, as well as being of extremely high scientific interest, would be to crystallise the AgrBD complex. No atomic-level curly arrow mechanism for AgrD cleavage and cyclisation has been put forwards in the literature. One is proposed overleaf figure 6.1.



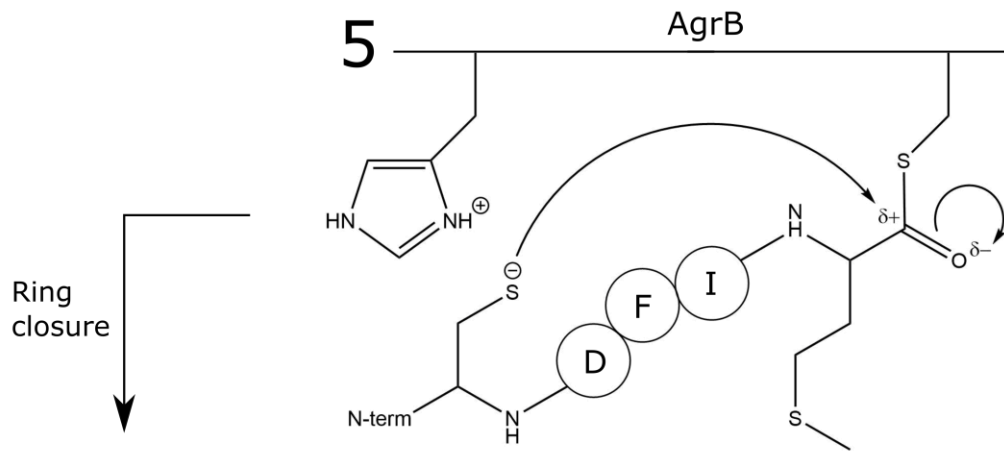


4

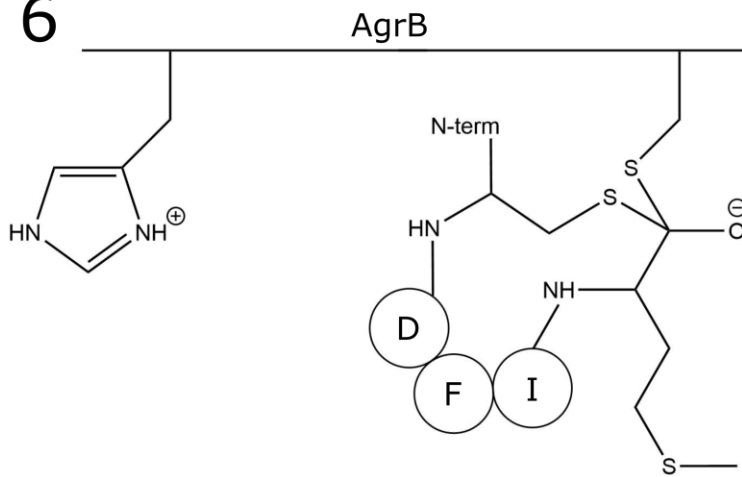


What activates AgrD Cys28?

5



6



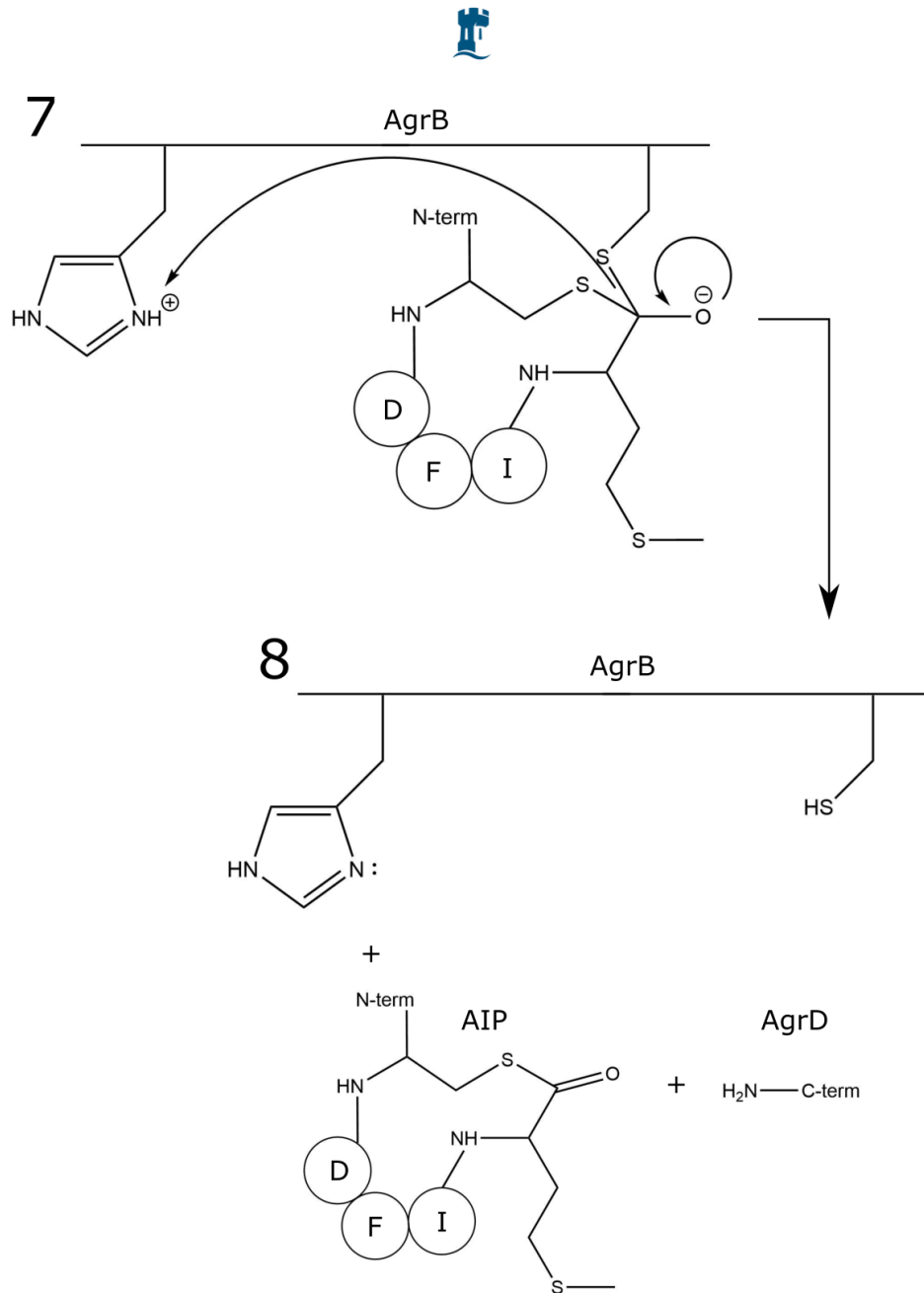


Figure 6.1. Potential curly arrow mechanism for the cleavage and cyclisation of AgrD by AgrB. (1). Activation of AgrB Cys84. (2). Nucleophilic attack of the AgrB Cys84 thiolate on AgrD backbone carbonyl at Met32. (3). Reformation of carbonyl group and C-terminal cleavage of AgrD. (4). The linear AgrBD thioester intermediate. (5). Nucleophilic attack of AgrD Cys28 on thioester carbon leading to ring closure and resulting in (6). (7). Reformation of carbonyl group and cleavage of cyclic AIP intermediate from AgrB to give the products (8): a cyclic AIP intermediate with the N-terminus attached and the AgrD C-terminal cleavage product. **The dashed blue and red circles represent steps at which it would be theoretically possible to biochemically intervene in the reaction and are discussed in the ensuing text.**



There are two steps that are of biochemical interest and worthy of consideration. The first concerns the activation of AgrD Cys28 so as the internal cyclisation/thioester exchange can occur. The catalytic AgrB His77 will have regained its capacity to act as a Lewis base once the C-terminal end of AgrD has been cleaved, so it could in theory deprotonate AgrD Cys28 provided that the two are spatially-proximal. This is how it has been drawn in step 4 of figure 6.1, with the relevant N atom highlighted in a dashed blue circle. Histidine residues are only present on the cleaved C-terminal end of AgrDs 1, 2 and 4, and it has been shown that rapid removal of this cleaved unit is required to drive the equilibrium towards AgrD turnover.^[44] Consequently, it does not seem feasible that AgrD self-activates prior to cyclisation/thioester exchange. Another possibility is that the second AgrB protomer of the dimer provides a second His77 to deprotonate AgrD Cys28.

The second concerns the possibility of interfering with the cyclisation/thioester exchange to prevent it from occurring, thus leaving the linear AgrBD intermediate available for purification. Preparing the AgrD_C28A mutant would in theory prevent formation of the thiolactone ring, and a linear AgrBD intermediate has been observed in cells in at least one (*in vivo*) study to-date.^[42] The chemical bond between the AgrB and AgrD units in the resulting linear intermediate would be a thioester, as highlighted in figure 5.1 step 4 in a dashed red circle. This may be susceptible to base hydrolysis upon purification, which may be mitigated by purifying the intermediate under mildly acidic conditions, provided sample stability is not impaired, however. A good place to start this line of study would be to place the coding regions of AgrB and AgrD_C28A under the control of a single promoter. AgrB could retain the C-terminal (His)₆ tag whilst AgrD_C28A could be left untagged or have a Strep-tag® placed at the N-terminus to aid removal of any excess AgrB. The two species could also be placed under separate promoters on (say) pETDuet™ if more control was needed over the expression of either.

It would also be interesting to prepare the AgrD_C28S mutant to see if AgrB is cable of installing a lactone ring into AgrD. The result would be interesting in itself and if the answer was found to be “yes” then the lactone-derivative of the AgrBD complex might be more stable in aqueous solutions than the thiolactone version and consequently a better option for structural studies.



Anticipated problems with this approach will probably relate to the toxicity associated with the PSM-like nature of AgrD, which could harm the host cells when overexpressed. Similarly, the accumulation of AgrD C-terminal cleavage product could harm the cells and retard the C-terminal AgrD cleavage event. Ultimately the strategy would need to be tested to see where the difficulties lay.

Experiments that probably constitute a waste of time are more *in surfo* crystallisation experiments on AgrB2. The lack of success in this area and the potential to make use of the technologies highlighted above suggest they should be avoided.

6.5.1: Additional Work

Over the PhD process aims change and some threads of research prove to be unrewarding or, where successful, do not neatly tie in with other data. This was so for some work related to *agr* and some work performed to help peers within the structural biology research group. Whilst the latter is not related to main body of work discussed in this thesis it, does rely on some relevant practical skills, acquisition of which is an important part PhD process. They are therefore worth a brief mention.

6.5.2: Work Related to *agr*

To validate the *in silico* docking results obtained for the IQS analogues, AgrA_Y229A and Y229M mutants were prepared in full-length AgrA_C199S and AgrA₁₃₅₋₂₃₈_C199S constructs. The planned experiment was to test these mutants in EMSAs using a P2 or P3 promoter probe in the presence of key IQS compounds. The experiment has not yet been performed due to the Covid-19 outbreak.

To try to obtain experimental kinetic binding data for the IQS analogues, preliminary ITC experiments were performed using full-length AgrA_C199S. The low affinity of the analogues for AgrA mandated high concentrations (~ 300 μ M) of AgrA however, and IQS concentrations approximately an order of magnitude higher than for AgrA. Given that the ITC sample cell is ~ 250 μ L, and the purification yields of full-length AgrA are modest, only a few experiments could be performed per purification. Furthermore, the IQS analogues were not soluble at elevated concentrations in buffers containing < 10 % v/v DMSO. Although buffer-mismatch between the IQS



and AgrA samples was avoided, the elevated concentrations of AgrA rendered it susceptible to the unfolding effects of DMSO at the temperatures at which the experiments were conducted (20 and 37 °C) with large amounts of AgrA precipitating upon titration of the IQS analogues. The scheme showing how IQS and AgrA samples were prepared so as to avoid buffer-mismatch is presented in appendix VIII and the thermograms for the experiments are given in appendix IX.

The AgrD1 gene from *S. aureus* was cloned via restriction enzyme methods into the pMALX(e) vector^[355] with a view to performing crystallisation experiments and AgrB activity assays. Although the correct sequence was obtained (data not shown) the construct could not be induced to express when transformed into *E. coli* cells.

6.5.3: Work Unrelated to *agr*

In gaining experience during the PhD process help was given to newer members of the research group or to those whose core experience lies in other topics and there are two significant examples of this. Firstly the experimenter was involved in helping crystallise and collecting the crystallographic data for *Leishmania mexicana* cysteine protease B in complex with a covalent azadipeptide nitrile inhibitor (PDB: 6p4e). Some help was also lent in improving the quality of the resulting manuscript. This work was eventually published in *Bioorganic and Medicinal Chemistry*.^[356]

The second example again concerns a crystal structure, this time of a bacterial pathogen DNase in complex with an immunity protein. Although the work on this was less involved, it did require locating the selenium atom K edge at the synchrotron station and collecting high redundancy, low dose data at the appropriate wavelength so as a set of *ab initio* phase estimates could be derived from the anomalous scattering data. This work is yet to be published.



Chapter 7: Methods

7.1: General Methods

7.2: SDS-PAGE Gel Preparation and Running Parameters

Requisite amount of sample was mixed with 3 x concentrate sodium dodecyl sulphate Poly Acrylamide Gel Electrophoresis (SDS-PAGE) loading buffer (tris, 150 mM, pH 6.5, NaCl, 300 mM, SDS, 6 % w/v, bromophenol blue, 0.3 % w/v and glycerol, 30 % v/v), usually 10 μ L sample with 5 μ L of 3 x SDS-PAGE buffer. A gel of appropriate percentage acrylamide (specified as appropriate in the discussion, see table 7.1 for recipes) was immersed in SDS-PAGE running buffer (tris, 25 mM, glycine, 192 mM and SDS, 0.1 % w/v) in Mini PROTEAN™ 3 Cell gel electrophoresis tank [Bio-Rad]. Electrophoresis was performed using a PowerPac™ basic 75 W power pack [Bio-Rad] for 45 mins – 1 hr at a constant voltage of 200 V. Gels were stained for 1 hr using InstantBlue™ protein stain [Sigma Aldrich] and brightened with H₂O. Gels were visualised and photographed on a Gel Doc™ XR+ [Bio-Rad] with Image Lab™ software (vers. 5.2) [Bio-Rad]. Note that the 30 % acrylamide solution was a 37.5:1 ratio of acrylamide:bis-acrylamide.

Component	Amount required for X % resolving gel	Amount required for stacking gel
Acrylamide, 30 %	0.5 x X mL	1.98 mL
Tris, 0.5 M, pH 6.5	0 mL	3.78 mL
Tris, 1.5 M, pH 8.8	3.75 mL	0 mL
SDS, 10 %	150 μ L	150 μ L
H ₂ O	11.03 – (0.5 x X) mL	9 mL
TEMED	7.5 μ L	15 μ L
APS, 10 %	75 μ L	75 μ L
Total volume, mL	15 mL	15 mL

Table 7.1. Summary of SDS-PAGE recipes for common percentage acrylamide gels. APS stands for ammonium persulphate. TEMED stands for N,N,N',N'-tetramethylethylenediamine. Gels were cast using 0.75 mm thickness using glass sandwich plates and casting equipment [Bio-Rad].



7.3: Western Blotting

Two identical SDS-PAGE gels were run for each blotting experiment. One gel was stained with InstantBlue™ [Sigma Aldrich] whilst proteins on the other gel were transferred to a polyvinylidene fluoride (PVDF) 0.45 µm membrane [Thermo Fisher] using a TransBlot™ Turbo Mini [Bio-Rad] for 30 mins at room temperature with 2 x Thick Blot Filter Papers 7.5 x 10 cm [Bio-Rad]. Prior to protein transfer, the PVDF membrane was cut to the appropriate size and activated by soaking in methanol, 50 % v/v, for 5 mins, followed by soaking in transfer buffer (tris, 25 mM, pH 8.3, glycine, 190 mM and methanol, 20 % v/v) for 5 mins. The membrane, after transference of proteins, was subsequently blocked with tris-buffed salt with tween 20 (TBSt) and milk (tris, 50 mM, pH 7.5, NaCl, 150 mM, tween 20, 0.1 % v/v, and powdered skimmed milk, 5 % w/v) overnight at 4 °C. Excess powdered skimmed milk was removed by washing the membrane with TBSt buffer without milk, 10 mL, for 5 mins. Primary antibody, mouse anti-tetrahistidine [Qiagen], was diluted one thousand-fold in TBSt plus powdered skimmed milk, 5 % w/v, 5 mL.

Primary antibody was then allowed to bind to the proteins on the membrane via incubation of the membrane with the antibody-containing buffer for 1 hr at room temperature. The membrane was washed three times with TBSt, 10 mL, and once with TBS (no tween 20), 10 mL. Each wash step lasted ~ 5 mins. The secondary antibody, goat anti-mouse horseradish peroxidase [Thermo Fisher] was diluted one thousand-fold in TBSt plus powdered skimmed milk, 5% w/v, 5 mL. Binding of the secondary antibody to the primary antibody was achieved via incubation of the membrane with this secondary antibody-containing buffer for 1 hr at room temperature. The membrane was again washed three times with TBSt, 10 mL, and once with TBS, 10 mL. Each wash step lasted ~ 5 mins. Chemiluminescence was initiated by wetting the membrane with Enhanced Chemiluminescence (ECL) reagent (tris, 100 mM, pH 8.5, luminol, 1.25 mM, *p*-coumaric acid, 200 µM, and 3 µL of 30 % v/v H₂O₂), 10 mL, and visualised using a C-DiGit western blot scanner scanner [LI-COR].



7.4: Agarose Gel Preparation

Sufficient volume (typically 1 L to suit communal needs) 50 x tris-acetate EDTA (TAE) buffer was prepared as per table 7.2. 50 and diluted fifty-fold with H₂O to a volume of 100 mL and agarose, 1 g, added in a 250 mL conical flask. This mixture was heated in a microwave for short 30 s bursts with intermittent hand stirring until all of the agarose had dissolved to generate a 1 % agarose solution. The solution was then left to cool at room temperature until H₂O was no longer evaporating.

Meanwhile, 5 µL of SYBRTM safe [Invitrogen] dye was pipetted onto the base of a 7 x 7 cm Mini-Sub CellTM GT agarose slab casting unit [Bio-Rad]. The sides of this unit were then sealed as per the equipment instructions. The agarose solution was poured over the dye to a depth of ~ 1 cm, and the dye gently mixed into the agarose solution using a Pasteur pipette. A 7-lane agarose gel comb [Bio-Rad] was placed into the tray of 1 % agarose in TAE, which was allowed to cool at room temperature until it had fully set. This took ~ 30 mins, after which the comb was gently removed and casting equipment disassembled. Gels were used immediately after preparation.

Component	Concentration, <i>M</i>	Concentration, amount L⁻¹
Tris base	2	242 g
Acetic acid	1	60.5 mL
EDTA disodium salt dihydrate	0.05	18.6 g

Table 7.2. Summary of reagents used to prepare 50 x TAE buffer.



7.5: Agarose Gel Electrophoresis

6 x Purple loading dye [NEB] was diluted six-fold into DNA in clear PCR tubes which were then briefly vortexed. A pre-prepared 1 % agarose gel was placed into a Sub-Cell™ agarose gel electrophoresis system [Bio-Rad]. The gel was then completely covered in 1 x TAE so as the electrophoresis bay was entirely filled with TAE buffer. Typically ~ 5 µL of each of the DNA samples/controls were loaded into each of the the wells of the gel. Adjustments were made to accommodate for different anticipated DNA yields/downstream experiments. Electrophoresis was performed at 100 V using a PowerPac™ basic 75 W [Bio-Rad] for 35 mins. Gels were visualised using a BioDoc-It™ transilluminator [UVP Imaging] with the appropriate exposure times.



7.6.1: AgrA

7.6.2: Construct Generation – Wild-Type AgrA, _C199S and D59E_C199S

The coding sequence for AgrA was cloned into the restrictions sites SacI and XbaI of plasmid pCOLD-I [TakaraBio] using standard PCR procedures (work performed by former PhD student Dr. Yanin Jaiyen). The point C199S mutant was introduced via inverse PCR to help mitigate the redox sensitivity of AgrA during purification. The D59E_C199S was also introduced via the same methods. All primers were purchased from Sigma Aldrich as dry, desalted samples. Phusion polymerase [NEB] was used for all amplification steps and all restriction enzymes were purchased as high fidelity variants [NEB]. The initial denaturation was at 98 °C for 5 mins, annealing was performed at 50 and 60 °C for 1 min, with duplicate reactions incubated at each temperature. Extension was performed at 72 °C for 5 mins, followed by denaturation at 97 °C for 30 secs. A total of 34 cycles of denaturation, annealing and extension were performed followed by a final extension at 72 °C for 8 mins. All PCR reactions were treated with DpnI for 1 hr at 37 °C to destroy methylated template DNA followed by transformation of PCR product, 1 – 5 µL, into NovaBlue™ competent cells [Merck] via heat shock (see AgrA expression methods section 7.6.6 for details). PCR reaction compositions are given in table 7.3 and summary of strains, primers and construct amino acid sequences are given in table 7.4.

Component	Volume, µL
5 x Phusion® HF buffer	10
50 ng µL ⁻¹ pCOLD-I_AgrA_C199S template	1
20 µM forwards primer	1
20 µM reverse primer	1
50 % DMSO	2
Milli-Q™ H ₂ O	34.5
Phusion® polymerase	0.5

Table 7.3. Summary of PCR reaction contents used to generate the AgrA_C199S constructs.



Restriction cloning: WT AgrA	
Source organism	<i>S. aureus</i> strain SH1000
DNA source	Genomic DNA
Forward primer	5'-AAAAAGAGCTCATGAAAATTTTCATTTGCGAAGACGATCC-3'
Reverse primer	5'-AAAAATCTAGATTATATTTTTTTAACGTTTCTACCGATGCATAGC-3'
Expression vector	pCOLD-I
Expression host	BL21-CodonPlus
Amino acid sequence of construct	MNHKVHHHHHHIEGRHMELMKIFICEDDPKQRENMVTIKNYIMIEEKPMIEAL ATDNPYEVLEQAKNMNDIGCYFLDIQLSTDINGIKLGSEIRKHDPVGNIIFVTSHS ELTYLTFVYKVAAMDFIFKDDPAELRTRIIDCLETAHTRLQLLSKDNSVETIELKR GSNSVYVQYDDIMFFESSTKSHRLIAHLDNRQIEFYGNLKELSQLDDRFFRCHN SFVVNRHNIESIDSKERIVYFKNKEHCYASVRNVKKI
Mutagenesis: C199S	
Forward primer	5'-CTTCAGATGTCATAATAGCTTTGTCGCAATCG-3'
Reverse primer	5'-CTATTATGACTTCTGAAGAAACGATCATCTAATTG-3'
Amino acid sequence of construct	MNHKVHHHHHHIEGRHMELMKIFICEDDPKQRENMVTIKNYIMIEEKPMIEAL ATDNPYEVLEQAKNMNDIGCYFLDIQLSTDINGIKLGSEIRKHDPVGNIIFVTSHS ELTYLTFVYKVAAMDFIFKDDPAELRTRIIDCLETAHTRLQLLSKDNSVETIELKR GSNSVYVQYDDIMFFESSTKSHRLIAHLDNRQIEFYGNLKELSQLDDRFFR <u>SHNS</u> FVVNRHNIESIDSKERIVYFKNKEHCYASVRNVKKI
Mutagenesis: D59E_C199S	
Forward primer	5'-CTGTACTTTTTAGAAATTCAACTTTCAACTGA-3'
Reverse primer	5'-GTTGAATTTCTAAAAAGTAACAGCCTATGTCATTC-3'
Amino acid sequence of construct	MNHKVHHHHHHIEGRHMELMKIFICEDDPKQRENMVTIKNYIMIEEKPMIEAL ATDNPYEVLEQAKNMNDIGCYFL <u>E</u> IQIQLSTDINGIKLGSEIRKHDPVGNIIFVTSHS ELTYLTFVYKVAAMDFIFKDDPAELRTRIIDCLETAHTRLQLLSKDNSVETIELKR GSNSVYVQYDDIMFFESSTKSHRLIAHLDNRQIEFYGNLKELSQLDDRFFR <u>SHNS</u> FVVNRHNIESIDSKERIVYFKNKEHCYASVRNVKKI

Table 7.4. Summary of AgrA constructs: amino acid sequences for: wild-type AgrA, AgrA_C199S and AgrA_D59E_C199S. Source organism, source DNA, primers, expression vector and expression host are also given. Mutated residues are underlined in red.



7.6.3: Construct Generation – SER Mutants

The primers shown in table 7.5 were ordered [Sigma Aldrich] and dissolved in Mili-Q™ H₂O to generate 200 μM stock solutions of each in accordance with the product information sheet supplied with each primer. PCR reactions were then set up according to the table 7.6.

The initial denaturation was at 98 °C for 5 mins, annealing was performed at 55 and 65 °C for 1 min with duplicate reactions incubated at each temperature. Extension was at 72 °C for 6 mins, followed by denaturation at 98 °C for 30 secs. A total of 35 cycles of denaturation, annealing and extension were performed, followed by a final extension at 72 °C for 8 mins.

All PCR reactions were treated with Dpn1 for 1 hr at 37 °C to destroy methylated template DNA followed by transformation of PCR product, 1 – 5 μL, into NovaBlue™ competent cells [Merck] via heat shock (See AgrA expression methods section 7.6.6 for details). All PCR reagents were purchased as high fidelity variants from NEB.

Transformant colonies were prepared exactly as per the AgrA Expression methods section 7.6.6 except no chloramphenicol was used since the NovaBlue™ cells do contain the pRARE plasmid. Transformant colonies were picked and inoculated into Lysogeny Broth (LB) growth medium, 10 mL, with ampicillin, 100 μg mL⁻¹, and cultured overnight with shaking at 180 rpm at 37 °C. Overnight cultures were plasmid miniprep using a GeneElute™ plasmid miniprep kit [Sigma Aldrich, catalogue number PLX50] and sent for Sanger DNA sequencing [Source Bioscience, Nottingham]. Several attempts at the PCR reaction were required to obtain the desired point mutations, all of which were eventually forthcoming. The resulting plasmids were stored at -20 °C for subsequent transformation into BL21.pRARE for overexpression. Construct amino acid sequences are given in table 7.7.



Construct	Primers: forward then reverse
AgrA_K101A_C199S	5'-AAAATAAAATCCATCGCTGCAACTGCGTAGACAAATGTAAATACGTAAGTTCAC-3' 5'-GTGAACTTACGTATTTAACATTTGTCTACGCAGTTGCAGCGATGGATTTTATTTT-3'
AgrA_Y100T_K101A_C199S	5'-GAGTCACAGTGAACCTACGTATTTAACATTTGTCCACCGCAGTTGCAGCGATGGATTTTATTTT-3' 5'-AAAATAAAATCCATCGCTGCAACTGCGGTGACAAATGTAAATACGTAAGTTCAGTGTGACTC-3'
AgrA_K77A_C199S	5'-GTATCAAATTAGGCAGTGAATTCGTGCGCATGACCCAGTTGGTAA-3' 5'-TTACCAACTGGGTCATGCGCACGAATTTCACTGCCTAATTTGATAC-3'
AgrA_K101T_C199S	5'-TAAAATAAAATCCATCGCTGCAACCGTGTAGACAAATGTAAATACGTAAGTTCAC-3' 5'-GTGAACTTACGTATTTAACATTTGTCTACACGGTTGCAGCGATGGATTTTATTTT-3'
AgrA_Y100T_K101T_C199S	5'-AAATAAAATCCATCGCTGCAACCGTGGTGACAAATGTAAATACGTAAGTTCAGTGTGACTC-3' 5'-GAGTCACAGTGAACCTACGTATTTAACATTTGTCCACCGTTGCAGCGATGGATTTTATTT-3'

Table 7.5. Summary of primers used to generate the AgrA_C199S SER constructs.

Component	Volume, μL
5 x Phusion® HF buffer	10
50 ng μL^{-1} pCOLD-I_AgrA_C199S template	1
20 μM forwards primer	1
20 μM reverse primer	1
50 % DMSO	4
Milli-Q™ H ₂ O	32.5
Phusion® polymerase	0.5

Table 7.6. Summary of PCR reaction contents used to generate the AgrA_C199S SER constructs.



AgrA_K101T_C199S	
Amino acid sequence of construct	MNHKVHHHHHHHIEGRHMELMKIFICEDDPKQRENMVTIKKNYIMIEEKPMIALA TDNPYEVLEQAKNMNDIGCYFLDIQLSTDINGIKLGSEIRKHDPVGNIFVTSHSEL TYLTFVY <u>I</u> VAAMDFIFKDDPAELRTRIIDCLETAHTRLQLLSKDNSVETIELKRGSN SVYVQYDDIMFFESSTKSHRLIAHLNDRQIEFYGNLKELSQLDDRFFRSHNSFVVN RHNIESIDSKERIVYXKNKEHCYASVRNVKKI
AgrA_K77A_C199S	
Amino acid sequence of construct	MNHKVHHHHHHHIEGRHMELMKIFICEDDPKQRENMVTIKKNYIMIEEKPMIALA TDNPYEVLEQAKNMNDIGCYFLDIQLSTDINGIKLGSEIR <u>A</u> HDPVGNIFVTSHSEL TYLTFVYKVAAMDFIFKDDPAELRTRIIDCLETAHTRLQLLSKDNSVETIELKRGSN SVYVQYDDIMFFESSTKSHRLIAHLNDRQIEFYGNLKELSQLDDRFFRCHNSFVV NRHNIESIDSKERIVYFKNKEHCYASVRNVKKI
AgrA_Y100T_K101T_C199S	
Amino acid sequence of construct	MNHKVHHHHHHHIEGRHMELMKIFICEDDPKQRENMVTIKKNYIMIEEKPMIALA TDNPYEVLEQAKNMNDIGCYFLDIQLSTDINGIKLGSEIRKHDPVGNIFVTSHSEL TYLTFV <u>TI</u> VAAMDFIFKDDPAELRTRIIDCLETAHTRLQLLSKDNSVETIELKRGSN SVYVQYDDIMFFESSTKSHRLIAHLNDRQIEFYGNLKELSQLDDRFFRSHNSFVVN RHNIESIDSKERIVYXKNKEHCYASVRNVKKI
AgrA_K101A_C199S	
	MNHKVHHHHHHHIEGRHMELMKIFICEDDPKQRENMVTIKKNYIMIEEKPMIALA TDNPYEVLEQAKNMNDIGCYFLDIQLSTDINGIKLGSEIRKHDPVGNIFVTSHSEL TYLTFVY <u>A</u> VAAMDFIFKDDPAELRTRIIDCLETAHTRLQLLSKDNSVETIELKRGSN SVYVQYDDIMFFESSTKSHRLIAHLNDRQIEFYGNLKELSQLDDRFFRSHNSFVVN RHNIESIDSKERIVYXKNKEHCYASVRNVKKI
AgrA_Y100A_K101A_C199S	
	MNHKVHHHHHHHIEGRHMELMKIFICEDDPKQRENMVTIKKNYIMIEEKPMIALA TDNPYEVLEQAKNMNDIGCYFLDIQLSTDINGIKLGSEIRKHDPVGNIFVTSHSEL TYLTFV <u>AA</u> VAAMDFIFKDDPAELRTRIIDCLETAHTRLQLLSKDNSVETIELKRGSN SVYVQYDDIMFFESSTKSHRLIAHLNDRQIEFYGNLKELSQLDDRFFRSHNSFVVN RHNIESIDSKERIVYXKNKEHCYASVRNVKKI

Table 7.7. Amino acid sequences for AgrA_C199S SER constructs: AgrA_K101T_C199S, AgrA_K77A_C199S, AgrA_Y100T_K101T_C199S, AgrA_K101A_C199S and AgrA_Y100A_K101A_C199S. Mutated residues are underlined in red.



7.6.4: Construct Generation – Individual AgrA Domains: AgrA₁₋₁₄₁ and AgrA₁₃₅₋₂₃₈_C199S

Construct generation was achieved via inverse PCR using the pCOLD-I_AgrA_C199S construct as the template DNA. PCR Reactions were established identically to those reactions used to generate the C199S and D59E_C199S double mutants but with the requisite primers. Methylated DNA was also destroyed in the exact same manner. Ligation of the PCR product was achieved by blunt-end ligation performed overnight at 10 °C. All primers were ordered with the 5' phosphate groups inserted [Sigma Aldrich]. Screening for transformant colonies was performed identically to the SER mutants. Construct amino acid sequences are given in table 7.8.

AgrA N-terminal domain only (AgrA₁₋₁₄₁)	
Forward primer	5'-[Phos]-TAATCTAGATAGGTAATCTCTGCTTAAAAG-3'
Reverse primer	5'-[Phos]-TTCAACGCTATTATCTTTTGATAATAATTGTAAGCG-3'
Amino acid sequence of construct	MNHKVHHHHHHIEGRHMELMKIFICEDDPKQRENMVTIKKNYIMIEEKPMIALA TDNPYEVLEQAKNMNDIGCYFLDIQLSTDINGIKLGSEIRKHDPVGNIIFVTSHSEL TYLTFVYKVAAMDFIFKDDPAELRTRIIDCLETATHRLQLLSKDNSVE
AgrA C-terminal domain only (AgrA₁₃₅₋₂₃₈_C199S)	
Forward primer	5'-[Phos]-AATAGCGTTGAAACGATTGAGTTAAAACGTGGCAGTAATTCAGTG-3'
Reverse primer	5'-[Phos]- CATGAGCTCCATATGCCTACCTTCGATATGATGATGATGATGATGCACTTTGTGA TTCAT-3'
Amino acid sequence of construct	MNHKVHHHHHHIEGRHMELMNSVETIELKRGSNVYVQYDDIMFFESSTKSHRL IAHLDNRQIEFYGNLKELSQLDDRFR <u>S</u> HNSFVVRHNIESIDSKERIVYFKNKEH CYASVRNVKKI

Table 7.8. Primer and amino acid sequences for AgrA constructs corresponding to each of the individual AgrA domains. Mutated residues are underlined in red.



7.6.5: SER Mutant Test Expression

5 x 10 mL overnight cultures of Terrific Broth (TB) with D-sorbitol, 1 M, ampicillin, 100 $\mu\text{g mL}^{-1}$ and chloramphenicol, 35 $\mu\text{g mL}^{-1}$, were inoculated with transformant BL21-CodonPlus [Agilent] colonies for each of the AgrA_C199S SER mutants. These were cultured overnight with shaking at 180 rpm at 37 °C. 3 mL of each of these overnight cultures was transferred separately to 50 mL of TB with D-sorbitol and antibiotics at the above concentrations. These in turn were cultured to an Optical Density at $\lambda = 600 \text{ nm}$ ($\text{OD}_{600\text{nm}}$) of 0.3 – 0.7. The 1 mL of the culture that was used to measure the optical density was kept as a pre-induction control sample, and isopropyl β -D-1-thiogalactopyranoside (IPTG), to 0.4 mM, added to each of the cultures. Overexpression of the SER mutants was allowed to proceed overnight with shaking at 180 rpm at 37 °C. Before collecting the cells via centrifugation, a 1 mL aliquot was taken as a “post-induction” test sample.

The pre- and post-induction samples were centrifuged at 17,000 g for 10 mins at room temperature to pellet the cells. The supernatant was discarded and 3 x SDS-PAGE loading buffer, 50 μL , added. The cells were gently resuspended and heated to 98 °C for 10 mins. A second identical centrifugation step was performed and 1.5 μL of the resulting supernatant was carefully loaded onto a 16 % polyacrylamide SDS-PAGE that was ran according to the General Methods procedure. This method for testing the expression of mutants was used throughout the PhD process even though not every test expression gel is presented in the Results and Discussion section.

7.6.6: Expression (For Full-Length, Wild-Type and All Mutants)

E. coli strain BL21-CodonPlus cells [Agilent] were transformed via heat shock. Typically, purified plasmid DNA, 1 μL , was gently mixed with competent cells, 50 μL . This mixture was subjected to a heat pulse of 42 °C for 60 sec, after which sterile LB, 0.5 mL, was added and the cells incubated at 37 °C for 45 mins. Transformant colonies were selected for by spreading ~ 125 μL of the transformant culture onto an LB agar plate containing ampicillin, 100 $\mu\text{g mL}^{-1}$, and chloramphenicol, 35 $\mu\text{g mL}^{-1}$, followed by overnight incubation at 37 °C.



Transformant colonies were inoculated into TB, 100 mL, containing D-sorbitol, 1 M, and antibiotics to the above concentration, and incubated overnight at 37 °C with shaking at 180 rpm. Overnight culture, 20 mL each, was re-inoculated into flasks containing TB, 1 L, with D-sorbitol, 1 M, and antibiotics to the above concentrations. Cells were cultured to an OD_{600nm} of ~ 0.8 and overexpression of AgrA_C199S induced by the addition of IPTG, 0.4 mM. Overexpression was allowed to proceed for 24 hrs at 16 °C with shaking at 200 rpm. Cells were collected by centrifugation at 4600 g for 40 mins at 7 °C and were stored at -20 °C.

7.6.7: Purification (For Full-Length, Wild-Type and All Mutants)

All of the following steps were performed on ice unless otherwise stated. 40 mL of lysis buffer (Na₂H/NaH₂PO₄, 20 mM, pH 7.0, NaCl, 300 mM, MgCl₂, 5mM and glycerol, 10 % v/v) was added to two cell pellets, each corresponding to biomass obtained from 1 L of TB overexpression culture to give an 80 mL resuspension. Serine protease phenylmethylsulfonyl fluoride (PMSF) was added to 5 mM, from a 100 mM stock in isopropanol. Cells were lysed via sonication at 15 μm with 30s on/off pulses for a total “on” time of 8 mins. Intact cells and cellular debris were pelleted via centrifugation at 24,000 g for 50 mins at 7 °C. The supernatant was loaded onto a 5 mL HiTrap™ Chelating HP Ni²⁺-affinity column [GE] using a P-1 pump [GE]. Bound sample was washed with 20 column volumes of wash buffer (Na₂H/NaH₂PO₄, 20 mM, pH 7.0, NaCl, 1.5 M, imidazole, 50 mM and glycerol, 10 % v/v) to remove weakly-bound sample from the column and weakly-bound DNA from the sample. Cleaned, bound sample was then washed into so-called buffer A (Na₂H/NaH₂PO₄, 20 mM, pH 7.0, NaCl, 300 mM, imidazole, 50 mM and glycerol, 10 % v/v). Elution of AgrA_C199S was achieved via 0 – 100 % buffer A – buffer B (Na₂H/NaH₂PO₄, 20 mM, pH 7.0, NaCl, 300 mM, imidazole, 500 mM and glycerol, 10 % v/v) gradient over 20 column volumes using an ÄKTA Purifier liquid chromatography unit [GE] using a flow rate of 2.0 mL min⁻¹. Dithiothreitol (DTT) was added, to 10 mM, immediately to peak fractions to prevent oxidative aggregation of AgrA which otherwise occurred within 10 mins of elution.



Pooled sample was spin-concentrated to ~ 5 mL using a Vivaspin™ concentrator [Thermo Fisher] with a 10 kDa molecular weight cut-off, at 5000 g at 7 °C. Concentrated sample was centrifuged at 17,000 g for 10 mins at 7 °C to pellet any aggregated protein before Size Exclusion Chromatography (SEC). Spun sample was loaded onto a 10 mL gel filtration loop and injected onto a HiLoad® Superdex® 75 16/600 gel filtration column [GE] equilibrated into running buffer (Na₂H/NaH₂PO₄, 20 mM, pH 7.0, NaCl, 300 mM, EDTA, 1mM, DTT, 10 mM and glycerol, 5 % v/v). In order to prevent degradation of the sample during the SEC step, the column was pre-equilibrated and chilled at 4 °C until immediately prior to use. SEC was carried out at a flow rate of 0.8 mL min⁻¹ and AgrA_C199S migrated as a monomer, eluting as a single resolved peak at ~ 70 mL.

Sample purity and integrity was analysed via reducing SDS-PAGE. Pooled sample was spin-concentrated to ~ 5 mL using a Vivaspin™ concentrator [Thermo Fisher] with a 10 kDa molecular weight cut-off at 3000 g at 7 °C to a volume of 5 mL. Sample concentration was determined using the Nanodrop™ 1000 [Thermo Fisher] at a wavelength of 280 nm using a theoretical extension coefficient = 14,900 M⁻¹ cm⁻¹ assuming all cysteine residues are reduced. Sample was flash frozen on N₂ (l) and stored at -80 °C.

7.6.8: AgrA₁₋₁₄₁ Purification (N-terminal Domain Only)

A construct encoding the N-terminal 141 AgrA residues was prepared using inverse PCR of the pCOLD-I_AgrA_C199S construct used throughout this document as template DNA, using back-to-back primers followed by blunt-end ligation at 10 °C. This work was carried out by a new PhD researcher Sara Zandomenighi under the guidance of Dr. Philip Bardelang. The AgrA₁₋₁₄₁ construct was expressed and purified by Ni²⁺-affinity by Sara using expression method and buffers developed by the experimenter for full-length AgrA_C199S.

A 2 mL aliquot of Ni²⁺-affinity purified sample was handed to the experimenter and was subject to SEC using a Hi-Load® Superdex® 75 16/600 gel filtration column [GE] equilibrated with running buffer (Na₂H/NaH₂PO₄, 20 mM, pH 7.0, NaCl, 300 mM, EDTA, 1mM, DTT, 10 mM and glycerol, 5 % v/v) at a flow rate of 0.8 mL min⁻¹. The sample eluted as a monomer with a retention volume of ~



78 mL. Pooled sample was spin-concentrated to ~ 5 mL using a Vivaspin™ concentrator [Thermo Fisher] with a 3 kDa molecular weight cut-off at 3000 g at 7 °C to a concentration of 178 μ M. Sample concentration was determined using the Nanodrop™ 1000 [Thermo Fisher] via measuring the absorbance maxima at a wavelength of 280 nm and using a theoretical extinction coefficient = 7450 $M^{-1} \text{ cm}^{-1}$ assuming all cysteine residues were reduced. A final centrifugation at 17,000 g at 7 °C for 10 mins was performed and the sample was used immediately to establish sparse matrix crystallisation trials as per section 7.6.12.

7.6.9: AgrA₁₃₅₋₂₃₈_C199S Purification (C-terminal Domain Only)

Expression and purification to AgrA₁₃₅₋₂₃₈_C199S (C-terminal domain only) was performed identically to full-length AgrA_C199S and using identical buffers, except for a few key differences. No reducing agents were required to stabilise the sample and the Ni²⁺-affinity chromatography and SEC steps could be performed on separate days with no sample degradation. Spin-concentrators with appropriate cut-offs were used to prevent loss of sample.

7.6.10: SEC Peak Shift DNA Binding Assay

AgrA_C199S, 0.5 mL, 62.5 μ M, was defrosted on ice and DTT added, to 10 mM. 2 molar equivalents of DNA with 5' A/T overhangs was added and allowed to bind to AgrA_C199S for 15 – 20 mins on ice. The sample was centrifuged at 17,000 g for 10 mins at 7 °C and loaded onto a HiLoad® Superdex® 75 16/600 column [GE]. SEC was performed at a flow rate of 0.8 mL min⁻¹ and the retention volume of the sample noted. DNA-free sample typically produced a single, Gaussian peak eluting at ~ 70 mL on the same column under identical running conditions during protein purification. The assay produced two peaks on SEC: one eluted at ~ 62 mL corresponding to the AgrA_C199S/DNA complex and the other eluted at ~ 69 mL, corresponding to excess DNA, which has a (relatively) large hydrodynamic radius owing to the linear conformation of a short segment dsDNA. Presence of protein and DNA in respective SEC peaks was confirmed by retention time comparison between DNA-bound and DNA-free SEC peaks, SDS-PAGE and Nanodrop™ 1000 measurements at 260 nm.



7.6.11: Oligoduplex Generation (P2_{fragment} Blunt-End, With 5' A/T Overhangs and Brominated)

The following oligonucleotides, corresponding to a single 9 bp (Base Pair) AgrA-binding site on the P2 promoter region with 3 bp flanking regions, were ordered [Sigma Aldrich] to 1 μ M as dry, desalted samples: 5'-TTTAACAGTTAAGTAT-3' and 5'-AATACTTAACTGTAA-3'. Each strand was made up to 3 mM in Milli-QTM H₂O and mixed together in equal volumes in a PCR tube. The mixture was heated to 98 °C for 5 mins and annealed via cooling at a rate of 10 °C hr⁻¹ to 4 °C in a PCR machine set to the appropriate reaction volume (~ 80 μ L) to generate a 1.5 M 15 bp oligoduplex with 1b 5' A/T overhangs stock solution in H₂O for co-crystallisation experiments. In the instances where bromated DNA was used to grow crystals, the above steps were repeated identically for oligonucleotides: 5'-TTTAACAGTTAAG[5-BrdU]AT-3' and 5'-AA[5-BrdU]ACTTAACTGTAA-3', where 5-BrdU corresponds to 5-bromouracil. This modification is available through Sigma Aldrich for an additional cost. In the instances where "blunt end" DNA was used to attempt to grow crystals, the above steps were repeated identically for oligonucleotides 5'-TTAACAGTTAAGTAT-3' and 5'-ATACTTAACTGTAA-3'. A summary of these oligonucleotides is given below in table 7.9.

Oligonucleotide	Sequence
P2 _{fragment} with 5' A/T overhangs forwards	5'-TTTAACAGTTAAGTAT-3'
P2 _{fragment} with 5' A/T overhangs reverse	5'-AATACTTAACTGTAA-3'
Brominated P2 _{fragment} with 5' A/T overhangs forwards	5'-TTTAACAGTTAAG[5-BrdU]AT-3'
Brominated P2 _{fragment} with 5' A/T overhangs reverse	5'-AA[5-BrdU]ACTTAACTGTAA-3'
Blunt-end P2 _{fragment} forwards	5'-TTAACAGTTAAGTAT-3'
Blunt-end P2 _{fragment} reverse	5'-ATACTTAACTGTAA-3'

Table 7.9: Summary of oligonucleotides used in the growth and optimisation of the AgrA_C199S/P2_{fragment} crystals.



7.6.12: Sparse Matrix Crystallisation Screening

Apo and DNA-bound AgrA_C199S were both used to perform sparse matrix vapour diffusion crystallisation experiments using the mosquito liquid dispensing unit [TTP Labtech]. 200 nL of purified sample was mixed with 200 nL of crystallisation solution, and incubated against an 80 μ L reservoir of crystallisation solution in a 96 x well crystallisation plate [Molecular Dimensions]. No crystals were ever observed for *apo* AgrA_C199S, which was treated identically to DNA-bound AgrA_C199S in all respects except the addition of DNA.

Extremely small ($< 10 \mu\text{m}$ for the longest axis) plate-like crystals were observed for DNA-bound AgrA_C199S. The sample used to obtain the initial crystals was “worked up” as follows. Unbound sample in SEC running buffer was thawed on ice. DTT, to 10 mM, and DNA with 5' A/T overhangs, 2 molar equivalents, were both added to AgrA_C199S. DNA was allowed to bind to AgrA_C199S for 15 – 20 mins on ice, after which the sample was concentrated to 250 μM in a VivaspinTM concentrator [Thermo Fisher] with a 10 kDa molecular weight cut-off at 5000 g at 7 °C. Serial dilutions were prepared using fresh SEC running buffer to obtain DNA-bound samples at 250, 125 and 62.5 μM . Crystals were observed for 1:1 mixtures of both the 125 and 62.5 μM sample with: bis-tris propane, 100 mM, pH 6.5, NaF, 200 mM and polyethylene glycol (PEG) 3350, 20 % w/v. The 62.5 μM sample produced the largest crystals. The temperature of the experiment was maintained at 10 °C throughout and crystals appeared after approximately two weeks. The commercial screen which afforded the initial crystallisation condition was PACT PremierTM [Molecular Dimensions].

7.6.13: Control Crystallisation Experiments

To ensure the observed crystals were not DNA or buffer component false positives prior to X-ray diffraction experiments, DNA with 5' A/T overhangs was diluted into fresh SEC running buffer to 250, 125 and 62.5 μM , and DTT added, to 10 mM. A negative control crystallisation experiment was then set up identically to that which produced the prospective AgrA_C199S/DNA crystals. Control experiments were inspected for crystals periodically over the ensuing weeks and months. Neither precipitate nor crystals ever appeared.



7.6.14.1 Crystal Optimisation 1 – 2D Chemical Grid Screening

As well as observing crystals NaF, 200 mM, crystals were also observed for conditions containing (in place of NaF): NaBr, NaI, NaOOH and NaOAc, all 200 mM, corresponding to conditions: F1, F2, F3, F6, and F7 of the PACT Premier™ screen [Molecular Dimensions]. Crystals in the NaF-containing condition were appreciably larger (although still very small) than the crystals observed in the presence of the other sodium salts. Crystals in the NaBr-containing condition were the second largest. Therefore, the concentration of these two salts was nominated to be variables in the initial crystal optimisation experiments.

The PACT Premier™ screen is designed such that, for rows E, F, G and H, only the identity of the sodium salt (and not its concentration) is varied going along the twelve conditions in each of the rows. All other reagents, namely bis-tris propane, and PEG 3350, are invariant and are maintained at 100 mM and 20 % w/v respectively. Going down each of the rows, the pH of the bis-tris propane is altered by one pH unit each time. That is to say all twelve conditions of row F are at pH 6.5, row G pH 7.5 and row H pH 8.5. As AgrA_C199S/P2_{fragment} crystals did not appear in any conditions in rows G and H, it was inferred that AgrA_C199S/P2_{fragment} crystallogenesis is highly pH-dependant and so pH was not chosen as a variable to alter in the initial crystal optimisation experiments.

The initial crystallisation condition was optimised by preparing two two-dimensional chemical grid screens, in which the concentration of the sodium salt was varied to: 150, 175, 200 and 225 mM, and the concentration of the PEG 3350 was varied to: 13, 15, 17, 19, 21, 23 %. This grid screen was prepared for both NaF and NaBr in duplicate in order to screen crystallisation temperature, which was performed at temperatures of 10 and 20 °C for each of the sodium salts NaF and NaBr.

Experiments were set up using the sitting drop vapour diffusion approach, with a 1:1 experimental drop containing 1 µL of AgrA_C199S/P2_{fragment} DNA, 62.5 µM in SEC running buffer, and 1 µL crystallisation solution, incubated against an 80 µL reservoir of crystallisation solution. Experiments were set up in a 48 well MRC Maxi crystallisation plate [Hampton Research]. DNA was introduced to AgrA_C199S as described for the initial crystallisation screening experiments. The largest crystals



were observed in NaF, 175mM, bis-tris propane, 100 mM, pH 6.5, PEG 3350, 21 %. This composition of crystallisation solution was therefore used in later optimisation experiments, and “crystallisation solution” is taken to mean this solution the in all Results and Discussion sections in which it is mentioned.

Additionally PEG 3350, 21 – 27 % w/v, was screened using a grid approach identical to that described above.

7.6.14.2: Crystal Optimisation 2 – Hanging Drop Experiments

Optimised crystallisation solution, 500 μ L, was added to the reservoir of a 24 well hanging drop crystallisation plate [MiTeGen]. The perimeter of each reservoir was lined with vacuum grease [Dow-Corning]. 2.5 μ L of crystallisation solution was mixed with 2.5 μ L AgrA_C199S/P2_{fragment} DNA on a 22 mm square siliconised glass cover slide [Hampton Research]. The cover slide was inverted and placed over the reservoir with gentle pressure applied so as the grease formed an air-tight seal. The AgrA_C199S/P2_{fragment} DNA concentration was screened as a variable in the hanging drop experiments, by taking AgrA_C199S, 62.5 μ M, in SEC running buffer, mixing it with 2 molar equivalents of DNA, adding DTT, to 10 mM, and then preparing serial dilutions in SEC running buffer, to 32.3 and 15.6 μ M AgrA_C199S (with 2 molar equivalents of P2_{fragment} DNA). Hanging drop vapour diffusion experiments, and all subsequent optimisation experiments, were performed at 20 °C.

7.6.14.3: Crystal Optimisation 3 – Streak Seeding

AgrA_C199S was diluted to 50 μ M in SEC running buffer and 2 molar equivalents of DNA added. Experimental drops to which the seeds were to be administered were set up using the sitting drop vapour diffusion approach, with a 1:1 experimental drop containing 2.5 μ L of AgrA_C199S/DNA, 50 μ M in SEC running buffer, and 2.5 μ L crystallisation solution, incubated against an 80 μ L reservoir of crystallisation solution. Experiments were set up in a 48 well MRC Maxi crystallisation plate [Hampton Research]. Sealed, unseeded drops were allowed to equilibrate for ~ 7 hrs before seeding was performed. Microseeds were administered by gently dragging a hair through the drop containing



the initial hit crystals and then straight through the newly-prepared drop. The experiment was then resealed. Dehydration of the exposed drops containing the seeds was prevented by placing a 22 mm siliconised glass cover slide [Hampton Research] over the top of the drops.

7.6.14.4: Crystal Optimisation 4 – Seeding from Seed Stocks

A drop into which microseeds were to be administered was prepared as per the streak seeding method, and a similar timeframe was allowed for the drop to equilibrate. Meanwhile, seed stocks were prepared by adding 3 μL of cold crystallisation solution to a crystal-containing experimental drop, many of which were afforded by the initial two-dimensional grid optimisation. A micro-needle [Hampton Research] was used to gently crush and smear the crystals. The entirety of the resulting mixture was transferred to an Eppendorf® tube [Sigma Aldrich] and a further 80 μL of cold crystallisation solution added to dilute the microseeds. 80 μL of this mixture was then transferred to a chilled Eppendorf® tube containing ceramic seed beads [Hampton Research]. The microseeds were obliterated by vortexing, 5 x 30 secs, with intermittent cooling, 30 secs on ice, between each vortex. 5 x ten-fold serial dilutions of the resulting “seed stock” were prepared by taking seed stock, 8 μL , and adding cold crystallisation solution, 72 μL , to achieve a dilution range of 10^{-1} – 10^{-5} of the original seed stock. Crystallisation experiments were prepared by mixing seed stock, 0.5 μL , 62.5 μM AgrA_C199S/P2_{fragment} DNA, 2.0 μL , and crystallisation solution, 2.5 μL . The resulting 5 μL experimental drop was equilibrated against either 80 μL of crystallisation solution in a 48 well sitting drop plate, or 500 μL of crystallisation solution, in a 24 well hanging drop plate in order to test both sitting and hanging drop methods in tandem with seed stocks. Typically, five replicate experimental drops were performed for each seed stock dilution.

7.6.14.5: Crystal Optimisation 5 – DNA Concentration Screening

AgrA_C199S in SEC running buffer, 62.5 μM , was incubated with, 1.0, 1.2, 1.5 and 2.0 molar equivalents of P2_{fragment} DNA in the presence of DTT, 10 mM. DNA was allowed to bind to AgrA_C199S for 15 – 20 mins in each case. Samples were then centrifuged at 17,000 g for 10 mins at 7 °C. Crystallisation experiments were subsequently set up via streak seeding microcrystals into



experimental hanging drops as outlined above. Each experimental condition was established with five redundant drops per condition in order to mitigate both the gross errors and the randomness that are associated with the streak seeding method.

7.6.14.6: Crystal Optimisation 6 – Additive Screening

Additive screening was performed using both the Angstrom [Molecular Dimensions] and Additive Screen [Hampton Research] additive screens, in a 96 well sitting drop vapour diffusion format using a 96 well 2 drop MRC crystallisation plate [Molecular Dimensions]. The reservoir of each experiment contained: NaF, 175 mM, bis-tris propane, 100 mM, pH 6.5, PEG 3350, 21 % w/v, 72 μ L, mixed with 8 μ L of each of the additive screen conditions (representing a ten-fold dilution of the latter) to give a total volume of 80 μ L. 0.5 μ L of the reservoir solution was mixed with AgrA_C199S/P2_{fragment}, 62.5 μ M, 0.5 μ L, to give a 1 μ L experimental drop. Crystallisation plates were sealed and incubated at 20 °C. Increases in crystal size were checked for by visually comparing results from the additive screen with those crystals which grew from the initial hit condition. L-rhamnose afforded the single largest increase in crystal size hence it was added to all subsequent crystallisation optimisations.

7.6.14.7: Crystal Optimisation 7 – Blunt-End, With 5' A/T Overhangs and Brominated DNA

The oligonucleotides (given in table 7.9) were purchased [Sigma Aldrich] with High-Performance Liquid Chromatography (HPLC) purity and annealed as outlined above. AgrA_C199S in SEC running buffer, 62.5 μ M, was mixed with 2 molar equivalents of each oligoduplex in the presence of DTT, 10 mM, and allowed to bind to AgrA_C199S for 15 – 20 mins. 2.5 μ L of each of the AgrA_C199S/DNA samples was mixed with 2.5 μ L of crystallisation solution containing L-rhamnose, 1.2 % w/v, in hanging drop vapour diffusion experiments, set up as described above. These experiments were subsequently streak seeded into also as described above. Experiments were sealed and incubated at 20 °C with periodic monitoring for crystallisation.



7.6.14.8: Crystal Optimisation 8 – PEG Screening

The 21 % w/v PEG 3350 in the optimised crystallisation solution was replaced with PEG 1500, PEG 3500 and PEG 4000, all to 21 % w/v, whilst all other reagents remained invariant. Hanging drop vapour diffusion experiments were then established and subsequently streak seeded into in replicate as described previously.

7.6.14.9: Crystal Optimisation 9 – *In Situ* Macroseeding

The best result after streak seeding into hanging drops (judged as to which drop contained just a few large crystals rather than larger number of smaller ones) was carefully opened. On a separate siliconised glass cover slide, 2.5 μL of freshly-defrosted 62.5 μM AgrA_C199S/P2_{fragment} in SEC running buffer and DTT, 10 mM, was mixed with 2.5 μL of fresh crystallisation solution. 2 – 3 μL of the crystal-containing drop mother liquor was removed and discarded 1 μL at a time with care being taken to avoid aspirating any of the crystals. The same volume of the fresh mother liquor mixture was then carefully injected back into the crystal-containing drop and the experiment resealed. This process was repeated as necessary to enlarge crystals initially grown via streak seeding methods.

7.6.14.10: Crystal Optimisation 10 – AgrA_D59E_C199S Crystallogenesis

AgrA_D59E_C199S was cloned and purified as described above. Crystallisation experiments were performed via streak seeding into hanging drops using the AgrA_C199S/P2_{fragment} crystals as microseeds. All other experimental procedures were identical to those already described. The resulting crystals were heavily twinned and of no use for X-ray diffraction experiments.

7.6.14.11: Crystal Optimisation 11 – $[\text{BeF}_3]^-$ as a Phosphate Mimetic. All Strategies

Fluoride adducts of beryllium, $[\text{BeF}_x]^{n-}$, may be formed spontaneously in aqueous solutions.^[228] Several approaches were used to prepare $[\text{BeF}_x]^{n-}$ and introduce it to AgrA_C199S. A) Appreciably-sized crystals in a 5 μL hanging drop were soaked, after removal of 1 μL of mother liquor, with 1 μL of crystallisation solution containing BeSO_4 , 25 mM, and MgCl_2 , 25 mM, thus achieving a final concentration of 5 mM for both additional components in the experimental drop. B) $[\text{BeF}_x]^{n-}$ was



prepared in the reservoir by mixing an aqueous solution of BeSO_4 , 400 mM, and MgCl_2 , 400, mM, 2 μL total, with crystallisation solution, 78 μL , to achieve a concentration of both additional components of 10 mM in the reservoir. 2.5 μL of the reservoir solution was then mixed with 2.5 μL of 62.5 μM AgrA_C199S/P2_{fragment} in SEC running buffer and DTT, 10 mM. Hanging drop vapour diffusion experiments were then established and subsequently streak seeded into. C) 62.5 μM AgrA_C199S/P2_{fragment} in SEC running buffer and DTT, 10 mM, 45 μL total, was mixed directly with an aqueous solution of BeSO_4 , 225 mM, and MgCl_2 , 225, mM, 1 μL total. This solution was then used to set up hanging drop vapour diffusion experiments which were streak seeded into. D) An aqueous solution of BeSO_4 , 225 mM, MgCl_2 , 225, mM, and NaF, 500 mM, was prepared and left to stand at room temperature for 30 mins. 1 μL of this solution was then mixed with 62.5 μM AgrA_C199S/P2_{fragment} in SEC running buffer and DTT, 10 mM, 45 μL total, and hanging drop vapour diffusion experiments established which were subsequently streak seeded into.

7.6.14.12: Crystal Optimisation 12 – Sugar Additive Screening

Since the additives which imparted the biggest increase in crystal size were invariably sugar molecules, a crystallisation experiment was devised to empirically test the gain in crystal size achieved from several of the best-performing sugars in the additive screen. Crystallisation solution was prepared with the addition of either: L-rhamnose, 1.5 % w/v, or D-glucose, trehalose, sucrose, or D-sorbitol, 3 % w/v (the same concentrations as in the additive screens). Seed stocks were prepared as described above and ten-fold serial dilutions generated over a range of $1^{-1} - 10^{-3}$. Crystallisation experiments were prepared by mixing the seed stock, 0.5 μL , with 62.5 μM AgrA_C199S/P2_{fragment}, 2.0 μL , and each of the sugar-containing crystallisation solutions, 2.5 μL . The resulting 5 μL experimental drops were equilibrated against a 500 μL reservoir in a 24 well hanging drop crystallisation tray as previously described, thus generating a “seed stock titrant vs. sugar additive” grid. This entire experiment was repeated, but with L-rhamnose at 1.0, 1.1, 1.2, 1.3, 1.4 and 1.5 % w/v, once it was deduced to be the sugar that imparted the largest increase in crystal size when compared to the others.



7.6.14.13: Crystal Optimisation 13 – Greased Experimental Wells

Sitting drops, into which streak seeding was to be performed, were set up as described above. To try to overcome the problems associated with crystals adhering to the bottom of the crystallisation well, a small amount of vacuum grease [Dow-Corning] was applied to the bottom of experimental wells using a disposable Kimwipe™ [Cole-Parmer]. Excess grease was wiped away using a fresh wipe, resulting in a fine layer of grease along the bottom of each well. Crystallisation experiments were then set up as described.

7.6.14.14: Crystal Optimisation 14 – Cryoprotectant Screening

The AgrA_C199S/DNA cryoprotectant used to prevent ice formation which ultimately facilitated the AgrA_C199S/P2_{fragment} structure solution was: NaF, 87.5 mM, bis-tris propane, 50 mM, pH 6.5, PEG 3350, 17.5 % w/v, L-rhamnose, 0.6 % w/v, and glycerol, 40 % v/v. This solution was prepared in Milli-Q™ H₂O and sterile filtered using a 0.22 µm syringe filter [Cole-Parmer]. > 5 µL of this solution was mixed, 1 µL at a time, into the crystal-containing drop whilst monitoring the crystals for damage. No visible damage was observed for any of the crystals. The crystal was then fished manually using an appropriately sized CryoLoop™ [Hampton Research] and cryocooled on N₂ (l) in a unipuck [Diamond Light Source]. Xylitol, 30 % w/v, trehalose, 30 % w/v, PEG 400, 40 % v/v, and ethylene glycol, 40 % v/v, were all also used as cryoprotecting agents (taking the place of glycerol in the above cryoprotectant solution) throughout the crystal optimisation process, whenever beam time at Diamond Light Source permitted. None of these prevented so-called ice rings.

7.6.15: Cryocooling, Data Collection, Model Building and Refinement

Crystals were cryocooled by adding 1 µL drops of bis-tris propane, 50 mM, pH 6.5, NaF, 87.5 mM, PEG 3350, 17.5 % w/v, L-rhamnose, 0.6 % w/v, and glycerol, 40 % v/v, until the experimental drop size had been doubled. Crystals were then fished and cryoprotected on N₂ (l). Crystal data were collected at Diamond Light Source beamline I04 using beam size of a 20 µm², an expose time of 0.2 s, a rotation per exposure of 0.2 ° and a transmission of 69.62 %. 900 images were collected over 180 ° of rotation. The crystals diffracted to a resolution of 4.75 Å with a spherical completeness of 99.3 %



in space group I222. Crystal data were integrated using DIALS (vers. 1.14.2),^[248] indexed using POINTLESS^[249] (vers. 1.11.19) and scaled and merged using AIMLESS^[199] (0.7.4) all available as part of the CCP4 software suite^[250] (vers. 7.0.072). An anisotropy correction was applied using the STARANISO server^[251] (vers. 3.315) [Global Phasing Ltd].

Initial phases were determined via molecular replacement (MR) using PHASER^[182] (vers. 2.7) [CCP4]. The DNA-binding domain in complex with DNA was located using the available structure of this region^[80] (PDB: 3bs1) with water atoms, ions and the five N-terminal residues MDNSV removed in PyMOLTM (vers. 2.2.0) [Schrödinger Inc].^[244] The resulting molecule was used as the search model. It returned a unique solution with a Log Likelihood Gain (LLG) of ~ 100. To build the rest of the sequence of the first 135 amino acids of AgrA (AgrA₁₋₁₃₅) was searched against the PDB using the BLAST server.^[243] The coordinates of the N-terminal domain of LytR from *Staphylococcus aureus*^[88] (PDB: 6m8o) were used to perform a subsequent molecular replacement MR step. Again all non-protein atoms were removed from the search model. This returned a unique solution with an LLG of ~ 250. To achieve the correct AgrA sequence, a homology fold of AgrA₁₋₁₃₅ was generated using the Phyre2 server.^[242] This homology fold was superposed over the PHASER-positioned LytR unit using Chimera^[245] (vers. 1.12.2rc) [UCSF] and the LytR atoms were deleted. This coordinate file was used to perform a third MR step which returned a unique solution with an LLG of 257. Inspection of this result revealed the asymmetric unit to be comprised of N- and C-terminal domains from two different protomers in the crystal. Shifting the N-terminal domain to a symmetry-equivalent position yielded a single AgrA_C199S polypeptide, and re-performing the MR step with it returned an LLG of 430.

The asymmetric unit was found to contain a single AgrA_C199S protein and a single DNA molecule. The solvent content was determined to be 66.63 % based on a Mathew's coefficient of $3.69 \text{ \AA}^3 \text{ Da}^{-1}$. Manual model editing was performed in COOT^[196] (vers. 0.8.9) [CCP4] and refinement was performed in PHENIX^[197] (vers. 1.17.1) with secondary structure restraints for the N-terminal domain (apart from $\alpha 4$), group B-factors and optimised X-ray/stereochemistry weights enabled. Structure validation was performed using MolProbity^[270] (vers 4.2) and the PDB validate server (vers. 4.4). The structure refined to an R and R_{free} of 30 and 36 % with a MolProbity score of 2.28. No major



validation issues (PDB validate server) are present in the final model. Nineteen residues, corresponding to the (His)₆ tag, factor Xa protease site and the pCOLD-I translation enhancing element were not modelled. These all reside at the N-terminus of the protein. Only poor electron density was observed for AgrA residues 136-140, which correspond to the linker region between the two domains and has the sequence KDNSV. Thus these residues were not modelled. Of the seven Ramachandran outliers, the two prolines, Pro30 and Pro40 and Gly148 are found in loops whilst Tyr41 immediately follows Pro40. Residues Ser90 and Val102 are also in loops and Ser63 is involved in a β -turn.

7.6.16: Preparation of AgrA_C199S Dimer on Full Promoter Regions

The oligonucleotides in table 7.10 were ordered [Sigma Aldrich] to 1 μ M as dry desalted samples. Each oligonucleotide was made up to 3 mM in Milli-Q™ H₂O and each forwards oligonucleotide was mixed in an equimolar ratio with its reverse oligonucleotide partner. Annealing was achieved by heating the mixtures up to 98 °C for 10 mins followed by cooling at a rate of 10 °C hr⁻¹ to achieve 1.5 mM stocks of both the P2 and P3 promoter regions with 5' A/T overhangs.

180 μ L of 125 μ M AgrA_C199S was defrosted on ice and fresh DTT, to 10 mM, added from a 1 M stock. This sample was then split into two equal fractions and 1 molar equivalent of P2 and P3 DNA from the 1.5 M stock was added to each and allowed to bind on ice for 15 – 20 mins. Separately, an aqueous solution of BeSO₄, 225 mM, and MgCl₂, 225 mM was prepared in Milli-Q™ H₂O and mixed with an equal volume of NaF, 500 mM, in Milli-Q™ H₂O. [BeF₃]⁻ was allowed to form by leaving this mixture on the bench for ~ 30 mins. The P2- and P3-bound AgrA_C199S samples were once more divided into two equal fractions to give a total of 4 x 45 μ L samples. 10 molar equivalents of the [BeF₃]⁻ solution was added to one of the AgrA_C199S/P2_{full} and AgrA_C199S/P3_{full} samples and allowed to bind on ice for 10 mins followed by centrifugation at 17,000 g for 10 mins at 7 °C. Half of each of the four samples was transferred to a clean Eppendorf® tube and diluted two-fold via addition of an equal volume of SEC running buffer.



These samples were then used to establish sitting drop crystallisation experiments in conditions corresponding to Morpheus® well A12 and MIDASPlus™ well A11 in 48 well sitting drop crystallisation plates with 2 µL experimental drops and 80 µL reservoirs. Incubation of the plates was at 20 °C.

Oligonucleotide	Sequence
P2 Forward	5'-ATACATTTAACAGTTAAGTATTTATTTTCCTACAGTTAGGCAATATAATG-3'
P2 Reverse	5'-TCATTATATTGCCTAACTGTAGGAAATAAATACTTAACTGTTAAATGTA-3'
P3 Forward	5'-TAATTTTTCTTAACTAGTCGTTTTTTATTCTTAACTGTAAATTTTT-3'
P3 Reverse	5'-AAAAAATTTACAGTTAAGAATAAAAAACGACTAGTTAAGAAAAATT-3'

Table 7.10. Summary of oligonucleotides used of AgrA_C199S DNA co-crystallisation experiments.



7.6.17: Strategy for Introducing $[\text{BeF}_3]^-$ to AgrA_C199S Without Precipitation

Purified AgrA samples at any desired concentration were treated with 0.2 – 1.0 molar excess of DNA (the sequence and length of the DNA depends on the downstream experiment. Full promoter sequences containing two AgrA binding sites generally afforded more protection from the harsh $[\text{BeF}_3]^-$ when used at 1.0 molar excess) and left to bind on ice for 10 – 20 mins. Separately an aqueous solution of BeSO_4 , 225 mM, and MgCl_2 , 225 mM was prepared in Milli-Q™ H_2O and mixed with an equal volume of NaF, 500 mM in Milli-Q™ H_2O . $[\text{BeF}_3]^-$ was allowed to form by leaving this mixture on the bench for 30 mins. Note that resulting solution, although assumed to be a 125 mM stock $[\text{BeF}_3]^-$ in 250 mM NaF, is more likely a mixture of Be^{2+} and F^- adducts, i.e. $[\text{BeF}_x]^{n-}$. 10 molar equivalents of the $[\text{BeF}_3]^-$ solution was added to the AgrA_C199S/DNA mixture and allowed to bind on ice for 10 mins. Downstream experiments were then performed. Using 100 molar equivalents of $[\text{BeF}_3]^-$ generally lead to sample precipitation, as did adding the $[\text{BeF}_3]^-$ before the DNA. It may be that $[\text{BeF}_3]^-$ can be delivered to AgrA_C199S without precipitation of the sample additions lower than ten-fold excess.

7.6.18: AgrA_C199S Dimer on P3_{full} DNA Preparation for SAXS

The oligonucleotides which have been demonstrated to bind AgrA^[82] corresponding to the P3 promoter region (given in table 7.10) were ordered [Sigma Aldrich] to 1 μM as dry, desalted samples and annealed identically to the DNA used in the crystallisation experiments. 250 μL of 62.5 μM AgrA_C199S was defrosted on ice and fresh DTT added, to 10 mM, followed by 1 molar equivalent of P3 DNA from the 1.5 M stock solution which was allowed to bind on ice for 15 – 20 mins. Separately an aqueous solution of BeSO_4 , 225 mM, and MgCl_2 , 225 mM, was prepared in Milli-Q™ H_2O and mixed with an equal volume of NaF, 500 mM, in Milli-Q™ H_2O . $[\text{BeF}_3]^-$ was allowed to form by leaving this mixture on the bench for ~ 30 mins. 10 molar equivalents of $[\text{BeF}_3]^-$ was added to the AgrA_C199S/P3_{full} mixture and allowed to bind on ice for 10 mins followed by centrifugation at 17,000 g for 10 mins at 7 °C. The sample was loaded onto a 500 μL gel filtration sample loop and SEC performed using a Superdex® 200 Increase 10/300 GL gel filtration column [GE] equilibrated



into running buffer ($\text{Na}_2\text{H}/\text{NaH}_2\text{PO}_4$, 20 mM, pH 7.0, NaCl, 300 mM, β -mercaptoethanol, 5 mM and glycerol, 5 % v/v) at a flow rate of 0.5 mL min^{-1} . The sample eluted as a single species at a retention volume of 13.6 mL, corresponding to an M_r of 140 kDa according to the column calibration curve given in appendix II. The true M_r of the dimer/promoter complex is 89 kDa and the discrepancy between the two mass values is because of the large hydrodynamic radius of short, rod-shaped DNA molecules. The sample was spin-concentrated to $73 \mu\text{M}$ using a VivaspinTM concentrator [Thermo Fisher] with a 10 kDa molecular weight cut-off at 5000 g at 7°C . Protein concentration in the sample was assumed to be twice the molarity of the DNA and was calculated using the absorbance maxima at a wavelength of 260 nm using a NanodropTM 1000 [Thermo Fisher]. Sample was not frozen before downstream experimentation but transported on ice and used within 48 hrs.

7.6.19: AgrA_C199S Dimer on P3_{full} DNA SAXS Data Collection

Data were collected at Diamond Light Source beamline B21 using their standard SEC-SAXS setup. To check the quality of the sample, three two-fold serial dilutions of the $73 \mu\text{M}$ were prepared and screened to check that R_g varied with sample concentration. Four datasets were subsequently taken: one for the $73 \mu\text{M}$ sample and one for each of the serial dilutions. Buffer scattering, measured from the column flow through away from the sample peak, was subtracted from the raw sample scattering data. The dataset corresponding to the $73 \mu\text{M}$ buffer-subtracted dataset was analysed using the ATSAS suite^[300] (vers. 3.0.1) [EMBL]. The Molecular envelope was determined using the DAMMIF^[298] package and the envelope constructed by DAMMAVER^[299] [both available as part of ATSAS] was used for modelling. Positioning of the AgrA N-terminal domain dimer was performed manually using PyMOLTM (vers. 2.2.0) [Schrödinger Inc].^[244] The sample returned good scattering data which exhibited a linear Guinier plot. The radius of gyration determined from the Guinier plot and the pair-distance distribution, P_r , plot agree and are 51.3 and 51.6 respectively and the maximum dimension of the complex, D_{max} , was determined as 176 Å. The scattering data provided readily-interpretable molecular envelopes.



7.6.20: Oxidising and Reducing EMSAs on AgrA_C199S

For the oxidising and reducing EMSAs, reaction and control conditions were set up as per table 7.15.

The 3 x loading/reaction buffer was set up as per table 7.11 below.

Stock solution	Vol, μL
Tris-HCl, pH 7.5, 1 M	20
EDTA, 0.5 M	4
NaCl, 5 M	7.5
Bovine Serum Albumin (BSA), 2 mg mL ⁻¹	10
CaCl ₂ , 100 mM	5
Milli-Q™ H ₂ O	200
Glycerol	80 (~ 105 mg)

Table 7.11. Summary of reagents used to prepare the 3 x loading buffer used for the oxidising and reducing EMSAs. The BSA acts as a crowding agent and results in clearer gels. No dye is used as it may interfere with the reaction.

The 6 μM AgrA_C199S stock solution was prepared as per table 7.12 below.

Stock solution	Vol, μL
AgrA_C199S, 53.1 μM	5.65
DTT/H ₂ O ₂ , 100 mM	1
Tris-HCl, pH 7.5, 100 mM, CaCl ₂ , 10 nM	5
Milli-Q™ H ₂ O	38.35

Table 7.12. Summary of 6 μM AgrA_C199S sample preparation. Two Samples were treated separately with DTT and H₂O₂. It is assumed that the concentration of H₂O₂ overcomes the reducing agents used in the purification procedure.



The cyanine5.5-labelled probe was prepared from the oligonucleotides [Sigma Aldrich] given in table 7.13 overleaf.

Oligonucleotide	Sequence
Forward	5'-[Cyanine5.5]-TTCTTAACTAGTCGTTTTTTATTCTTAACTGTAA-3'
Reverse	5'-TTACGTTAAGAATAAAAAACGACTAGTTAAGAA-3'

Table 7.13. Primers used to generate the cyanine5.5-labelled EMSA probe. Oligonucleotides were annealed by heating an equimolar mixture (200 nM each) of both to 98 °C for 10 mins followed by cooling at 10 °C hr⁻¹ to 4 °C in a PCR machine to achieve a 100 nM stock solution which was stored away from light at -20 °C.

The gel tris/borate/EDTA (TBE buffer) was prepared by dissolving the reagent masses given in table 7.14 below, in dH₂O, to 1 L, to generate a 10 x stock.

Component	Mass, g
Tris base	121.1
Boric acid	61.8
EDTA disodium salt	7.4

Table 7.14. 10 x TBE Buffer recipe.

These components were then used to prepare the 30 µL EMSA reactions presented in table 7.15 overleaf.



Tube No	250 mM AcPO ₄	3 x Buff.	H ₂ O	100 nM DNA	6 μM AgrA	200 mM DTT	200 mM H ₂ O ₂
Vol, μL							
1	6	10	10	3	0	1	0
2	6	10	8.75	3	1.25	1	0
3	6	10	7.50	3	2.50	1	0
4	6	10	6.25	3	3.75	1	0
5	6	10	5.00	3	5.00	1	0
6	6	10	3.75	3	6.25	1	0
7	6	10	2.05	3	7.50	1	0
8	6	10	0	3	10.0	1	0
9	6	10	10	3	0	0	1
10	6	10	8.75	3	1.25	0	1
11	6	10	7.50	3	2.50	0	1
12	6	10	6.25	3	3.75	0	1
13	6	10	5.00	3	5.00	0	1
14	6	10	3.75	3	6.25	0	1
15	6	10	2.05	3	7.50	0	1
16	6	10	0	3	10.0	0	1

Table 7.15. Summary of oxidising and reducing EMSA reaction compositions. AcPO₄ is shorthand for lithium potassium acetyl phosphate [Sigma Aldrich]. The water used was Milli-Q™. The 100 nM DNA is the 5' cyanin5.5-labelled probe. The reagents were added in the following order: 3 x loading/reaction buffer, 6 μM AgrA_C199S, DTT/H₂O₂, AcPO₄, H₂O, P3 DNA probe. After the addition of the AcPO₄ the tubes were gently rocked for 10 mins at room temperature, as they were also after the addition of the P3 DNA.

The EMSA gel was allowed to stand in 0.5 x TBE at 4 °C whilst the reactions were being prepared. Gels were run in 0.5 x TBE at 200 volts for 12 mins, then at 120 volts for 50 mins at 4 °C. Visualisation was performed using a Li-COR® Odyssey® [Li-COR] at an excitation wavelength of 683 nm.



7.6.21: EMSA Gel Preparation

Gels (6 % polyacrylamide) for the oxidising and reducing EMSAs assays were prepared native PAGE-style, i.e. not agarose gels, according to table 7.16 below.

Component	Stock concentration, %	Vol for 2 x gels
Acrylamide:bis-acrylamide (19:1)	30	2 mL
Glycerol	100	250 μ L
5 x TBE	500	1 mL
Milli-Q TM H ₂ O	100	6.65 mL
APS	10	150 μ L
TEMED	100	15 μ L

Table 7.16. Gel recipe for the EMSA assays. APS stands for ammonium persulphate. TEMED stands for N,N,N',N'-tetramethylethylenediamine. Gels were kept wet and cold and used within two days of preparation. No stacking gel was used in order to facilitate entry of the native AgrA_C199S/P3_{full} complex onto the gel. Gels were cast using 0.75 mm thickness using glass sandwich plates and casting equipment [Bio-Rad].

7.6.22: IQS Analogue Docking to AgrA_C Using AutoDock Vina

Solvent of crystallisation and symmetry copies were removed from the receptor molecule AgrA_C (PDB: 4g4k) in PyMOLTM (vers. 2.2.0) [Schrödinger Inc]^[244] to leave only a single copy of the *apo* AgrA protomer. Polar hydrogens were added using MGL Tools^[310] (vers. 1.5.6) [MGL lab, Scripps Institute] and the resulting molecule was exported as a .pdbqt file. IQS analogue ligands were drawn in Chem3D (vers. 16.0) [PerkinElmer] with polar hydrogens added and exported as .mol2 files. The .mol2 files were imported into MGL Tools and the rotatable bonds were specified. The resulting molecules were exported as .pdbqt files.

The search space for the docking experiment was set to cover the entire surface of the receptor molecule, and was specified with the following coordinates and dimensions: center_x = 11.6, center_y



= 3.8 and center_z = 39.9; size_x = 30, size_y = 34 and size_z = 30. The exhaustiveness of the experiment was set to 10.

Docking experiments were performed on each of the fifteen IQS analogue molecules against the receptor AgrA_C molecule using AutoDock Vina^[309] (vers. 1.1.2) [MGL lab, Scripps Institute]. Results were visualised in PyMOL[™]. Schematic diagrams of the ligand binding modes were prepared using LigPlus^[262] (vers. 2.2) [EMBL] and converted from postscript files to .png images using GhostView (vers. 6.0).

7.6.23: *agrP3::lux* Bioreporter Assays and Growth Curves

All experimentation via these assays was performed by Dr. Ewan Murray. Assays were performed identically to the protocols published previously^[302] except that the plate reader used was a Tecan F200 Pro in combination with i-control[™] software [Tecan]. The AgrA gene in the ROJ143 reporter was mutated via site-directed mutagenesis. All primer sequences are available upon request.



7.7.1: AgrB

7.7.2: Construct Generation

AgrB1_(His)₆	
DNA source	Synthetic gene
Expression vector	pET24b(+)
Expression host	C41[De3]
Amino acid sequence of construct	MNYFDNKIDQFATYLQKRNNLDHIQFLQVRLGMQVLAKNIGKLIVMYTIAYILNI FLFTLITNLTFFYLIRRHAHGAHAPSSFWCYVESIILFILLPLVIVNFHINFLIMIILTVIS LGVISVYAPAATKKKPIPVRLIKRRKKYYAIIIVSLTLFIITLIIKEPFAQFIQLGIIIIEAITL LPIFFIKEDLKEHHHHHHH
AgrB1_(His)₁₀	
Amino acid sequence of construct	MNYFDNKIDQFATYLQKRNNLDHIQFLQVRLGMQVLAKNIGKLIVMYTIAYILNI FLFTLITNLTFFYLIRRHAHGAHAPSSFWCYVESIILFILLPLVIVNFHINFLIMIILTVIS LGVISVYAPAATKKKPIPVRLIKRRKKYYAIIIVSLTLFIITLIIKEPFAQFIQLGIIIIEAITL LPIFFIKEDLKEHHHHHHHHHHH
AgrB2_(His)₆	
Amino acid sequence of construct	MNYFDNKIDQFATYLQKRNNLDHIQFLQVRLGMQIIVGNFFKILVTYSISIFLSVFL FTLVTHLSYMLIRYNAHGAHAKSSILCYIQSILTFVFPYFLINIDINFTYLLALSIG LISVVIYAPAATKKQPIPIKLVKRKKYLSIIMYLLVLILSLIHPFYAQFMLLGILVESIT LLPIFFPKEDLEHHHHHHH
AgrB2_(His)₁₀	
Amino acid sequence of construct	MNYFDNKIDQFATYLQKRNNLDHIQFLQVRLGMQIIVGNFFKILVTYSISIFLSVFL FTLVTHLSYMLIRYNAHGAHAKSSILCYIQSILTFVFPYFLINIDINFTYLLALSIG LISVVIYAPAATKKQPIPIKLVKRKKYLSIIMYLLVLILSLIHPFYAQFMLLGILVESIT LLPIFFPKEDLEHHHHHHHHHHH

Table 7.17. Summary of AgrB constructs: Source DNA, expression vector, expression host and amino acid sequences. AgrB genes were synthesized and cloned into the expression vector pET24(b)+ [Genscript].



7.7.3: Expression

E. coli strain C41[DE3] [Lucigen] electrocompetent cells were transformed by electroporation. Typically, purified plasmid DNA, 1 μL , was gently mixed with competent cells, 50 μL . This mixture was transferred to a 2 mm electroporation cuvette and a 1.5 kV electric pulse administered to the cells. Transformed cells were suspended in LB, 1 mL, and allowed to recover for 45 mins at 37 °C. Transformant colonies were selected for by spreading $\sim 125 \mu\text{L}$ of the electroporated culture onto an LB agar plate containing kanamycin [Thermo Fisher], 50 $\mu\text{g mL}^{-1}$, followed by incubation overnight at 37 °C.

Transformant colonies were inoculated into TB containing kanamycin, 50 $\mu\text{g mL}^{-1}$, and incubated overnight at 37 °C with shaking at 180 rpm. 10 mL of this overnight culture was re-inoculated into flasks containing TB, 1 L, and cultured for 16 – 17 hrs overnight at 37°C with shaking at 180 rpm. Overnight culture, 10 mL each, was re-inoculated into flasks containing TB, 1 L, with kanamycin, 50 $\mu\text{g mL}^{-1}$. Flasks were cultured at 37 °C until an $\text{OD}_{600\text{nm}}$ of ~ 0.6 was reached. Overexpression of AgrB2 was induced by the addition of IPTG, 0.4 mM, and allowed to proceed for 16 – 20 hrs at 22 °C with shaking at 200 rpm. Cells were collected by centrifugation at 4600 g for 40 mins at 7 °C and were stored at -20 °C.

7.7.4: Purification

All of the following steps were performed on ice unless otherwise stated. 30 mL of lysis buffer ($\text{K}_2\text{H/KH}_2\text{PO}_4$, 50 mM, pH 7.5 and NaCl, 100 mM) was added to two cell pellets each containing biomass obtained from a 1 L flask of TB, to give a 60 mL total resuspension. Cells were lysed via sonication at 15 μm with 30s on/off pulses for a total “on” time of 8 mins. Cell debris and intact cells were pelleted by centrifugation at 24,000 g for 50 mins at 7 °C. Cell lysate was split equally six ultracentrifuge tubes [Thermo Fisher catalogue number 338455A] and cell membranes pelleted by ultracentrifugation [Hitachi Tabletop Micro catalogue number CS150NX] at 200,000 g for 1 hr at 4 °C. The supernatant was discarded and membranous pellets gently suspended into extraction buffer ($\text{K}_2\text{H/KH}_2\text{PO}_4$, 50 mM, pH 7.5, NaCl, 100 mM and lauryl- β -D-maltoside (DDM), 2 % w/v), 1 mL per



ultracentrifuge tube, by gentle agitation with a bacterial colony picker. Once the membrane pellets were broken into small enough pieces, the end was cut from a 1000 μ L blue pipette tip and each 1 mL extract was aspirated and transferred to clean PyrexTM beaker. Bubbles were carefully removed using a pipette to preserve the integrity of the sample. The entire process hitherto outlined was then repeated owing to the limited capacity (\sim 30 mL) of the ultracentrifuge. The combined 12 mL membrane suspension was gently stirred overnight at 4 °C to allow extraction of AgrB2 into DDM micelles.

The overnight extract was split equally between four ultracentrifuge tubes and depleted membranous material pelleted by ultracentrifugation at 100,000 g for 35 mins at 4 °C. The crude membrane extract was transferred to a 27 mL Econo-ColumnTM [Bio-Rad catalogue number 7374156] containing 2.5 mL of TALON SuperflowTM Co²⁺ Immobilised Metal Affinity Chromatography (IMAC) resin [GE catalogue number 28-9575-99] equilibrated with 10 column volumes of wash buffer (K₂H/KH₂PO₄, 50 mM, pH 7.5, NaCl, 100 mM, imidazole, 50 mM, beta-mercaptoethanol, 4 mM and DDM, 0.1 % w/v). Crude membrane extract was diluted 1:1 with lysis buffer on the column to decrease the concentration of excess DDM micelles, which sterically hinder binding of the sample (His)₆ affinity tag to the column. The TALON SuperflowTM resin was gently suspended into the sample using a 10 mL stripette [Sigma Aldrich] and the column, sealed, set to tumble end-over-end for 3 – 5 hrs at 4 °C to allow DDM_AgrB2 to bind to the Co²⁺ resin.

Weakly-bound proteins were washed away by the addition of 10 column volumes of wash buffer. DDM_AgrB2 was eluted using elution buffer (K₂H/KH₂PO₄, 50 mM, pH 7.5, NaCl, 100 mM, imidazole, 500 mM and DDM, 0.1 % w/v). 5 x 2 mL additions. Dithiothreitol (DTT) was immediately added, to 1 mM, to each of the 5 eluted fractions.

IMAC extracts were further purified by SEC using a preparative grade HiLoad® Superdex® 200 16/600 column [GE] equilibrated with running buffer (K₂H/KH₂PO₄, 50 mM, pH 7.5, NaCl, 50 mM, DTT, 1 mM and DDM, 0.02 % w/v). Pooled sample was concentrated two-fold, diluted two-fold with running buffer, concentrated ten-fold, diluted ten-fold in running buffer to steadily reduce the concentration of imidazole prior to loading the sample onto the injection loop. DDM_AgrB2 showed



a propensity to precipitate out of solution at concentrations $> 2 \text{ mg mL}^{-1}$ in the presence of 500 mM imidazole. Sample was then concentrated to final volume of $\sim 2.5 \text{ mL}$. All concentration steps after SEC were carried out using a VivaspinTM concentrator [Thermo Fisher] with a 100 kDa cut-off at 5000 g at $7 \text{ }^\circ\text{C}$. Concentrated DDM_AgrB2 was centrifuged at $17,000 \text{ g}$ at $7 \text{ }^\circ\text{C}$ for 10 mins prior to loading onto a 5 mL injection loop. SEC was carried out at a flow rate of 0.8 mL min^{-1} . DDM_AgrB2 migrated as an unknown oligomer eluting at $\sim 58 \text{ mL}$ with a broad, Gaussian peak.

In the case of detergent screening experiments, initial extraction of AgrB2 was always into DDM. Whilst DDM_AgrB2 was bound to the IMAC column, a 50 mL wash buffer containing: imidazole, 50 mM , DDM, 0.1 \% w/v , plus above wash buffer components was performed. This was followed by a 50 mL wash in a buffer containing: imidazole, 50 mM , test detergent at $30 \times$ Critical Micelle Concentration (CMC), plus above wash buffer components. Elution was into a buffer containing imidazole, 500 mM , test detergent at $2 \times$ CMC and, if the detergent was non-ionic, NaCl, 50 mM . If the detergent was ionic, 100 mM NaCl was used. This was because of the observed precipitation of AgrB2 when solubilised in fos-choline-12 (FC-12, which is zwitterionic) at 50 mM NaCl. All other buffer components were maintained as described above. The SEC running buffer was identical the elution buffer in each case, only with no imidazole included, and the inclusion of DTT, 1 mM , with the exception of lauryl maltose neopentyl glycol (LMNG), in which no LMNG was present in the SEC running buffer.

7.7.5: Thermofluor Assay

IMAC-purified DDM_AgrB2 in elution buffer (tris, 50 mM , pH 7.0, NaCl, 100 mM , imidazole, 500 mM , TCEP-HCl, 1 mM and DDM, 0.1 \% w/v) was concentrated to 0.35 mg mL^{-1} . $1000 \times$ SYPROTM orange dye [Sigma Aldrich] was diluted one hundred-fold to generate a $10 \times$ stock in H_2O . These components were mixed as per table 7.18, given overleaf, to generate a series of samples and controls.



Sample ID	Vol. 0.35 mg mL ⁻¹ DDM_AgrB2, μ L	Vol. 10 x SYPRO TM , μ L	Vol. buff., μ L
1	30	5	15
2	30	5	15
3	30	5	15
Control ID			
1	0	5	45
2	0	5	45
3	0	5	45

Table 7.18. Summary of reaction and control sample composition for the thermofluor assay.

Assuming that the absorbance of DDM_AgrB2 accounts for ~ 50% of the absorbance exhibited by the crude sample, the samples 1 – 3 have a final DDM_AgrB2 mass of ~ 10 μ g per reaction.

50 μ L Samples were applied to clear PCR tubes and heated through 25 – 90 °C in an Applied Biosystems 7500 quantitative qPCR [Thermo Fisher] machine equipped with a ROX orange filter to select for emissions near the emission maxima of SYPROTM orange at 570 nm.



7.7.6: T7-AgrD Cleavage Assay

T7-AgrD2^{Synthetic}, 0.00021g, was dissolved in dimethylformamide (DMF), 75 μL , 50 mM TCEP-HCl, 30 μL , and H₂O, 645 μL to generate a $\sim 28 \mu\text{g mL}^{-1}$ substrate stock. 250 mg mL⁻¹ of 1,2-dioleoyl-sn-glycero-3-phospho-rac-(1-glycerol) sodium salt (DOPG) [Sigma Aldrich], 500 μL in chloroform, was dried under N₂ (g). Dried powder was dissolved in 250 μL of reaction buffer (tris, 25 mM, pH 7.4, glycerol, 10 % v/v, and MgCl₂, 5 mM) to generate a fresh 500 mg mL⁻¹ stock which was bath sonicated to generate liposomes. This stock was diluted ten-fold into the same buffer to generate a 50 mg mL⁻¹ DOPG liposome stock. Purified DDM_AgrB2 was concentrated to $\sim 50 \mu\text{M}$ in SEC running buffer (see AgrB purification method section 7.7.4) and frozen at -20°C. AgrB2 was thawed on ice and centrifuged at 17,000 g at room temperature for 10 mins to pellet any aggregates. The supernatant protein concentration was checked using a NanodropTM 1000 [Thermo Fisher] using a theoretical extension coefficient = 17,880 M⁻¹ cm⁻¹. No aggregation was observed. The four components, T7-AgrD2^{Synthetic}, DOPG, DDM_AgrB2 and buffer were combined as per table 7.19 into PCR tubes to generate four reactions: two experimental reactions with DDM_AgrB2 at both 2.5 and 5.0 μM and two negative controls, one lacking DOPG liposomes and one lacking DDM_AgrB2.

Reaction tubes were incubated at 37 °C with shaking at 60 rpm for 3 hrs. 30 μL of each reaction tube was added to 2 x Laemmli buffer [Novagen], 30 μL . Three reference markers: T7-AgrD2^{Synthetic}, T7-AgrD1^{Synthetic} and $\Delta\text{C-terminus-T7-AgrD1}^{\text{synthetic}}$ (positive control), were generated by diluting 3 μL of each substrate into H₂O, 27 μL , and adding to 2 x Laemmli buffer, 30 μL . β -mercaptoethanol, 6 %, was added to experimental and control tricine-SDS-PAGE samples immediately prior to gel electrophoresis. A 16 % tris-tricine gel was used for electrophoresis at 40 V for 20 mins and subsequently, 140 V for 100 mins. Proteins were transferred to an immobilon® PSQ membrane [Sigma Aldrich] at 90 A for 90 mins. This membrane was blocked overnight with blocking reagent [Qiagen], 0.2 g, in 1 x blocking reagent buffer [Qiagen], 1 mL, with tween 20, 2 % v/v, at 4 °C. The membrane was washed three times with TBSt buffer (NaCl, 137 mM, KCl, 2.7 mM, tris, 19 mM, pH 7.4 and tween 20, 0.1 % v/v). After washing, the membrane was incubated in TBSt, 20 mL, with



casein, 0.2 g, [Sigma Aldrich] and anti-T7-HRP conjugate antibody [Novagen catalogue number 69048], 1.2 μL . The antibody was allowed to bind at room temperature for 3 hrs with shaking at 50 rpm. The membrane was rinsed with 6 additions of TBSt, 10 mL, for 8 mins each. Rinsed membrane was wetted with ECL prime [GE] with mixing for 5 mins before being rinsed with H_2O and transferred into cling film. Air and residual ECL prime was thoroughly forced out by hand to avoid background chemiluminescence. Antibody chemiluminescence was visualized by 5, 10 and 20 minute exposures of the membrane to photographic film in dark room, which was then developed and visualised using standard latent image development procedures.

Volume of component in each reaction, μL				
Component	Reaction 1	Reaction 2	(-) DOPG	(-) AgrB2
Reaction buffer	81	76	85	86
50 μM DDM_AgrB2	5	10	5	0
28 $\mu\text{g mL}^{-1}$ T7-AgrD2 ^{synthetic}	10	10	10	10
50 mg mL^{-1} DOPG	4	4	0	4
Total volume, μL	100	100	100	100

Table 7.19. Summary of T7-AgrD cleavage assay reaction composition.

Electrophoresis and gel buffers were prepared as per table 7.20. 1L of each of the 3 x electrophoresis buffers was prepared whilst the 3 x gel buffer was prepared in 50 mL.

Component and amount	10 x Anode buffer	10 x Cathode buffer	3 x Gel buffer
Tris, M	1.0	1.0	3.0
Tricine, M	0	1.0	0
HCl, M	0.225	0	1.0
SDS, % w/v	0	1.0	0.3
Final pH	8.9	8.25	8.5

Table 7.20. Summary of buffer recipes used to perform the T7-AgrD cleavage assay.

16 % tris-tricine gels were prepared as per the Nature Protocols method^[357] and ran as described above before western blotting.



7.7.7: Circular Dichroism Measurements on DDM_AgrB2

Circular dichroism (CD) data were collected with the help of Dr. Alex Slater [University of Birmingham] using a ChirscanPlus spectrophotometer [Applied Photophysics] purged of O₂. DDM_AgrB2 in SEC running buffer was spin-concentrated to 5.0 mg mL⁻¹ and applied to a clean 1 mm quartz cuvette [Applied Photophysics]. A single CD spectrum was obtained for the sample at 4 °C. As this spectrum contained unacceptable levels of noise, the sample was diluted four-fold to 1.25 mg mL⁻¹ in SEC running buffer and another CD spectrum collected. Since this concentration provided a less noisy spectrum, it was used for a variable temperature CD data collection to quantify the thermostability of DDM_AgrB2 in SEC running buffer. The sample was steadily heated from 4 – 90 °C and CD spectra collected periodically at accurately-known temperatures.

7.7.8: Negative Staining

Four drops of uranyl acetate stain, 5 µL each, were pipetted in a row along a strip of parafilm. The hydrophobic nature of parafilm forced each drop to tighten up into a little ball. Meanwhile, an Electron Micrograph (EM) grid was glow discharged (that is placed in a low pressure vestibule and exposed to an electrical current, to produce a layer of negatively charged, water-attracting ions on the grid surface). Glow discharging took several minutes, after which the grid was carefully gripped at one edge with a pair of tweezers, which was subsequently locked closed using a small clamp, to make the grid easier to handle. 0.5 mg mL⁻¹ LMNG_AgrB2, 5 µL, was centrifuged at 17,000 g for 10 mins at room temperature and pipetted directly onto each grid and was allowed to “settle” for 20 – 30 secs. The bolus of liquid on the grid was subsequently dotted against tissue paper which immediately wicked all of the excess liquid away. The grid was washed with H₂O, 5 µL, to prevent formation of uranyl phosphate crystals which can occur when uranyl acetate stain comes into contact with a phosphate buffer system. After this washing step, the grid was dotted against each of the four uranyl acetate drops in turn with each drop wicked before exposure to the next drop. Each drop immediately adhered to the grid and each exposure to uranyl acetate was for ~ 20 – 30 secs total. After the final wicking away of uranyl acetate, the grid was ready for insertion into the microscope. Grids were

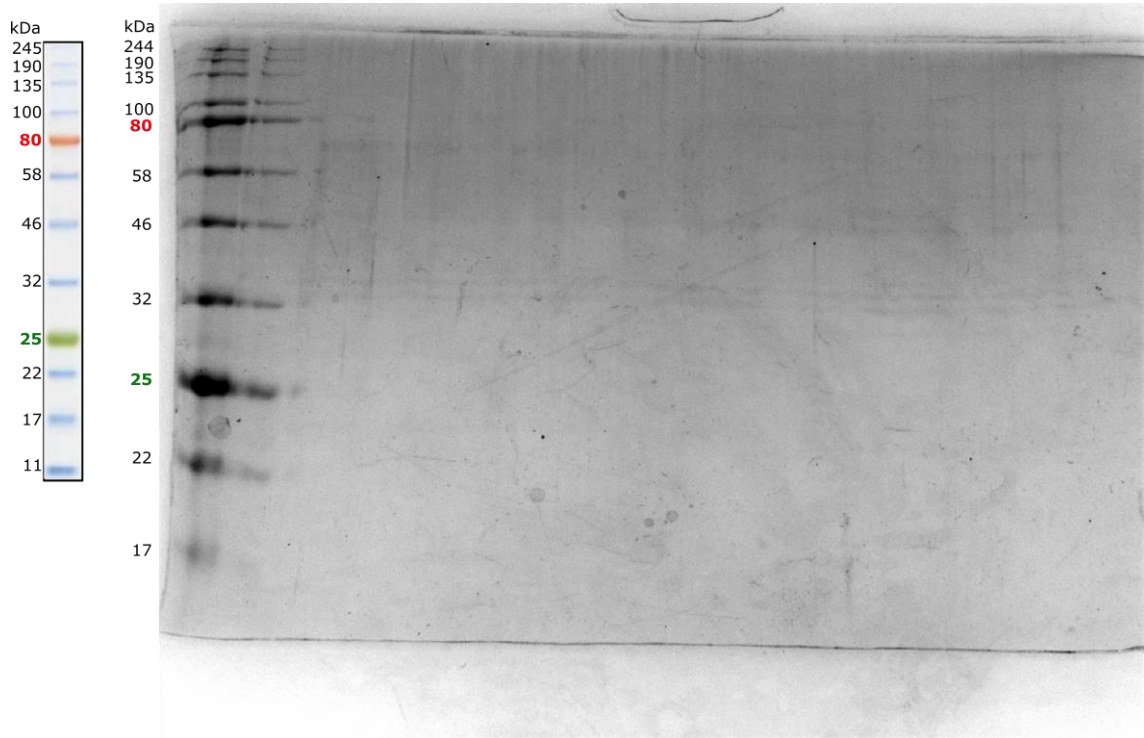


visualised on a 200 kV JEM-2100Plus Transmission Electron Microscopy (TEM) microscope [JEOL] fitted with an OneView camera [Gatan] by Dr. Saskia Bakker [University of Warwick]. All reagents quoted were those used by the microscopy lab at the University of Warwick. Raw micrographs were converted into .png format using the Fiji software package.^[358]



Appendices

8.1: Appendix I – SDS-PAGE Gel Marker

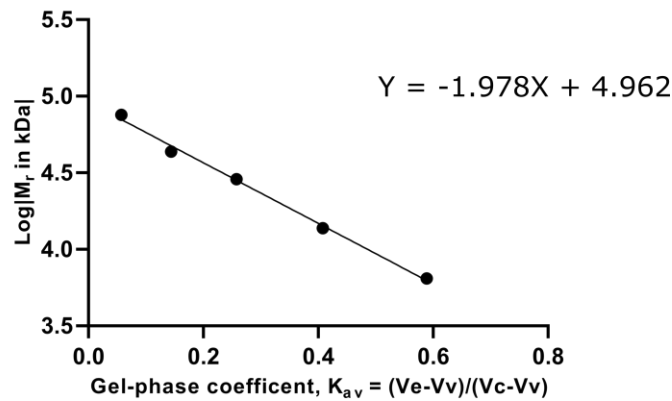


The same SDS-PAGE marker, Color Prestained Protein Standard, Broad Range (11 – 245 kDa) [NEB catalogue number P7712], was used throughout the entirety of this thesis. The graphical key as provided by the manufacturer is given above along with a 4 μ L of this marker loaded onto a blank 16 % polyacrylamide gel.

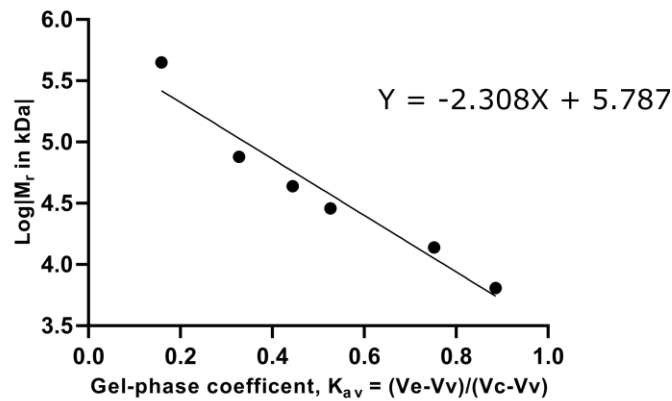


8.2: Appendix II – Column Calibration Curves

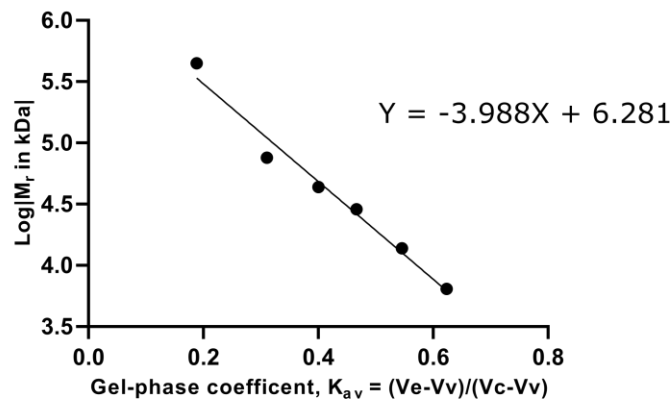
Hi-Load Superdex 75 16/600



Hi-Load Superdex 200 16/600



Superdex 200 Increase 10/300 GL



V_e = Analyte retention volume, V_v = column void volume, V_c = Stationary phase/accessible volume. Column void volumes for the Hi-Load® columns are ~ 45 mL, and for the Increase column ~ 13 mL. The accessible volumes are 120 and 24 mL respectively. Calibration standards used were: ferritin (not ran on Increase column), conalbumin, ovalbumin, carbonic anhydrase, ribonuclease A and aprotinin.



8.3: Appendix III – Preliminary PDB Validation Report for AgrA_C199S/P2_{fragment} Structure



Preliminary Full wwPDB X-ray Structure Validation Report ⓘ

Mar 2, 2020 – 10:07 AM EST

This is a Preliminary Full wwPDB X-ray Structure Validation Report.

This report is produced by the standalone wwPDB validation server.
The structure in question has not been deposited to the wwPDB.
This report should not be submitted to journals.

We welcome your comments at validation@mail.wwpdb.org

A user guide is available at

<https://www.wwpdb.org/validation/2017/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Xtrriage (Phenix)	:	1.13
EDS	:	2.9
Percentile statistics	:	20171227.v01 (using entries in the PDB archive December 27th 2017)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.9



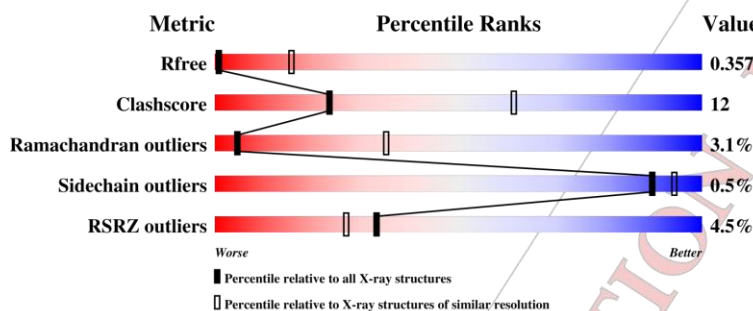
1 Overall quality at a glance [i](#)

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 4.76 Å.

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



Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	111664	1100 (5.82-3.70)
Clashscore	122126	1026 (5.76-3.76)
Ramachandran outliers	120053	1114 (5.82-3.70)
Sidechain outliers	120020	1095 (5.82-3.70)
RSRZ outliers	108989	1025 (5.90-3.62)

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

Mol	Chain	Length	Quality of chain
1	A	233	
2	B	16	
3	C	16	



2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 2575 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	233	1925	1221	327	366	11	0	1	0

- Molecule 2 is a DNA chain called DNA (5'-D(*TP*TP*TP*AP*AP*CP*AP*GP*TP*TP*AP*AP*GP*TP*AP*T)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	B	16	326	159	57	95	15	0	0	0

- Molecule 3 is a DNA chain called DNA (5'-D(*AP*AP*TP*AP*CP*TP*TP*AP*AP*CP*TP*GP*TP*TP*AP*A)-3').

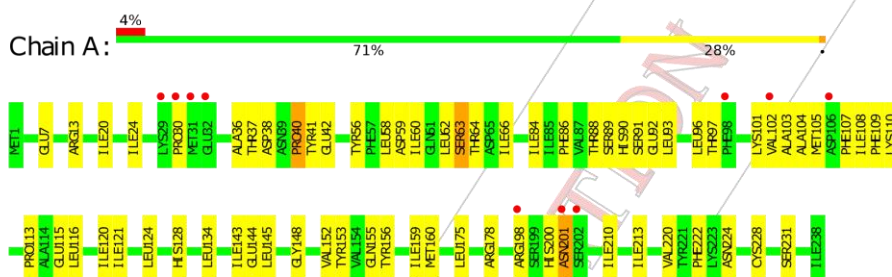
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
3	C	16	324	158	58	93	15	0	0	0



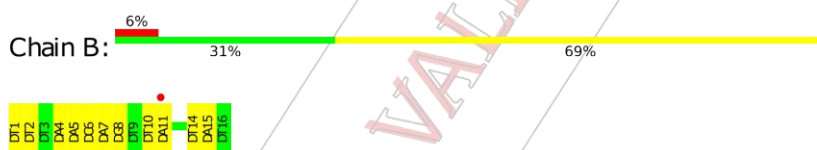
3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density ($RSRZ > 2$). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

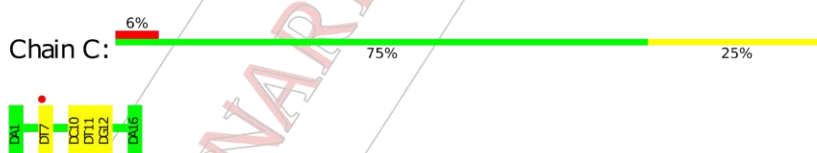
- Molecule 1:



- Molecule 2: DNA (5'-D(*TP*TP*TP*AP*AP*CP*AP*GP*TP*TP*AP*AP*GP*TP*AP*AT)-3')



- Molecule 3: DNA (5'-D(*AP*AP*TP*AP*CP*TP*TP*AP*AP*CP*TP*GP*TP*TP*AP*A)-3')





4 Data and refinement statistics

Property	Value	Source
Space group	I 2 2 2	Depositor
Cell constants a, b, c, α , β , γ	61.11Å 96.95Å 195.70Å 90.00 90.00 90.00	Depositor
Resolution (Å)	54.12 - 4.76 54.12 - 4.76	Depositor EDS
% Data completeness (in resolution range)	75.1 (54.12-4.76) 75.1 (54.12-4.76)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.39 (at 4.86Å)	Xtrige
Refinement program	phenix.refine 1.17.1_3660, PHENIX 1.17.1_3660	Depositor
R , R_{free}	0.297 , 0.361 0.297 , 0.357	Depositor DCC
R_{free} test set	136 reflections (5.77%)	wwPDB-VP
Wilson B-factor (Å ²)	134.9	Xtrige
Anisotropy	0.239	Xtrige
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.30 , 114.1	EDS
L-test for twinning ²	$\langle L \rangle = 0.45$, $\langle L^2 \rangle = 0.28$	Xtrige
Estimated twinning fraction	No twinning to report.	Xtrige
F_o, F_c correlation	0.80	EDS
Total number of atoms	2575	wwPDB-VP
Average B, all atoms (Å ²)	155.0	wwPDB-VP

Xtrige's analysis on translational NCS is as follows: The largest of -origin peak in the Patterson function is 5.98% of the height of the origin peak. No significant pseudotranslation is detected.

¹ Intensities estimated from amplitudes.

² Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.



5 Model quality [oi](#)

5.1 Standard geometry [oi](#)

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	# $ Z > 5$	RMSZ	# $ Z > 5$
1	A	0.24	0/1962	0.44	0/2639
2	B	0.57	0/365	1.04	0/562
3	C	0.57	0/363	0.97	0/558
All	All	0.36	0/2690	0.66	0/3759

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts [oi](#)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	1925	0	1915	47	0
2	B	326	0	185	8	0
3	C	324	0	184	5	0
All	All	2575	0	2284	59	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 12.

All (59) close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:96:LEU:HD13	1:A:97:THR:N	1.75	1.00

Continued on next page...



Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:96:LEU:HD13	1:A:97:THR:H	1.37	0.85
3:C:11:DT:H2'	3:C:12:DG:C8	2.16	0.81
1:A:37:THR:HG22	1:A:38:ASP:H	1.49	0.77
2:B:11:DA:H61	3:C:7:DT:H3	1.33	0.75
1:A:175:LEU:HB2	1:A:178:ARG:HB2	1.71	0.71
2:B:6:DC:H2'	2:B:7:DA:C8	2.29	0.68
1:A:160:MET:SD	1:A:224:ASN:ND2	2.68	0.67
1:A:103:ALA:O	1:A:105:MET:N	2.26	0.66
2:B:14:DT:H2''	2:B:15:DA:H5'	1.77	0.66
1:A:93:LEU:HD12	1:A:96:LEU:HB3	1.80	0.61
1:A:90:HIS:O	1:A:92:GLU:N	2.35	0.59
1:A:88:THR:O	1:A:110:LYS:NZ	2.26	0.57
1:A:13:ARG:HH11	1:A:36:ALA:HB1	1.71	0.55
1:A:210:ILE:HG12	1:A:222:PHE:HE1	1.72	0.53
1:A:200:HIS:CD2	1:A:231:SER:HB3	2.43	0.53
2:B:1:DT:C6	2:B:2:DT:H72	2.44	0.52
1:A:145:LEU:HB2	1:A:152:VAL:HG13	1.91	0.51
1:A:40:PRO:O	1:A:42:GLU:N	2.35	0.51
3:C:10:DC:H2'	3:C:11:DT:C6	2.46	0.51
1:A:108:ILE:HG23	1:A:116:LEU:HD22	1.93	0.49
2:B:10:DT:H2''	2:B:11:DA:N7	2.27	0.49
1:A:59:ASP:HB3	1:A:62:LEU:HD21	1.95	0.49
1:A:96:LEU:CD1	1:A:97:THR:N	2.64	0.48
3:C:11:DT:H2'	3:C:12:DG:H8	1.67	0.48
1:A:7:GLU:H	1:A:13:ARG:NH2	2.11	0.48
1:A:37:THR:HG22	1:A:38:ASP:N	2.23	0.48
1:A:93:LEU:CD1	1:A:96:LEU:HD12	2.43	0.48
2:B:7:DA:H2''	2:B:8:DG:H8	1.79	0.48
1:A:93:LEU:HD11	1:A:96:LEU:HD12	1.94	0.47
1:A:104:ALA:HB1	1:A:107:PHE:CE1	2.50	0.47
1:A:121:ILE:HA	1:A:124:LEU:HD12	1.98	0.45
1:A:143:ILE:HB	1:A:156:TYR:HE1	1.82	0.45
1:A:63:SER:OG	1:A:64:THR:N	2.49	0.45
3:C:10:DC:H3'	3:C:11:DT:H71	1.98	0.45
1:A:160:MET:HB3	1:A:222:PHE:CD2	2.52	0.45
1:A:201:ASN:ND2	1:A:201:ASN:O	2.48	0.45
1:A:58:LEU:HD12	1:A:86:PHE:HE1	1.82	0.44
1:A:62:LEU:HD13	1:A:66:ILE:HD11	1.99	0.44
1:A:89:SER:O	1:A:110:LYS:HB2	2.17	0.44
1:A:101:LYS:O	1:A:103:ALA:N	2.50	0.43
1:A:213:ILE:HG12	1:A:220:VAL:HG22	2.00	0.43

Continued on next page...



Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:124:LEU:O	1:A:128:HIS:ND1	2.41	0.43
1:A:7:GLU:O	1:A:13:ARG:NH2	2.52	0.43
2:B:4:DA:H2''	2:B:5:DA:C8	2.53	0.43
1:A:198:ARG:HH21	1:A:201:ASN:HD22	1.67	0.43
2:B:5:DA:C4	2:B:6:DC:C5	3.07	0.43
1:A:220:VAL:HB	1:A:228:CYS:SG	2.60	0.42
1:A:113:PRO:O	1:A:115:GLU:N	2.44	0.42
1:A:144:GLU:HG3	1:A:153:TYR:CE1	2.54	0.42
1:A:20:ILE:O	1:A:24:ILE:HG12	2.20	0.42
1:A:104:ALA:HB1	1:A:107:PHE:CZ	2.55	0.42
1:A:60:ILE:HD12	1:A:88:THR:HA	2.03	0.41
1:A:96:LEU:HG	1:A:109:PHE:CZ	2.55	0.41
1:A:134:LEU:H	1:A:134:LEU:HD23	1.85	0.41
1:A:86:PHE:HB2	1:A:107:PHE:CD1	2.56	0.40
1:A:155:GLN:O	1:A:159:ILE:HG13	2.22	0.40
1:A:56:TYR:HB2	1:A:84:ILE:HG23	2.04	0.40
1:A:120:ILE:O	1:A:124:LEU:HG	2.22	0.40

There are no symmetry-related clashes.

5.3 Torsion angles [i](#)

5.3.1 Protein backbone [i](#)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
1	A	230/233 (99%)	191 (83%)	32 (14%)	7 (3%)	5 35

All (7) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	63	SER
1	A	91	SER
1	A	148	GLY
1	A	41	TYR

Continued on next page...



Continued from previous page...

Mol	Chain	Res	Type
1	A	40	PRO
1	A	102	VAL
1	A	30	PRO

5.3.2 Protein sidechains [i](#)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	A	219/218 (100%)	218 (100%)	1 (0%)	90 94

All (1) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	201	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

5.6 Ligand geometry [i](#)

There are no ligands in this entry.



5.7 Other polymers [i](#)

There are no such residues in this entry.

5.8 Polymer linkage issues [i](#)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
1	A	1

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	A	135:SER	C	141:GLU	N	12.17

PRELIMINARY VALIDATION REPORT



6 Fit of model and data [oi](#)

6.1 Protein, DNA and RNA chains [oi](#)

In the following table, the column labelled '#RSRZ > 2' contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95th percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled 'Q < 0.9' lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ > 2	OWAB (Å ²)	Q < 0.9
1	A	233/233 (100%)	0.13	10 (4%) 35 28	78, 147, 205, 253	0
2	B	16/16 (100%)	0.75	1 (6%) 20 16	158, 166, 204, 220	0
3	C	16/16 (100%)	0.72	1 (6%) 20 16	147, 174, 197, 208	0
All	All	265/265 (100%)	0.20	12 (4%) 33 27	78, 153, 205, 253	0

All (12) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	30	PRO	7.3
1	A	31	MET	5.7
1	A	32	GLU	4.5
1	A	201	ASN	4.2
2	B	11	DA	2.6
1	A	202	SER	2.2
3	C	7	DT	2.2
1	A	106	ASP	2.2
1	A	98	PHE	2.1
1	A	198	ARG	2.1
1	A	102	VAL	2.1
1	A	29	LYS	2.1

6.2 Non-standard residues in protein, DNA, RNA chains [oi](#)

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates [oi](#)

There are no carbohydrates in this entry.



6.4 Ligands [oi](#)

There are no ligands in this entry.

6.5 Other polymers [oi](#)

There are no such residues in this entry.

PRELIMINARY VALIDATION REPORT



8.4: Appendix IV – IQS Analogue Docking Log Files. IQS1

```
#####  
# If you used AutoDock Vina in your work, please cite:      #  
#                                                           #  
# O. Trott, A. J. Olson,                                    #  
# AutoDock Vina: improving the speed and accuracy of docking #  
# with a new scoring function, efficient optimization and    #  
# multithreading, Journal of Computational Chemistry 31 (2010) #  
# 455-461                                                    #  
#                                                           #  
# DOI 10.1002/jcc.21334                                     #  
#                                                           #  
# Please see http://vina.scripps.edu for more information. #  
#####
```

```
WARNING: The search space volume > 27000 Angstrom^3 (See FAQ)  
Output will be IQS1_out.pdbqt  
Detected 6 CPUs  
Reading input ... done.  
Setting up the scoring function ... done.  
Analyzing the binding site ... done.  
Using random seed: -1137268088  
Performing search ... done.  
Refining results ... done.
```

```
mode |  affinity | dist from best mode  
      | (kcal/mol) | rmsd l.b.| rmsd u.b.  
-----+-----+-----+-----  
   1 |    -5.1 |    0.000 |    0.000  
   2 |    -5.0 |    4.598 |    5.430  
   3 |    -5.0 |   11.916 |   14.057  
   4 |    -4.9 |    1.974 |    2.668  
   5 |    -4.8 |    1.208 |    1.370  
   6 |    -4.8 |   19.498 |   20.420  
   7 |    -4.8 |    1.801 |    5.452  
   8 |    -4.7 |    3.082 |    3.426  
   9 |    -4.7 |    3.301 |    4.668
```

```
Writing output ... done.
```



8.4: Appendix IV – IQS Analogue Docking Log Files. IQS2

```
#####  
# If you used AutoDock Vina in your work, please cite:      #  
#                                                           #  
# O. Trott, A. J. Olson,                                    #  
# AutoDock Vina: improving the speed and accuracy of docking #  
# with a new scoring function, efficient optimization and    #  
# multithreading, Journal of Computational Chemistry 31 (2010) #  
# 455-461                                                    #  
#                                                           #  
# DOI 10.1002/jcc.21334                                     #  
#                                                           #  
# Please see http://vina.scripps.edu for more information. #  
#####
```

```
WARNING: The search space volume > 27000 Angstrom^3 (See FAQ)  
Output will be IQS2_out.pdbqt  
Detected 6 CPUs  
Reading input ... done.  
Setting up the scoring function ... done.  
Analyzing the binding site ... done.  
Using random seed: 1509174032  
Performing search ... done.  
Refining results ... done.
```

```
mode |  affinity | dist from best mode  
      | (kcal/mol) | rmsd l.b. | rmsd u.b.  
-----+-----+-----+-----  
  1      -5.4      0.000      0.000  
  2      -5.4     16.959     18.816  
  3      -5.4     17.531     19.939  
  4      -5.3     18.607     19.505  
  5      -5.3     12.628     15.111  
  6      -5.2     17.260     18.378  
  7      -5.2     15.093     17.226  
  8      -5.1     18.838     20.337  
  9      -5.1     16.584     18.507
```

```
Writing output ... done.
```



8.4: Appendix IV – IQS Analogue Docking Log Files. IQS3

```
#####  
# If you used AutoDock Vina in your work, please cite:      #  
#                                                           #  
# O. Trott, A. J. Olson,                                    #  
# AutoDock Vina: improving the speed and accuracy of docking #  
# with a new scoring function, efficient optimization and    #  
# multithreading, Journal of Computational Chemistry 31 (2010) #  
# 455-461                                                    #  
#                                                           #  
# DOI 10.1002/jcc.21334                                     #  
#                                                           #  
# Please see http://vina.scripps.edu for more information. #  
#####
```

```
WARNING: The search space volume > 27000 Angstrom^3 (See FAQ)  
Output will be IQS3_out.pdbqt  
Detected 6 CPUs  
Reading input ... done.  
Setting up the scoring function ... done.  
Analyzing the binding site ... done.  
Using random seed: 139858052  
Performing search ... done.  
Refining results ... done.
```

mode	affinity	dist from best mode	
	(kcal/mol)	rmsd l.b.	rmsd u.b.
1	-5.3	0.000	0.000
2	-5.3	4.363	5.805
3	-5.3	16.653	18.500
4	-5.2	17.195	18.310
5	-5.2	18.601	20.514
6	-5.1	14.830	16.856
7	-5.1	3.446	7.909
8	-5.1	16.303	18.386
9	-5.1	13.991	16.477

```
Writing output ... done.
```



8.4: Appendix IV – IQS Analogue Docking Log Files. IQS4

```
#####  
# If you used AutoDock Vina in your work, please cite:      #  
#                                                            #  
# O. Trott, A. J. Olson,                                    #  
# AutoDock Vina: improving the speed and accuracy of docking #  
# with a new scoring function, efficient optimization and    #  
# multithreading, Journal of Computational Chemistry 31 (2010) #  
# 455-461                                                    #  
#                                                            #  
# DOI 10.1002/jcc.21334                                     #  
#                                                            #  
# Please see http://vina.scripps.edu for more information. #  
#####
```

```
WARNING: The search space volume > 27000 Angstrom^3 (See FAQ)  
Output will be IQS4_out.pdbqt  
Detected 6 CPUs  
Reading input ... done.  
Setting up the scoring function ... done.  
Analyzing the binding site ... done.  
Using random seed: -1745975616  
Performing search ... done.  
Refining results ... done.
```

```
mode |  affinity | dist from best mode  
      | (kcal/mol) | rmsd l.b. | rmsd u.b.  
-----+-----+-----+-----  
  1      -4.7      0.000      0.000  
  2      -4.7      4.357      5.386  
  3      -4.6     18.707     19.689  
  4      -4.6     11.076     13.360  
  5      -4.5      2.184      5.622  
  6      -4.5      3.994      4.512  
  7      -4.4      1.674      1.810  
  8      -4.4     19.557     21.622  
  9      -4.4     20.685     21.363
```

```
Writing output ... done.
```




8.4: Appendix IV – IQS Analogue Docking Log Files. IQS5

```
#####  
# If you used AutoDock Vina in your work, please cite:      #  
#                                                           #  
# O. Trott, A. J. Olson,                                    #  
# AutoDock Vina: improving the speed and accuracy of docking #  
# with a new scoring function, efficient optimization and    #  
# multithreading, Journal of Computational Chemistry 31 (2010) #  
# 455-461                                                    #  
#                                                           #  
# DOI 10.1002/jcc.21334                                     #  
#                                                           #  
# Please see http://vina.scripps.edu for more information. #  
#####
```

```
WARNING: The search space volume > 27000 Angstrom^3 (See FAQ)  
Output will be IQS5_out.pdbqt  
Detected 6 CPUs  
Reading input ... done.  
Setting up the scoring function ... done.  
Analyzing the binding site ... done.  
Using random seed: 692666144  
Performing search ... done.  
Refining results ... done.
```

mode	affinity	dist from best mode	
	(kcal/mol)	rmsd l.b.	rmsd u.b.
1	-4.7	0.000	0.000
2	-4.6	1.372	1.401
3	-4.6	18.636	19.700
4	-4.5	19.663	20.875
5	-4.5	4.645	5.433
6	-4.4	15.922	17.128
7	-4.3	2.791	3.073
8	-4.3	2.070	5.364
9	-4.3	3.315	5.352

```
Writing output ... done.
```



8.4: Appendix IV – IQS Analogue Docking Log Files. IQS6

```
#####  
# If you used AutoDock Vina in your work, please cite:      #  
#                                                            #  
# O. Trott, A. J. Olson,                                    #  
# AutoDock Vina: improving the speed and accuracy of docking #  
# with a new scoring function, efficient optimization and    #  
# multithreading, Journal of Computational Chemistry 31 (2010) #  
# 455-461                                                    #  
#                                                            #  
# DOI 10.1002/jcc.21334                                     #  
#                                                            #  
# Please see http://vina.scripps.edu for more information. #  
#####
```

```
WARNING: The search space volume > 27000 Angstrom^3 (See FAQ)  
Output will be IQS6_out.pdbqt  
Detected 6 CPUs  
Reading input ... done.  
Setting up the scoring function ... done.  
Analyzing the binding site ... done.  
Using random seed: -1585091664  
Performing search ... done.  
Refining results ... done.
```

```
mode |  affinity | dist from best mode  
      | (kcal/mol) | rmsd l.b. | rmsd u.b.  
-----+-----+-----+-----  
  1      -4.9      0.000      0.000  
  2      -4.8      3.745      6.046  
  3      -4.8      3.494      4.886  
  4      -4.8      4.023      5.151  
  5      -4.7      1.826      2.040  
  6      -4.6      3.547      4.646  
  7      -4.6      1.795      2.301  
  8      -4.6      2.598      3.215  
  9      -4.5      4.487      7.410
```

```
Writing output ... done.
```



8.4: Appendix IV – IQS Analogue Docking Log Files. IQS7

```
#####  
# If you used AutoDock Vina in your work, please cite:      #  
#                                                           #  
# O. Trott, A. J. Olson,                                    #  
# AutoDock Vina: improving the speed and accuracy of docking #  
# with a new scoring function, efficient optimization and    #  
# multithreading, Journal of Computational Chemistry 31 (2010) #  
# 455-461                                                    #  
#                                                           #  
# DOI 10.1002/jcc.21334                                     #  
#                                                           #  
# Please see http://vina.scripps.edu for more information. #  
#####
```

```
WARNING: The search space volume > 27000 Angstrom^3 (See FAQ)  
Output will be IQS7_out.pdbqt  
Detected 6 CPUs  
Reading input ... done.  
Setting up the scoring function ... done.  
Analyzing the binding site ... done.  
Using random seed: 213548640  
Performing search ... done.  
Refining results ... done.
```

mode	affinity	dist from best mode	
	(kcal/mol)	rmsd l.b.	rmsd u.b.
1	-4.9	0.000	0.000
2	-4.8	5.531	6.924
3	-4.8	7.793	9.756
4	-4.7	11.234	14.726
5	-4.7	7.078	8.519
6	-4.7	13.196	14.080
7	-4.6	7.402	9.040
8	-4.6	10.279	11.373
9	-4.6	4.708	7.375

```
Writing output ... done.
```



8.4: Appendix IV – IQS Analogue Docking Log Files. IQS8

```
#####  
# If you used AutoDock Vina in your work, please cite:      #  
#                                                           #  
# O. Trott, A. J. Olson,                                    #  
# AutoDock Vina: improving the speed and accuracy of docking #  
# with a new scoring function, efficient optimization and    #  
# multithreading, Journal of Computational Chemistry 31 (2010) #  
# 455-461                                                    #  
#                                                           #  
# DOI 10.1002/jcc.21334                                     #  
#                                                           #  
# Please see http://vina.scripps.edu for more information. #  
#####
```

```
WARNING: The search space volume > 27000 Angstrom^3 (See FAQ)  
Output will be IQS8_out.pdbqt  
Detected 6 CPUs  
Reading input ... done.  
Setting up the scoring function ... done.  
Analyzing the binding site ... done.  
Using random seed: 1039072416  
Performing search ... done.  
Refining results ... done.
```

```
mode |  affinity | dist from best mode  
      | (kcal/mol) | rmsd l.b. | rmsd u.b.  
-----+-----+-----+-----  
  1      -4.9      0.000      0.000  
  2      -4.8      1.451      6.316  
  3      -4.8     17.373     18.424  
  4      -4.7      3.580      5.087  
  5      -4.7      4.029      5.662  
  6      -4.7      2.292      6.521  
  7      -4.5      2.145      2.442  
  8      -4.5      4.115      6.766  
  9      -4.5     16.638     18.258
```

```
Writing output ... done.
```



8.4: Appendix IV – IQS Analogue Docking Log Files. IQS9

```
#####  
# If you used AutoDock Vina in your work, please cite:      #  
#                                                           #  
# O. Trott, A. J. Olson,                                    #  
# AutoDock Vina: improving the speed and accuracy of docking #  
# with a new scoring function, efficient optimization and    #  
# multithreading, Journal of Computational Chemistry 31 (2010) #  
# 455-461                                                    #  
#                                                           #  
# DOI 10.1002/jcc.21334                                     #  
#                                                           #  
# Please see http://vina.scripps.edu for more information. #  
#####
```

```
WARNING: The search space volume > 27000 Angstrom^3 (See FAQ)  
Output will be IQS9_out.pdbqt  
Detected 6 CPUs  
Reading input ... done.  
Setting up the scoring function ... done.  
Analyzing the binding site ... done.  
Using random seed: -473163404  
Performing search ... done.  
Refining results ... done.
```

mode	affinity	dist from best mode	
	(kcal/mol)	rmsd l.b.	rmsd u.b.
1	-4.8	0.000	0.000
2	-4.7	17.394	18.432
3	-4.6	0.994	1.744
4	-4.3	15.638	17.711
5	-4.3	2.976	4.851
6	-4.2	4.997	8.616
7	-4.2	12.968	15.220
8	-4.2	17.421	18.959
9	-4.1	17.325	19.043

```
Writing output ... done.
```



8.4: Appendix IV – IQS Analogue Docking Log Files. IQS10

```
#####  
# If you used AutoDock Vina in your work, please cite:      #  
#                                                            #  
# O. Trott, A. J. Olson,                                     #  
# AutoDock Vina: improving the speed and accuracy of docking #  
# with a new scoring function, efficient optimization and    #  
# multithreading, Journal of Computational Chemistry 31 (2010) #  
# 455-461                                                    #  
#                                                            #  
# DOI 10.1002/jcc.21334                                     #  
#                                                            #  
# Please see http://vina.scripps.edu for more information. #  
#####
```

```
WARNING: The search space volume > 27000 Angstrom^3 (See FAQ)  
Output will be IQS10_out.pdbqt  
Detected 6 CPUs  
Reading input ... done.  
Setting up the scoring function ... done.  
Analyzing the binding site ... done.  
Using random seed: -1983675136  
Performing search ... done.  
Refining results ... done.
```

```
mode |  affinity | dist from best mode  
      | (kcal/mol) | rmsd l.b. | rmsd u.b.  
-----+-----+-----+-----  
  1      -5.0      0.000      0.000  
  2      -5.0     11.270     13.429  
  3      -5.0      1.836      5.539  
  4      -4.9     19.526     20.087  
  5      -4.9      2.701      3.777  
  6      -4.8      4.065      5.069  
  7      -4.8      2.621      5.521  
  8      -4.8      2.589      5.949  
  9      -4.8     17.749     18.550
```

```
Writing output ... done.
```



8.4: Appendix IV – IQS Analogue Docking Log Files. IQS11

```
#####  
# If you used AutoDock Vina in your work, please cite:      #  
#                                                           #  
# O. Trott, A. J. Olson,                                    #  
# AutoDock Vina: improving the speed and accuracy of docking #  
# with a new scoring function, efficient optimization and    #  
# multithreading, Journal of Computational Chemistry 31 (2010) #  
# 455-461                                                    #  
#                                                           #  
# DOI 10.1002/jcc.21334                                     #  
#                                                           #  
# Please see http://vina.scripps.edu for more information. #  
#####
```

```
WARNING: The search space volume > 27000 Angstrom^3 (See FAQ)  
Output will be IQS11_out.pdbqt  
Detected 6 CPUs  
Reading input ... done.  
Setting up the scoring function ... done.  
Analyzing the binding site ... done.  
Using random seed: -2102767168  
Performing search ... done.  
Refining results ... done.
```

mode	affinity	dist from best mode	
	(kcal/mol)	rmsd l.b.	rmsd u.b.
1	-5.1	0.000	0.000
2	-5.0	2.918	4.017
3	-5.0	15.627	16.577
4	-4.9	2.038	7.263
5	-4.9	2.022	7.179
6	-4.9	3.067	4.285
7	-4.8	1.984	2.595
8	-4.8	13.465	16.412
9	-4.6	2.350	7.195

```
Writing output ... done.
```



8.4: Appendix IV – IQS Analogue Docking Log Files. IQS12

```
#####  
# If you used AutoDock Vina in your work, please cite:      #  
#                                                           #  
# O. Trott, A. J. Olson,                                    #  
# AutoDock Vina: improving the speed and accuracy of docking #  
# with a new scoring function, efficient optimization and   #  
# multithreading, Journal of Computational Chemistry 31 (2010) #  
# 455-461                                                    #  
#                                                           #  
# DOI 10.1002/jcc.21334                                     #  
#                                                           #  
# Please see http://vina.scripps.edu for more information. #  
#####
```

```
WARNING: The search space volume > 27000 Angstrom^3 (See FAQ)  
Output will be IQS12_out.pdbqt  
Detected 6 CPUs  
Reading input ... done.  
Setting up the scoring function ... done.  
Analyzing the binding site ... done.  
Using random seed: 538798712  
Performing search ... done.  
Refining results ... done.
```

```
mode |  affinity | dist from best mode  
      | (kcal/mol) | rmsd l.b. | rmsd u.b.  
-----+-----+-----+-----  
  1      -4.9      0.000      0.000  
  2      -4.8      3.331      5.635  
  3      -4.8      1.056      2.321  
  4      -4.8      2.102      2.792  
  5      -4.6     18.183     19.721  
  6      -4.6      2.315      3.185  
  7      -4.5     15.480     17.077  
  8      -4.5      7.048      8.852  
  9      -4.5     16.209     17.364
```

```
Writing output ... done.
```




8.4: Appendix IV – IQS Analogue Docking Log Files. IQS13

```
#####  
# If you used AutoDock Vina in your work, please cite:      #  
#                                                            #  
# O. Trott, A. J. Olson,                                    #  
# AutoDock Vina: improving the speed and accuracy of docking #  
# with a new scoring function, efficient optimization and    #  
# multithreading, Journal of Computational Chemistry 31 (2010) #  
# 455-461                                                    #  
#                                                            #  
# DOI 10.1002/jcc.21334                                     #  
#                                                            #  
# Please see http://vina.scripps.edu for more information. #  
#####
```

```
WARNING: The search space volume > 27000 Angstrom^3 (See FAQ)  
Output will be IQS13_out.pdbqt  
Detected 6 CPUs  
Reading input ... done.  
Setting up the scoring function ... done.  
Analyzing the binding site ... done.  
Using random seed: 946657976  
Performing search ... done.  
Refining results ... done.
```

mode	affinity	dist from best mode	
	(kcal/mol)	rmsd l.b.	rmsd u.b.
1	-4.8	0.000	0.000
2	-4.7	17.809	19.209
3	-4.6	15.751	17.375
4	-4.6	16.456	17.956
5	-4.6	12.956	15.431
6	-4.4	17.145	18.441
7	-4.4	18.699	19.461
8	-4.4	17.096	18.361
9	-4.4	15.443	17.059

```
Writing output ... done.
```



8.4: Appendix IV – IQS Analogue Docking Log Files. IQS14

```
#####  
# If you used AutoDock Vina in your work, please cite:      #  
#                                                            #  
# O. Trott, A. J. Olson,                                     #  
# AutoDock Vina: improving the speed and accuracy of docking #  
# with a new scoring function, efficient optimization and    #  
# multithreading, Journal of Computational Chemistry 31 (2010) #  
# 455-461                                                    #  
#                                                            #  
# DOI 10.1002/jcc.21334                                     #  
#                                                            #  
# Please see http://vina.scripps.edu for more information. #  
#####
```

```
WARNING: The search space volume > 27000 Angstrom^3 (See FAQ)  
Output will be IQS14_out.pdbqt  
Detected 6 CPUs  
Reading input ... done.  
Setting up the scoring function ... done.  
Analyzing the binding site ... done.  
Using random seed: 1316724152  
Performing search ... done.  
Refining results ... done.
```

```
mode |  affinity | dist from best mode  
      | (kcal/mol) | rmsd l.b. | rmsd u.b.  
-----+-----+-----+-----  
  1      -4.8      0.000      0.000  
  2      -4.7      1.241      1.819  
  3      -4.6     17.219     18.696  
  4      -4.6      3.844      5.826  
  5      -4.6      3.889      5.841  
  6      -4.5      6.779      8.322  
  7      -4.5     10.357     11.953  
  8      -4.4      1.846      2.339  
  9      -4.4     12.624     14.260
```

```
Writing output ... done.
```



8.4: Appendix IV – IQS Analogue Docking Log Files. IQS15

```
#####  
# If you used AutoDock Vina in your work, please cite:      #  
#                                                           #  
# O. Trott, A. J. Olson,                                    #  
# AutoDock Vina: improving the speed and accuracy of docking #  
# with a new scoring function, efficient optimization and    #  
# multithreading, Journal of Computational Chemistry 31 (2010) #  
# 455-461                                                    #  
#                                                           #  
# DOI 10.1002/jcc.21334                                     #  
#                                                           #  
# Please see http://vina.scripps.edu for more information. #  
#####
```

```
WARNING: The search space volume > 27000 Angstrom^3 (See FAQ)  
Output will be IQS15_out.pdbqt  
Detected 6 CPUs  
Reading input ... done.  
Setting up the scoring function ... done.  
Analyzing the binding site ... done.  
Using random seed: -1136605076  
Performing search ... done.  
Refining results ... done.
```

mode	affinity	dist from best mode	
	(kcal/mol)	rmsd l.b.	rmsd u.b.
1	-4.8	0.000	0.000
2	-4.6	1.751	2.463
3	-4.5	16.827	17.901
4	-4.5	15.794	17.018
5	-4.5	19.079	19.893
6	-4.5	14.973	17.037
7	-4.4	14.766	16.236
8	-4.4	15.017	16.952
9	-4.4	17.213	18.961

```
Writing output ... done.
```



8.5: Appendix V – Summary of *In Surfo* AgrB Crystallisation Trials (Starts Overleaf)



Protein	Sample concentration, mg mL ⁻¹	Final crystallisation sample components	Screen size	Crystallogenesis method	Screen name	Protein and screen soln. aliquot vol., nL	Temp., °C
AgrB2 (His) ₆	4.0	Tris (20mM), NaCl, (100mM), Fos-choline-12, (0.14% wt/v), TCEP-HCl, (1mM), pH 7.55	96 well	Sitting Drop Vapour Diffusion	MPD	200:200	20
AgrB2 (His) ₆	4.0	Tris (20mM), NaCl, (100mM), Fos-choline-12, (0.14% wt/v), TCEP-HCl, (1mM), pH 7.55	96 well	Sitting Drop Vapour Diffusion	ProComplex	200:200	20
AgrB2 (His) ₆	4.0 and 4.5	Tris (20mM), NaCl, (100mM), Fos-choline-12, (0.14% wt/v), TCEP-HCl, (1mM), pH 7.53	26 well	Sitting Drop Vapour Diffusion	Custom optimisation	1000:1000	20
AgrB2 (His) ₆	3.5	Tris (20mM), NaCl, (100mM), Fos-choline-12, (0.14% wt/v), TCEP-HCl, (1mM), pH 7.55	96 well	Sitting Drop Vapour Diffusion	Morpheus	150:150	20
	4.0	Tris (20mM), NaCl, (100mM), Fos-choline-12, (0.14% wt/v), TCEP-HCl, (1mM), pH 7.55					
AgrB2 (His) ₆	3.5	Tris (20mM), NaCl, (100mM), Fos-choline-12, (0.14% wt/v), TCEP-HCl, (1mM), pH 7.55	96 well	Sitting Drop Vapour Diffusion	PACT	150:150	20
	4.0	Tris (20mM), NaCl, (100mM), Fos-choline-12, (0.14% wt/v), TCEP-HCl, (1mM), pH 7.55					
AgrB2 (His) ₆	2.5	Tris (20mM), NaCl, (100mM), Fos-choline-12, (0.14% wt/v), TCEP-HCl, (1mM), pH 7.55	4 well	Sitting Drop Vapour Diffusion	PCT Test	1000:1000	20
AgrB2 (His) ₆	1.0	Tris (20mM), NaCl, (100mM), Dodecyl maltoside, (0.03% wt/v), TCEP-HCl (1mM), pH 7.0	96 well	Sitting Drop Vapour Diffusion	MemGold	200:200	20
AgrB2 (His) ₆	0.5	Tris (20mM), NaCl, (100mM), Dodecyl maltoside, (0.03% wt/v), TCEP-HCl (1mM), pH 7.0	96 well	Sitting Drop Vapour Diffusion	MemGold	200:200	20
	1.0	Tris (20mM), NaCl, (100mM), Dodecyl maltoside, (0.03% wt/v), TCEP-HCl (1mM), pH 7.0	96 well	Sitting Drop Vapour Diffusion	MemGold	200:200	20
AgrB2 (His) ₆	0.5	Tris (20mM), NaCl, (100mM), Dodecyl maltoside, (0.03% wt/v), TCEP-HCl (1mM), pH 7.0	96 well	Sitting Drop Vapour Diffusion	MemGold2	200:200	20
	1.0	Tris (20mM), NaCl, (100mM), Dodecyl maltoside, (0.03% wt/v), TCEP-HCl (1mM), pH 7.0	96 well	Sitting Drop Vapour Diffusion	MemGold2	200:200	20
AgrB2 (His) ₆	0.5	Tris (20mM), NaCl, (100mM), Dodecyl maltoside, (0.03% wt/v), TCEP-HCl (1mM), pH 7.0	96 well	Sitting Drop Vapour Diffusion	MemStart + MemSys	200:200	2



Protein	Sample concentration, mg mL ⁻¹	Final crystallisation sample components	Screen size	Crystallogenesis method	Screen name	Protein and screen soln. aliquot vol., nL	Temp., °C
AgrB2 (His) ₆	0.5	Tris (20mM), NaCl, (100mM), Dodecyl maltoside, (0.03% wt/v), TCEP-HCl (1mM), pH 7.0	96 well	Sitting Drop Vapour Diffusion	MemStart + MemSys	200:200	2
	1.0	Tris (20mM), NaCl, (100mM), Dodecyl maltoside, (0.03% wt/v), TCEP-HCl (1mM), pH 7.0	96 well	Sitting Drop Vapour Diffusion	MemStart + MemSys	200:200	20
AgrB2 (His) ₆	2.0	Tris (20mM), NaCl, (100mM), Dodecyl maltoside, (0.03% wt/v), TCEP-HCl (1mM), pH 7.0	96 well	Sitting Drop Vapour Diffusion	MemGold	200:200	20
AgrB2 (His) ₆	2.0	Tris (20mM), NaCl, (100mM), Dodecyl maltoside, (0.03% wt/v), TCEP-HCl (1mM), pH 7.0	96 well	Sitting Drop Vapour Diffusion	MemGold2	200:200	20
AgrB2 (His) ₆	2.0	Tris (20mM), NaCl, (100mM), Dodecyl maltoside, (0.03% wt/v), TCEP-HCl (1mM), pH 7.0	96 well	Sitting Drop Vapour Diffusion	MemStart + MemSys	200:200	20
AgrB2 (His) ₆	1.0	Tris (20mM), NaCl, (100mM), Dodecyl maltoside, (0.03% wt/v), TCEP-HCl (1mM), pH 7.0	96 well	Sitting Drop Vapour Diffusion	Custom optimisation	1000:1000	1
AgrB2 (His) ₆	0.1, 0.5, 1.5 and 2.0	Tris (20mM), NaCl, (100mM), Dodecyl maltoside, (0.03% wt/v), TCEP-HCl (1mM), pH 7.0	96 well	Sitting Drop Vapour Diffusion	Custom optimisation	1000:1000	20
AgrB2 (His) ₆	0.1	Tris (20mM), NaCl, (100mM), Dodecyl maltoside, (0.03% wt/v), TCEP-HCl (1mM), pH 7.0	96 well	Sitting Drop Vapour Diffusion	MemGold	200:200	20
AgrB2 (His) ₆	0.1	Tris (20mM), NaCl, (100mM), Dodecyl maltoside, (0.03% wt/v), TCEP-HCl (1mM), pH 7.0	96 well	Sitting Drop Vapour Diffusion	MemGold2	200:200	20
AgrB2 (His) ₆	0.1	Tris (20mM), NaCl, (100mM), Dodecyl maltoside, (0.03% wt/v), TCEP-HCl (1mM), pH 7.0	96 well	Sitting Drop Vapour Diffusion	MemStart + MemSys	200:200	20
AgrB2 (His) ₆	1.0	Tris (20mM), NaCl, (100mM), Dodecyl maltoside, (0.03% wt/v), TCEP-HCl (1mM), pH 7.0	96 well	Sitting Drop Vapour Diffusion	MemGold	400:400	4
AgrB2 (His) ₆	1.0	Tris (20mM), NaCl, (100mM), Dodecyl maltoside, (0.03% wt/v), TCEP-HCl (1mM), pH 7.0	96 well	Sitting Drop Vapour Diffusion	MemGold2	400:400	4
AgrB2 (His) ₆	1.0	Tris (20mM), NaCl, (100mM), Dodecyl maltoside, (0.03% wt/v), TCEP-HCl (1mM), pH 7.0	96 well	Sitting Drop Vapour Diffusion	MemStart + MemSys	400:400	4
AgrB2 (His) ₆	8.9	KP (20mM), NaCl, (50mM), Dodecyl maltoside (>0.02% wt/v), TCEP-HCl (1mM), pH 7.5	96 well	Sitting Drop Vapour Diffusion	MemStart + MemSys	150:150	4



Protein	Sample concentration, mg mL ⁻¹	Final crystallisation sample components	Screen size	Crystallogenesis method	Screen name	Protein and screen soln. aliquot vol., nL	Temp., °C
AgrB2 (His) ₆	8.9	KP (20mM), NaCl, (50mM), Dodecyl maltoside (>0.02% wt/v), TCEP-HCl (1mM), pH 7.6	96 well	Sitting Drop Vapour Diffusion	MemGold	150:150	4
AgrB2 (His) ₆	4.5	KP (20mM), NaCl, (50mM), Dodecyl maltoside (>0.02% wt/v), TCEP-HCl (1mM), pH 7.7	96 well	Sitting Drop Vapour Diffusion	MemGold 2	150:150	4
AgrB2 (His) ₆	2.5	KP (20mM), NaCl, (50mM), Dodecyl maltoside (>0.02% wt/v), TCEP-HCl (1mM), pH 7.8	96 well	Sitting Drop Vapour Diffusion	MemGold	150:150	20
AgrB2 (His) ₆	2.5	KP (20mM), NaCl, (50mM), Dodecyl maltoside (>0.02% wt/v), TCEP-HCl (1mM), pH 7.9	96 well	Sitting Drop Vapour Diffusion	MemGold 2	150:150	20
AgrB2 (His) ₆	5.0	KP (20mM), NaCl, (50mM), Dodecyl maltoside (>0.02% wt/v), TCEP-HCl (1mM), pH 7.10	96 well	Sitting Drop Vapour Diffusion	MemGold 2	150:150	20
AgrB2 (His) ₆	2.5	KP (20mM), NaCl, (50mM), Dodecyl maltoside (>0.02% wt/v), TCEP-HCl (1mM), pH 7.11	96 well	Sitting Drop Vapour Diffusion	MemStart + MemSys	150:150	20
AgrB2 (His) ₆	2.5	KP (20mM), NaCl, (50mM), Dodecyl maltoside (>0.02% wt/v), TCEP-HCl (1mM), pH 7.12	96 well	Sitting Drop Vapour Diffusion	PACT	150:150	20
AgrB2 (His) ₆	2.5	KP (20mM), NaCl, (50mM), Dodecyl maltoside (>0.02% wt/v), TCEP-HCl (1mM), pH 7.13	96 well	Sitting Drop Vapour Diffusion	Morpheus	150:150	20
AgrB2 (His) ₆	2.5	KP (20mM), NaCl, (50mM), Dodecyl maltoside (>0.02% wt/v), TCEP-HCl (1mM), pH 7.14	96 well	Sitting Drop Vapour Diffusion	jCSG	150:150	20
AgrB2 (His) ₆	5.0	KP (20mM), NaCl, (50mM), 1mM DTT, pH 7.5 (LMNG_AgrB)	96 well	Sitting Drop Vapour Diffusion	MemTrans	200:200	20
AgrB2 (His) ₆	5.0	KP (20mM), NaCl, (50mM), 1mM DTT, pH 7.5 (LMNG_AgrB)	96 well	Sitting Drop Vapour Diffusion	MemGold	200:200	20
AgrB2 (His) ₆	5.0	KP (20mM), NaCl, (50mM), 1mM DTT, pH 7.5 (LMNG_AgrB)	96 well	Sitting Drop Vapour Diffusion	MemGold 2	200:200	20
AgrB2 (His) ₆	5.0	KP (20mM), NaCl, (100mM), 1mM DTT, 0.046% LDAO, pH 7.5	96 well	Sitting Drop Vapour Diffusion	MemGold	200:200	20
AgrB2 (His) ₆	5.0	KP (20mM), NaCl, (100mM), 1mM DTT, 0.046% LDAO, pH 7.5	96 well	Sitting Drop Vapour Diffusion	MemGold 2	200:200	20
AgrB2 (His) ₆	5.0	KP (20mM), NaCl, (100mM), 1mM DTT, 0.046% LDAO, pH 7.5	96 well	Sitting Drop Vapour Diffusion	MemTrans	200:200	20



Protein	Sample concentration, mg mL ⁻¹	Final crystallisation sample components	Screen size	Crystallogenesis method	Screen name	Protein and screen soln. aliquot vol., nL	Temp., °C
AgrB2 (His) ₆	5.0	KP (20mM), NaCl, (100mM), 1mM DTT, 0.046% LDAO, pH 7.5	96 well	Sitting Drop Vapour Diffusion	MemTrans	200:200	20
AgrB2 (His) ₆	2.5	KP (20mM), NaCl, (100mM), 1mM DTT, 0.046% LDAO, pH 7.5	96 well	Sitting Drop Vapour Diffusion	MemTrans	200:200	20
AgrB2 (His) ₆	1.0	KP (20mM), NaCl, (100mM), 1mM DTT, 0.046% LDAO, pH 7.5	96 well	Sitting Drop Vapour Diffusion	MemGold	200:200	20
AgrB2 (His) ₆	1.0	KP (20mM), NaCl, (100mM), 1mM DTT, 0.046% LDAO, pH 7.5	96 well	Sitting Drop Vapour Diffusion	MemGold 2	200:200	20
AgrB2 (His) ₆	1.0	KP (20mM), NaCl, (100mM), 1mM DTT, 0.046% LDAO, pH 7.5	96 well	Sitting Drop Vapour Diffusion	MemStart + MemSys	200:200	20
AgrB2 (His) ₆	1.0	KP (20mM), NaCl, (100mM), 1mM DTT, 0.046% LDAO, pH 7.5	96 well	Sitting Drop Vapour Diffusion	MemTrans	200:200	20
AgrB2 (His) ₆	1.0	KP (20mM), NaCl, (100mM), 1mM DTT, 0.046% LDAO, pH 7.5	96 well	Sitting Drop Vapour Diffusion	MemMeso	200:200	20
AgrB2 (His) ₆	1.0	KP (20mM), NaCl, (100mM), 1mM DTT, 0.046% LDAO, pH 7.5	96 well	Sitting Drop Vapour Diffusion	MemWizard	200:200	20
AgrB (His) ₁₀	7.8	KP (20mM), NaCl, (50mM), Dodecyl maltoside (>0.02% wt/v), pH 7.5	96 well	Sitting Drop Vapour Diffusion	MemTrans	100:100	20
AgrB (His) ₁₀	7.8	KP (20mM), NaCl, (50mM), Dodecyl maltoside (>0.02% wt/v), pH 7.6	96 well	Sitting Drop Vapour Diffusion	PACT Eco	100:100	20
AgrB (His) ₁₀	7.8	KP (20mM), NaCl, (50mM), Dodecyl maltoside (>0.02% wt/v), pH 7.7	96 well	Sitting Drop Vapour Diffusion	MemTrans	200:200	20
AgrB (His) ₁₀	7.8	KP (20mM), NaCl, (50mM), Dodecyl maltoside (>0.02% wt/v), pH 7.8	96 well	Sitting Drop Vapour Diffusion	MemTrans B11 (1.1x conc.) w/ MemAdvantage (1in10 dilution into reservoir	200:200	20
AgrB (His) ₁₀	7.8 , 3.9 and 2.0	KP (20mM), NaCl, (50mM), Dodecyl maltoside (>0.02% wt/v), pH 7.9	46 well	Sitting Drop Vapour Diffusion	MemTrans B11 (1.0x conc.) w/ Calixar C2B 1 in 8 dilution into drop	1500:500:2000	20

Table 8.1. Summary of AgrB crystallisation trials.



8.6: Appendix VI – Instruct Application Documents (Abstract)

Structural Studies of the Staphylococcal AgrB Protease Substrate Selectivity

Staphylococcus aureus is a Gram-positive human commensal bacteria which is notoriously rapid at acquiring resistance to new antibiotic drugs. The ability of *S.aureus* to rapidly acquire resistance to antibiotics is compounded by the high number and severity of infections it is capable of producing. These include meningitis, toxic shock syndrome, pneumonia and endocarditis. The host immune response and virulence factor production by *S.aureus* is controlled by a Quorum Sensing (QS) circuit called the accessory gene regulator (*agr*).^[1] The extracellular signalling (autoinducer) molecule of this QS circuit is a small thiolactone-containing signal peptide, abbreviated AIP. AgrB is a 22kDa integral membrane protease critical for the processing of a linear precursor peptide, AgrD, into AIP. Pharmacological inhibition of AgrB would therefore block this essential process within the QS system by disabling production of the AIP signal molecule. This would ultimately attenuate virulence and inhibit biofilm formation. Furthermore this mode of inhibition would be non-lethal to the cell and thus transcend the capabilities of currently-available antibiotics.

Our group aims to solve the structure of AgrB to elucidate the function, selectivity and mechanism of AgrB action and inhibition. Since AgrB contains only small hydrophilic loops, the type II (loop-to-loop) packing of the detergent-solubilised protein is likely to be intrinsically challenging without substantial modification. The prospect of a rigid nanobody crystallisation chaperone is therefore exciting to us since it would provide a substantial surface residing *exo*- to the detergent belt surrounding the purified AgrB molecules. This methodology was found to be critical to the structure determination of the mouse serotonin 5-HT₃ receptor^[2] and by attending your course and generating a nanobody against AgrB we hope to emulate this success.

Assorted References

[1] Targeting *Staphylococcus aureus* Quorum Sensing with Nonpeptidic Small Molecule Inhibitors. *J. Med. Chem.*, 2014, **57**(6), 2813-2819.

[2] X-ray Structure of the Mouse Serotonin 5-HT₃ Receptor. *Nature*, 2014, **512**, 276-281.



8.7: Appendix VII – Instruct Application Documents (Motivation Letter)

Centre for Biomolecular Sciences
University of Nottingham
University Park, Nottingham
NG7 2RD

22nd February, 2018

Dear Maia De Kerpel,

As outlined in the abstract provided, our group aims to solve the structure the AgrB protease found in *S.aureus*. We are one of only several groups in the world working on this ambitious target and we aim to achieve our goal by pursuing as many different routes to structure determination as possible.

Not only is AgrB an attractive target due to the pharmacological implications as a target for novel antibiotics, it also bears no homology with other protease classes and no structure has yet been solved for any AgrB or so-called AgrB-like protein. In addition to these, little is known about the exact proteolytic mechanism of AgrB except it being absolutely necessary for generating the auto inducer peptide in the *S.aureus* quorum sensing system. It must install a thiolactone ring in the precursor peptide AgrD, and this itself must be preceded by, or concomitant with C-terminal cleavage of AgrD. Whilst published evidence exists to support these observations they do not contain any atomic-level mechanistic detail. Furthermore, it is not yet understood how the processed AgrD is delivered across the cell membrane and whether AgrB, or AgrB oligomer, plays a role in this process also.

Our group is capable of purifying large quantities of pure, folded and monodisperse AgrB in both β -dodecylmaltoside and Styrene Maleic Acid Lipid Particles (SMALPs) via affinity and size exclusion chromatography. Standard vapour diffusion crystal trials have been set up with detergent-solubilised AgrB and we are currently establishing *in mesophase* crystallisation experiments and CryoEM trials also. Furthermore our group has the chemical expertise to synthesize an inhibitory molecule for cocrystallisation studies.

It is our desire to capitalise on this opportunity provided by Instruct to increase our chances of successful structure determination by learning about nanobody generation. Ultimately we would use these expertise and the isolated nanobody to pursue the following three targets:

1. Nanobody + AgrB
2. Nanobody + AgrB + AgrD substrate
3. Nanobody + AgrB + inhibitor

We hope you find our target a desirable inclusion to your nanobody generation project.

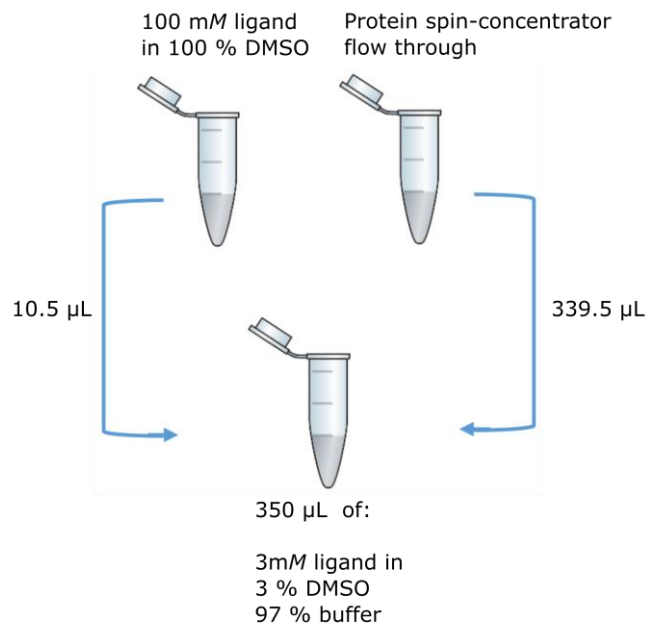
Yours sincerely,

Thomas Warwick

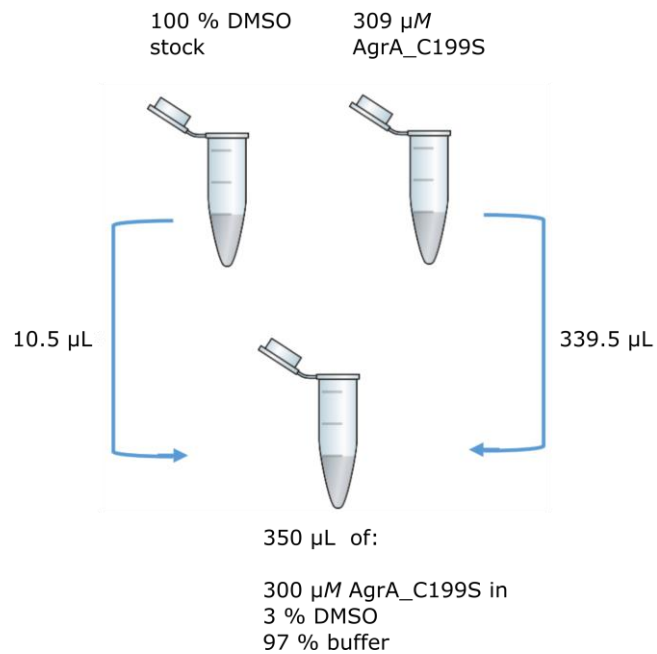


8.8: Appendix VIII – ITC Sample Preparation Schematic

Ligand preparation



Host preparation

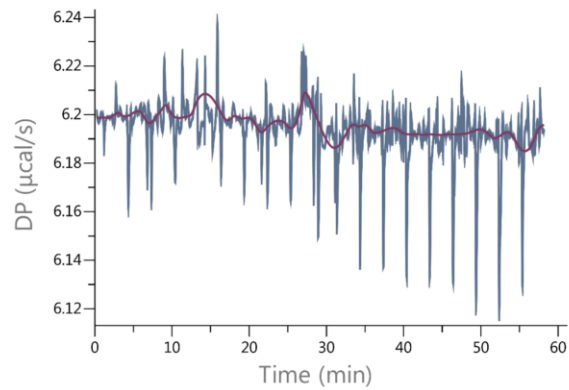


Schematic of the strategy used to prepare the ligand and host samples for ITC. Although no binding of IQS analogues to AgrA was observed, no buffer mismatch was observed either. Thus the strategy is of general use and may be adapted to design other experiments.

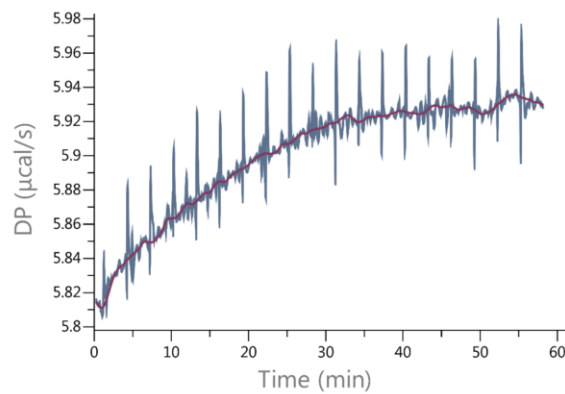


8.9: Appendix IX – Preliminary ITC Thermograms for IQS3 Titrated into AgrA_C199S

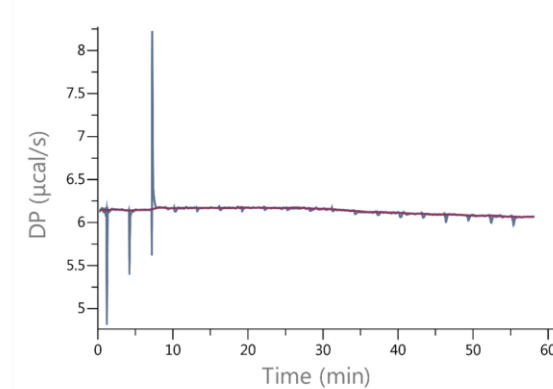
Ligand titrated into host buffer



3 mM ligand (IQS3) titrated into 300 µM AgrA_C199S at 20 °C



3 mM ligand (IQS3) titrated into 300 µM AgrA_C199S at 37 °C



Data obtained from the preliminary ITC experiments. The ligand-to-buffer sample (top) although noisy does not show buffer mismatch. No sigmoidal binding curves were obtained from titration of the ligand, IQS3, into sample. At 37 °C (bottom) the sample precipitated out of solution.



8.10: Appendix X – Key Parameters for All Constructs Discussed

Construct	Residues, N ^o	M _r , kDa	ϵ , ^a dm ³ mol ⁻¹ cm ⁻¹	ϵ , ^b mL mg ⁻¹ cm ⁻¹	Theoretical pI, ^c pH	Tryptophan, Y/N
WT AgrA	257	30.305	14,900	0.492	6.24	N
AgrA_C199S	257	30.229	14,900	0.492	6.24	N
AgrA_D59E_C199S	257	30.303	14,900	0.492	6.24	N
AgrA_K77A_C199S	257	30.248	14,900	0.493	6.14	N
AgrA_K101A_C199S	257	30.196	14,900	0.493	6.14	N
AgrA_Y100T_K101A _C199S	257	30.104	13,410	0.445	6.14	N
AgrA_Y100T_K101T _C199S	257	30.164	13,410	0.445	6.14	N
AgrA_Y100A_K101A	257	30.104	13,410	0.445	6.14	N
AgrA ₁₋₁₄₁	160	18.664	7,450	0.399	5.60	N
AgrA ₁₃₅₋₂₃₈ _C199S	121	14.604	7,450	0.51	8.72	N
AgrB1_(His) ₆	197	22.995	18,920	0.822	9.77	Y
AgrB1_(His) ₁₀	201	23.543	18,910	0.803	9.77	Y
AgrB2_(His) ₆	195	22.745	17,880	0.786	9.47	N
AgrB1_(His) ₁₀	199	23.294	17,880	0.768	9.47	N

Table 8.2. Key parameters for constructs discussed in this thesis. All parameters calculated using ProtParam [Swiss Institute of Bioinformatics] and include all of the residues that are expressed in the respective constructs (see relevant methods sections for vectors). ^aThe molar absorption coefficient. ^bThe absorption coefficient of a 0.1 % w/v solution (1 g L⁻¹) which is calculated by dividing the molar absorption coefficient by the M_r (i.e. dm³ mol⁻¹ cm⁻¹ / g mol⁻¹ = dm³ g⁻¹ cm⁻¹ ≡ mL mg⁻¹ cm⁻¹). ^cIsoelectric point (pH at which the sample has no net charge).



Bibliography

1. Merbs CF. A new world of infectious disease. *Am J Phys Anthropol.* 1992;35(15S):3-42. doi:10.1002/ajpa.1330350603
2. Baxby D. Jenner, Edward (1749–1823), surgeon and pioneer of smallpox vaccination. 2009;(May):1-10.
<https://www.oxforddnb.com/view/10.1093/ref:odnb/9780198614128.001.0001/odnb-9780198614128-e-14749>
3. Rihová M. Sir Alexander Fleming. *Cas Lek Cesk.* 2008;147(6):351-352.
4. Tan SY, Ponstein N. Jonas Salk (1914–1995): A vaccine against polio. *Singapore Med J.* 2019;60(1):9-10. doi:10.11622/smedj.2019002
5. Public Health England. Trends in HIV testing, new diagnoses and people receiving HIV-related care in the United Kingdom: data to the end of December 2019. 2020;14(20):1-16.
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/541111/hiv-testing-diagnoses-and-care-2019.pdf
6. Schackman BR, Gebo KA, Walensky RP, Losina E, Muccio T, Sax PE, Weinstein MC, Seage GR III, Moore RD, Freedberg KA. The Lifetime Cost of Current Human Immunodeficiency Virus Care in the United States. *Med Care.* 2006;44(11).
https://journals.lww.com/lww-medicalcare/Fulltext/2006/11000/The_Lifetime_Cost_of_Current_Human.5.aspx
7. Moore RD, Chaisson RE. Natural history of HIV infection in the era of combination antiretroviral therapy. *Aids.* 1999;13(14):1933-1942. doi:10.1097/00002030-199910010-00017
8. Holten KB, Onusko EM. Appropriate prescribing of oral beta-lactam antibiotics. *Am Fam Physician.* 2000;62(3):611-620.
9. Wolf J, Bruno S, Eichberg M, Jannat R, Rudo S, VanRheenen S, Collier BA. Applying lessons from the Ebola vaccine experience for SARS-CoV-2 and other epidemic pathogens. *npj Vaccines.* 2020;5(1). doi:10.1038/s41541-020-0204-7
10. Wang W, Arshad MI, Khurshid M, Rasool MH, Nisar MA, Aslam MA, Qamar MU. Antibiotic resistance: a rundown of a global crisis. *Infect Drug Resist.* Published online 2018:1645-1658.
11. National Institute for Care Excellence. Putting NICE guidance into practice Resource impact report: Sepsis: the recognition, diagnosis and early. 2016;(July).
<https://www.nice.org.uk/guidance/ng51/resources/resource-impact-report-pdf-2549846269>



12. Morgan S, Grootendorst P, Lexchin J, Cunningham C, Greyson D. The cost of drug development: A systematic review. *Health Policy (New York)*. 2011;100(1):4-17. doi:10.1016/j.healthpol.2010.12.002
13. Blumenthal KG, Peter JG, Trubiano JA, Phillips EJ. Antibiotic allergy. *Lancet*. 2019;393(10167):183-198. doi:10.1016/S0140-6736(18)32218-9
14. Love BL, Mann JR, Hardin JW, Lu ZK, Cox C, Amrol DJ. Antibiotic prescription and food allergy in young children. *Allergy, Asthma Clin Immunol*. 2016;12(1):1-8. doi:10.1186/s13223-016-0148-7
15. Lavebratt C, Yang LL, Giacobini MB, Forsell Y, Schalling M, Partonen T, Gissler M. Early exposure to antibiotic drugs and risk for psychiatric disorders: a population-based study. *Transl Psychiatry*. 2019;9(1). doi:10.1038/s41398-019-0653-9
16. Schneewind O. Classic spotlight: Molecular biology of methicillin resistance in *Staphylococcus aureus*. *J Bacteriol*. 2016;198(14):1903-1903. doi:10.1128/JB.00277-16
17. Jevons MP. "Celbenin" - resistant Staphylococci. *Br Med J*. 1961;1(5219):124-125.
18. Harkins CP, Pichon B, Doumith M, Parkhill J, Westh H, Tomasz A, de Lencastre H, Bentley SD, Kearns AM, Holden MTG. Methicillin-resistant *Staphylococcus aureus* emerged long before the introduction of methicillin into clinical practice. *Genome Biol*. 2017;18(1):1-11. doi:10.1186/s13059-017-1252-9
19. Rubinstein E, Keynan Y. Vancomycin revisited - 60 years later. *Front Public Heal*. 2014;2(OCT):1-7. doi:10.3389/fpubh.2014.00217
20. Cong Y, Yang S, Rao X. Vancomycin resistant *Staphylococcus aureus* infections: A review of case updating and clinical features. *J Adv Res*. 2020;21:169-176. doi:10.1016/j.jare.2019.10.005
21. Farber BF, Moellering RC. Retrospective study of the toxicity of preparations of vancomycin from 1974 to 1981. *Antimicrob Agents Chemother*. 1983;23(1):138-141. doi:10.1128/AAC.23.1.138
22. Pantosti A, Sanchini A, Monaco M. Mechanisms of antibiotic resistance in *Staphylococcus aureus*. *Future Microbiol*. 2007;2(3):323-334. doi:10.2217/17460913.2.3.323
23. Tunkel AR, Hartman BJ, Kaplan SL, Kaufman BA, Roos KL, Scheld WM, Whitley RJ. Practice guidelines for the management of bacterial meningitis. *Clin Infect Dis*. 2004;39(9):1267-1284. doi:10.1086/425368
24. Mishra AK, Yadav P, Mishra A. A Systemic Review on Staphylococcal Scalded Skin Syndrome (SSSS): A Rare and Critical Disease of Neonates. *Open Microbiol J*. 2016;10(1):150-159. doi:10.2174/1874285801610010150



25. Low DE. Toxic shock syndrome: major advances in pathogenesis, but not treatment. *Crit Care Clin.* 2013;29(3):651-675. doi:10.1016/j.ccc.2013.03.012
26. Prina E, Ranzani OT, Torres A. Community-acquired pneumonia. *Lancet.* 2015;386(9998):1097-1108. doi:10.1016/S0140-6736(15)60733-4
27. Holland TL, Baddour LM, Bayer AS, Hoen B, Miro JM, Fowler Jr VG. Infective endocarditis. *Nat Rev Dis Prim.* 2016;2:16059. doi:10.1038/nrdp.2016.59
28. Brown DR, Pattee PA. Identification of a chromosomal determinant of alpha-toxin production in *Staphylococcus aureus*. *Infect Immun.* 1980;30(1):36-42. doi:10.1128/IAI.30.1.36-42.1980
29. Janzon L, Löfdahl S, Arvidson S. Evidence for a coordinate transcriptional control of alpha-toxin and protein A synthesis in *Staphylococcus aureus*. *FEMS Microbiol Lett.* 1986;33(2-3):193-198. doi:https://doi.org/10.1111/j.1574-6968.1986.tb01270.x
30. Graille M, Stura EA, Corper AL, Sutton BJ, Taussig MJ, Charbonnier J-B, Silverman GJ. Crystal structure of a *Staphylococcus aureus* protein A domain complexed with the Fab fragment of a human IgM antibody: Structural basis for recognition of B-cell receptors and superantigen activity. *Proc Natl Acad Sci.* 2000;97(10):5399-5404. doi:10.1073/pnas.97.10.5399
31. Bhakdi S, Tranum-Jensen J. Alpha-toxin of *Staphylococcus aureus*. *Microbiol Rev.* 1991;55(4):733-751.
32. Kornblum J. *agr*: a polycistronic locus regulating exoprotein synthesis in *Staphylococcus aureus*. *Mol Biol staphylococci.* Published online 1990:373-402.
33. Novick RP, Ross HF, Projan SJ, Kornblum J, Kreiswirth B, Moghazeh S. Synthesis of staphylococcal virulence factors is controlled by a regulatory RNA molecule. *EMBO J.* 1993;12(10):3967-3975.
34. Nikolskaya AN, Galperin MY. A novel type of conserved DNA-binding domain in the transcriptional regulators of the AlgR/AgrA/LytR family. *Nucleic Acids Res.* 2002;30(11):2453-2459. doi:10.1093/nar/30.11.2453
35. Morfeldt E, Panova-Sapundjieva I, Gustafsson B, Arvidson S. Detection of the response regulator AgrA in the cytosolic fraction of *Staphylococcus aureus* by monoclonal antibodies. *FEMS Microbiol Lett.* 1996;143(2-3):195-201. doi:10.1016/0378-1097(96)00310-2
36. Novick RP, Projan SJ, Kornblum J, Ross HF, Ji G, Kreiswirth B, Vandenesch F, Moghazeh S. The *agr* P2 operon: An autocatalytic sensory transduction system in *Staphylococcus aureus*. *Mol Gen Genet.* 1995;248(4):446-458. doi:10.1007/BF02191645
37. Balaban I, Novick RP. Autocrine regulation of toxin synthesis by *Staphylococcus aureus*. *Proc Natl Acad Sci U S A.* 1995;92(5):1619-1623. doi:10.1073/pnas.92.5.1619



38. Ji G, Beavis R, Novick RP. Bacterial interference caused by autoinducing peptide variants. *Science*. 1997;276(5321):2027-2030. doi:10.1126/science.276.5321.2027
39. Nealson KH, Platt T, Hastings JW. Cellular control of the synthesis and activity of the bacterial luminescent system. *J Bacteriol*. 1970;104(1):313-322. doi:10.1128/jb.104.1.313-322.1970
40. Ahmer BMM. Cell-to-cell signalling in *Escherichia coli* and *Salmonella enterica*. *Mol Microbiol*. 2004;52(4):933-945. doi:10.1111/j.1365-2958.2004.04054.x
41. Ahmer BMM, Van Reeuwijk J, Timmers CD, Valentine PJ, Heffron F. *Salmonella typhimurium* encodes an SdiA homolog, a putative quorum sensor of the LuxR family, that regulates genes on the virulence plasmid. *J Bacteriol*. 1998;180(5):1185-1193. doi:10.1128/jb.180.5.1185-1193.1998
42. Thoendel M, Horswill AR. Identification of *Staphylococcus aureus* AgrD residues required for autoinducing peptide biosynthesis. *J Biol Chem*. 2009;284(33):21828-21838. doi:10.1074/jbc.M109.031757
43. Zhang L, Lin J, Ji G. Membrane anchoring of the AgrD N-terminal amphipathic region is required for its processing to produce a quorum-sensing pheromone in *Staphylococcus aureus*. *J Biol Chem*. 2004;279(19):19448-19456. doi:10.1074/jbc.M311349200
44. Wang B, Zhao A, Novick RP, Muir TW. Key driving forces in the biosynthesis of autoinducing peptides required for staphylococcal virulence. *Proc Natl Acad Sci U S A*. 2015;112(34):10679-10684. doi:10.1073/pnas.1506030112
45. Cheung GYC, Joo H-S, Chatterjee SS, Otto M. Phenol-soluble modulins--critical determinants of staphylococcal virulence. *FEMS Microbiol Rev*. 2014;38(4):698-719. doi:10.1111/1574-6976.12057
46. Schwartz K, Sekedat MD, Syed AK, O'Hara B, Payne DE, Lamb A, Boles BR. The AgrD N-terminal leader peptide of *Staphylococcus aureus* has cytolytic and amyloidogenic properties. *Infect Immun*. 2014;82(9):3837-3844. doi:10.1128/IAI.02111-14
47. Carpenter EP, Beis K, Cameron AD, Iwata S. Overcoming the challenges of membrane protein crystallography. *Curr Opin Struct Biol*. 2008;18(5):581-586. doi:10.1016/j.sbi.2008.07.001
48. Luckey M. Chapter 1 Introduction to the Structural Biology of Membrane Proteins. In: *Computational Biophysics of Membrane Proteins*. The Royal Society of Chemistry; 2017:1-18. doi:10.1039/9781782626695-00001
49. Tiefenauer L, Demarche S. Challenges in the Development of Functional Assays of Membrane Proteins. *Materials (Basel)*. 2012;5(11):2205-2242. doi:10.3390/ma5112205



50. Seddon AM, Curnow P, Booth PJ. Membrane proteins, lipids and detergents: not just a soap opera. *Biochim Biophys Acta - Biomembr.* 2004;1666(1):105-117. doi:<https://doi.org/10.1016/j.bbamem.2004.04.011>
51. Zhang L, Gray L, Novick RP, Ji G. Transmembrane topology of AgrB, the protein involved in the post-translational modification of AgrD in *Staphylococcus aureus*. *J Biol Chem.* 2002;277(38):34736-34742. doi:10.1074/jbc.M205367200
52. Qiu R, Pei W, Zhang L, Lin J, Ji G. Identification of the putative staphylococcal AgrB catalytic residues involving the proteolytic cleavage of AgrD to generate autoinducing peptide. *J Biol Chem.* 2005;280(17):16695-16704. doi:10.1074/jbc.M411372200
53. Gray B, Hall P, Gresham H. Targeting *agr*- and *agr*-Like quorum sensing systems for development of common therapeutics to treat multiple gram-positive bacterial infections. *Sensors (Basel).* 2013;13(4):5130-5166. doi:10.3390/s130405130
54. Kavanaugh JS, Thoendel M, Horswill AR. A role for type I signal peptidase in *Staphylococcus aureus* quorum sensing. *Mol Microbiol.* 2007;65(3):780-798. doi:10.1111/j.1365-2958.2007.05830.x
55. Todd DA, Parlet CP, Crosby HA, Malone CL, Heilmann KP, Horswill AR, Cech NB. Signal biosynthesis inhibition with ambuic acid as a strategy to target antibiotic-resistant infections. *Antimicrob Agents Chemother.* 2017;61(8):1-11. doi:10.1128/AAC.00263-17
56. Li JY, Harper JK, Grant DM, Tombe BO, Bashyal B, Hess WM, Strobel GA. Ambuic acid, a highly functionalized cyclohexenone with antifungal activity from *Pestalotiopsis* spp. and *Monochaetia* sp. *Phytochemistry.* 2001;56(5):463-468. doi:10.1016/s0031-9422(00)00408-8
57. Nakayama J, Uemura Y, Nishiguchi K, Yoshimura N, Igarashi Y, Sonomoto K. Ambuic acid inhibits the biosynthesis of cyclic peptide quormones in gram-positive bacteria. *Antimicrob Agents Chemother.* 2009;53(2):580-586. doi:10.1128/AAC.00995-08
58. Novick RP. Autoinduction and signal transduction in the regulation of staphylococcal virulence. *Mol Microbiol.* 2003;48(6):1429-1449. doi:10.1046/j.1365-2958.2003.03526.x
59. Ji G, Pei W, Zhang L, Qiu R, Lin J, Benito Y, Lina G, Novick RP. *Staphylococcus intermedius* produces a functional *agr* autoinducing peptide containing a cyclic lactone. *J Bacteriol.* 2005;187(9):3139-3150. doi:10.1128/JB.187.9.3139-3150.2005
60. Håvarstein LS, Coomaraswamy G, Morrison DA. An unmodified heptadecapeptide pheromone induces competence for genetic transformation in *Streptococcus pneumoniae*. *Proc Natl Acad Sci U S A.* 1995;92(24):11140-11144. doi:10.1073/pnas.92.24.11140



61. Parsek MR, Greenberg EP. Acyl-homoserine lactone quorum sensing in gram-negative bacteria: a signaling mechanism involved in associations with higher organisms. *Proc Natl Acad Sci U S A*. 2000;97(16):8789-8793. doi:10.1073/pnas.97.16.8789
62. Cottrell TL. *The Strengths of Chemical Bonds*. 2nd ed. Butterworths Scientific Publications; 1958.
63. Chan WC, Coyle BJ, Williams P. Virulence regulation and quorum sensing in staphylococcal infections: competitive AgrC antagonists as quorum sensing inhibitors. *J Med Chem*. 2004;47(19):4633-4641. doi:10.1021/jm0400754
64. Wright JS, Jin R, Novick RP. Transient interference with staphylococcal quorum sensing blocks abscess formation. *Proc Natl Acad Sci U S A*. 2005;102(5):1691-1696. doi:10.1073/pnas.0407661102
65. Lyon GJ, Wright JS, Muir TW, Novick RP. Key determinants of receptor activation in the *agr* autoinducing peptides of *Staphylococcus aureus*. *Biochemistry*. 2002;41(31):10095-10104. doi:10.1021/bi026049u
66. Fleming V, Feil E, Sewell AK, Day N, Buckling A, Massey RC. *agr* Interference between Clinical *Staphylococcus aureus* Strains in an Insect Model of Virulence. *J Bacteriol*. 2006;188(21):7686-7688. doi:10.1128/JB.00700-06
67. Jarraud S, Lyon GJ, Figueiredo AMS, Gérard L, Vandenesch F, Etienne J, Muir TW, Novick RP. Exfoliatin-Producing Strains Define a Fourth *agr* Specificity Group in *Staphylococcus aureus*. *J Bacteriol*. 2000;182(22):6517-6522. doi:10.1128/JB.182.22.6517-6522.2000
68. Lina G, Jarraud S, Ji G, Greenland T, Pedraza A, Etienne J, Novick RP, Vandenesch F. Transmembrane topology and histidine protein kinase activity of AgrC, the *agr* signal receptor in *Staphylococcus aureus*. *Mol Microbiol*. 1998;28(3):655-662. doi:10.1046/j.1365-2958.1998.00830.x
69. Grebe TW, Stock JB. The histidine protein kinase superfamily. *Adv Microb Physiol*. 1999;41:139-227. doi:10.1016/s0065-2911(08)60167-8
70. George Cisar EA, Geisinger E, Muir TW, Novick RP. Symmetric signalling within asymmetric dimers of the *Staphylococcus aureus* receptor histidine kinase AgrC. *Mol Microbiol*. 2009;74(1):44-57. doi:10.1111/j.1365-2958.2009.06849.x
71. Srivastava SK, Rajasree K, Fasim A, Arakere G, Gopal B. Influence of the AgrC-AgrA Complex on the Response Time of *Staphylococcus aureus* Quorum Sensing. *J Bacteriol*. 2014;196(15):2876-2888. doi:10.1128/JB.01530-14
72. Clardy J, Fischbach MA, Currie CR. The natural history of antibiotics. *Curr Biol*. 2009;19(11):R437-R441. doi:10.1016/j.cub.2009.04.001
73. Chandra N, Kumar S. Antibiotics and Antibiotics Resistance Genes in Soils: Monitoring, Toxicity, Risk Assessment and Management. In: Hashmi MZ, Strezov V,



Varma A, eds. *Antibiotics Producing Soil Microorganisms*. Springer International Publishing; 2017:1-18. doi:10.1007/978-3-319-66260-2_1

74. Papp-Wallace KM, Endimiani A, Taracila MA, Bonomo RA. Carbapenems: Past, present, and future. *Antimicrob Agents Chemother*. 2011;55(11):4943-4960. doi:10.1128/AAC.00296-11
75. Harrison CJ, Bratcher D. Cephalosporins. *Pediatr Rev*. 2008;29(8):264-273. doi:10.1542/pir.29-8-264
76. Vasquez JK, Tal-Gan Y, Cornilescu G, Tyler KA, Blackwell HE. Simplified AIP-II Peptidomimetics Are Potent Inhibitors of *Staphylococcus aureus* AgrC Quorum Sensing Receptors. *ChemBioChem*. 2017;18(4):413-423. doi:10.1002/cbic.201600516
77. Tal-Gan Y, Ivancic M, Cornilescu G, Yang T, Blackwell HE. Highly Stable, Amide-Bridged Autoinducing Peptide Analogues that Strongly Inhibit the AgrC Quorum Sensing Receptor in *Staphylococcus aureus*. *Angew Chem Int Ed Engl*. 2016;55(31):8913-8917. doi:10.1002/anie.201602974
78. Galperin MY. Structural classification of bacterial response regulators: diversity of output domains and domain combinations. *J Bacteriol*. 2006;188(12):4169-4182. doi:10.1128/JB.01887-05
79. Galperin MY. Telling Bacteria: Do Not LytTR. *Structure*. 2008;16(5):657-659. doi:10.1016/j.str.2008.04.003
80. Sidote DJ, Barbieri CM, Wu T, Stock AM. Structure of the *Staphylococcus aureus* AgrA LytTR domain bound to DNA reveals a beta fold with an unusual mode of binding. *Structure*. 2008;16(5):727-735. doi:10.1016/j.str.2008.02.011
81. Koenig RL, Ray JL, Maleki SJ, Smeltzer MS, Hurlburt BK. *Staphylococcus aureus* AgrA binding to the RNAlII-*agr* regulatory region. *J Bacteriol*. 2004;186(22):7549-7555. doi:10.1128/JB.186.22.7549-7555.2004
82. Morfeldt E, Tegmark K, Arvidson S. Transcriptional control of the *agr*-dependent virulence gene regulator, RNAlII, in *Staphylococcus aureus*. *Mol Microbiol*. 1996;21(6):1227-1237. doi:10.1046/j.1365-2958.1996.751447.x
83. Brenowitz M, Senear DF, Kingston RE. DNase I Footprint Analysis of Protein-DNA Binding. *Curr Protoc Mol Biol*. 1989;7(1):12.4.1-12.4.16. doi:https://doi.org/10.1002/0471142727.mb1204s07
84. van der Vliet PC, Verrijzer CP. Bending of DNA by transcription factors. *Bioessays*. 1993;15(1):25-32. doi:10.1002/bies.950150105
85. Rajasree K, Fasim A, Gopal B. Conformational features of the *Staphylococcus aureus* AgrA-promoter interactions rationalize quorum-sensing triggered gene expression. *Biochem Biophys Reports*. 2016;6:124-134. doi:10.1016/j.bbrep.2016.03.012



86. Stock AM, Mottonen JM, Stock JB, Schutt CE. Three-dimensional structure of CheY, the response regulator of bacterial chemotaxis. *Nature*. 1989;337(6209):745-749. doi:10.1038/337745a0
87. Stock JB, Lukat GS, Stock AM. Bacterial chemotaxis and the molecular logic of intracellular signal transduction networks. *Annu Rev Biophys Biophys Chem*. 1991;20:109-136. doi:10.1146/annurev.bb.20.060191.000545
88. Shala A, Patel KH, Golemi-Kotra D, Audette GF. Expression, purification, crystallization and preliminary X-ray analysis of the receiver domain of *Staphylococcus aureus* LytR protein. *Acta Crystallogr Sect F*. 2013;69(Pt 12):1418-1421. doi:10.1107/S1744309113030972
89. Park AK, Moon JH, Lee KS, Chi YM. Crystal structure of receiver domain of putative NarL family response regulator spr1814 from *Streptococcus pneumoniae* in the absence and presence of the phosphoryl analog beryll fluoride. *Biochem Biophys Res Commun*. 2012;421(2):403-407. doi:https://doi.org/10.1016/j.bbrc.2012.04.035
90. Boudes M, Sanchez D, Graille M, van Tilbeurgh H, Durand D, Quevillon-Cheruel S. Structural insights into the dimerization of the response regulator ComE from *Streptococcus pneumoniae*. *Nucleic Acids Res*. 2014;42(8):5302-5313. doi:10.1093/nar/gku110
91. Narayanan A, Kumar S, Evrard AN, Paul LN, Yernool DA. An asymmetric heterodomain interface stabilizes a response regulator-DNA complex. *Nat Commun*. 2014;5:1-9. doi:10.1038/ncomms4282
92. Leonard PG, Bezar IF, Sidote DJ, Stock AM. Identification of a hydrophobic cleft in the LytTR domain of AgrA as a locus for small molecule interactions that inhibit DNA binding. *Biochemistry*. 2012;51(50):10035-10043. doi:10.1021/bi3011785
93. Bezar IF, Mashruwala AA, Boyd JM, Stock AM. Drug-like Fragments Inhibit *agr*-Mediated Virulence Expression in *Staphylococcus aureus*. *Sci Rep*. 2019;9(1):1-14. doi:10.1038/s41598-019-42853-z
94. Greenberg M, Kuo D, Jankowsky E, Long L, Hager C, Bandi K, Ma D, Manoharan D, Shoham Y, Harte W, Ghannoum MA, Shoham M. Small-molecule AgrA inhibitors F12 and F19 act as antivirulence agents against Gram-positive pathogens. *Sci Rep*. 2018;8(1):1-12. doi:10.1038/s41598-018-32829-w
95. Sully EK, Malachowa N, Elmore BO, Alexander SM, Femling JK, Gray BM, DeLeo FR, Otto M, Cheung AL, Edwards BS, Sklar LA, Horswill AR, Hall PR, Gresham HD. Selective Chemical Inhibition of *agr* Quorum Sensing in *Staphylococcus aureus* Promotes Host Defense with Minimal Impact on Resistance. *PLOS Pathog*. 2014;10(6):e1004174. https://doi.org/10.1371/journal.ppat.1004174
96. Shopsin B, Eaton C, Wasserman GA, Mathema B, Adhikari RP, Agolory S, Altman DR, Holzman RS, Kreiswirth BN, Novick RP. Mutations in *agr* do not persist in natural populations of methicillin-resistant *Staphylococcus aureus*. *J Infect Dis*. 2010;202(10):1593-1599. doi:10.1086/656915



97. Khodaverdian V, Pesho M, Truitt B, Bollinger L, Patel P, Nithianantham S, Yu G, Delaney E, Jankowsky E, Shoham M. Discovery of antivirulence agents against methicillin-resistant *Staphylococcus aureus*. *Antimicrob Agents Chemother*. 2013;57(8):3645-3652. doi:10.1128/AAC.00269-13
98. Queck SY, Jameson-Lee M, Villaruz AE, Bach T-HL, Khan BA, Sturdevant DE, Ricklefs SM, Li M, Otto M. RNAIII-independent target gene control by the *agr* quorum-sensing system: insight into the evolution of virulence regulation in *Staphylococcus aureus*. *Mol Cell*. 2008;32(1):150-158. doi:10.1016/j.molcel.2008.08.005
99. Blevins JS, Beenken KE, Elasri MO, Hurlburt BK, Smeltzer MS. Strain-dependent differences in the regulatory roles of *sarA* and *agr* in *Staphylococcus aureus*. *Infect Immun*. 2002;70(2):470-480. doi:10.1128/iai.70.2.470-480.2002
100. Baba T, Takeuchi F, Kuroda M, Yuzawa H, Aoki K, Oguchi A, Nagai Y, Iwama N, Asano K, Naimi T, Kuroda H, Cui L, Yamamoto K, Hiramatsu K. Genome and virulence determinants of high virulence community-acquired MRSA. *Lancet*. 2002;359(9320):1819-1827. doi:10.1016/s0140-6736(02)08713-5
101. Somerville GA, Proctor RA. At the crossroads of bacterial metabolism and virulence factor synthesis in Staphylococci. *Microbiol Mol Biol Rev*. 2009;73(2):233-248. doi:10.1128/MMBR.00005-09
102. Wang R, Braughton KR, Kretschmer D, Bach THL, Queck SY, Li M, Kennedy AD, Dorward DW, Klebanoff SJ, Peschel A, DeLeo FR, Otto M. Identification of novel cytolytic peptides as key virulence determinants for community-associated MRSA. *Nat Med*. 2007;13(12):1510-1514. doi:10.1038/nm1656
103. Janzon L, Löfdahl S, Arvidson S. Identification and nucleotide sequence of the delta-lysin gene, *hld*, adjacent to the accessory gene regulator (*agr*) of *Staphylococcus aureus*. *Mol Gen Genet MGG*. 1989;219(3):480-485. doi:10.1007/BF00259623
104. Huntzinger E, Boisset S, Saveanu C, Benito Y, Geissmann T, Namane A, Lina G, Etienne J, Ehresmann B, Ehresmann C, Jacquier A, Vandenesch F, Romby P. *Staphylococcus aureus* RNAIII and the endoribonuclease III coordinately regulate *spa* gene expression. *EMBO J*. 2005;24(4):824-835. doi:10.1038/sj.emboj.7600572
105. Boisset S, Geissmann T, Huntzinger E, Fechter P, Bendridi N, Possedko M, Chevalier C, Helfer AC, Benito Y, Jacquier A, Gaspin C, Vandenesch F, Romby P. *Staphylococcus aureus* RNAIII coordinately represses the synthesis of virulence factors and the transcription regulator Rot by an antisense mechanism. *Genes Dev*. 2007;21(11):1353-1366. doi:10.1101/gad.423507
106. Mcnamara PJ, Milligan-Monroe KC, Khalili S, Proctor RA. Identification, cloning, and initial characterization of *rot*, a locus encoding a regulator of virulence factor expression in *Staphylococcus aureus*. *J Bacteriol*. 2000;182(11):3197-3203. doi:10.1128/JB.182.11.3197-3203.2000



107. Saïd-Salim B, Dunman PM, McAleese FM, Macapagal D, Murphy E, McNamara PJ, Arvidson S, Foster TJ, Projan SJ, Kreiswirth BN. Global regulation of *Staphylococcus aureus* genes by Rot. *J Bacteriol.* 2003;185(2):610-619. doi:10.1128/JB.185.2.610-619.2003
108. Dunman PM, Murphy E, Haney S, Palacios D, Tucker-Kellogg G, Wu S, Brown EL, Zagursky RJ, Shlaes D, Projan SJ. Transcription profiling-based identification of *Staphylococcus aureus* genes regulated by the *agr* and/or *sarA* loci. *J Bacteriol.* 2001;183(24):7341-7353. doi:10.1128/JB.183.24.7341-7353.2001
109. Cheung AL, Eberhardt KJ, Chung E, Yeaman MR, Sullam PM, Ramos M, Bayer AS. Diminished virulence of a *sar-lagr-* mutant of *Staphylococcus aureus* in the rabbit model of endocarditis. *J Clin Invest.* 1994;94(5):1815-1822. doi:10.1172/JCI117530
110. McNamara PJ, Bayer AS. A rot mutation restores parental virulence to an *agr*-null *Staphylococcus aureus* strain in a rabbit model of endocarditis. *Infect Immun.* 2005;73(6):3806-3809. doi:10.1128/IAI.73.6.3806-3809.2005
111. Tegmark K, Morfeldt E, Arvidson S. Regulation of *agr*-dependent virulence genes in *Staphylococcus aureus* by RNAIII from coagulase-negative staphylococci. *J Bacteriol.* 1998;180(12):3181-3186. doi:10.1128/jb.180.12.3181-3186.1998
112. Benito Y, Lina G, Greenland T, Etienne J, Vandenesch F. trans-Complementation of a *Staphylococcus aureus agr* mutant by *Staphylococcus lugdunensis agr* RNAIII. *J Bacteriol.* 1998;180(21):5780-5783. doi:10.1128/jb.180.21.5780-5783.1998
113. Dinges MM, Orwin PM, Schlievert PM. Exotoxins of *Staphylococcus aureus*. *Clin Microbiol Rev.* 2000;13(1):16-34. doi:10.1128/cmr.13.1.16-34.2000
114. Valeva A, Palmer M, Bhakdi S. Staphylococcal alpha-toxin: formation of the heptameric pore is partially cooperative and proceeds through multiple intermediate stages. *Biochemistry.* 1997;36(43):13298-13304. doi:10.1021/bi971075r
115. Huntzinger E, Boisset S, Saveanu C, Benito Y, Geissmann T, Namane A, Lina G, Etienne J, Ehresmann B, Ehresmann C, Jacquier A, Vandenesch F, Romby P. *Staphylococcus aureus* RNAIII and the endoribonuclease III coordinately regulate *spa* gene expression. *EMBO J.* 2005;24(4):824-835. doi:10.1038/sj.emboj.7600572
116. Morfeldt E, Taylor D, von Gabain A, Arvidson S. Activation of alpha-toxin translation in *Staphylococcus aureus* by the trans-encoded antisense RNA, RNAIII. *EMBO J.* 1995;14(18):4569-4577.
117. Yarwood JM, McCormick JK, Schlievert PM. Identification of a novel two-component regulatory system that acts in global regulation of virulence factors of *Staphylococcus aureus*. *J Bacteriol.* 2001;183(4):1113-1123. doi:10.1128/JB.183.4.1113-1123.2001
118. Liang X, Zheng L, Landwehr C, Lunsford D, Holmes D, Ji Y. Global regulation of gene expression by ArlRS, a two-component signal transduction regulatory system of *Staphylococcus aureus*. *J Bacteriol.* 2005;187(15):5486-5492. doi:10.1128/jb.187.15.5486-5492.2005



119. Schmidt KA, Manna AC, Gill S, Cheung AL. SarT, a repressor of alpha-hemolysin in *Staphylococcus aureus*. *Infect Immun*. 2001;69(8):4749-4758. doi:10.1128/IAI.69.8.4749-4758.2001
120. Manna AC, Cheung AL. Expression of SarX, a negative regulator of *agr* and exoprotein synthesis, is activated by MgrA in *Staphylococcus aureus*. *J Bacteriol*. 2006;188(12):4288-4299. doi:10.1128/JB.00297-06
121. Cheung AL, Koomey JM, Butler CA, Projan SJ, Fischetti VA. Regulation of exoprotein expression in *Staphylococcus aureus* by a locus (*sar*) distinct from *agr*. *Proc Natl Acad Sci*. 1992;89(14):6462-6466. doi:10.1073/pnas.89.14.6462
122. Chien Y, Cheung AL. Molecular interactions between two global regulators, *sar* and *agr*, in *Staphylococcus aureus*. *J Biol Chem*. 1998;273(5):2645-2652. doi:10.1074/jbc.273.5.2645
123. Rehtin TM, Gillaspay AF, Schumacher MA, Brennan RG, Smeltzer MS, Hurlburt BK. Characterization of the SarA virulence gene regulator of *Staphylococcus aureus*. *Mol Microbiol*. 1999;33(2):307-316. doi:10.1046/j.1365-2958.1999.01474.x
124. Chien Y, Manna AC, Projan SJ, Cheung AL. SarA, a global regulator of virulence determinants in *Staphylococcus aureus*, binds to a conserved motif essential for *sar*-dependent gene regulation. *J Biol Chem*. 1999;274(52):37169-37176. doi:10.1074/jbc.274.52.37169
125. Schumacher MA, Hurlburt BK, Brennan RG. Crystal structures of SarA, a pleiotropic regulator of virulence genes in *S. aureus*. *Nature*. 2001;409(6817):215-219. doi:10.1038/35051623
126. Nicod SS, Weinzierl ROJ, Burchell L, Escalera-Maurer A, James EH, Wigneshweraraj S. Systematic mutational analysis of the LytTR DNA binding domain of *Staphylococcus aureus* virulence gene transcription factor AgrA. *Nucleic Acids Res*. 2014;42(20):12523-12536. doi:10.1093/nar/gku1015
127. Guédon E, Serror P, Ehrlich SD, Renault P, Delorme C. Pleiotropic transcriptional repressor CodY senses the intracellular pool of branched-chain amino acids in *Lactococcus lactis*. *Mol Microbiol*. 2001;40(5):1227-1239. doi:https://doi.org/10.1046/j.1365-2958.2001.02470.x
128. Majerczyk CD, Sadykov MR, Luong TT, Lee C, Somerville GA, Sonenshein AL. *Staphylococcus aureus* CodY negatively regulates virulence gene expression. *J Bacteriol*. 2008;190(7):2257-2265. doi:10.1128/JB.01545-07
129. Majerczyk CD, Dunman PM, Luong TT, Lee CY, Sadykov MR, Somerville GA, Bodi K, Sonenshein AL. Direct Targets of CodY in *Staphylococcus aureus*. *J Bacteriol*. 2010;192(11):2861-2877. doi:10.1128/JB.00220-10
130. den Hengst CD, van Hijum SAFT, Geurts JMW, Nauta A, Kok J, Kuipers OP. The *Lactococcus lactis* CodY regulon: identification of a conserved cis-regulatory element. *J Biol Chem*. 2005;280(40):34332-34342. doi:10.1074/jbc.M502349200



131. Belitsky BR, Sonenshein AL. Genetic and Biochemical Analysis of CodY-Binding Sites in *Bacillus subtilis*. *J Bacteriol.* 2008;190(4):1224-1236. doi:10.1128/JB.01780-07
132. Ma C, Yang X, Lewis PJ. Bacterial Transcription as a Target for Antibacterial Drug Development. *Microbiol Mol Biol Rev.* 2016;80(1):139-160. doi:10.1128/MMBR.00055-15
133. Santos R, Ursu O, Gaulton A, Bento AP, Donadi RS, Bologa CG, Karlsson A, Al-Lazikani B, Hersey A, Oprea TI, Overington JP. A comprehensive map of molecular drug targets. *Nat Rev Drug Discov.* 2016;16(1):19-34. doi:10.1038/nrd.2016.230
134. White SH. Biophysical dissection of membrane proteins. *Nature.* 2009;459(7245):344-346. doi:10.1038/nature08142
135. Marroquin S, Gimza B, Tomlinson B, Stein M, Frey A, Keogh RA, Zapf R, Todd DA, Cech NB, Carroll RK, Shaw LN. MroQ Is a Novel Abi-Domain Protein That Influences Virulence Gene Expression in *Staphylococcus aureus* via Modulation of *Agr* Activity. Freitag NE, ed. *Infect Immun.* 2019;87(5):e00002-19. doi:10.1128/IAI.00002-19
136. Bronner S, Monteil H, Prévost G. Regulation of virulence determinants in *Staphylococcus aureus*: complexity and applications. *FEMS Microbiol Rev.* 2004;28(2):183-200. doi:10.1016/j.femsre.2003.09.003
137. Cheung AL, Eberhardt KJ, Chung E, Yeaman MR, Sullam PM, Ramos M, Bayer AS. Diminished virulence of a *sar*-/*agr*- mutant of *Staphylococcus aureus* in the rabbit model of endocarditis. *J Clin Invest.* 1994;94(5):1815-1822. doi:10.1172/JCI117530
138. Kolár M, Urbánek K, Látal T. Antibiotic selective pressure and development of bacterial resistance. *Int J Antimicrob Agents.* 2001;17(5):357-363. doi:10.1016/s0924-8579(01)00317-x
139. Kepler J. *Strena Seu de Nive Sexangula*. 1st ed. Godefrid Tampach; 1611.
140. Miller WH. *A Treatise on Crystallography*. 1st ed. Deighton; 1839.
141. Röntgen WC. Ueber eine neue Art von Strahlen. *Sitzungsberichte der Phys Gesellschaft zu Wuerzbg.* 1895;29:132-141.
142. Bragg WH, Bragg WL. The reflection of X-rays by crystals. *Proc R Soc London.* 1913;88(605):428-438. doi:10.1098/rspa.1913.0040
143. Glazer AM. *Crystallography: A Very Short Introduction*. Oxford University Press; 2016. doi:10.1093/actrade/9780198717591.001.0001
144. Curiel E. The many definitions of a black hole. *Nat Astron.* 2019;3(1):27-34. doi:10.1038/s41550-018-0602-1



145. Rayleigh, Lord. X. On the electromagnetic theory of light. *Philos Mag.* 1881;12(73):81-101. doi:10.1080/14786448108627074
146. Strutt JW. LVIII. On the scattering of light by small particles. *Philos Mag.* 1871;41(275):447-454. doi:10.1080/14786447108640507
147. Mie G. Beiträge zur Optik trüber Medien, speziell kolloidaler Metallösungen. *Ann Phys.* 1908;330(3):377-445. doi:https://doi.org/10.1002/andp.19083300302
148. Lamb TD, Collin SP, Pugh Jr EN. Evolution of the vertebrate eye: opsins, photoreceptors, retina and eye cup. *Nat Rev Neurosci.* 2007;8(12):960-976. doi:10.1038/nrn2283
149. Strutt JW. XV. On the light from the sky, its polarization and colour. *Philos Mag.* 1871;41(271):107-120. doi:10.1080/14786447108640452
150. Compton AH. A Quantum Theory of the Scattering of X-rays by Light Elements. *Phys Rev.* 1923;21(5):483-502. doi:10.1103/PhysRev.21.483
151. Sliney DH. What is light? The visible spectrum and beyond. *Eye.* 2016;30(2):222-229. doi:10.1038/eye.2015.252
152. Speight J. *Lange's Handbook of Chemistry*. 17th ed. McGraw-Hill Education; 2017.
153. Bracewell RN. *The Fourier Transform and Its Applications*. 3rd ed. McGraw-Hill Education; 1986.
154. Taylor G. The phase problem. *Acta Crystallogr Sect D.* 2003;59(11):1881-1890. doi:10.1107/S0907444903017815
155. Polycarpou C, Cassemiro KN, Venturi G, Zavatta A, Bellini M. Adaptive Detection of Arbitrarily Shaped Ultrashort Quantum Light States. *Phys Rev Lett.* 2012;109(5):53602. doi:10.1103/PhysRevLett.109.053602
156. Monroe D. Measuring the shape of a photon. *Physics.* 2012;5(86).
157. Maxwell JC. *A Treatise on Electricity and Magnetism*. 1st ed. Clarendon press; 1873.
158. Becker U. Matter-wave interference made clear. *Nature.* 2011;474(7353):586-587. doi:10.1038/474586a
159. Hook JR, Hall HE. *Solid State Physics*. 2nd ed. Wiley; 1991.
160. Cowtan K. Phase Problem in X-ray Crystallography, and Its Solution. *Encycl Life Sci.* Published online 2003:1-5. doi:10.1038/npg.els.0002722
161. Rhodes G. *Crystallography Made Crystal Clear*. 3rd ed. Academic Press; 2006.



162. Wynne SA, Crowther RA, Leslie AGW. The Crystal Structure of the Human Hepatitis B Virus Capsid no obvious change in the morphology of the particle (Zlotnick et al The structure of the T 4 capsid containing 240 cop. *Mol Cell*. 1999;3:771-780.
163. Coxeter HSM. *Introduction to Geometry*. 2nd ed. Wiley Classics Library; 1889.
164. Dana JD. *Dana's Manual of Mineralogy*. 17th ed. Chapman and Hall; 1959.
165. Cotton AF. *Chemical Applications of Group Theory*. 3rd ed. Wiley; 1990.
166. Wukovitz SW, Yeates TO. Why protein crystals favour some space-groups over others. *Nat Struct Biol*. 1995;2(12):1062-1067. doi:10.1038/nsb1295-1062
167. Friedel G. Sur les symétries cristallines que peut révéler la diffraction des rayons Röntgen. *Comptes Rendus*. 1913;157:1533-1536.
168. Laue M von. Kritische Bemerkungen zu den Deutungen der Photogramme von Friederich und Knipping. *Phys Zeitschrift*. 1913;14:421-423.
169. Harburn G, Taylor CA, Welberry TR. *Atlas of Optical Transforms*. 1st ed. G. Bell and Sons; 1975.
170. Ewald PP. Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Ann Phys*. 1920;369(3):253-287.
171. Ewald PP. Introduction to the dynamical theory of X-ray diffraction. *Acta Crystallogr Sect A*. 1969;25(1):103-108. doi:10.1107/S0567739469000155
172. McKie D, McKie C. *Essentials of Crystallography*. Blackwell Scientific Publications; 1992.
173. Prince E, ed. *International Tables for Crystallography. Volume C, Mathematical, Physical and Chemical Tables*. Kluwer Academic Publishers; 2004. <https://search.library.wisc.edu/catalog/999703926102121>
174. Deybe P. Interferenz von Röntgenstrahlen und Wärmebewegung. *Ann Phys*. 1913;348(1):49-92.
175. Carugo O. How large B-factors can be in protein crystal structures. *BMC Bioinformatics*. 2018;19(1):61. doi:10.1186/s12859-018-2083-8.
176. Hauptman H. The Direct Methods of X-ray Crystallography. *Science*. 1986;233(4760):178-183. doi:10.1126/science.233.4760.178
177. Sheldrick GM. A short history of *SHELX*. *Acta Crystallogr Sect A*. 2008;64(1):112-122. doi:10.1107/S0108767307043930



178. Patterson AL. A Direct Method for the Determination of the Components of Interatomic Distances in Crystals. *Zeitschrift für Krist - Cryst Mater.* 90(1):517-542. doi:<https://doi.org/10.1524/zkri.1935.90.1.517>
179. Oszlányi G, Süto A. *Ab initio* structure solution by charge flipping. *Acta Crystallogr Sect A.* 2004;60(2):134-141. doi:10.1107/S0108767303027569
180. Sheldrick GM. *SHELXT* - Integrated space-group and crystal-structure determination. *Acta Crystallogr Sect A.* 2015;71(1):3-8. doi:10.1107/S2053273314026370
181. Navaza J. AMoRe: an automated package for molecular replacement. *Acta Crystallogr Sect A.* 1994;50(2):157-163. doi:10.1107/S0108767393007597
182. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. *PHASER* crystallographic software. *J Appl Crystallogr.* 2007;40(4):658-674. doi:10.1107/S0021889807021206
183. Hendrickson WA. Anomalous diffraction in crystallographic phase evaluation. *Q Rev Biophys.* 2014;47(1):49-93. doi:10.1017/S0033583514000018
184. Barkla CG. XXXIX. The spectra of the fluorescent Röntgen radiations. *Philos Mag.* 1911;22(129):396-412. doi:10.1080/14786440908637137
185. Glocker R, Schreiber H. Quantitative Röntgenspektralanalyse mit Kalterregung des Spektrums. *Ann Phys.* 1928;390(8):1089-1102. doi:<https://doi.org/10.1002/andp.19283900805>
186. Sliz P, Harrison SC, Rosenbaum G. How does Radiation Damage in Protein Crystals Depend on X-Ray Dose? *Structure.* 2003;11(1):13-19. doi:[https://doi.org/10.1016/S0969-2126\(02\)00910-3](https://doi.org/10.1016/S0969-2126(02)00910-3)
187. Sheldrick GM. Experimental phasing with *SHELXC / D / E*: combining chain tracing with density modification. *Acta Crystallogr Sect D.* 2010;66(4):479-485. doi:10.1107/S0907444909038360
188. Read RJ. Pushing the boundaries of molecular replacement with maximum likelihood. *Acta Crystallogr Sect D.* 2001;57(10):1373-1382. doi:10.1107/S0907444901012471
189. Tronrud DE. Introduction to macromolecular refinement. *Acta Crystallogr Sect D.* 2004;60(12 Part 1):2156-2168. doi:10.1107/S090744490402356X
190. Wlodawer A, Minor W, Dauter Z, Jaskolski M. Protein crystallography for aspiring crystallographers or how to avoid pitfalls and traps in macromolecular structure determination. *FEBS J.* 2013;280(22):5705-5736. doi:10.1111/febs.12495
191. Jansen J, Tang D, Zandbergen HW, Schenk H. *MSLS*, a Least-Squares Procedure for Accurate Crystal Structure Refinement from Dynamical Electron Diffraction Patterns. *Acta Crystallogr Sect A.* 1998;54(1):91-101. doi:10.1107/S0108767397010489



192. Pannu NS, Read RJ. Improved Structure Refinement Through Maximum Likelihood. *Acta Crystallogr Sect A*. 1996;52(5):659-668. doi:10.1107/S0108767396004370
193. Murshudov GN, Skubák P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, Winn MD, Long F, Vagin AA. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr Sect D*. 2011;67(4):355-367. doi:10.1107/S0907444911001314
194. Kovalevskiy O, Nicholls RA, Murshudov GN. Automated refinement of macromolecular structures at low resolution using prior information. *Acta Crystallogr Sect D*. 2016;72:1149-1161. doi:10.1107/S2059798316014534
195. Immirzi A. Constraints and restraints in crystal structure analysis. *J Appl Crystallogr*. 2009;42(Pt 2):362-364. doi:10.1107/S0021889808044142
196. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of *Coot*. *Acta Crystallogr Sect D*. 2010;66(4):486-501. doi:10.1107/S0907444910007493
197. Liebschner D, Afonine PV, Baker ML, Bunkóczi G, Chen VB, Croll, TI, Hintze BJ, Hung LW, Jain S, McCoy AJ, Moriarty NW, Oeffner RD, Poon BK, Prisant MG, Read RJ, Richardson JS, Richardson DC, Sammito MD, Sobolev OV, Stockwell DH, Terwilliger TC, Urzhumstev AG, Videau LL, Williams CJ, Adams PD. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in *Phenix*. *Acta Crystallogr Sect D*. 2019;75(10):861-877. doi:10.1107/S2059798319011471
198. Moriarty NW, Grosse-Kunstleve RW, Adams PD. electronic Ligand Builder and Optimization Workbench (*eLBOW*): a tool for ligand coordinate and restraint generation. *Acta Crystallogr Sect D*. 2009;65(10):1074-1080. doi:10.1107/S0907444909029436
199. Evans PR, Murshudov GN. How good are my data and what is the resolution? *Acta Crystallogr Sect D*. 2013;69(7):1204-1214. doi:10.1107/S0907444913000061
200. Groom CR, Bruno IJ, Lightfoot MP, Ward SC. The Cambridge Structural Database. *Acta Crystallogr Sect B*. 2016;72(2):171-179. doi:10.1107/S2052520616003954
201. Safaei M, Foroughi MM, Ebrahimipour N, Jahani S, Omid A, Khatami M. A review on metal-organic frameworks: Synthesis and applications. *TrAC Trends Anal Chem*. 2019;118:401-425. doi:https://doi.org/10.1016/j.trac.2019.06.007
202. Weckhuysen BM, Yu J. Recent advances in zeolite chemistry and catalysis. *Chem Soc Rev*. 2015;44(20):7022-7024. doi:10.1039/C5CS90100F
203. Wang J. Estimation of the quality of refined protein crystal structures. *Protein Sci*. 2015;24(5):661-669. doi:10.1002/pro.2639
204. Hodel A, Kim S-H, Brünger AT. Model bias in macromolecular crystal structures. *Acta Crystallogr Sect A*. 1992;48(6):851-858. doi:10.1107/S0108767392006044



205. Brünger AT. Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*. 1992;355(6359):472-475. doi:10.1038/355472a0
206. Terwilliger TC, Grosse-Kunstleve RW, Afonine P V, Moriarty NW, Adams PD, Read RJ, Zwart PH, Hung L-W. Iterative-build OMIT maps: map improvement by iterative model building and refinement without model bias. *Acta Crystallogr Sect D*. 2008;64(Pt 5):515-524. doi:10.1107/S0907444908004319
207. Takara. pCold DNA cold-shock expression system. Published online 2020. <https://www.takarabio.com/products/protein-research/expression-vectors-and-systems/protein-folding-kits/pcold-dna-cold-shock>
208. Sun F, Liang H, Kong X, Xie S, Cho H, Deng X, Ji Q, Zhang H, Alvarez S, Hicks LM, Bae T, Luo C, Jiang H, He C. Quorum-sensing *agr* mediates bacterial oxidation response via an intramolecular disulfide redox switch in the response regulator AgrA. *Proc Natl Acad Sci*. 2012;109(23):9095-9100. doi:10.1073/pnas.1200603109
209. Robichon C, Luo J, Causey TB, Benner JS, Samuelson JC. Engineering *Escherichia coli* BL21(DE3) Derivative Strains To Minimize *E. coli* Protein Contamination after Purification by Immobilized Metal Affinity Chromatography. *Appl Environ Microbiol*. 2011;77(13):4634-4646. doi:10.1128/AEM.00119-11
210. Bolanos-Garcia VM, Davies OR. Structural analysis and classification of native proteins from *E. coli* commonly co-purified by immobilised metal affinity chromatography. *Biochim Biophys Acta*. 2006;1760(9):1304-1313. doi:10.1016/j.bbagen.2006.03.027
211. Gao R, Bouillet S, Stock AM. Structural Basis of Response Regulator Function. *Annu Rev Microbiol*. 2019;73(1):175-197. doi:10.1146/annurev-micro-020518-115931
212. Cho H, Wang W, Kim R, Yokota H, Damo S, Kim SH, Wemmer D, Kustu S, Yan D. BeF(3)(-) acts as a phosphate analog in proteins phosphorylated on aspartate: structure of a BeF(3)(-) complex with phosphoserine phosphatase. *Proc Natl Acad Sci U S A*. 2001;98(15):8525-8530. doi:10.1073/pnas.131213698
213. Stetefeld J, McKenna SA, Patel TR. Dynamic light scattering: a practical guide and applications in biomedical sciences. *Biophys Rev*. 2016;8(4):409-427. doi:10.1007/s12551-016-0218-6
214. Kikhney AG, Svergun DI. A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. *FEBS Lett*. 2015;589(19PartA):2570-2577. doi:https://doi.org/10.1016/j.febslet.2015.08.027
215. Brakke MK. Density Gradient Centrifugation: A New Separation Technique. *J Am Chem Soc*. 1951;73(4):1847-1848. doi:10.1021/ja01148a508
216. Raghunathan K, Harris PT, Arvidson DN. Trial by fire: are the crystals macromolecules? *Acta Crystallogr Sect F*. 2010;66(Pt 5):615-620. doi:10.1107/S1744309110012078



217. Durst RA, Staples BR. Tris/Tris·HCl: A Standard Buffer for Use in the Physiologic pH Range. *Clin Chem*. 1972;18(3):206-208. doi:10.1093/clinchem/18.3.206
218. Hou H, Shi M, Hu S-Y, Ahmad F, Zhang B, Chen Z-H, Yin D-C. A systematic comparison of sitting and hanging-drop crystallization using traditional and cross-diffusion microbatch crystallization plates. *J Cryst Growth*. 2019;521:1-8. doi:https://doi.org/10.1016/j.jcrysgro.2019.05.011
219. Bergfors T. Seeds to crystals. *J Struct Biol*. 2003;142(1):66-76. doi:10.1016/s1047-8477(03)00039-x
220. Sousa R. Use of glycerol, polyols and other protein structure stabilizing agents in protein crystallization. *Acta Crystallogr Sect D*. 1995;51(Pt 3):271-277. doi:10.1107/S0907444994014009
221. Miermont A, Waharte F, Hu S, McClean MN, Bottani S, Léon S, Hersen P. Severe osmotic compression triggers a slowdown of intracellular signaling, which can be explained by molecular crowding. *Proc Natl Acad Sci U S A*. 2013;110(14):5725-5730. doi:10.1073/pnas.1215367110
222. Oganesyanyan N, Ankoudinova I, Kim S-H, Kim R. Effect of osmotic stress and heat shock in recombinant protein overexpression and crystallization. *Protein Expr Purif*. 2007;52(2):280-285. doi:10.1016/j.pep.2006.09.015
223. Volkman BF, Lipson D, Wemmer DE, Kern D. Two-state allosteric behavior in a single-domain signaling protein. *Science*. 2001;291(5512):2429-2433. doi:10.1126/science.291.5512.2429
224. Corrêa F, Gardner KH. Basis of Mutual Domain Inhibition in a Bacterial Response Regulator. *Cell Chem Biol*. 2016;23(8):945-954. doi:10.1016/j.chembiol.2016.07.010
225. Siam R, Marczyński GT. Glutamate at the phosphorylation site of response regulator CtrA provides essential activities without increasing DNA binding. *Nucleic Acids Res*. 2003;31(6):1775-1779. doi:10.1093/nar/gkg271
226. Lan CY, Igo MM. Differential expression of the OmpF and OmpC porin proteins in *Escherichia coli* K-12 depends upon the level of active OmpR. *J Bacteriol*. 1998;180(1):171-174. doi:10.1128/JB.180.1.171-174.1998
227. Klose KE, Weiss DS, Kustu S. Glutamate at the site of phosphorylation of nitrogen-regulatory protein NTRC mimics aspartyl-phosphate and activates the protein. *J Mol Biol*. 1993;232(1):67-78. doi:10.1006/jmbi.1993.1370
228. Mesmer RE, Baes CF. Fluoride complexes of beryllium(II) in aqueous media. *Inorg Chem*. 1969;8(3):618-626. doi:10.1021/ic50073a042
229. Jin Y, Molt RW, Blackburn GM. Metal Fluorides: Tools for Structural and Computational Analysis of Phosphoryl Transfer Enzymes. *Top Curr Chem*. 2017;375(2):36. doi:10.1007/s41061-017-0130-y



230. Hastings CA, Lee S-Y, Cho HS, Yan D, Kustu S, Wemmer DE. High-Resolution Solution Structure of the Beryll fluoride-Activated NtrC Receiver Domain. *Biochemistry*. 2003;42(30):9081-9090. doi:10.1021/bi0273866
231. Lan T-Y, Li P-N, Lee T-K. Method to enhance the resolution of x-ray coherent diffraction imaging for non-crystalline bio-samples. *New J Phys*. 2014;16(3):33016. doi:10.1088/1367-2630/16/3/033016
232. Matthews BW. Solvent content of protein crystals. *J Mol Biol*. 1968;33(2):491-497. doi:10.1016/0022-2836(68)90205-2
233. Kantardjieff KA, Rupp B. Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. *Protein Sci*. 2003;12(9):1865-1871. doi:10.1110/ps.0350503
234. Weichenberger CX, Afonine P V, Kantardjieff K, Rupp B. The solvent component of macromolecular crystals. *Acta Crystallogr D Biol Crystallogr*. 2015;71(Pt 5):1023-1038. doi:10.1107/S1399004715006045
235. Bellamy HD, Snell EH, Lovelace J, Pokross M, Borgstahl GEO. The high-mosaicity illusion: Revealing the true physical characteristics of macromolecular crystals. *Acta Crystallogr Sect D*. 2000;56(8):986-995. doi:10.1107/S0907444900007356
236. Boese R, Antipin MY, Bläser D, Lyssenko KA. Molecular Crystal Structure of Acetylacetone at 210 and 110 K: Is the Crystal Disorder Static or Dynamic? *J Phys Chem B*. 1998;102(44):8654-8660. doi:10.1021/jp980121+
237. Thorn A, Parkhurst J, Emsley P, Nicholls RA, Vollmar M, Evans G, Murshudov GN. AUSPEX: a graphical tool for X-ray diffraction data analysis. *Acta Crystallogr Sect D*. 2017;73(Pt 9):729-737. doi:10.1107/S205979831700969X
238. Garman EF, Owen RL. Cryocooling and radiation damage in macromolecular crystallography. *Acta Crystallogr Sect D*. 2006;62(1):32-47. doi:10.1107/S0907444905034207
239. Mitchell EP, Garman EF. Flash freezing of protein crystals: investigation of mosaic spread and diffraction limit with variation of cryoprotectant concentration. *J Appl Crystallogr*. 1994;27(6):1070-1074. doi:10.1107/S0021889894008629
240. Senda M, Hayashi T, Hatakeyama M, Takeuchi K, Sasaki AT, Senda T. Use of Multiple Cryoprotectants to Improve Diffraction Quality from Protein Crystals. *Cryst Growth Des*. 2016;16(3):1565-1571. doi:10.1021/acs.cgd.5b01692
241. Oeffner RD, Afonine P V, Millán C, Sammito M, Usón I, Read RJ, McCoy AJ. On the application of the expected log-likelihood gain to decision making in molecular replacement. *Acta Crystallogr Sect D*. 2018;74(Pt 4):245-255. doi:10.1107/S2059798318004357



242. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc.* 2015;10(6):845-858. doi:10.1038/nprot.2015.053
243. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990;215(3):403-410. doi:10.1016/S0022-2836(05)80360-2
244. Schödinger LLC. The PyMOL Molecular Graphics System. Published online 2010.
245. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem.* 2004;25(13):1605-1612. doi:10.1002/jcc.20084
246. Berg JM, Tymoczko JL, Stryer L. *Biochemistry.* 5th ed. W H Freeman; 2002.
247. Vonrhein C, Flensburg C, Keller P, Sharff A, Smart O, Paciorek W, Womack T, Bricogne G. Data processing and analysis with the *autoPROC* toolbox. *Acta Crystallogr Sect D.* 2011;67(4):293-302. doi:10.1107/S0907444911007773
248. Beilsten-Edmands J, Winter G, Gildea R, Parkhurst J, Waterman D, Evans G. Scaling diffraction data in the DIALS software package: algorithms and new approaches for multi-crystal scaling. *Acta Crystallogr Sect D.* 2020;76(4):385-399. doi:10.1107/S2059798320003198
249. Evans P. Scaling and assessment of data quality. *Acta Crystallogr Sect D.* 2006;62(1):72-82. doi:10.1107/S0907444905036693
250. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evens PR, Keegan RM, Krissinel EB, Leslie AGW, McCoy A, McNicholas SJ, Murhsudov GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A, Wilson KS. Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr.* 2011;67(Pt 4):235-242. doi:10.1107/S0907444910045749
251. Tickle IJ, Flensburg C, Keller P, Paciorek W, Sharf A, Vonrhein C, Bricogne G. STARANISO. (<http://staraniso.globalphasing.org/cgi-bin/staraniso.cgi>). Cambridge, United Kingdom: Global Phasing Ltd.
252. He X, Wang L, Wang S. Structural basis of DNA sequence recognition by the response regulator PhoP in *Mycobacterium tuberculosis*. *Sci Rep.* 2016;6(1):24442. doi:10.1038/srep24442
253. Hoogsteen K. The crystal and molecular structure of a hydrogen-bonded complex between 1-methylthymine and 9-methyladenine. *Acta Crystallogr.* 1963;16(9):907-916. doi:10.1107/S0365110X63002437
254. Volz K, Matsumura P. Crystal structure of *Escherichia coli* CheY refined at 1.7 Å resolution. *J Biol Chem.* 1991;266(23):15511-15519. doi:10.2210/pdb3chy/pdb



255. Das S, Ramakumar S, Pal D. Identifying functionally important cis-peptide containing segments in proteins and their utility in molecular function annotation. *FEBS J*. 2014;281(24):5602-5621. doi:<https://doi.org/10.1111/febs.13100>
256. Wedemeyer WJ, Welker E, Narayan M, Scheraga HA. Disulfide Bonds and Protein Folding. *Biochemistry*. 2000;39(15):4207-4216. doi:10.1021/bi992922o
257. Wemmer DE, Kern D. Beryllofluoride binding mimics phosphorylation of aspartate in response regulators. *J Bacteriol*. 2005;187(24):8229-8230. doi:10.1128/JB.187.24.8229-8230.2005
258. Yan D, Cho HS, Hastings CA, Igo MM, Lee SY, Pelton JG, Stewart V, Wemmer DE, Kustu S. Beryllofluoride mimics phosphorylation of NtrC and other bacterial response regulators. *Proc Natl Acad Sci U S A*. 1999;96(26):14789-14794. doi:10.1073/pnas.96.26.14789
259. Arnott S, Chandrasekaran R, Birdsall DL, Leslie AG, Ratliff RL. Left-handed DNA helices. *Nature*. 1980;283(5749):743-745. doi:10.1038/283743a0
260. Mohr CD, Martin DW, Konyecsni WM, Govan JR, Lory S, Deretic V. Role of the far-upstream sites of the *algD* promoter and the *algR* and *rpoN* genes in environmental modulation of mucoidy in *Pseudomonas aeruginosa*. *J Bacteriol*. 1990;172(11):6576-6580. doi:10.1128/jb.172.11.6576-6580.1990
261. Birck C, Mourey L, Gouet P, Fabry B, Schumacher J, Rousseau P, Kahn D, Samama JP. Conformational changes induced by phosphorylation of the FixJ receiver domain. *Structure*. 1999;7(12):1505-1515. doi:10.1016/S0969-2126(00)88341-0
262. Wallace AC, Laskowski RA, Thornton JM. LIGPLOT: a program to generate schematic diAgrams of protein-ligand interactions. *Protein Eng*. 1995;8(2):127-134. doi:10.1093/protein/8.2.127
263. Krissinel E, Henrick K. Inference of macromolecular assemblies from crystalline state. *J Mol Biol*. 2007;372(3):774-797. doi:10.1016/j.jmb.2007.05.022
264. Krissinel E. Crystal contacts as nature's docking solutions. *J Comput Chem*. 2010;31(1):133-143. doi:10.1002/jcc.21303
265. Kozakov D, Hall DR, Xia B, Porter KA, Padhorny D, Yueh C, Beglov D, Vajda S. The ClusPro web server for protein-protein docking. *Nat Protoc*. 2017;12(2):255-278. doi:10.1038/nprot.2016.169
266. Vajda S, Yueh C, Beglov D, Bohnuud T, Mottarella SE, Xia B, Hall DR, Kozakov D. New additions to the ClusPro server motivated by CAPRI. *Proteins*. 2017;85(3):435-444. doi:10.1002/prot.25219
267. Kozakov D, Beglov D, Bohnuud T, Mottarella SE, Xia B, Hall DR, Vajda S. How good is automated protein docking? *Proteins*. 2013;81(12):2159-2166. doi:10.1002/prot.24403



268. Vekilov PG, Feeling-Taylor AR, Yau ST, Petsev D. Solvent entropy contribution to the free energy of protein crystallization. *Acta Crystallogr D Biol Crystallogr*. 2002;58(Pt 10 Pt 1):1611-1616. doi:10.1107/s0907444902014312
269. Vekilov PG. Solvent entropy effects in the formation of protein solid phases. *Methods Enzymol*. 2003;368:84-105. doi:10.1016/S0076-6879(03)68006-7
270. Williams CJ, Headd JJ, Moriarty NW, Prisant MG, Videau LL, Dies LN, Verma V, Keedy DA, Hintze BJ, Chen VB, Jain S, Lewis SM, Arendall 3rd WB, Snoeyink J, Adams PD, Lovell SC, Richardson JS, Richardson DC. MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci*. 2018;27(1):293-315. doi:10.1002/pro.3330
271. Ramachandran GN, Ramakrishnan C, Sasisekharan V. Stereochemistry of polypeptide chain configurations. *J Mol Biol*. 1963;7(1):95-99. doi:https://doi.org/10.1016/S0022-2836(63)80023-6
272. Krieger F, Möglich A, Kiefhaber T. Effect of Proline and Glycine Residues on Dynamics and Barriers of Loop Formation in Polypeptide Chains. *J Am Chem Soc*. 2005;127(10):3346-3352. doi:10.1021/ja042798i
273. Baker ML, Ju T, Chiu W. Identification of Secondary Structure Elements in Intermediate-Resolution Density Maps. *Structure*. 2007;15(1):7-19. doi:https://doi.org/10.1016/j.str.2006.11.008
274. Dal Palù A, He J, Pontelli E, Lu Y. Identification of alpha-helices from low resolution protein density maps. *Comput Syst bioinformatics Comput Syst Bioinforma Conf*. Published online 2006:89-98.
275. Arribas-Bosacoma R, Kim S-K, Ferrer-Orta C, Blanco AG, Pereira PJB, Gomis-Rüth FX, Wanner BL, Coll M, Solà M. The X-ray Crystal Structures of Two Constitutively Active Mutants of the *Escherichia coli* PhoB Receiver Domain Give Insights into Activation. *J Mol Biol*. 2007;366(2):626-641. doi:https://doi.org/10.1016/j.jmb.2006.11.038
276. Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, Basutkar P, Tivey ARN, Potter SC, Finn RD, Lopez R. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res*. 2019;47(W1):W636-W641. doi:10.1093/nar/gkz268
277. Thornton J. Structural genomics takes off. *Trends Biochem Sci*. 2001;26(2):88-89. doi:10.1016/s0968-0004(00)01765-5
278. Perrot P. *A to Z of Thermodynamics*. 1st ed. Oxford University Press; 1998.
279. Berezovsky IN, Chen WW, Choi PJ, Shakhnovich EI. Entropic Stabilization of Proteins and Its Proteomic Consequences. *PLOS Comput Biol*. 2005;1(4):e47. https://doi.org/10.1371/journal.pcbi.0010047



280. Lins L, Thomas A, Brasseur R. Analysis of accessible surface of residues in proteins. *Protein Sci.* 2003;12(7):1406-1417. doi:10.1110/ps.0304803
281. Baud F, Karlin S. Measures of residue density in protein structures. *Proc Natl Acad Sci.* 1999;96(22):12494-12499. doi:10.1073/pnas.96.22.12494
282. Dasgupta S, Iyer GH, Bryant SH, Lawrence CE, Bell JA. Extent and nature of contacts between protein molecules in crystal lattices and between subunits of protein oligomers. *Proteins Struct Funct Bioinforma.* 1997;28(4):494-514. doi:https://doi.org/10.1002/(SICI)1097-0134(199708)28:4<494::AID-PROT4>3.0.CO;2-A
283. Iyer GH, Dasgupta S, Bell JA. Ionic strength and intermolecular contacts in protein crystals. *J Cryst Growth.* 2000;217(4):429-440. doi:https://doi.org/10.1016/S0022-0248(00)00503-0
284. Jones S, Thornton JM. Principles of protein-protein interactions. *Proc Natl Acad Sci.* 1996;93(1):13-20. doi:10.1073/pnas.93.1.13
285. Conte L Lo, Chothia C, Janin J. The atomic structure of protein-protein recognition sites. *J Mol Biol.* 1999;285(5):2177-2198. doi:https://doi.org/10.1006/jmbi.1998.2439
286. Doye JPK, Louis AA, Vendruscolo M. Inhibition of protein crystallization by evolutionary negative design. *Phys Biol.* 2004;1(1):P9-P13. doi:10.1088/1478-3967/1/1/p02
287. Slack J. *Genes: A Very Short Introduction.* 1st ed. Oxford University Press; 2014.
288. Sundd P, Gladwin MT, Novelli EM. Pathophysiology of Sick Cell Disease. *Annu Rev Pathol.* 2019;14:263-292. doi:10.1146/annurev-pathmechdis-012418-012838
289. Derewenda ZS. Rational protein crystallization by mutational surface engineering. *Structure.* 2004;12(4):529-535. doi:10.1016/j.str.2004.03.008
290. Murray CW, Callaghan O, Chessari G, Cleasby A, Congreve M, Frederickson M, Hartshorn MJ, McMenemy R, Patel S, Wallis N. Application of Fragment Screening by X-ray Crystallography to β -Secretase. *J Med Chem.* 2007;50(6):1116-1123. doi:10.1021/jm0611962
291. Makabe K, Biancalana M, Yan S, Tereshko V, Gawlak G, Miller-Auer H, Meredith SC, Koide S. High-Resolution Structure of a Self-Assembly-Competent Form of a Hydrophobic Peptide Captured in a Soluble β -Sheet Scaffold. *J Mol Biol.* 2008;378(2):459-467. doi:https://doi.org/10.1016/j.jmb.2008.02.051
292. Munshi S, Hall DL, Kornienko M, Darke PL, Kuo LC. Structure of apo, unactivated insulin-like growth factor-1 receptor kinase at 1.5 Å resolution. *Acta Crystallogr Sect D.* 2003;59(10):1725-1730. doi:10.1107/S0907444903015415



293. Goldschmidt L, Cooper DR, Derewenda ZS, Eisenberg D. Toward rational protein crystallization: A Web server for the design of crystallizable protein variants. *Protein Sci.* 2007;16(8):1569-1576. doi:10.1110/ps.072914007
294. Karginov A, Agaphonov M. A simple enrichment procedure improves detection of membrane proteins by immunoblotting. *Biotechniques.* 2016;61(5):260-261. doi:10.2144/000114474
295. Guinier A, Fournet G. *Small-Angle Scattering of X-Rays.* 1st ed. Wiley; 1955.
296. Putnam CD. Guinier peak analysis for visual and automated inspection of small-angle X-ray scattering data. *J Appl Crystallogr.* 2016;49(Pt 5):1412-1419. doi:10.1107/S1600576716010906
297. Grant TD, Luft JR, Carter LG, Matsui T, Weiss TM, Martel A, Snell EH. The accurate assessment of small-angle X-ray scattering data. *Acta Crystallogr Sect D.* 2015;71(1):45-56. doi:10.1107/S1399004714010876
298. Franke D, Svergun DI. *DAMMIF*, a program for rapid *ab-initio* shape determination in small-angle scattering. *J Appl Crystallogr.* 2009;42(Pt 2):342-346. doi:10.1107/S0021889809000338
299. Volkov V V, Svergun DI. Uniqueness of *ab initio* shape determination in small-angle scattering. *J Appl Crystallogr.* 2003;36(3 Part 1):860-864. doi:10.1107/S0021889803000268
300. Manalastas-Cantos K, Konarev P V, Hajizadeh NR, Kikhney AG, Petoukhov M V, Molodenskiy DS, Panjkovich A, Mertens HDT, Gruzinov A, Borges C, Jeffries CM, Svergun DI, Franke D. *ATSAS 3.0*: expanded functionality and new tools for small-angle scattering data analysis. *J Appl Crystallogr.* 2021;54(1). doi:10.1107/S1600576720013412
301. Svergun DI, Koch MHJ. Small-angle scattering studies of biological macromolecules in solution. *Reports Prog Phys.* 2003;66(10):1735-1782. doi:10.1088/0034-4885/66/10/R05
302. Jensen RO, Winzer K, Clarke SR, Chan WC, Williams P. Differential Recognition of *Staphylococcus aureus* Quorum-Sensing Signals Depends on Both Extracellular Loops 1 and 2 of the Transmembrane Sensor AgrC. *J Mol Biol.* 2008;381(2):300-309. doi:https://doi.org/10.1016/j.jmb.2008.06.018
303. Haack A, Aragão GG, Novaes MRCG. Pathophysiology of cystic fibrosis and drugs used in associated digestive tract diseases. *World J Gastroenterol.* 2013;19(46):8552-8561. doi:10.3748/wjg.v19.i46.8552
304. Machan ZA, Taylor GW, Pitt TL, Cole PJ, Wilson R. 2-Heptyl-4-hydroxyquinoline N-oxide, an antistaphylococcal agent produced by *Pseudomonas aeruginosa*. *J Antimicrob Chemother.* 1992;30(5):615-623. doi:10.1093/jac/30.5.615



305. Melter O, Radojevič B. Small colony variants of *Staphylococcus aureus*--review. *Folia Microbiol (Praha)*. 2010;55(6):548-558. doi:10.1007/s12223-010-0089-3
306. Qazi S, Middleton B, Muharram SH, Cockayne A, Hill P, O'Shea P, Chhabra SR, Cámara M, Williams P. N-acylhomoserine lactones antagonize virulence gene expression and quorum sensing in *Staphylococcus aureus*. *Infect Immun*. 2006;74(2):910-919. doi:10.1128/IAI.74.2.910-919.2006
307. Cornelis P, Dingemans J. *Pseudomonas aeruginosa* adapts its iron uptake strategies in function of the type of infections. *Front Cell Infect Microbiol*. 2013;3:75. doi:10.3389/fcimb.2013.00075
308. Carmi R, Carmeli S, Levy E, Gough FJ. (+)-(S)-dihydroaeruginolic acid, an inhibitor of *Septoria tritici* and other phytopathogenic fungi and bacteria, produced by *Pseudomonas fluorescens*. *J Nat Prod*. 1994;57(9):1200-1205. doi:10.1021/np50111a002
309. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem*. 2010;31(2):455-461. doi:10.1002/jcc.21334
310. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J Comput Chem*. 2009;30(16):2785-2791. doi:10.1002/jcc.21256
310. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J Comput Chem*. 2009;30(16):2785-2791. doi:10.1002/jcc.21256
311. Jensen HM, Eng T, Chubukov V, Herbert RA, Mukhopadhyay A. Improving membrane protein expression and function using genomic edits. *Sci Rep*. 2017;7(1):13030. doi:10.1038/s41598-017-12901-7
312. Miroux B, Walker JE. Over-production of Proteins in *Escherichia coli*: Mutant Hosts that Allow Synthesis of some Membrane Proteins and Globular Proteins at High Levels. *J Mol Biol*. 1996;260(3):289-298. doi:https://doi.org/10.1006/jmbi.1996.0399
313. Baumgarten T, Schlegel S, Wagner S, Löw M, Eriksson J, Bonde I, Herrgård MJ, Heipieper HJ, Nørholm MHH, Slotboom DJ, de Gier J-W. Isolation and characterization of the *E. coli* membrane protein production strain Mutant56(DE3). *Sci Rep*. 2017;7(1):45089. doi:10.1038/srep45089
314. Kaur J, Bachhawat AK. A modified Western blot protocol for enhanced sensitivity in the detection of a membrane protein. *Anal Biochem*. 2009;384(2):348-349. doi:https://doi.org/10.1016/j.ab.2008.10.005
315. Bai N, Roder H, Dickson A, Karanicolas J. Isothermal Analysis of ThermoFluor Data can readily provide Quantitative Binding Affinities. *Sci Rep*. 2019;9(1):2650. doi:10.1038/s41598-018-37072-x



316. Dart ML, Machleidt T, Jost E, Schwinn MK, Robers MB, Shi C, Kirkland TA, Killoran MP, Wilkinson JM, Hartnett JR, Zimmerman K, Wood K V. Homogeneous Assay for Target Engagement Utilizing Bioluminescent Thermal Shift. *ACS Med Chem Lett.* 2018;9(6):546-551. doi:10.1021/acsmchemlett.8b00081
317. Kelly SM, Jess TJ, Price NC. How to study proteins by circular dichroism. *Biochim Biophys Acta - Proteins Proteomics.* 2005;1751(2):119-139. doi:https://doi.org/10.1016/j.bbapap.2005.06.005
318. Robertson JL. The lipid bilayer membrane and its protein constituents. *J Gen Physiol.* 2018;150(11):1472-1483. doi:10.1085/jgp.201812153
319. Ollila S, Hyvönen MT, Vattulainen I. Polyunsaturation in lipid membranes: dynamic properties and lateral pressure profiles. *J Phys Chem B.* 2007;111(12):3139-3150. doi:10.1021/jp065424f
320. Marsh D. Lateral pressure profile, spontaneous curvature frustration, and the incorporation and conformation of proteins in membranes. *Biophys J.* 2007;93(11):3884-3899. doi:10.1529/biophysj.107.107938
321. Marsh D. Lateral pressure in membranes. *Biochim Biophys Acta.* 1996;1286(3):183-223. doi:10.1016/s0304-4157(96)00009-3
322. Jamshad M, Lin Y-P, Knowles TJ, Parslow RA, Harris C, Wheatley M, Poyner DR, Bill RM, Thomas ORT, Overduin M, Dafforn TR. Surfactant-free purification of membrane proteins with intact native membrane environment. *Biochem Soc Trans.* 2011;39(3):813-818. doi:10.1042/BST0390813
323. Wei Y, Thyparambil AA, Latour RA. Protein helical structure determination using CD spectroscopy for solutions with strong background absorbance from 190 to 230nm. *Biochim Biophys Acta.* 2014;1844(12):2331-2337. doi:10.1016/j.bbapap.2014.10.001
324. Li H, Hu D, Liang F, Huang X, Zhu Q. Influence factors on the critical micelle concentration determination using pyrene as a probe and a simple method of preparing samples. *R Soc Open Sci.* 2020;7(3):192092. doi:10.1098/rsos.192092
325. Allen TM, Romans AY, Kercret H, Segrest JP. Detergent removal during membrane reconstitution. *Biochim Biophys Acta - Biomembr.* 1980;601:328-342. doi:https://doi.org/10.1016/0005-2736(80)90537-4
326. Sivars U, Tjerneld F. Mechanisms of phase behaviour and protein partitioning in detergent/polymer aqueous two-phase systems for purification of integral membrane proteins. *Biochim Biophys Acta - Gen Subj.* 2000;1474(2):133-146. doi:https://doi.org/10.1016/S0304-4165(99)00208-1
327. Eriks LR, Mayor JA, Kaplan RS. A strategy for identification and quantification of detergents frequently used in the purification of membrane proteins. *Anal Biochem.* 2003;323(2):234-241. doi:https://doi.org/10.1016/j.ab.2003.09.002



328. Strop P, Brunger AT. Refractive index-based determination of detergent concentration and its application to the study of membrane proteins. *Protein Sci.* 2005;14(8):2207-2211. doi:10.1110/ps.051543805
329. Ben-Naim A, Stillinger FH. Critical micelle concentration and the size distribution of surfactant aggregates. *J Phys Chem.* 1980;84(22):2872-2876. doi:10.1021/j100459a008
330. Herrmann KW. Non-ionic-cationic micellar properties of dimethyldodecylamine oxide. *J Phys Chem.* 1962;66(2):295-300. doi:10.1021/j100808a025
331. Chung KY, Kim TH, Manglik A, Alvares R, Kobilka BK, Prosser RS. Role of detergents in conformational exchange of a G protein-coupled receptor. *J Biol Chem.* 2012;287(43):36305-36311. doi:10.1074/jbc.M112.406371
332. Chaptal V, Delolme F, Kilburg A, Magnard S, Montigny S, Picard M, Prier C, Monticelli L, Bornert O, Agez M, Ravaid S, Orelle C, Wagner R, Jawhari A, Broutin I, Pebau-Peyroula E, Jault J-M, Kaback HR, le Maire M, Falson P. Quantification of Detergents Complexed with Membrane Proteins. *Sci Rep.* 2017;7(1):41751. doi:10.1038/srep41751
333. Fan X, Wang J, Zhang X, Yang Z, Zhang J-C, Zhao L, Peng H-L, Lei J, Wang H-W. Single particle cryo-EM reconstruction of 52 kDa streptavidin at 3.2 Angstrom resolution. *Nat Commun.* 2019;10(1):2386. doi:10.1038/s41467-019-10368-w
334. Oliver RC, Lipfert J, Fox DA, Lo RH, Doniach S, Columbus L. Dependence of micelle size and shape on detergent alkyl chain length and head group. *PLoS One.* 2013;8(5):e62488. doi:10.1371/journal.pone.0062488
335. Timmins PA, Leonhard M, Weltzien HU, Wacker T, Welte W. A physical characterization of some detergents of potential use for membrane protein crystallization. *FEBS Lett.* 1988;238(2):361-368. doi:10.1016/0014-5793(88)80513-1
336. Birch J, Axford D, Foadi J, Meyer A, Eckhardt A, Thielmann Y, Moraes I. The fine art of integral membrane protein crystallisation. *Methods.* 2018;147:150-162. doi:https://doi.org/10.1016/j.ymeth.2018.05.014
337. Landau EM, Rosenbusch JP. Lipidic cubic phases: a novel concept for the crystallization of membrane proteins. *Proc Natl Acad Sci.* 1996;93(25):14532-14535. doi:10.1073/pnas.93.25.14532
338. Caffrey M. A comprehensive review of the lipid cubic phase or in meso method for crystallizing membrane and soluble proteins and complexes. *Acta Crystallogr Sect F.* 2015;71(Pt 1):3-18. doi:10.1107/S2053230X14026843
339. Cherezov V, Peddi A, Muthusubramaniam L, Zheng YF, Caffrey M. A robotic system for crystallizing membrane and soluble proteins in lipidic mesophases. *Acta Crystallogr Sect D.* 2004;60(Pt 10):1795-1807. doi:10.1107/S0907444904019109



340. Qiu H, Caffrey M. The phase diAgrAm of the monoolein/water system: metastability and equilibrium aspects. *Biomaterials*. 2000;21(3):223-234. doi:10.1016/s0142-9612(99)00126-x
341. Caffrey M. On the Mechanism of Membrane Protein Crystallization in Lipidic Mesophases. *Cryst Growth Des*. 2008;8(12):4244-4254. doi:10.1021/cg800693r
342. Caffrey M. Kinetics and mechanism of transitions involving the lamellar, cubic, inverted hexagonal, and fluid isotropic phases of hydrated monoacylglycerides monitored by time-resolved X-ray diffraction. *Biochemistry*. 1987;26(20):6349-6363. doi:10.1021/bi00394a008
343. Caffrey M, Cherezov V. Crystallizing membrane proteins using lipidic mesophases. *Nat Protoc*. 2009;4(5):706-731. doi:10.1038/nprot.2009.31
344. Wang Y, Huang Y, Wang J, Cheng C, Huang W, Lu P, Xu Y-N, Wang P, Yan N, Shi Y. Structure of the formate transporter FocA reveals a pentameric aquaporin-like channel. *Nature*. 2009;462(7272):467-472. doi:10.1038/nature08610
345. Ritter B, Denisov AY, Philie J, Deprez C, Tung EC, Gehring K, McPherson PS. Two WXXF-based motifs in NECAPs define the specificity of accessory protein binding to AP-1 and AP-2. *EMBO J*. 2004;23(19):3701-3710. doi:10.1038/sj.emboj.7600378
346. Holst M, Saied F. Multigrid solution of the Poisson—Boltzmann equation. *J Comput Chem*. 1993;14(1):105-113. doi:https://doi.org/10.1002/jcc.540140114
347. Jurrus E, Engel D, Star K, Monson M, Brandi J, Felberg LE, Brookes DH, Wilson L, Chen J, Liles K, Chun M, Li P, Gohara DW, Dolinsky T, Konecny R, Koes DR, Nielsen JE, Head-Gordon T, Geng W, Krasny R, Wei G-W, Holst MJ, McCammon JA, Baker NA. Improvements to the APBS biomolecular solvation software suite. *Protein Sci*. 2018;27(1):112-128. doi:https://doi.org/10.1002/pro.3280
348. Thoendel M, Horswill AR. Random mutagenesis and topology analysis of the autoinducing peptide biosynthesis proteins in *Staphylococcus aureus*. *Mol Microbiol*. 2013;87(2):318-337. doi:10.1111/mmi.12100
349. Forrest LR. Structural Symmetry in Membrane Proteins. *Annu Rev Biophys*. 2015;44:311-337. doi:10.1146/annurev-biophys-051013-023008
350. Levy ED, Pereira-Leal JB, Chothia C, Teichmann SA. 3D complex: a structural classification of protein complexes. *PLoS Comput Biol*. 2006;2(11):e155-e155. doi:10.1371/journal.pcbi.0020155
351. Zhang L, Ji G. Identification of a staphylococcal AgrB segment(s) responsible for group-specific processing of AgrD by gene swapping. *J Bacteriol*. 2004;186(20):6706-6713. doi:10.1128/JB.186.20.6706-6713.2004
352. Weiss GA, Watanabe CK, Zhong A, Goddard A, Sidhu SS. Rapid mapping of protein functional epitopes by combinatorial alanine scanning. *Proc Natl Acad Sci*. 2000;97(16):8950 LP - 8954. doi:10.1073/pnas.160252097



353. Alexandrov AI, Mileni M, Chien EYT, Hanson MA, Stevens RC. Microscale fluorescent thermal stability assay for membrane proteins. *Structure*. 2008;16(3):351-359. doi:10.1016/j.str.2008.02.004
354. Gless BH, Bojer MS, Peng P, Baldry M, Ingmer H, Olsen CA. Identification of autoinducing thiopeptides from staphylococci enabled by native chemical ligation. *Nat Chem*. 2019;11(5):463-469. doi:10.1038/s41557-019-0256-3
355. Moon AF, Mueller GA, Zhong X, Pedersen LC. A synergistic approach to protein crystallization: Combination of a fixed-arm carrier with surface entropy reduction. 2010;19:901-913. doi:10.1002/pro.368
356. Ribeiro JFR, Cianni L, Li C, Warwick TG, de Vita D, Rosini F, dos Reis Rocho F, Martins FCP, Kenny PW, Lameira J, Leitão A, Emsley J, Montanari CA. Crystal structure of *Leishmania mexicana* cysteine protease B in complex with a high-affinity azadipeptide nitrile inhibitor. *Bioorg Med Chem*. 2020;28(22):115743. doi:<https://doi.org/10.1016/j.bmc.2020.115743>
357. Schägger H. Tricine–SDS-PAGE. *Nat Protoc*. 2006;1(1):16-22. doi:10.1038/nprot.2006.4
358. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, Tinevez J-Y, White DJ, Hartenstein V, Eliceiri K, Tomancak P, Cardona A. Fiji: an open-source platform for biological-image analysis. *Nat Methods*. 2012;9(7):676-682. doi:10.1038/nmeth.2019