

Curtin Medical School

**Investigating the structure and binding mechanism of QseM,
a novel dual-target protein-inhibitor**

**Drew Anthony Hall
0000-0001-6562-2335**

**This thesis is presented for the Degree of
Doctor of Philosophy
of
Curtin University**

August 2021

Declaration

To the best of my knowledge and belief this thesis contains no material previously published by any other person except where due acknowledgment has been made.

This thesis contains no material which has been accepted for the award of any other degree or diploma in any university.

Signature:

Date: 20-08-2021
.....

Acknowledgments

I acknowledge Curtin University and the Australian Government which supported this research through an APA-CRS scholarship.

The work in this thesis contains significant contributions from members of the Ronson lab at the University of Otago, Mackay and Matthews lab at the University of Sydney, members of the Bond lab and the Centre for Microscopy, Characterisation and Analysis (CMCA) at the University of Western Australia and the Australian Synchrotron staff and facilities. Specific credits to individuals are included throughout this thesis where appropriate.

I would like to acknowledge my two supervisors, Dr Josh Ramsay and Prof. Charles Bond, who have dedicated a huge amount of work and time in teaching and supporting me throughout my PhD. I need to thank the whole Ramsay lab for their friendship and help, particularly Callum Verdonk and Tahlia Blastholm. I also wish to thank the Bond lab, with a special thanks to Dr Karina Yui Eto, Dr Jason Schimdburger and Dr Gavin Knott.

I also want to thank Prof. Clive Ronson, Dr. John Sullivan, Will Jowsey and Calum Morris for their contributions towards our shared research goals. I also need to acknowledge Prof. Mark Howard, Dr. Gareth Nealon, Prof. Joel Mackay and Paul Solomon for their teaching, assistance and work done towards our NMR experiments.

This thesis was made possible by continued support and talent from all the people mentioned above. On a more personal note, over the course of this PhD my life has been filled with family and friends that, in one way or another, contributed to my growth, both as a student and person. I want to thank my parents, Jeff and Sue, who have given me an amazing life filled with opportunities and unwavering support. To Percy, thank you for being the grandfather you were. Carmen, we don't speak too much, but I know that you are there if I need you.

Elena, you have been one of the best things that has happened to me. In just a few short years we have formed a loving home with our perfect little girl. You have been there to help me get through some hard times and made my life so

amazing. Finally, dear Elizaveta, one day I hope you find your own path, driven by curiosity and commitment to achieve great things.

“Far better it is to dare mighty things, to win glorious triumphs, even though checkered by failure, than to take rank with those poor spirits that neither enjoy much nor suffer much, because they live in the grey twilight that knows neither victory nor defeat” - Theodore Roosevelt

ABSTRACT

Transcriptional activator proteins bind DNA and promote gene transcription. These proteins often contain a helix-turn-helix (HTH) motif to enable sequence-specific DNA binding within the DNA major groove. Bacterial communication systems called quorum sensing (QS) have evolved to regulate specific gene expression based on population density. The most well characterised QS system is the LuxR/I family found in Gram-negative bacterial species. LuxR proteins of these QS systems contain a HTH motif that can only bind and activate transcription in the presence of specific membrane-diffusible signaling molecules.

Mezorhizobium loti R7A is Gram-negative bacterium that forms a symbiosis with plant roots and converts atmospheric nitrogen into ammonium. Genes required for this nitrogen-fixing symbiosis are located on a 502 kilobase (kb) mobile integrative and conjugative element (ICE) called ICE M/Sym^{R7A} . Chromosomal excision and conjugative transfer of ICE M/Sym^{R7A} is regulated by transcriptional response protein (TraR) and transcriptional inducer protein (TraI1) which comprise a LuxR/I QS system. TraR activity is tightly regulated by a quorum sensing master protein (QseM), which inhibits the function of TraR and is thus termed an antiactivator. Uniquely, QseM also inhibits the function of a second transcriptional activator FseA, which shares no primary sequence similarity with TraR. QseM also exhibits no sequence similarity to previously characterised antiactivator proteins.

QseM and FseA both contain a domain of unknown function (DUF) 2285. DUF2285-domain proteins are present across 330 proteobacterial species, however, at the time work in this thesis was completed, no structure had been solved for any DUF2285 protein. FseA is a 266 amino-acid protein that encodes a DUF2285 domain within its C-terminus, while the DUF2285 domain spans the length of the entire 84 amino-acid QseM protein. The aim of this thesis was to determine the molecular structure of QseM to provide insight into the DUF2285 domain and help expose how QseM achieves antiactivation of both TraR and FseA. QseM was heterologously expressed, purified and used in X-ray crystallography experiments. We collected 44 datasets for QseM

crystals and achieved X-ray diffraction data with $< 3 \text{ \AA}$ resolution. 2 complete datasets were collected for native crystals and 42 were collected using seleniomethionine-incorporated QseM crystals to determine phase using single/multi-wavelength anomalous dispersion (SAD/MAD) methods. Despite obtaining datasets required for phase determination by both SAD and MAD, phase was not successfully resolved and a crystallographic solution for QseM was not solved.

Nuclear magnetic resonance (NMR) was used as an alternative approach to obtain three-dimensional structure information for QseM. Unlabeled and $^{13}\text{C}/^{15}\text{N}$ isotope labeled QseM protein was used to collect NMR data in various one, two or three dimensional (1D, 2D, 3D) experiments. Backbone assignment was completed to 88% and full atom assignment was completed to 77% of all proton, carbon and nitrogen atoms in the protein. The majority of unassigned atoms resided in the extreme C and N terminal sections of the protein. Other unassignable atoms included those located on the far ends of long sidechains. Inability to complete atom assignment came from poor resonance signal for those atoms. This reduced signal likely came from high mobility or conformation positions of those atoms relative to the majority of the protein backbone.

Structural predictions based on distance restraints showed QseM formed three alpha helices (H1, H2, H3) that together formed a stable globular protein. Comparisons with known protein structures revealed the fold between H2 and H3 resembled that of the HTH motif, but with a distinct extended loop between H2 and H3. Structural predictions based on amino-acid coevolution and homology models based on a recently solved RovC (distantly related DUF2285 domain protein) crystal structure were consistent with the determined NMR structure and small angle X-ray scattering data. Unfortunately, attempts to use QseM NMR ensemble or *ab initio* predicted structure for molecular replacement to solve the crystal structure of QseM was unsuccessful.

Mutational analyses of QseM and FseA carried out by colleagues (Calum Morris, Will Jowsey, Ronson Lab, University of Otago) in parallel with this work,

revealed that mutations in the HTH region of QseM had little effect on its ability to bind or inhibit FseA or TraR. In contrast, mutations in H1 and the C-terminus of H3, which together form a surface on the opposite side of the QseM structure, revealed residues in this region critical for binding both TraR and FseA. It is likely that the C-terminal region on FseA (homologous to QseM) interacts with the N-terminus of another FseA chain to create homodimers. The homologous region on FseA likely forms extensive interactions with N-terminus of FseA. Mutations in the HTH region of FseA reduced or abolished its ability to bind DNA. We hypothesize that dimeric DUF2285 proteins such as FseA are a divergent family of DNA-binding HTH proteins but monomeric DUF2285-domain proteins such as QseM, have lost the ability to bind DNA and instead antagonize the activity of transcriptional activators through heterodimer interactions requiring the H1/H3 surface.

Future experiments would be focused on successfully solving the crystal structure of QseM followed by co-crystallisation of QseM with TraR or FseA. A co-crystallised structure would explicitly show the binding mechanism between QseM and TraR or FseA.

Contents:

List of Figures:.....	xi
List of Tables:.....	xiv
Abbreviations:.....	xv
Chapter 1:.....	17
Literature Review.....	17
1.1: Transcriptional regulators	18
1.2: Bacterial Quorum Sensing.....	19
1.3: N-acyl-homoserine lactone (HSL) synthesis.....	21
1.4: QS pathogens.....	24
1.5: Plant associated QS	25
1.6: LuxR-Family transcriptional regulators	28
1.7: QS inhibition	31
1.8: Antiactivation of LuxR-family proteins.....	32
1.9: Horizontal transfer of genes for nitrogen-fixing symbiosis in <i>Mesorhizobium loti</i> R7A.....	38
1.10: Regulation of symbiosis island transfer by quorum sensing	39
1.12: Aims of this thesis.....	45
Chapter 2:.....	46
Materials and Methods	46
2.1 Materials:	47
2.2 Chemicals:.....	48
2.3 Strains and vectors:	49
2.4 Buffers and solutions	52
2.5 Methods:.....	55
2.5.1 Plasmid construction	55
2.5.2 Expression of QseM and FseA.....	55

2.5.3 Cell lysis	56
2.5.4 Purification of QseM	56
2.5.5 Purification of FseA	57
2.5.6 Protein concentration.....	58
2.5.7 Analysis of protein sample.....	58
2.6 Crystallisation	59
2.6.1 Crystallisation screening.....	59
2.6.2 Observation of crystallisation.....	60
2.6.3 Creating a crystal micro-seeding stock.....	60
2.6.4 Treating crystals with cryoprotectant solution	60
2.7 Crystallographic data collection and processing	61
2.7.1 X-ray diffraction data collection	61
2.7.2 Anomalous phasing	62
2.7.3 Molecular replacement phasing.....	62
2.8 NMR data collection, atom assignment and structure calculation:.....	63
2.8.1 Protein sample preparation and handling for NMR.....	63
2.8.2 Initial one-dimensional spectra analysis	64
2.8.3 Single and double labelled QseM NMR spectra collection	64
2.8.4 Atom assignment.....	66
2.8.5 Dihedral angle calculations.....	67
2.8.6 Structure calculations	67
2.9 Additional protein characterisation methods.	67
2.9.1 Small angle X-ray scattering (SAXS).....	67
2.9.2 Analytical size exclusion chromatography	68
2.9.3 Structural alignment of QseM with other proteins.....	68
2.9.4 <i>de novo</i> protein structure prediction	69
Chapter 3:.....	70

Recombinant protein expression and purification	70
3.1 Introduction:	71
3.2 Results and Discussion:	73
3.2.1 Purification of QseM and FseA.....	73
3.2.2 Purification of untagged QseM	78
3.2.3 Purification of 6H-MBP-TEV-FseA	80
3.2.4 Confirmation of protein post-purification via mass spectroscopy..	82
3.3 Conclusions	88
Chapter 4:.....	89
Structural analysis via X-ray crystallography	89
4.1 Introduction:.....	90
4.2 Results and Discussion:	94
4.2.1 Observation of crystallisation.....	94
4.2.2 Optimisation of 6H-QseM crystallisation.....	98
4.2.3 Crystallisation of Se-6H-QseM	101
4.2.4 Cryoprotection and cryocooling crystals.....	101
4.2.5 Crystallisation of FseA.....	103
4.2.6 Diffraction data collection and analysis.....	104
4.2.7 Space Group Assignment.....	112
4.2.8 Attempts to determine phase information	115
4.2.9 Attempts to solve structure using MR	116
4.3 Conclusion:.....	117
Chapter 5:.....	119
QseM Structure Determination by NMR	119
5.1 Introduction:.....	120
5.2 Results and Discussion:	125
5.2.1 Testing QseM suitability for NMR	125

5.2.2 Atom assignment.....	129
5.2.3 NMR structure of 6H-Qsem	139
5.2.4 Preliminary structure validation report	144
5.2.5 Structures of 6H-QseM.....	145
5.2.6 Refinement of NMR structure for molecular replacement.....	147
5.2.7 Supporting structural information.....	148
5.3 Conclusions	153
Chapter 6:.....	156
Structure-function analysis of QseM and DUF2285 proteins.....	156
6.1 Introduction:.....	157
6.2 Results and Discussion:	159
6.2.1 Alignment of QseM structure to known proteins	159
6.2.2 Structure determination of FseA.....	166
6.2.3 Structure comparison between QseM and FseA	166
6.2.4 Residue conservation of QseM and FseA	170
6.2.5 Mutagenesis analysis of QseM and FseA	174
6.3 Conclusion:.....	186
Chapter 7:.....	188
Final discussion and conclusions	188
Chapter 8:.....	196
References	196

List of Figures:

Figure 1.1: Lambda repressor bound to DNA target.	19
Figure 1.2: The <i>Vibrio fischeri</i> QS system.....	20
Figure 1.3: N-acyl homoserine lactone structure and target.	22
Figure 1.4: of QS regulation in <i>R. leguminosarum</i>	28
Figure 1.5: TraR of <i>A. tumefaciens</i> binding DNA.	30
Figure 1.6: LasR ligand binding pocket.....	31
Figure 1.7: TraM-TraR heterotetramer.....	34
Figure 1.8: TraR-TraM binding region.....	35
Figure 1.9: QsIA homodimer binding ligand-binding domain (LBD) of LasR.	36
Figure 1.10: QsIA-QsIA-LasR heterotrimer hydrophobic surface interface. .	37
Figure 1.11: ICE excision and transfer.....	39
Figure 1.12: QS activation in <i>M. loti</i> R7A.	41
Figure 1.13: QS inactivation in <i>M. loti</i> R7A by the antiactivator QseM.	42
Figure 1.14: Circular dichroism profiles of various secondary structures.	43
Figure 1.15: Circular dichroism of 6H-QseM.....	44
Figure 2.1: Vector map of pQe80L.....	50
Figure 2.2: Vector map of pET11a.....	50
Figure 3.1: Purification of 6H-QseM.....	75
Figure 3.2: Purification of Se-6H-QseM.	76
Figure 3.3: Purification of isotope-labelled 6H-QseM.....	77
Figure 3.4: Purification of 6H-TEV-QseM.....	79
Figure 3.5: Purification of 6H-MBP-TEV-FseA.....	81
Figure 4.1: Attempts to crystallise 6H-QseM in PBS buffer.....	95
Figure 4.2: Crystal screening of 6H-QseM.....	97
Figure 4.3: pH and PEG effects on crystallisation of 6H-QseM in D11 condition.	99
Figure 4.4: growth over time of 6H-QseM crystals.....	99
Figure 4.5: 6H-QseM crystallisation optimisation.....	101
Figure 4.6: Cryocooled 6H-QseM crystal mounted at AS MX1 beamline...	103
Figure 4.7: 6H-QseM diffraction comparison.....	105
Figure 4.8: Diffraction data quality and intensity	108
Figure 4.9: Ctruncate L-Test twinning analysis.	109

Figure 4.10: Intensity vs index for systematic absence determination.	114
Figure 4.11: Rotational relationship between the unit cell dimensions of 6H-QseM diffraction data index results.....	115
Figure 5.1: 1D ¹ H NMR analysis of 6H-QseM.....	126
Figure 5.2: 1D ¹ H spectra of 6H-QseM before and after freeze-drying.	127
Figure 5.3: ¹⁵ N-HSQC of 6H-QseM over time.	128
Figure 5.4: ¹⁵ N-HSQC of precipitated 6H-QseM.	129
Figure 5.5: Backbone assignment method.....	131
Figure 5.6: The typical C _α and C _β chemical shifts of each amino acid.....	132
Figure 5.7: First full assignment of 6H-QseM linked to the ¹⁵ N-HSQC.....	133
Figure 5.8: Initial backbone assignment of 6H-QseM.	134
Figure 5.9: ¹⁵ N-HSQC of 6H-QseM.....	135
Figure 5.10: HCCH-TOCSY assignment for sidechain carbons and protons.	136
Figure 5.11: ¹³ C-HSQC of 6H-QseM.....	137
Figure 5.12: Backbone and sidechain assignment.....	138
Figure 5.13: CYANA structure calculation.....	141
Figure 5.14: Final 6H-QseM NMR ensemble.	142
Figure 5.15: NMR structure of 6H-QseM with CYANA-derived distance restraints.	143
Figure 5.16: preliminary structure quality of 6H-QseM.....	144
Figure 5.17: Random coil index plot (RCI) indicating ordered region of 6H-QseM.	145
Figure 5.18: 6H-QseM structures generated.....	146
Figure 5.19: Ensemble of the top 10 structures generated from the <i>ROSETTA</i> after refinement.....	148
Figure 5.20: SAXS analysis of 6H-QseM.	151
Figure 5.21: <i>CRYSQL</i> comparison between SAXS scattering and NMR ensemble of 6H-QseM.	152
Figure 5.22: DAMMIN surface model with NMR ensemble.....	152
Figure 6.1: Helix-turn-helix motif of QseM.....	158
Figure 6.2: <i>DALI</i> matches structure aligned to 6H-QseM.....	162
Figure 6.3: RMSD analysis from <i>DALI</i> matches.....	163
Figure 6.4: Residue conservation in structure alignment.	164

Figure 6.5: Sequence alignment of converged structure region.....	164
Figure 6.6: 6H-QseM-NMR structure compared to DUF2285 domain of FseA.	167
Figure 6.7: Q-dependent antitermination protein binding DNA.	168
Figure 6.8: Electrostatic surface of QseM and DUF2285 domain of FseA.	169
Figure 6.9: Residue conservation of QseM homologues.	172
Figure 6.10: Residue conservation across the entire DUF2285 protein family.	173
Figure 6.11: Preliminary sequence alignment of QseM homologues from various species.	174
Figure 6.12: Summary of QseM residue substitution in function analysis. .	176
Figure 6.13: Mapped alanine substitution results on 6H-QseM (antiactivation against FseA).....	178
Figure 6.14: Mapping alanine substitution results on QseM structure (antiactivation against TraR).	179
Figure 6.15: Overview of binding site and HTH motif of 6H-QseM.....	180
Figure 6.16: Conserved residues of 6H-QseM likely associated with antiactivation (low mobility).	182
Figure 6.17: Conserved residues of 6H-QseM likely associated with antiactivation (high mobility).....	182
Figure 6.18: Surface-exposed conserved residues of 6H-QseM likely required for antiactivation, Trp82, Met38 and Met83.....	183
Figure 6.19: Residues Arg39 and Asp42 in the 6H-QseM NMR ensemble.	184
Figure 6.20: Residues of 6H-QseM likely required for antiactivation, Gly59, Asp61 and Pro62.	185
Figure 7.1: Overall observations of newly determined 6H-QseM structure.	193
Figure 7.2: Proposed antiactivation method of QseM.	195

List of Tables:

Table 1.1: List of known bacteria phenotypes regulated by QS.	23
Table 2.1: Overview of materials used.	47
Table 2.2: Overview of chemicals used.	48
Table 2.3: Strain details.	49
Table 2.4: Vector details.	49
Table 2.5: Primer details.	49
Table 2.6: Medium composition.	52
Table 2.7: Antibiotic selection for strains and vectors.	53
Table 2.8: Nickel-purification buffer composition.....	53
Table 2.9: SEC purification/crystallisation buffer.....	53
Table 2.10: Minimal salt buffer.	53
Table 2.11: Protein extinction coefficients.....	54
Table 2.12: Buffers for protein gel electrophoresis.....	54
Table 2.13: Protein loading dye (4 x).....	54
Table 2.14: SDS-PAGE Coomassie blue stain and de-stain.....	54
Table 2.15: List of NMR experiments run for structure determination (required isotope-labelled protein).	64
Table 2.16: Experimental collection details.	65
Table 3.1: Mascot search parameters.....	83
Table 3.2: Mascot search results for mass spectrometry analysis.	83
Table 3.3: 6H-QseM mass spectrometry analysis data.....	84
Table 3.4: 6H-MBP-TEV-FseA mass spectrometry analysis data.....	85
Table 4.1: 6H-QseM crystal producing conditions.....	96
Table 4.2: Concentrations and effects of cryoprotectants used.	102
Table 4.3: Determined MAD wavelengths for Se-6H-QseM.....	106
Table 4.4: Statistics for best diffraction datasets of crystallised 6H-QseM.	111
Table 4.5: Laue group and space group determination for top datasets (from <i>POINTLESS</i>).....	113
Table 5.1: NMR experiments and usage.....	122
Table 5.2: Chemical shift assignment completeness of 6H-QseM.	138
Table 5.3: SAXS data collection and analysis for 6H-QseM.	149
Table 6.1: list of primary <i>DALI</i> matches to 6H-QseM NMR input.	160

Abbreviations:

13C	Carbon (isotope)
15N	Nitrogen (isotope)
6H	Hexa-histidine
Ala	Alanine
Arg	Arginine
Asn	Asparagine
Asp	Aspartic acid
Cys	Cystine
ddH2O	Double distilled H2O
DNA	Deoxyribonucleic acid
DUF2285	Domain of unknown function 2285
EDTA	Ethylenediaminetetraacetic acid disodium
EG	Ethylene glycol
EPS	Extracellular polymeric substance
FPLC	Fast protein liquid chromatography
Glu	Glutamic acid
Glu	Glutamine
Gly	Glycine
His	Histidine
HSQC	Heteronuclear single quantum coherence
IDT	Integrated DNA Technologies
Ile	Isoleucine
IPTG	Isopropyl β -D-1-thiogalactopyranoside
ITC	Isothermal titration calorimetry
LB	Lysogeny broth
Leu	Leucine
Lys	Lysine
M9	Minimal medium
MAD	Multiple anomalous dispersion
MBP	Maltose binding protein
Met	Methionine
MST	Micro scale thermophoresis
MW	Molecular weight
Ni-NTA	Nickel-nitrilotriacetic acid
NMR	Nuclear magnetic resonance
PBS	Phosphate buffer saline
PCR	Polymerase chain reaction
PDB	Protein data bank
PEG	Polyethylene glycol
Phe	Phenylalanine
Pro	Proline
QS	Quorum sensing
RMSD	Root mean squared deviation
RNA	Ribonucleic acid

SAD	Single anomalous dispersion
SDS-PAGE	Sodium dodecyl sulphate - polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
SeMet	L-selenomethionine
Ser	Serine
TEV	Tobacco etch virus
Thr	Threonine
Tris	Trisaminomethane
Trp	Tryptophan
Tyr	Tyrosine
UV	Ultraviolet
Val	Valine

Chapter 1:

Literature Review

1.1: Transcriptional regulators

Genes encoded by DNA are transcribed into RNA, often to be further translated into proteins (1). Gene transcription is carefully regulated as organisms need to tune the expression of specific genes and proteins in order to respond to changes in their environment (2). In bacteria, transcription is carried out by multi-subunit RNA polymerase that is recruited to the gene to be transcribed via a secondary protein called the sigma factor (3). Sigma factors are DNA-binding proteins that recognise conserved DNA sequences upstream of genes and together with RNA polymerase (RNAP), initiate transcription (4). Additional DNA-binding proteins are often required for the formation of an active RNAP complex. These transcription factors (TF) can bind specific DNA sequences often positioned adjacent to sigma-factor binding sites. TFs often require additional ligands or phosphorylation for activation and consequently TFs can enable transcriptional regulation of target gene in response to some stimulating molecule or event (5).

Protein structure determination using X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy has revealed that most prokaryotic TFs contain either a helix-turn-helix (HTH) (Figure 1.1) or a winged helix (wHTH) motif (6). These motifs are necessary for DNA binding and function by forming a series of bonds to exposed nucleotide bases in the major groove of DNA (7). Both HTH and wHTH motifs require at least two alpha helices. One helix recognises specific DNA sequences and is termed the recognition helix (8), while a second helix helps to stabilise the interaction but does not contribute significantly to DNA recognition. HTHs can be further classified by total number of helices present (9, 10). Tri-helical are the simplest, consisting of three alpha helices with the third being the recognition helix. Tri-helical HTH are found across the domains of Archaea, Bacteria and Eukarya. Tetra-helical HTH motif proteins have an additional C-terminal alpha helix. Similar to tri-helical HTH, the third helix recognises and binds the DNA target sequence. There are several families of TFs classified as tetra-helical HTH proteins, however, some of these are only found in Prokaryotes, including the LuxR-family present in N-acyl-homoserine lactone quorum-sensing systems (11). Winged HTH TFs contain a β -strand hairpin that extends from the C-terminal

tri-helical HTH core. Variations to these Winged HTH exist, typically with additional C-terminal β -strands or helices. Both HTH and wHTH TFs typically form dimers that target short DNA repeat sequences (12-14). Both chains in the dimer usually target identical short DNA sequences. Disruption to dimerisation of these TFs often results in inactivation of the TF (15-17).

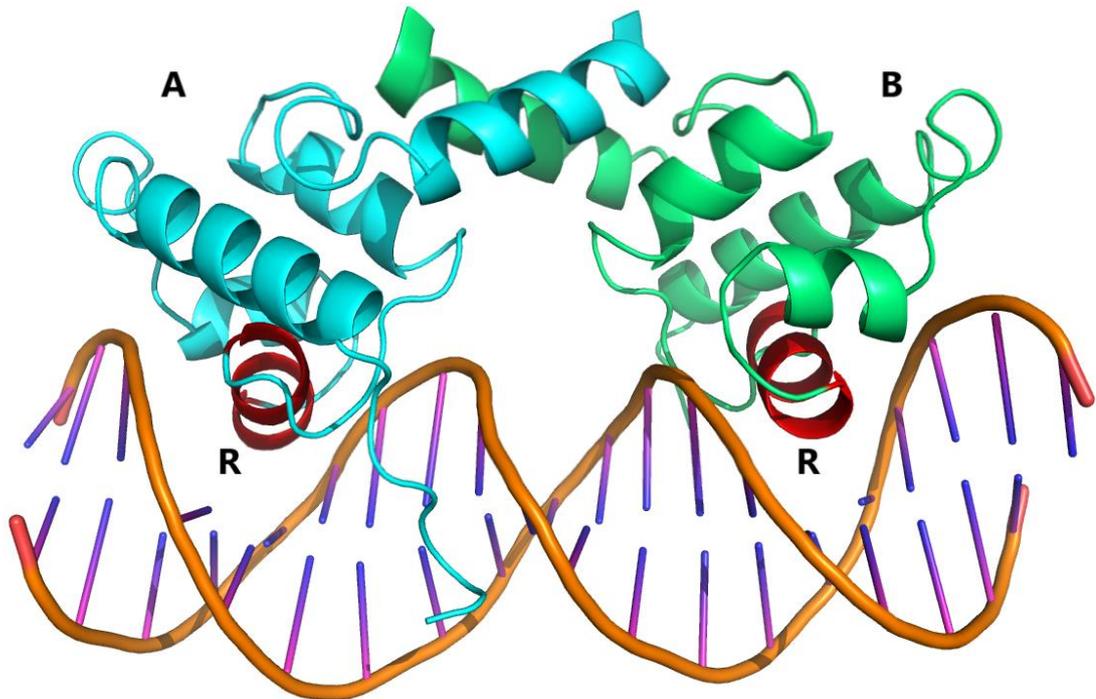


Figure 1.1: Lambda repressor bound to DNA target. Chains A (blue) and B (green) of the lambda repressor dimer are shown with the recognition helix (R) of each chain positioned into the major groove of the DNA target sequence. Structure determined by X-ray crystallography. RCSB PDB: 3BDN (18).

1.2: Bacterial Quorum Sensing

Quorum sensing (QS) is a type of cell-cell communication employed by many bacteria to coordinate expression of phenotypes across a population (19, 20). QS was discovered through investigation of the bioluminescent marine bacterium *Vibrio fischeri* (21, 22). *V. fischeri* bioluminescence only occurs in stationary-phase cultures, however, bioluminescence can be induced in log-phase cultures of *V. fischeri* if they are grown in the presence of pre-

conditioned medium containing extracellular material from a previously bioluminescent active culture (23). A membrane-diffusible chemical signalling molecule, termed an autoinducer, is excreted into the environment by *V. fischeri* (21). This signal molecule, called an N-acyl homoserine lactone (HSL) (24), is produced constitutively at a low level by uninduced cells and gradually increases in concentration with cell density (21). The LuxR transcriptional regulator (25, 26) responds to the accumulation of N-(3-oxo-hexanoyl)-L-homoserine lactone (referred to as 3-oxo-C6-HSL herein), forming an active homodimer that recognises the 20-base pair inverted-repeat DNA target sequence (27, 28), commonly referred to as the *lux*-box (29). Binding of activated LuxR-3-oxo-C6-HSL to the *lux*-box results in transcription of *LuxICDABEG* operon (30), inducing 3-oxo-C6-HSL synthesis and bioluminescence (Figure 1.2). The positive feedback loop created by LuxR activating transcription of the 3-oxo-C6-HSL synthase gene *luxI* ensures QS and bioluminescence are stimulated throughout the cell population.

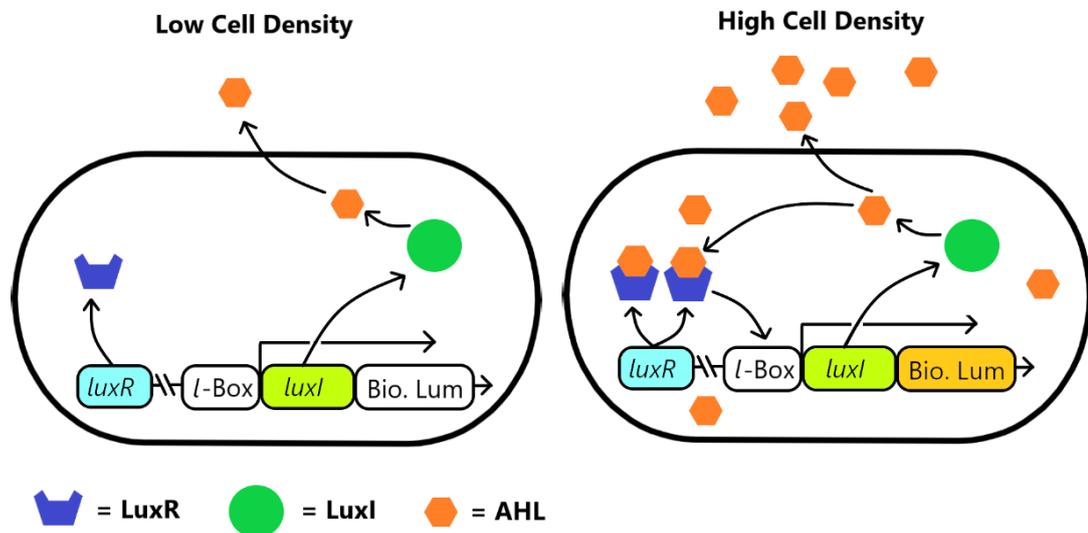


Figure 1.2: The *Vibrio fischeri* QS system. Transitioning from ‘off’ state in low cell density to high cell density results in accumulation of HSL signal molecule (produced by LuxI), formation of LuxR-homodimer and activation of bioluminescence.

Since characterisation of QS in *Vibrio* species, including *V. harveyi* and *V. cholerae*, many QS systems have been identified in various species that regulate a range of phenotypes (19, 31). These include biofilm formation, virulence factors, pigmentation, motility and horizontal gene transfer (20, 32-35). Proteins involved are often homologous with *V. fischeri* LuxR and LuxI but can synthesise and respond to a wide range of distinct N-acyl homoserine signal molecules (Table 1.1). The majority of QS systems within Gram negative species utilise LuxR homologues that respond to HSL signal molecules (31, 33, 36, 37). Gram positive species differ, with signal molecules often being peptides that function in two-component QS systems involving transmembrane histidine kinases (38, 39). Gram positive QS systems have been reviewed in detail in other publications (32, 34).

1.3: N-acyl-homoserine lactone (HSL) synthesis

Signal molecule synthases can be divided into three types, with LuxI homologues the most common (40) and are present in hundreds of species (33). The second type includes LuxM and its homologues AinS and VanM and the third type includes HdtS. LuxI and LuxM homologues metabolise substrates S-adenosyl-methionine and an acylated-acyl carrier protein (derived from fatty acid biosynthesis) for the synthesis of HSLs (36, 41-43). How the HdtS homologues produce their respective signal molecules is not well understood (40). HSLs vary by length, between 4 and 18 carbons in the fatty acyl side chain, and oxidation at the 3rd carbon position (43, 44) (Figure 1.3). Some variations to substrate sources have been noted (33, 45), but are uncommon. HSL synthesis is often highly specific, with each HSL synthase producing a unique HSL for specific QS systems within the same species under autoregulation (37, 46), however, some exceptions exist. *Yersinia enterocolitica* has one HSL synthase (YenI) which produces 3-oxo-C6-HSL, C6-HSL and 3-oxo-C7-HSL disproportionately, with 3-oxo-C6-HSL being most prevalent. This system is not under autoinduction as HSL synthesis is not regulated by either LuxR-homologue TFs YenR or YcoR present within the *Y. enterocolitica* (47, 48).

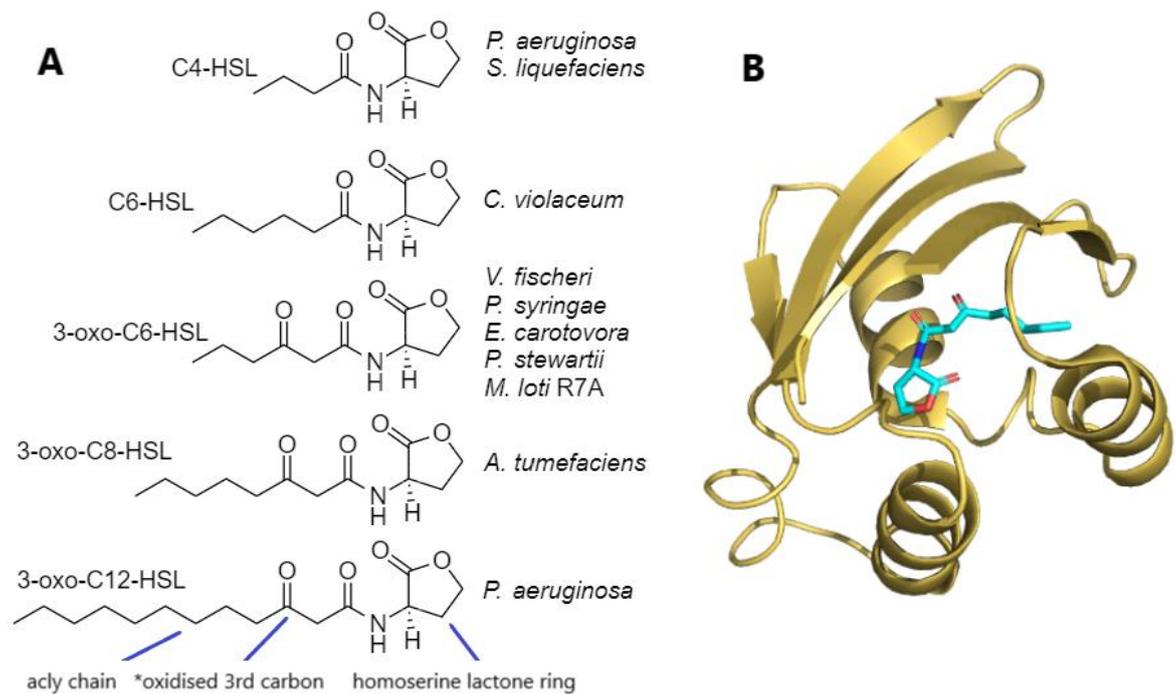


Figure 1.3: N-acyl homoserine lactone structure and target. A list of various N-acyl homoserine lactone (HSL) QS signal molecules with the respective species that it is produced in (A). All HSLs have a similar structure, comprising an acyl chain and a homoserine lactone ring. Some HSLs have the 3rd carbon of the acyl chain oxidised, denoted with a “3-oxo”. The 3-oxo-C12-HSL of *P. aeruginosa* (blue) is embedded in a pocket within the ligand binding domain of LasR (B).

Table 1.1: List of known bacteria phenotypes regulated by QS. Adapted from (49)

Organism	Phenotypes	TF/s	Synthase (signal molecule)	Reference
<i>P. aeruginosa</i>	Exoenzymes, virulence, biofilm formation, motility, iron acquisition, pyocyanin production	RhlR, LasR*, QscR, PqsR	RhlI, LasI, PqsA-D, PqsH (C4-HSL, 3-oxo-C12-HSL, HHQ, PQS)	(50, 51)
<i>P. syringae</i>	EPS, plant colonisation	AhIR	AhII (3-oxo-C6-HSL)	(52)
<i>E. carotovora</i>	Carbapenem, exoenzymes, virulence	CarR, ExpR, VirR	Carl (3-oxo-C6-HSL)	(53)
<i>P. stewartii</i>	Adhesion, EPS, plant colonisation	EsaR	Esal (3-oxo-C6-HSL)	(52)
<i>A. tumefaciens</i>	Ti plasmid conjugation, virulence	TraR*	TraI (3-oxo-C8-HSL)	(52)
<i>C. violaceum</i>	Exoenzymes, antibiotic production (violacein)	CviR	Cvil (C6-HSL)	(54)
<i>S. liquefaciens</i>	Swarming motility, biofilm formation	SwrR	SwrI (C4-HSL)	(55)
<i>V. harveyi</i>	Bioluminescence, siderophores, protease/EPS production, virulence	LuxN, LuxP, CqsS	LuxM, LuxS, CqsA (3-OH-C4-HSL, AI-2, CAI-1)	(37, 56)
<i>V. cholerae</i>	Virulence, biofilm formation, EPS	LuxP, CqsS	LuxS, CqsA (AI-2, CAI-1)	(37, 56)
<i>V. fischeri</i>	Bioluminescence, host colonisation, motility	LuxR, AinR, LuxP, SdiA	LuxI, AinS, LuxS (3-oxo-C6-HSL C8-HSL, AI-2,)	(56)
<i>S. aureus</i>	Virulence, exotoxins, biofilm dispersal	AgrC	<i>agrD^d</i> (AIP)	(57)
<i>E. faecalis</i>	Gelatinase, protease, adhesion, conjunction	FsrC, PrgX	<i>frsD^b</i> , <i>ccfA^b</i> (GBAP, cCF10)	(58, 59)
<i>S. pneumoniae</i>	Competence, virulence, autolysis	ComD	<i>comC^b</i> (CSP)	(60-62)
<i>B. thuringiensis</i>	Exoenzymes, toxins, sporulation, necrotrophism	PlcR, NprR	<i>papR^b</i> , <i>nprRB^{b,c}</i> (PapR, NprX)	(59, 63, 64)
<i>X. campestris</i>	Virulence, biofilm dispersal, EPS	RpfC	PrfB, RpfF (DSF)	(65)

*TF with characterised antagonistic proteins that suppress activation of these proteins (16, 66).

1.4: QS pathogens

Phenotypes involved in pathogenesis and infection, such as exoprotease expression and secretion, EPS production and biofilm formation are often regulated by LuxR-homologue QS systems. For example, *A. hydrophila* is a pathogen commonly implicated in 'motile *Aeromonas* septicaemia' in aquaculture (67), however, *Aeromonas* species are also associated with human intestinal infections (68). There are several virulence factors of *A. hydrophila* which include adhesins, EPS and extracellular excretions, with exoprotease production under QS regulation (69). This QS system includes the LuxR-homologue AhyR which responds to C4-HSL produced by AhyI (70). *P. aeruginosa* is a common human pathogen which notably causes chronic lung infections in people with cystic fibrosis. The LuxR-homologues LasR and RhIR respond to 3-oxo-C12-HSL and C4-HSL respectively (71) and promote the expression of various genes with some overlap. These two QS systems are linked because LasR is responsible for the promotion of RhIR, thereby making LasR a master regulator in the QS regulon of *P. aeruginosa* (72). Use of high-density DNA microarrays has determined that this LasR/RhIR QS regulon influences the expression of over 300 genes, which approximates to 6% of the genome (73). This includes key pathogenic phenotypes like extracellular virulence factor expression and secretion and biofilm formation, however metabolic functions are also under its influence which suggests that QS regulation has been heavily employed to adapt cells to high-cell-density environments (73). Both *Escherichia* and *Salmonella* genera are common human pathogens and express a LuxR-homologue called SdiA (74). This transcriptional activator responds to HSLs like many others in the LuxR-family, however neither genera have so far been found to express their own HSL synthase protein and rely on external sources for HSL. SdiA has a diverse range of HSL sensitivity, with responses to 3-oxo-C8-HSL, C8-HSL and C6-HSL (75). Both lack of HSL synthase and a broad sensitivity suggest that this QS system has evolved to detect high cell density in heterogenous populations (74).

Y. enterocolitica causes gastroenteritis in humans and regulates swimming motility and host cell attachment via a LuxR-homologue QS system (48, 76).

This system includes the LuxR-homologue YenR, which, unlike most LuxR-homologues, loses its DNA-binding function in the presence of its cognate HSL 3-oxo-C6-HSL, produced by YenI. Active YenR in low cell densities promotes the expression of the *yenI*-repressor, YenS, however in high cell density this repression is inhibited due to accumulation of HSL, resulting in sustained HSL synthesis and motility and cell adhesion phenotypes (48).

1.5: Plant associated QS

HSL screening of Gram-negative plant-associated bacteria by Cha et al (77) found the majority of *Agrobacterium*, *Rhizobium* and *Pantoea* isolates, along with about half of *Erwinia* and *Pseudomonas* isolates produce HSLs. Interestingly, there exists some correlation between HSL type and genera, with *Pseudomonas* and *Erwinia* species often produce 3-oxo-C6-HSL and all *Agrobacterium* tested produce 3-oxo-C8-HSL. However, *Rhizobium* species showed significant diversity, with some producing single HSLs and others producing up to seven different molecules.

The plant pathogen *A. tumefaciens* causes Crown Gall disease by infecting the host with Ti-plasmid T-DNA (78). QS is used to regulate Ti-plasmid conjugative transfer between *Agrobacterium* spp.. This QS system includes the LuxR-homologue TraR which responds to 3-oxo-C8-HSL produced by TraI (79, 80). This system also employs a TraR-binding protein called TraM that binds and represses TraR. Since the repression function of TraM is entirely dependent on the activation by TraR, TraM is termed an antiactivator. Antiactivation by TraM leads to QS suppression, which is discussed in more detail in section 1.7.

P. carotovorum is a plant pathogen that causes soft rot disease in a range of species including both vegetables and ornamentals. As is characteristic of plant pathogens that degrade plant tissues, *P. carotovorum* produces a range of plant cell wall degrading enzymes (PCWDE), leading to soft rot disease (81, 82). PCWDE production is under QS regulation. At low cell density the expression of PCWDE remains uninduced, however, once a higher cell density is attained and the population is sufficiently large enough to overcome the host

response, QS induces production of PCWDE. The QS system includes the LuxR-homologue ExpR1 which responds to the HSL 3-oxo-C6-HSL produced by ExpI. A secondary QS system involving the LuxR homologue CarR regulates expression of the β -lactam antibiotic carbapenem. CarR is activated by 3-oxo-C8-HSL produced by CarI. Both ExpRI and CarRI systems function independently since knockout of ExpI does not prevent CarR from functioning, nor does knockout of CarI prevent ExpR from functioning. This model eloquently highlights the specificity of HSL recognition in separate QS systems within the same cell (53, 83).

P. stewartii causes Stewart's wilt disease which afflicts maize. Expression of the virulence factor stewartan is under QS regulation. The LuxR-homologue in this system is EsaR which responds to the EsaI-produced 3-oxo-C6-HSLHSL (84). EsaR is similar to YenR in that it becomes inactivated by its cognate HSL. Inactive EsaR is unable to repress expression of the transcriptional activator RcsA, resulting in expression of stewartan (85).

Several plant symbiosis-associated QS-regulated systems are found in plant-associated bacteria. For example, *P. putida* competes with plant pathogens making it beneficial to the plant host (86). *P. putida* has its biofilm formation under QS regulation. This QS system includes the LuxR and LuxI homologues, PpuR and PpuI. PpuR shares significant sequence similarity (62%) to LasR of *P. aeruginosa* and respond to the same HSL, 3-oxo-C12-HSL.

The plant legume family Fabaceae contains many significant species to agriculture including harvestable crops, livestock feed and for soil nitrogen enrichment. Legumes can form symbiotic relationships with nitrogen-fixing, Gram negative bacteria, commonly referred to as Rhizobia. *R. leguminosarum* bv. *viciae* RhiR (87-89) is a heavily investigated Rhizobial member which forms such symbiotic interactions with legume species. This interaction requires a large set of genes necessary for establishing plant-bacterial communication, subsequent root infection and nodulation. Bacterial populations within the plant root nodules are then able to fix atmospheric nitrogen and supply the host with ammonium and in return receive a supply of dicarboxylic acids, primary malate and succinate (88, 90). These genes are

often present on large plasmids that can be horizontally transferred to other bacterial species.

Rhizobium leguminosarum bv. *viciae* encodes a LuxR-type QS system that regulates both host nodulation and horizontal gene transfer (87). The plasmid pRL1JI (91) carries genes required for plant symbiosis and nitrogen fixation (92, 93). *R. leguminosarum* bv. *Viciae* has four distinct QS systems, including the TraRI system that regulates pRL1JI conjugative transfer (Figure 1.4) (94). The chromosomal CinRI system regulates growth and is linked to the TraRI system via an intermediate transcription regulator, BisR (94, 95). A RhiRI system located on pRL1JI influences nodulation (87) and the RaiRI system is on the secondary plasmid pIJ9001 (96), however the role of the RaiRI system is not well understood.

Similar to QS systems of *A. Tumefaciens* and *P. aeruginosa*, the QS system of *R. leguminosarum* incorporates antiactivators called CinS (97) and TraM. Bacterial two-hybrid assays demonstrate CinS binds (97) and inhibits PraR, a transcriptional repressor of both *rhiR* and *raiR*, allowing both QS systems to be active (98, 99). TraM binds and inhibit TraR. TraM has been investigated in great molecular detail and is described in section 1.7. QS systems are also present in other Rhizobial genera, including *Bradyrhizobium*, *Mesorhizobium* and *Sinorhizobium*; all three showing QS-regulated nodulation and plasmid transfer (99). Almost all utilise a LuxR-homologue and HSL signal molecule synthase for QS activation of these phenotypes.

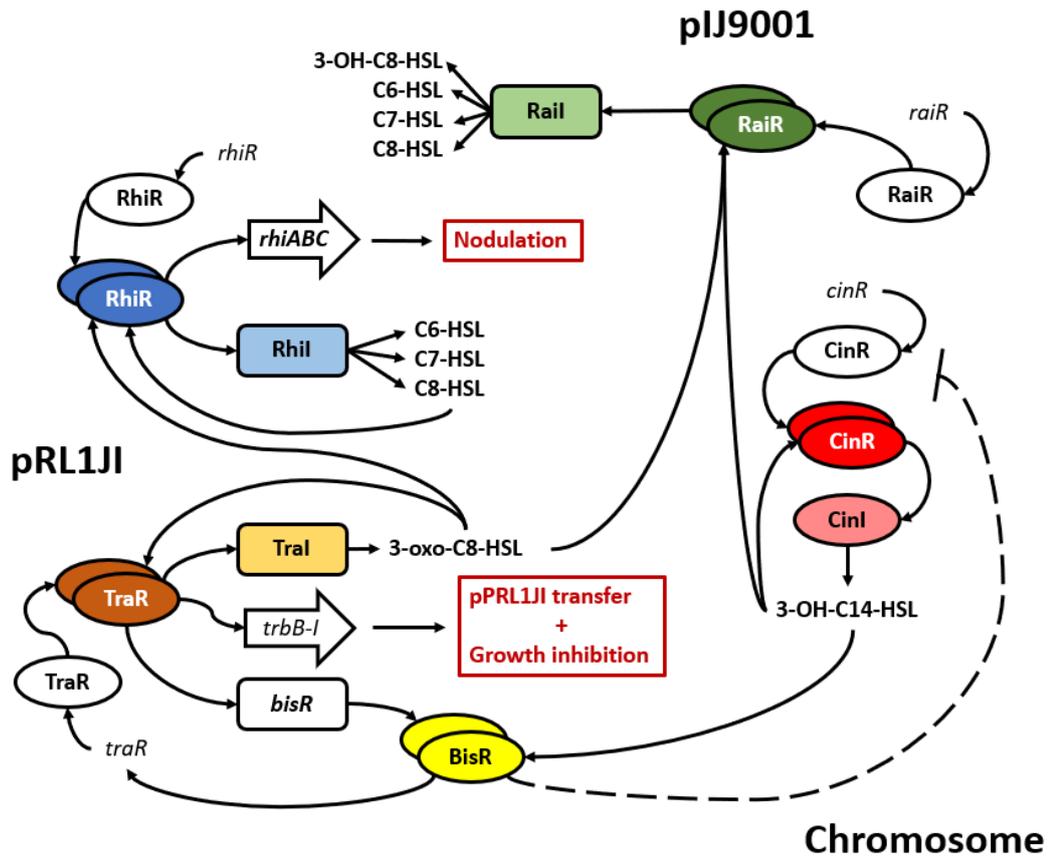


Figure 1.4: of QS regulation in *R. leguminosarum*. The QS systems TraRI and RhiRI located on pRL1JI regulate conjugal transfer, growth inhibition and nodulation. Chromosomal located CinRI system acts as a regulator to the TraR system. The RaiRI system on pIJ9001 produces several HSLs but its function is not known. Adapted from (94).

1.6: LuxR-Family transcriptional regulators

LuxR-family HSL receptor proteins function as transcriptional regulators (31) and each share a similar protein domain architecture (33). LuxR proteins are generally ~250 amino acids long and contain an N-terminal signal molecule recognition domain and C-terminal DNA-binding helix-turn-helix (HTH) motif (100). The HTH DNA-binding sequence is highly specific to each LuxR regulator. The DBD of TraR targets a 9-bp sequence using 5 residues located on the recognition helix (101, 102). In particular, two adjacent arginine residues (Arg206 and Arg210) make base-specific hydrogen bonds. The DNA recognition helix also exhibits an overall positive charge on its surface which

complements the negatively charged DNA it interacts with. Binding is further established by having a second, identical but inverted, 9-bp sequence existing immediately downstream. The complete 18-bp inverted repeat accommodates a TraR dimer with two DBDs with two-fold rotational symmetry(103). This 18-bp inverted repeat is called the *tra*-box. The positioning of the DBDs of the TraR dimer and the distances between targeted bases within the *tra*-box are essential for TraR binding. Additional bases inserted into the centre of the inverted repeat prevent TraR from binding.

All LuxR-homologues that have been characterised appear to target similarly structured DNA sequences as homodimers (33). Most LuxR-homologues only form homodimers after binding their cognate HSL molecules which is evident in FPLC size exclusion chromatography of TraR (17), MtrR and crystallographic structures of both TraR (102) (Figure 1.5) and LasR (104). However, there are some QS transcriptional regulators that bind DNA in the absence of HSL and are deactivated in the presence of their cognate HSL. One of the most investigated examples of these is EsaR, which is unusual in that it represses a phenotype (EPS production) but only in the absence of its cognate HSL 3-oxo-C6-HSL (105). There are several other transcriptional regulators that are deactivated by their cognate HSLs, all of which are closely related to EsaR. Dimerisation is also critical in these EsaR-like LuxR-homologues since it is often the disassociation of the dimer in the presence of the cognate HSL that causes loss of DNA binding function and reduced protein stability.

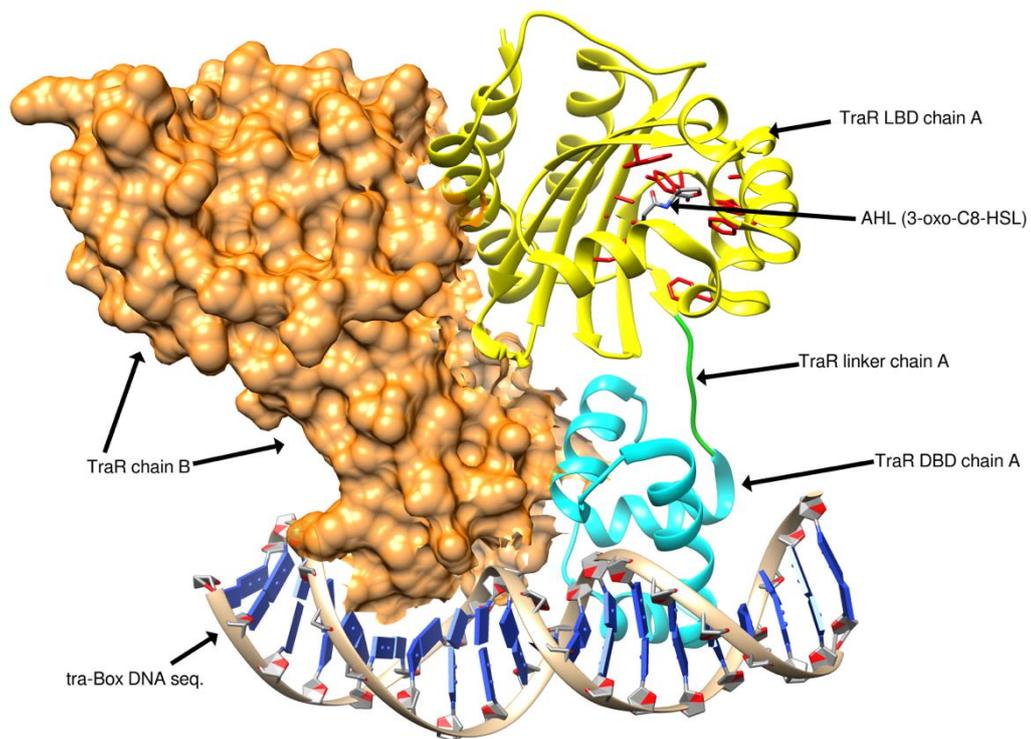


Figure 1.5: TraR of *A. tumefaciens* binding DNA. TraR in its active form is an HSL-bound homodimer that can bind its target (*tra*-Box) DNA sequence. The protein can be divided into three segments, the ligand binding domain (LBD) (yellow), the DNA binding domain (DBD) (cyan), and the linker between these (green). When dimerised (second chain shown as orange surface map) the structure can bind DNA. This structure was generated by X-ray crystallography. RSCB PDB: 1H0M (101).

LuxR-homologues typically respond to a narrow range of HSL molecules, often with a single dominate type, however, some exceptions exist. As mentioned previously, SdiA and RhiR are examples of LuxR proteins with sensitivity to a broad range of HSLs. An example of this is LasR which has strongest affinity to 3-oxo-C12-HSL (its cognate HSL synthesised by LasI), followed by 3-oxo-C14-HSL and in decreasing affinity that correlates to acyl-side chain length, C10, C8 and C6 (106).

HSL specificity by LuxR-homologues is imposed by some structural features of the LBD (104). Many LuxR-homologue LBD are pockets built by hydrophobic beta-sheet and adjacent alpha-helices. The third loop (L3),

adjacent to the first β -sheet of the LBD, covers the surface of this binding pocket (Figure 1.6). This loop has some flexibility when no HSL is present in the LBD pocket, resulting in increased internal solvent exposure which destabilises the LuxR-homologue. The presence of a HSL reduces internal solvent exposure and stabilises the protein, however, HSLs with short fatty acid chains may still not prevent destabilisation. It is evident that LBD specificity to HSL's is heavily associated with the length of the L3 loop. Furthermore, specific hydrophobic residues presented in the binding pocket have greater affinity to specific HSL types.

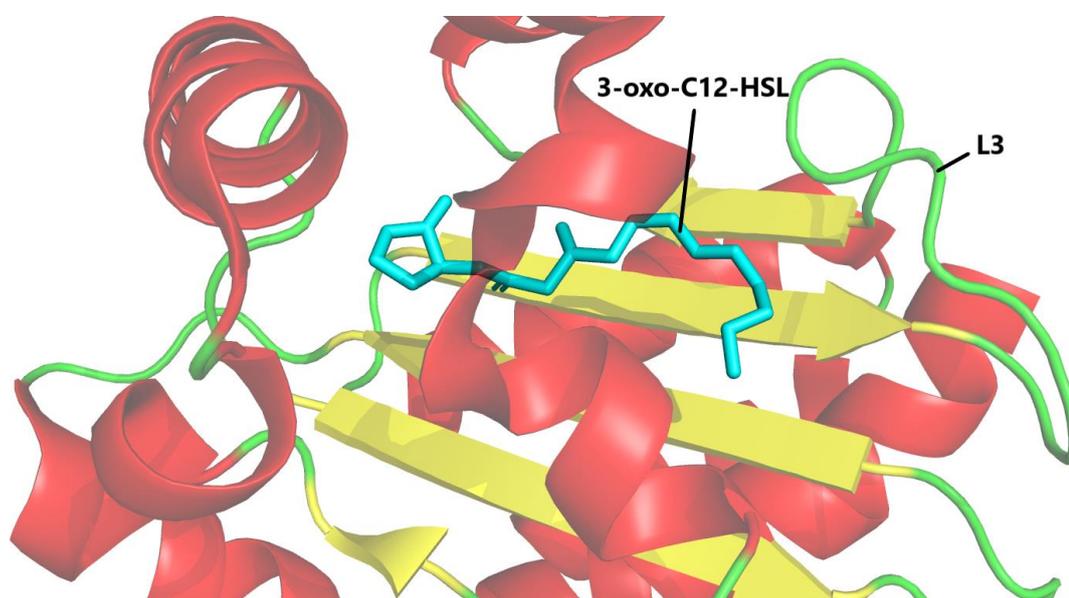


Figure 1.6: LasR ligand binding pocket. The LBD of LasR (coloured by secondary structure) with the 3-oxo-C12-HSL (cyan) inside the ligand binding pocket along with the loop (L3) that interacts with the acyl chain, blocking solvent entry into the pocket and stabilising the LBD of LasR. RCSB PDB ID: 3IX3 (107).

1.7: QS inhibition

Inhibition of QS has been observed in several species and frequently occurs through interference or degradation of either the transcriptional activator protein or its cognate signal molecule. Signal molecule degradation can be

enzymatically achieved by lactonases or decarboxylases that hydrolyse the HSL ring or acylases and deaminases that cleave the acyl side chain. Both enzymes render the HSL molecule ineffective in signalling (108). HSL acylases may target distinct HSL species or exhibit a broad spectrum of activity. Acylases expressed by various *Streptomyces* species predominantly target HSLs with sidechains having 6 or more carbons whereas the acylase of *P. aeruginosa* cleave 3-oxo-C6-HSL, but not the C4-HSL used by its the RhIR-QS system. The *P. aeruginosa* LasR-QS system uses 3-oxo-C6-HSL and therefore acylases expressed lowers the concentration of 3-oxo-C6-HSL and inhibits QS activation (109). *A. tumefaciens* K588 and M103 expresses a lactonase that reduces 3-oxo-C8-HSL, preventing QS activation required for transfer of its Ti plasmid (110, 111).

Since many QS systems are associated with pathogenesis it is unsurprising that eukaryotes have also evolved mechanisms of QS inhibition. This has been characterised mostly in plants which are susceptible to QS-regulated disease and often results in signal molecule degradation via lactonases. Additionally, in mammals, antibodies are generated against 3-oxo-C12-HSL, directly reducing QS response in *P. aeruginosa* reducing QS-activated virulence, similarly, antibodies are generated against the autoinducing peptide-4 (AIP-4) of *S. aureus*, also reducing QS-activated virulence (112).

1.8: Antiactivation of LuxR-family proteins

QS systems can be inhibited by proteins that antagonise LuxR function. These antagonistic proteins directly interfere with their targeted LuxR-family transcriptional activator to suppress its function and are termed antiactivators. Although few antiactivators have been characterised, the modes of action for the TraR-antiactivator TraM of *A. tumefaciens* and the LasR-antiactivator QsIA of *P. aeruginosa* have been investigated in detail using molecular and biochemical experiments together with crystal structure determination. The TraM interaction with TraR was first demonstrated using yeast two-hybrid assays (113) without the presence of TraR's 3-oxo-C8-HSL signal molecule. Gel filtration and isothermal titration calorimetry (ITC) experiments

demonstrated TraR and TraM are present in a 1:1 ratio in complexes. The complexes are frequently heterotetramers, indicating TraM can bind TraR dimers (114, 115). It was also determined through ITC that a single TraM bound to TraR can recruit a second TraM monomer, accounting for the preferential heterotetramer complex composed of two TraR and two TraM proteins. Interestingly, TraM expression is promoted by active TraR, which creates a negative feedback system (116), suggesting that there is a selective pressure against spontaneous QS activation or lingering activation if cell density is not high enough. It is also worth noting that TraR expression is often stimulated in *Agrobacterium* species by signal molecules from the cells of the plant itself, termed opines. Opines and TraM therefore function antagonistically to tightly regulate QS activation and the threshold for induction of Ti plasmid transfer (117).

Yeast two-hybrid assays (113, 115) indicate TraM targets TraR between amino-acid residues 121 to 234 of TraR. This range includes both the C-terminal end of the LBD (186-234) and the linking region between the LBD and the DBD. TraM does not interact with the DNA-binding domain of TraR. TraM increases rate of proteolytic degradation of TraR (118). TraM is not itself degraded in the TraR-TraM interaction. TraM likely achieves degradation of TraR by destabilising TraR structure and exposing it to proteolytic enzymes in the cell (119).

The structure of TraM in complex with TraR was solved using by X-ray crystallography and shed further light on the mechanism of TraM-mediated TraR antiactivation (66) (Figure 1.7). TraM is a 11.2 kDa protein containing two antiparallel helices. Supporting yeast two hybrid assays, TraM was found complexed with TraR through interactions between the middle-to-C-terminal portion of TraR, between the N-terminal LBD and C-terminal DBD. In these structures TraM appears to distort the dimerisation and DNA-binding domains that are required for TraR to form a stable complex with the *tra*-box. TraM can bind TraR in complex with DNA or not; TraR is removed from its DNA target when bound by TraM.

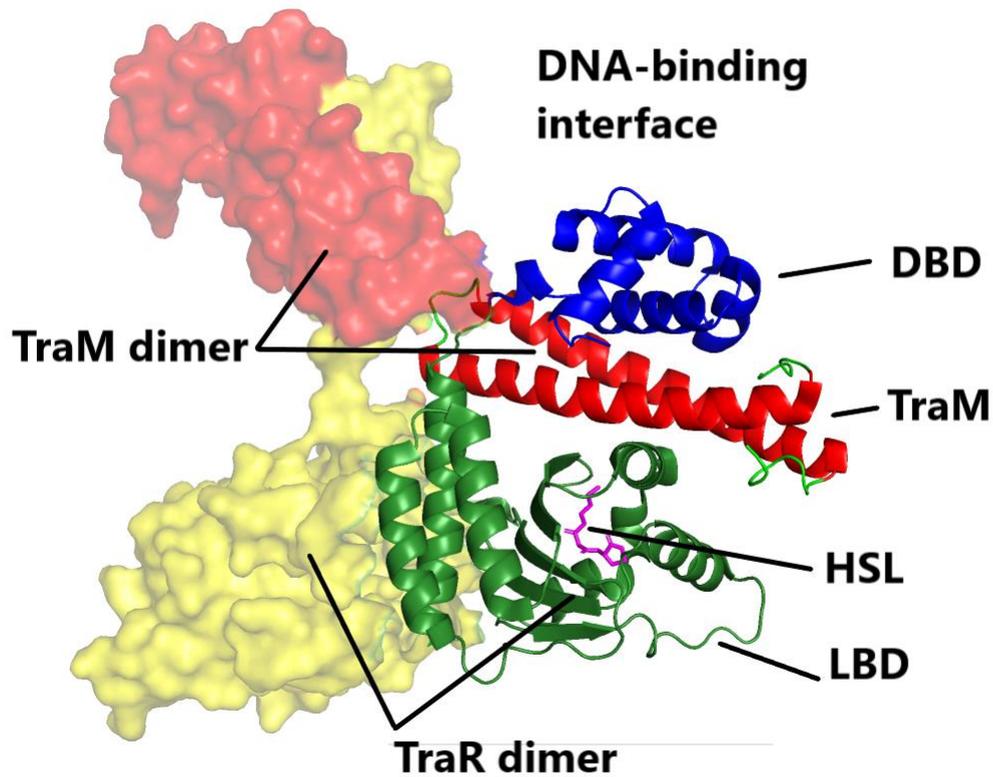


Figure 1.7: TraM-TraR heterotetramer. TraM (red) binding between the DBD (blue) and LBD (green) of TraR. TraR is its active homodimer complex (partner yellow) bound to its HSL ligand (magenta). TraM moves the DBD of both TraR chains apart, preventing the DNA-binding interface from functioning. RCSB PDB ID: 2Q00 (Uniprot ID, Q57471) (66).

TraR and TraM both have hydrophobic surfaces which are positioned in close proximity with each other in the cocrystal structure (Figure 1.8). This region includes the residues Ala32-Ile36 and Leu88-Val98 of TraM and residues Leu170 to Leu199 of TraR. Residue Trp186 and the LuxR-homologue conserved Leu199 of TraR seem to be critical to the interaction between TraR and TraM. The overall hydrophobic surfaces of this area (120) is likely required to form a stable interaction with TraR. A Hydrophobic surface area on the antiactivator of *Klebsiella pneumoniae* (NifL) (121) is critically required to bind the hydrophobic region in the C terminus of the LuxR homologue NifA.

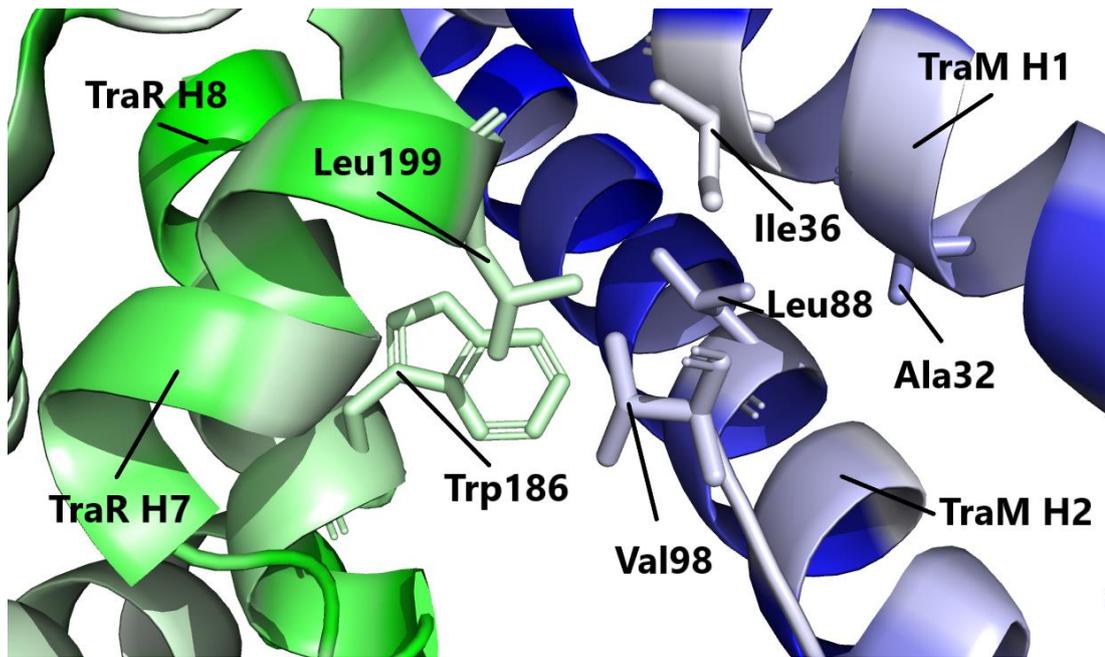


Figure 1.8: TraR-TraM binding region. Residues coloured by hydrophobicity (white is most hydrophobic), TraM (blue/white) alpha helices (TraM H1/H2), TraR (green/white) helices (TraR H7/H8), and residues labelled. Hydrophobicity is conserved at the binding interface where the critical residues of TraM (Ala32, Ile36, Leu88, Val98) and TraR (Trp186, Leu199) are located. RCSB PDB ID: 4NG2.

Three distinct antiactivators have been identified in *Pseudomonas* spp.. The 11.8 kDa antiactivator QsIA inhibits LasR, QteE (21.8 kDa) inhibits RhIR (27.6 kDa) and QscR inhibits both LasR and RhIR (122). QsIA is the most well-characterised of the *Pseudomonas* spp.. antiactivators. QsIA forms a heterotrimer with LasR. Analytical ultracentrifugation has shown that a QsIA dimer binds a single LasR protein forming a QsIA-QsIA-LasR heterotrimer. The QsIA dimer targets the ligand binding domain (LBD) of LasR, preventing LasR dimerisation and QS activation (16, 123). QsIA does not share any primary sequence similarity with TraM, indicating it has evolved independently. Electrophoretic mobility shift assays (EMSA) showed that LasR disassociates from its DNA in the presence of QsIA (123). Similar to TraM-TraR interactions, two-hybrid assays demonstrate QsIA binds the LBD of LasR (16) (Figure 1.9).

Mutations to the hydrophobic region of QsIA that span residues 74-107 prevent dimerisation of QsIA and antiactivation of LasR (16).

Structure determination of QsIA showed it is a predominately alpha helical protein like TraM but adopts a significantly different structure (123). Structure solution of QsIA-LasR in complex confirmed the QsIA-QsIA-LasR heterotrimer complex, with the binding interface located on the LBD of LasR. The QsIA dimer blocks the space required for two LasR proteins to bind and form a dimer required for QS activation. QsIA does not appear to prevent LasR from binding its 3-oxo-C12-HSL ligand, suggesting that QsIA physically blocking the binding region required by LasR to form an active dimer is the primary means by which antiactivation is achieved.

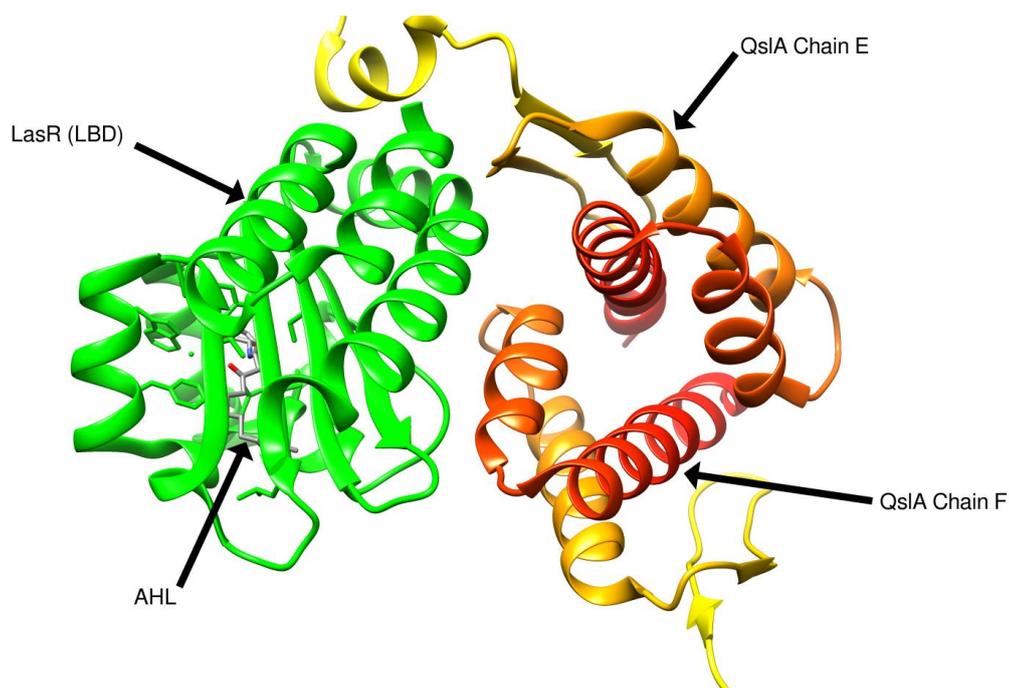


Figure 1.9: QsIA homodimer binding ligand-binding domain (LBD) of LasR. The crystal structural RSCB PDB ID, 4NG2, showing LasR (chain A) and QsIA (chain E and F). The lasR LBD (green), with bound HSL, interacting with two QsIA chains (N to C terminal, yellow to red). QsIA the dimer inhibits LasR by binding its LBD (16).

The structure solution of QsIA-QsIA-LasR heterotrimer reveals extensive hydrophobic surfaces where the QsIA dimer targets the LBD of LasR (Figure 1.10). This region includes residues Leu3, Val4, Phe7, Leu30, Phe143, Val147-Pro149, Trp152 and Met153 of LasR and Val20, Ile23, Ile24, Tyr27, Leu30, Leu45-Val49 and Met89 of QsIA. The exposed aromatics are of particular interest in this region since they have the potential to form extensive hydrogen bonds with LasR. Both QsIA and TraM achieve QS antiactivation by interfering with the LuxR-homologue, however, the mechanism through which antiactivation is achieved differ. This variation suggests antiactivation has evolved independently in different QS systems.

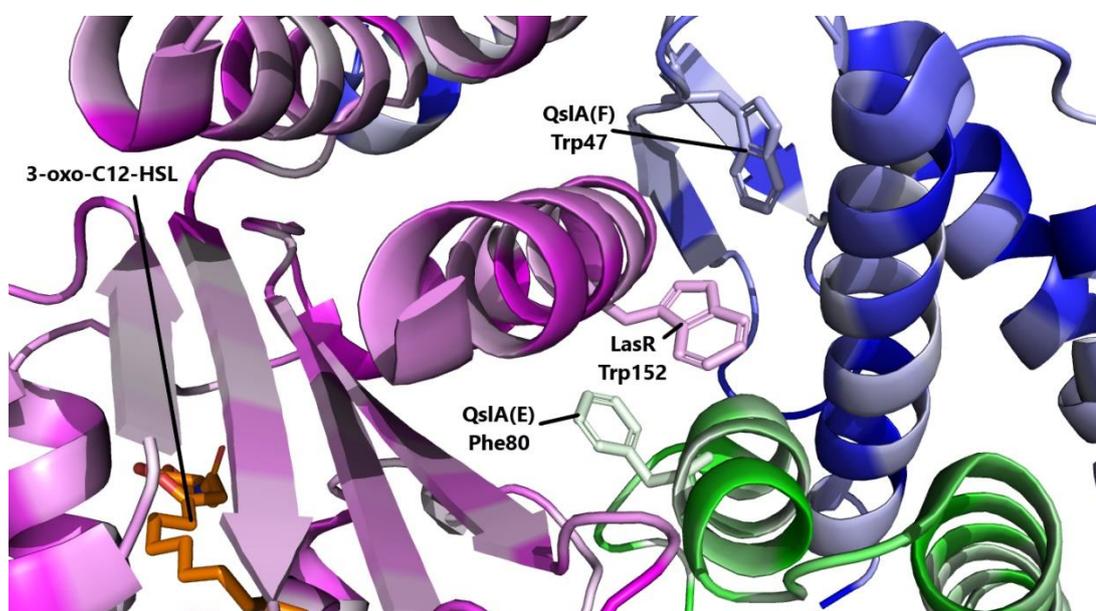


Figure 1.10: QsIA-QsIA-LasR heterotrimer hydrophobic surface interface. LasR (chain A, magenta/white) and QsIA (chain E, green/white and F, blue/white). Residues coloured by hydrophobicity (most hydrophilic white) with residues Trp152 of LasR (chain A), Phe80 of QsIA (chain E) and Trp47 of QsIA (chain F) all sit in the proposed binding space. 3-oxo-C12-HSL present within the LBD of LasR and not effected by QsIA binding. Structure RSCB PDB ID 4NG2 (123).

1.9: Horizontal transfer of genes for nitrogen-fixing symbiosis in *Mesorhizobium loti* R7A

Mesorhizobium species are nitrogen-fixing plant symbionts like *R. leguminosarum* described in section 1.5. Nitrogen fixation and nodulation by *Mesorhizobium* and *Rhizobium* species is similar, however, plant hosts targeted for nodulation differ. *Mesorhizobium* species nodulate a diverse array of legumes, with the *M. loti* specifically nodulating *Lotus corniculatus*. The genes for nitrogen fixation in *Mesorhizobium* spp. are generally encoded on the chromosomally-integrated mobile genetic elements instead of plasmids (124). These elements are termed integrative and conjugative elements (ICEs). The first symbiosis ICE was discovered in *Mesorhizobium loti* strain R7A in New Zealand soils. Strain R7A was used in trials as an inoculant for *Lotus corniculatus* in a field site originally devoid of symbiotic mesorhizobia capable of forming a symbiosis with *Lotus* spp. However, these field sites also contained several, previously uncharacterised non-symbiotic *Mesorhizobium* species which subsequently acquired nodulation and nitrogen-fixation genes from *M. loti* R7A through horizontal gene transfer. The new symbiotic species were found within the root nodules of *L. corniculatus* plants harvested several years after *M. loti* R7A was introduced to the field site. Further investigation of *M. loti* strain R7A revealed that its nodulation and nitrogen fixation genes are encoded in a 501.8 kb integrative and conjugative element (ICE), herein referred to as ICE M/Sym^{R7A} (125) (Figure 1.11) (126). ICE M/Sym^{R7A} integrates within the chromosomally-located *phe*-tRNA gene (124) in *Mesorhizobium* spp.. ICE M/Sym^{R7A} only horizontally transfers when cultures are at high cell density and the quorum-sensing regulators TraR and Tral1 control induction of ICE M/Sym^{R7A} excision and conjugative transfer (127).

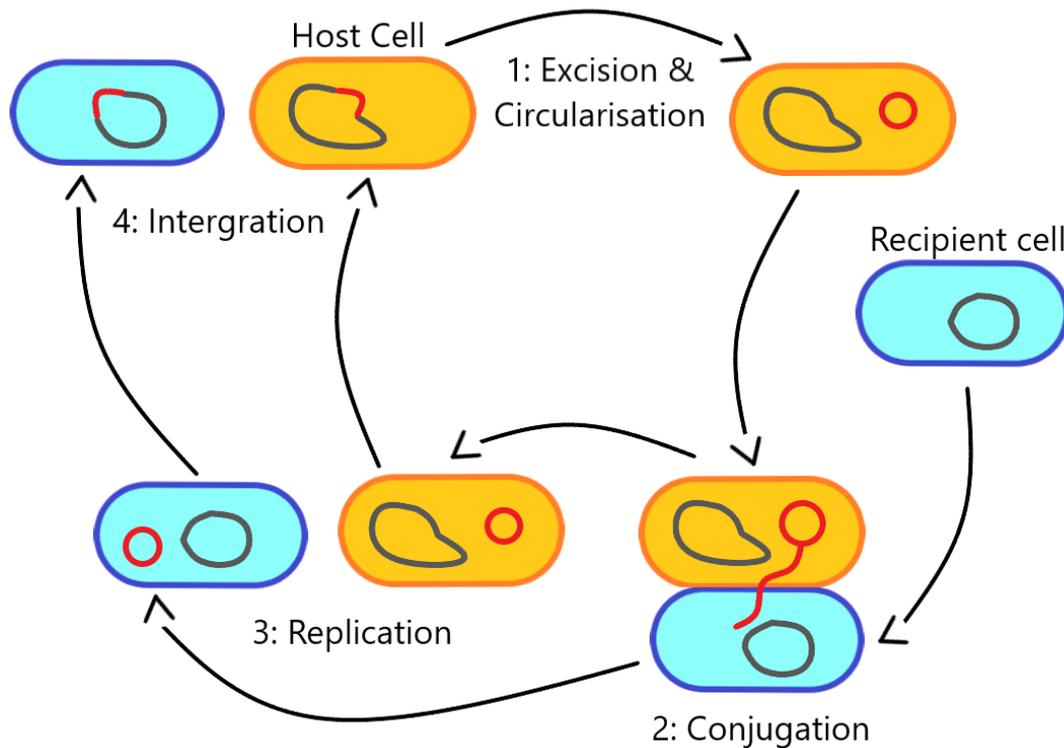


Figure 1.11: ICE excision and transfer. An example of horizontal gene transfer where a mobile genetic element (coloured red) transfers from one bacterial chromosome to another through conjugation (Figure adapted from (128)). The genetic element excises from the host chromosome and circularises (1) before a strand of the dsDNA is transferred to an adjacent cell through a pore structure (2). The remaining ssDNA strands in either cell undergoes replication (3) to synthesize a matching strand before the completed dsDNA genetic element is integrated into the host chromosome (4).

1.10: Regulation of symbiosis island transfer by quorum sensing

Integration and excision of ICE M/Sym^{R7A} is catalysed IntS, a phage-P4-family tyrosine recombinase that catalyses site-specific recombination between sequence called attachment sites (*att*). The bacterial chromosome integration *phe*-tRNA gene contains the bacterial attachment site (*attB*, which recombines with the ICE M/Sym^{R7A}) and the product attachment site (*attP*, present on the circular ICE M/Sym^{R7A}). During ICE M/Sym^{R7A} integration, recombination between *attB* and *attP* sites generates two new hybrid attachment sites *attL* and *attR* that form a direct repeat sequence flanking integrated ICE M/Sym^{R7A} .

In the absence of other factors, IntS-mediated recombination heavily favours the formation of *attL* and *attR* and drives ICEM/Sym^{R7A} chromosomal integration. In the vast majority of cells ICEM/Sym^{R7A} is maintained in an integrated state (125). However, in stationary-phase cultures ICEM/Sym^{R7A} excision and transfer occur at 10-100-fold greater frequency (125).

The recombination directionality factor (RDF) of *M. loti* R7A, RdfS (129), induces the integrase IntS to catalyse excision of ICEM/Sym^{R7A} (130). RDFs are typically small proteins containing a winged HTH (wHTH) motif. RdfS expression is increased in stationary phase cultures. The HSL-synthases TraI1 produces 3-oxo-C6-HSL as the main product. When TraR is activated by 3-oxo-C6-HSL it stimulates the expression of genes *msi172* and *msi171* and the apparent pseudogene, *tral2*. The *msi172* and *msi171* open-reading-frames produce a single polypeptide product through ribosomal frameshifting, called FseA (131, 132). This frameshifting motif is located within the 3' end of the *msi172* open-reading-frame (133). FseA activates expression from the *rdfS* promoter which controls expression of *rdfS*, *traF* (encoding a protease for pilis protein maturation), *msi107* (encoding a predicted murein hydrolase) and the conjugative relaxase gene *rlxS*. The *traF*, *msi107* and *rlxS* gene products are essential ICEM/Sym^{R7A} conjugation, thus bringing horizontal transfer of ICEM/Sym^{R7A} under QS control (Figure 1.12) (35, 125, 130, 134). The R7A TraR and TraI1 proteins are most closely related to TraR and TraI of *A. tumefaciens* and *R. leguminosarum* bv. *viciae* (79, 95). Introduction of an additional copy of R7A *traR* on a plasmid increases ICEM/Sym^{R7A} excision frequency to 100% (35), suggesting that TraR concentrations are limiting in wild-type R7A cultures.

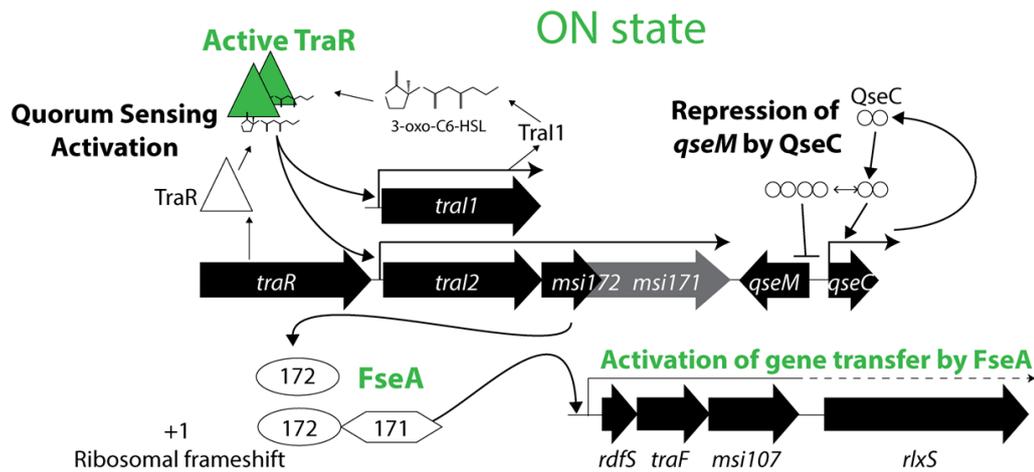


Figure 1.12: QS activation in *M. loti* R7A. Quorum sensing regulation of symbiosis island transfer in *M. loti* R7A homologous to the LuxR/I system. TraR becomes an active homodimer in the presence of 3-oxo-C6-HSL (synthesized by Tra1). Active TraR promotes expression of Msi172 and Msi171 which frequently exist as a fused product due to a ribosomal frameshift motif. The Msi172-Msi171 fused product, FseA, is a transcriptional activator that promotes the expression of genes required for gene transfer. Figure adapted from (134).

1.11: The master antiactivator QseM The ICEM/Sym^{R7A} QS and transfer system is repressed by the QS antiactivator QseM (Figure 1.13). The *qseM* gene is transcribed towards *msi172-msi171*. QseM expression is under regulation of the 'quorum sensing and excision modulator' (QseC) protein, a 69-residue protein belonging to the Xre family of HTH DNA binding proteins, expressed adjacently to *qseM* (130). R7AΔ*qseM* mutants are derepressed for QS and exhibit 100% excision of ICEM/Sym^{R7A} (130). Bacterial two-hybrid assays demonstrate QseM binds TraR. QseM only binds TraR in the presence of 3-oxo-C6-HSL, suggesting that QseM only targets activated TraR, conceivably when TraR is in a dimerised state. It is noteworthy that TraM of *A. tumefaciens* binds either the monomeric or dimeric forms of TraR, potentially indicating that QseM and TraR are dissimilar in their mechanism of antiactivation (115). Two-hybrid assays also demonstrate that QseM binds the transcriptional activator FseA, so QseM is distinct from all other characterised

antiactivators in that it binds two different proteins (135). By targeting both TraR and FseA, QseM is able suppress ICEM/Sym^{R7A} excision and transfer on two levels; antiactivation of QS by binding TraR and by binding FseA it also inhibits *rdfS*, *traF*, *msi107* and *rlxS* expression.

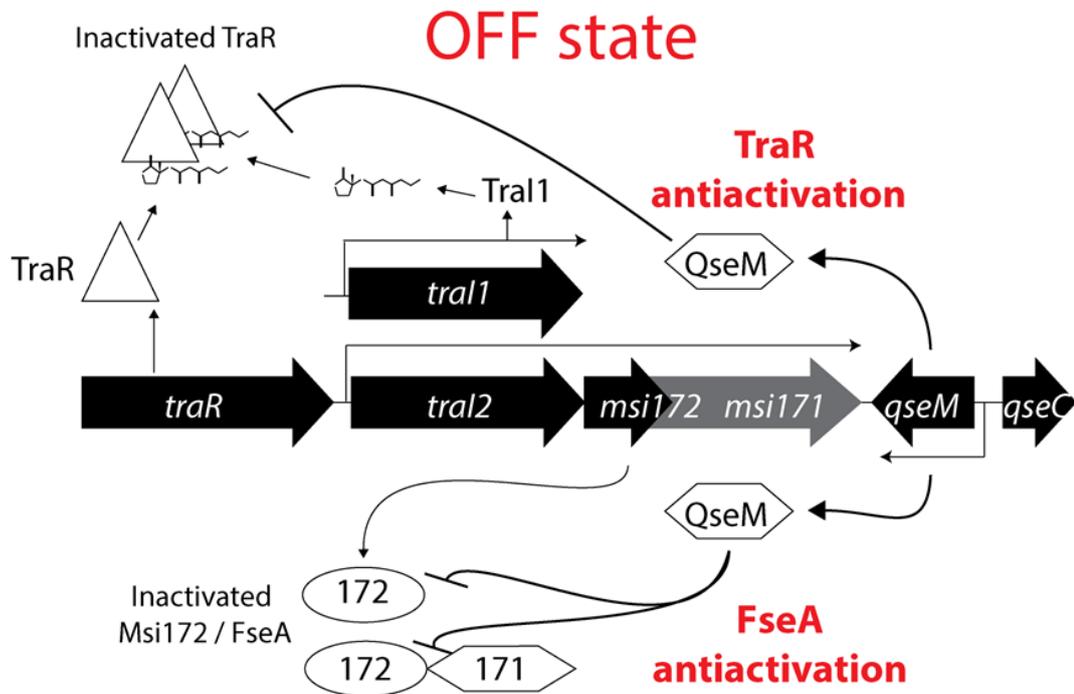


Figure 1.13: QS inactivation in *M. loti* R7A by the antiactivator QseM. QseM binds the TraR homodimer and FseA. QseM prevents either TraR or FseA from promoting gene expression, effectively inhibiting QS activation and gene transfer. QseM is targets the Msi172, either as by itself or when expressed as FseA (Msi172-Msi171 fusion). Figure adapted from (134).

QseM is an 83-amino acid long protein and thus similar in size to both TraM and QsIA, however, it shares no sequence similarity with either. Amino acid sequence analysis of QseM shows it is distantly related to the C-terminal portion of FseA encoded by *msi171*. Automated annotations of both QseM and Msi171 classify these proteins as containing a domain of unknown function (DUF) 2285 (COG5419 pfam10074) (130).

Circular dichroism (CD) experiments on 6H-QseM were carried out by Dr. Joshua Ramsay as part of early structure analysis of the QseM (unpublished

data). CD is an effective method for determination of secondary structure composition of a protein based on how it absorbs certain light. The use of CD for protein structure analysis has been published showing how different protein secondary structures appear (Figure 1.14) (136). CD of hexahistadine-tagged QseM protein (6H-QseM) revealed a high ΔE (difference of light absorption) between wavelengths 190-196 along with a low ΔE at wavelength 205-220 before gradually returning to baseline (Figure 1.15). This profile is consistent with 6H-QseM being predominantly alpha-helical. No other structural analysis of QseM had been carried out prior to this thesis.

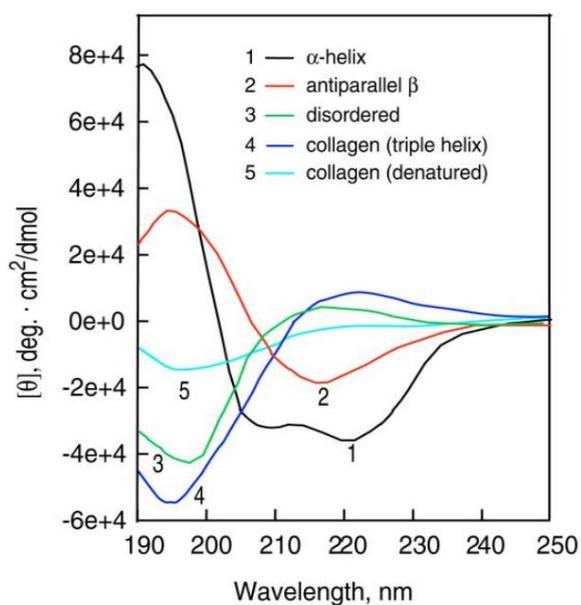


Figure 1.14: Circular dichroism profiles of various secondary structures. Light absorption plotted on the y axis against wavelength on the x axis. Figure sourced from (136).

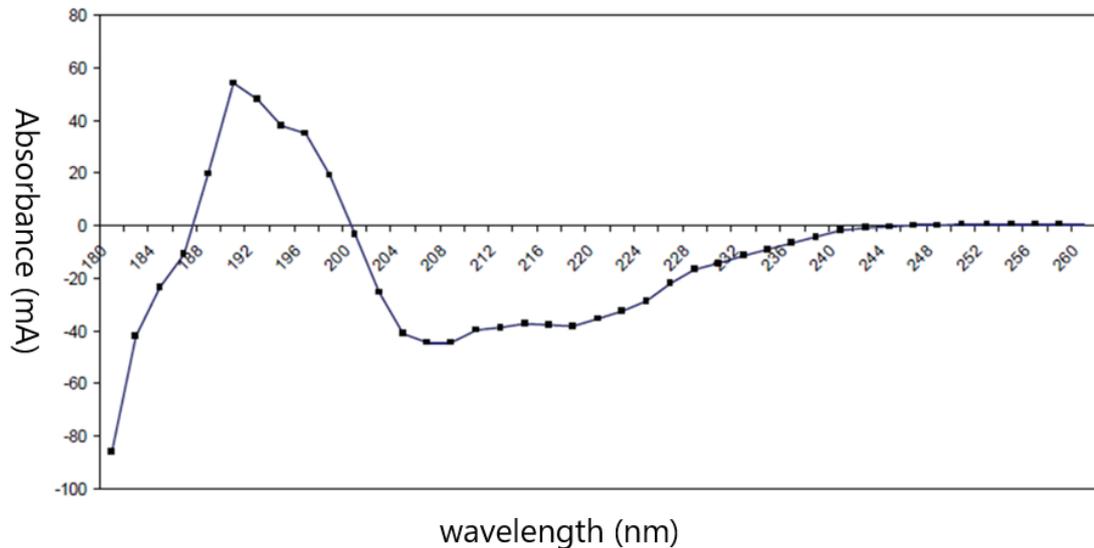


Figure 1.15: Circular dichroism of 6H-QseM. The Circular dichroism plot of 6H-QseM displays a $\Delta\epsilon$ pattern that indicates it has a predominantly alpha-helical secondary structure.

Domain of Unknown Function (DUF) protein families are broadly distributed homologous proteins for which a function has not been attributed (137). These DUFs constitute (as of 2013) more than 20% of all protein domains, with ~1,500 DUFs in Eukaryotes and 2,700 in Prokaryotes (138). The DUF2285 protein family is widely distributed among the Alphaproteobacteria phylum, with half of all DUF2285-domain proteins found in either Rhizobiales or Sphingomonadales. There are 298 Alphaproteobacteria species encode DUF2285-domain within their genome (139). The QseM sequence is entirely comprised by the DUF2285-domain, while the FseA sequence is larger at 266 residues and encodes the DUF2285 domain in the last 70 residues of its C-terminus. Aside from this, proteins with the DUF2285 domain do not appear to be fused to any other domain. Interestingly, two-hybrid assays have demonstrated that QseM binds to the N-terminus of FseA and not the DUF2285 domain of FseA

1.12: Aims of this thesis.

QseM is a DUF2285 protein that can bind and inhibit both FseA and TraR. The mechanism by which QseM binds and inactivates TraR and FseA is not known, nor are there any characterised DUF2285 proteins that can shed light on this.

Given that QseM shares no primary sequence similarity to QslA and TraM, it seems likely that QseM may interact with its partner activator proteins TraR and FseA through a distinct mechanism. Both QslA and TraM were characterised by structure determination in conjunction with protein-interaction assays. We propose that a similar approach taken with QseM will lead to a better understand of how it functions as an antiactivator, and potentially uncover characteristics of the DUF2285 protein family in general.

The aim of this thesis was to biochemically characterise and determine the structure of QseM using X-ray crystallography and nuclear magnetic resonance (NMR).

Chapter 2:

Materials and Methods

2.1 Materials:

Table 2.1: Overview of materials used.

Material	Supplier/manufacture
Microcentrifuge tubes (various size)	SSlbio
Freeloads Pipette tips	Interpath
electro cuvette (1 mm, 90 μ L)	BTX
Centrifuge tubes (various sizes)	Greiner & Nunc
Universal tubes (30 mL)	Techno Plas
Petri dish (94 mm)	Techno Plas
PCR tubes (0.2 mL)	SSlbio
Serological Pipettes (10 and 30 mL)	Nunc
Cryostorage vials (2 mL)	Cryo.S
Syringe Filter (0.22 and 0.44 μ M)	Millex
Syringe (plastic, Luer Slip 10 mL)	Terumo
Plasmid extraction kit	Favorgen
Gel/PCR purification kit	Favorgen
Centrifuge bottles (30 and 250 mL)	Nalgene
Amicon concentrator tubes (various volume/filter size)	Millipore
Vivaspin centrifuge concentrators (various volume/filter size)	GE Healthcare
Mini-Protean TGX gels (4 – 15%)	Bio-Rad
HisTrap HP Ni-NTA column (5 mL)	GE Healthcare
Membrane filters (0.22 and 0.44 μ M)	Pall
Filtration/degasser (500 mL)	Nalgene
SEC 16/600 HiLoad Superdex 200	GE Healthcare
SEC 10/300 GL Superdex 200	GE Healthcare
NMR tubes (5 mm)	Shigemi
Intelli-plate crystallisation tray (3 drop, 96 well)	Hampton Research
Sitting-drop crystallisation tray (24 well)	Hampton Research
Crystal suspension loops (nylon, various sizes)	MiTeGen
Crystallisation screens (Index, Crystal, PEG/ion)	Hampton Research
10-200 kDa protein ladder	NEB

2.2 Chemicals:

Table 2.2: Overview of chemicals used.

Chemical name	Supplier/manufacturer
Potassium phosphate monobasic	Chem-Supply
di-Sodium hydrogen orthophosphate	Chem-Supply
Sodium dihydrogen orthophosphate	Chem-Supply
Sodium chloride	Astral Scientific
Ethylenediaminetetraacetic acid (EDTA)	Perth Scientific
Iron (III) chloride hexahydrate	Perth Scientific
Zinc chloride	Chem-Supply
Copper (II) sulfate pentahydrate	Perth Scientific
Cobalt (II) chloride hexahydrate	Perth Scientific
Boric acid	Perth Scientific
Manganese (II) sulfate monohydrate	Perth Scientific
Calcium chloride dihydrate	Perth Scientific
Magnesium sulfate heptahydrate	Perth Scientific
Thiamine hydrochloride	MP Biomedicals
Biotin	Astral Scientific
Ammonium chloride	Perth Scientific
D-Glucose	Perth Scientific
Tryptone (casein peptone)	Amyl Media
Yeast extract	Sigma-Aldrich
DL-selenomethionine	Chem-Supply
L-Leucine	Astral Scientific
L-Phenylalanine	Astral Scientific
L-Threonine	Astral Scientific
L-Isoleucine	Astral Scientific
L-Valine	Astral Scientific
L-Lysine monohydrochloride	Astral Scientific
Glycerol	Perth Scientific
Tris-HCl	Astral Scientific
Imidazole	Astral
D-Glucose (U- ¹³ C6)	Novachem
Ammonium chloride (15N)	Novachem
Isopropyl β- d-1-thiogalactopyranoside	Astral Scientific
Methanol	Perth Scientific
Glacial acetic acid	Chem-Supply
Coomassie brilliant blue (R-250)*	Chem-supply

2.3 Strains and vectors:

Table 2.3: Strain details.

<i>E. Coli</i> Strains	Description	Source/reference
EPI300	Electro-competent strain, high transformation efficiency and suitable for DNA library storage.	Ramsay Lab, Curtin University
BL21(DE3) (pLysS)	Optimised protein expression strain, pLysS vector to reduce non-IPTG induced expression.	(140)

Table 2.4: Vector details.

Plasmids	Description	Source
pQE-80L	Protein expression (T7/lac induced), ampicillin resistance. T7/lac induced.	Qiagen
pET-11a	Protein expression (T7/lac induced), kanamycin resistance, adds a TEV cleavable 6His tag sequence (upstream) to insert gene.	Novagen
pETM-41	Protein expression (T7/lac induced), kanamycin resistance, adds a TEV cleavable 6His-MBP tag sequence (upstream) to insert gene.	Novagen
pQe80_FseA	Protein expression (T7/lac induced), ampicillin resistance. T7/lac induced. Expresses 6H-FseA	Ramsay Lab, Curtin University
pQe80_QseM	Protein expression (T7/lac induced), ampicillin resistance. T7/lac induced. Expresses 6H-QseM	Ramsay Lab, Curtin University

Table 2.5: Primer details.

Primer name	Sequence	Use
pQe80_fwd	TCTTCACCTCGAGAAATCAT	Sequencing pQe80 constructs
pQe80_rev	TGAGGTCATTACTGGATCTATCA	Sequencing pQe80 constructs
pETM11_fwd	TGATGTCGGCGATATAGGCGC	Sequencing all pET constructs

pETM11_rev	GTTATGCTAGTTATTGCTCAGCGG	Sequencing all pET constructs
FseAco_NcoI_fwd	ATATCCATGGCTATGATCGGTAACGAC	PCR amplify FseA, add RE site
FseAco_BamHI_rev	TATAGGATCCTTAAACCAGGAAGTCACG	PCR amplify FseA, add RE site

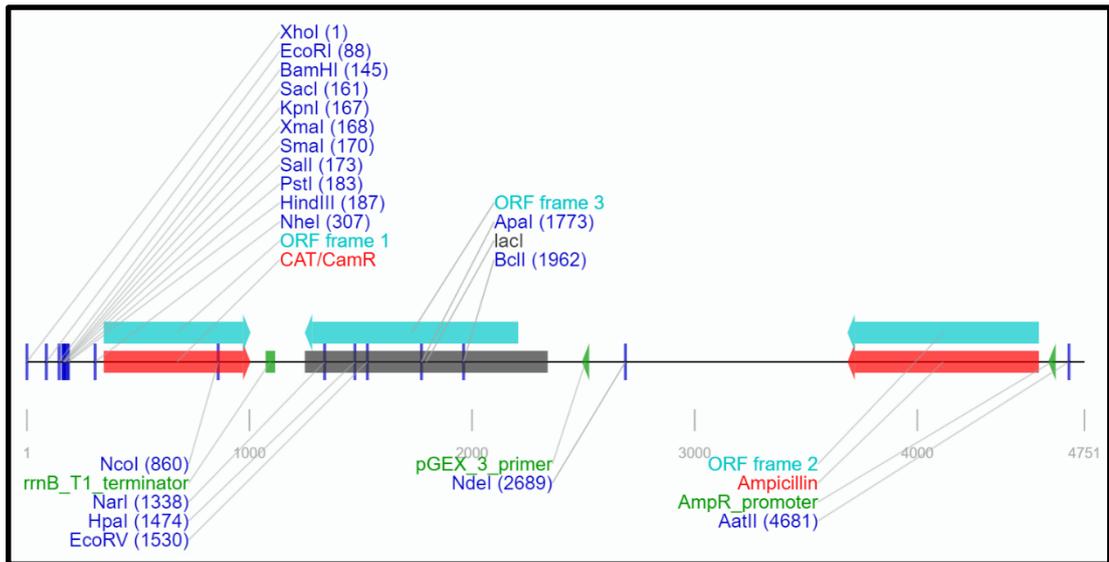


Figure 2.1: Vector map of pQe80L (addgene sequence analyser).

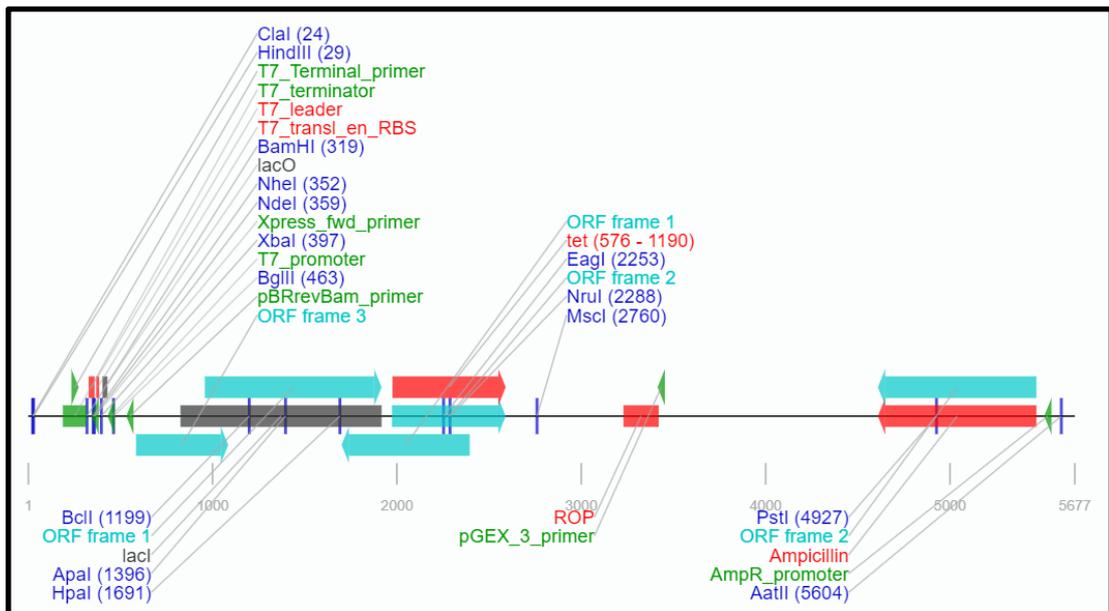


Figure 2.2: Vector map of pET11a (addgene sequence analyser).

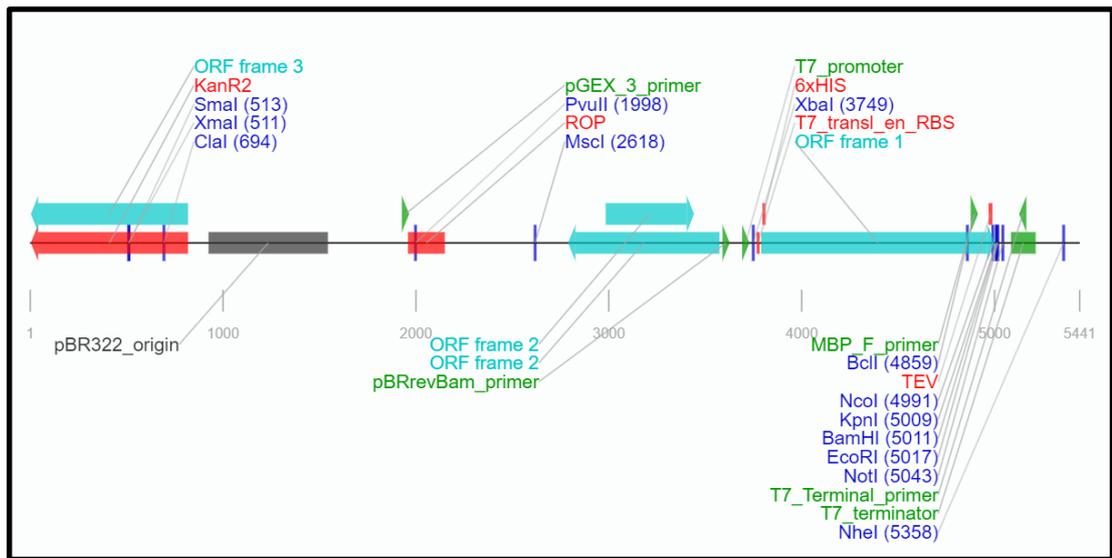


Figure 2.3: Vector map of pETM41 (addgene sequence analyser).

2.4 Buffers and solutions

Note, all solutions are made up in ultra-pure water (filtered through an ultrapure system, Ibis Technology) unless specified otherwise.

Table 2.6: Medium composition.

Lysogeny Broth (rich) media	
• Tryptone	10 g/L
• NaCl	10 g/L
• Yeast Extract	5 g/L
M9 (minimal) media	
• M9 salt stock solution (10X)	100 mL/1 L media
• <i>Na₂HPO₄</i>	60.0 g
• <i>KH₂PO₄</i>	30.0 g
• <i>NaCl</i>	5.0 g
• Trace element solution (100X)	10 mL/1 L media
• <i>EDTA</i>	5.000g
• <i>FeCl₃.6H₂O</i>	0.830 g
• <i>ZnCl₂</i>	0.084 g
• <i>CuCl₂.5H₂O</i>	0.015 g
• <i>CoCl₂.6H₂O</i>	0.010 g
• <i>H₃BO₃ (boric acid)</i>	0.010 g
• <i>MnSO₄.H₂O</i>	0.002 g
• 0.34 M CaCl₂ (dissolved in ddH₂O)	1 mL/ L
• 1 M MgSO₄ (dissolved in ddH₂O)	1 mL/ L
• 1 mg/mL Thiamin (dissolved in ddH₂O)	1 mL/ L
• 1 mg/mL Biotin (dissolved in ddH₂O)	1 mL/ L
• NH₄Cl*	0.5 g/L
• 20% glucose (dissolved in ddH₂O)*	20.0 mL/L
• Selenomethionine mix (dissolved in ddH₂O with addition on HCl)#	50 mg/per amino acid/1 L medium (L-selenomethionine, L-isoleucine, L-leucine, L-valine) 100mg/per amino acid/ 1 L medium (L-phenylalanine, L-lysine, L-threonine)

*For isotope incorporation, ¹⁵NH₄Cl and/or L-¹³C-glucose were used

#For selenium incorporation. Added to expression medium after inoculation of broth with culture in early log-phase but before addition of IPTG

Table 2.7: Antibiotic selection for strains and vectors.

Strain/vector	Antibiotic concentration
EPI300	No strain selection
BL21(DE3) (pLysS)	Chloramphenicol: 100 µg/mL
pQE-80L	Ampicillin: 100 µg/mL
pET-11a	Ampicillin: 100 µg/mL
pETM-41	Kanamycin: 50 µg/mL

Table 2.8: Nickel-purification buffer composition. Final buffers were pH adjusted to 7.5 using NaOH/HCl.

Component	Buffer A	Buffer B
NaCl	300 mM	300 mM
Na ₂ HPO ₄ /NaH ₂ PO ₄	100 mM	100 mM
Glycerol	5-10 %	5-10 %
Imidazole (C ₃ N ₂ H ₄)	100 mM	800 mM

Table 2.9: SEC purification/crystallisation buffer. Final buffer was pH adjusted to 7.5 using NaOH/HCl.

Component	Concentration
NaCl	100 mM
Tris-HCl	50 mM
Glycerol	0-2 %

Table 2.10: Minimal salt buffer. Final pH was adjusted to 7.5 using NaOH/HCl.

Component	Concentration
NaCl	20 mM
NaH ₂ PO ₄	10 mM

Table 2.11: Protein extinction coefficients.

Protein	Molecular mass (kDa)	Ext. coefficient ($M^{-1} cm^{-1}$)	Residue number	pI
QseM	9.780	15470	83	4.87
6H-QseM	10.960	15470	91	5.85
6H-TEV-QseM	12.704	18450	108	5.50
6H-MBP-TEV-FseA	75.219	121810	667	7.06

Values calculated with ProtParam (141). (<https://web.expasy.org/protparam/>)

Table 2.12: Buffers for protein gel electrophoresis.

Tris-glycine 10× stock (1 L)	
Component	Concentration
Glycine	144.0 g
Tris	30.4 g
Gel running buffer (1 L)	
Component	Concentration
Tris-glycine (10× stock)	100 mL
SDS (10% in ddH ₂ O)	10 mL
ddH ₂ O	890 mL

Table 2.13: Protein loading dye (4 ×).

Component	Concentration
Tris-HCl	50 mM
Sodium dodecyl sulphate	2.0 %
Bromophenol blue	0.1 %
Glycerol	30 % v/v
Dithiothreitol (5 M in water)	20 μ L/1 mL

Table 2.14: SDS-PAGE Coomassie blue stain and de-stain.

Components	Concentration
Methanol	50%
Glacial acetic acid	10%
Coomassie brilliant blue (R-250) *	0.1%

* Only present in stain, de-stain was dye-free.

2.5 Methods:

2.5.1 Plasmid construction

DNA to be cloned was created either by PCR amplification from previously constructed vectors (Table 2.4, pQe80_FseA and pQe80_QseM) or synthetically as gene blocks (IDT). PCR amplified products and expression vectors (Table 2.4, pQe80L, pET-11a and pETM-41) were restriction-digested and cleaned using a Gel/PCR purification kit (Favorgen) prior to ligation. After ligation and clean-up of product using a Plasmid extraction kit (Favorgen), vectors were transformed into *E. coli* EPI300 strain (Table 2.3) by electroporation. Electroporation was achieved by mixing ~50 µg/mL post clean-up vectors with a thawed but still cold 40 µL aliquot of bacteria (frozen mid-log phase). The mixture was inserted into a pre-chilled (on ice) electrocuvette (BTX) before being loaded into a MicroPulser Electroporator (Bio-Rad) and exposed to a brief burst of 1.8 kV at 9 A current. Immediately after electroporation, cell-DNA mixture was suspended in 1 mL LB (Lysogeny Broth, refer to Table 2.6) media, pre-warmed to 37 °C, and allowed to recover for 30 - 60 minutes at 37 °C before being used to inoculate a LB-medium plate containing antibiotics selecting for both the strain and vector. Successful clones were identified by agarose gel electrophoresis of restriction-enzyme digests and Sanger sequencing (AGRF) using appropriate sequencing primers (refer to Table 2.5).

2.5.2 Expression of QseM and FseA

Expression plasmids containing N-terminally fused hexahistidine (6H) tagged QseM and FseA were previously constructed (Dr. Josh Ramsay). A second expression plasmid was constructed (as per 2.5.1) to express 6H-QseM with a tobacco etch virus (TEV) protease site between the hexahistidine tag and the protein, referred to as 6H-TEV-QseM. Additionally, a plasmid was constructed to express 6H-TEV-MBP-FseA. These vectors were electroporated into BL21(DE3) pLysS and resulting strains were used to inoculate a 5 mL LB starter culture which was incubated at 37 °C overnight under antibiotic selection for the strain and the expression vector (refer to Table 2.7). Once the starter culture reached mid-log phase ($OD_{600} \sim 0.2$), the

entire culture was used to inoculate 1 L LB medium containing the same antibiotics and incubated with shaking (180 rpm) at 37° C. All expression vectors used were isopropyl β -D-1-thiogalactopyranoside (IPTG) inducible (142) and were induced to express protein with 0.2 mM IPTG once the culture reached log phase (\sim OD₆₀₀ 0.4). The culture was further incubated with shaking for 12-18 h after which cells were collected by centrifugation at 10,000 \times g (Beckman Coulter Avanti J-E) in 250 mL centrifuge bottles (Nalgene), transferred to 50 mL tubes (Greiner) and stored at -80 °C or used immediately for purification.

For incorporation of L-selenomethionine, and carbon and nitrogen isotope labelling, strains were cultured in M9 minimal medium (Table 2.6, M9 (minimal) medium). For SeMet incorporation, 50 mg/L SeMet was added. Additionally, 100 mg/L phenylalanine, lysine and threonine and 50 mg/L leucine, isoleucine and valine were added to inhibit methionine biosynthesis plus (143). For NMR, cells were grown in M9 medium containing ¹³C-containing D-glucose (Novachem) and/or ¹⁵N-containing ammonium chloride (Novachem) as sole carbon or nitrogen sources, respectively.

2.5.3 Cell lysis

Cell pellets were suspended in a PBS buffer (buffer A, Figure 2.8) and mechanically lysed using a constant flow system (Constant Systems, CF1) maintained at 5 °C via refrigerated water jacket. The lysed solution was centrifuged (Beckman Coulter Avanti J-E) at 24,000 \times g for 45 minutes at 4 °C in 50ml centrifuge bottles (Nalgene) to separate insoluble components. The soluble fraction was filtered through a 0.22 μ m syringe filter (Millex) using a sterile syringe (Terumo). Filtered soluble fractions were used immediately for purification.

2.5.4 Purification of QseM

6H-QseM protein was purified from lysed *E. coli* suspension in a two or three step process depending on the construct used. Throughout the entire process, where possible, temperature of the protein was kept at or below 4 °C. The

soluble fraction was loaded immediately into a 5 mL Ni-NTA column (GE Healthcare) at a flow rate of 1 – 2 mL/min using a peristaltic pump (EP-1 Econo Pump, Bio-Rad) in binding buffer (Buffer A, Table 2.8). Purification was carried out using fast protein liquid chromatography (FPLC) (GE Äkta Purifier, or Bio-Rad DuoFlow) and protein elution was monitored by measuring absorbance at 280 nm using an in-line UV probe. The Ni-NTA column containing bound 6H-QseM was washed with Buffer A and then Buffer B (buffer A + 800 mM imidazole, Table 2.8) was added with a linear gradient increase of 32mM imidazole/ml. All proteins were eluted from the Ni-NTA column with less than 800 mM imidazole. Size exclusion chromatography (SEC) (Superdex 200 16/600, GE Healthcare) was then used to purify 6H-QseM and replace purification buffer with crystallisation buffer (Table 2.9) or an NMR-suitable minimal-salt buffer (Table 2.10). The Superdex 200 16/600 and Superdex 200 10/300 GL (GE Healthcare) columns for SEC had void volumes of ~45 ml and ~7.5 ml respectively.

For purification of 6H-TEV-QseM an additional step was added after the nickel-affinity step. Nickel-purified 6H-TEV-QseM was mixed in a 10:1 molar concentration ratio with recombinant 6H-tagged TEV protease and incubated for 12-16 h at 4 °C with 1mM dithiothreitol (DTT). The mixture was then loaded into a 5 mL HisTrap nickel column. The flow-through containing cleaved QseM protein was collected. The cleaved QseM protein sample was then purified using SEC as above.

2.5.5 Purification of FseA

6H-MBP-FseA was purified in a similar method to 6H-QseM, however the buffers used during the Ni-NTA and SEC purification process were adjusted to pH 6.5. A dedicated Ni-NTA column was used exclusively for 6H-MBP-FseA to avoid cross-contamination between protein samples. 6H-MBP-tagged FseA was used immediately after purification due to rapid aggregation.

2.5.6 Protein concentration

Purified protein was concentrated using centrifugal filtration tubes (GE Healthcare, Millipore) with a mass cut-off of 10 kDa. Concentration of protein samples was done in centrifugation increments of no more than 10 minutes at a time at 2,000 × g at 4 °C. Between centrifugations, the sample was gently mixed by pipette and absorbance at 280 nm was measured (absorbance on Thermo Scientific NanoDrop 1000) and the concentration was calculated using ProtParam (141) (ExPasy) with the appropriate extinction coefficient (detailed in Table 2.11). Protein loss during concentration was monitored by measuring protein absorbance of the column flow-through. Once the desired concentration was reached, the sample was transferred to a new tube (SSIBio) and used immediately for experimentation, stored in short term on ice or at 4 °C, or frozen in liquid nitrogen and kept at -80 °C for long term storage.

2.5.7 Analysis of protein sample

Visible clouding of the protein suspension was used as an indicator that protein became insoluble in that solution at that concentration or temperature. After centrifugation to pellet any insoluble protein that was visible or not, a significant drop in a 280 nm absorbance signal in the soluble fraction (Thermo Scientific NanoDrop 1000) was also used as an indicator that protein precipitation and/or aggregation was occurring.

Denaturing SDS-PAGE analysis was used throughout the purification process to confirm presence and molecular weight of the protein. For SDS-PAGE analysis, protein was mixed with a loading dye (Table 2.13), heated to 95 °C for 10 minutes then loaded into a pre-cast SDS-PAGE gel (4 – 15% Mini-Protean TGX gel, Bio-Rad) mounted in a protein electrophoresis vertical chamber (Mini-Protean, Bio-Rad) submerged in gel running buffer (Table 2.12). Gels were run using a 250 W 1 kV 500 mA DC power supply (PowerPac 1000, Bio-Rad) for 30 minutes at 90 V followed by a further 60 minutes at 120 V (or until loading dye band reached < 5 mm from end of gel). Gels were stained for a minimum of 2 h, followed by a de-staining in water for 12-16 h (stain and de-stain detailed in Table 2.14). Gels were imaged using Gel Doc

(Bio-Rad). Identity of proteins in excised gel bands was confirmed using mass spectrometry analysis (Proteomics international).

2.6 Crystallisation

2.6.1 Crystallisation screening

Initial crystallisation screening was carried out using a variety of commercial sparse-matrix screens (Index, Crystal 1 and 2, Natrix, and PEG/ion screens, Hampton Research), which tested variations in precipitant, pH, salt, and other additives. Screens were set up using a Phoenix robot (Art Robbins Scientific) fitted with a 96-syringe head and single nano dispenser, in sitting-drop vapour diffusion format in 96-well Intelli-plate 96-3 crystallisation trays (Hampton Research). Each 96-well plate contained 96 unique crystallisation solutions with each well containing 3 individual drops of protein and crystallisation solution undergoing vapour diffusion. The 3 drops in each well had different protein:crystallisation solution ratios (1:2, 1:1, and 2:1). The total drop volumes were either 0.3 μL (for 1:2 and 2:1 ratios) or 0.4 μL (for 1:1 ratio) and the diffusion reservoir volume was 70.0 μL . A layer of HD Clear (Duck) tape was used to seal the tray immediately after all drops and reservoirs were filled in the tray. All trays were assembled and stored in a temperature-controlled room at a constant 20 °C where vapour diffusion took place.

Crystallisation conditions that produced initial crystals were optimised using larger drop sizes in 24-well sitting-drop Cryschem plates (Hampton Research), either with or without the addition of a crystal seed stock. Drop volumes were either 4 μL (2 μL of protein and solution each), 5 μL (2 μL of protein and solution each and 1 μL of crystal seed stock), 8 μL (4 μL of protein and solution each) or 9 μL (4 μL of protein and solution each and 1 μL crystal seed stock). All had a reservoir volume of 450 μL . Wells were sealed using HD Clear (Duck) tape and stored in a temperature-controlled room at a constant 20 °C.

2.6.2 Observation of crystallisation

Crystal trays were typically monitored using a BD-60ST stereo microscope (Boshida Optical) at 1, 2, 4 and 7 days (then once every additional 7 days) from setup. Some trays were selected for automated monitoring using a Minstrel HT crystal imaging system with an attached plate gallery storage unit (Rigaku) with images taken of each drop at the following times (0, 1, 2, 4, 7 days, and every subsequent 7 days). Crystals observed using visible light were then selected for UV fluorescence measurements.

2.6.3 Creating a crystal micro-seeding stock

A micro-seeding stock was created by using previously formed crystals. Crystals were collected from the crystallisation drop of a crystal tray and transferred using a crystal loop (MiTeGen) to a 50 μ L aliquot of mother liquor. Fine fragmentation of the crystal was achieved using a Seed Bead Kit (Hampton Research, PTFE seed type). The resulting stock was stored at 4 °C and used to make diluted 'seeder' solution, typically at a 1:100 concentration. Seeder solutions were added to the crystallisation drops directly after the addition to the protein solution and the crystallisation buffer. Typically, a 0.1 μ L amount would be added to the 96 well screening trays or 1 μ L would be added to larger volume 24-well optimisation trays.

2.6.4 Treating crystals with cryoprotectant solution

Cryoprotectants were screened for suitability by suspending crystals in each cryoprotectant and observing immediate effects on the crystals, freezing in liquid nitrogen and observing the freezing and diffraction quality during diffraction data collection. Cryoprotectants were selected based on similarity with compounds that were within the crystallisation solution to avoid significant environmental changes to the crystals during harvesting and cooling. Crystals were manipulated using Cryo Loops (nylon loops attached to a pin fixed to a magnetic base suitable for mounting on a goniometer, Hampton Research). Crystals were individually collected and transferred into a drop of the tested cryoprotectant (e.g. glycerol, PEG, glucose or EG mixed with the original

crystallisation buffer), placed on a glass cover slip to soak for typically 1-2 minutes (maximum 5 minutes). During soaking, the suspended crystals were observed using a BD-60ST stereo microscope (Boshida Optical). Crystals were then removed from the cryoprotectant unless visible defects in the crystals or significant evaporation was observed. After soaking, crystals were cooled and stored in liquid nitrogen. Concentrations of cryoprotectants were optimised over several iterations of crystal-cooling, looking for improved diffraction quality across vary concentrations (typically in increments of 5% cryoprotectant concentration). Mounted crystals were placed in a transfer puck (UniPuck, Crystal Positioning Systems) and transported to the Australian Synchrotron in a liquid nitrogen dewar (Taylor Wharton).

2.7 Crystallographic data collection and processing

2.7.1 X-ray diffraction data collection

Diffraction data were collected from crystals using the Australian Synchrotron MX1 and MX2 beamlines using the Blu-Ice Software (144) under remote or on-site user access. Data were automatically processed using XDS with *POINTLESS* (145) and *AIMLESS* (146) used to determine space group and to scale and merge data. Dataset quality and resolution limit was evaluated by the signal:noise ratio ($I/\sigma(I) \geq 1.5$) and half-dataset correlation coefficient ($CC_{1/2} \geq 0.3$). Diffraction data were routinely reprocessed in-house using iMosflm (147) or XDS (148). Following integration, *POINTLESS* and *AIMLESS* were used to determine probable point/space groups and to merge and scale data. Further uses of *POINTLESS* and *AIMLESS* were used within the *CCP4 suite*(149).

Collection of anomalous diffraction data required tuning the X-ray wavelength to wavelengths close to the X-ray absorption edge of selenium, $\sim 0.9795 \text{ \AA}$ ($\sim 12658 \text{ eV}$). 'Peak' and 'inflection point' wavelengths of selenium were initially determined as 0.9793 \AA ($12,660 \text{ eV}$) and 0.9809 \AA ($12,640 \text{ eV}$) respectively. These peak and inflection point wavelengths were based on theoretical values (http://skuld.bmsc.washington.edu/scatter/AS_index.html (150)). MAD (multi anomalous diffraction) fluorescence scans carried out at the Australian

Synchrotron on the MX1 beamline determined a selenium peak wavelength of 0.9636 Å (12,867 eV) and inflection point wavelength of 0.9772 Å (12,688 eV) from 6H-QseM with L-selenomethionine. Remote datasets were collected at ~500 above or below peak for high (0.9393 Å, 13,200) or low (1.0163 Å, 12,200).

2.7.2 Anomalous phasing

CCP4 suite-hosted *SHELX C/D/E* (151) and *PHASER* (152) (153) software pipelines were used to calculate phase information either by SAD (single-wavelength anomalous dispersion) using peak wavelength datasets or MAD (multi-wavelength multiple anomalous dispersion) using peak, inflection and remote wavelength datasets. *SHELX C/D/E* was run using default settings with a phase solvent content of 0.5 and set to search for 6 unique selenium (Se) atoms. *PHASER* was run using default settings and set to search for 6 unique Se atoms with asymmetric unit components defined as a single 6H-QseM molecule (specified by input amino-acid sequence file). If the space group was chiral, both enantiomorphs were tested. The asymmetric unit composition was determined by *MATTHEWS* (149) using the pre-defined space group and unit cell dimensions derived from *POINTLESS/AIMLESS*. The crystallographic reflection files were MTZ formatted files containing either as a single scaled and merged peak SeMet dataset for SAD, or as a merged (using the *CCP4 suite*, *CAD* (149)) set containing native, peak, inflection point and/or remote SeMet datasets for MAD. These mtz files were the inputs for both *SHELX C/D/E* and *PHASER* when attempting to solve the phase problem.

2.7.3 Molecular replacement phasing

Various protein prediction models using *ROSETTA*, *ROBETTA* (154, 155) and *CS-ROSETTA* (156-159) and NMR-determined models of 6H-QseM were used by *PHASER-MR* (160) to attempt to successfully solve a crystallographic structure of 6H-QseM by molecular replacement to circumvent the necessity to solve the phase using anomalous phasing. How these models were generated are detailed in Section 2.9.4. All resultant structures were PDB

formatted files. The structure PDB and crystallographic MTZ files were inputted to *PHASER-MR*.

2.8 NMR data collection, atom assignment and structure calculation:

2.8.1 Protein sample preparation and handling for NMR

Expression and purification of ^{15}N - ^{13}C -labelled 6H-QseM for NMR outlined in 2.5, however, protein was transferred into a minimal salt (Table 2.10) buffer during SEC purification. For protein concentration the procedure detailed in 2.3.5. was followed. In most instances, protein solution was maintained in a liquid state at $\sim 4\text{ }^{\circ}\text{C}$ or on ice when used within a 24-hour period. For longer-term storage batches of 500 μL at $< 2\text{ mg/mL}$ concentration were frozen using liquid nitrogen and stored at $-80\text{ }^{\circ}\text{C}$.

6H-QseM was freeze-dried prior to transportation or when the water was substituted with D_2O . Samples with a concentration of 1-2 mg/mL were aliquoted into 200-300 μL volumes and frozen in liquid nitrogen. Frozen protein was stored in a vacuum container mounted on a lyophilizer (Martin Christ, Alpha 3-4 LSCbasic) at room temperature for $\sim 12\text{ h}$. The lyophilised sample was then resuspended in D_2O using the same volume of water as the original solution to maintain consistent protein and buffer concentration. Each sample was spiked with 1% molar ratio 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) as a chemical shift standard prior to NMR experiments. DSS was used as the reference throughout all 6H-QseM NMR experiments to correctly align all spectra. All NMR data were processed using *TopSpin* (Bruker) and analysed using the *CCPN analysis* (Bruker) software. NMR data were collected on either a Bruker AVANCE III 600 MHz spectrometer with a room temperature TXI probe (Centre for Microscopy, Characterisation & Analysis [CMCA], University of Western Australia) or a Bruker AVANCE III 800 MHz spectrometer with a cryogenic TCI probe (School of Life and Environmental Sciences, University of Sydney).

2.8.2 Initial one-dimensional spectra analysis

Observations of one-dimensional (1D) spectra were used to determine buffer suitability and some secondary structure information as well as overall protein stability. Ideally, the buffer would not produce any signal except for a heavily suppressed water peak and the DSS reference peak. The relative concentration of amide proton signals either downfield (α -helical, 8.0-9.0 ppm) or upfield (β -sheet, 4.8-5.9 ppm) gave a general insight into secondary structure composition of the protein (161). Range and resolution of chemical shift peaks were used to determine protein order, with disorder resulting in reduced number of unique chemical shift signals. These 1D spectra were also collected after most experiments to ensure that substantial resonance changes were not observed between 2D and 3D experiments run on one sample or between all 6H-QseM samples prepared for NMR data collection.

2.8.3 Single and double labelled QseM NMR spectra collection

Incorporation of ^{15}N and/or ^{13}C into 6H-QseM was required for most 2D/3D NMR experiments (Table 2.15 and Table 2.16). Expression and purification are detailed in 2.5. ^{15}N -HSQC experiments were used to measure possible resonance shifts during data collection.

Table 2.15: List of NMR experiments run for structure determination (required isotope-labelled protein).

2D experiments	3D experiments
^{15}N -HSQC	HNCACB*
^{13}C -HSQC	CBCA(CO)NH*
^{13}C -HSQC (aromatic)	HNCO
^1H -TOCSY	HBHA(CO)NH
^1H -COSY	HCCH-TOCSY
^1H -NOESY	HCCH-COSY
^{15}N -HSQC (histidine)	^{13}C -NOESY
HBCBCGCDHD	^{13}C -NOESY (aromatic)
HBCBCGCDCEHE	^{15}N -NOESY

*Only HNCACB and CBCA(CO)NH were collected with non-uniform sampling, all others were collected using uniform sampling.

Table 2.16: Experimental collection details.

Experiments	Spectral Width			Data points			Scans per T1 increment	Ref.
	F1	F2	F3	T1	T2	T3		
¹⁵ N-HSQC	5.6	28.0	-	2048	512	-	1	(162-164)
¹⁵ N-HSQC (<i>histidine</i>)	5.7	140. 1	-	2048	256	-	1	(162, 165)
¹³ C-HSQC	7.1	70.0	-	2048	512	-	1	(162)
¹³ C-HSQC (<i>aromatic</i>)	13. 9	40.0	-	2048	256	-	1	(162)
¹ H-TOCSY	13. 9	13.9	-	2048	512	-	1	(166)
¹ H-COSY	13. 9	13.9	-	2048	512	-	1	(166)
¹ H-NOESY	13. 9	13.9	-	2048	512	-	1	(166)
HBCBCGCDHD	12. 0	25.0	-	2048	512	-	1	(167)
HBCBCGCDCE HE	12. 0	25.0	-	2048	512	-	1	(167)
HNCACB	6.0	28.0	70.0	512	256	256	16	(168)
CBCA(CO)NH	6.0	28.0	70.0	512	256	256	16	(168)
HNCO	6.0	28.0	70.0	512	256	256	16	(168, 169)
HBHA(CO)NH	6.0	28.0	7.0	512	256	256	16	(170)
HCCH-TOCSY	12. 0	70.0	12.0	1024	256	256	32	(171)
HCCH-COSY	12. 0	70.0	12.0	1024	256	256	32	(171)
¹³ C-NOESY	7.5	38.0	7.5	2048	256	512	5	(172)
¹³ C-NOESY (<i>aromatic</i>)	6.0	30.0	12.0	2048	128	256	4	(172)
¹⁵ N-NOESY	5.4	25.0	12.0	2048	128	256	29	(173, 174)

2.8.4 Atom assignment

Protein backbone atom assignment was completed prior to sidechain atom assignment. After indexing all possible ^{15}N -HSQC peaks, the 3D HNCACB/CACB(CO)NH spectra were used to find the $\text{C}\alpha$ and $\text{C}\beta$ chemical shifts of preceding residues based on methods outlined in <https://www.protein-nmr.org.uk/solution-nmr/assignment-theory/triple-resonance-backbone-assignment/>. Polypeptide segments were identified using this method enabled confident assignment of residue identity and associated H-N peaks found in ^{15}N -HSQC spectra. The HNCO spectrum was also used to find the corresponding C-O backbone resonances linked to assigned N-H peaks. Attempts to assign the sidechains began once there was significant completion of the ^{15}N -HSQC (>60% of all possible residues of the protein sequence, excluding those not found in the HSQC). At this stage only glycine residues had complete atom assignment.

Backbone assignment gave resonance information relating to N-H, C-O, $\text{C}\alpha$ and $\text{C}\beta$ where possible (proline has no N-H, glycine has no $\text{C}\beta$). Following on from this, the HBHA(CO)NH could be used to find the associated $\text{C}\alpha/\text{C}\beta$ -attached protons to enable potential complete atom assignment of all alanine residues. HCCH TOCSY/COSY were used for atom assignment beyond the C/ $\text{H}\beta$. This facilitated potential full atom assignment of cysteine (excluding S-H group), serine and valine (excluding O-H group), aspartate and glutamate, isoleucine, leucine, lysine, proline, arginine and valine residues.

Through-space coupling methods ^1H -NOESY/TOCSY/COSY, ^{15}N -NOESY, ^{13}C -NOESY and aromatic ^{13}C -NOESY (both requiring an assigned ^{13}C -HSQC) were used to assign more difficult atoms. Assignments were guided using like $\text{H}\beta$ - $\text{H}\delta/\text{H}\epsilon$ correlation experiments HBCBCGCDHD/HBCBCGCDCEHE and histidine-focused HSQC experiments. This enabled the potential assignment of the remaining atoms of the phenylalanine, histidine, methionine, glutamine, tryptophan and tyrosine residues.

2.8.5 Dihedral angle calculations

Dihedral angles were calculated using assigned chemical shift information either using *DANGLE* (175) or through the TALOS+ (176) server packages within *CCPN Analysis* (<https://spin.niddk.nih.gov/bax/nmrserver/talos/>). Dihedral information and secondary structure were predicted using PSIPRED (177, 178). TALOS+ and PSIPRED generated a list of dihedral angles where backbone assignments were completed and predicted dihedral information in regions where backbone assignments were missing. The TALOS+ dihedral output was used as a dihedral constraint list in further structure calculations.

2.8.6 Structure calculations

A preliminary structure calculation of 6H-QseM was made using the calculated dihedral angles in the *CS-ROSETTA* sever. The output provided initial 6H-QseM models. *CYANA* (179, 180) was used to run structure calculations. *CYANA* uses through-space NOEs to create distance restraints between pairs of protons. *CYANA* structure calculations were also supplemented with dihedral restraints. For *CYANA* calculations, the peaks, chemical shifts, dihedral and sequence information were exported by using the 'format converter utility' within the *CCPN Analysis* software. The peak lists were exported in *CYANA* format and the sequence and chemical shifts were exported in Xeasy format. The dihedral restraints were exported in TALOS format.

2.9 Additional protein characterisation methods.

2.9.1 Small angle X-ray scattering (SAXS)

Solution SAXS is a low-resolution structure analysis method using X-ray scattering, yielding information on the size and shape of monodisperse samples (181). Protein examined by SAXS was expressed, purified and concentrated as outlined in 2.5. SAXS data was collected at the Australian Synchrotron SAXS/WAXS beamline (182, 183). Data was collected using both static capillary and SEC methods (HPLC system, Shimadzu) with assistance from Dr. Jason Schmidberger and Dr. Karina Yui Eto. Scattering was collected

on a 1 M Pilatus detector. Background buffer scattering subtraction was automatically completed for all static samples, giving a subtracted DAT formatted file list readable by the *PRIMUS* (184) software. SEC-SAXS data was buffer-subtracted using *CHROMIXS* (185). Quality of data, distance distribution and calculated molecular weights was determined within *PRIMUS/qt* and *GNOM* (*ATSAS* suite (186)). A low-resolution model was calculated using *DAMMIN* (187). 6H-QseM SAXS scattering profile and 6H-QseM models (predicted or NMR calculated) were compared using *CRY SOL* (188).

2.9.2 Analytical size exclusion chromatography

Analytical size exclusion chromatography (SEC) using a GE Superdex 200 10/300 GL column (GE Healthcare) was used to assess if 6H-QseM and 6H-MBP-FseA formed a stable complex in solution. The column was mounted on an ÄKTA purifier system maintained at a 4 °C. All proteins used were purified as detailed in Section 2.5 prior to analytical SEC experiments. Two-hundred to five-hundred microlitres of protein sample was used, in cases involving a mixture of two proteins, samples were mixed in a 1:1 molar ratio. Protein was detected by UV absorbance at 280 nm, absorbances at 260 nm were also monitored to detect presence of nucleic acid. A 260:280 nm ratio < 1.0 in a single UV peak was interpreted as a protein sample free of significant nucleic acid contaminants. Peaks in the 280 nm UV absorbance were collected in separate fractions of 500 µL and analysed using SDS-PAGE as detailed in Section 2.5.7.

2.9.3 Structural alignment of QseM with other proteins

The *DALI* server was used for detection of structurally similar proteins to the NMR 6H-QseM model (<http://ekhidna2.biocenter.helsinki.fi/dali/>) (189). A global ('Heuristic PDB') search for comparable models in the RCSB PDB used with an input model truncated to exclude regions of disorder. The *COFACTOR* alignment algorithm (190, 191) was also used with the same truncated input model used for *DALI* to also search for similar models using default settings.

2.9.4 *de novo* protein structure prediction

Structure predictions using the full 6H-QseM sequence were generated using the *I-TASSER* (<https://zhanglab.ccmb.med.umich.edu/I-TASSER/>) (192, 193). Default settings were used with no additional restraints, structure templates or secondary structure information. *ROSETTA*-derived protein structure prediction was generated using default settings with an input of the 6H-QseM amino acid sequence. *ROBETTA*-derived protein structure prediction was generated using default settings with an input of the 6H-QseM amino acid sequence and a predicted distance restraint list generated from *GREMLIN2* (194-196) to assist in tertiary structure accuracy. This distance restraints were derived from a sequence alignment of proteins within the DUF2285 family, which QseM belongs to. The *CS-ROSETTA*-derived protein structure prediction was generated using default settings with an input of the 6H-QseM amino acid sequence and an NMR-derived torsion-angle list (detailed in chapter 5) to define secondary structure.

Chapter 3:

Recombinant protein expression and purification

3.1 Introduction:

Recombinant protein expression and purification is necessary for protein crystallography and nuclear magnetic resonance experiments (NMR). Protein expression is often carried out in a bacterial expression strain transformed with an inducible expression vector encoding the desired protein. Over-expression using *Escherichia coli* strains is commonplace for bacterial proteins that require no species-specific post-translational modifications. As an example, the popular *E. coli* BL21(DE3) pLysS strain can be used (197), as it is suitable for the expression of proteins from pET family vectors which use the T7 induction system and is selectively induced using Isopropyl β - d-1-thiogalactopyranoside (IPTG). Protein expression can be adjusted by modifying the medium the culture is grown in. For NMR, incorporation of nitrogen (^{15}N) and/or carbon (^{13}C) isotopes into the protein is often required and is achieved by expression in minimal medium (198) with ^{15}N or ^{13}C enriched nitrogen and carbon sources. Similarly, protein expressed with atoms that have X-ray absorption edges like selenium for anomalous diffraction collection (detailed in chapter 4) can be incorporated by expressing in L-selenomethionine enriched media.

Purification of protein post expression requires separation of the desired protein from all other cellular material and often transferred to a buffered solvent suitable for *in vitro* experimental methods like crystallography or NMR. Nickel-affinity purification of hexahistidine-tagged (6H) proteins is a common first step in recombinant protein purification workflows. Fusions with other tags like maltose-binding protein (MBP) or glutathione-S-transfer (GST) can improve stability or solubility of the desired protein product. MBP and GST also offer alternative affinity purification steps. MBP is commonly used to tackle expression, solubility and stability difficulties in various protein fusions ranging from 20 to 599 amino acids (199-201). Appropriately designed tags can also be removed by targeted protease cleavage after their function is no longer needed. Enzymes like the tobacco etch virus (TEV) are often used to remove N or C terminal fusions after affinity protein purification (202). Size exclusion chromatography (SEC) is an orthogonal secondary purification, based on molecular size and shape, that yields a high purity protein sample in a desired

buffered solvent. The chromatogram from SEC can also provide initial size and oligomeric state information on the protein in solution. SEC can also identify protein-protein complexes in heterogeneous protein mixtures. Analytical SEC is more precise in determining oligomeric state and size of a protein sample.

Prior work demonstrated QseM can be expressed and purified (203). Expression and purification methods were optimised for crystallographic and NMR experiments. It is also noteworthy that the 6H N-terminal tag fused to QseM does not inhibit its ability to bind TraR or FseA *in vivo* (Will Jowsey and Calum Morris, unpublished).

3.2 Results and Discussion:

3.2.1 Purification of QseM and FseA

QseM was expressed either as a fixed-tagged (6H-QseM) or cleavable 6H-tagged (6H-TEV-QseM) and L-selenomethionine incorporated (Se-6H-QseM) for crystallographic experiments detailed in Chapter 4. Preliminary work to express and purify 6H-QseM with and without L-selenomethionine was conducted prior to starting this thesis (Drew Hall honours dissertation, unpublished). QseM was also expressed with carbon-13 and nitrogen-15 isotopes ($^{13}\text{C}/^{15}\text{N}$ -6H-QseM) for single and double-labelled NMR experiments detailed in Chapter 5. FseA was expressed as a 6H-MBP-tagged construct with a TEV cleavage site for tag removal (6H-MBP-TEV-FseA). Protein expression methods are detailed in section 2.5.2. Purification methods are detailed in methods 2.5.3 – 2.5.5. Protein analysis by SDS-PAGE detailed in methods 2.5.7.

Purification and SDS-PAGE analysis was done for 6H-QseM (Figure 3.1), Se-6H-QseM (Figure 3.2), $^{15}\text{N}/^{13}\text{C}$ -6H-QseM (Figure 3.3). All target proteins eluted from the nickel-affinity column at 530 mM imidazole (as detected by significant absorbance at $\text{UV}_{280\text{nm}}$) and observed in the corresponding SDS-PAGE as prominent protein bands at ~10 kDa. This protein band at ~10 kDa corresponds with the mass of a 6H-QseM monomer, 10.960 kDa (Table 2.11). Although only one peak was observed in the Ni-NTA $\text{UV}_{280\text{nm}}$ absorbance chromatogram, SDS-PAGE analysis revealed additional protein bands indicating that Ni-NTA purification alone did not produce a pure sample. The subsequent SEC purification of the Ni-NTA-purified elution observed by SDS-PAGE maintained the 6H-QseM protein band at ~10 kDa and removed almost all observable contaminating protein bands expect for one with a molecular mass of ~15 kDa. This unidentified band at ~15 kDa may be an artefact since both NMR and SAXS of 6H-QseM (detailed in Chapter 5) presented results consistent with a homogeneous protein sample. This contaminating band was not excised or analysed further. 6H-QseM, Se-6H-QseM and $^{15}\text{N}/^{13}\text{C}$ -6H-QseM eluted at 95 mL volume point of the Superdex 200 16/600 SEC column, which is roughly consistent with the manufacturer (GE Healthcare) showing ribonuclease A (13 kDa) to elute at 90-100 mL on the same column under

similar pressure, temperature and buffer conditions. Calibration with protein standards of the SEC column used to purify 6H-QseM was not done, and exact molecular mass was only determined by SDS-PAGE during purification. Oligomeric state of the protein was later confirmed by both NMR and SAXS analysis detailed in chapter 5.

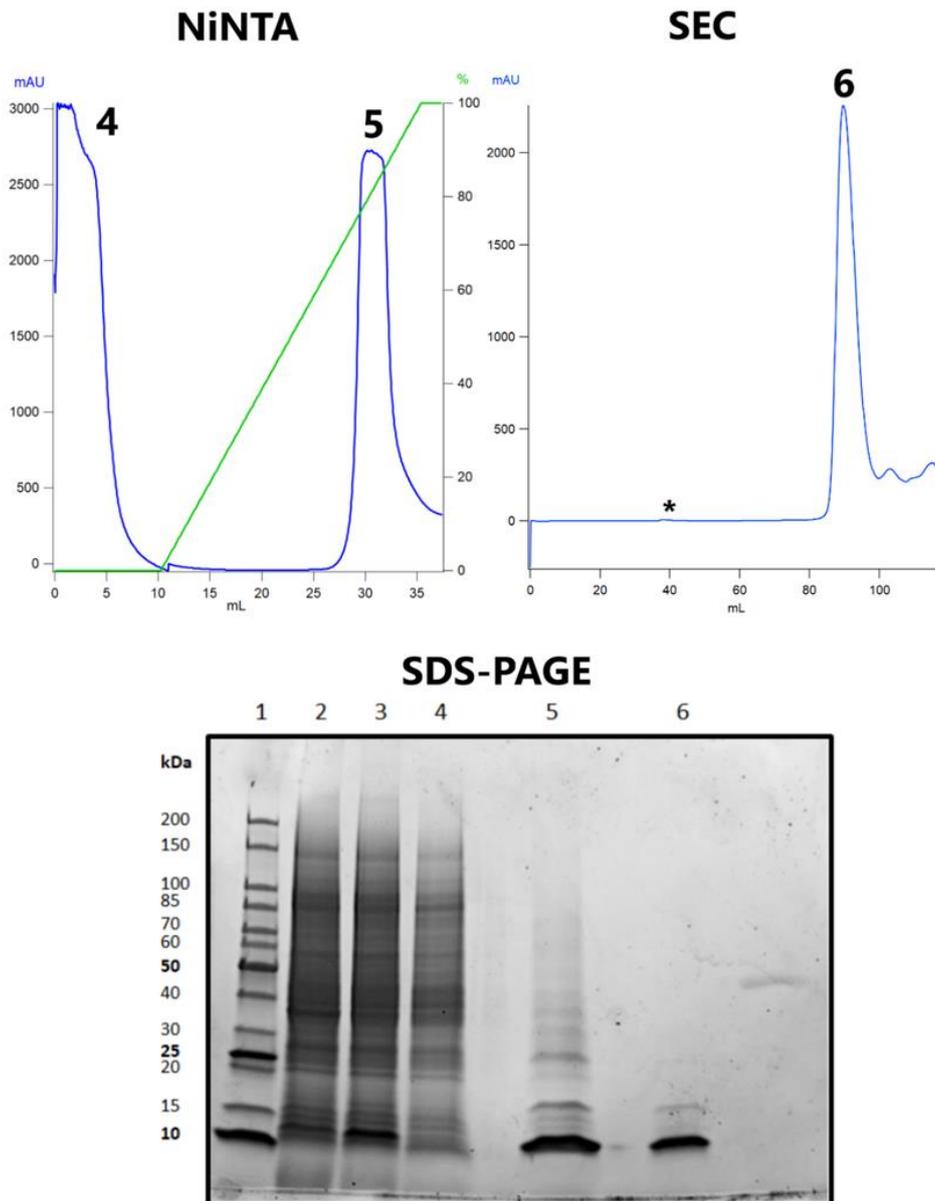


Figure 3.1: Purification of 6H-QseM. Ni-NTA affinity purification flow-through protein peak [4] and eluted protein peak [5], followed by subsequent SEC purification resulting in a single peak [6]. UV_{280nm} absorbance (mAU) presented as blue trace and imidazole concentration presented as green trace as buffer transitions from 50 mM (0%) to 800 mM (100%) imidazole. UV absorbance and imidazole concentration measured across total volume (mL) pumped through columns. All protein peak fractions were analysed on SDS-PAGE with cell lysate [2] and soluble protein fraction [3] against protein standard [1]. SEC column void volume [*] is ~40 mL, suspected aggregation peak observed at this point. Note that the UV_{280nm} trace for the Ni-NTA has been baseline corrected.

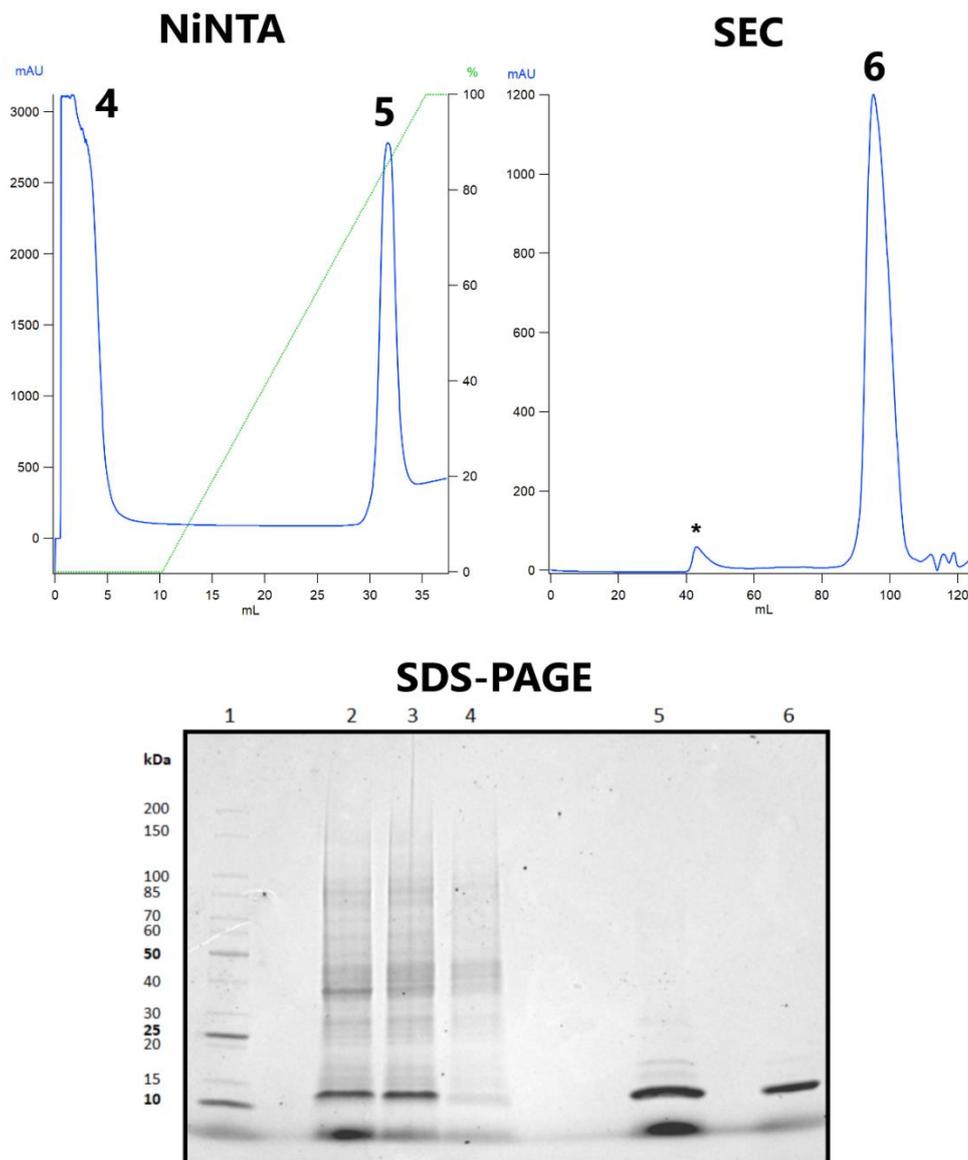


Figure 3.2: Purification of Se-6H-QseM. Ni-NTA affinity purification flow-through protein peak [4] and eluted protein peak [5], followed by subsequent SEC purification resulting in a single peak [6]. UV_{280nm} absorbance (mAU) presented as blue trace and imidazole concentration presented as green trace as buffer transitions from 50 mM (0%) to 800 mM (100%) imidazole. UV absorbance and imidazole concentration measured across total volume (mL) pumped through columns. All protein peak fractions were analysed on SDS-PAGE with cell lysate [2] and soluble protein fraction [3] against protein standard [1]. SEC column void volume [*] is ~40 mL, suspected aggregation peak observed at this point.

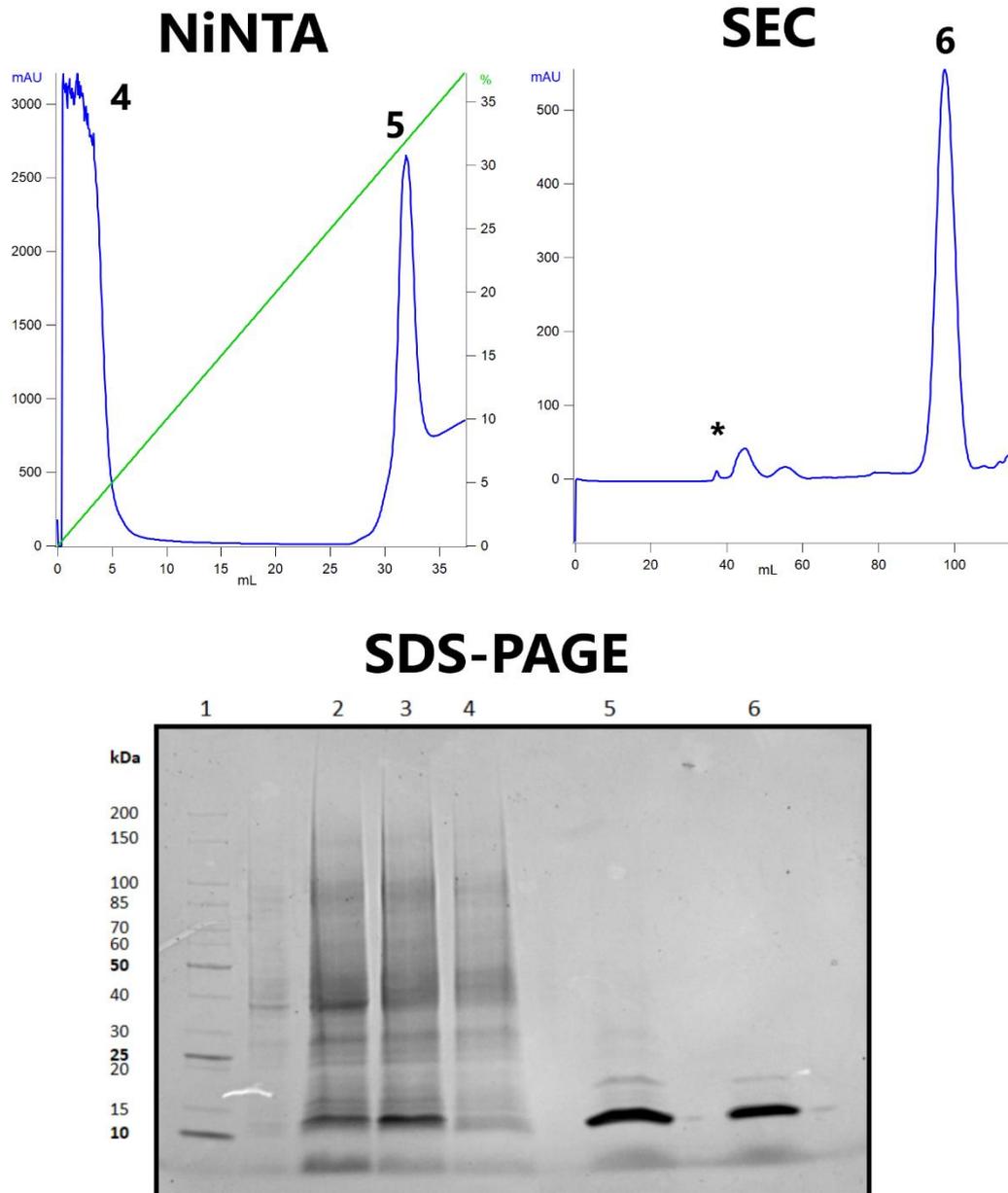


Figure 3.3: Purification of isotope-labelled 6H-QseM. Ni-NTA affinity purification flow-through protein peak [4] and eluted protein peak [5], followed by subsequent SEC purification resulting in a single peak [6]. UV_{280nm} absorbance (mAU) presented as blue trace and imidazole concentration presented as green trace as buffer transitions from 50 mM (0%) to 800 mM (100%) imidazole. UV absorbance and imidazole concentration measured across total volume (mL) pumped through columns. All protein peak fractions were analysed on SDS-PAGE with cell lysate [2] and soluble protein fraction [3] against protein standard [1]. SEC column void volume [*] is ~40 mL, suspected aggregation peak observed at this point.

3.2.2 Purification of untagged QseM

6H-TEV-QseM was purified identically to 6H-QseM with the addition of a TEV digestion (detailed in 2.5.4) and secondary Ni-NTA purification step to remove the cleaved 6H tag and any remaining 6H-TEV-QseM prior to SEC (Figure 3.4). Similar to 6H-QseM, 6H-TEV-QseM protein eluted from the nickel-affinity column at 530 mM imidazole and was observed in the corresponding SDS-PAGE as prominent protein bands at ~10 kDa. This protein band was between 10 and 15 kDa, consistent with the mass of a 6H-TEV-QseM at 12.704 kDa (Table 2.11). Digestion of 6H-TEV-QseM by TEV protease (detailed in 2.5.4) produced two protein bands on SDS-PAGE. The two bands from this digestion both had masses less than the 6H-TEV-QseM product before digestion, with the higher mass band (B1) likely to be QseM and the lower mass band (B2) likely to be the 6H-tag. This was further substantiated when a subsequent Ni-NTA purification step removed B2 but did not remove B1. SEC purification of B1 revealed a single peak at 90-105 mL on the SEC UV_{280nm} chromatogram and a single protein band at on SDS-PAGE at < 10 kDa, consistent with the expected mass of QseM, 9.780 kDa (Table 2.11). The final purified yield of QseM was ~5 mg/L of culture.

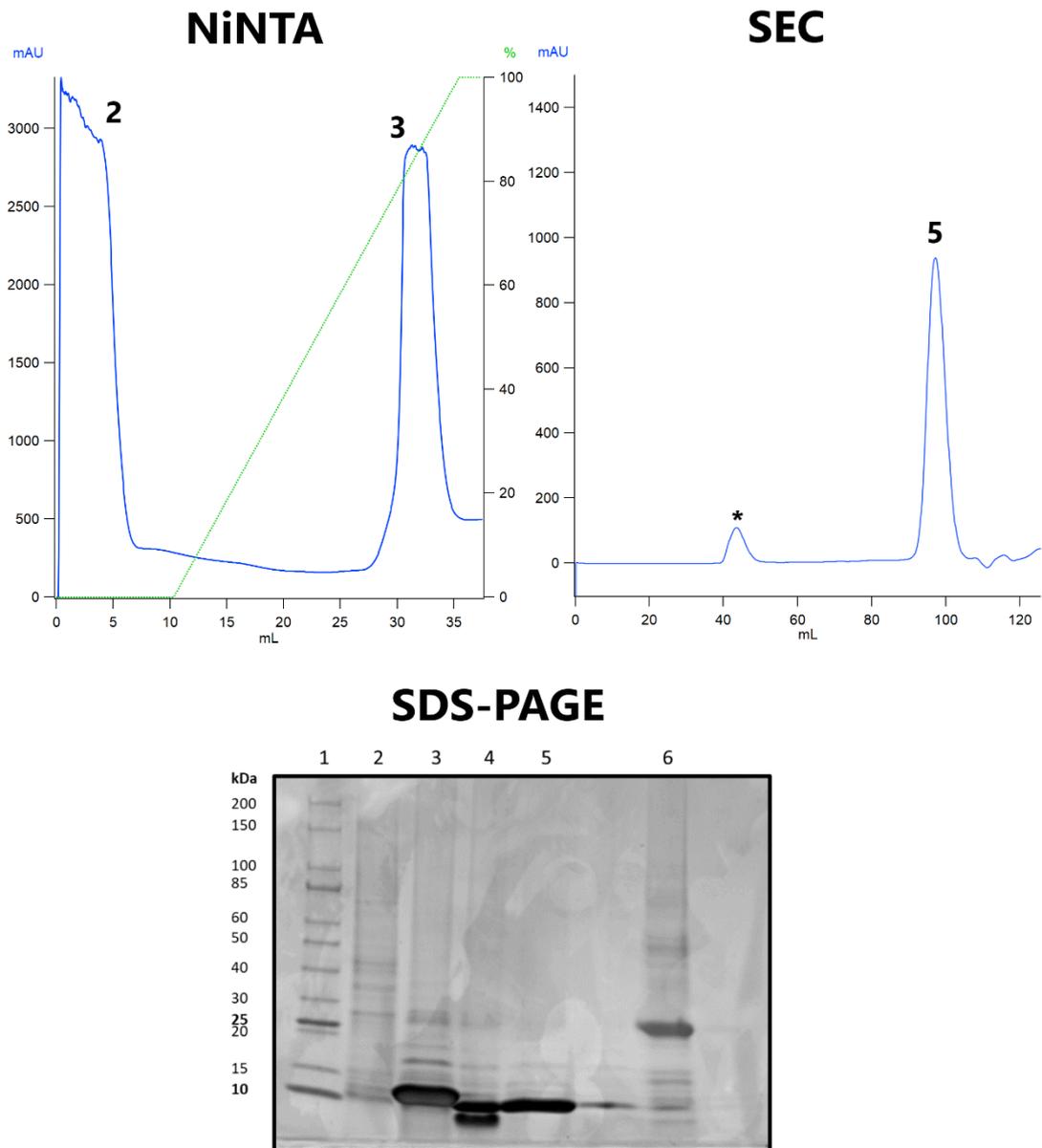
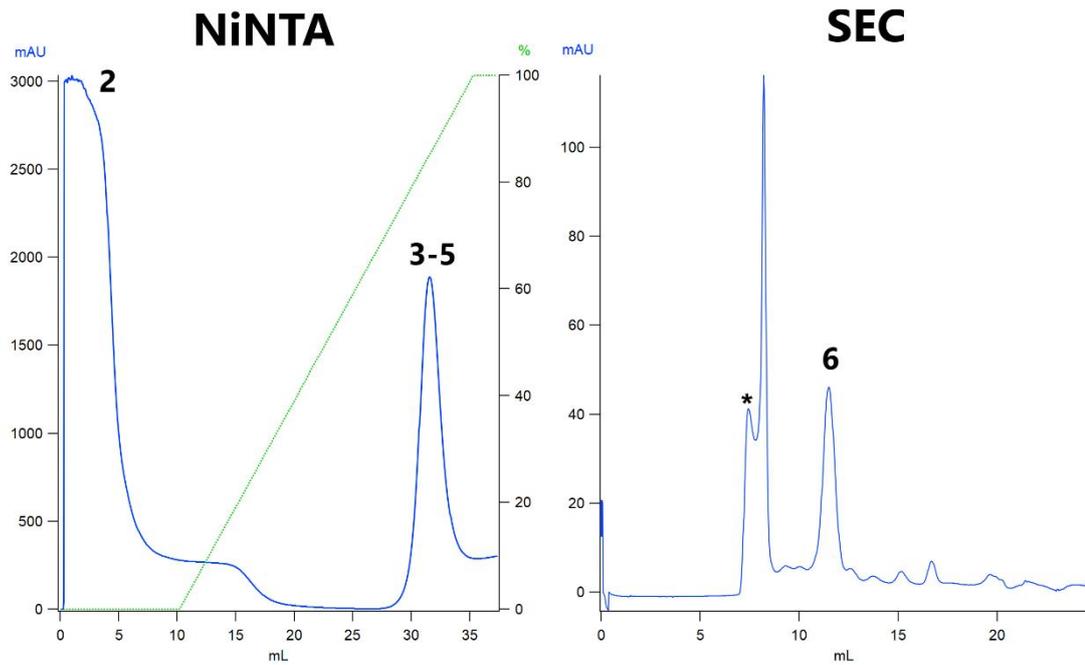


Figure 3.4: Purification of 6H-TEV-QseM. Ni-NTA affinity purification flow-through protein peak [2] and eluted protein peak [3]. A TEV protease cleaved 6H-QseM (now untagged) Ni-NTA purified sample [4]. UV_{280nm} absorbance (mAU) presented as blue trace and imidazole concentration presented as green trace as buffer transitions from 50 mM (0%) to 800 mM (100%) imidazole. UV absorbance and imidazole concentration measured across total volume (mL) pumped through columns. A secondary Ni-NTA affinity purification followed by SEC purification resulting in a single peak [5]. All protein peak fractions were analysed on SDS-PAGE against a protein standard [1] and the TEV protease sample used for cleavage [6]. SEC column void volume [*] is ~40 mL, suspected aggregation peak observed at this point.

3.2.3 Purification of 6H-MBP-TEV-FseA

Previous attempts to purify 6H-tagged FseA have been largely unsuccessful, possibly due to instability or insolubility of FseA (Will Jowsey, unpublished). A construct expressing FseA as a maltose-binding-protein fusion was therefore used in an attempt to increase solubility/stability of FseA. The 6H-MBP-TEV-FseA was able to be purified, however, within 24 h at 4 °C a significant aggregation peak was observed. 6H-MBP-FseA eluted at 11.5 mL volume of the Superdex 200 10/300 GL SEC column. An aggregation peak was also observed during SEC at 7.5-8.5 mL. Precipitation was observable within 2 h when the purified protein sample was stored at room temperature (20-30 °C). Purification and subsequent use of the protein was therefore carried out within a 24-hour period and, where possible, maintained at 4 °C to minimise loss of protein through aggregation or precipitation. 6H-MBP-TEV-FseA observable aggregation during SEC was lessened marginally by using a tris or phosphate with 100 mM NaCl and 2% v/v glycerol buffered solution that had a pH of 6.5. This pH also appeared to extend the time before visible precipitation was observed to 24-48 h. 6H-MBP-TEV-FseA was concentrated (2.5.6) to a maximum of 10 mg/mL but was diluted to 1-2 mg/mL and rapidly frozen in liquid nitrogen prior to storage at -80 °C. To limit potential aggregation and precipitation, 6H-MBP-TEV-FseA was thawed on ice and used within 24 h. Further purification attempts of 6H-MBP-FseA revealed that it forms a dimer in solution (Will Jowsey, unpublished). This fits with the expectation that, like other transcriptional activators, FseA may only be stable and function in a dimerised state. CD analysis of FseA was not attempted and SAXS data collection of FseA failed due to loss of protein by aggregation or degradation before X-ray exposure and acquisition could take place.



SDS-PAGE

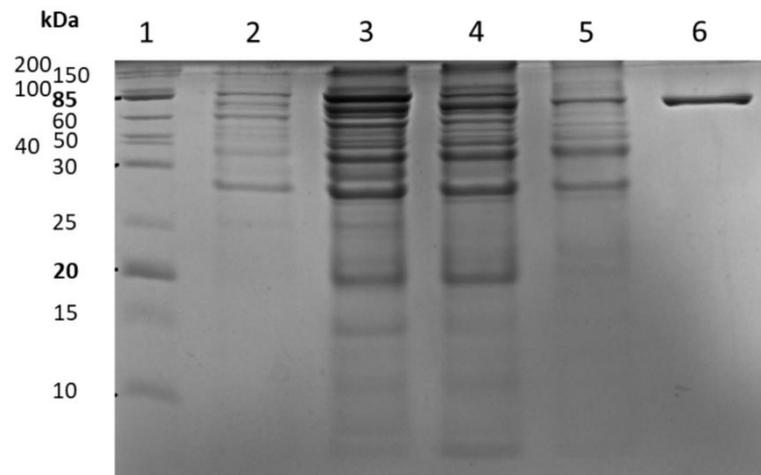


Figure 3.5: Purification of 6H-MBP-TEV-FseA. Ni-NTA affinity purification flow-through protein peak [2] and eluted protein peak [3-5] and SEC purified final product [6]. UV280nm absorbance (mAU) presented as blue trace and imidazole concentration presented as green trace as buffer transitions from 50 mM (0%) to 800 mM (100%) imidazole. UV absorbance and imidazole concentration measured across total volume (mL) pumped through columns. All protein peak fractions were analysed on SDS-PAGE against a protein standard [1]. SEC column (Superdex S200 10/300 GL, GE Healthcare) void volume [*] is ~7.5 mL, suspected aggregation peak observed at this point.

A yield of ~6 mg of purified 6H-QseM, Se-6H-QseM and $^{15}\text{N}/^{13}\text{C}$ -6H-QseM protein was obtained from 1 L of rich (LB) media, whereas a yield of ~4 mg of purified 6H-QseM protein was obtained from 1 L of minimal medium (for medium composition refer to Table 2.6). This difference was likely attributed to reduced cell number and/or nutrient availability of strains grown in minimal media. Purified 6H-QseM, Se-6H-QseM and $^{15}\text{N}/^{13}\text{C}$ -6H-QseM were concentrated (method 2.5.6) to a maximum of 15 mg/mL. For both crystallography and NMR experiments, protein concentration was ≤ 10 mg/ml (discussed in Chapters 4 and 5). Occasionally, protein was diluted to a 1-2 mg/ml concentration in 200-500 μL batches, frozen in liquid nitrogen and stored at -80 °C. 6H-QseM, Se-6H-QseM and $^{15}\text{N}/^{13}\text{C}$ -6H-QseM solubility consistently remained soluble at a concentration of ≤ 10 mg/mL in both tris or phosphate with 100 mM NaCl and 2% v/v glycerol buffers used in purification (Table 2.8 and 2.9) (using methods outlined in 2.5.7) at both room temperature and at 4 °C for at least 72 h from purification. 6H-QseM solubility was reduced in minimal salt buffer (Table 2.10), whereby precipitation was observed at concentrations >8 mg/mL within 24 h from purification.

3.2.4 Confirmation of protein post-purification via mass spectroscopy

We confirmed protein identity of 6H-QseM and 6H-MBP-TEV-FseA via mass spectrometry (as per method 2.5.7). Mass spectrometry results were searched using *MASCOT* (Matrix Science, (204, 205)) (Table 3.1), confirming 6H-QseM and 6H-MBP-TEV-FseA presence based on sequence matches created in multiple queries and calculated MOWSE (206) probability score (Table 3.2). Raw mass spectrometry analysis data shown in Table 3.3 and Table 3.4

Table 3.1: Mascot search parameters.

Search Type	MS/MS ion search
Enzyme	Trypsin
Variable modifications	Oxidation (M)
Mass values	Monoisotopic
Protein mass	Unrestricted
Peptide mass tolerance	± 0.2 Da
Fragment mass tolerance	± 0.2 Da
Max missed cleavages	1
Instrument type	ESI-QUAD

Table 3.2: Mascot search results for mass spectrometry analysis.

Protein	Matching queries (Total number)	Protein score*
6H-QseM	84 (302)	464
6H-MBP-TEV-FseA	123 (251)	1859

*Protein score based on individual query ion scores using MOWSE method

Table 3.3: 6H-QseM mass spectrometry analysis data.

6H-QseM Mass: 10653 Score: 464 Queries matched: 84 emPAI: 117.68									
Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
1	314.1603	626.3061	626.3136	-0.0075	0	-9	2.7	1	R.SHLDR.A
2	314.1653	626.316	626.3136	0.0023	0	14	0.88	1	R.SHLDR.A
3	314.1653	626.316	626.3136	0.0023	0	-5	6.1	1	R.SHLDR.A
4	318.1787	634.3429	634.3438	-0.001	0	17	0.64	1	K.ELFAR.-
22	585.7827	1169.551	1169.554	-0.0032	0	65	7.50E-06	1	R.ANWMVTTGYK.E
23	585.7827	1169.551	1169.554	-0.0032	0	-60	2.40E-05	1	R.ANWMVTTGYK.E
24	585.7827	1169.551	1169.554	-0.0032	0	-60	2.40E-05	1	R.ANWMVTTGYK.E
25	585.7827	1169.551	1169.554	-0.0032	0	-59	2.60E-05	1	R.ANWMVTTGYK.E
26	593.7809	1185.547	1185.549	-0.0016	0	-47	0.00045	1	R.ANWMVTTGYK.E + Oxidation (M)
27	593.7809	1185.547	1185.549	-0.0016	0	-43	0.0013	1	R.ANWMVTTGYK.E + Oxidation (M)
28	593.7809	1185.547	1185.549	-0.0016	0	-45	0.0008	1	R.ANWMVTTGYK.E + Oxidation (M)
29	593.7809	1185.547	1185.549	-0.0016	0	-55	7.60E-05	1	R.ANWMVTTGYK.E + Oxidation (M)
79	893.9474	1785.88	1785.887	-0.0069	1	-79	2.40E-07	1	R.ANWMVTTGYKELFAR.-
80	893.9474	1785.88	1785.887	-0.0069	1	-86	4.20E-08	1	R.ANWMVTTGYKELFAR.-
81	893.9474	1785.88	1785.887	-0.0069	1	-86	4.20E-08	1	R.ANWMVTTGYKELFAR.-
82	596.301	1785.881	1785.887	-0.0059	1	-51	0.00015	1	R.ANWMVTTGYKELFAR.-
83	596.301	1785.881	1785.887	-0.0059	1	-49	0.00025	1	R.ANWMVTTGYKELFAR.-
84	596.301	1785.881	1785.887	-0.0059	1	-49	0.00021	1	R.ANWMVTTGYKELFAR.-
85	596.301	1785.881	1785.887	-0.0059	1	-51	0.00015	1	R.ANWMVTTGYKELFAR.-
86	901.9453	1801.876	1801.882	-0.0061	1	99	2.10E-09	1	R.ANWMVTTGYKELFAR.- + Oxidation (M)
87	901.9453	1801.876	1801.882	-0.0061	1	-83	9.10E-08	1	R.ANWMVTTGYKELFAR.- + Oxidation (M)
88	901.9453	1801.876	1801.882	-0.0061	1	-85	4.80E-08	1	R.ANWMVTTGYKELFAR.- + Oxidation (M)
89	601.6344	1801.881	1801.882	-0.0007	1	-51	0.00014	1	R.ANWMVTTGYKELFAR.- + Oxidation (M)
90	601.6344	1801.881	1801.882	-0.0007	1	-67	3.30E-06	1	R.ANWMVTTGYKELFAR.- + Oxidation (M)
91	601.6344	1801.881	1801.882	-0.0007	1	-54	7.40E-05	1	R.ANWMVTTGYKELFAR.- + Oxidation (M)
92	601.6344	1801.881	1801.882	-0.0007	1	-56	3.80E-05	1	R.ANWMVTTGYKELFAR.- + Oxidation (M)
167	1338.125	2674.235	2674.247	-0.0119	0	-5	3.1	1	R.LLDASADDASEDEMAQLVLGIDPMR.E
168	892.4209	2674.241	2674.247	-0.0058	0	57	2.20E-05	1	R.LLDASADDASEDEMAQLVLGIDPMR.E
177	972.77	2915.288	2915.307	-0.0192	0	41	0.00059	1	K.VQDEVPWSDSLTAYDNEHFTIYMR.L
178	972.77	2915.288	2915.307	-0.0192	0	-26	0.017	1	K.VQDEVPWSDSLTAYDNEHFTIYMR.L
179	972.77	2915.288	2915.307	-0.0192	0	-36	0.0017	1	K.VQDEVPWSDSLTAYDNEHFTIYMR.L
182	972.77	2915.288	2915.307	-0.0192	0	-33	0.0036	1	K.VQDEVPWSDSLTAYDNEHFTIYMR.L
183	978.107	2931.299	2931.302	-0.0031	0	-32	0.0049	1	K.VQDEVPWSDSLTAYDNEHFTIYMR.L +
184	978.107	2931.299	2931.302	-0.0031	0	-16	0.19	1	K.VQDEVPWSDSLTAYDNEHFTIYMR.L +
185	761.8557	3043.394	3043.402	-0.0086	1	-71	5.10E-07	1	R.KVQDEVPWSDSLTAYDNEHFTIYMR.L
186	761.8557	3043.394	3043.402	-0.0086	1	-66	1.60E-06	1	R.KVQDEVPWSDSLTAYDNEHFTIYMR.L
187	761.8557	3043.394	3043.402	-0.0086	1	72	3.70E-07	1	R.KVQDEVPWSDSLTAYDNEHFTIYMR.L
188	761.8557	3043.394	3043.402	-0.0086	1	-66	1.40E-06	1	R.KVQDEVPWSDSLTAYDNEHFTIYMR.L
189	761.8557	3043.394	3043.402	-0.0086	1	-13	0.32	1	R.KVQDEVPWSDSLTAYDNEHFTIYMR.L
190	761.8557	3043.394	3043.402	-0.0086	1	-7	1.3	1	R.KVQDEVPWSDSLTAYDNEHFTIYMR.L
191	761.8557	3043.394	3043.402	-0.0086	1	-6	1.6	1	R.KVQDEVPWSDSLTAYDNEHFTIYMR.L
192	761.8557	3043.394	3043.402	-0.0086	1	-19	0.075	1	R.KVQDEVPWSDSLTAYDNEHFTIYMR.L
193	761.8557	3043.394	3043.402	-0.0086	1	-23	0.034	1	R.KVQDEVPWSDSLTAYDNEHFTIYMR.L

Type of search	MS/MS Ion Search
Enzyme	Trypsin
Variable modifications	Oxidation (M)
Mass values	Monoisotopic
Protein Mass	Unrestricted
Peptide Mass Tolerance	± 0.2 Da
Fragment Mass Tolerance	± 0.2 Da
Max Missed Cleavages	1
Instrument type	ESI-QUAD
Number of queries	302

194	761.8557	3043.394	3043.402	-0.0086	1	-17	0.13	1	R.KVQDEVWSDSLTAYDNEHFTIYMR.L
195	761.8557	3043.394	3043.402	-0.0086	1	-7	1.4	1	R.KVQDEVWSDSLTAYDNEHFTIYMR.L
196	761.8557	3043.394	3043.402	-0.0086	1	-5	2	1	R.KVQDEVWSDSLTAYDNEHFTIYMR.L

Table 3.4: 6H-MBP-TEV-FseA mass spectrometry analysis data.

6H-MBP-TEV-FseA Mass: 75171 Score: 1859 Queries matched: 123 emPAI: 5.02										
Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide	Type of search
8	435.2290	868.4434	868.4443	-0.0009	1	21	0.19	1	R.GYRDFLV.-	MS/MS Ion Search
9	435.2290	868.4434	868.4443	-0.0009	1	(21)	0.2	1	R.GYRDFLV.-	Enzyme Trypsin
10	435.2290	868.4434	868.4443	-0.0009	1	(21)	0.2	1	R.GYRDFLV.-	Variable modifications Oxidation (M)
11	435.2290	868.4434	868.4443	-0.0009	1	(21)	0.19	1	R.GYRDFLV.-	Mass values Monoisotopic
19	447.2570	892.4995	892.5018	-0.0023	0	(33)	0.015	1	K.DLLPNPPK.T	Protein Mass Unrestricted
20	447.2570	892.4995	892.5018	-0.0023	0	34	0.011	1	K.DLLPNPPK.T	Peptide Mass Tolerance ± 0.2 Da
24	473.2243	944.4341	944.4563	-0.0223	0	(5)	10	1	K.DVGVDNAGAK.A	Fragment Mass Tolerance ± 0.2 Da
25	473.2339	944.4533	944.4563	-0.0030	0	10	3.1	1	K.DVGVDNAGAK.A	Max Missed Cleavages 1
26	473.2339	944.4533	944.4563	-0.0030	0	(5)	11	1	K.DVGVDNAGAK.A	Instrument type ESI-QUAD
28	480.2658	958.5171	958.5196	-0.0025	0	(23)	0.14	1	R.TAVINAASGR.Q	Number of queries 302
29	480.2658	958.5171	958.5196	-0.0025	0	39	0.0039	1	R.TAVINAASGR.Q	
30	480.2658	958.5171	958.5196	-0.0025	0	(38)	0.0052	1	R.TAVINAASGR.Q	
31	480.2658	958.5171	958.5196	-0.0025	0	(12)	2.1	1	R.TAVINAASGR.Q	
34	493.7931	985.5717	985.5742	-0.0026	0	(33)	0.014	1	R.QIAMALLAR.Q	
35	493.7931	985.5717	985.5742	-0.0026	0	34	0.013	1	R.QIAMALLAR.Q	
36	501.7910	1001.5675	1001.5692	-0.0017	0	(28)	0.053	1	R.QIAMALLAR.Q + Oxidation (M)	
37	501.7910	1001.5675	1001.5692	-0.0017	0	(28)	0.047	1	R.QIAMALLAR.Q + Oxidation (M)	
38	501.7910	1001.5675	1001.5692	-0.0017	0	(31)	0.024	1	R.QIAMALLAR.Q + Oxidation (M)	
39	339.1693	1014.4861	1014.4851	0.0010	1	1	25	1	R.RGHMLMDR.G	
40	509.7697	1017.5248	1017.5277	-0.0029	0	(54)	0.00013	1	R.AALMITADGR.Q	
41	509.7697	1017.5248	1017.5277	-0.0029	0	55	0.00013	1	R.AALMITADGR.Q	
44	538.7699	1075.5252	1075.5332	-0.0079	0	18	0.45	1	R.IAATMENAQK.G	
45	538.7699	1075.5252	1075.5332	-0.0079	0	(8)	4.9	1	R.IAATMENAQK.G	
46	538.7699	1075.5252	1075.5332	-0.0079	0	(13)	1.5	1	R.IAATMENAQK.G	
50	595.3609	1188.7073	1188.7118	-0.0045	0	64	1e-005	1	K.AGLTFLVDLIK.N	
51	595.3609	1188.7073	1188.7118	-0.0045	0	(51)	0.00023	1	K.AGLTFLVDLIK.N	
52	595.3609	1188.7073	1188.7118	-0.0045	0	(57)	5.8e-005	1	K.AGLTFLVDLIK.N	
53	595.3609	1188.7073	1188.7118	-0.0045	0	(37)	0.0048	1	K.AGLTFLVDLIK.N	
55	601.3064	1200.5983	1200.6026	-0.0043	0	52	0.00014	1	K.TWEEIPALDK.E	
56	601.3064	1200.5983	1200.6026	-0.0043	0	(19)	0.32	1	K.TWEEIPALDK.E	
57	610.8179	1219.6213	1219.6250	-0.0038	1	27	0.056	1	R.RGWAWFLR.R	
58	610.8179	1219.6213	1219.6250	-0.0038	1	(25)	0.074	1	R.RGWAWFLR.R	
59	621.8251	1241.6356	1241.6404	-0.0048	0	68	4.1e-006	1	R.LSVDAIWPADR.M	

60	621.8251	1241.6356	1241.6404	-0.0048	0	(68)	4.2e-006	1	R.LSVDAIWPADR.M
63	634.3257	1266.6369	1266.6397	-0.0028	0	(35)	0.0079	1	K.LYPFTWDADR.Y
64	634.3257	1266.6369	1266.6397	-0.0028	0	35	0.0079	1	K.LYPFTWDADR.Y
68	437.2124	1308.6154	1308.6211	-0.0057	0	19	0.31	1	R.HALEHAFAER.W
69	437.2124	1308.6154	1308.6211	-0.0057	0	(13)	1.2	1	R.HALEHAFAER.W
70	437.2124	1308.6154	1308.6211	-0.0057	0	(5)	7.6	2	R.HALEHAFAER.W
71	668.8206	1335.6267	1335.6306	-0.0039	1	(39)	0.0025	1	K.SYEEELAKDPR.I
72	668.8206	1335.6267	1335.6306	-0.0039	1	41	0.0019	1	K.SYEEELAKDPR.I
73	668.8206	1335.6267	1335.6306	-0.0039	1	(39)	0.0026	1	K.SYEEELAKDPR.I
74	668.8206	1335.6267	1335.6306	-0.0039	1	(10)	2.2	1	K.SYEEELAKDPR.I
75	446.2173	1335.6301	1335.6306	-0.0005	1	(15)	0.74	1	K.SYEEELAKDPR.I
76	446.2173	1335.6301	1335.6306	-0.0005	1	(23)	0.12	1	K.SYEEELAKDPR.I
77	446.2173	1335.6301	1335.6306	-0.0005	1	(24)	0.096	1	K.SYEEELAKDPR.I
78	446.2173	1335.6301	1335.6306	-0.0005	1	(17)	0.48	1	K.SYEEELAKDPR.I
80	669.3740	1336.7334	1336.7391	-0.0057	0	104	8.9e-010	1	K.VNYGVTVLPFTK.G
81	669.3740	1336.7334	1336.7391	-0.0057	0	(89)	2.6e-008	1	K.VNYGVTVLPFTK.G
82	669.3740	1336.7334	1336.7391	-0.0057	0	(90)	2.5e-008	1	K.VNYGVTVLPFTK.G
84	446.5855	1336.7346	1336.7391	-0.0045	0	(5)	7.2	1	K.VNYGVTVLPFTK.G
91	710.3748	1418.7350	1418.7405	-0.0056	0	(85)	7.5e-008	1	K.ALSPETLSFDLAR.L
92	710.3748	1418.7350	1418.7405	-0.0056	0	85	7.2e-008	1	K.ALSPETLSFDLAR.L
93	475.2491	1422.7255	1422.7354	-0.0099	1	6	4.9	1	K.VTVEHPDKLEEK.F
97	716.4299	1430.8452	1430.8497	-0.0045	1	(38)	0.0031	1	K.AGLTFLVDLIKKN.H
98	716.4299	1430.8452	1430.8497	-0.0045	1	(77)	4.4e-007	1	K.AGLTFLVDLIKKN.H
99	716.4299	1430.8452	1430.8497	-0.0045	1	77	4.4e-007	1	K.AGLTFLVDLIKKN.H
100	716.4299	1430.8452	1430.8497	-0.0045	1	(50)	0.00024	1	K.AGLTFLVDLIKKN.H
101	477.9557	1430.8453	1430.8497	-0.0044	1	(49)	0.00025	1	K.AGLTFLVDLIKKN.H
102	477.9557	1430.8453	1430.8497	-0.0044	1	(52)	0.00013	1	K.AGLTFLVDLIKKN.H
103	477.9557	1430.8453	1430.8497	-0.0044	1	(49)	0.00026	1	K.AGLTFLVDLIKKN.H
104	477.9557	1430.8453	1430.8497	-0.0044	1	(1)	16	1	K.AGLTFLVDLIKKN.H
109	536.6038	1606.7896	1606.7964	-0.0068	1	25	0.066	1	R.VQADWTDPRNHLR.D
110	402.7050	1606.7910	1606.7964	-0.0055	1	(25)	0.069	1	R.VQADWTDPRNHLR.D
111	543.2884	1626.8434	1626.8478	-0.0044	0	(82)	1.2e-007	1	R.FVLQALDGSLLGGASHR.Q
112	543.2884	1626.8434	1626.8478	-0.0044	0	(81)	1.7e-007	1	R.FVLQALDGSLLGGASHR.Q
113	543.2884	1626.8434	1626.8478	-0.0044	0	95	6.3e-009	1	R.FVLQALDGSLLGGASHR.Q
114	543.2884	1626.8434	1626.8478	-0.0044	0	(76)	5.4e-007	1	R.FVLQALDGSLLGGASHR.Q
120	598.3446	1792.0120	1792.0207	-0.0087	0	37	0.0036	1	R.DLQLVVEGAEVLRPVR.L
121	897.0135	1792.0124	1792.0207	-0.0083	0	(28)	0.028	1	R.DLQLVVEGAEVLRPVR.L
122	619.6478	1855.9216	1855.9257	-0.0042	1	60	1.9e-005	1	K.AFQDKLYPFTWDADR.Y
123	619.6478	1855.9216	1855.9257	-0.0042	1	(44)	0.00076	1	K.AFQDKLYPFTWDADR.Y
133	631.0343	1890.0809	1890.0866	-0.0056	0	95	4.1e-009	1	K.LIAYPIAVEALSIIYNK.D
134	631.0343	1890.0809	1890.0866	-0.0056	0	(82)	7.2e-008	1	K.LIAYPIAVEALSIIYNK.D
135	631.0343	1890.0809	1890.0866	-0.0056	0	(61)	9.2e-006	1	K.LIAYPIAVEALSIIYNK.D
136	631.0343	1890.0809	1890.0866	-0.0056	0	(55)	4.3e-005	1	K.LIAYPIAVEALSIIYNK.D
137	946.0479	1890.0813	1890.0866	-0.0052	0	(49)	0.00015	1	K.LIAYPIAVEALSIIYNK.D
138	946.0479	1890.0813	1890.0866	-0.0052	0	(56)	2.9e-005	1	K.LIAYPIAVEALSIIYNK.D
139	946.0479	1890.0813	1890.0866	-0.0052	0	(34)	0.0052	1	K.LIAYPIAVEALSIIYNK.D
140	946.0479	1890.0813	1890.0866	-0.0052	0	(3)	6.2	1	K.LIAYPIAVEALSIIYNK.D
141	475.0134	1896.0244	1896.0330	-0.0086	1	(56)	4e-005	1	R.LRFVLQALDGSLLGGASHR.Q

142	633.0156	1896.0250	1896.0330	-0.0079	1	73	7.5e-007	1	R.LRFVLQALDGSGLGGASHR.Q
143	633.2844	1896.8314	1896.8312	0.0003	0	68	2.5e-006	1	K.HMNADTDYSIAEAAFNK.G
148	682.7050	2045.0932	2045.0946	-0.0013	1	41	0.0011	1	K.LVIWINGDKGYNGLAEVGK.K
150	1049.0198	2096.0250	2096.0313	-0.0063	0	72	8.6e-007	1	K.EFLENYLLTDEGLEAVNK.D
151	1049.0198	2096.0250	2096.0313	-0.0063	0	(64)	5.3e-006	1	K.EFLENYLLTDEGLEAVNK.D
153	1058.5825	2115.1505	2115.1576	-0.0071	1	(22)	0.094	1	K.DVGVDNAGAKAGLTFLVDLIK.N
154	706.0583	2115.1531	2115.1576	-0.0045	1	(59)	1.8e-005	1	K.DVGVDNAGAKAGLTFLVDLIK.N
155	706.0583	2115.1531	2115.1576	-0.0045	1	60	1.3e-005	1	K.DVGVDNAGAKAGLTFLVDLIK.N
156	706.0583	2115.1531	2115.1576	-0.0045	1	(7)	2.6	1	K.DVGVDNAGAKAGLTFLVDLIK.N
157	713.7215	2138.1425	2138.1484	-0.0059	0	(90)	1.4e-008	1	K.GQPSKPFVGVLSAGINAASP.NK.E
158	713.7215	2138.1425	2138.1484	-0.0059	0	90	1.3e-008	1	K.GQPSKPFVGVLSAGINAASP.NK.E
159	713.7215	2138.1425	2138.1484	-0.0059	0	(77)	2.7e-007	1	K.GQPSKPFVGVLSAGINAASP.NK.E
160	713.7215	2138.1425	2138.1484	-0.0059	0	(90)	1.4e-008	1	K.GQPSKPFVGVLSAGINAASP.NK.E
161	1071.5092	2141.0038	2141.0074	-0.0036	0	49	0.00019	1	K.GEIMPNIQMSAFWYAVR.T + 2 Oxidation (M)
166	568.5704	2270.2526	2270.2607	-0.0081	1	31	0.0076	1	R.NAHRDLQLVVEGAELRVPVR.L
171	785.1067	2352.2982	2352.3092	-0.0110	1	27	0.018	1	R.YNGKLIAYPIAVEALSILYK.D
176	592.7989	2367.1665	2367.1760	-0.0095	1	31	0.009	1	R.HALEHAFAERWFVVELVR.S
182	645.8569	2579.3985	2579.4071	-0.0086	1	44	0.00039	1	K.GQPSKPFVGVLSAGINAASP.NKELAK.E
183	645.8569	2579.3985	2579.4071	-0.0086	1	(36)	0.0022	1	K.GQPSKPFVGVLSAGINAASP.NKELAK.E
187	922.5308	2764.5707	2764.5778	-0.0071	1	(13)	0.27	1	K.LIAYPIAVEALSILYKDLLPNPK.T
188	922.5308	2764.5707	2764.5778	-0.0071	1	(47)	9.4e-005	1	K.LIAYPIAVEALSILYKDLLPNPK.T
189	922.5308	2764.5707	2764.5778	-0.0071	1	(46)	0.00012	1	K.LIAYPIAVEALSILYKDLLPNPK.T
190	922.5308	2764.5707	2764.5778	-0.0071	1	(47)	9.1e-005	1	K.LIAYPIAVEALSILYKDLLPNPK.T
191	692.1503	2764.5721	2764.5778	-0.0058	1	(49)	6.9e-005	1	K.LIAYPIAVEALSILYKDLLPNPK.T
192	692.1503	2764.5721	2764.5778	-0.0058	1	(39)	0.00056	1	K.LIAYPIAVEALSILYKDLLPNPK.T
193	692.1503	2764.5721	2764.5778	-0.0058	1	53	2.5e-005	1	K.LIAYPIAVEALSILYKDLLPNPK.T
194	754.9158	3015.6342	3015.6505	-0.0163	1	4	2.2	1	R.DLQLVVEGAELRVPVRLSVDIWPADR.M
195	1017.5132	3049.5179	3049.5266	-0.0087	1	(11)	0.6	1	K.GEIMPNIQMSAFWYAVRTAVINAASGR.Q
196	1017.5132	3049.5179	3049.5266	-0.0087	1	(6)	1.6	1	K.GEIMPNIQMSAFWYAVRTAVINAASGR.Q
197	1017.5132	3049.5179	3049.5266	-0.0087	1	(17)	0.15	1	K.GEIMPNIQMSAFWYAVRTAVINAASGR.Q
198	763.3873	3049.5202	3049.5266	-0.0064	1	25	0.022	1	K.GEIMPNIQMSAFWYAVRTAVINAASGR.Q
199	1022.8458	3065.5156	3065.5215	-0.0059	1	(4)	2.7	1	K.GEIMPNIQMSAFWYAVRTAVINAASGR.Q + Oxidation (M)
200	1022.8458	3065.5156	3065.5215	-0.0059	1	(11)	0.53	1	K.GEIMPNIQMSAFWYAVRTAVINAASGR.Q + Oxidation (M)
202	1028.1757	3081.5051	3081.5164	-0.0113	1	(6)	2	1	K.GEIMPNIQMSAFWYAVRTAVINAASGR.Q + 2 Oxidation (M)
203	773.1631	3088.6232	3088.6332	-0.0099	1	(36)	0.0017	1	K.EFLENYLLTDEGLEAVNKDKPLGAVALK.S
204	773.1631	3088.6232	3088.6332	-0.0099	1	(37)	0.0013	1	K.EFLENYLLTDEGLEAVNKDKPLGAVALK.S
205	773.1631	3088.6232	3088.6332	-0.0099	1	(50)	6.1e-005	1	K.EFLENYLLTDEGLEAVNKDKPLGAVALK.S
206	773.1631	3088.6232	3088.6332	-0.0099	1	(8)	1.2	1	K.EFLENYLLTDEGLEAVNKDKPLGAVALK.S
207	1030.5485	3088.6238	3088.6332	-0.0094	1	58	1.2e-005	1	K.EFLENYLLTDEGLEAVNKDKPLGAVALK.S
208	1030.5485	3088.6238	3088.6332	-0.0094	1	(56)	1.7e-005	1	K.EFLENYLLTDEGLEAVNKDKPLGAVALK.S
209	1030.5485	3088.6238	3088.6332	-0.0094	1	(42)	0.00041	1	K.EFLENYLLTDEGLEAVNKDKPLGAVALK.S
211	1056.5211	3166.5415	3166.5402	0.0014	1	34	0.0026	1	R.IAATMENAQKGEIMPNIQMSAFWYAVR.T

3.3 Conclusions

Prior experimentation showed that the N-terminal fused 6H tag did not appear to affect QseM activity *in vivo* (Calum Morris, unpublished). We concluded that if QseM's function was not impacted by the presence of the N-terminal tag then the 6H-QseM structure would be similar to native QseM. We found purified 6H-QseM, Se-6H-QseM and $^{15}\text{N}/^{13}\text{C}$ -6H-QseM remained soluble at 10 mg/mL in tris or phosphate with 100 mM NaCl and 2% v/v glycerol buffered solutions at pH 7 – 8 at room temperature, 4 °C or frozen and stored at -80 °C. No observable precipitation was detected.

We also purified an untagged TEV-cleavable variant of QseM. Tobacco etch virus (TEV) protease digestion removed the 6H tag from 6H-TEV-QseM, leaving a single N-terminal serine attached to the native QseM amino acid sequence. It is possible that the cleaved QseM protein may have reduced stability or solubility compared to the 6H-QseM since a proportionally larger aggregation peak to soluble peak was observed during SEC for QseM, however, solubility or stability for QseM were not further investigated.

6H-MBP-TEV-FseA was also purified by nickel-affinity and SEC. We were only able to purify FseA in SDS-PAGE observable quantities when fused to an MPB tag. Precipitation of 6H-MBP-TEV-FseA was observed visually following purification and a substantial proportion of the protein was observed as aggregate in SEC. This suggests 6H-MBP-TEV-FseA may be unstable and prone to degradation and/or aggregation. Solubility and protein stability were not explicitly examined for 6H-MBP-TEV-FseA. We concluded that precipitation and/or aggregation were the most likely factors that lessened total purification yield of 6H-MBP-TEV-FseA. Crystallisation screens were completed immediately after purification and before precipitation was observed (discussed in Chapter 4). Further investigation is required to determine a variant or formulation of FseA that is less prone to aggregation and precipitation that would be more suitable in future *in vitro* experiments. This investigation would likely include small scale purification to optimise method and buffer against FseA variants.

Chapter 4:

Structural analysis via X-ray crystallography

4.1 Introduction:

The DUF2285 domain is often located in small (<150 residues) proteins, with it being the only recognised domain present in the protein, often comprising most of the amino acid sequence. Some proteins (~10% of proteins characterised with DUF2285 in PFAM) also contain a second uncharacterised DUF6499 domain (*PFAM*).

DUF2285 and DUF6499 are almost exclusive to the Alphaproteobacteria phylum, with both predominantly found in either Rhizobiaceae or Sphingomonadaceae families. Until recently, QseM and homologous of it are the only annotated proteins that have the DUF2285 domain, including the C-terminus half of FseA (Msi171) (130) (134).

There is an association with DUF2285, and QS systems given that QseM homologous are found near QS proteins like TraR/TraI and homologous thereof, often all being present on mobile genetic elements. Additionally, a DUF2285 domain containing protein has been implicated to have a role in QS regulation of the type VI secretion system of *Yersinia pseudotuberculosis*, one of the rare instances DUF2285 is found outside of Alphaproteobacteria (207). This *Y. pseudotuberculosis* YPIII expresses a protein called RovC which contains the DUF2285 domain. Only recently a structure was solved for the RovC protein (208), revealing it to function as a hexameric transcription factor that promotes the expression of T6SS4. The C-terminus 181-247 residues of RovC which are form the DUF2285 domain fold into 4 alpha-helices, distinctly resembling a HTH motif.

The structure of RovC presents the DUF2285 domain as a HTH motif and directly interacts with DNA to promote gene transcription. This closely resembles FseA since it is of similar size (30 kDa) and function. QseM appears distinctly different in that it does not seem to bind DNA and only interacts with FseA. There is a possibility that structural variance of the DUF2285 domain may be responsible for these distinctly different functions which may be elucidated by structure determination of QseM.

X-ray crystallography is a gold-standard method to determine 3D arrangement of atoms within a molecule, in a repeating lattice, such as a crystal. X-ray crystallography has been used for over a century to solve the structure of small biomolecules and the first protein structures were resolved in the late 1950s (209, 210). The method relies on X-ray waves being scattered in distinct ways by an object such as a crystalline body. Crystals are made up of a repeating pattern of identical molecules and this lattice of ordered molecules cause X-rays to be diffracted into specific directions (“reflections”) which can be measured by highly sensitive photon detectors. The intensity and position of these diffracted reflections are collected and measured, which together with the phase information, allow obtaining the electron density map and ultimately the three-dimensional structure of the protein under investigation. In protein X-ray crystallography it is this electron density map that is used along with information about the protein’s primary structure, to create a completed structure, with resolution high enough to give atom-level detail (211-213).

Protein crystallisation was first observed around 1840 (214), scientifically investigated in the 1880s (215) but became critically associated with determining the atomic structure of macromolecules in the 1930s (216) after the characterisation of X-ray crystal diffraction (217, 218). In 1960 the first protein crystal structure was published. Protein crystallisation presented a significant obstacle to the types of proteins that could be investigated through X-ray diffraction, however, development of methods (219) to improve protein crystallisation over the past 50 years has paved the way to solve more than 160,000 structures currently in the Protein Data Bank (PDB) by X-ray crystallography as of 2021 (220). These methods often work by mixing a purified protein sample with a crystallisation solution and sealed in a chamber with a higher concentration crystallisation solution and due to vapour diffusion, the protein begins to slowly come out of solution into an ordered, repeating pattern, thus forming a crystal. The crystallisation solution requires a mixture of various precipitants, buffers, and salts. When this process of crystallisation is successful, many trillions of copies of the protein will be arranged in a repeating 3D crystal lattice, with individual crystals being large enough to be observed under a light microscope. Typically, a set of uniform crystals are

used in a single crystallisation experiment. Crystals are harvested by manually “scooping” them from a crystallisation drop using a small nylon loop mounted on a pin. Often these crystals are transferred to a cryoprotectant solution prior to submerging in liquid nitrogen to limit ice formation during cryocooling. Poorly cooled crystals often form ice crystals within the protein crystal, resulting in damage that translates to poor diffraction data. Additionally, excessive ice formation can obscure the protein crystal, potentially hindering positioning the crystal in the beamline. Poorly positioned crystals may rotate out of the beamline during data collection creating incomplete datasets.

After crystals are mounted and cryocooled, they are transferred to an X-ray source (typically at a synchrotron) to be exposed to a focused X-ray beam for small periods of time. The photon detector is located behind the crystal and during exposure it records diffracted beams traveling from the crystal. During each exposure the crystal is rotated by a small angle (1° or less) before another exposure is made. Depending on the symmetry of the crystal, a larger or smaller wedge of data is necessary to ensure a complete dataset is captured. For the lowest possible symmetry (triclinic) 180° of data is required for standard data, and 360° for anomalous data. Radiation damage will accumulate as multiple exposures take place through a common plane in a crystal. Excessive radiation damage will reduce the quality of diffraction data and is a limiting factor for number of complete datasets that can be collected. It is specifically because of this damage that cooling and maintaining the crystal at low temperatures (typically at 100K) and using multiple crystals is common practice for any protein X-ray crystallography experiment.

After diffraction data collection the unit cell is determined and the diffraction images are processed into a single file that contains all the information about each diffraction point, including its Miller indices, reflection intensity and the error in this intensity and finally, space group parameters are determined. These initial pieces of information are essential to describe the number and orientation of molecules that exist in the asymmetric unit.

Further processing and computation, or further experimentation is required to obtain phase information, which cannot be collected directly during diffraction

as only structure factor amplitudes can be derived from the experiment. The phase component of the structure factors is required for calculating electron density map. Molecular replacement is a common method used for determining phase information, but it requires a model of a protein of similar structure. However, since no model existed for QseM or any other DUF2285-domain protein at the time of this work, phase information was measured experimentally. Phase information can be experimentally measured using anomalous diffraction from atoms that can absorb X-ray radiation which are incorporated into the crystal. When diffraction data is collected with X-rays close to the absorption edge of an element present in the crystal, the intensities of reflections are altered slightly in an effect called anomalous diffraction. In this experiment, the structure factor amplitudes of pairs of reflections which would be identical in a normal experiment are now slightly different, and these differences encode information about the positions of the anomalously scattering atoms in the crystal lattice. These positions can be used to derive phases, which can be combined with the diffraction structure factor amplitudes to generate an electron density map. Incorporation of anomalous scattering atoms can be achieved post crystallisation using heavy atom 'soaking' (221) or, a more convenient and less invasive approach, is by expressing the protein in a defined minimal medium that facilitates the incorporation of the non-standard amino acid L-selenomethionine in lieu of sulfur-containing L-methionine (222, 223). The selenium atom of L-selenomethionine has an X-ray absorption edge at 12.66 keV which is in the useful range for diffraction experiments. The crystallographic structures of *Agrobacterium tumefaciens* TraR and TraM were solved this way (102, 224).

4.2 Results and Discussion:

4.2.1 Observation of crystallisation

Crystallisation screens using 6H-QseM in tris with 100 mM NaCl and 2% v/v glycerol buffer were carried out using the Hampton Research Index screen along with 'Crystal 1' and 'Crystal 2', 'PEG/ion' and 'Natrix' screens (2.6.1). Initial crystallisation screening led to the identification of 2 conditions exhibiting crystal formation (Figure 4.1). However, examination under UV radiation indicated they were likely salt crystals. The initial screens were carried out in a phosphate buffer (PBS buffer A in Table 2.8) and phosphate is known to rapidly crystallise with divalent cations such as Zn^{2+} . The phosphate-based buffer was exchanged with a tris-based one during SEC of QseM (Table 2.9). Subsequent crystallisation attempts did not appear to produce the salt crystals observed in PBS buffer.

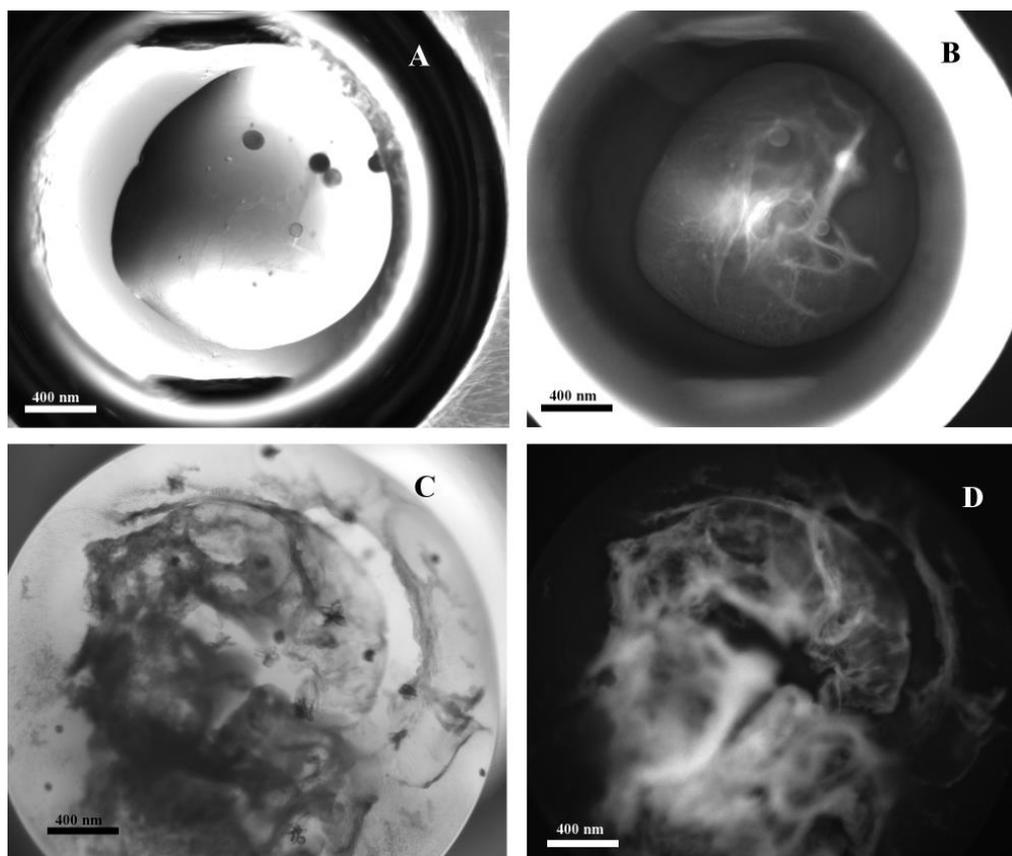


Figure 4.1: Attempts to crystallise 6H-QseM in PBS buffer. (A) Visible light image showing circular structures in 0.05 M Zinc Acetate di-hydrate and 20% polyethylene glycol 3350 and (B) its UV fluorescence equivalent. (C) Visible light image showing star shaped non-uniformed crystals and heavy precipitation in Tacsimate™ (Hampton research) at a pH 7.0 and (D) its UV fluorescence equivalent. Fluorescence was observed only in precipitate, suggesting that the crystals were salt and not protein.

Crystallisation experiments were attempted using protein solution at concentrations 1.5, 2.0, 2.5, 3.0, 3.2, 4.0, 5.0, 6.5, 8.0, 10.0 and 12.0 mg/ml. Crystals were only observed between 4-10 mg/mL produced crystalline formations across the crystallisation screens after 7 days. Concentrations at 3 mg/mL or lower did not produce crystalline formations and concentrations \geq 10 mg/mL often precipitated. Four conditions (Table 4.1 and Figure 4.2) generated crystalline formations. All screens contained crystals with similar morphologies, however, many involved multiple crystals fused together. The Index screen containing 0.1 M bis-tris pH 6.5, 28% (v/v) polyethylene glycol

monomethyl ether (PEG MME) 2,000 (Crystal screen ID D11) was selected as the most promising condition to pursue (Figures 4.2 and 4.3) as it mostly contained single crystals. Crystallisation was observed with all three ratios of protein:reservoir (1:2, 1:1 and 2:1), however the 1:2 protein:reservoir ratio predominantly gave the largest crystal size and greatest number of crystals.

Table 4.1: 6H-QseM crystal producing conditions.

Crystal screen	Salt/s	Buffer/s	Precipitant/s	Average pH
Index (D11)	-	0.1 M bis-tris (pH 6.5)	28% PEG MME 500	6.5
Index (G7)	0.2 M NH ₄ CH ₃ CO ₂	0.1 M bis-tris (pH 6.5)	25% PEG 3350	6.5
Crystal (F9)	0.1 M NaH ₂ PO ₄ ·H ₂ O 0.1 M NaH ₂ PO ₄	0.1 M MES monohydrate (pH 6.5)	2.0 M NaCl	5.4
Natrix (C7)	0.005 M MgCl ₂ ·6H ₂ O	0.05 M HEPES sodium (pH 7.0)	25% PEG MME 500	5.7

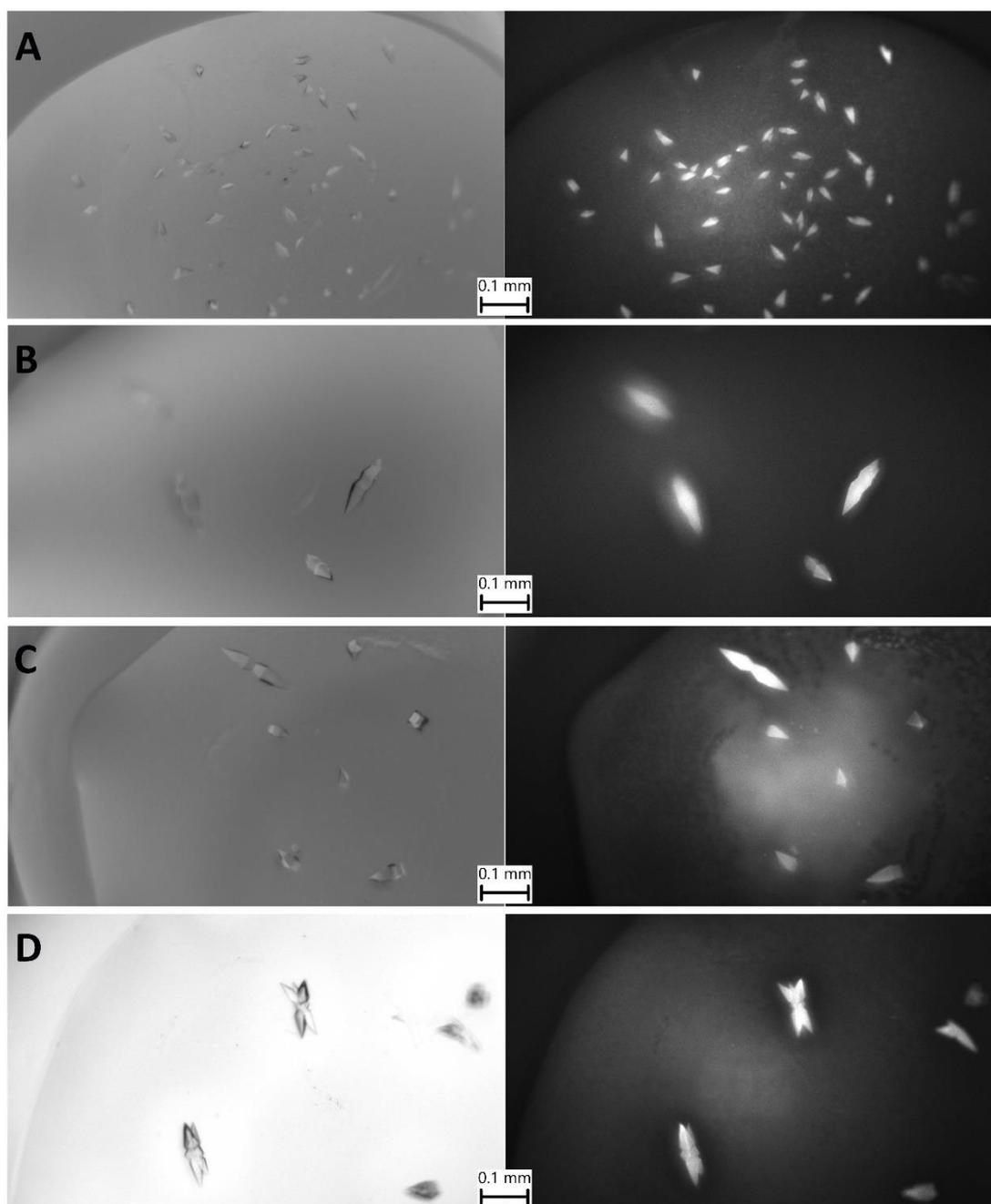


Figure 4.2: Crystal screening of 6H-QseM. Crystal condition Index D11 (A), Index G7 (B), Crystal F9 (C) and Natrix (C7) results shown as visible light (left) and matching UV (right) images (imaged by Rigaku Minstrel crystal imager, see 2.6.2). All images are from drops of 1:2 ratio of protein:reservoir except B, which had a 2:1 ratio. All have similar morphologies and strongly fluorescence under UV.

4.2.2 Optimisation of 6H-QseM crystallisation

Optimisation of the D11 crystallisation condition was attempted in 24-well sitting-drop trays (as detailed in 2.6.1) using a 1:2 protein:reservoir ratio. Optimisation trays exploring a pH range from 6.0 to 8.0 and PEG MME 2000 concentrations between 20% and 30% were carried out. A pH between 6.0 and 6.2 and a PEG MME 2000 concentration of 26-28% was determined to be the most optimal crystallisation condition. Crystals formed just outside these pH and PEG concentration ranges had lower consistency in crystal number, size and morphology (Figure 4.3). Initially, the optimisation drops contained 2 μL of protein and 4 μL of reservoir, this was later adjusted to 3 μL protein and 6 μL reservoir, as the larger drop size was less distorted by the loop during crystal harvesting. Drop sizes below a total volume of 6 μL were not attempted to avoid pipetting inconsistencies which may occur with protein component < 2 μL . Total drop volumes were not >9 μL due to limited protein sample volume available. A 1 μL volume of micro-seeding stock (Section 2.6.3) was added to the 9 μL protein:reservoir drop in later crystallisation trays. Micro-seeding gave a consistent increase in number and size of crystals formed per drop (see method 2.6.3). The protein concentration for these larger optimisation drops was 2.0 ± 0.5 mg/mL. Attempts to use protein concentration in optimisation >2.5 mg/mL often resulted in drops with no observable crystal formations or precipitation. The lowest concentration used was that produced crystals was 1.5 mg/mL. Drops with a protein concentration of 1 mg/mL did not produce any observable crystal formations or any precipitation.

No further crystallisation or increases in crystal size were observed after 14 days (Figure 4.4). After this time crystals began to visibly deteriorate. Crystals were therefore harvested for X-ray data collection within 14 days to maximise growth while minimising possible degradation.

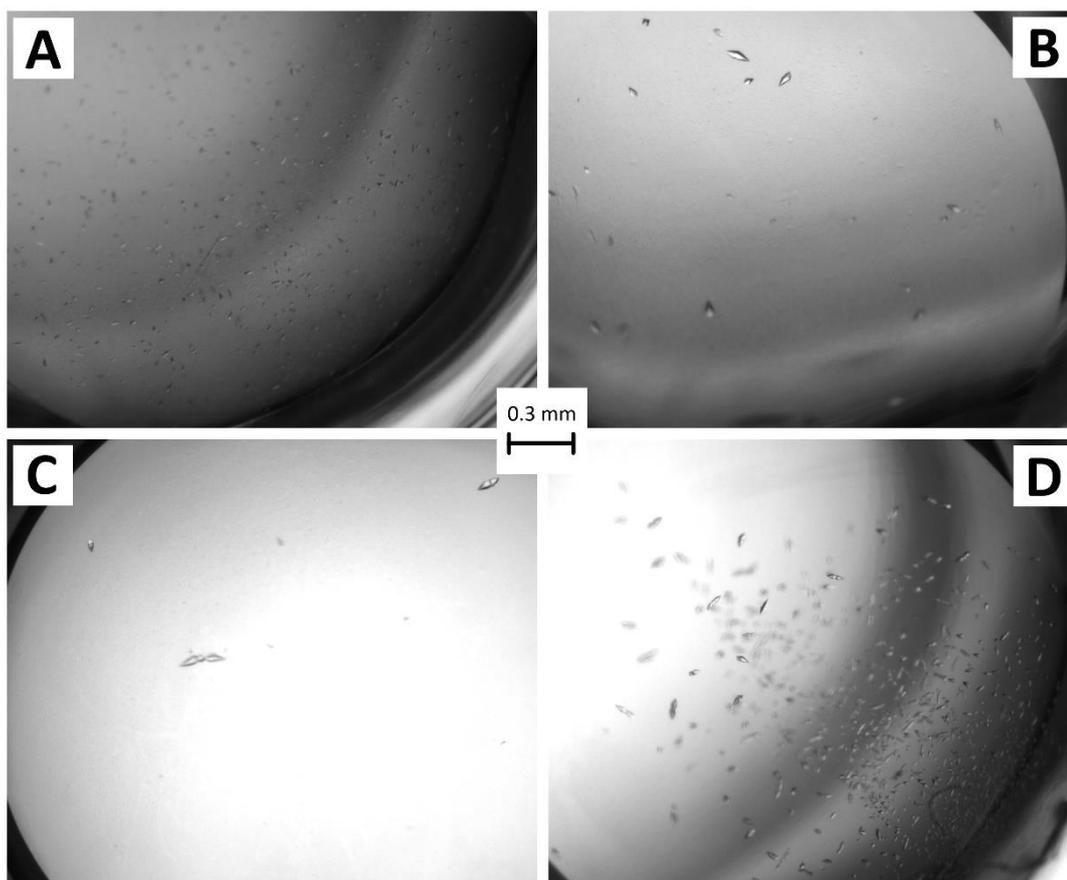


Figure 4.3: pH and PEG effects on crystallisation of 6H-QseM in D11 condition. Showing typical crystallisation results in the following conditions: (A) pH 6.0 - 6.2 and 25 % PEG; (B) pH 6.0 - 6.2 and 27% PEG; (C) pH 6.4 - 6.6 and 27% PEG and (D) and pH 6.0 - 6.2 and 28% PEG.

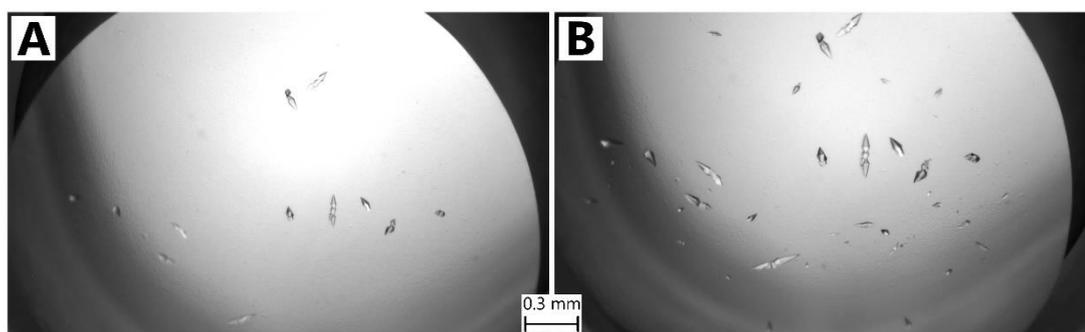


Figure 4.4: growth over time of 6H-QseM crystals. Images of the same crystallisation drop after 4 days (A) and 12 days (B). Little change was noted after 12 days.

The largest crystals were selected for X-ray diffraction experiments as they were easier to collect and provided a larger internal volume for increased diffraction signal intensity. Multiple crystals growing from a shared nucleation point occurred more frequently in larger crystals (Figure 4.5a), these crystals often broke into smaller fragments during harvesting, often presenting as single pyramid shapes. These pyramid shaped fragments were considered to be a single crystal but were often heavily damaged during X-ray beam exposure and were not used. The majority of 6H-QseM crystals formed a 'diamond' shape and did not break into small pyramids (Figure 4.5b). We believed that these diamond shaped crystals were likely two or more crystals fused together.

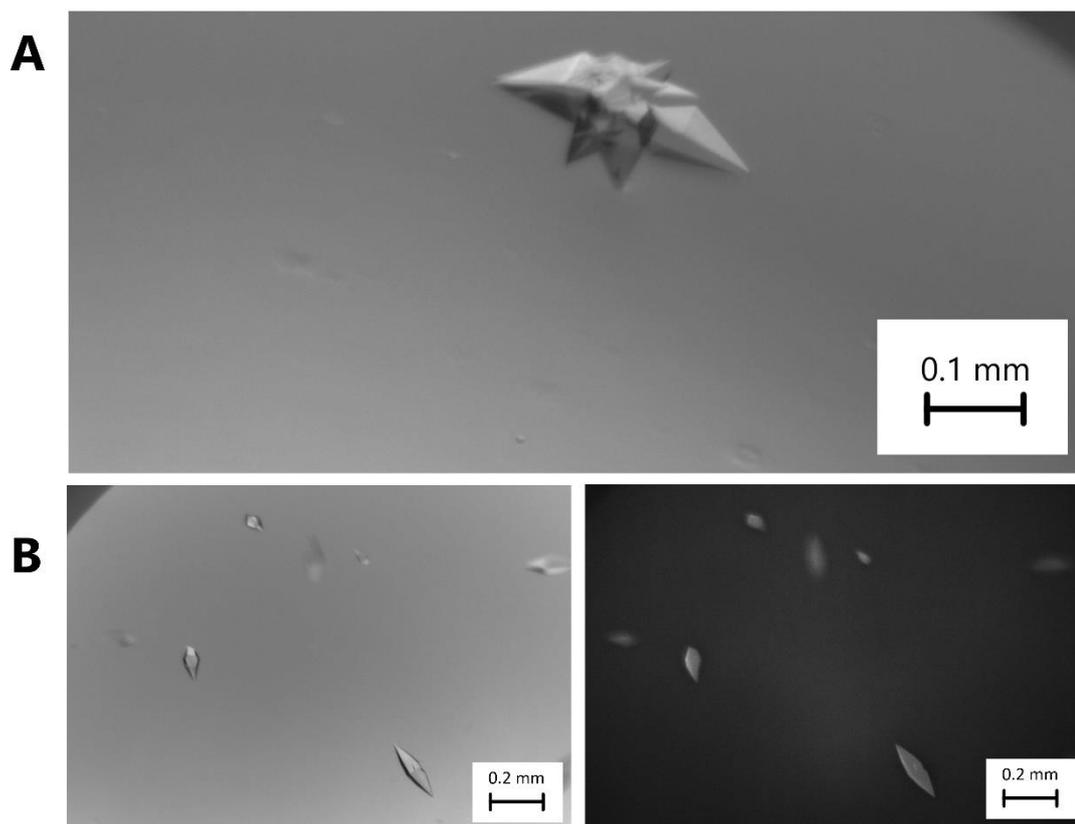


Figure 4.5: 6H-QseM crystallisation optimisation. Crystals formed in a 24-well optimisation-tray drop in 0.1 M bis-tris pH 6.5, 28% (v/v) PEG MME 2,000. Multi-fusion crystal bodies were commonly found in some drops (A), however, many optimisation drops reliably formed ‘diamond’ (likely double crystals from a shared nucleation point), these were often the largest and were imaged under light and UV (B).

4.2.3 Crystallisation of Se-6H-QseM

Crystals of Se-6H-QseM were grown in identical conditions to 6H-QseM. The crystal morphology was similar and there was no indication that incorporation of SeMet into 6H-QseM had any noticeable effect on crystallisation.

4.2.4 Cryoprotection and cryocooling crystals

Five cryoprotectants were tested on multiple crystals (Section 2.6.4), each with variable soaking times (Table 4.2). Since the 6H-QseM crystals were formed in a PEG MME 2,000 solution, we focused on using polyethylene glycol 500

and ethylene glycol as the cryoprotectant. Glycerol was also used as it is a component in the purification buffers and is also present in many of the crystallisation screen formulations. Glucose and 2-methyl-2,4-pentanediol (MPD) were also tested. Ethylene glycol, PEG 500 and glycerol all had reduced icing compared to non-cryoprotected crystals. Glucose and MPD did not appear to reduce icing compared to non-cryoprotected crystals. Cryocooled crystals were only observed for icing after transport and mounting on the beamline and it is not known if it occurred immediately, in transit, or when mounted on the beamline. Reduced diffraction quality was noted for crystals when significant icing occurred (to the extent that crystals were difficult to visualise on the beamline). Cryoprotectant screening was not exhausted, however, we deemed that soaking 6H-QseM crystals in ethylene glycol, PEG 500 or glycerol all resulted in equally improved cryocooling and employed in all subsequent cryocooling attempts.

Table 4.2: Concentrations and effects of cryoprotectants used.

Solution	Percentage (v/v)	Notes
Ethylene glycol	10-35%	Glassy, clear appearance, no obvious crystal deterioration if soaked for a maximum of 3 minutes
Polyethylene glycol 500	20-40%	Glassy, clear appearance up to 30%. Degradation avoided at 40% by pre-soaking in a 20-30% solution prior
Glycerol	20-40%	Glassy, clear appearance at 30%, degradation appearing at >5 minutes at 40%
Glucose	10-50%	No observable degradation but icing occurred during cryocooling
MPD (2-methyl-2,4-pentanediol)	10-30%	Visible degradation occurred within a minute at >15%, however $\leq 10\%$ resulted in icing during cryocooling.

We suspected that these crystals may be multiples with two crystal lattices growing outwards from a point. Crystals were frozen in a length-ways orientation (Figure 4.6). This orientation was least likely to collect datasets through multiple crystals. This orientation also allowed us to maximise number of individual datasets that could be collected per crystal.



Figure 4.6: Cryocooled 6H-QseM crystal mounted at AS MX1 beamline. Cryoprotected crystal positioned lengthways to collect multiple datasets from different areas of the crystal (image taken with AS beamline sample camera, using AreaVision interface).

4.2.5 Crystallisation of FseA

We attempted to crystallise 6H-MBP-TEV-FseA using similar methods to those for 6H-QseM. While few MBP-protein fusions have been crystallised, those that have (199, 225) are often small proteins or peptides. Given the poor solution behaviour of untagged FseA, we decided to attempt crystallisation of the MBP fusion protein. Index and Crystal screens (Hampton) were used with protein concentrations of 3.0, 4.0 and 8.0 mg/mL, however no crystallisation was not observed. We attributed this lack of crystallisation partly due to either the solubilising effect of the MBP tag, the poor stability of the FseA protein itself, or the inherent flexibility of the fusion protein, despite the inclusion of 17 amino acid rigged linker (including the TEV recognition site).

4.2.6 Diffraction data collection and analysis

Diffraction data obtained using 6H-QseM crystals were collected at the MX1 and MX2 beamlines at the Australian Synchrotron (AS). The native and initial SeMet datasets were collected on the ADSC Quantum 315r detector, whereas later diffraction data from SeMet crystals was collected on an Eiger2 16M detector (Figure 4.7). Diffraction data from the ADSC detector was integrated using iMosflim, whereas the Eiger2 diffraction data was integrated using XDS (methods outlined in Section 2.6.6).

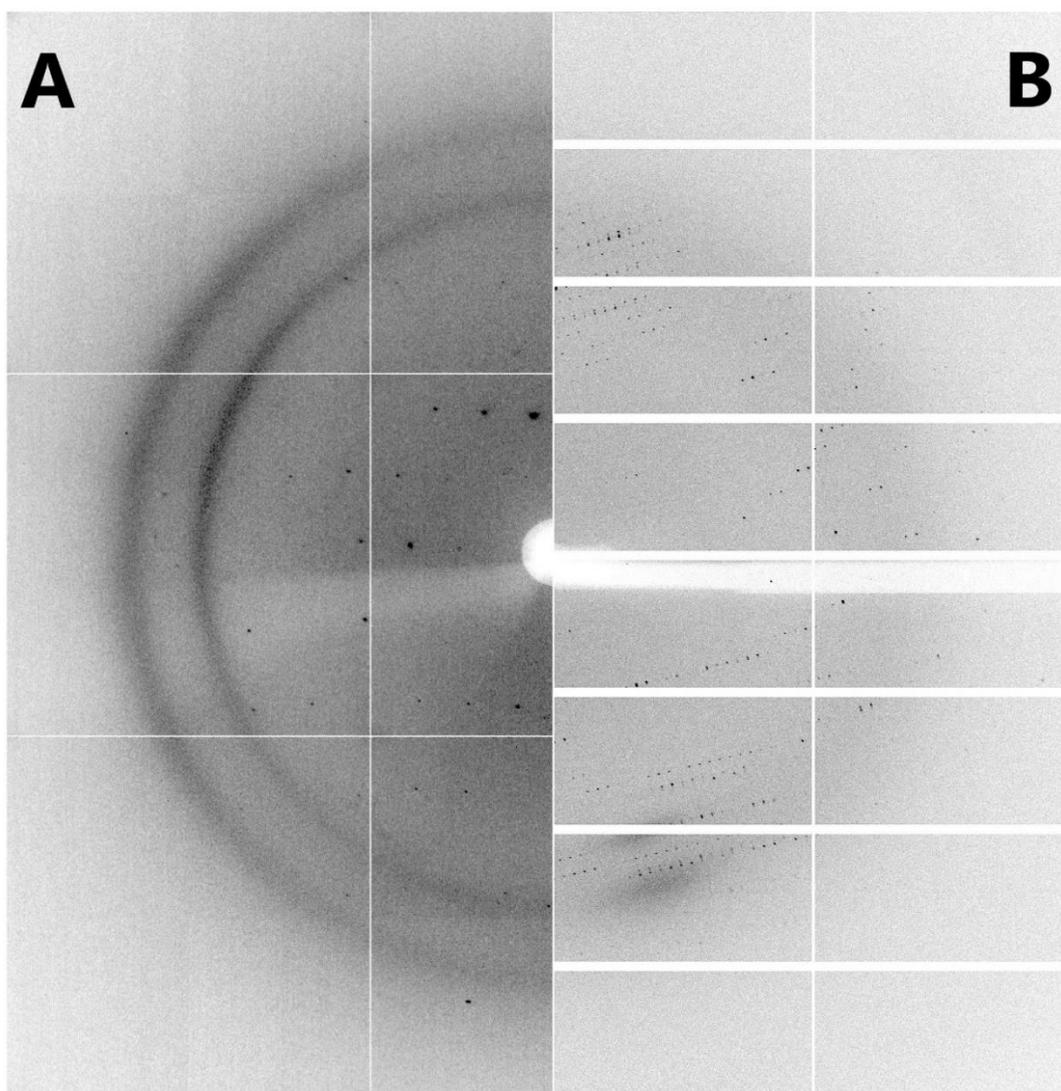


Figure 4.7: 6H-QseM diffraction comparison. 6H-QseM crystal diffraction using SeMet-peak wavelength on an ADSC Quantum 315r (A) at a crystal-detector distance of 399.97 mm (max. resolution of 2.5 Å) and on the Eiger2 16M (B) at a crystal-detector distance of 384.02 mm (max. resolution of 2.4 Å). The SeMet-optimised crystals produced diffraction spots further out from the centre on the Eiger2 16M detector, corresponding to higher resolution data. Less exposure time reduced background noise, potentially revealing weak diffraction spots. Improved crystallisation freezing eliminated a diffraction ‘ring’ from ice surrounding the protein crystal. Higher diffraction spot resolution reveals potential ‘spot splitting’ which may indicate multiple crystals in a single dataset or crystal splitting. Note that improved resolution and reduced ice diffraction are independent of detector change, however spot separation may have only been apparent on the Eiger2 16M detector.

Two diffraction datasets were collected from native 6H-QseM crystals and 42 were collected from Se-6H-QseM crystals. Peak, inflection point and remote wavelengths were determined as per methods 2.7.1 (Table 4.3). Of the 42 Se-6H-QseM crystal diffraction datasets, 13 were collected at inflection point (12688 eV), 26 at peak (12367 eV) and 3 at high remote (13200 eV) wavelengths. The quality of data was determined by half-dataset correlation coefficient ($CC_{1/2}$) and R_{merge} scores and diffraction limit (defined by $1/\sigma(I)$) (as per Section 2.7.1). The majority of poor datasets had diffraction limit >3.5 Å attributed to advanced crystal age or damage during harvesting and cryocooling. These problems were reduced by harvesting crystals more delicately at an earlier stage in formation and using cryoprotectants prior to cryocooling. These methods improved the quality of data by increasing diffraction limit and giving higher $CC_{1/2}$ and lower R_{merge} scores at higher resolution. Unusual doubling of diffraction spots was apparent in some of these higher resolution images.

Table 4.3: Determined MAD wavelengths for Se-6H-QseM.

	Energy (eV)	Wavelength (Å)	Theoretical (eV)
Se Inflection point	12688	0.9772	12655
Se Peak	12867	0.9636	12660
Se High remote	13200	0.9393	13200

SeMet crystals generated using optimised methods referred to as SeMet-optimised hereon. The datasets for the SeMet crystals were divided into peak and inflection point wavelengths when determining the data quality. Peak datasets had better quality compared to inflection point datasets since the later were collected from the same crystal after peak datasets and may have been impacted by radiation damage accumulated during peak dataset collection.

The output of *AIMLESS* (Figure 4.8) shows quality of data (measured as a $CC_{1/2}$ score) dramatically falling at < 3.5 Å for SeMet-optimised inflection point data, < 3.0 Å for SeMet-optimised peak data and < 2.8 Å for native data. R_{merge}

was worse with all SeMet data exceeding a score of 1.0 at 4 Å and the native data exceeding a score of 1.0 at 3.5 Å. Structure factor intensity (measured as mean $I/\sigma(I)$) fell below 2.0 at a resolution of 3.5 Å for SeMet data, 3.2 Å for SeMet-optimised inflection data, 3.0 Å for SeMet-optimised peak data and 2.9 Å for native data.

All datasets had completeness (from *AIMLESS*, overall completeness value) of over 91% for the resolution range collected (excluding a single outlier SeMet peak dataset with 76.5% overall completeness and 73.5% anomalous completeness). The native data was the most complete at 99.9%. The average SeMet peak, inflection point, and remote data had a 99.6% completeness (anomalous at 99.7% completeness). The average SeMet-optimised inflection point data had a completeness of 88.2% and the average SeMet-optimised peak data had a completeness of 93.8% (anomalous at 91.5% completeness).

Additionally, there was indication of twinning occurring in some of the datasets. Native and initial SeMet diffraction datasets appeared untwinned. Higher resolution inflection point and peak datasets collected from SeMet crystals were more likely to be twinned. Twinning likelihood was determined by the Ctruncate L-Test (226) (Figure 4.9). Almost all datasets were indexed as tetragonal, however, a few were indexed as monoclinic (detailed in 4.2.7). These occurrences of monoclinic indexing were mostly associated with higher resolution, SeMet-optimised datasets. Indication of twinning occurred only in these SeMet-optimised datasets, in particular, the poorer quality inflection point datasets. It is possible that these lower symmetry space groups are giving false positive twinning results.

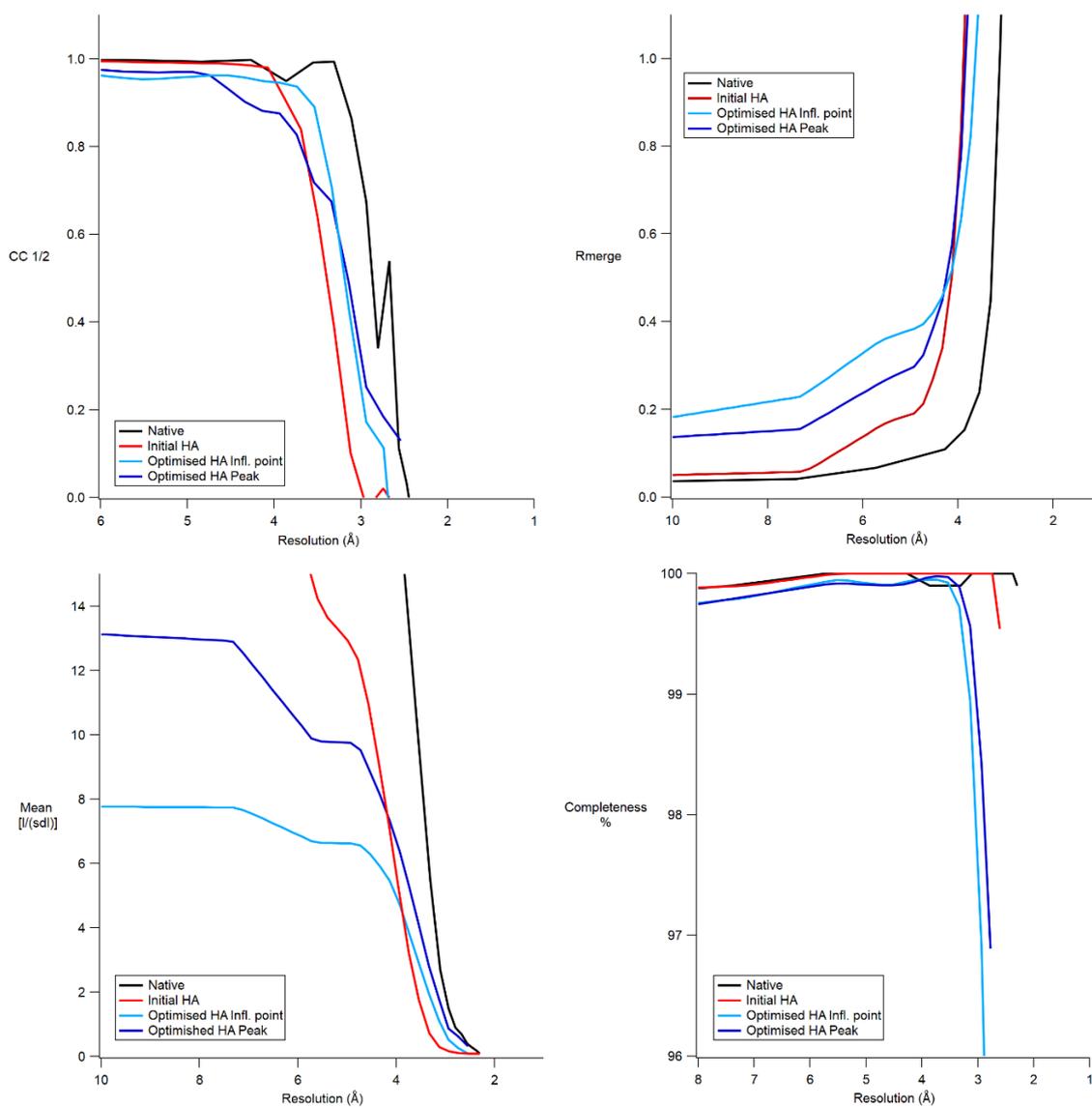


Figure 4.8: Diffraction data quality and intensity. Comparison of native (black), initial SeMet (red), optimised SeMet 6H-QseM inflection point (light blue) and optimised SeMet 6H-QseM peak (dark blue) diffraction dataset statistics. $CC_{1/2}$ (A) and R_{merge} (B) data quality scores and diffraction signal (C) (as mean $I/\sigma(I)$) and completeness (D) are plotted against resolution (Å). Native data was the best in completeness and quality compared to all other dataset types. SeMet-optimised peak data quality and completeness was better than the inflection point data, likely due to radiation damage for the collection of the inflection point datasets. SeMet dataset $CC_{1/2}$ and completeness were better than the SeMet-optimised datasets, likely from reading split spots from the SeMet-optimised diffraction as unique structure factors.

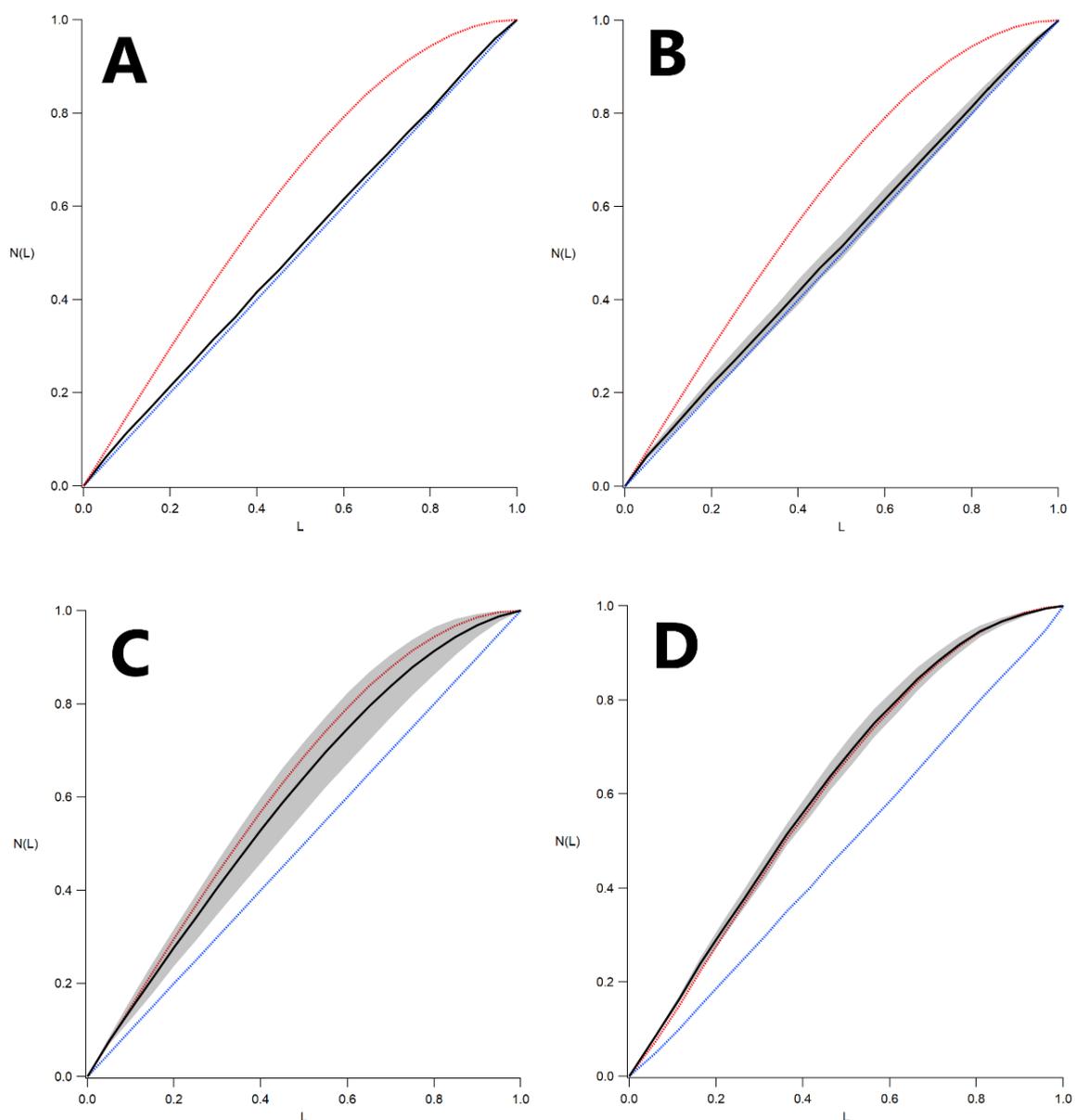


Figure 4.9: Ctruncate L-Test twinning analysis. Native (A), initial SeMet (B), SeMet-optimised peak (C) and SeMet-optimised inflection point (D) L-test scores plotted against ideal un-twinned (blue) and perfectly twinned (red) lines. Results are presented as a black line average where multiple datasets are concerned, with grey shade showing individual dataset variation ranges. Probability for twinning is higher for the SeMet-optimised datasets, particularly the inflection point subset.

Four individual datasets were assessed based on quality, with the best resolution native, SeMet peak, SeMet-optimised peak and SeMet-optimised inflection point datasets used to solve the phase problem (Table 4.4). *SHELXC/D/E* and *PHASER* were both used in parallel to solve the phase problem. These datasets were also used in attempts to solve the phase problem using molecular replacement as described in 5.2.7

Table 4.4: Statistics for best diffraction datasets of crystallised 6H-QseM (scaled and merged in AIMLESS).

	Energy (wavelength)	Resolution Range	Space group/s	Unit cell dimensions	Completeness*	$I/\sigma(I)^*$	CC $\frac{1}{2}*\dagger$	Anomalous resolution	Anomalous Completeness
Native	13.002 KeV (0.95360)	34.97 – 2.80 Å (2.95 – 2.80 Å)	P 4 ₁ 2 ₁ 2 P 4 ₃ 2 ₁ 2	a = 34.97, b = 34.97, c = 118.33 $\alpha = 90^\circ, \beta = 90^\circ, \gamma = 90^\circ$	100.0 (100.0)	13.6 (1.2)	0.98 (0.619)	N/A	N/A
Se Peak ‡ (initial)	12.867 KeV (0.96360)	35.70 – 3.20 Å (3.46 – 3.20 Å)	P 4 ₁ 2 ₁ 2 P 4 ₃ 2 ₁ 2	a = 35.70, b = 35.70, c = 119.79 $\alpha = 90^\circ, \beta = 90^\circ, \gamma = 90^\circ$	100.0 (100.0)	10.7 (0.8)	0.999 (0.505)	4.63 Å	100.0 (100.0)
Se Peak ‡ (optimised)	(13.000 KeV) (0.95373)	39.83 – 2.80 Å (2.95 – 2.80Å)	P 4 ₁ 2 ₁ 2 P 4 ₃ 2 ₁ 2	a = 36.95, b = 36.95, c = 119.48 $\alpha = 90^\circ, \beta = 90^\circ, \gamma = 90^\circ$	100.0 (100.0)	7.3 (2.6)	0.992 (0.835)	5.44 Å	100.0 (100.0)
Se Inflection point	12.689 KeV (0.97710)	36.75 – 3.00 Å (3.18 – 3.00 Å)	P 4 ₁ 2 ₁ 2 P 4 ₃ 2 ₁ 2	a = 36.75, b = 36.75, c = 119.59 $\alpha = 90^\circ, \beta = 90^\circ, \gamma = 90^\circ$	100.0 (100.0)	4.6 (1.3)	0.993 (0.410)	N/A	N/A

* Values in parenthesis refer to the highest resolution shell.

† *AIMLESS* resolution maximum truncated to ensure a minimum CC $\frac{1}{2}$ scores >0.30

‡ There are two SeMet peak datasets that both warrant representation as the SeMet peak (initial) has significantly higher anomalous resolution limit, however the SeMet peak dataset (optimised) has overall greatest diffraction quality statistics compared to all 6H-QseM diffraction data. Anomalous resolution determined by *AIMLESS* as a maximum before CC $\frac{1}{2}$ scores >0.30 for anomalous signal.

4.2.7 Space Group Assignment

Almost all 6H-QseM crystals were automatically determined to be, and subsequently indexed as tetragonal with unit cell dimensions $a=36 \text{ \AA}$, $b=36 \text{ \AA}$, $c=119 \text{ \AA}$. Scaling tests in all tetragonal Laue groups by *POINTLESS* demonstrated that the crystals belong to group 4/mmm (Table 4.5). The candidate space groups in this Laue group are $P 4 2 2$, $P 4_1 2 2$, $P 4_2 2 2$, $P 4_3 2 2$, $P 4 2_1 2$, $P 4_1 2_1 2$, $P 4_2 2_1 2$ and $P 4_3 2_1 2$, and *MATTHEWS* (149) analysis indicates a single 6H-QseM molecule at a solvent content of 32.6%. Systematic absences indicated a two-fold screw (2_1 – only $k00$ reflections with $k=2n$ are present) on the short axis, and a four-fold screw (4_1 or 4_3 – only $h00$ reflection with $h=4n$ are present) on the long axis (Figure 4.10). The native and initial SeMet peak datasets had distinct systematic absences, however, both SeMet-optimised peak and inflection point datasets had weaker structure factor intensities and irregular systematic absences which coincided with reduced probability scores for space group determination for these datasets (Table 4.5). All datasets processed through *POINTLESS* indicated that the space group of 6H-QseM crystals was either $P 4_1 2_1 2$ or $P 4_3 2_1 2$ (average probability >71 %) compared to the next possible space groups (average probability <23 %).

Table 4.5: Laue group and space group determination for top datasets (from *POINTLESS*).

	Laue Group	Likelihood	Reindex operator	Space group	Prob.	Prob. (Sys. Abs.)
Native	P 4/m m	0.994	[h, k, l]	P 41 21 2	0.721	0.725
	m	0.004	[h+k, -h+k, l]	P 43 21 2	0.721	0.725
	C m m m	0.002	[h, k, l]	P 41 2 2	0.191	0.192
	P m m m			P 43 2 2	0.191	0.192
				P 42 21 2	0.038	0.038
				P 4 21 2	0.027	0.028
				P 42 2 2	0.010	0.010
				P 4 2 2	0.007	0.007
Se Peak (initial)	P 4/m m	0.998	[h, k, l]	P 41 21 2	0.809	0.811
	m	0.001	[h, k, l]	P 43 21 2	0.809	0.811
	P m m m			P 41 2 2	0.133	0.133
				P 43 2 2	0.133	0.133
				P 42 21 2	0.027	0.027
				P 4 21 2	0.022	0.022
Se Peak (optimised)	P 4/m m	0.960	[h, k, l]	P 4 21 2	0.632	0.659
	m	0.020	[h, k, l]	P 4 2 2	0.311	0.324
	P m m m	0.005	[h+k, -h+k, l]	P 42 21 2	0.011	0.012
	C m m m	0.005	[h-k, h+k, l]			
	C 1/2m 1	0.004	[h, k, l]			
	P 4/m	0.003	[-k, -h, -l]			
	P 1 2/m 1	0.002	[h, k, l]			
	P 1 2/m 1	0.001	[k, l, h]			
Se Inflection point	C 1 2/m 1	0.591	[h, k, l]	P 4 21 2	0.665	0.665
	P -1	0.234	$[-\frac{1}{2}h-\frac{1}{2}k,$	P 4 2 2	0.268	0.268
	P 1 2/m 1	0.045	$\frac{1}{2}h+\frac{1}{2}k, -l]$	P 42 41 2	0.042	0.042
	P 1 2/m 1	0.041	$[-\frac{1}{2}h-\frac{1}{2}k,$	P 42 2 2	0.017	0.017
	C 1 2/m 1	0.041	$\frac{1}{2}h+\frac{1}{2}k, -l]$			
	P 1 2/m 1	0.023	$[-\frac{1}{2}h-\frac{1}{2}k, -l, \frac{1}{2}h-$			
	C m m m	0.018	$\frac{1}{2}k]$			
	P 4/m	0.005	[-k, -h, -l]			
	P m m m	0.001	$[-\frac{1}{2}h-\frac{1}{2}k,$			
			$\frac{1}{2}h+\frac{1}{2}k, l]$			
			[h, k, l]			

Accepted laue/space group is in bold. Likelihood, probability (Prob.) and probability adjusted with observed systematic absences (Prob. Abs.) are scored between 1 and 0. In cases of equal probability, multiple space groups are accepted and processed in parallel in *AIMLESS*.

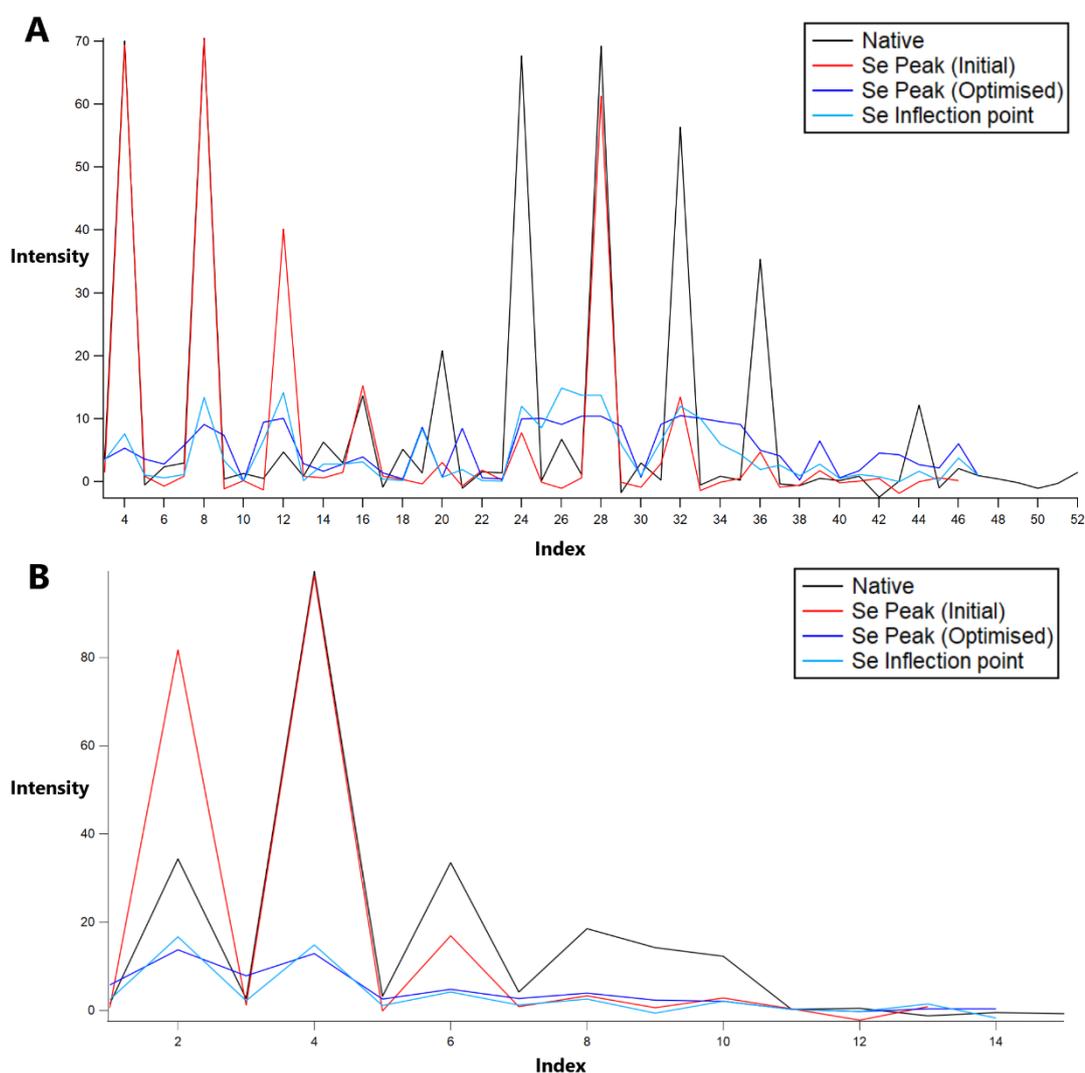


Figure 4.10: Intensity vs index for systematic absence determination. Strong intensities of both Native and SeMet Peak (Initial) datasets indicate that $h=4$ (A) and $k=2$ (B). SeMet Peak (Optimised) and SeMet Inflection point are significantly weaker yet appear to fit this trend.

On a number of occasions, the lower symmetry (C-centred monoclinic) space group $C 2$ was identified (formally, $C 1 2 1$) with the cell dimensions of $a=52 \text{ \AA}$, $b=52 \text{ \AA}$, $c=119 \text{ \AA}$ ($\alpha=90^\circ$, $\beta=90^\circ$, $\gamma=90^\circ$) instead of the previously described tetragonal ($a=36 \text{ \AA}$, $b=36 \text{ \AA}$, $c=119 \text{ \AA}$). These two space groups are mathematically related. The primary tetragonal space group dimensions of $a = 36 \text{ \AA}$, $b = 36 \text{ \AA}$ has a hypotenuse of $\sim 52 \text{ \AA}$, corresponding to the a or b length of the monoclinic space group (Figure 4.11). This leaves the c dimension

unchanged, which corresponds to what we observed between both tetragonal and monoclinic unit cell parameters. It is possible that poor distinction in systematic absences, coupled with diffraction spot doubling described in 4.2.6 may have given rise to datasets being indexed as C-centred monoclinic space group. We were able to process all monoclinic datasets in $P 4_{1/3} 2_1 2$ on the presumption that the $C 2$ indexing is a 'red herring' due to poorer quality data, but also retained scaled and truncated data in $C 2$ in case the lower symmetry could be real. These lower symmetry indexed datasets were used in molecular replacement (method 2.7.3) as described in 5.2.7.

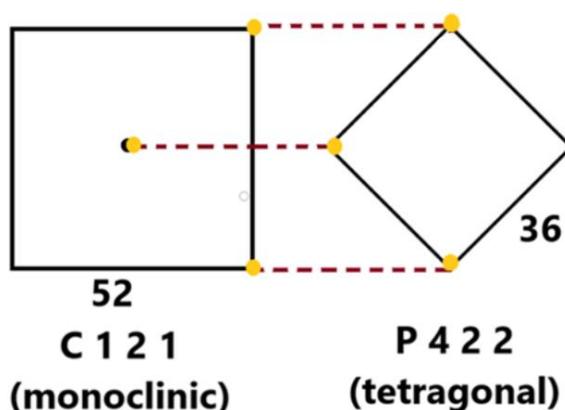


Figure 4.11: Rotational relationship between the unit cell dimensions of 6H-QseM diffraction data index results. Lattice points exist on the four corners of the tetragonal space group and the four corners and centre of the monoclinic space group (in 2D space). Three corner lattice points (highlighted yellow) of the tetragonal (unit cell length of 36 Å) and two corner and centre lattice points of the monoclinic (unit cell length 52 Å) space groups align when they are super-imposed on top of each other (visualised by dashed lines).

4.2.8 Attempts to determine phase information

We could not successfully obtain phase with the best datasets available using either anomalous dispersion (*SHELX C/D/E* or *PHASER*) pipelines. *PHASER-EP* pipeline uses *SHELXD* for heavy atom location prior to phasing, followed by *PARROT* (227) (density modification) and *BUCCANNEER* (149, 228) (model building). Both programs had acceptable occupancy scores (values

>0.5) for the expected number of Se atoms per asymmetric unit. We expect all 6 of the 6 methionine residues present in the 6H-QseM protein to be substituted with selenomethionine. A low CC (correlation coefficient) score of <0.3 when determining heavy atom substructure in *SHELX D* and a low contrast (<0.4) and connectivity (<0.75) during phasing and density modification in *SHELX E* after 100 cycles for both steps was observed for all attempts to solve the phase problem. No significant variation of CC and contrast and connectivity scores were seen for either P4₁ or P4₃ enantiomorphs which, with the overall low scores, suggested that no conclusive solution could be determined via *SHELX C/D/E*. Quality of success for *PHASER* was measured by figure of merit and R-factor scores (153). The overall phasing results were poor, with the best Se peak data set having a low figure of merit (total) of 0.279 and high R-factor of 74.3. Consequently, the density modification and model building steps that followed were considered unsuccessful. This failure to phase correctly was likely attributed to the low completion of the final model set generated via *BUCCANEER*, having only a 18.1% completeness of the total 6H-QseM amino acid sequence. This completeness was also heavily fragmented, existing in 3 disconnected chains.

4.2.9 Attempts to solve structure using MR

Models were used with collected crystallographic data to attempt structure solution by molecular replacement. Best SeMet peak and native crystallographic data were used (Table 4.3) with models generated via NMR or *ab initio* (method 2.7.3, ROSETTA/ROBETTA) predictive software in *PHASER-MR*. The best result from *PHASER-MR* determined 16 possible solutions with the top scoring solution having a TFZ score of 6.5 (below the default '8.0' value considered successful (229)). Attempts to refine solutions using *REFMAC5* (230) did not improve R-factors (231), with R_{work} and R_{free} (232) remaining >0.45.

4.3 Conclusion:

Crystallisation was optimised from the original crystallographic results of 6H-QseM with best diffraction to $< 3 \text{ \AA}$ resolution. Optimisation of crystal growth time, pH and PEG precipitant, micro-seeding and the cryoprotectant used each appeared critical for improvement of diffraction data. Crystal morphology and unit cell dimensions appeared consistent between multiple crystallisation experiments of unlabelled and Se-labelled 6H-QseM. The space group was determined to be a tetragonal $P 4 2_{1/3} 2_1 2$, although some weak evidence suggested that it may be a lower symmetry monoclinic $C 2$ space group.

We were unable to solve the phase problem using SeMet-based anomalous phasing. While the morphology of the crystals seemed to contain more than one lattice, diffraction data analysis displayed no evidence of twinning. Nevertheless, a combination of ambiguity in space group determination and multiple crystals may have caused some problems yet to be fully characterised. The anomalous signal measured in the diffraction data was not strong for any of the datasets, raising the possibility of either problems with twinning or low selenium incorporation into crystals. The level of SeMet incorporation was not explicitly measured due to time constraints, however, anomalous diffraction during X-ray diffraction at the absorption wavelength of Se was consistently detected.

Molecular replacement was also attempted using predicted or NMR-derived 6H-QseM structure. This was also unsuccessful, suggesting that either there are unresolved issues in the diffraction data or that the structures are not close enough to the crystal structure for successful molecular replacement. This latter situation is common and often requires refinement of the structure before molecular replacement is successful (233). This refinement primarily includes more realistic forcefield parameters that account for solvent that is interacting with the protein, something that is not well accounted for in NMR but integral to its structure within a crystal lattice. Future experiments should focus on finding new crystallisation conditions, crystallising the untagged QseM or crystallising QseM homologues.

Attempts to crystallise 6H-MBP-TEV-FseA were unsuccessful and no diffraction data was collected. Due to the stabilising nature of the MBP-tag, protein purification was improved but crystallisation may have been potentially hindered. It is possible that the MBP-tagged FseA construct is inherently difficult to crystallise and may need to be removed or substituted with a different tag before crystallisation is successful, an observation found in other recombinant proteins used in crystallisation (234).

Chapter 5:
QseM Structure
Determination by NMR

5.1 Introduction:

This chapter contains work that was completed with assistance of Dr Mark Howard (School of Chemistry, University of Leeds, UK) and Dr Gareth Nealon (Centre for Microscopy, Characterisation & Analysis [CMCA], University of Western Australia) using a Bruker AVANCE IIHD 600MHz spectrometer at UWA CMCA. Further work was completed with assistance of Prof. Joel Mackay and Paul Solomon (School of Life and Environmental Sciences, University of Sydney) on a Bruker AVANCE III 800 MHz spectrometer.

Nuclear Magnetic Resonance (NMR) spectroscopy can be used to determine structural information from macromolecules. Over 13,000 protein structures solved by NMR have been deposited to date in the RCSB PDB (rcsb.org (235)). Data acquisition consists of 1-, 2- and 3-dimensional NMR experiments on a homogenous protein sample, most often in solution. Proteins amenable to solution NMR structure determination are generally less than 20 kDa, stable and soluble at ≥ 0.3 mM concentration at temperatures between 4 and 40 °C. Proteins need to be stable in a low (< 50 mM) salt buffer solution free of organic agents like tris, HEPES and glycerol. Experimental run times can exceed 24 h of continuous signal acquisition for multidimensional NMR experiments (236, 237). These factors limit the range of proteins that can be structurally determined by NMR spectroscopy (238, 239).

The signals observed in multi-dimensional NMR experiments arise from either through-bond (240) or through-space (nuclear Overhauser, or NOE) interactions (241). Through-bond NMR experiments detect interactions between atoms that are 'coupled' through chemical bonds and NOE experiments detect interactions between atoms that are coupled due to their proximity in the three-dimensional structure of the folded protein. Through-bond NMR experiments are used for backbone and sidechain resonance assignment and the NOE based experiments are critical to assigning distance restraints between the assigned atoms. These distance restraints are necessary to determine a protein's folded tertiary structure (242, 243). NMR experiments can be further categorised by the number of atom types used and how they are coupled (Table 5.1). The simplest experiments are 1D, where a single atom type signal is collected, often ^1H . Two-dimensional experiments

collect data connecting two atoms of either the same (homonuclear) or different (heteronuclear) types. These include ^1H -TOCSY/COSY/NOESY and ^{13}C - or ^{15}N -separated NOESY and HSQC and less common, HBCBCGCDHD and HBCBCGCDCEHE. Three-dimensional experiments (spectra collected with three measured frequencies) include more specific resonance couplings for experiments like HNCACB, CBCA(CO)NH, HNCO, HBHA(CO)NH. The majority of backbone assignments are made using the HNCACB, CBCA(CO)NH and HNCO experiments. Sidechain assignments are made with HCCH-TOCSY/COSY, HBHA(CO)NH, HBCBCGCDHD, HBCBCGCDCEHE, ^{15}N - and ^{13}C -NOESY. Successfully assigning the majority of atoms, particularly those in hydrophobic sidechains that form the core of a protein, is required to attempt structure determination.

Table 5.1: NMR experiments and usage

Atoms	Dimensions	Name	Use
^1H	1	1D ^1H spectrum	Simple analysis of all protons in sample, sample quality
^1H	2	^1H -TOCSY/COSY	Through bond, give signals for H atoms within 2-3 (COSY) or more (TOCSY) bonds of each other
^1H	2	^1H -NOESY	Through space, gives signal for H atoms within a few Å of proximity to each other
^1H , ^{15}N	2	^{15}N -HSQC	Gives a signal for H-N pairs, for backbone or sidechain H and N assignments, used with ^{15}N -NOESY
^1H , ^{13}C	2	^{13}C -HSQC	Gives a signal for H-C pairs for backbone/sidechain H and C assignments, used with ^{13}C -NOESY
^1H , ^{13}C	2	HCCH-TOCSY/COSY	Gives a signal for H-C-C-H atoms either 2-3 (COSY) or more (TOCSY) bonds, used for sidechain H and C assignments
^1H , ^{13}C	2	HBCBCGCDHD, HBCBCGCDCEHE	Gives a signal for the HB-HD or HB-HE pairs, used for long sidechain H assignments, linking aromatic groups to backbone
^1H , ^{15}N , ^{13}C	3	HNCACB, CBCA(CO)NH, HNCO, HBHA(CO)NH	Gives a signal for the H, C and N atoms across the backbone and start of the sidechain of a residue, used for backbone assignment and linking adjacent residues.
^1H , ^1H , ^{15}N	3	^{15}N -NOESY	Gives a signal for protons in proximity, can be linked back to all assigned N atoms, used for sidechain assignment and distance restraints
^1H , ^1H , ^{13}C	3	^{13}C -NOESY	Gives a signal for proton in proximity, can be linked back to all assigned C atoms, used for sidechain assignment and distance restraints

A chemical shift list can be used to predict the backbone ϕ and ψ dihedral angles. Full structure calculations use both these dihedral angle predictions and through-space NOE assignments, from which distance restraints are derived. Other information, including paramagnetic relaxation enhancement data (244) and residual dipolar couplings (245), are also used in some cases. Determining distance restraints and/or generating a protein structure can be automated using software such as *CYANA* and *CNS*. Structure calculations are typically carried out multiple times and the structure taken to be the ensemble of the ~20 conformers with the lowest energy as calculated by the program. A structure solution is deemed successful when there is consistent convergence between different structures, reflected as a minimal root-mean-squared-deviation (RMSD) score (180, 246, 247).

In principle, well-determined NMR structures can be used in molecular replacement for phasing-solving crystallographic diffraction data. However, since most NMR data lack long-range constraints, NMR structures are often poor templates for molecular replacement. Significant deviation between the NMR structure and the crystallographic data can render the template incompatible for MR structure calculation. Recent improvements in molecular dynamics-based refinement of NMR structures have given rise to higher success rates in MR (248). Given the difficulties in phasing the QseM crystallographic data described in the previous chapter, a QseM NMR structure refined by such molecular-dynamics methods may prove useful for MR structure determination.

NMR data can also be used to supplement protein prediction methods if distance restraints are incomplete or unobtainable. Chemical shift data without NOE information can be used in *CS-ROSETTA* to generate a structure with secondary structure determined using experimental data. The output can also be a rough validation model to compare with the full NMR calculated ensemble. Convergence between both would suggest the structure is likely to be correct.

In this chapter, multidimensional NMR experiment data were collected on 6H-QseM. Following atom assignment, structure calculations resulted in the first experimentally derived structure of 6H-QseM.

5.2 Results and Discussion:

Protein purification was discussed in chapter 3 and 6H-QseM was prepared for NMR experiments by methods outlined in 2.8.1. Details of NMR experiments are listed in Table 2.15 and Table 2.16.

5.2.1 Testing QseM suitability for NMR

1D ^1H experiments were collected as described in 2.8.2. The 1D spectrum was similar at 50 μM and 500 μM concentrations, suggesting that QseM does not undergo concentration dependent aggregation or denaturation (249) during data collection (Figure 5.1).

For ^1H -TOCSY/COSY/NOESY, and ^{13}C -NOESY, experiments 6H-QseM was suspended in buffer made up in >99% H_2O or >99% D_2O . Transfer to D_2O was achieved via freeze-drying/resuspension methods outlined in 2.8.1. Freeze drying and resuspending appeared to have little impact on the integrity of the protein as comparisons of spectra before and after freeze-drying revealed few differences (Figure 5.2). Minor changes in the spectra were likely associated with slight differences between buffers; however, it is also possible the freeze-dried protein solution was more susceptible to aggregation.

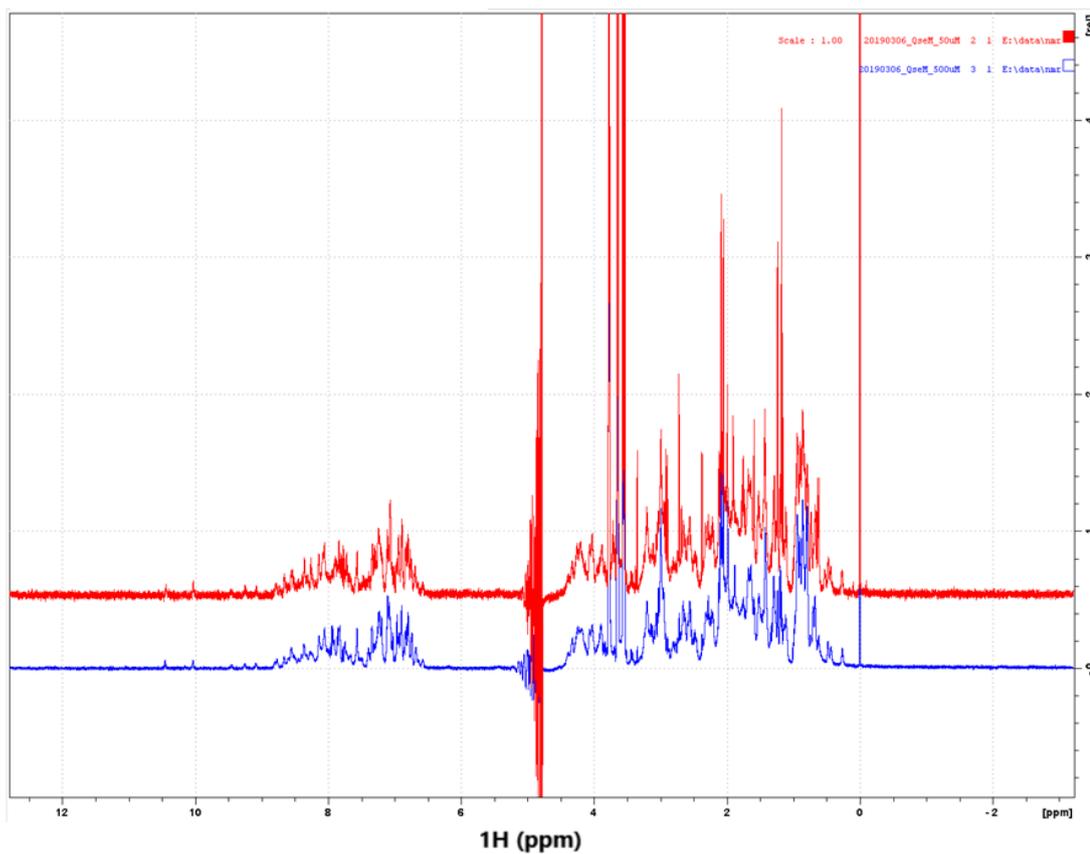


Figure 5.1: 1D ¹H NMR analysis of 6H-QseM. Spectra of 6H-QseM at two concentrations, namely 500 μM (~5 mg/ml) (blue) and 50 μM (red), were recorded. The spectra were referenced using sodium trimethylsilypropanesulfonate (DSS) (set at 0 ppm).

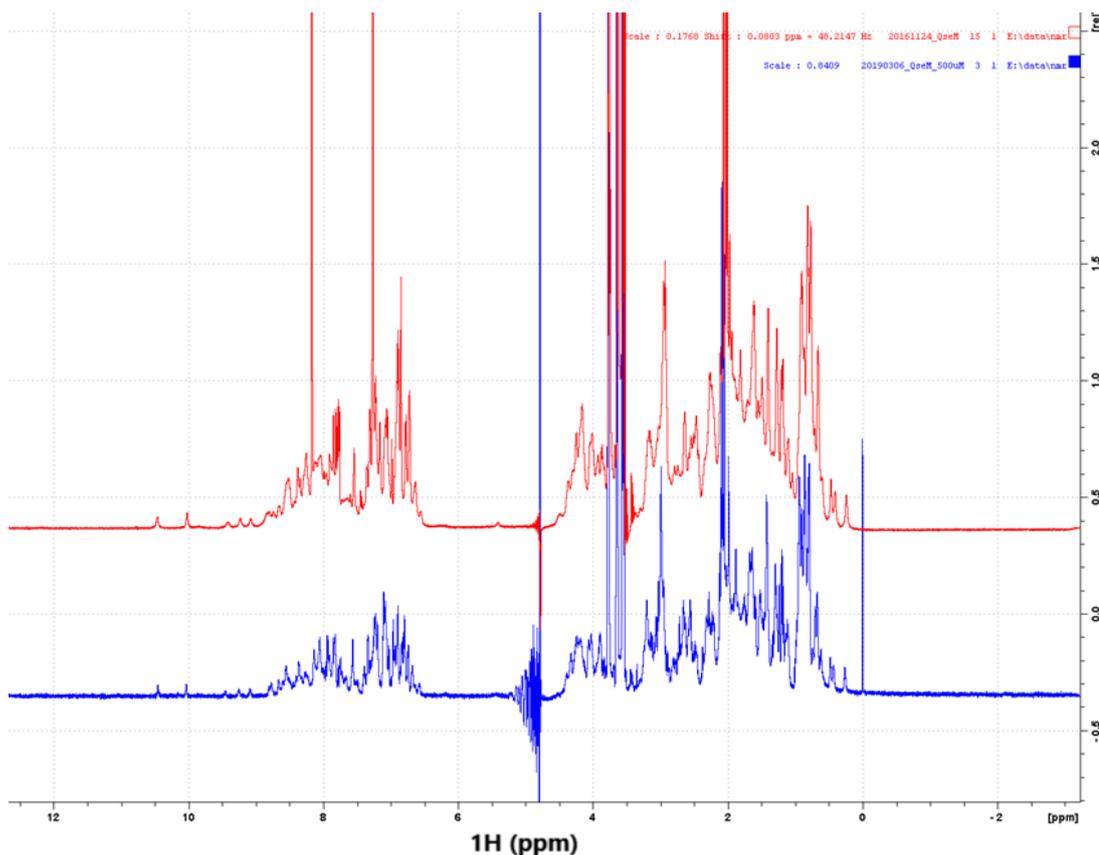


Figure 5.2: 1D ^1H spectra of 6H-QseM before and after freeze-drying. Before (red) and after (blue) freeze-drying. Sample concentration 500 μM .

The 6H-QseM sample was tested for consistent behaviour between experimental data collection by running ^{15}N -HSQC spectra periodically. Consistent HSQC profiles indicated the sample did not change in any significant way between experiments. Additional signals did begin to appear in the HSQC profile after 88 h in the spectrometer (Figure 5.3). The 88 h mark was the timeframe in which the majority of 3 dimensional experiments needed for atom assignment were run (excluding ^{13}C NOESY and HSQC). Several sample replicates were used to run all NMR experiments within similar timeframes. These signals were in positions that are consistent with either degradation or unfolding of the protein. A protein sample that has significant degradation or precipitation often has all amide signals condensed into a narrow chemical shift range (~ 7.8 - 8.6 ppm). Some 6H-QseM samples had visible precipitation before NMR data were collected; this precipitation had a negative effect on data collection. For these samples, a ^{15}N -HSQC (Figure 5.4)

would reveal a severely condensed spectrum and no further NMR data were collected.

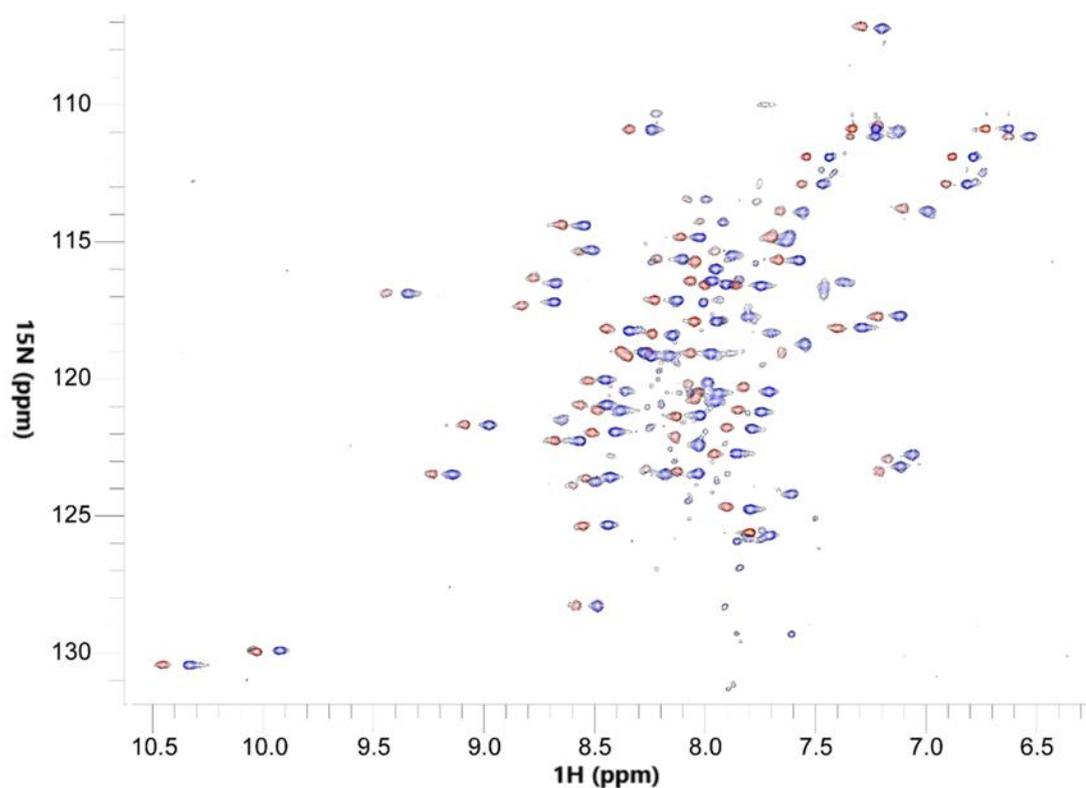


Figure 5.3: ^{15}N -HSQC of 6H-QseM over time. ^{15}N -HSQC of the same 200 μM 6H-QseM sample taken as soon as sample was loaded into the spectrometer (red) and 88 h later (blue). The +88 h ^{15}N -HSQC peaks were manually shifted 0.1 ppm lower along the ^1H axis for clarity. The ^{15}N -HSQC profile remains consistent; however, weak peaks begin to emerge in the ^1H 7.5-8.5 ppm range. These weak peaks may indicate sample degradation is increasing and may suggest that 88 h is near the limit of sample lifespan for NMR data collection.

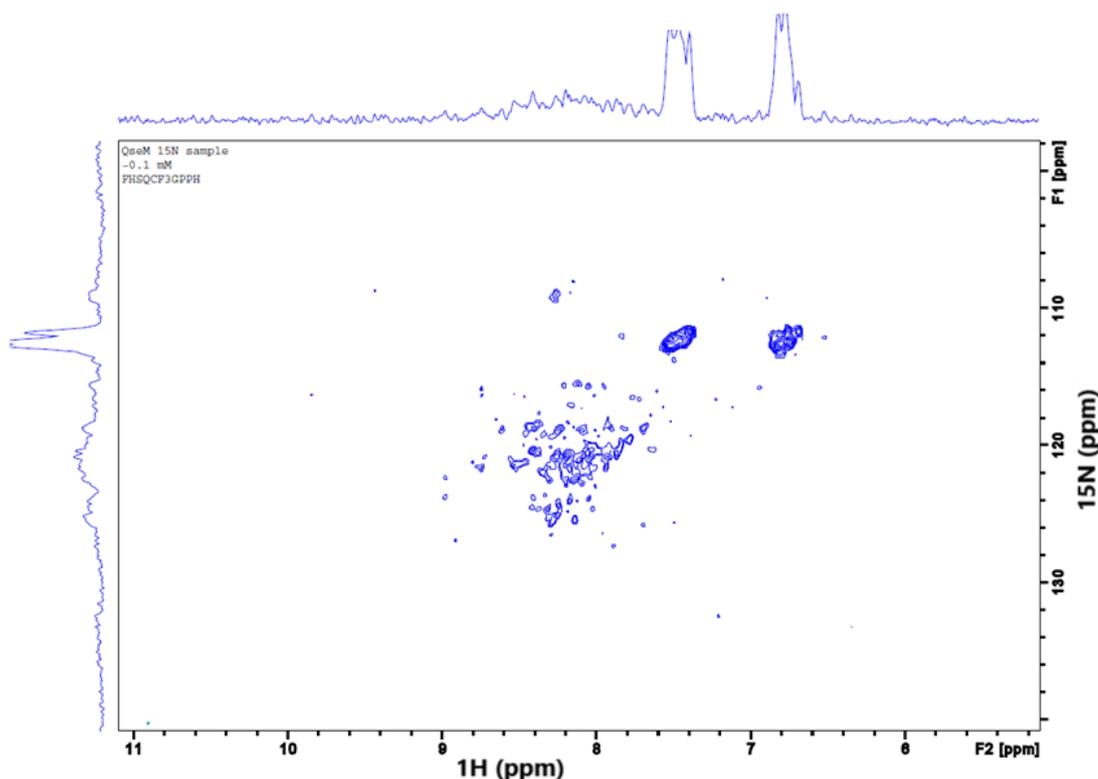


Figure 5.4: ^{15}N -HSQC of precipitated 6H-QseM. Some protein samples prepared for NMR data collection showed visible precipitation before any NMR data were collected. ^{15}N -HSQC spectra of such samples showed no peak separation, suggesting aggregation and/or unfolding.

5.2.2 Atom assignment

The ^{15}N -HSQC displayed a single peak for most residues in the 94 amino acid sequence of 6H-QseM. Note that the first 12 amino acids of the 6H-QseM construct compose the N-termini 6H-tag, the remaining 13-94 are identical to the 2-83 amino acids of native QseM. The HNCACB experiment gave signals for all $\text{C}\alpha$ and $\text{C}\beta$ atoms linked to the N-H of the same residue. The HNCACB experiment also gave progressively weaker $\text{C}\alpha$ and $\text{C}\beta$ signals for many previous residues linked to this N-H. The CBCA(CO)NH experiment gave signals for all $\text{C}\alpha$ and $\text{C}\beta$ atoms of the residue preceding the residue for which the signal was observed in the ^{15}N -HSQC. These two experiments allowed identification of the $\text{C}\alpha$ and $\text{C}\beta$ chemical shifts of each backbone amide group found in the ^{15}N -HSQC as well as the $\text{C}\alpha$ and $\text{C}\beta$ chemical shift positions of the preceding residue allowing the assigner to link each peak in

the ^{15}N -HSQC to the previous one (Figure 5.5). Some residues, like alanine, serine and threonine have distinct $\text{C}\alpha$ and $\text{C}\beta$ chemical shifts (Figure 5.6) that can be used as starting points for sequence-specific assignment. When a unique sequence has been determined, all atoms in that sequence can be assigned with a high degree of confidence.

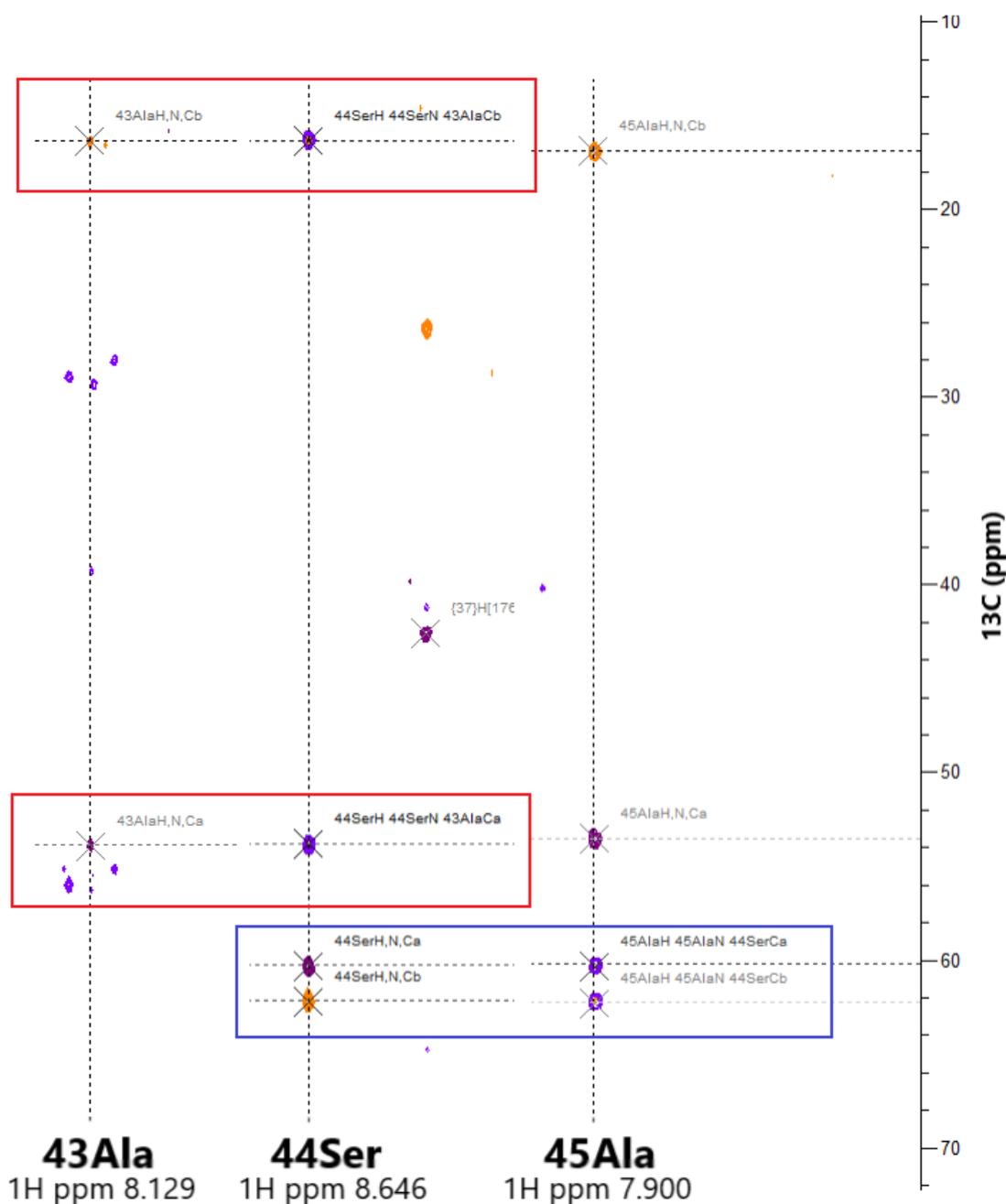


Figure 5.5: Backbone assignment method. The C α and C β of 43Ala has a chemical shift pair linked with the 44Ser which in turn has a chemical shift link to the 45Ala. The 1H ppm shift position for these residues is given. Since the sequence 'ASA' only occurs once in the 6H-QseM sequence, we could assign these atoms with confidence.

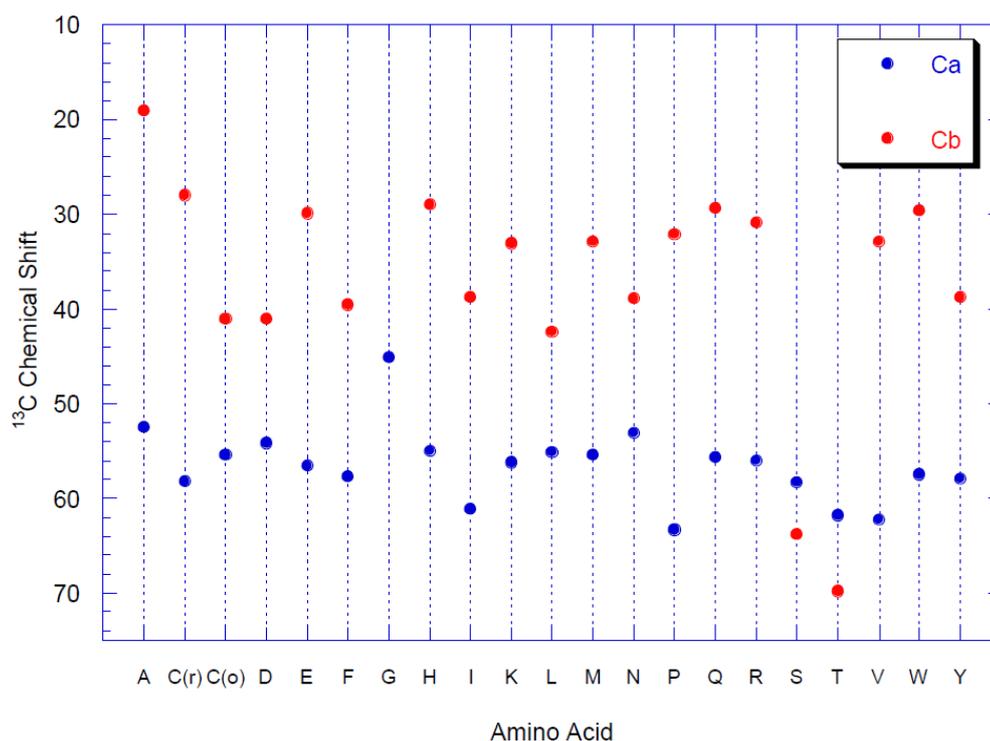


Figure 5.6: The typical C α and C β chemical shifts of each amino acid. Note that with serine and threonine, C α and C β chemical shifts are inverted, and alanine has a significantly lower C β chemical shift than all other residues. The remaining residues can be loosely distinguished into either ‘low’ (around 30 ppm) and ‘high’ (around 40 ppm).

Assignment completeness was mapped out on the corresponding ^{15}N -HSQC (Figure 5.7). This method of backbone atom assignment was used to assign 67 of the 94 residues of the 6H-QseM sequence (Figure 5.8) and was completed with assistance from Dr Mark Howard and Dr Gareth Nealon. We excluded the first N-terminal 10 residues that correspond to the 6H-tag; signals from these residues were not observed in the ^{15}N -HSQC, likely due to high mobility. Assignment was completed for 67 residues out of the possible 84, or 80%. The remaining residues that could not be assigned included residues 11-15, 21, 23, 26, 62, 66, 84-90. The majority of unassignable residues were found in the 11-15 region and 84-90 region. These residues had extremely weak signals. It is possible the weak signals are the result of higher mobility of these residues. Residues 21, 62 and 66 were prolines and do not have a presence on the ^{15}N -HSQC.

<i>MRGSHHHHHH</i>	<i>GSKR</i> KVQDEV	PWSD <i>SLTAYD</i>	30
NEHFTIYMRL	LDASADDASE	DEMAQLVLGI	60
DPMRE <i>PERAR</i>	MAVRSHLDRA	NWM <i>VTTGYKE</i>	90
LFAR			94

Figure 5.8: Initial backbone assignment of 6H-QseM. Backbone assignment completed residues highlighted blue.

Further NMR experimental work was carried out with assistance of Prof. Joel Mackay and Paul Solomon at the University of Sydney. These experiments included a new backbone assignment in the hopes that the 800 MHz spectrometer may lead to higher assignment coverage. Only on the 800 MHz, double labelled sample suspended in D₂O was used to acquire ¹³C/¹⁵N NOESY and ¹³C HSQC experiments, essential to completing sidechain atom assignment necessary for generating distance restraints. These NMR data were collected to start backbone assignment as detailed above, with assignments mapped on the ¹⁵N-HSQC (Figure 5.9). Additional experiments included HCCH-TOCSY/COSY (Figure 5.10) and ¹³C-HSQC (Figure 5.11) in order to complete assignment of all atoms in the sidechain of each residue. These included carbon and proton atoms that extended past the C β ; in particular, the methyl groups that would generate the most useful through-space NOEs, which are key in generating a protein structure.

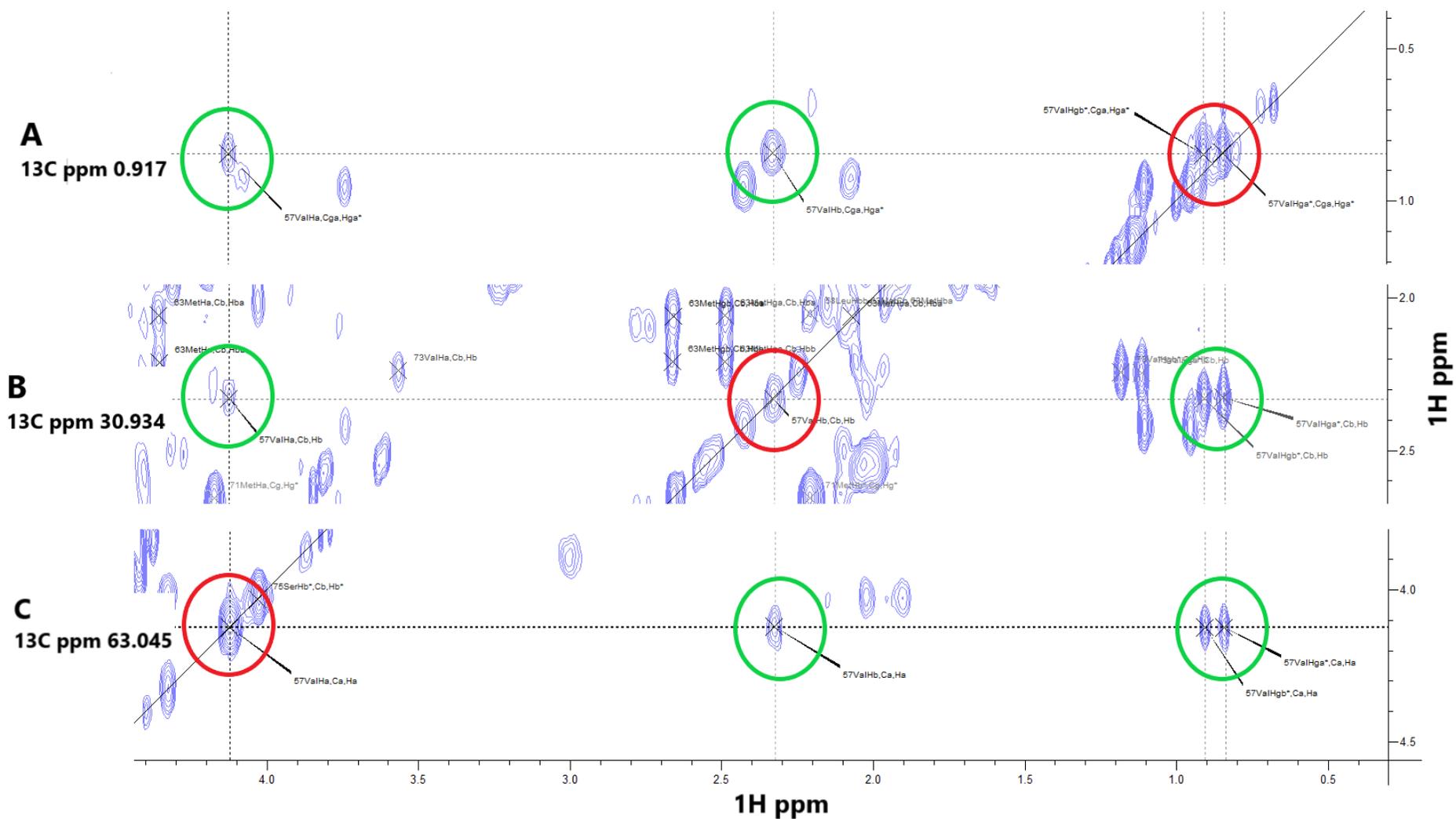


Figure 5.10: HCCH-TOCSY assignment for sidechain carbons and protons. Val57 has a C α (C), C β (B) and two C γ (A). An HCCH-TOCSY can show these as resonance sets. In each ^{13}C plane, a H-C diagonal peak shows cross peaks (connections) to other protons in the same sidechain (circled in red).

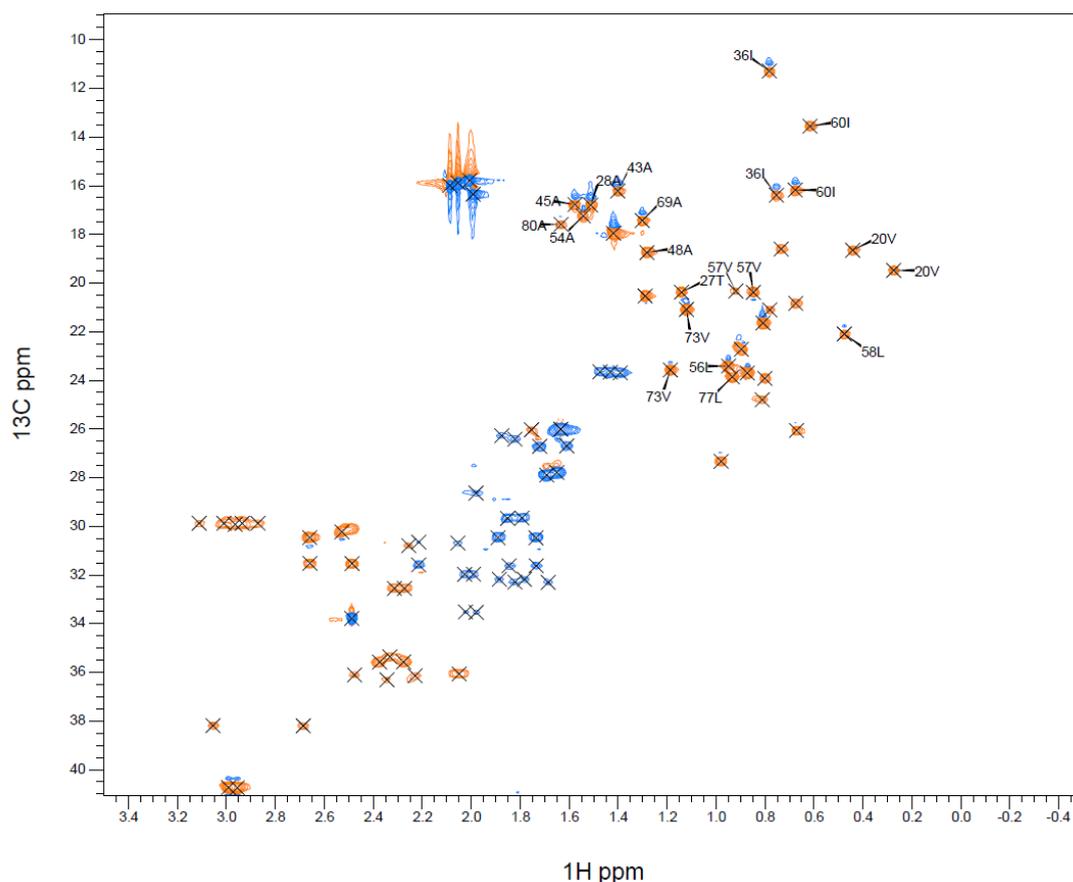


Figure 5.11: ^{13}C -HSQC of 6H-QseM. This ^{13}C -HSQC displays resonance peaks of C-H pairs. The majority of assignments displayed are the methyl groups that are critical when assigning through-space NOEs.

This second set of NMR experiments resulted in similar backbone assignment completion to the first but also included extensive proton assignment of those residues (Figure 5.12). In total, 63 residues (75% of total residues in 6H-QseM excluding the histidine tag) were backbone assigned with missing residues in similar areas to those of the initial assignment attempt shown in Figure 5.8. Of those residues, 50 (79%) had full sidechain proton assignment, 15 (24%) had most of the sidechain protons assigned and 4 (6%) had only a few sidechain protons assigned. These assignments were complete enough to begin structure generation via *CYANA*.

MRGSHHHHHH GSKRKVQDEV PWSDSL TAYD 30
 NEHFTIYMRL LDASADDASE DEMAQLVLGI 60
 DPMREPERAR MAVRSHLDRA NWMVTTGYKE 90
 LFAR 94

Figure 5.12: Backbone and sidechain assignment. Proton, carbon and nitrogen atoms that were fully assigned (highlighted blue), mostly assigned but with a few missing protons (highlighted green) and missing most protons (highlighted yellow).

The similar missing regions of assignment between two independently collected NMR data sets reinforces that these residues without assignment exhibit weak atom resonances, likely due to mobility.

Further assistance from Prof. Joel Mackay and Paul Solomon improved the assignment completeness to 88% of backbone atoms, 77% of sidechain atoms and 62% of atoms associated with aromatic rings in sidechains (Table 5.2).

Table 5.2: Chemical shift assignment completeness of 6H-QseM.

	Total atoms	¹ H	¹³ C	¹⁵ N
Backbone	291/329 (88%)	116/131 (89%)	118/134 (88%)	57/64 (89%)
Sidechain	308/437 (70%)	188/256 (73%)	116/159 (73%)	4/22 (18%)
Aromatic	39/63 (62%)	26/33 (79%)	11/26 (42%)	2/4 (50%)
Overall	625/814 (77%)	323/412 (78%)	241/314 (77%)	61/88 (69%)

5.2.3 NMR structure of 6H-Qsem

The *CYANA* software was used to generate structures using experimentally determined distance restraints as outlined in 2.8.6. *CYANA* by default has a total of 7 cycles which each output a set of 20 lowest-energy structures (those that best fit the experimental data and the basic forcefield that describes the geometry of the protein chain) that are then used to seed the next iteration. Significant convergence of structures in the ensemble was seen by the third iteration, at which point the three helices of 6H-QseM emerge. The remaining 4 iterations produced some minor improvement mostly regarding the N and C terminal ends of the protein becoming more converged (Figure 5.13). Convergence was significant for the single well-defined core (residues 18-83) with an average backbone (atoms C α , N, H) RMSD of 0.36 Å across this range. Convergence was also noted for sidechains within this core region, with lowest RMSD for core residues (Figure 5.14). Of interest were residues that are apparently surface exposed yet have low variance in sidechain orientation, potentially highlighting areas of the protein where structure is more tightly ordered (discussed in more detail in chapter 6). Comparatively, the extreme N- and C-terminal regions exhibited much greater variability. Significant disorder was apparent for residues 1-17 and 89-94 due to low structure convergence in these areas. Additionally, residues Arg39, Arg64 and Arg74 present in the 'well ordered' region had a >4 Å sidechain RMSD, however, this is unsurprising since arginine and lysine residues often are incompletely assigned due to inherent disorder.

A concentrated set of distance restraints was observed around the small, second helix (residues 50 – 59), suggesting stable interactions between adjacent helices (1 and 3). The N terminal Val20 and the C terminal Tyr88 appeared to be restrained in these otherwise disordered regions. Numerous identified distance constraints (Figure 5.15) around the Leu41 and Trp82 (which are in proximity to each other), suggested these residues are potentially critical to maintaining structure. There was an absence of distance constraints around residues Ala42, Ser43 and Met53 where we would expect multiple distance constraints to exist as they are clearly internalised and in proximity to similarly aliphatic residues like Val72 and Leu77. Inaccuracies or

incompleteness in resonance assignment might explain this lack of distance restraints, however there is also the possibility that this space may be more mobile or can accommodate a ligand.

The final NMR ensemble revealed 6H-QseM to have three alpha helices, H1, H2, and H3, with H1 and H2 in an antiparallel arrangement, the is H3 perpendicular to H1/H2. The N and C termini are at opposite polar ends of the overall protein structure. The structure is explored in significant detail in chapter 6, including residues associated with function, designation of a helix-turn-helix motif and comparisons with other structures.

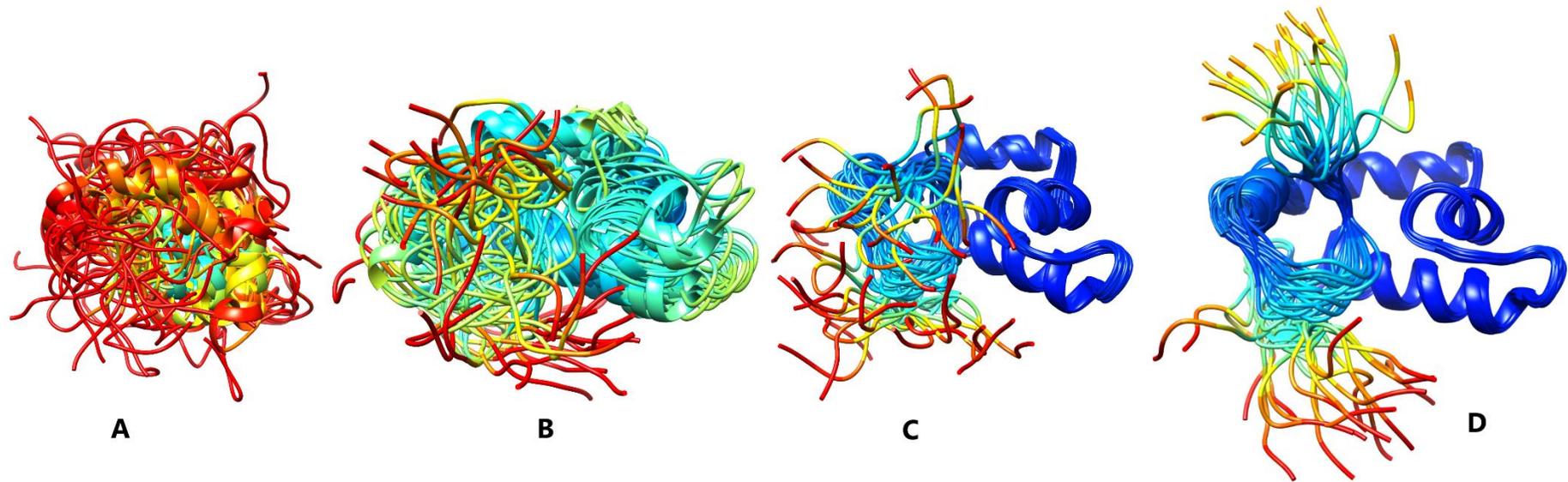


Figure 5.13: CYANA structure calculation. The first three iterations of ensemble generation are shown sequentially (A, B and C) followed by the last ensemble (D). Ensemble displays residues 11-94. Structure ensemble is coloured by backbone RMSD to chain that is most similar to the majority of structures in the ensemble (model of best representative), where 0-5 Å is dark blue to cyan, 5-10 Å is cyan to yellow, 10-15 Å is yellow to red, all >15 Å are red.

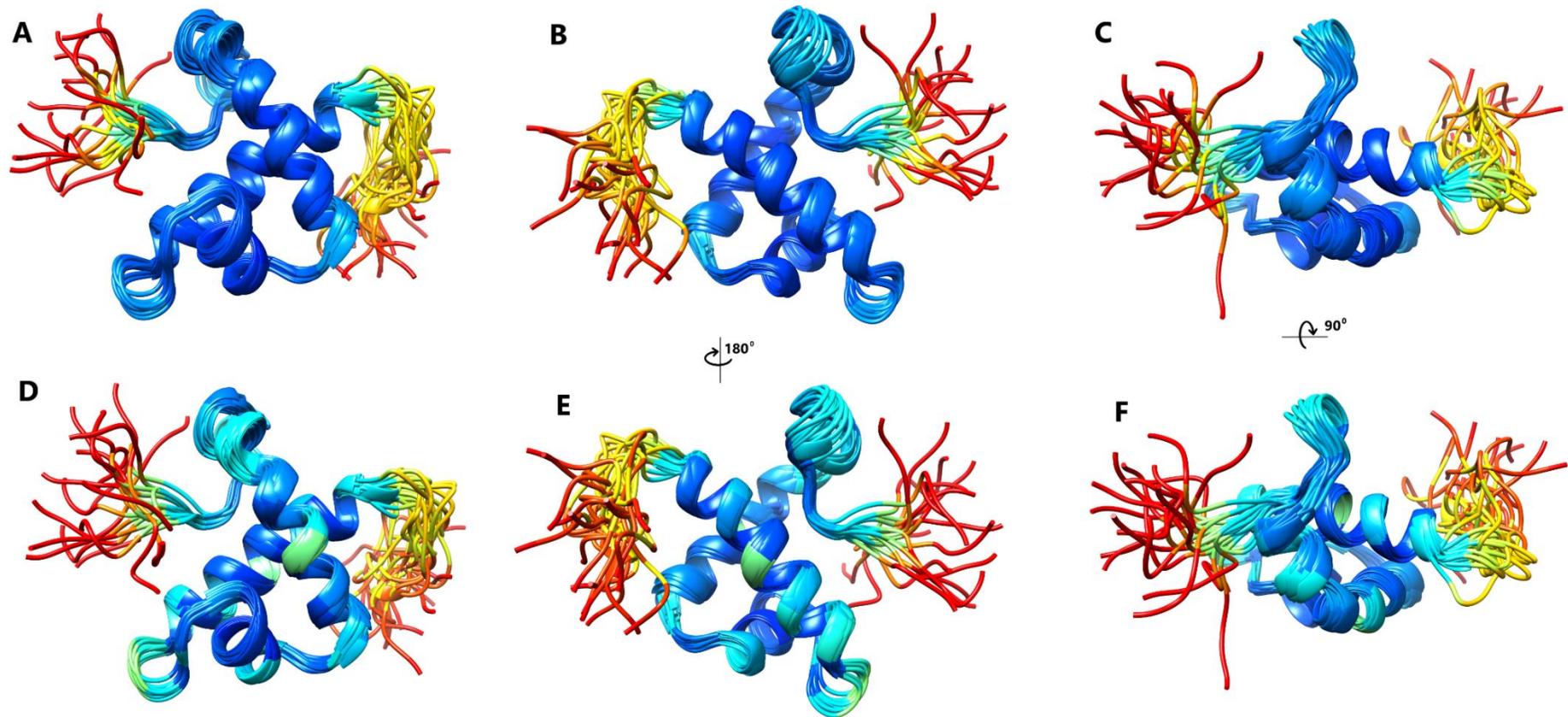


Figure 5.14: Final 6H-QseM NMR ensemble. Three orientations shown, front (A, D), back (B, E) and top (C, F) of the same ensemble (with an N to C terminal transition from left to right in the front orientation). Ensemble displays residues 11-94. Coloured by RMSD, from 0 - 2 Å (dark blue to cyan), 2 - 5 Å (cyan to yellow), 5 - 10+ Å (yellow to red) for backbone (A, B and C) and with sidechain (D, E and F).

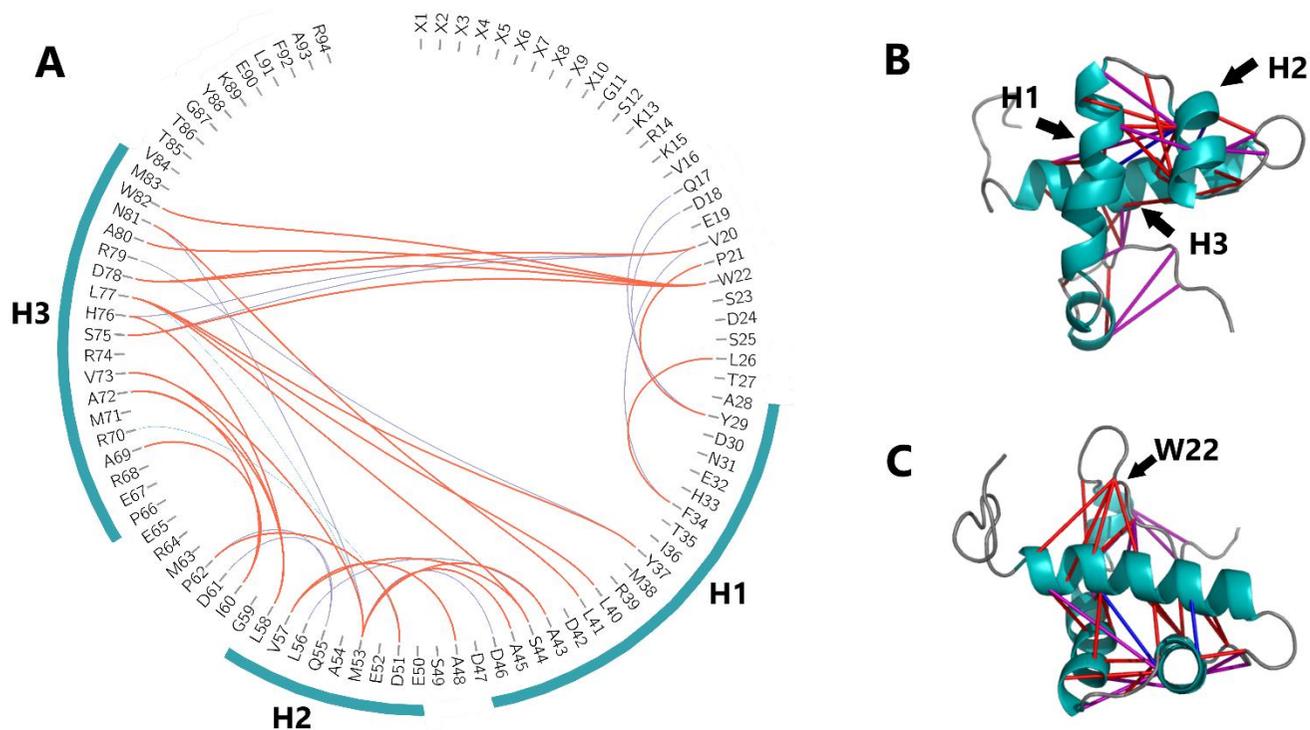


Figure 5.15: NMR structure of 6H-QseM with CYANA-derived distance restraints. Sequence of 6H-QseM displayed showing inter-residue contacts that are $>4.5 \text{ \AA}$ and more than 5 residue positions apart coloured based on distance restraint probability (0-1) as either >0.8 (red), $0.5-0.8$ (purple) or <0.5 (blue) (A) generated in *Circos* (250). The same restraints mapped on an NMR 6H-QseM structure from a 'front' (B) and 'top' (C) orientation. Helices coloured teal. Contacts worth noting include the residues 75-82 that have strong links to residues 20, 22, 37, 40-41, 44.

5.2.4 Preliminary structure validation report

A preliminary full wwPDB NMR structure validation report was created as a final analysis of the 6H-QseM structure. The 'Overall quality at a glance' figure supplied by the report presents the 6H-QseM structure as relatively good compared to other NMR-derived structures in the PDB (Figure 5.16).

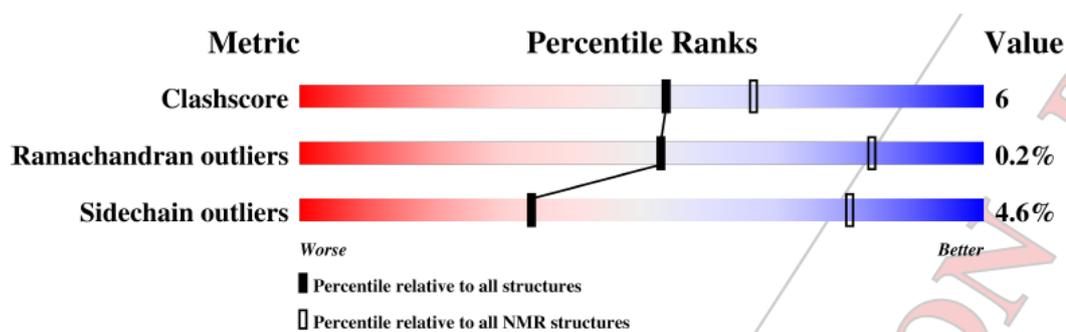


Figure 5.16: preliminary structure quality of 6H-QseM. 6H-QseM has reasonably low clash scores represented as the number of atoms that have steric conflicts. The Ramachandran outliers (torsion angles that are unusual, potentially incorrect) are represented as a residue percentage against total residues in structure. Sidechain outliers (sidechains that have unlikely conformations based on backbone torsion angles for that residue) are similarly represented as a percentage against the total number of residues in the structure.

The report further indicated a single ordered 'core' region between residues 18 – 88 (Figure 5.17). This ordered region had an average backbone RMSD of 0.36 Å between the medoid structure (most similar to other structures) and the remaining structures in the ensemble. This medoid structure was used as the 'most representative model' in further structure analysis work. More than half of the disordered N-terminal region flanking the 'core' (residues 1-17) included the 6H-tag (residues 1-10). This tag region was often omitted in further structure analysis.

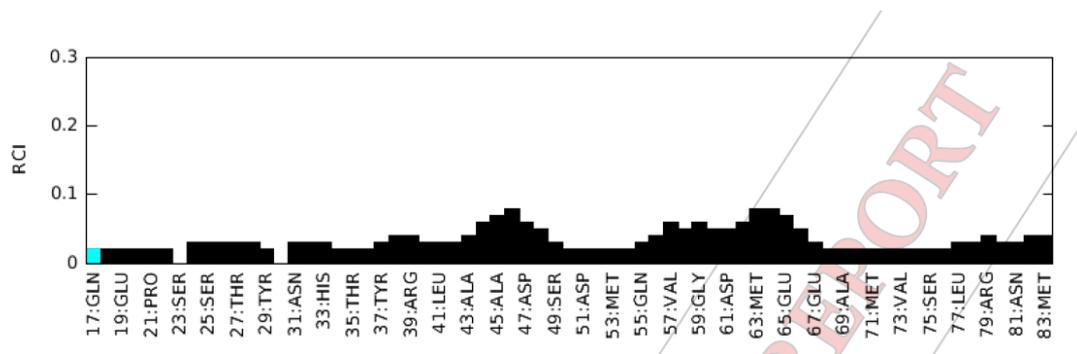


Figure 5.17: Random coil index plot (RCI) indicating ordered region of 6H-QseM. The RCI determines the probability that a residue is disordered based on its chemical shift information. Values above 0.2 indicated high probability of disorder. The 18-83 core region is shown to have a low RCI score. 17Gln is ill-defined due to limited chemical shift information.

5.2.5 Structures of 6H-QseM

Structure determination involved dihedral angles and secondary structure derived from methods outlined in 2.8.5. Two methods other than CYANA were used to generate a structure of 6H-QseM (Figure 5.18). One method employed *CS-ROSETTA* (<https://spin.niddk.nih.gov/bax/software/CSROSETTA/>) (156-159) with chemical shift information from NMR backbone assignments. The structure (6H-QseM-CSROSETTA) contained three main helices. A fourth helix was observed at the N-terminal, however, the majority of residues in it belong to the 6H tag and therefore can be seen as not part of the native protein structure.

The other method employed *GREMLIN2* (194-196) server (<https://gremlin2.bakerlab.org/>) to predict distance restraints based on co-evolution of residues in the DUF2285 protein family. These distance restraints were used by *ROBETTA* to generate a structure of 6H-QseM (6H-QseM-CoEvo) (work done by Calum Morris). This structure also consisted of three alpha helices.

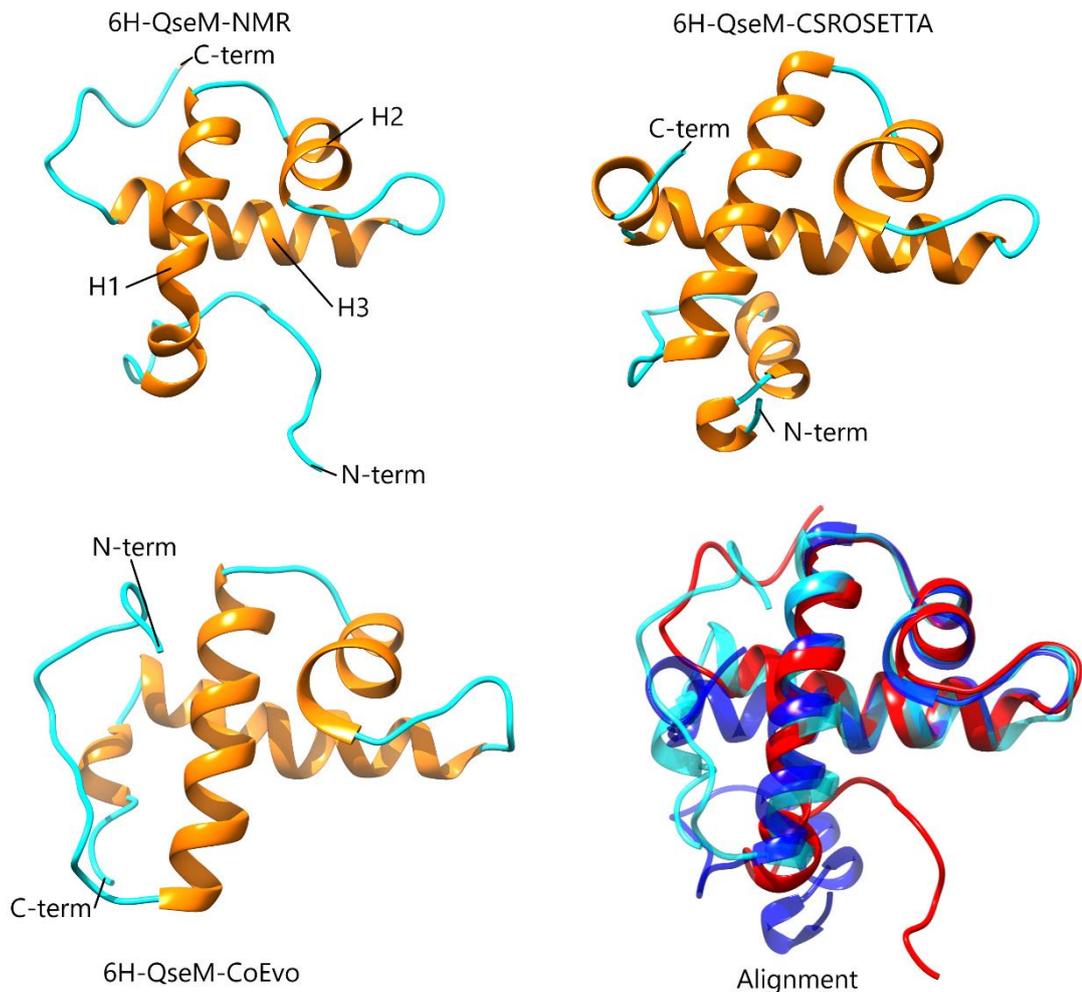


Figure 5.18: 6H-QseM structures generated. All three 6H-QseM structures show include ‘6H-QseM-NMR’ (best representative from *CYANA* ensemble), 6H-QseM-CSROSETTA (*CS-ROSETTA* predicted with NMR chemical shift data) and the 6H-QseM-CoEvo (*ROBETTA* predicted model using *GREMLIN* predicted distance restraints). The 6H-QseM-NMR structure has its three helices marked (H1, H2, H3) and its C and N terminal ends. All other structures have only the C and N terminal ends marked. The Alignment ensemble shows all three structures superimposed.

Structure alignment shows that the 6H-QseM-NMR, 6H-QseM-CSROSETTA and 6H-QseM-CoEvo structures share a consistent triple-helix core with deviations only becoming significant towards the C and N terminal regions. The middle of the first helix (H1) through to the middle of the third helix (H3) of 6H-QseM-NMR had a RMSD $<2\text{\AA}$ to either the 6H-QseM-CSROSETTA or the 6H-QseM-CoEvo. This observation is in line with the fact that predicted protein

structures often closely resemble the NMR or X-ray crystal derived structures of the same protein (156, 251). The overall alignment of all three structures also reinforces that 6H-QseM is a triple-helical protein with the first and second helix in an anti-parallel positioning, with the third helix running perpendicular to both. All three structures were used unsuccessfully with crystallographic data in attempts to solve the structure through molecular replacement, described in Chapter 4.

5.2.6 Refinement of NMR structure for molecular replacement

A successfully solved crystal structure of QseM may reveal the structure in finer detail than what has been achieved by NMR, particularly regarding the 'disordered' termini regions that have no definitive structural information. The NMR 6H-QseM structure was refined using *ROSETTA* (Figure 5.19), which includes force-field algorithms in *ROSETTA* to improve suitability for molecular replacement (248, 252). For refinement, the best representative model from the *CYANA* ensemble was used as a starting template together with NOE and dihedral restraints. Regions of high disorder were defined between residues 1-30 and 70-94. *ROSETTA* was given freedom to adjust any part of the structure, including designated disorder regions. The refinement resulted in an ensemble of 6H-QseM which had more rigid secondary structure compared to the pre-refined NMR structure input, particularly towards the N and C terminal ends where there was less curvature of the helices. The motif formed by the end of the first helix, second helix and start of the first helix had an RMSD < 2 Å compared to the pre-refined 6H-QseM-NMR, 6H-QseM-CSROSETTA and 6H-QseM-CoEvo structures.

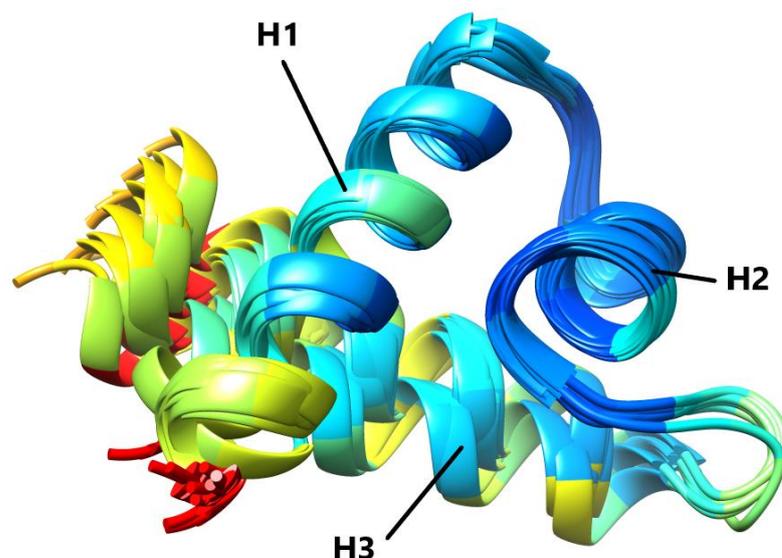


Figure 5.19: Ensemble of the top 10 structures generated from the *ROSETTA* after refinement. Coloured by average RMSD of backbone and sidechains, where 0 - 2 Å (dark blue – cyan), 2 - 4 Å (cyan – yellow) and 4 - 5 Å (yellow to red), with >5 Å remaining red. N and C terminal regions of high disorder have been removed for clarity. The three helices are shown as H1, H2 and H3 (from N to C terminal as first, second and third helix).

5.2.7 Supporting structural information

Small angle X-ray scattering (SAXS) analysis of 6H-QseM was carried out to determine mass, size and shape analysis of 6H-QseM in solution. SAXS data was used to collect low-resolution structural information of QseM in solution, parallel to crystallography structure determination attempts, before an NMR structure was generated. Although this SAXS model was made redundant after success with NMR, comparison between both reinforced that the 6H-QseM-NMR structure was accurate. This work was completed at the Australian synchrotron with assistance from Dr. Jason Schmidberger and Dr. Karina Yui Eto as per method 2.8.1. X-ray scattering, Guinier, Kratky and distance distribution plots were generated (Figure 5.20), followed by molecular weight calculation (Table 5.3) using *PRIMUSQT*. The scattering profile indicated that 6H-QseM is globular. The data from both Guinier and Kratky indicated that the overall quality of data is sufficient to gain an accurate

structural information. The D_{\max} (maximal radial distance of the protein) was ~ 56.94 Å and the mass was calculated to be 10.40 kDa, which is within reasonable range of the theoretical mass of 6H-QseM at 10.96 kDa (Table 2.11). The SAXS data conclusively proves that 6H-QseM exists in a monomeric, globular form.

Table 5.3: SAXS data collection and analysis for 6H-QseM.

Data collection	Instrument	AS	SAXS/WAXS
	q range (Å ⁻¹)	0.006 – 0.375	beamline
Structural parameters	P(r) I(0) (cm ⁻¹)	00.01 ± 0.00	
	P(r) Rg (Å)	15.57 ± 0.01	
	Guinier I(0) (cm ⁻¹)	00.01 ± 0.00	
	Guinier Rg (Å)	15.56 ± 0.01	
	Dmax (Å)	56.94	
	P(r) Quality estimate	0.97	
Molecular mass	Calculated MW (Da)*	10400	
	MW probability (%)	69.94	
	Credibility Interval (Da)	9050 – 10850	
	Credibility interval (%)	96.56	
	Theoretical MW (Da)‡	10960	
	Apparent stoichiometry (n)	monomeric	

* determined by Bayesian Inference

‡ determined by sequence

Scattering data were compared to the 6H-QseM-NMR structure using *CRY SOL* followed by a *DAMMIN* model calculation as detailed in 2.8.1. Both *CRY SOL* and *DAMMIN* were run online through *ATSAS* at <https://www.embl-hamburg.de/biosaxs/atsas-online/>. The analysis produced by *CRY SOL* indicated a strong fit of theoretical scattering of the input structure (best representative of 6H-QseM NMR structure) with a χ^2 score of 0.302 (Figure 5.21; note that for data collected at the Australian Synchrotron, a χ^2 of 0.25 represents perfect fit). It is worth noting, however, that for a small, globular (and effectively spherical) protein, there is an increased tolerance for fitting a

structure to the scattering curve compared to a larger more distinct protein structure that requires a more specific scattering profile.

The NMR ensemble of 6H-QseM was fitted into the *DAMMIN* surface model using inbuilt fitting algorithm in UCSF Chimera (Fit in Map) (Figure 5.22). The surface space and curvature fitted well with the NMR ensemble, despite some void areas present which may represent capacity for possible dynamic conformation in solution. It is also possible that some of this extra space was related to the C-terminal 10 residues that constitute the 6H tag which are removed from the NMR ensemble but would contribute to scattering during SAXS. SAXS surface models are fairly limited in presenting fine detail and so have reduced value compared to the *CRY SOL* results.

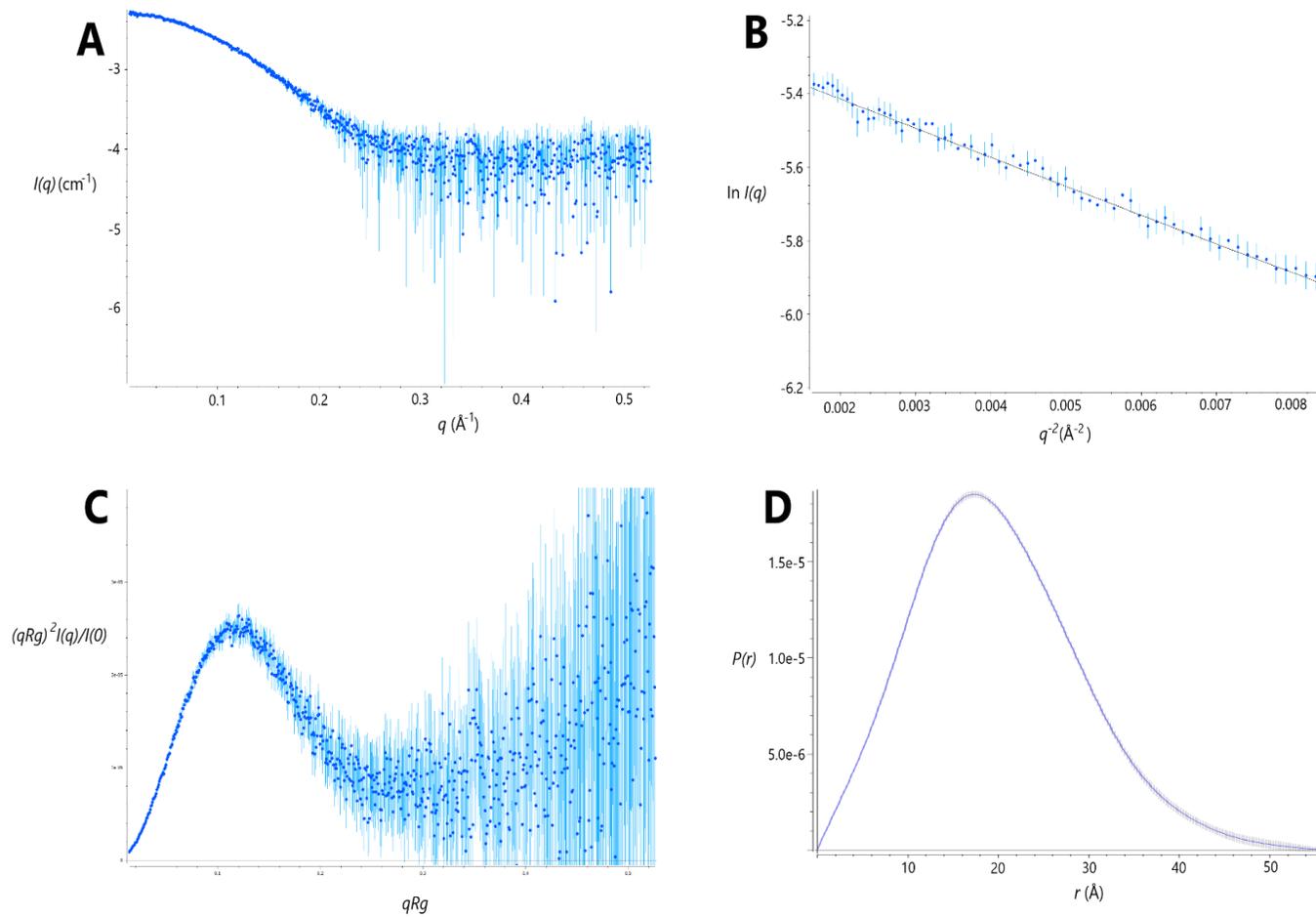


Figure 5.20: SAXS analysis of 6H-QseM. Showing the scattering (A), Guinier plot (B), Kratky plot (C) and distance distribution (D). Averaged values are shown as blue dots in A, B and C, with error bars in light blue.

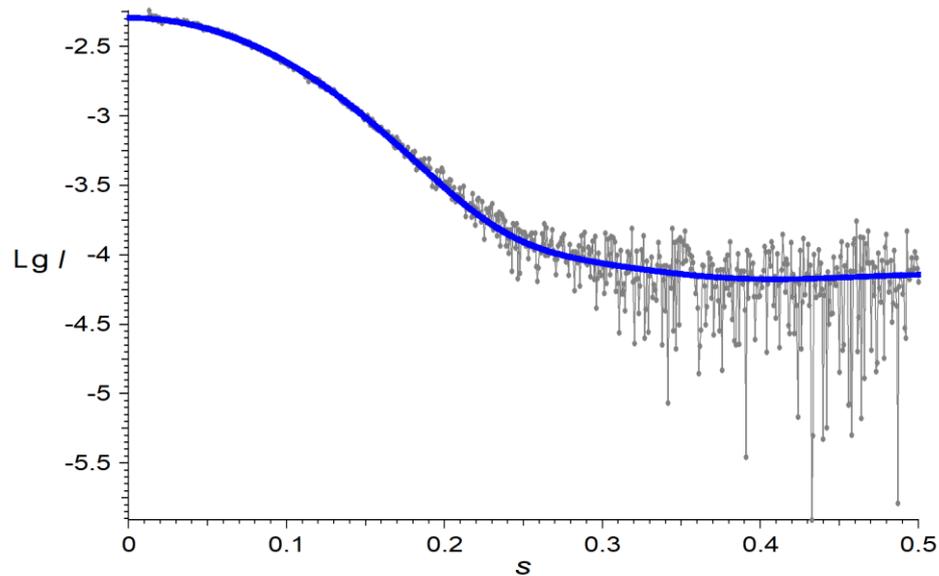


Figure 5.21: CRY SOL comparison between SAXS scattering and NMR ensemble of 6H-QseM. Theoretical scattering of NMR input structure (blue) fitted to collected SAXS scattering (grey dots, with error bars).

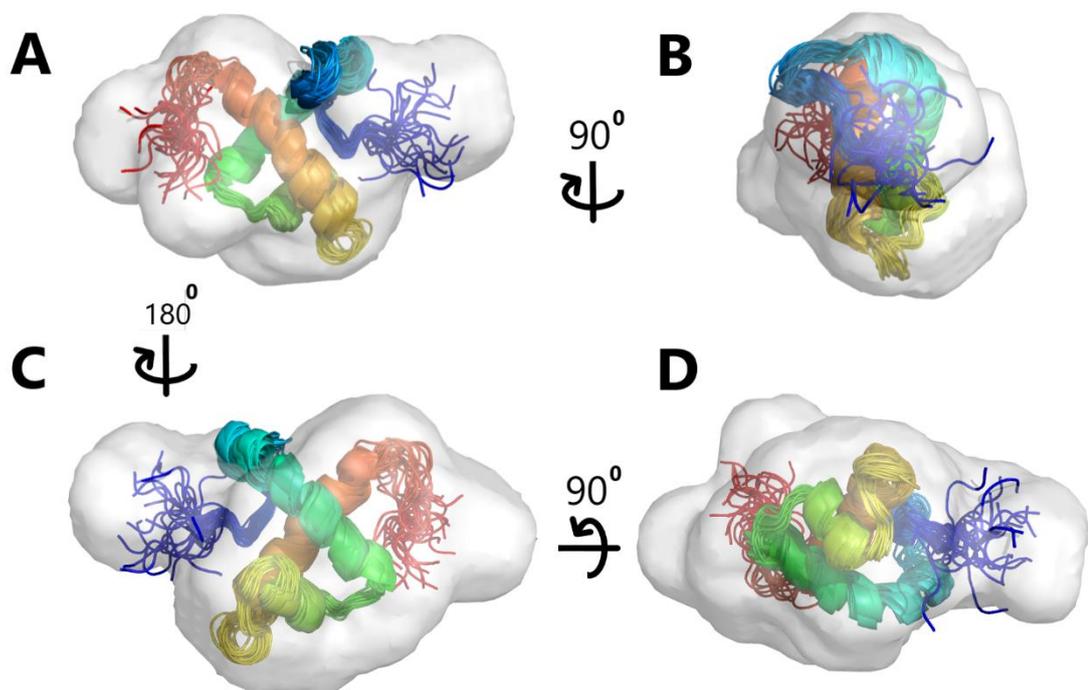


Figure 5.22: DAMMIN surface model with NMR ensemble. 6H-QseM NMR ensemble coloured N to C terminal (blue to red), fitted in SAXS surface map determined by *DAMMIN*.

5.3 Conclusions

Previous attempts to solve the structure of 6H-QseM by X-ray crystallography were unsuccessful largely to difficulties in phase problem solving. Forgoing crystallised 6H-QseM, structure prediction was attempted, however, we could not conclusively assign these structure predictions as the true structure of 6H-QseM in solution. Parallel experiments during crystallographic work identified 6H-QseM as a suitable protein for nuclear magnetic resonance (NMR) experiments, subsequently leading to solving its structure.

We attempted backbone assignment of the protein using two separate datasets. The second dataset included experiments for full sidechain assignment for structure determination. Both backbone assignments resulted in similar completion levels with conserved areas missing assignment. These unassigned residues are likely in variable conformations or otherwise significantly more mobile compared to other parts of the protein since their resonance signals were too weak to locate.

A structure ensemble of 6H-QseM was successfully created using the second NMR dataset. 1D, 2D and 3D spectra were collected, with a final backbone assignment completeness to 88% and overall atom assignment completed to 77% using ^{15}N - and ^{13}C -labelled 6H-QseM. ^{13}C -HSQC/NOESY experiments required protein suspended in D_2O instead of H_2O , requiring an intermediate freeze-drying and re-suspending step. There was no apparent degradation to the protein during this process, however, these samples were often lower in concentration (compared to initial unlabelled and ^{15}N only labelled protein) due to reduced $^{13}\text{C}/^{15}\text{N}$ -labelled protein availability. The lower concentration likely benefited sample solubility over the longer spectra collection run times often required with these $^{13}\text{C}/^{15}\text{N}$ -labelled protein experiments. Most sidechain assignments were achieved using HCCH TOCSY/NOESY experiments, however ^{13}C -HSQC/NOESY were critical for long sidechain assignments that dramatically improved distance restraints. CYANA was used to generate an ensemble with significant convergence of structures for residues 17-84. This NMR ensemble presents 6H-QseM as a triple-helix, globular protein that is discussed in detail in chapter 6.

The NMR structure aligned with predicted structures using *CS-ROSETTA* and *GREMLIN2*. This alignment was highest around the end of the first helix through to the middle of the third helix (residues 35-77) with an average backbone RMSD $<2 \text{ \AA}$. However, these models diverged towards the N and C termini. It is likely that these regions are more mobile since the overall residue composition was hydrophilic and likely exposed to solvent, particularly the N terminal residues (1-17) which include the 6H-tag (which was present in NMR data collection but trimmed from structure calculations). As the 6H-QseM-CoEvo model draws theoretical distance restraints from sequence alignment of the whole DUF2285 protein family, it is likely three-helix motif observed in 6H-QseM is conserved among all other DUF2285 proteins.

The 6H-QseM-NMR ensemble was further refined using *ROSETTA* to improve its suitability in molecular replacement experiments with previously collected crystallographic data. Unfortunately, this refined 6H-QseM-NMR model nor 6H-QseM-CSROSETTA or 6H-QseM-CoEvo were successful in solving a structure of 6H-QseM using the currently available X-ray crystallographic data via molecular replacement. We initially suspected that molecular replacement was unsuccessful due to the poorly defined structure of the termini in the NMR ensemble, however, applying just the well-defined structural core comprising the three helices was also unsuccessful. Furthermore, 6H-QseM predicted structures also failed to succeed in molecular replacement. We concluded that either the crystallographic data is inadequate for molecular replacement due to incorrectly designated space group or unit cell dimensions or the in-solution structure of 6H-QseM differs from the crystal structure to a degree that prevented molecular replacement to succeed.

We also compared our 6H-QseM-NMR ensemble with collected small angle X-ray scattering (SAXS) data of 6H-QseM. The SAXS data results indicated that 6H-QseM is monomeric in solution and globular in shape. This globular shape agrees with our structures of 6H-QseM. The SAXS data was also used to generate a low-resolution representation of 6H-QseM in-solution which agrees with our 6H-QseM NMR ensemble. Comparisons were made between the 6H-QseM-NMR ensemble and the SAXS data via *CRY SOL* which resulted in low χ^2 score, confirming a positive match between the two.

Further NMR experiments with 6H-QseM could be attempted in the future. Ligand-HSQC experiments between 6H-QseM and FseA or TraR could reveal residues critical to binding on 6H-QseM since the ^{15}N -HSQC has been mostly assigned.

Chapter 6:
Structure-function
analysis of QseM and
DUF2285 proteins

6.1 Introduction:

This chapter contains QseM and FseA mutagenesis work conducted by Calum Morris, Will Jowsey and Prof. Clive Ronson at the University of Otago. Author contribution has been defined where appropriate throughout the chapter.

Helix-turn-helix (HTH) DNA-binding motif, in its simplest form, is composed of two helices in a specific spatial arrangement with a critical loop between them (Figure 6.1). Often this HTH motif includes more than two helices and, as is the case of winged HTH, also includes beta-strands. The HTH motif is able to bind DNA by positioning itself into the major groove of the DNA target site. Some HTH motifs use an extended turn feature in DNA binding (253).

As described in chapter 1, Prokaryotic transcription factors often exhibit a HTH motif that is essential for DNA recognition and binding. Structure determination of 6H-QseM by NMR detailed in chapter 5 has shown that QseM contains a HTH motif, however, QseM is not known to bind DNA (Calum Morris, unpublished), only the transcriptional activation proteins TraR and FseA. It is therefore unexpected that QseM retains a HTH motif. It may be that proteins with the majority of DUF2285 domain contain a HTH motif for DNA binding, like FseA and RovC, however, some variation in QseM has made it incapable of binding DNA. We know that QseM is an antiactivator and perhaps its non-functional HTH domain is critical to that function. It is also worth noting that many HTH motifs function in proteins that form a quaternary structure, such as the dimer of TraR and the hexamer of RovC. FseA may also form a superstructure that may be prohibited from binding DNA by interference from QseM. This high specific interaction is supported by the fact that QseM is only present in genomes that have FseA.

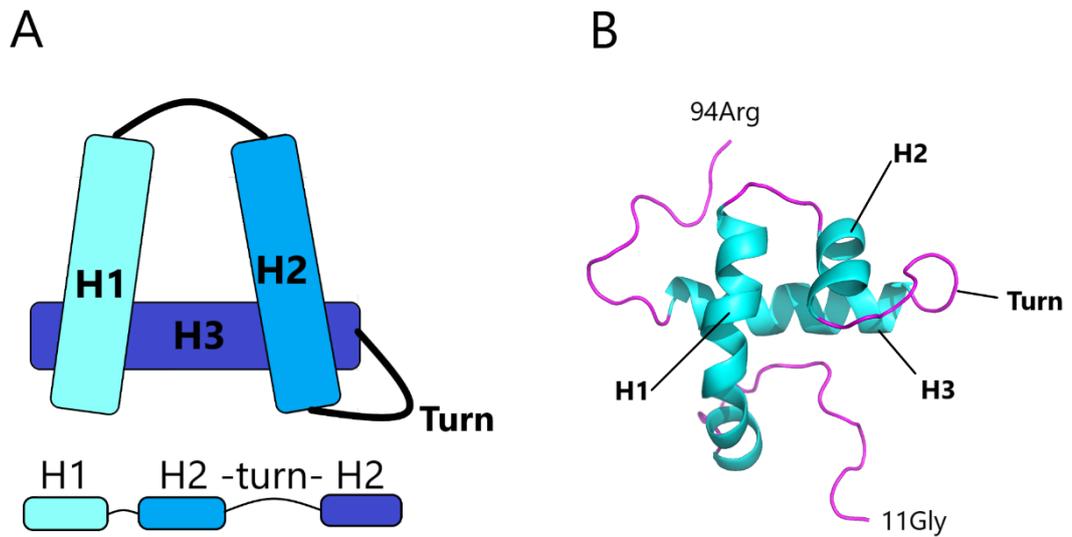


Figure 6.1: Helix-turn-helix motif of QseM. A typical HTH motif contains 3 helices, with two forming the critical HTH feature required for DNA binding (A). 6H-QseM-NMR structure features a HTH motif across the 2nd helix, loop and 3rd helix secondary structure (B).

6.2 Results and Discussion:

Chapter 5 details structure of 6H-QseM solved by NMR. To summarise, 6H-QseM is a globular protein with three distinct alpha helices, the loop between H2 and H3 forms a HTH motif. Both N and C termini are positioned away from the HTH motif and are disordered. It is worth highlighting in chapter 5 we conclude that 6H-QseM-NMR structure is an accurate representation of the DUF2285 domain.

6.2.1 Alignment of QseM structure to known proteins

Our 6H-QseM-NMR ensemble was used with both *DALI* and *COFACTOR* servers to find similar protein structures across the PDB. We used a 'model of best representation' of our CYANA generated ensemble for structure alignments (as per method 2.9.3), however, regions of disorder prevented both the *DALI* and *COFACTOR* algorithms for finding comparable protein structures and were trimmed from the structure during the search. *DALI* found a total of 1216 structure matches (before a cut-off Z-score of 2.0) with 11 unique structure alignments $\geq 20\%$ sequence identity and a Z-score ≥ 3.0 (Table 6.1). With the highest Z-score at 3.5 (sequence identity of 20%) match being RNA polymerase sigma factor CNRH (4cxf-A) and the highest sequence identity match at 24% (with a Z-score of 3.3) being RNA polymerase sigma factor (3hug-E).

Table 6.1: list of primary DALI matches to 6H-QseM NMR input.

Rank	Chain ID	Z-score	RMSD	Lalign	Nres	%ID	Description
1	4cxf-A	3.5	2.1	44	154	20	RNA polymerase sigma factor CNRH
2	2h27-A	3.4	2.4	45	71	20	RNA polymerase sigma E factor
3	3vep-D	3.3	2.5	45	68	22	Uncharacterised protein RV3413C/MT3522
4	3hug-E	3.3	2.3	45	73	24	RNA polymerase sigma factor
5	2vbx-A	3.3	2.7	47	149	21	Transcriptional regulatory protein
6	5cy2-F	3.2	3.0	45	180	20	Transposon TN3 (subunit)
7	6dv9-F	3.1	2.4	45	174	24	DNA-directed RNA polymerase subunit alpha
8	1xsv-A	3.0	2.4	45	106	22	Hypothetical UPF0122 protein SAC1236
9	5z7i-B	3.0	2.2	42	45	24	Cell cycle regulatory protein GCRA
10	2w48-C	3.0	2.1	44	299	23	Sorbitol operon regulator
11	1i1g-A	3.0	2.8	47	140	21	Transcriptional regulator LRPA

Chain ID associated with RSCB PDB entry. **Z-score** a determination of fit, with >2.0 being significant. **Lalign** is the number of residues of entry which are part of structure alignment with 6H-QseM query. **Nres** is the number of total residues in entry. **%ID** is the sequence identity between entry and 6H-QseM query.

The DALI matches 4cxf-A and 3hug-E had the last 45 residues aligned with the 6H-QseM-NMR structure (Figure 6.2), but divergent at the C and N termini. This alignment was characterised by comparing RMSD (Figure 6.3) and sequence alignment (Figure 6.4 and Figure 6.5). RMSD analysis indicated that the core triple-helix region is conserved, with divergence occurring in at the C

and N termini, as well as the turn region between the second and third helix, which appears slightly extended in 6H-QseM-NMR structure compared to all other *DALI* matches. It is noteworthy that many HTH motifs, like those reflected in Figure 6.2, feature a much shorter connection between H2 and H3. The extended loop between H2 and H3 in QseM is consistent with other predicted and calculated models of 6H-QseM (6H-QseM-CSROSETTA, 6H-QseM-CoEvo described in chapter 5). All but one of the *DALI* matches outlined in Table 6.1 displayed similar alignments to those shown in Figure 6.2. This different alignment match was 5z7i-B (Cell cycle regulatory protein GCRA), which only had 45 residues and occupied the entire conserved structure of 6H-QseM.

There was limited residue conservation between 6H-QseM-NMR and *DALI* matches. These residues included Ser49, Ala54, Gly59 and Val73. These are likely embedded residues that are conserved for structural purposes. Similarly, many conservative substitutions are often located in embedded positions and many non-conserved residues are surface-exposed.

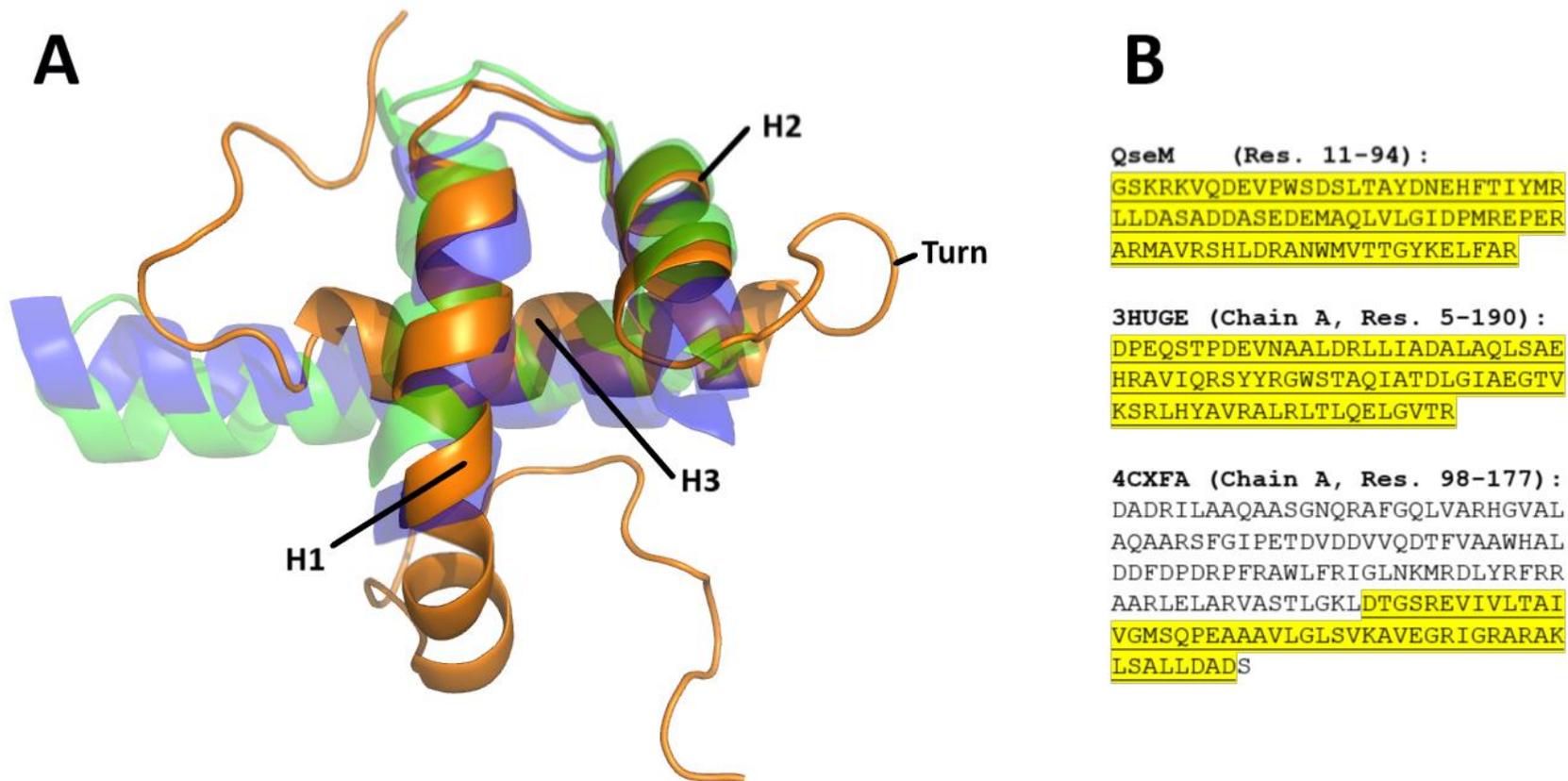


Figure 6.2: DALI matches structure aligned to 6H-QseM. 6H-QseM (orange) structure aligned with RNA polymerase sigma factor CNRH [4cxf-A] (blue) and RNA polymerase sigma factor [3hug-E] (green) (A). Amino acid sequence of these proteins with region displayed in part A highlighted and underlined. Hidden parts of the sequences fail to align with the C or N terminal ends of 6H-Qsem.

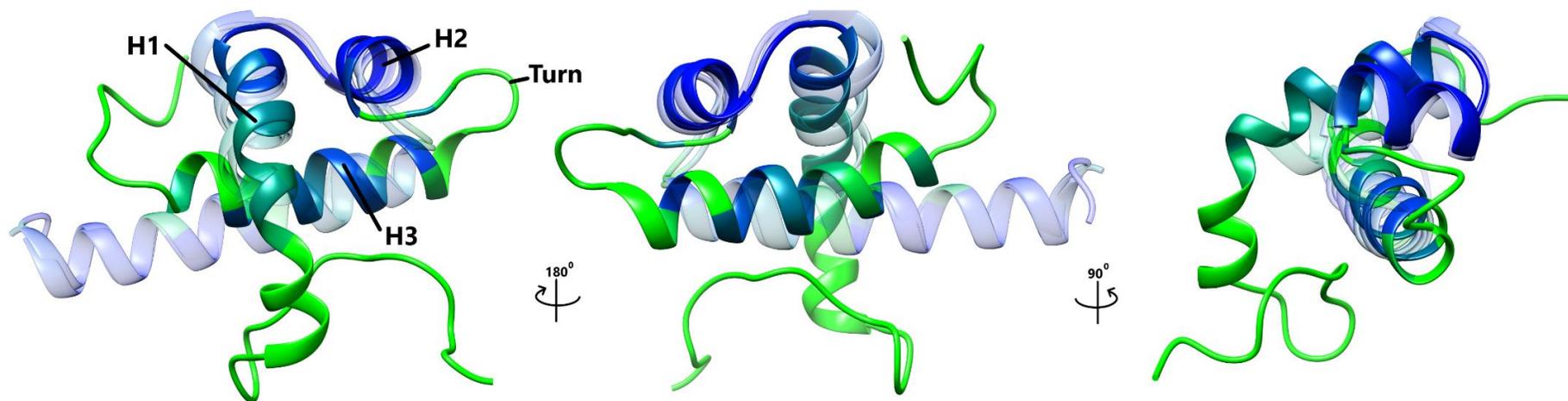


Figure 6.3: RMSD analysis from *DALI* matches. RMSD coloured from most aligned (blue) to least (green). The two matching structures are transparent. Greatest convergence occurs where the three helices (H1, H2 and H3) are in close proximity to each other, large divergence occurring towards the C and N terminal ends of 6H-QseM-NMR.

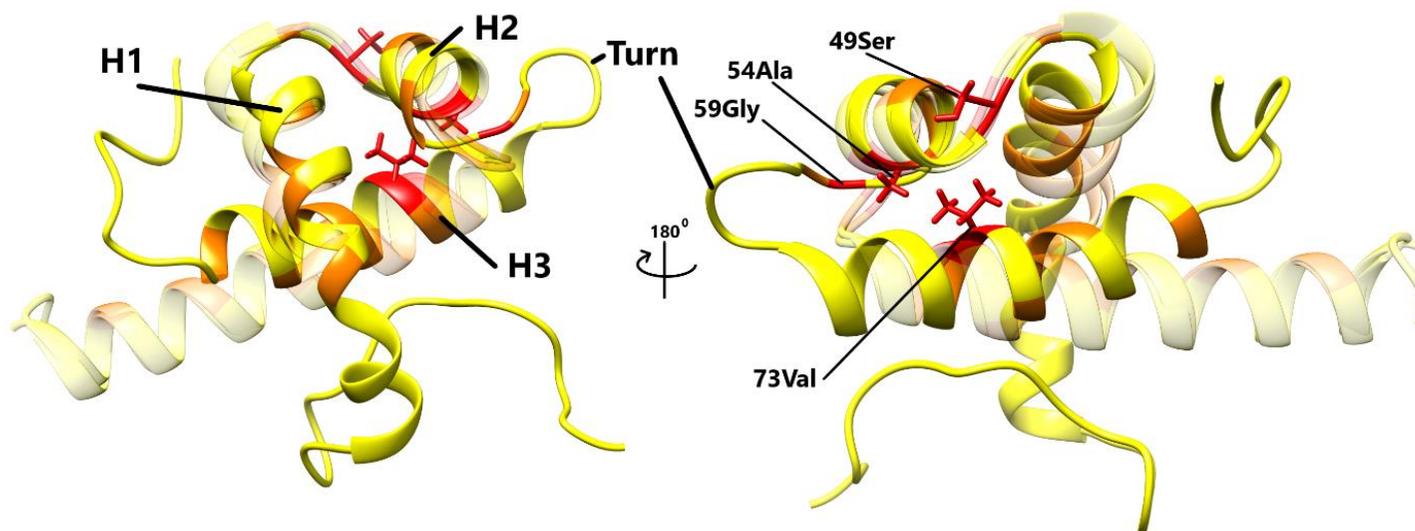


Figure 6.4: Residue conservation in structure alignment. Fully conserved residues (red), along with similar residues (orange) and non-conserved residues (yellow). Three helices noted (H1, H2, H3) and the turn.

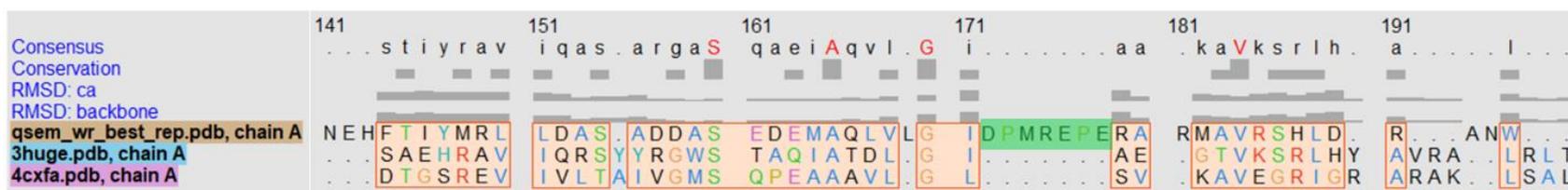


Figure 6.5: Sequence alignment of converged structure region. Sequence in orange boxes are areas where the structures (Qs align, as visualised in Figure 6.4). The turn (highlighted green) feature of 6H-QseM between the 2nd and 3rd helix is apparent in this alignment.

COFACTOR generated similar results to *DALI*, giving a list of the top 10 matches ranked by a generated TM-score (percentage of the query 6H-QseM structure residues which align with known protein structures). The top TM-score was 54%, and the lowest was 46%, suggesting approximately half of the residues in the 6H-QseM-NMR structure form a similar structure to the given matches. Again, similar to the *DALI* results, the most closely matching areas that align to the first helix connects to the second and where the second helix connects to the third in 6H-QseM. Similar to the matches identified by *DALI*, divergence between *COFACTOR*-identified structures is most prominent towards the C and N termini of the query structure. For the matching regions (defined by the TM-score), an average RMSD was calculated, which ranged between 2.5 and 4 Å across the 10 matches. The residue sequence identity between the QseM query and the matches was very low, ranging between 4 and 12.5%, which was also fairly consistent with the *DALI* results.

COFACTOR also attempted to determine some preliminary predicted gene ontology and ligand binding sites. The predicted gene ontology results had low Cscores for molecular function prediction with only one potential association with amidine-lyase activity, a Cscore of 0.37. For predicted biological process there were multiple matches which included protein oligomerisation and ribonucleotide processes, all had a low Cscore of 0.37. Multiple matches with identical Cscore values suggest no definitive answer among the set, with it likely that none are correct. Predicted cellular component gave a more confident result, placing the protein in the cytoplasm with a Cscore of 0.75, which matches with our current hypothesis of QseM functioning as an antiactivator within the cytoplasm and not membrane embedded or excreted into extracellular space. No predicted ligand binding sites could be determined as no homology model was found with known ligand binding sites. This does not definitively suggest that QseM has no potential ligand, only that no supporting evidence was found.

Use of *3DLigandSite* using the 6H-QseM-NMR structure could not determine accurate binding sites for small molecules or ions due to low number of homology pairs with ligand information. We also attempted using just a sequence input, relying on the software's use of the *Phyre* server (254) for

structure prediction to create a structure and then scan for homology models, however, this also did not reveal any matches.

6.2.2 Structure determination of FseA

FseA structure was predicted using *GREMLIN*, *RoseTTAfold* (255) and *Alphafold2* (256). These *ab initio* predictions were generated using structure prediction algorithms supplemented with distance restraints derived from co-evolution data. All three software methods produced similar FseA structures, with the most similar having an average backbone RMSD of 1.9 Å (between *RoseTTAfold* and *Alphafold2*). Co-evolution data was based sequence alignment of ~4,500 FseA homologues. It is worth noting that C-terminal DUF2285 domain of FseA is incorrectly annotated by excluding the N-terminal domain of FseA homologues, when in reality, both domains are often expressed as a single polypeptide due to the +1 programmed ribosomal frameshift motif detailed in section 10.1. FseA structure predictions were completed by Will Jowsey.

6.2.3 Structure comparison between QseM and FseA

Both QseM and the FseA DUF2285 domain models exhibited three helices (H1, H2, H3) with a HTH motif forming between the H2 and H3 of 6H-QseM-NMR (Figure 6.6). The HTH of the FseA DUF2285 likely exists between H2 and H3, with a smaller helix, H2b within that turn. Existence of extra secondary structure in the turn of FseA draws parallels to the HTH motif of Q antitermination factor in *E. coli* (A novel Helix-turn-[loop]-helix motif was recently identified in the *E. coli* Q antitermination factor (253)) which has a small antiparallel beta sheet feature within the turn (Figure 6.7). As is typical of HTH motifs, a α -helix of the HTH in Q antitermination factor is for sequence recognition and fits into the major groove of the DNA target within the regulatory gene cassette, Q. The Q antitermination factor is required to maintain a stable interaction between the RNAP and the DNA of Q, to that end, the additional secondary structure of the loop, which includes B1/B2 β -sheets ensure a strong interaction. This extended loop feature is what makes

the HTH motif of Q antitermination novel. QseM and FseA display an extended HTH loop, however neither exhibit β -sheets, with the loop of FseA forming a distinct α -helix (H2b). It is for this reason that the DUF2285 domain HTH motif is novel and may be distinct in the loop for added stability. It is also noteworthy that QseM's HTH loop is far simpler to that of FseA, potentially revealing the reason why only the DUF2285 of FseA is capable of binding DNA.

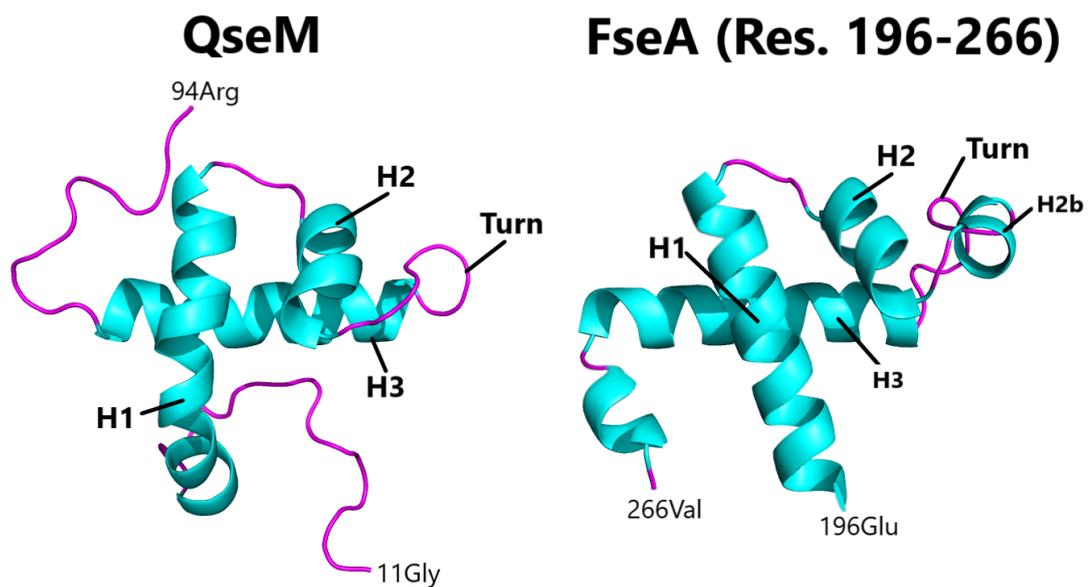


Figure 6.6: 6H-QseM-NMR structure compared to DUF2285 domain of FseA. The QseM structure taken from NMR ensemble, FseA DUF2285 domain taken from the predicted structure of FseA. Structures are coloured by secondary structure with alpha helices coloured cyan and loops coloured purple.

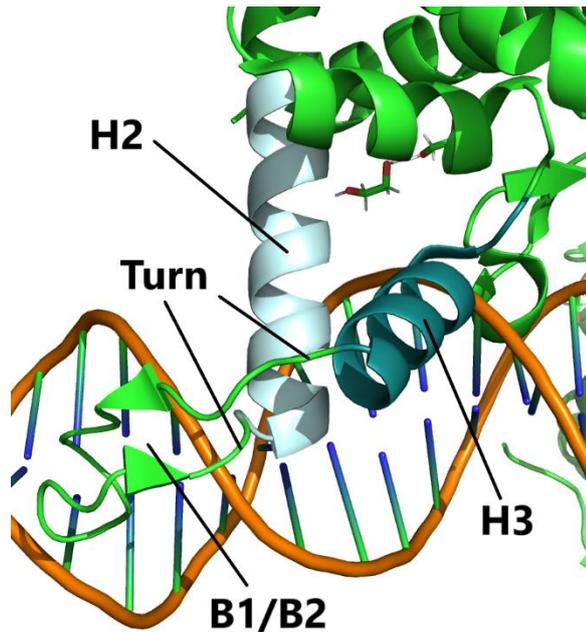


Figure 6.7: Q-dependent antitermination protein binding DNA. HTH motif shown including the 2nd and 3rd helices (H2 and H3) the turn and the two beta-strands (B1/B2) within that turn motif. RSCB PDB ID: 6P18 (253).

The electrostatic surface of both 6H-QseM-NMR and FseA (residue 196-266) was examined to reveal distinguishing features between either protein (Figure 6.8). The surface of 6H-QseM-NMR has more acid residues around the area where H1 connects to H2 whereas it has more basic residues around the C and N termini. There are also some base residues flanking the acid-rich region around the H1-H2 connection. FseA (residues 196-266) has a many more basic residues than acidic residues on the surface. The majority of these basic residues are located at the end of the turn and most of H3 as well as areas where H2 and H1 are close to H3.

The turn motif of a HTH motif is involved in binding DNA, however, the turn in 6H-QseM-NMR contains an even balance of basic and acidic residues. In contrast the HTH turn in FseA has an overall greater number of basic than acidic residues and corresponds with what is expected in a DNA-binding HTH turn motif. This surface charge discrepancy between 6H-QseM-NMR and FseA may be a critical factor as to why it appears only FseA is able to bind DNA.

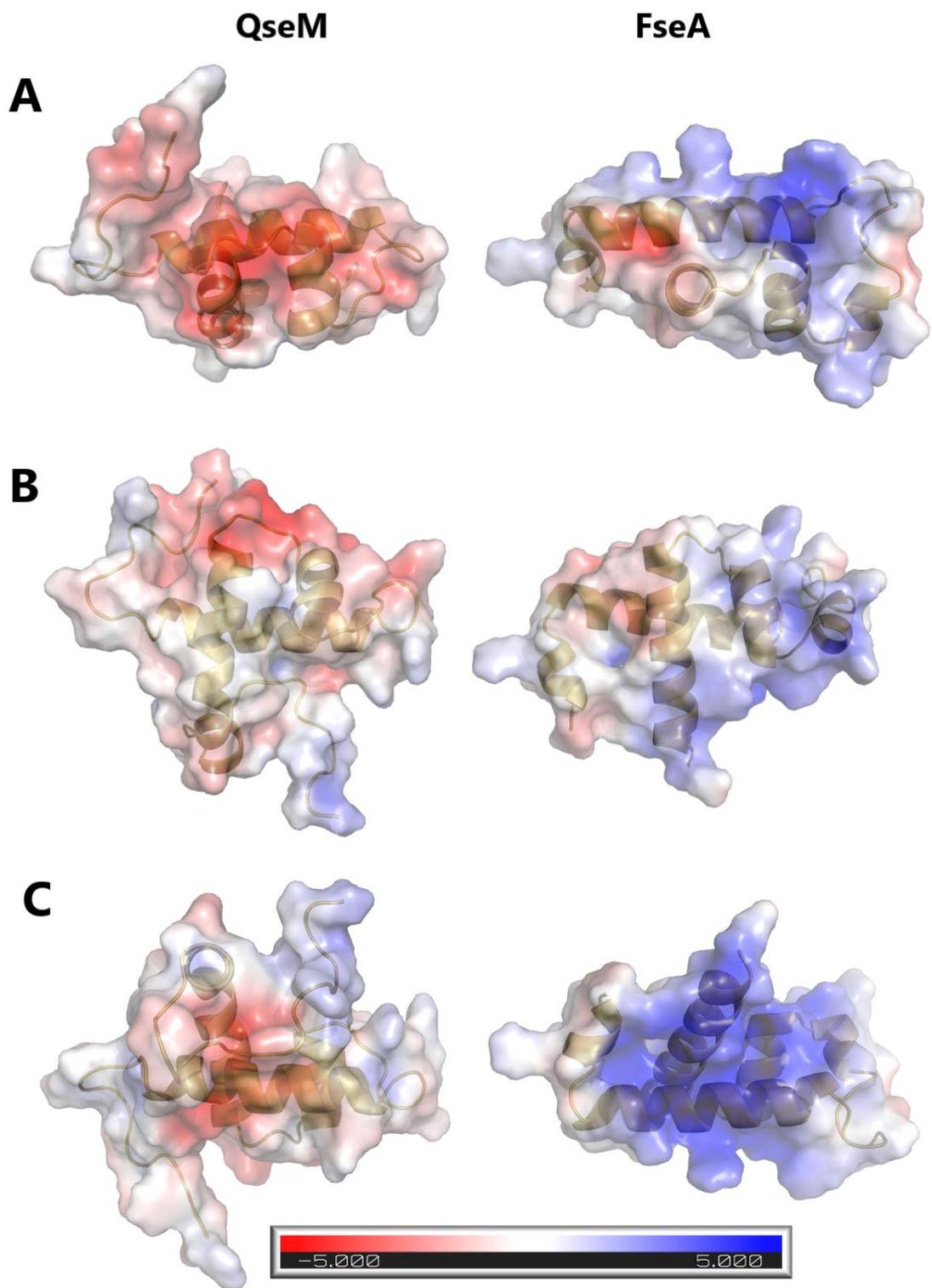


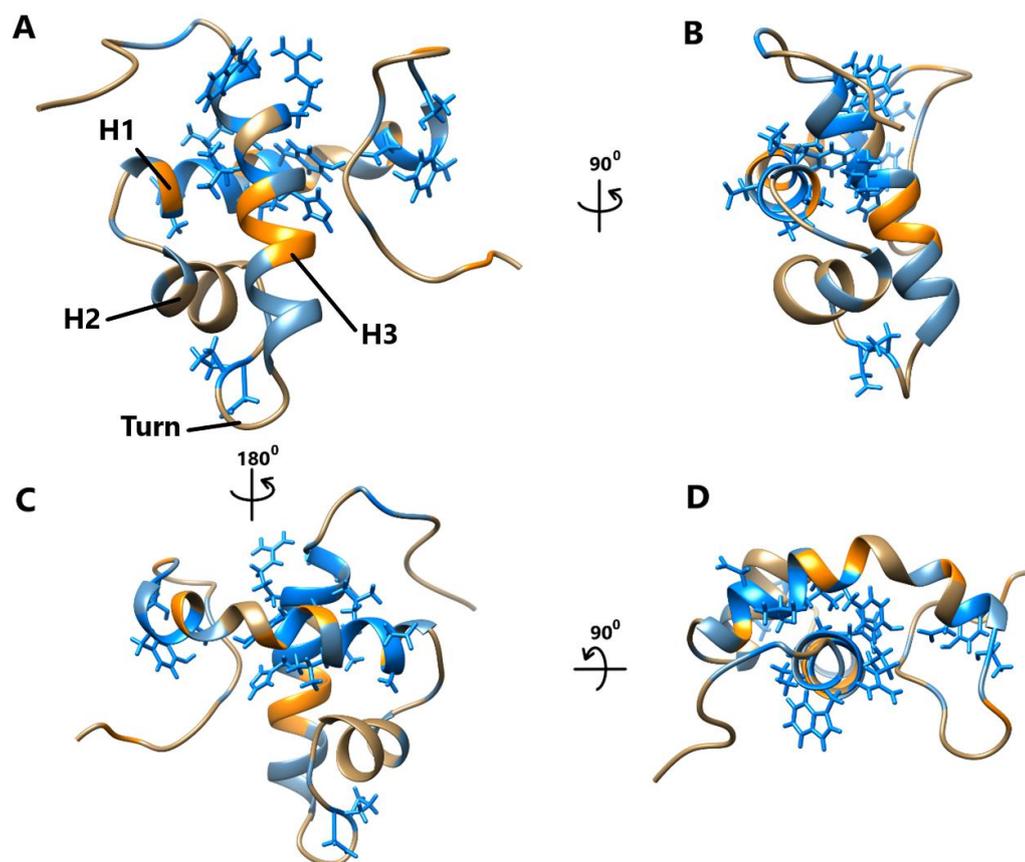
Figure 6.8: Electrostatic surface of QseM and DUF2285 domain of FseA. Surface map coloured by charge of 6H-QseM and FseA (residues 196-266) in (A) top, (B) front and (C) bottom orientations for both. Negative (red), natural (white) and positive (blue) electrostatic charge. Electrostatic surface generated using *APBS Electrostatic* plugin on PyMOL.

6.2.4 Residue conservation of QseM and FseA

A list of protein sequences similar to QseM was generated by a non-redundant blastp search (QseM sequence as query) using default settings. This list contained a total of 100 sequences with identify scores ranging from 100 to 73.49%. Sequence alignment was carried out on this list, herein referred to as QseM homologues, and residue conservation was mapped onto the 6H-QseM-NMR structure (Figure 6.9). Highly conserved residues clustered in H1 (residues 27-43) and H3 (residues 76-87), which are in close proximity to each other in the structure. These residues may be critical to antiactivation function since they are mostly solvent exposed and are heavily conserved. The extreme N- and C-terminal ends, which appear disordered in the 6H-QseM-NMR ensemble, had less residue conservation suggesting these regions may be less critical for QseM function. Interestingly, low residue conservation exists in the embedded area around H1, H2 and H3 (residues 40-80) despite having the greatest structural alignment between all 6H-QseM structures (detailed in chapter 5). However, the residues in this embedded area were often changes to chemically similar amino acids, which is expected if the overall structure of the protein was to be maintained.

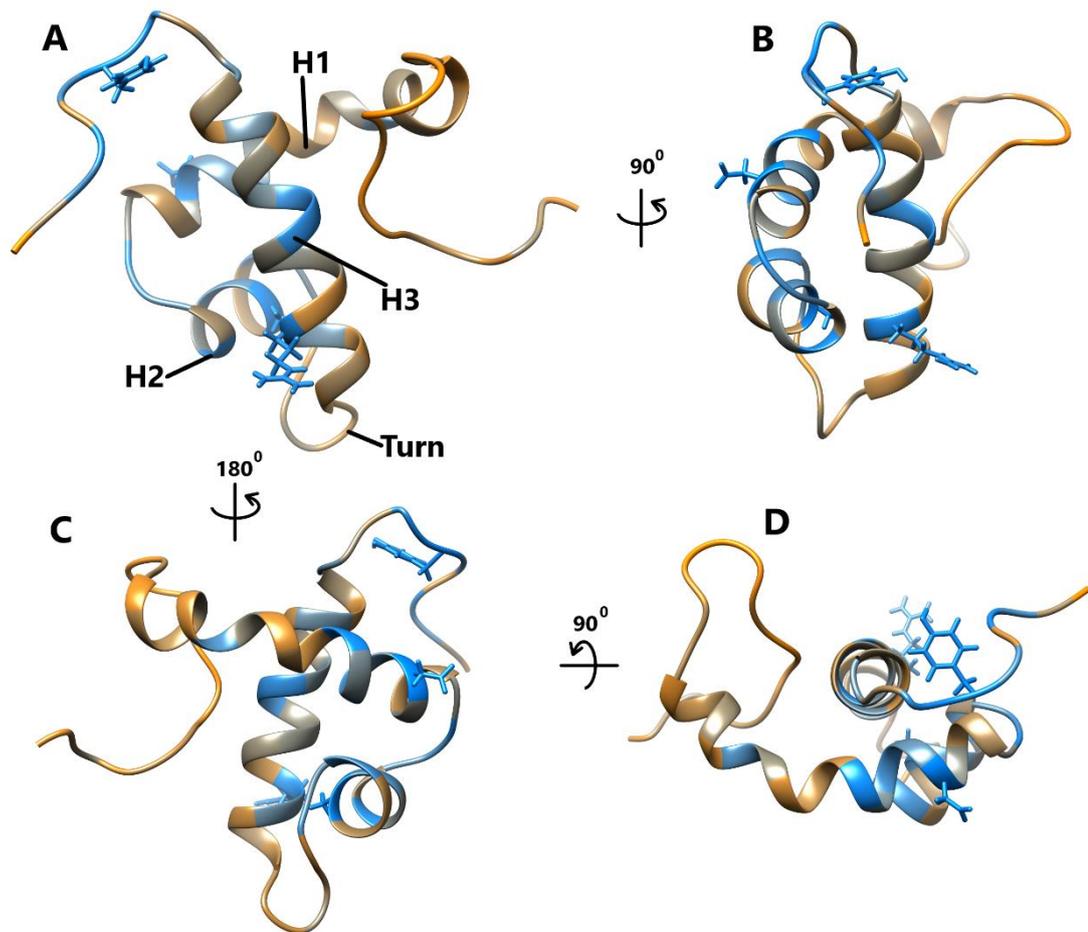
A much larger sequence alignment against QseM was created using all DUF2285 proteins identified by Pfam using default settings on Pfam (Pfam family ID PF10074) (<http://pfam.xfam.org/>). This sequence alignment, herein referred to as DUF2285 family, contained a total of 959 sequences aligned to QseM. Conserved residues were mapped on the 6H-QseM-NMR structure (Figure 6.10). Due to the larger range of sequences used, residue conservation was weaker compared to the QseM homologues; however, areas of conservation were still detectable, mostly existing as chemically similar, rather than identical residues. In general, these conserved areas were more apparent in the buried hydrophobic centre of the protein where H1, H2 and H3 were in close contact. The surface-exposed residues were significantly less conserved than internal ones. The residue conservation suggests that DUF2285 domain of proteins likely share a similar structure that is built using similar residues. The N and C termini show the least residue conservation in the DUF2285 family sequence alignment. This result is in contrast to that of

the sequence alignment of just QseM homologues, where the C and N termini residues are significantly more conserved. This could mean that the residues on these termini are more likely associated with the conserved function among QseM homologues, not the HTH motif. The DUF2285 sequence alignment presents a residue conservation pattern that agrees with the 6H-QseM-NMR structure since areas which can accommodate residue changes, like those that are surface exposed, to have more freedom to change compared to those that are internalised.



*MRGSHHHHHH GSKRKVQDEV PWSDSL**TAYD** NEHF**TIY**MRL **LDAS**ADDASE
 DEMAQLVLGI **DPM**REPERAR MAVR**SHLD**RA N**WMV**TT**GY**KE L**FAR***

Figure 6.9: Residue conservation of QseM homologues. NMR structure (best representative from ensemble) coloured by conservation (high is blue, low is orange) from multiple sequence alignment of QseM homologues. Residue position with complete conservation have sidechains shown and are highlighted in the sequence presented (Thr27, Try29, Asp30, Try37, Leu40, Leu41, Asp42, Ala43, Asp61, Pro62, His76, Leu77, Arg79, Ala80, Trp82, Gly87).



MRGSHHHHHH GSKRKVQDEV PWSDSLTAID NEHFTIYMRL LDASADDASE
DEMAQLVLGI DPMREPERAR MAVRSHLDRA NWMVTTGYKE L FAR

Figure 6.10: Residue conservation across the entire DUF2285 protein family. NMR structure (best representative of ensemble) coloured by conservation (high is blue, low is orange) from multiple sequence alignment of DUF2285 protein family. Residue position with highest conservation have sidechains shown and highlight in the sequence provided (Asp42, Ala54, Arg70, Try88).

6.2.5 Mutagenesis analysis of QseM and FseA

Preliminary sequence alignment of several QseM homologues found in various species revealed substantial residue conservation (work done by Calum Morris, unpublished) (Figure 6.11). The QseM sequence of *Sphingomonas* sp. (SKA58, ICE1) appears most divergent compared to QseM from *M. loti* R7A, *Stappia indica* SBBC, *Xanthobacter autotrophicus* Py2 and *Caulobacter* sp. FWC2. Each QseM homologue antiactivation function tested against FseA and TraR of *M. loti* R7A. Many of these QseM homologues showed antiactivation function against FseA but none (aside from *M. loti* R7A) were able to antiactivate TraR. The QseM of *Sphingomonas* sp. (SKA58, ICE1) and, to a lesser extent, (SKA58, ICE2) are appear unable to antiactivate FseA.

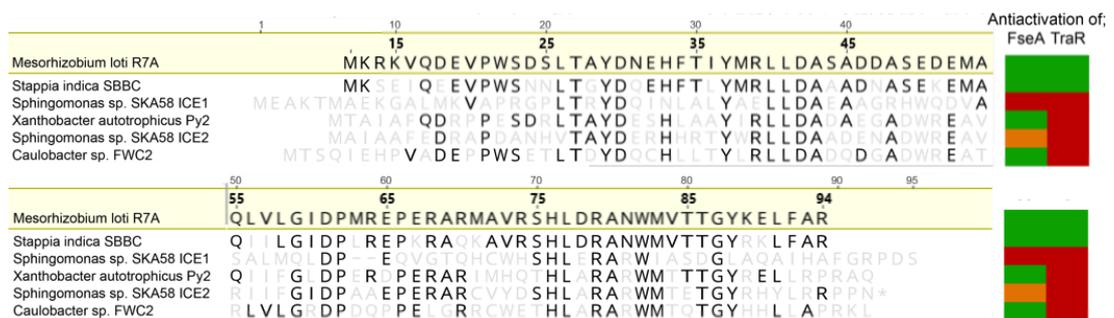


Figure 6.11: Preliminary sequence alignment of QseM homologues from various species. Sequences are aligned to QseM of *M. loti* R7A (residue position numbers correspond to the 6H-QseM sequence). Areas of high conservation can be observed between residues across the QseM sequence, in particular, regions 25- 29, 37- 43, 58- 61 and 75- 87.

Variants of *qseM* were constructed with single or double alanine substitutions of 28 residues that were highly conserved among QseM homologues. These were synthesised and cloned into overexpression plasmid pPR3. These expression vectors were subsequently transformed into the *M. japonicum* strain R7ANS (R7A cured of ICEMISymR7A) carrying pSDzfseA, which expresses *fseA* by way of the *P_{lac}* promoter and the *lacZ* gene from the FseA-activated promoter *PrdfS*. QseM variants unable to antiactivate FseA result in β -galactosidase activity from *PrdfS* in *M. japonicum* strain R7ANS. Wild-type QseM abolished FseA-dependent β -galactosidase activity from *PrdfS*. Variants of *qseM* with identical alanine substitutions were cloned into expression vectors pBT. FseA was cloned into the expression vector pTRG. Both pBT and pTRG were transformed into an *E. coli* histidine auxotroph. This transformed strain was grown on medium lacking histidine. Strains with positive protein-protein binding between QseM variants and FseA grew significantly more than QseM variants that failed to bind FseA (normalised for background growth and compared to positive and negative controls). Similar *lacZ* and bacterial two-hybrid assays were conducted for QseM variants to detect their ability to antiactivate and bind TraR (Figure 6.12). Work completed by Calum Morris and Will Jowsey.

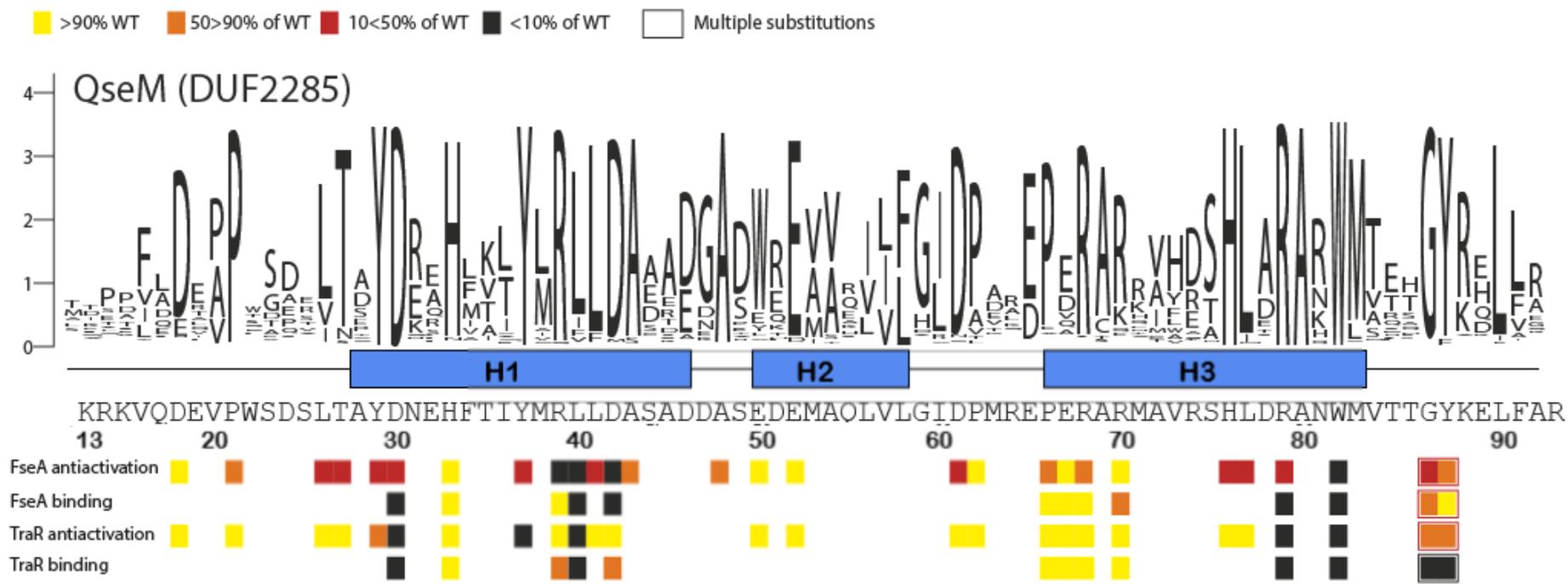


Figure 6.12: Summary of QseM residue substitution in function analysis. Data and figure produced by Calum Morris and Josh Ramsay. Alanine substitutions resulted in >90%, (yellow), 50>90% (orange), 10<50% (red) and <10% (black) binding or antiactivation of TraR or FseA relative to wild type QseM. Antiactivation function or protein-protein binding was assayed via bacterial two-hybrid or LacZ assays (Calum Morris, unpublished).

Analysis of residues isolated in substitution assays and subsequently mapped on the structure highlighted specific residues and structural regions that are likely to be critical to antiactivation function against FseA (Figure 6.13). These residues included Leu26, Thr27, Tyr29, Asp30, Tyr37, Arg39, Leu40, Leu41, Asp42, Asp61, His76, Leu77, Arg79 and Trp82. Of these, Asp30, Tyr37, Leu40, Arg79 and Trp82 were also critical for antiactivation against TraR (Figure 6.14). Of these residues, Arg79 and Trp82 appear surface exposed on the 6H-QseM-NMR structure. Furthermore, both these residues are highly conserved among QseM homologues.

Overall, these results are in agreement with the NMR structure, as the residues 26-45 are in proximity with residues 76-90 and together strongly suggest that this area of the 6H-QseM structure is where binding takes place to for antiactivation to occur, unrelated to the HTH motif (Figure 6.15).

In contrast, conserved residues in other regions of the protein appear to have no noticeable effect on function when substituted as an alanine. This matches with the NMR structure of 6H-QseM since it positions these 'inconsequential' residues on the surface, where mutations would not significantly alter the structure of the protein. However, it is also possible that the substitutions were mostly alanine and so in some cases they would not be expected to significantly disrupt hydrophobic areas and may not break its native structure if these substitutions were embedded in the structure. Similarly, alanine substitutions to residues located on the surface may not significantly disrupt surface charge to the point where loss of function is noticeable.

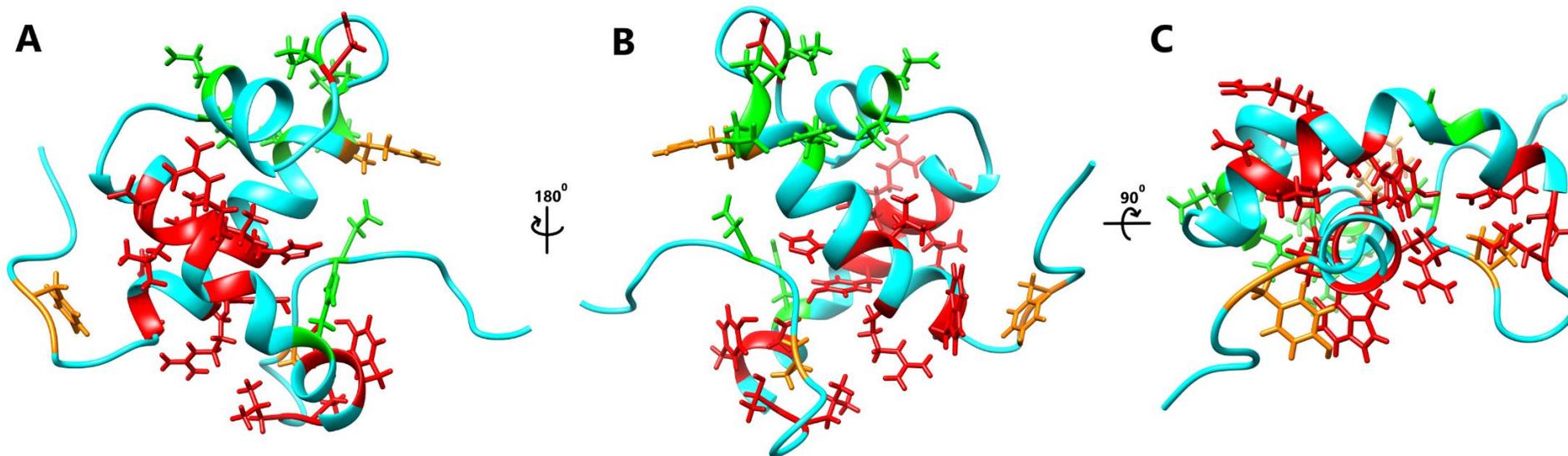


Figure 6.13: Mapped alanine substitution results on 6H-QseM (antiactivation against FseA). Residues coloured by ability to knockout antiactivation function (red), reduce function (orange), and have no apparent effect on antiactivation function (green). A distinct concentration of residues required for antiactivation is present in half of the protein (H1 and C-terminal end of H3).

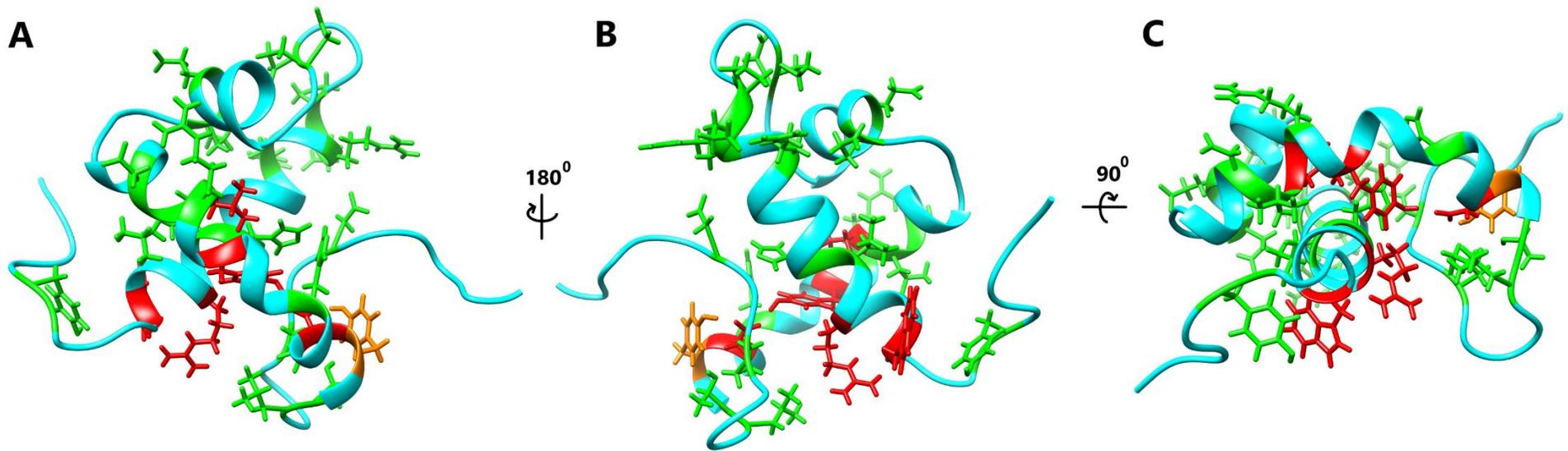


Figure 6.14: Mapping alanine substitution results on QseM structure (antiactivation against TraR). Residues coloured by ability to knockout antiactivation function (red), reduce function (orange), and have no apparent effect on antiactivation function (green). Compared to mutagenesis results regarding QseM against FseA, there appears to be far less residues critical to antiactivation. These critical residues are concentrated in roughly the same region as those critical to antiactivation of FseA.

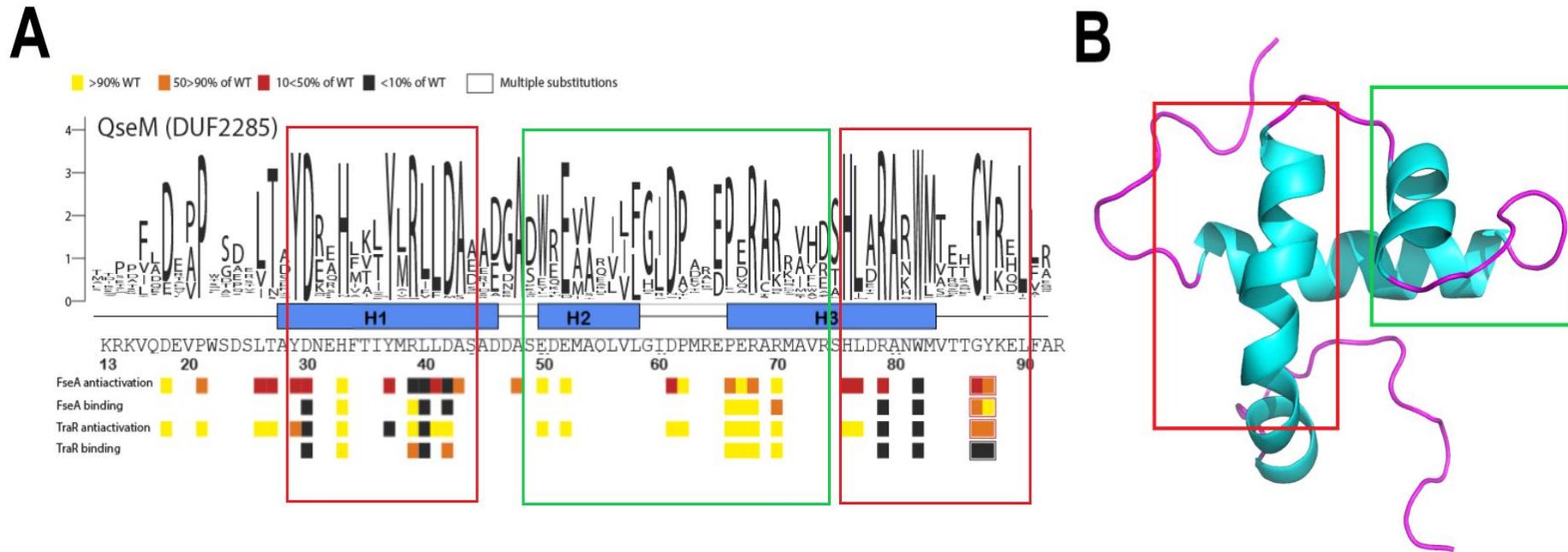


Figure 6.15: Overview of binding site and HTH motif of 6H-QseM. Regions critical to antiactivation and protein binding locations (red boxes) on both mutational analysis figure (A) and NMR structure (B). The HTH motif (green boxes) on both mutational analysis figure and NMR structure. It is evident that residues critical to antiactivation and binding are separate from the HTH motif. Section A generated by Calum Morris.

Analysis of a superimposed complete NMR ensemble was useful in highlighting residues with variable or consistent sidechain positioning. Residues that are internalised in the protein are likely to have a greater number of distance restraints and reduced steric freedom and are expected to appear in consistent positions across the 20-structure NMR ensemble. Residues that have low mobility yet appeared to be surface exposed are likely limited by distance restraints only. Critically residues include Leu26, Leu41, Asp42 and Trp82. These critical residues all appear somewhat surface exposed, conserved in QseM homologues and required for antiactivation of FseA. Additionally, these critical residues have minimal side-chain range of mobility across the entire 20-structure NMR ensemble, suggesting that their position is conserved within the QseM structure and thus further suggesting they are critical to antiactivation or protein binding. It's possible that the Leu26 and Leu41 are necessary to sustain structure. The Asp42 is conserved for charge and contributes to the negatively charged surface area (described in Figure 6.16) which may be disrupted by an alanine substitution, potentially highlighting the critical importance of this negatively charged area either in function or maintaining structure. Finally, the Trp82 is a large residue, highly conserved in the DUF2285 and QseM homologues and is a common residue involved in extensive hydrogen bonding. It is likely that this particular residue may be of critical importance to function. Significant hydrophobic regions, supported by extensive distance restraints on the opposing side of the helix which Trp82 exists on, strongly suggests that it remains surface exposed. Residues Asp30, Tyr37, Leu40, Val73, His76 and Leu77 are conserved and have low mobility because they are internalised. These residues likely have no direct input to QseM's antiactivation function, however, they may be critical to maintain specific protein structure which may contribute to residue positioning and surface motifs necessary for function.

Residues Thr27, Tyr29, Arg39, Asp61 and Arg79 are conserved, surface exposed and apparently critical to antiactivation function against FseA (Figure 6.17). Arg76 is of particular interest since it is in close proximity to Trp82, a residue already flagged as being potentially critical to antiactivation against both FseA and TraR.

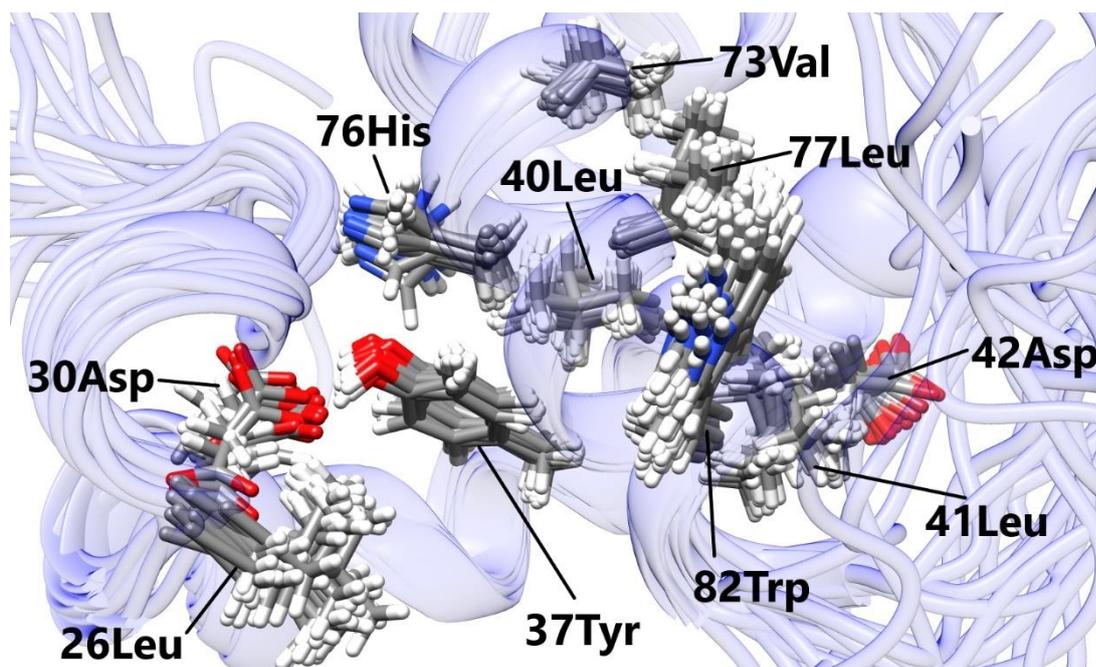


Figure 6.16: Conserved residues of 6H-QseM likely associated with antiactivation (low mobility). Resides shown have low sidechain variance within the ensemble. Residues Leu26, Leu41, Asp42 and Trp82 are surface exposed and may be critical to antiactivation function. Mobility can be observed by higher positional range the sidechains can exist in.

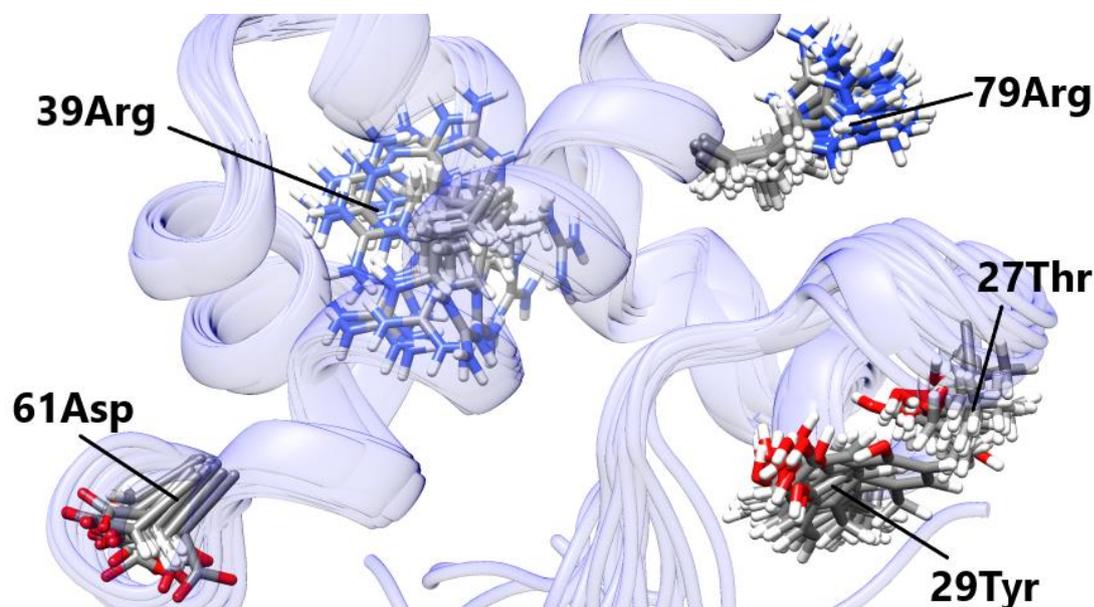


Figure 6.17: Conserved residues of 6H-QseM likely associated with antiactivation (high mobility). All residues are conserved with extensive steric freedoms and are surface exposed. Mobility can be observed by higher positional range the sidechains can exist in.

Not tested, but of interest were residues Met38 and Met83 (Figure 6.18). Both are solvent exposed, are residues commonly involved in protein-protein interactions and share a similar surface space with the Trp82. Compared to the Trp82, both Met residues have more mobility, however they are conserved within the QseM homologues and create a distinct surface space that may be the critical interface for binding and inhibiting both FseA and TraR. Met83 is conserved in all QseM homologues except *Sphingomonas sp.* (SKA58, ICE1), as previously noted, is unable to inactivate FseA and has an isoleucine in this position instead of a methionine. This suggests that Met83 may participate in inactivation.

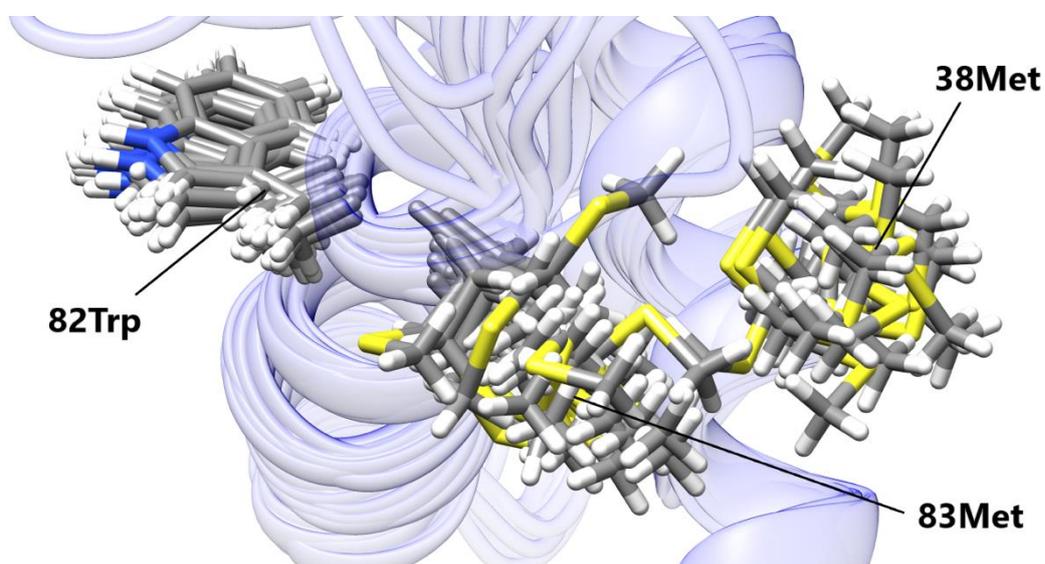


Figure 6.18: Surface-exposed conserved residues of 6H-QseM likely required for inactivation, Trp82, Met38 and Met83. Residues Met38, Trp82 and Met83 are displayed in all structures of the NMR ensemble. Mobility can be observed by higher positional range the sidechains can exist in.

Similar to Met83, a similar observation is seen with residue Arg39, which is conserved in all QseM homologues except *Sphingomonas sp.* (SKA58, ICE1). This particular residue has high mobility as seen in Figure 6.18 and is in proximity to the hydrophobic Met38 and Thr35 which may contribute to protein-binding (Figure 6.19). The proximal Asp42 and Ala43 are highly conserved, with possibility the Arg39 and Asp42 may form part of a salt-bridge exchange for protein-binding, however the Ala43 is likely more

necessary for maintaining the consistent 'core' helix motif described in 5.2.5. Aside from Leu41, all other remaining residues of the highly conserved 37-YMRLLEDA-43 region of 6H-QseM appear to be embedded and reinforce the surface exposed positioning of Met38, Arg39 and Asp42.

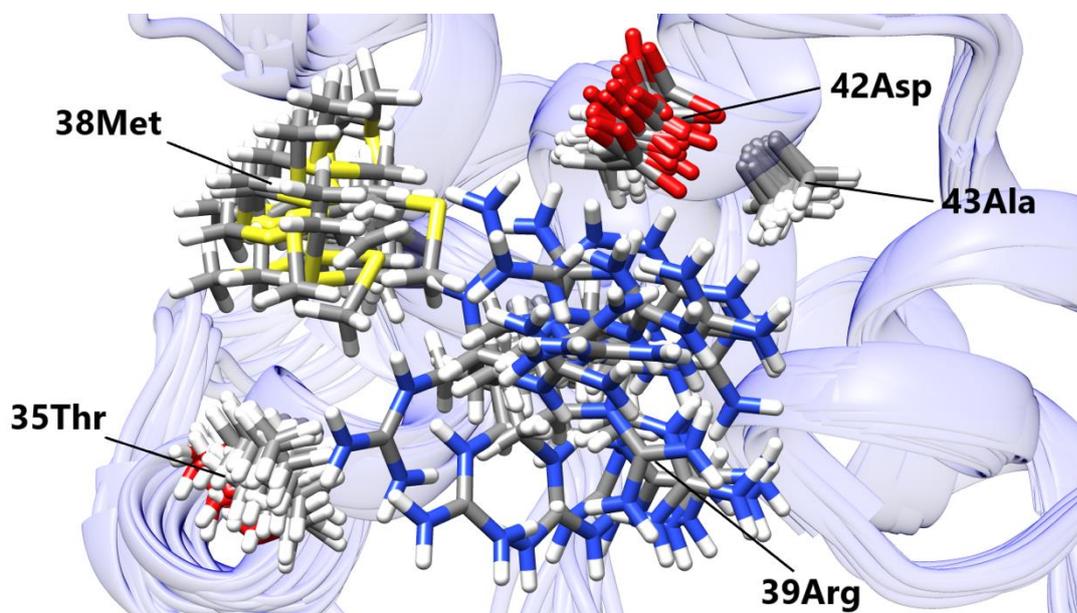


Figure 6.19: Residues Arg39 and Asp42 in the 6H-QseM NMR ensemble. Surface exposed residues Met38, Arg39 and Asp42 may be critical for antiactivation function.

Additionally, on the opposite end of the 6H-QseM, Asp61 is surface exposed and necessary for antiactivation against FseA (Figure 6.20). Both Asp61 and Pro62 are conserved among all QseM homologues, however alanine substitution of only Asp61 appeared to interfere with antiactivation function. It is likely that the Pro62 is conserved to maintain secondary structure as a common alpha helical initiator or terminator, however the alanine substitution may not severely disrupt this function and resulted in no obvious effect on function. The neighbouring Gly59 is also conserved in all QseM homologues except *Spingomonas sp.* (SKA58, ICE1). It is likely this residue is conserved to either minimise steric effects, maintaining specific structure motif or to correctly position the apparently critical Asp61 surface exposed for antiactivation function. Similar to Asp42, it is possible that Asp61 may form salt bridges for protein-protein binding or contribute in antiactivation in some other way.

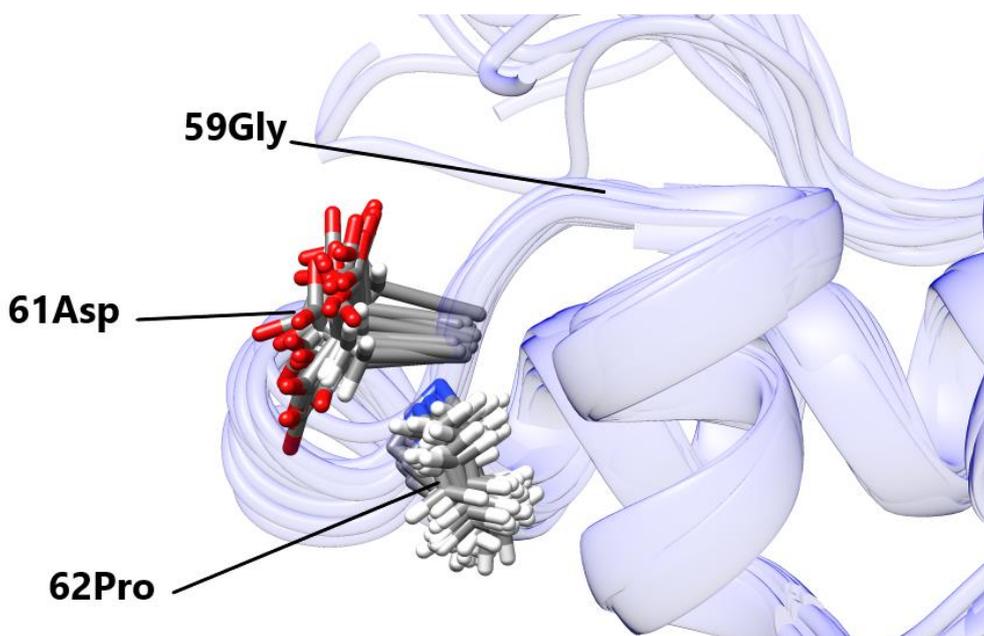


Figure 6.20: Residues of 6H-QseM likely required for antiactivation, Gly59, Asp61 and Pro62. Asp61 is surface exposed and necessary for antiactivation against FseA.

Due to purification difficulties of FseA as outlined in chapter 3, we were unable to utilise ^{15}N HSQC profile of 6H-QseM to highlight residues likely critical to binding with FseA. This ^{15}N HSQC profile would likely show chemical shift in residues on 6H-QseM that are directly interacting with FseA. We would expect those residues with chemical shift changes to overlap between residues highlighted in the mutagenesis analysis, however, additional residues may be revealed, like the suspected Met38, Arg39 and Asp42. It may also show some residues that have been determined critical to function in the mutagenesis assays but don't have chemical shift changes because they are for structural purpose rather than directly interacting with FseA.

6.3 Conclusion:

We have shown that 6H-QseM contains three alpha helices (H1, H2, H3) with a HTH motif existing between H2 and H3. This HTH motif is associated with DNA-binding proteins like transcriptional activators. QseM antiactivates FseA, a transcriptional activator, and both proteins belong to the DUF2285 protein family. Structural predictions of FseA indicate the DUF2285 domain containing portion of FseA (residues 196-266) shares high structural similarity with all generated models of 6H-QseM. Residues critical to DNA binding by FseA are located on its HTH motif (Will Jowsey, unpublished). Both proteins display a HTH motif, however, the 'turn' in FseA is much larger than that of QseM. The surface charge of the DUF2285 portion of FseA which includes the HTH motif is also predominantly positive. However, QseM has an overall neutral surface charge, with a slightly more negative surface charge area located around the H1 and H2 connection (residues 42-52) that would normally be associated with DNA binding by a protein with a HTH motif. We concluded that the differences in the 'turn' feature and surface charge between QseM and FseA critically influence their ability to interact with DNA.

Mutagenesis work on QseM was completed by Calum Morris and mutagenesis work on FseA was completed by Will Jowsey. Targeted single residue mutations of both QseM and FseA were tested for function using bacterial two-hybrid and lacZ assays. A region on QseM where protein binding likely occurs was identified based on a concentrated number surface exposed residues critical to antiactivation function. Specifically, TraR antiactivation is abolished when residues Asp29, Tyr37, Leu39, Arg79, Trp82 are substituted to alanine. FseA antiactivation is abolished when Arg38, Leu39, Asp42, Trp82 are substituted to alanine and significantly reduced when Leu26, Thr27, Asp29, Tyr37, Leu40, Asp61, His76, Leu77, Arg79 are substituted to alanine. Many of these residues that significantly reduce FseA antiactivation have no observable effect on QseM's ability to antiactivate TraR. A number of alanine substitutions outside of this suspected binding-site also resulted in knocking out antiactivation function, however, when mapped on the NMR ensemble, it became apparent that these are likely required to maintain structure, which could be affected during these substitutions. None of the residues determined to be critical to the antiactivation function of QseM are associated with the HTH motif. Since a greater number of QseM's residues appear to be critical to antiactivate FseA instead of TraR,

there is a possibility that QseM has evolved to primarily antiactivate FseA and gained antiactivation of TraR as a secondary function. This is supported by the observation that QseM antiactivates FseA to a greater effect than that of TraR (Calum Morris, unpublished).

Substitution assays are not complete, with the majority that have been done being limited to alanine substitutions. There is potential to use the generated 6H-QseM structures to direct further substitutions. These include targeting residues that are in proximity to the already flagged critical residues, Trp82, Arg39 and Asp61. By successfully guiding experimental work based on our structure and generating expected results, we may also reinforce the validity of the 6H-QseM structure. Future experimental work should also include co-crystallisation of QseM bound to either FseA or TraR. This method was employed for both QslA and TraM to demonstrate their mechanism of antiactivation against their respective target proteins.

Although the HTH motif is heavily investigated to date, we note that some variations to the HTH exist. An example of this is the Q antitermination factor of *E. coli* as discussed in this chapter. The 'turn' component of most HTH motifs often exists as a simple short chain loop between antiparallel α -helices, instead Q antitermination factor incorporates additional secondary structure comprising two antiparallel β -sheets in making it novel. Similarly, the HTH of DUF2285 in FseA and QseM has an extended turn relative to many other HTH motifs (as observed by *DALI* structure alignments). But unlike Q antitermination factor, the extended turn of DUF2285 is only a loop in QseM, and more strikingly, a small α -helix in FseA. This shows that the HTH of DUF2285 is novel in structure. It is also likely that the significant variations to the turn between QseM and FseA is critically relevant to functional discrepancy of the DUF2285 in FseA being able to bind DNA whereas the DUF2285 in QseM is not. It may be that FseA is a more accurate representation of DUF2285 since it also more closely resembles RovC, a distant DUF2285 relative. To that end, we conclude that QseM is an evolutionarily paralog of FseA with a DUF2285 with a novel HTH motif that is incapable of binding DNA that functions as an FseA antiactivator.

Chapter 7:

Final discussion and conclusions

Several LuxR/I-homologue QS systems have evolved antiactivation proteins like TraM and QIsA that bind and inhibit LuxR proteins. Antiactivators prevent their transcriptional activators from activating transcription from target promoters. The QS system of *M. luti* R7A encodes an antiactivator, QseM, that binds and inhibits two different transcriptional activators, TraR and FseA. The work presented in this thesis has detailed attempts to solve the structure of QseM by X-ray crystallography and NMR and investigate the structure of QseM in conjunction with residue-mutagenesis results completed by collaborators. Our findings have revealed QseM to contain a unique HTH motif. The QseM protein shares sequence similarity and structural similarity to the predicted DUF2285 domain of FseA. However, QseM lacks a significantly positively charged surface in the HTH region characteristic of DNA-binding proteins. Mutagenesis by colleagues revealed that substitutions in amino acids core to the structure of QseM abolished function as expected but substitutions in residues surrounding the 'turn' region and possible DNA-recognition helix of the QseM structure had little impact on antiactivation of TraR and FseA. Instead, substitutions in the exposed Asp31 and Trp71, at the opposite side of the intersection between H1 and H3 were critical for QseM antiactivation and binding to both TraR and FseA. In summary, QseM while resembles a HTH motif in structure it seems unlikely to be a DNA-binding protein and likely uses the opposite side HTH structure to interface with both TraR and FseA.

We found that 6H-QseM could be expressed and purified, with purity and yield to levels required for X-ray crystallography and NMR. This includes 6H-QseM with incorporated L-selenomethionine and $^{13}\text{C}/^{15}\text{N}$ enrichment. 6H-QseM was purified and concentrated for NMR and crystallisation experiments. Future experiments involving purified QseM constructs may benefit from further buffer optimisation to either improve crystallisation or increase the apparent solubility limit for increased protein concentration or experimental runtime available for NMR data collection. Removal of the 6H tag of QseM was discussed in chapter 3 but was not thoroughly investigated. Modification of the QseM construct applied to crystallisation screens is likely to have significant effects on crystal formation conditions and morphology. Future crystallisation attempts with QseM should extend to a more focused attempt at tag removal, changing attachment from the N-terminal to the C-terminal of QseM. Further experiments with QseM in solution should include analytical SEC to provide stoichiometric information of QseM

or QseM+FseA or QseM+TraR samples and complement attempts to solve the structure of QseM bound to FseA or QseM. Similarly, isothermal titration colometry (ITC) and microscale thermophoresis (MST) are comparatively simple experiments to attempt in combination with analytical SEC.

Despite belonging to the same DUF2285 family as QseM, FseA proved to be difficult to express and purify. Quantities of stable FseA protein required for X-ray crystallography were unattainable, however, improvements to FseA purification methods subsequently revealed FseA to purify as a dimer (Will Jowsey, unpublished). Further attempts at purified FseA should include creating fusions with other tags like glutathione S-transferase (GST), green fluorescent protein (GFP) or thioredoxin (Trx), changing the tag fusion site or expressing the N- and C- terminal domains of FseA separately. Since QseM and the C-terminal domain of FseA are both DUF2285 members of similar size, the FseA DUF2285 domain may more closely behave like QseM during expression and purification. Crystallographic structure determination of the FseA DUF2285 domain may succeed where QseM did not and could further add to structural characterisation of the DUF2285.

Only one crystal morphology was observed for 6H-QseM. One condition was optimised to give the best diffracting 6H-QseM crystals. We were unable to solve the structure of 6H-QseM by X-ray crystallography, with the primary reason being attributed to poor diffraction data. It appeared that Se incorporation was present based on collected anomalous signal, however, this anomalous signal was not as strong as expected. Diffraction data quality was negatively impacted by a several factors. As crystallisation was optimised, diffraction resolution increased, however, higher resolution revealed evidence of crystal-splitting or multiple crystals being present within a single dataset. These crystal flaws likely reduced anomalous signal strength and negatively impacted the overall diffraction data collected. Occasional instances of lower symmetry space group indexing within the higher resolution SeMet-optimised datasets. It is likely that this lower symmetry indexing may have occurred from difficulties determining systematic absences in these datasets. The corresponding lower symmetry may have resulted in false positive twinning results.

Ultimately, a structure was not solved via X-ray crystallography using anomalous diffraction data. Further attempts were made to solve the structure using predicted

models of 6H-QseM by molecular replacement also failed to produce a solution. Failure using predicted phase information reinforced the theory that crystal morphology may be the primary issue or that protein prediction methods were not accurately modelling QseM. Further attempts at crystallisation of QseM, in conjunction with modifications to constructs expressed and purification considerations discussed in 7.1, require re-screening crystallisation conditions. Generation of different crystals with similar or different morphologies could give overall better diffraction data and perhaps reduce issues observed with 6H-QseM crystals generated in this thesis. Further attempts should be made to crystallise QseM homologues, FseA, the FseA DUF2285 domain, FseA homologues or some other DUF2285 protein. A successful crystal structure from any of these proteins may contribute to our understanding of QseM, FseA, or other proteins containing DUF2285 domains. Successful crystallisation of QseM or FseA may also be the catalyst for co-crystallisation of both proteins bound, revealing details regarding the binding-interface of QseM.

We determined that 6H-QseM was a suitable candidate for collecting NMR data based on its stability in low ionic strength buffers and protein concentrations amenable to protein NMR experiments. Initial NMR experiments focused on backbone assignment before attempting full sidechain proton assignment required for subsequent structure generation. The X-ray crystallography data could not be phased using this structure. This was not improved by using *ROSETTA* to refine the initial NMR ensemble.

Our NMR-derived structure, 6H-QseM-NMR, was structurally similar to co-evolution (6H-QseM-CoEvo) and *CS-ROSETTA* (6H-QseM-CSROSETTA) predicted structures. All three methods generated structures of 6H-QseM that have three alpha helices (H1, H2, H3) between residues 26 and 82 with a HTH motif spanning the H2 and H3. The agreement between all three solution structures suggests that the true structure of QseM is likely a triple-helix protein with a HTH motif with all three helices in the orientations determined in this thesis. Structural variations exist in the extreme termini and either remain poorly defined or are highly mobile elements of the protein and are of less significance regarding structure determination. Additionally, 6H-QseM was determined to be monomeric in solution by NMR and SAXS data. However, this does not imply that QseM is not capable of forming a homo-complex during antiactivation as observed with the characterised antiactivators, TraM or QsIA.

Further NMR experiments should build off the annotated ^{15}N HSQC of 6H-QseM. Titration of FseA or TraR+HSL into an NMR sample containing 6H-QseM would likely result chemical shifts of atoms associated with residues in QseM that are in contact with either FseA or TraR. It would be expected that these chemical shifts will correlate with residues already established as critical for antiactivation, however, additional residues may also be revealed.

Based on investigation of the structure of 6H-QseM, we can make some final observations. The first is that conservation of residues between QseM and all DUF2285 domains identified by Pfam seem to be mostly associated with maintaining structure, specifically, a HTH motif that may exist in all DUF2285 domains.

Secondly, there is very low conservation of residues that are solvent exposed on the 6H-QseM-NMR structure. Importantly, this low conservation is shown when looking at alignments of all proteins across the DUF2285 family, however, some of these solvent exposed residues were revealed to be conserved when only QseM homologues were aligned. Residues conserved within the sub-group of only QseM homologues are likely related to the specific function of QseM and its homologues. This is in agreement with our proposed binding-site location on 6H-QseM-NMR.

Thirdly, three distinct regions of 6H-QseM structure were flagged as being important to antiactivation function. Mutagenesis analysis of QseM in protein-protein assays found residues Leu26, Thr27, Asp29, Tyr37, Arg38, Leu39, Leu40, Asp42, Asp61, His76, Leu77, Arg79, Trp82, are critical to antiactivation against FseA and TraR (determined by Calum Moris and Will Jowsey, unpublished work). These were mapped on the 6H-QseM structure revealing a likely protein binding site required for antiactivation (Figure 7.1). Neighbouring this was a region of significant negative surface charge which may be critical to binding FseA and TraR or disrupting the DNA binding of either of these.

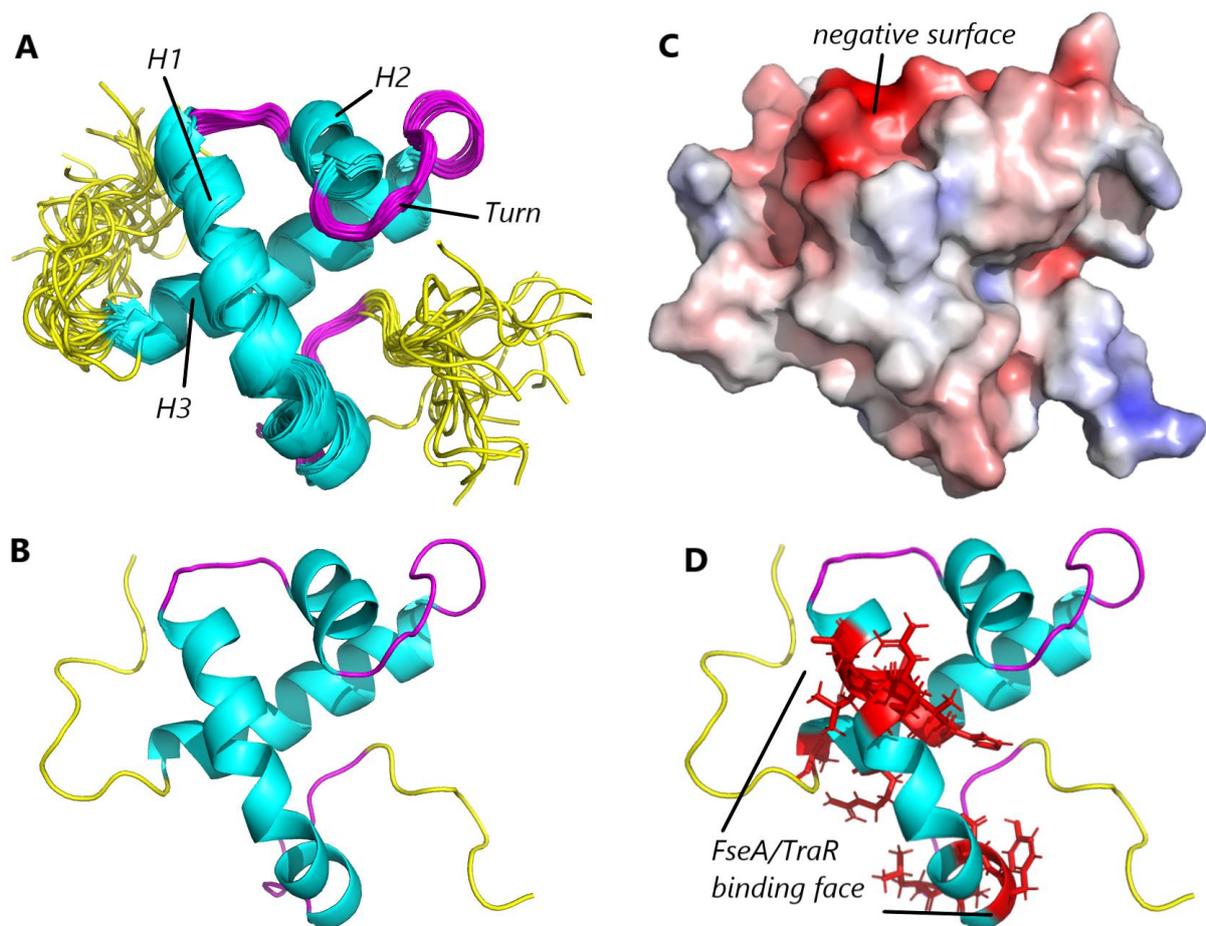


Figure 7.1: Overall observations of newly determined 6H-QseM structure. Structure ensemble generated via *CYANA* with N-terminal tag residues 1-10 removed. (A) Ensemble coloured by secondary structure (helix as cyan, coil as magenta). Residues 11-18 and 83-94 are discarded (yellow). The overall structure contains 3 helices (H1, H2 and H3). (B) The H2, Turn and H3 form a helix-turn-helix motif. The best representative structure of the ensemble is shown. (C) A surface model of the best representative structure is shown with a prominent negatively charged surface at the C-terminal end of H1 and the coil leading the H2. (D) Residues critical to antiactivation against FseA (red) are shown and present a likely FseA/TraR binding face that spans H1 and the C-terminal end of H3.

Finally, structural alignment of results from *DALI* and *COFACTOR* suggests that QseM and other DUF2285 members are most closely related to DNA binding proteins. This is in agreement with FseA being a transcriptional activator and may suggest that QseM has lost its DNA-binding function during evolution. A co-evolution-derived structure prediction of FseA was generated by Will Jowsey. The DUF2285 domain of FseA contains a similar triple-helix motif between residues 196-266, also displaying a HTH

motif. The 'turn' of this HTH motif in FseA is extended and contains an additional small helix. Our research suggests that the DUF2285 is a domain that contains a novel HTH motif for transcription regulation within the Alphaproteobacteria phylum.

The means by which QseM antiactivates FseA has not been explicitly defined, however, a theory can be extrapolated from the findings of this thesis. Both QseM and the DUF2285 domain of FseA likely bind the N-terminal domain (Msi172) of FseA. A FseA homodimer may form when the C-terminal domain of one chain binds the N-terminal domain of the other. However, QseM could prevent this dimerisation by binding the N-terminal domain of FseA. If FseA is unable to dimerise it may be unable to function as a transcriptional activator since many transcriptional regulators often require a dimerised state to successfully dock to their respective DNA target sequences (Figure 7.2).

QseM targets TraR in the presence of its signal molecule, 3-oxo-C6-HSL. It is presumed that TraR is in a dimer state when this signal molecule is present, therefore, QseM only binds a TraR homodimer. Extremely similar residues in QseM are required for the binding and antiactivation of both FseA and TraR and thus that the binding face of QseM is the same for both, however, the site on TraR where this face docks remains elusive.

To conclude, structure determination of QseM by NMR revealed the DUF2285 domain to be a HTH motif capable of antiactivation or, as is the case of FseA and RovC, promote gene transcription. Discrepancy in function between these proteins is likely attributed to distinct structural changes to the HTH motif. This thesis has set the scene for further investigation of the dual-target antiactivator QseM and provide insight into the structure and function of the DUF2285 domain.

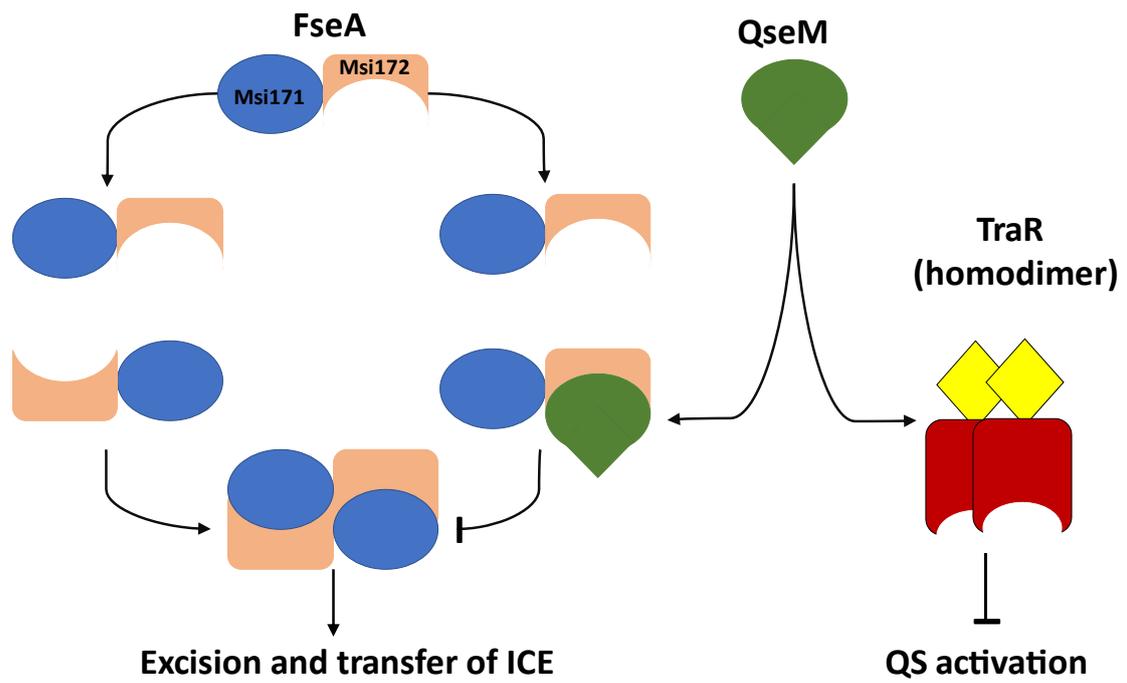


Figure 7.2: Proposed antiactivation method of QseM. QseM (green) prevents FseA forming a dimer capable of promoting gene expression for ICE excision and transfer. How QseM binds TraR is not known, however, it only targets TraR in a dimer state.

Chapter 8:

References

1. Reményi A, Schöler HR, Wilmanns M. Combinatorial control of gene expression. *Nature structural & molecular biology*. 2004;11(9):812-5.
2. Feugeas JP, Tourret J, Launay A, Bouvet O, Hoede C, Denamur E, et al. Links between Transcription, Environmental Adaptation and Gene Variability in *Escherichia coli*: Correlations between Gene Expression and Gene Variability Reflect Growth Efficiencies. *Molecular biology and evolution*. 2016;33(10):2515-29.
3. Paget MS. Bacterial Sigma Factors and Anti-Sigma Factors: Structure, Function and Distribution. *Biomolecules*. 2015;5(3):1245-65.
4. Saecker RM, Record MT, Jr., Dehaseth PL. Mechanism of bacterial transcription initiation: RNA polymerase - promoter binding, isomerization to initiation-competent open complexes, and initiation of RNA synthesis. *J Mol Biol*. 2011;412(5):754-71.
5. van Hijum SAFT, Medema MH, Kuipers OP. Mechanisms and Evolution of Control Logic in Prokaryotic Transcriptional Regulation. *Microbiology and Molecular Biology Reviews*. 2009;73(3):481.
6. Aravind L, Anantharaman V, Balaji S, Babu MM, Iyer LM. The many faces of the helix-turn-helix domain: Transcription regulation and beyond*. *FEMS microbiology reviews*. 2005;29(2):231-62.
7. Matthews BW, Ohlendorf DH, Anderson WF, Takeda Y. Structure of the DNA-binding region of lac repressor inferred from its homology with cro repressor. *Proceedings of the National Academy of Sciences of the United States of America*. 1982;79(5):1428-32.
8. Brennan RG, Matthews BW. The helix-turn-helix DNA binding motif. *The Journal of biological chemistry*. 1989;264(4):1903-6.
9. Wintjens R, Rooman M. Structural classification of HTH DNA-binding domains and protein-DNA interaction modes. *J Mol Biol*. 1996;262(2):294-313.
10. Huffman JL, Brennan RG. Prokaryotic transcription regulators: more than just the helix-turn-helix motif. *Current Opinion in Structural Biology*. 2002;12(1):98-106.
11. Hinrichs W, Kisker C, Duvel M, Muller A, Tovar K, Hillen W, et al. Structure of the Tet repressor-tetracycline complex and regulation of antibiotic resistance. *Science*. 1994;264(5157):418.
12. Funnell AP, Crossley M. Homo- and heterodimerization in transcriptional regulation. *Advances in experimental medicine and biology*. 2012;747:105-21.
13. Amoutzias GD, Robertson DL, Van de Peer Y, Oliver SG. Choose your partners: dimerization in eukaryotic transcription factors. *Trends in Biochemical Sciences*. 2008;33(5):220-9.
14. Marianayagam NJ, Sunde M, Matthews JM. The power of two: protein dimerization in biology. *Trends Biochem Sci*. 2004;29(11):618-25.
15. Weiss MA, Pabo CO, Karplus M, Sauer RT. Dimerization of the operator binding domain of phage λ repressor. *Biochemistry*. 1987;26(3):897-904.
16. Fan H, Dong Y, Wu D, Bowler MW, Zhang L, Song H. QsIA disrupts LasR dimerization in antiactivation of bacterial quorum sensing. *Proceedings of the National Academy of Sciences of the United States of America*. 2013;110(51):20765-70.
17. Qin Y, Luo ZQ, Smyth AJ, Gao P, Beck von Bodman S, Farrand SK. Quorum-sensing signal binding results in dimerization of TraR and its release from membranes into the cytoplasm. *The EMBO journal*. 2000;19(19):5212-21.
18. Beamer LJ, Pabo CO. Refined 1.8 Å crystal structure of the lambda repressor-operator complex. *J Mol Biol*. 1992;227(1):177-96.
19. Miller MB, Bassler BL. Quorum Sensing in Bacteria. *Annual Review of Microbiology*. 2001;55(1):165-99.
20. Waters CM, Bassler BL. QUORUM SENSING: Cell-to-Cell Communication in Bacteria. *Annual Review of Cell and Developmental Biology*. 2005;21(1):319-46.
21. Nealson KH. Autoinduction of bacterial luciferase. Occurrence, mechanism and significance. *Archives of microbiology*. 1977;112(1):73-9.
22. Dunn AK. Chapter Two - *Vibrio fischeri* Metabolism: Symbiosis and Beyond. In: Poole RK, editor. *Advances in Microbial Physiology*. 61: Academic Press; 2012. p. 37-68.

23. Eberhard A. Inhibition and Activation of Bacterial Luciferase Synthesis. *J Bacteriol.* 1972;109(3):1101.
24. Eberhard A, Burlingame AL, Eberhard C, Kenyon GL, Nealson KH, Oppenheimer NJ. Structural identification of autoinducer of *Photobacterium fischeri* luciferase. *Biochemistry.* 1981;20(9):2444-9.
25. Fuqua WC, Winans SC, Greenberg EP. Quorum sensing in bacteria: the LuxR-LuxI family of cell density-responsive transcriptional regulators. *J Bacteriol.* 1994;176(2):269-75.
26. Qin N, Callahan SM, Dunlap PV, Stevens AM. Analysis of LuxR Regulon Gene Expression during Quorum Sensing in *Vibrio fischeri*. *J Bacteriol.* 2007;189(11):4127.
27. Tsai CS, Winans SC. LuxR-type quorum-sensing regulators that are detached from common scents. *Mol Microbiol.* 2010;77(5):1072-82.
28. Devine JH, Shadel GS, Baldwin TO. Identification of the operator of the lux regulon from the *Vibrio fischeri* strain ATCC7744. *Proceedings of the National Academy of Sciences of the United States of America.* 1989;86(15):5688-92.
29. Devine JH, Shadel GS, Baldwin TO. Identification of the operator of the lux regulon from the *Vibrio fischeri* strain ATCC7744. *Proceedings of the National Academy of Sciences.* 1989;86(15):5688.
30. Engebrecht J, Silverman M. Identification of genes and gene products necessary for bacterial bioluminescence. *Proceedings of the National Academy of Sciences of the United States of America.* 1984;81(13):4154-8.
31. de Kievit TR, Iglewski BH. Bacterial Quorum Sensing in Pathogenic Relationships. *Infection and immunity.* 2000;68(9):4839.
32. Banerjee G, Ray AK. Quorum-sensing network-associated gene regulation in Gram-positive bacteria. *Acta microbiologica et immunologica Hungarica.* 2017;64(4):439-53.
33. Papenfort K, Bassler BL. Quorum sensing signal-response systems in Gram-negative bacteria. *Nat Rev Microbiol.* 2016;14(9):576-88.
34. Rutherford ST, Bassler BL. Bacterial quorum sensing: its role in virulence and possibilities for its control. *Cold Spring Harbor perspectives in medicine.* 2012;2(11).
35. Ramsay JP, Sullivan JT, Jambari N, Ortori CA, Heeb S, Williams P, et al. A LuxRI-family regulatory system controls excision and transfer of the *Mesorhizobium loti* strain R7A symbiosis island by activating expression of two conserved hypothetical genes. *Molecular Microbiology.* 2009;73(6):1141-55.
36. Schaefer AL, Val DL, Hanzelka BL, Cronan JE, Greenberg EP. Generation of cell-to-cell signals in quorum sensing: acyl homoserine lactone synthase activity of a purified *Vibrio fischeri* LuxI protein. *Proceedings of the National Academy of Sciences.* 1996;93(18):9505-9.
37. Ng WL, Bassler BL. Bacterial quorum-sensing network architectures. *Annu Rev Genet.* 2009;43:197-222.
38. Bacon Schneider K, Palmer TM, Grossman AD. Characterization of comQ and comX, two genes required for production of ComX pheromone in *Bacillus subtilis*. *J Bacteriol.* 2002;184(2):410-9.
39. Zhao L, Xue T, Shang F, Sun H, Sun B. *Staphylococcus aureus* AI-2 Quorum Sensing Associates with the KdpDE Two-Component System To Regulate Capsular Polysaccharide Synthesis and Virulence. *Infection and immunity.* 2010;78(8):3506-15.
40. Frederix M, Downie JA. Chapter 2 - Quorum Sensing: Regulating the Regulators. In: Poole RK, editor. *Advances in Microbial Physiology.* 58: Academic Press; 2011. p. 23-80.
41. Eberhard A, Longin T, Widrig CA, Stranick SJ. Synthesis of the lux gene autoinducer in *Vibrio fischeri* is positively autoregulated. *Archives of Microbiology.* 1991;155(3):294-7.
42. Moré MI, Finger LD, Stryker JL, Fuqua C, Eberhard A, Winans SC. Enzymatic Synthesis of a Quorum-Sensing Autoinducer Through Use of Defined Substrates. *Science.* 1996;272(5268):1655-8.
43. Liao L, Schaefer AL, Coutinho BG, Brown PJB, Greenberg EP. An aryl-homoserine lactone quorum-sensing signal produced by a dimorphic prosthecate bacterium. *Proceedings of the National Academy of Sciences.* 2018;115(29):7587-92.

44. Whiteley M, Diggle SP, Greenberg EP. Progress in and promise of bacterial quorum sensing research. *Nature*. 2017;551(7680):313-20.
45. Schaefer AL, Greenberg EP, Oliver CM, Oda Y, Huang JJ, Bittan-Banin G, et al. A new class of homoserine lactone quorum-sensing signals. *Nature*. 2008;454(7204):595-9.
46. Fuqua C, Parsek MR, Greenberg EP. Regulation of Gene Expression by Cell-to-Cell Communication: Acyl-Homoserine Lactone Quorum Sensing. *Annu Rev Genet*. 2001;35(1):439-68.
47. Throup JP, Camara M, Briggs GS, Winson MK, Chhabra SR, Bycroft BW, et al. Characterisation of the *yenI/yenR* locus from *Yersinia enterocolitica* mediating the synthesis of two N-acylhomoserine lactone signal molecules. *Mol Microbiol*. 1995;17(2):345-56.
48. Ng YK, Grasso M, Wright V, Garcia V, Williams P, Atkinson S. The Quorum Sensing System of *Yersinia enterocolitica* 8081 Regulates Swimming Motility, Host Cell Attachment, and Virulence Plasmid Maintenance. *Genes*. 2018;9(6).
49. LaSarre B, Federle MJ. Exploiting quorum sensing to confuse bacterial pathogens. *Microbiol Mol Biol Rev*. 2013;77(1):73-111.
50. Dubern JF, Diggle SP. Quorum sensing by 2-alkyl-4-quinolones in *Pseudomonas aeruginosa* and other bacterial species. *Molecular bioSystems*. 2008;4(9):882-8.
51. Jimenez PN, Koch G, Thompson JA, Xavier KB, Cool RH, Quax WJ. The multiple signaling systems regulating virulence in *Pseudomonas aeruginosa*. *Microbiol Mol Biol Rev*. 2012;76(1):46-65.
52. Von Bodman SB, Bauer WD, Coplin DL. Quorum sensing in plant-pathogenic bacteria. *Annu Rev Phytopathol*. 2003;41:455-82.
53. Barnard AM, Salmond GP. Quorum sensing in *Erwinia* species. *Analytical and bioanalytical chemistry*. 2007;387(2):415-23.
54. McClean KH, Winson MK, Fish L, Taylor A, Chhabra SR, Camara M, et al. Quorum sensing and *Chromobacterium violaceum*: exploitation of violacein production and inhibition for the detection of N-acylhomoserine lactones. *Microbiology*. 1997;143(12):3703-11.
55. Labbate M, Queck SY, Koh KS, Rice SA, Givskov M, Kjelleberg S. Quorum sensing-controlled biofilm development in *Serratia liquefaciens* MG1. *J Bacteriol*. 2004;186(3):692-8.
56. Milton DL. Quorum sensing in vibrios: complexity for diversification. *International journal of medical microbiology : IJMM*. 2006;296(2-3):61-71.
57. Thoendel M, Kavanaugh JS, Flack CE, Horswill AR. Peptide signaling in the staphylococci. *Chemical reviews*. 2011;111(1):117-51.
58. Nakayama J, Cao Y, Horii T, Sakuda S, Akkermans AD, de Vos WM, et al. Gelatinase biosynthesis-activating pheromone: a peptide lactone that mediates a quorum sensing in *Enterococcus faecalis*. *Mol Microbiol*. 2001;41(1):145-54.
59. Rocha-Estrada J, Aceves-Diez AE, Guarneros G, de la Torre M. The RNPP family of quorum-sensing proteins in Gram-positive bacteria. *Appl Microbiol Biotechnol*. 2010;87(3):913-23.
60. Guiral S, Mitchell TJ, Martin B, Claverys JP. Competence-programmed predation of noncompetent cells in the human pathogen *Streptococcus pneumoniae*: genetic requirements. *Proceedings of the National Academy of Sciences of the United States of America*. 2005;102(24):8710-5.
61. Havarstein LS, Coomaraswamy G, Morrison DA. An unmodified heptadecapeptide pheromone induces competence for genetic transformation in *Streptococcus pneumoniae*. *Proceedings of the National Academy of Sciences of the United States of America*. 1995;92(24):11140-4.
62. Kowalko JE, Seibert ME. The *Streptococcus pneumoniae* competence regulatory system influences respiratory tract colonization. *Infection and immunity*. 2008;76(7):3131-40.
63. Dubois T, Faegri K, Perchat S, Lemy C, Buisson C, Nielsen-LeRoux C, et al. Necrotrophism is a quorum-sensing-regulated lifestyle in *Bacillus thuringiensis*. *PLoS pathogens*. 2012;8(4):e1002629.
64. Rocha J, Flores V, Cabrera R, Soto-Guzman A, Granados G, Juaristi E, et al. Evolution and some functions of the NprR-NprRB quorum-sensing system in the *Bacillus cereus* group. *Appl Microbiol Biotechnol*. 2012;94(4):1069-78.

65. Deng Y, Wu J, Tao F, Zhang LH. Listening to a new language: DSF-based quorum sensing in Gram-negative bacteria. *Chemical reviews*. 2011;111(1):160-73.
66. Vannini A, Volpari C, Di Marco S. Crystal structure of the quorum-sensing protein TraM and its interaction with the transcriptional regulator TraR. *The Journal of biological chemistry*. 2004;279(23):24291-6.
67. Stratev D, Odeyemi OA. An overview of motile *Aeromonas* septicaemia management. *Aquaculture International*. 2017;25(3):1095-105.
68. Thornley J, Shaw J, Gryllos I, Eley A. Virulence properties of clinically significant *Aeromonas* species: Evidence for pathogenicity. *Reviews in Medical Microbiology*. 1997;8.
69. Swift S, Lynch MJ, Fish L, Kirke DF, Tomás JM, Stewart GS, et al. Quorum sensing-dependent regulation and blockade of exoprotease production in *Aeromonas hydrophila*. *Infection and immunity*. 1999;67(10):5192-9.
70. Swift S, Karlyshev AV, Fish L, Durant EL, Winson MK, Chhabra SR, et al. Quorum sensing in *Aeromonas hydrophila* and *Aeromonas salmonicida*: identification of the LuxRI homologs AhyRI and AsaRI and their cognate N-acylhomoserine lactone signal molecules. *J Bacteriol*. 1997;179(17):5271-81.
71. Pearson JP, Pesci EC, Iglewski BH. Roles of *Pseudomonas aeruginosa* las and rhl quorum-sensing systems in control of elastase and rhamnolipid biosynthesis genes. *J Bacteriol*. 1997;179(18):5756-67.
72. Dekimpe V, Déziel E. Revisiting the quorum-sensing hierarchy in *Pseudomonas aeruginosa*: the transcriptional regulator RhIR regulates LasR-specific factors. *Microbiology*. 2009;155(3):712-23.
73. Schuster M, Peter Greenberg E. A network of networks: Quorum-sensing gene regulation in *Pseudomonas aeruginosa*. *International Journal of Medical Microbiology*. 2006;296(2):73-81.
74. Michael B, Smith JN, Swift S, Heffron F, Ahmer BM. SdiA of *Salmonella enterica* is a LuxR homolog that detects mixed microbial communities. *J Bacteriol*. 2001;183(19):5733-42.
75. Nguyen Y, Nguyen NX, Rogers JL, Liao J, MacMillan JB, Jiang Y, et al. Structural and Mechanistic Roles of Novel Chemical Ligands on the SdiA Quorum-Sensing Transcription Regulator. *MBio*. 2015;6(2):e02429-14.
76. Atkinson S, Sockett RE, Cámara M, Williams P. Quorum sensing and the lifestyle of *Yersinia*. *Current issues in molecular biology*. 2006;8(1):1-10.
77. Cha C, Gao P, Chen Y-C, Shaw PD, Farrand SK. Production of Acyl-Homoserine Lactone Quorum-Sensing Signals by Gram-Negative Plant-Associated Bacteria. *Molecular Plant-Microbe Interactions*[®]. 1998;11(11):1119-29.
78. Gordon JE, Christie PJ. The *Agrobacterium* Ti Plasmids. *Microbiol Spectr*. 2014;2(6):10.1128/microbiolspec.PLAS-0010-2013.
79. Piper KR, Beck von Bodman S, Farrand SK. Conjugation factor of *Agrobacterium tumefaciens* regulates Ti plasmid transfer by autoinduction. *Nature*. 1993;362(6419):448-50.
80. Zhang L, Murphy PJ, Kerr A, Tate ME. *Agrobacterium* conjugation and gene regulation by N-acyl-L-homoserine lactones. *Nature*. 1993;362(6419):446-8.
81. Toth IK, Birch PRJ. Rotting softly and stealthily. *Current Opinion in Plant Biology*. 2005;8(4):424-9.
82. Pöllumaa L, Alamäe T, Mäe A. Quorum sensing and expression of virulence in pectobacteria. *Sensors (Basel)*. 2012;12(3):3327-49.
83. Bainton NJ, Stead P, Chhabra SR, Bycroft BW, Salmond GP, Stewart GS, et al. N-(3-oxohexanoyl)-L-homoserine lactone regulates carbapenem antibiotic production in *Erwinia carotovora*. *The Biochemical journal*. 1992;288 (Pt 3)(Pt 3):997-1004.
84. Ramachandran R, Stevens AM. Proteomic Analysis of the Quorum-Sensing Regulon in *Pantoea stewartii* and Identification of Direct Targets of EsaR. *Appl Environ Microbiol*. 2013;79(20):6244.

85. Beck von Bodman S, Farrand SK. Capsular polysaccharide biosynthesis and pathogenicity in *Erwinia stewartii* require induction by an N-acylhomoserine lactone autoinducer. *J Bacteriol.* 1995;177(17):5000-8.
86. Dubern JF, Lugtenberg BJ, Bloemberg GV. The ppul-rsaL-ppuR quorum-sensing system regulates biofilm formation of *Pseudomonas putida* PCL1445 by controlling biosynthesis of the cyclic lipopeptides putisolvins I and II. *J Bacteriol.* 2006;188(8):2898-906.
87. Rodelas B, Lithgow JK, Wisniewski-Dye F, Hardman A, Wilkinson A, Economou A, et al. Analysis of quorum-sensing-dependent control of rhizosphere-expressed (rhi) genes in *Rhizobium leguminosarum* bv. *viciae*. *J Bacteriol.* 1999;181(12):3816-23.
88. van Rhijn P, Vanderleyden J. The *Rhizobium*-plant symbiosis. *Microbiological Reviews.* 1995;59(1):124-42.
89. Giordano W. Rhizobial Extracellular Signaling Molecules and Their Functions in Symbiotic Interactions with Legumes. *Quorum Sensing vs Quorum Quenching: A Battle with no end in Sight.* 2015:123-32.
90. Laranjo M, Alexandre A, Oliveira S. Legume growth-promoting rhizobia: an overview on the *Mesorhizobium* genus. *Microbiological research.* 2014;169(1):2-17.
91. Cubo MT, Economou A, Murphy G, Johnston AW, Downie JA. Molecular characterization and regulation of the rhizosphere-expressed genes rhiABCR that can influence nodulation by *Rhizobium leguminosarum* biovar *viciae*. *J Bacteriol.* 1992;174(12):4026-35.
92. Zahran HH. Plasmids impact on rhizobia-legumes symbiosis in diverse environments. *Symbiosis.* 2017;73(2):75-91.
93. Masson-Boivin C, Sachs JL. Symbiotic nitrogen fixation by rhizobia—the roots of a success story. *Current Opinion in Plant Biology.* 2018;44:7-15.
94. González JE, Marketon MM. Quorum sensing in nitrogen-fixing rhizobia. *Microbiol Mol Biol Rev.* 2003;67(4):574-92.
95. Wilkinson A, Danino V, Wisniewski-Dyé F, Lithgow JK, Downie JA. N-acyl-homoserine lactone inhibition of rhizobial growth is mediated by two quorum-sensing genes that regulate plasmid transfer. *J Bacteriol.* 2002;184(16):4510-9.
96. Wisniewski-Dyé F, Jones J, Chhabra SR, Downie JA. railR genes are part of a quorum-sensing network controlled by cinI and cinR in *Rhizobium leguminosarum*. *J Bacteriol.* 2002;184(6):1597-606.
97. Frederix M, Edwards A, McAnulla C, Downie JA. Co-ordination of quorum-sensing regulation in *Rhizobium leguminosarum* by induction of an anti-repressor. *Mol Microbiol.* 2011;81(4):994-1007.
98. Edwards A, Frederix M, Wisniewski-Dyé F, Jones J, Zorreguieta A, Downie JA. The cin and rai quorum-sensing regulatory systems in *Rhizobium leguminosarum* are coordinated by ExpR and CinS, a small regulatory protein coexpressed with CinI. *J Bacteriol.* 2009;191(9):3059-67.
99. Calatrava-Morales N, McIntosh M, Soto MJ. Regulation Mediated by N-Acyl Homoserine Lactone Quorum Sensing Signals in the *Rhizobium*-Legume Symbiosis. *Genes (Basel).* 2018;9(5):263.
100. Hanzelka BL, Greenberg EP. Evidence that the N-terminal region of the *Vibrio fischeri* LuxR protein constitutes an autoinducer-binding domain. *J Bacteriol.* 1995;177(3):815-7.
101. Zhang RG, Pappas KM, Brace JL, Miller PC, Oulmassov T, Molyneaux JM, et al. Structure of a bacterial quorum-sensing transcription factor complexed with pheromone and DNA. *Nature.* 2002;417(6892):971-4.
102. Vannini A, Volpari C, Gargioli C, Muraglia E, Cortese R, De Francesco R, et al. The crystal structure of the quorum sensing protein TraR bound to its autoinducer and target DNA. *The EMBO journal.* 2002;21(17):4393-401.
103. White CE, Winans SC. The quorum-sensing transcription factor TraR decodes its DNA binding site by direct contacts with DNA bases and by detection of DNA flexibility. *Molecular Microbiology.* 2007;64(1):245-56.
104. Zou Y, Nair SK. Molecular basis for the recognition of structurally distinct autoinducer mimics by the *Pseudomonas aeruginosa* LasR quorum-sensing signaling receptor. *Chem Biol.* 2009;16(9):961-70.

105. Minogue TD, Wehland-von Trebra M, Bernhard F, von Bodman SB. The autoregulatory role of EsaR, a quorum-sensing regulator in *Pantoea stewartii* ssp. *stewartii*: evidence for a repressor function. *Mol Microbiol.* 2002;44(6):1625-35.
106. McCready AR, Paczkowski JE, Henke BR, Bassler BL. Structural determinants driving homoserine lactone ligand selection in the *Pseudomonas aeruginosa* LasR quorum-sensing receptor. *Proceedings of the National Academy of Sciences.* 2019;116(1):245.
107. Zou Y, Nair, S.K. LasR-OC12 HSL complex. 2009.
108. Kalia VC. Quorum sensing inhibitors: An overview. *Biotechnology Advances.* 2013;31(2):224-45.
109. Sio CF, Otten LG, Cool RH, Diggle SP, Braun PG, Bos R, et al. Quorum quenching by an N-acyl-homoserine lactone acylase from *Pseudomonas aeruginosa* PAO1. *Infection and immunity.* 2006;74(3):1673-82.
110. Zhang H-B, Wang L-H, Zhang L-H. Genetic control of quorum-sensing signal turnover in *Agrobacterium tumefaciens*. *Proceedings of the National Academy of Sciences.* 2002;99(7):4638.
111. Haudecoeur E, Faure D. A fine control of quorum-sensing communication in *Agrobacterium tumefaciens*. *Communicative & integrative biology.* 2010;3(2):84-8.
112. Jiang Q, Chen J, Yang C, Yin Y, Yao K. Quorum Sensing: A Prospective Therapeutic Target for Bacterial Diseases. *Biomed Res Int.* 2019;2019:2015978-.
113. Hwang I, Smyth AJ, Luo Z-Q, Farrand SK. Modulating quorum sensing by antiactivation: TraM interacts with TraR to inhibit activation of Ti plasmid conjugal transfer genes. *Molecular Microbiology.* 1999;34(2):282-94.
114. Chen G, Jeffrey PD, Fuqua C, Shi Y, Chen L. Structural basis for antiactivation in bacterial quorum sensing. *Proceedings of the National Academy of Sciences of the United States of America.* 2007;104(42):16474-9.
115. Luo ZQ, Qin Y, Farrand SK. The antiactivator TraM interferes with the autoinducer-dependent binding of TraR to DNA by interacting with the C-terminal region of the quorum-sensing activator. *The Journal of biological chemistry.* 2000;275(11):7713-22.
116. Fuqua C, Burbea M, Winans SC. Activity of the *Agrobacterium* Ti plasmid conjugal transfer regulator TraR is inhibited by the product of the traM gene. *J Bacteriol.* 1995;177(5):1367-73.
117. Piper KR, Farrand SK. Quorum sensing but not autoinduction of Ti plasmid conjugal transfer requires control by the opine regulon and the antiactivator TraM. *J Bacteriol.* 2000;182(4):1080-8.
118. Turgay K, Hahn J, Burghoorn J, Dubnau D. Competence in *Bacillus subtilis* is controlled by regulated proteolysis of a transcription factor. *The EMBO journal.* 1998;17(22):6730-8.
119. Costa ED, Chai Y, Winans SC. The quorum-sensing protein TraR of *Agrobacterium tumefaciens* is susceptible to intrinsic and TraM-mediated proteolytic instability. *Mol Microbiol.* 2012;84(5):807-15.
120. Swiderska A, Berndtson AK, Cha MR, Li L, Beaudoin GM, 3rd, Zhu J, et al. Inhibition of the *Agrobacterium tumefaciens* TraR quorum-sensing regulator. Interactions with the TraM anti-activator. *The Journal of biological chemistry.* 2001;276(52):49449-58.
121. Sidoti C, Harwood G, Ackerman R, Coppard J, Merrick M. Characterisation of mutations in the *Klebsiella pneumoniae* nitrogen fixation regulatory gene *nifL* which impair oxygen regulation. *Archives of microbiology.* 1993;159(3):276-81.
122. Asfahl KL, Schuster M. Additive Effects of Quorum Sensing Anti-Activators on *Pseudomonas aeruginosa* Virulence Traits and Transcriptome. *Front Microbiol.* 2017;8:2654.
123. Seet Q, Zhang L-H. Anti-activator QslA defines the quorum sensing threshold and response in *Pseudomonas aeruginosa*. *Molecular Microbiology.* 2011;80(4):951-65.
124. Sullivan JT, Ronson CW. Evolution of rhizobia by acquisition of a 500-kb symbiosis island that integrates into a phe-tRNA gene. *Proceedings of the National Academy of Sciences of the United States of America.* 1998;95(9):5145-9.

125. Ramsay JP, Sullivan JT, Stuart GS, Lamont IL, Ronson CW. Excision and transfer of the *Mesorhizobium loti* R7A symbiosis island requires an integrase IntS, a novel recombination directionality factor RdfS, and a putative relaxase RlxS. *Mol Microbiol.* 2006;62(3):723-34.
126. Burrus V, Pavlovic G, Decaris B, Guédon G. Conjugative transposons: the tip of the iceberg. *Mol Microbiol.* 2002;46(3):601-10.
127. Sullivan JT, Trzebiatowski JR, Cruickshank RW, Gouzy J, Brown SD, Elliot RM, et al. Comparative sequence analysis of the symbiosis island of *Mesorhizobium loti* strain R7A. *J Bacteriol.* 2002;184(11):3086-95.
128. Wozniak RA, Waldor MK. Integrative and conjugative elements: mosaic mobile genetic elements enabling dynamic lateral gene flow. *Nat Rev Microbiol.* 2010;8(8):552-63.
129. Verdonk CJ, Sullivan JT, Williman KM, Nicholson L, Bastholm TR, Hynes MF, et al. Delineation of the integrase-attachment and origin-of-transfer regions of the symbiosis island ICEMISymR7A. *Plasmid.* 2019;104:102416.
130. Ramsay JP, Major AS, Komarovskiy VM, Sullivan JT, Dy RL, Hynes MF, et al. A widely conserved molecular switch controls quorum sensing and symbiosis island transfer in *Mesorhizobium loti* through expression of a novel antiactivator. *Mol Microbiol.* 2013;87(1):1-13.
131. Gesteland RF, Weiss RB, Atkins JF. Recoding: reprogrammed genetic decoding. *Science.* 1992;257(5077):1640-1.
132. Poole ES, Major LL, Cridge AG, Tate WP. The Mechanism of Recoding in Pro- and Eukaryotes. *Protein Synthesis and Ribosome Structure.* 2004:397-428.
133. Ramsay JP, Ronson CW. Silencing quorum sensing and ICE mobility through antiactivation and ribosomal frameshifting. *Mob Genet Elements.* 2015;5(6):103-8.
134. Ramsay JP, Tester LGL, Major AS, Sullivan JT, Edgar CD, Kleffmann T, et al. Ribosomal frameshifting and dual-target antiactivation restrict quorum-sensing-activated transfer of a mobile genetic element. *Proceedings of the National Academy of Sciences.* 2015;112(13):4104.
135. Ramsay JP, Major As Fau - Komarovskiy VM, Komarovskiy Vm Fau - Sullivan JT, Sullivan Jt Fau - Dy RL, Dy Rl Fau - Hynes MF, Hynes Mf Fau - Salmond GPC, et al. A widely conserved molecular switch controls quorum sensing and symbiosis island transfer in *Mesorhizobium loti* through expression of a novel antiactivator. *Molecular Microbiology.* 2013(1365-2958 (Electronic)).
136. Greenfield NJ. Using circular dichroism spectra to estimate protein secondary structure. *Nat Protoc.* 2006;1(6):2876-90.
137. Bateman A, Coggill P, Finn RD. DUFs: families in search of function. *Acta Crystallogr Sect F Struct Biol Cryst Commun.* 2010;66(Pt 10):1148-52.
138. Goodacre NF, Gerloff DL, Uetz P. Protein domains of unknown function are essential in bacteria. *MBio.* 2013;5(1):e00744.
139. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, et al. Pfam: the protein families database. *Nucleic Acids Res.* 2014;42(Database issue):D222-D30.
140. Studier FW, Moffatt BA. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *Journal of Molecular Biology.* 1986;189(1):113-30.
141. Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, Appel RD, et al. Protein identification and analysis tools in the Expasy server. *Methods in molecular biology (Clifton, NJ).* 1999;112:531-52.
142. Rosano GL, Ceccarelli EA. Recombinant protein expression in *Escherichia coli*: advances and challenges. *Frontiers in Microbiology.* 2014;5:172.
143. Walden H. Selenium incorporation using recombinant techniques. *Acta Crystallographica Section D.* 2010;66(4):352-7.
144. McPhillips TM, McPhillips SE, Chiu HJ, Cohen AE, Deacon AM, Ellis PJ, et al. Blu-Ice and the Distributed Control System: software for data acquisition and instrument control at macromolecular crystallography beamlines. *Journal of synchrotron radiation.* 2002;9(Pt 6):401-6.
145. Evans P. Scaling and assessment of data quality. *Acta crystallographica Section D, Biological crystallography.* 2006;62(Pt 1):72-82.

146. Evans PR, Murshudov GN. How good are my data and what is the resolution? *Acta crystallographica Section D, Biological crystallography*. 2013;69(Pt 7):1204-14.
147. Battye TG, Kontogiannis L, Johnson O, Powell HR, Leslie AG. iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta crystallographica Section D, Biological crystallography*. 2011;67(Pt 4):271-81.
148. Kabsch W. XDS. *Acta crystallographica Section D, Biological crystallography*. 2010;66(Pt 2):125-32.
149. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, et al. Overview of the CCP4 suite and current developments. *Acta crystallographica Section D, Biological crystallography*. 2011;67(Pt 4):235-42.
150. Cromer DT, Liberman DA. Anomalous dispersion calculations near to and on the long-wavelength side of an absorption edge. *Acta Crystallographica Section A*. 1981;37(2):267-8.
151. Sheldrick GM. A short history of SHELX. *Acta crystallographica Section A, Foundations of crystallography*. 2008;64(Pt 1):112-22.
152. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. *Journal of applied crystallography*. 2007;40(Pt 4):658-74.
153. Bunkóczy G, McCoy AJ, Echols N, Grosse-Kunstleve RW, Adams PD, Holton JM, et al. Macromolecular X-ray structure determination using weak, single-wavelength anomalous data. *Nature methods*. 2015;12(2):127-30.
154. Raman S, Vernon R, Thompson J, Tyka M, Sadreyev R, Pei J, et al. Structure prediction for CASP8 with all-atom refinement using Rosetta. *Proteins*. 2009;77 Suppl 9(0 9):89-99.
155. Song Y, DiMaio F, Wang RY, Kim D, Miles C, Brunette T, et al. High-resolution comparative modeling with RosettaCM. *Structure*. 2013;21(10):1735-42.
156. Shen Y, Lange O, Delaglio F, Rossi P, Aramini JM, Liu G, et al. Consistent blind protein structure generation from NMR chemical shift data. *Proceedings of the National Academy of Sciences of the United States of America*. 2008;105(12):4685-90.
157. Shen Y, Vernon R, Baker D, Bax A. De novo protein structure generation from incomplete chemical shift assignments. *Journal of biomolecular NMR*. 2009;43(2):63-78.
158. Shen Y, Bryan PN, He Y, Orban J, Baker D, Bax A. De novo structure generation using chemical shifts for proteins with high-sequence identity but different folds. *Protein Sci*. 2010;19(2):349-56.
159. Lange OF, Rossi P, Sgourakis NG, Song Y, Lee HW, Aramini JM, et al. Determination of solution structures of proteins up to 40 kDa using CS-Rosetta with sparse NMR data from deuterated samples. *Proceedings of the National Academy of Sciences of the United States of America*. 2012;109(27):10873-8.
160. McCoy AJ. Solving structures of protein complexes by molecular replacement with Phaser. *Acta crystallographica Section D, Biological crystallography*. 2007;63(Pt 1):32-41.
161. Mielke SP, Krishnan VV. Characterization of protein secondary structure from NMR chemical shifts. *Prog Nucl Magn Reson Spectrosc*. 2009;54(3-4):141-65.
162. Bodenhausen G, Ruben DJ. Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chemical Physics Letters*. 1980;69(1):185-9.
163. Palmer AG, Cavanagh J, Wright PE, Rance M. Sensitivity improvement in proton-detected two-dimensional heteronuclear correlation NMR spectroscopy. *Journal of Magnetic Resonance (1969)*. 1991;93(1):151-70.
164. Kay L, Keifer P, Saarinen T. Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *J Am Chem Soc*. 1992;114(26):10663-5.
165. Tettamanzi MC, Keeler C, Meshack S, Hodsdon ME. Analysis of site-specific histidine protonation in human prolactin. *Biochemistry*. 2008;47(33):8638-47.
166. Rinaldi PL, Monwar M. Two-Dimensional NMR. In: Lindon JC, Tranter GE, Koppenaal DW, editors. *Encyclopedia of Spectroscopy and Spectrometry (Third Edition)*. Oxford: Academic Press; 2017. p. 475-87.

167. Yamazaki T, Forman-Kay JD, Kay LE. Two-dimensional NMR experiments for correlating carbon-13. beta. and proton. delta. / . epsilon. chemical shifts of aromatic residues in 13C-labeled proteins via scalar couplings. *J Am Chem Soc.* 1993;115(23):11054-5.
168. Grzesiek S, Bax A. Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR. *J Am Chem Soc.* 1992;114(16):6291-3.
169. Kay LE, Ikura M, Tschudin R, Bax A. Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. *Journal of Magnetic Resonance (1969).* 1990;89(3):496-514.
170. Grzesiek S, Bax A. Amino acid type determination in the sequential assignment procedure of uniformly 13C/15N-enriched proteins. *Journal of biomolecular NMR.* 1993;3(2):185-204.
171. Bax A, Clore GM, Gronenborn AM. 1H · 1H correlation via isotropic mixing of 13C magnetization, a new three-dimensional approach for assigning 1H and 13C spectra of 13C-enriched proteins. *Journal of Magnetic Resonance (1969).* 1990;88(2):425-31.
172. Muhandiram DR, Farrow NA, Xu GY, Smallcombe SH, Kay LE. A Gradient 13C NOESY-HSQC Experiment for Recording NOESY Spectra of 13C-Labeled Proteins Dissolved in H2O. *Journal of Magnetic Resonance, Series B.* 1993;102(3):317-21.
173. Marion D, Driscoll PC, Kay LE, Wingfield PT, Bax A, Gronenborn AM, et al. Overcoming the overlap problem in the assignment of proton NMR spectra of larger proteins by use of three-dimensional heteronuclear proton-nitrogen-15 Hartmann-Hahn-multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy: application to interleukin 1. beta. *Biochemistry.* 1989;28(15):6150-6.
174. Marion D, Kay LE, Sparks SW, Torchia DA, Bax A. Three-dimensional heteronuclear NMR of nitrogen-15 labeled proteins. *J Am Chem Soc.* 1989;111(4):1515-7.
175. Cheung MS, Maguire ML, Stevens TJ, Broadhurst RW. DANGLE: A Bayesian inferential method for predicting protein backbone dihedral angles and secondary structure. *Journal of magnetic resonance (San Diego, Calif : 1997).* 2010;202(2):223-33.
176. Shen Y, Delaglio F, Cornilescu G, Bax A. TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *Journal of biomolecular NMR.* 2009;44(4):213-23.
177. Buchan DWA, Jones DT. The PSIPRED Protein Analysis Workbench: 20 years on. *Nucleic Acids Res.* 2019;47(W1):W402-w7.
178. Jones DT. Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol.* 1999;292(2):195-202.
179. Guntert P, Mumenthaler C, Wuthrich K. Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J Mol Biol.* 1997;273(1):283-98.
180. Guntert P. Automated NMR structure calculation with CYANA. *Methods in molecular biology (Clifton, NJ).* 2004;278:353-78.
181. Skou S, Gillilan RE, Ando N. Synchrotron-based small-angle X-ray scattering of proteins in solution. *Nat Protoc.* 2014;9(7):1727-39.
182. Kirby N, Cowieson N, Hawley AM, Mudie ST, McGillivray DJ, Kusel M, et al. Improved radiation dose efficiency in solution SAXS using a sheath flow sample environment. *Acta crystallographica Section D, Structural biology.* 2016;72(Pt 12):1254-66.
183. Ryan T, Trewella J, Murphy J, Keown J, Casey L, Pearce FG, et al. An optimized SEC-SAXS system enabling high X-ray dose for rapid SAXS assessment with correlated UV measurements for biomolecular structure analysis. *Journal of applied crystallography.* 2018;51:97-111.
184. Konarev PV, Volkov VV, Sokolova AV, Koch MHJ, Svergun DI. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *Journal of applied crystallography.* 2003;36(5):1277-82.
185. Panjkovich A, Svergun DI. CHROMIXS: automatic and interactive analysis of chromatography-coupled small-angle X-ray scattering data. *Bioinformatics.* 2018;34(11):1944-6.

186. Franke D, Petoukhov MV, Konarev PV, Panjkovich A, Tuukkanen A, Mertens HDT, et al. ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *Journal of applied crystallography*. 2017;50(Pt 4):1212-25.
187. Svergun DI. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys J*. 1999;76(6):2879-86.
188. Svergun D, Barberato C, Koch MHJ. CRY SOL– a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *Journal of applied crystallography*. 1995;28(6):768-73.
189. Holm L. Benchmarking fold detection by DaliLite v.5. *Bioinformatics*. 2019;35(24):5326-7.
190. Roy A, Yang J, Zhang Y. COFACTOR: an accurate comparative algorithm for structure-based protein function annotation. *Nucleic Acids Res*. 2012;40(Web Server issue):W471-W7.
191. Zhang C, Freddolino PL, Zhang Y. COFACTOR: improved protein function prediction by combining structure, sequence and protein-protein interaction information. *Nucleic Acids Res*. 2017;45(W1):W291-W9.
192. Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc*. 2010;5(4):725-38.
193. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. The I-TASSER Suite: protein structure and function prediction. *Nature methods*. 2015;12(1):7-8.
194. Ovchinnikov S, Kamisetty H, Baker D. Robust and accurate prediction of residue–residue interactions across protein interfaces using evolutionary information. *eLife*. 2014;3:e02030.
195. Kamisetty H, Ovchinnikov S, Baker D. Assessing the utility of coevolution-based residue–residue contact predictions in a sequence- and structure-rich era. *Proceedings of the National Academy of Sciences*. 2013;110(39):15674.
196. Balakrishnan S, Kamisetty H, Carbonell JG, Lee S-I, Langmead CJ. Learning generative models for protein fold families. *Proteins: Structure, Function, and Bioinformatics*. 2011;79(4):1061-78.
197. Studier FW. Protein production by auto-induction in high density shaking cultures. *Protein expression and purification*. 2005;41(1):207-34.
198. Cai M, Huang Y, Craigie R, Clore GM. A simple protocol for expression of isotope-labeled proteins in *Escherichia coli* grown in shaker flasks at high cell density. *Journal of biomolecular NMR*. 2019;73(12):743-8.
199. Waugh DS. Crystal structures of MBP fusion proteins. *Protein Sci*. 2016;25(3):559-71.
200. Kapust RB, Waugh DS. *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Sci*. 1999;8(8):1668-74.
201. Fox JD, Routzahn KM, Bucher MH, Waugh DS. Maltodextrin-binding proteins from diverse bacteria and archaea are potent solubility enhancers. *FEBS letters*. 2003;537(1-3):53-7.
202. Waugh DS. An overview of enzymatic reagents for the removal of affinity tags. *Protein expression and purification*. 2011;80(2):283-93.
203. Ramsay JP. Regulation of excision and transfer of the *Mesorhizobium loti* R7A symbiosis island (Thesis, Doctor of Philosophy). 2008.
204. Perkins DN, Pappin DJC, Creasy DM, Cottrell JS. Probability-based protein identification by searching sequence databases using mass spectrometry data. *ELECTROPHORESIS*. 1999;20(18):3551-67.
205. Koenig T, Menze BH, Kirchner M, Monigatti F, Parker KC, Patterson T, et al. Robust Prediction of the MASCOT Score for an Improved Quality Assessment in Mass Spectrometric Proteomics. *Journal of Proteome Research*. 2008;7(9):3708-17.
206. Pappin DJ, Hojrup P, Bleasby AJ. Rapid identification of proteins by peptide-mass fingerprinting. *Current biology : CB*. 1993;3(6):327-32.
207. Zhang W, Xu S, Li J, Shen X, Wang Y, Yuan Z. Modulation of a thermoregulated type VI secretion system by AHL-dependent quorum sensing in *Yersinia pseudotuberculosis*. *Archives of microbiology*. 2011;193(5):351-63.

208. Knittel V, Sadana P, Seekircher S, Stolle AS, Körner B, Volk M, et al. RovC - a novel type of hexameric transcriptional activator promoting type VI secretion gene expression. *PLoS pathogens*. 2020;16(9):e1008552.
209. Kendrew JC, Bodo G, Dintzis HM, Parrish RG, Wyckoff H, Phillips DC. A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature*. 1958;181(4610):662-6.
210. Kendrew JC, Dickerson RE, Strandberg BE, Hart RG, Davies DR, Phillips DC, et al. Structure of myoglobin: A three-dimensional Fourier synthesis at 2 Å resolution. *Nature*. 1960;185(4711):422-7.
211. Smyth MS, Martin JH. x ray crystallography. *Mol Pathol*. 2000;53(1):8-14.
212. 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):5705-36.
213. Lieberman RL, Peek ME, Watkins JD. Determination of soluble and membrane protein structures by X-ray crystallography. *Methods in molecular biology (Clifton, NJ)*. 2013;955:475-93.
214. Giegé R. A historical perspective on protein crystallization from 1840 to the present day. *FEBS J*. 2013;280(24):6456-97.
215. Hofmeister T. Über die Darstellung von krystallisiertem Eiralbumin und die Krystallisirbarkeit colloider Stoffe. *Z Physiol Chem*. 1890;14:165.
216. Stroud RM. Present at the flood: How structural biology came about, by Richard E. Dickerson. *Protein Sci*. 2007;16(1):135-6.
217. Bragg WL, Bragg WH. The structure of some crystals as indicated by their diffraction of X-rays. *Proceedings of the Royal Society of London Series A, Containing Papers of a Mathematical and Physical Character*. 1913;89(610):248-77.
218. Bragg WH, Bragg WL. The reflection of X-rays by crystals. *Proceedings of the Royal Society of London Series A, Containing Papers of a Mathematical and Physical Character*. 1913;88(605):428-38.
219. McPherson A, Gavira JA. Introduction to protein crystallization. *Acta Crystallogr F Struct Biol Commun*. 2014;70(Pt 1):2-20.
220. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. *Nucleic Acids Res*. 2000;28(1):235-42.
221. Pike ACW, Garman EF, Krojer T, von Delft F, Carpenter EP. An overview of heavy-atom derivatization of protein crystals. *Acta crystallographica Section D, Structural biology*. 2016;72(Pt 3):303-18.
222. Hendrickson WA, Horton JR, LeMaster DM. Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *The EMBO Journal*. 1990;9(5):1665-72.
223. Walden H. Selenium incorporation using recombinant techniques. *Acta crystallographica Section D, Biological crystallography*. 2010;66(Pt 4):352-7.
224. Chen G, Malenkos JW, Cha M-R, Fuqua C, Chen L. Quorum-sensing antiactivator TraM forms a dimer that dissociates to inhibit TraR. *Molecular Microbiology*. 2004;52(6):1641-51.
225. Smyth DR, Mrozkiewicz MK, McGrath WJ, Listwan P, Kobe B. Crystal structures of fusion proteins with large-affinity tags. *Protein Sci*. 2003;12(7):1313-22.
226. Padilla JE, Yeates TO. A statistic for local intensity differences: robustness to anisotropy and pseudo-centering and utility for detecting twinning. *Acta crystallographica Section D, Biological crystallography*. 2003;59(Pt 7):1124-30.
227. Cowtan K. Recent developments in classical density modification. *Acta crystallographica Section D, Biological crystallography*. 2010;66(Pt 4):470-8.
228. Cowtan K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta crystallographica Section D, Biological crystallography*. 2006;62(Pt 9):1002-11.
229. Oeffner RD, Bunkóczi G, McCoy AJ, Read RJ. Improved estimates of coordinate error for molecular replacement. *Acta crystallographica Section D, Biological crystallography*. 2013;69(Pt 11):2209-15.

230. Kovalevskiy O, Nicholls RA, Long F, Carlon A, Murshudov GN. Overview of refinement procedures within REFMAC5: utilizing data from different sources. *Acta crystallographica Section D, Structural biology*. 2018;74(Pt 3):215-27.
231. Morris AL, MacArthur MW, Hutchinson EG, Thornton JM. Stereochemical quality of protein structure coordinates. *Proteins*. 1992;12(4):345-64.
232. Brünger AT. Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*. 1992;355(6359):472-5.
233. Qian B, Raman S, Das R, Bradley P, McCoy AJ, Read RJ, et al. High-resolution structure prediction and the crystallographic phase problem. *Nature*. 2007;450(7167):259-64.
234. Smyth DR, Mrozkiewicz MK, McGrath WJ, Listwan P, Kobe B. Crystal structures of fusion proteins with large-affinity tags. *Protein Sci*. 2003;12(7):1313-22.
235. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. *Nucleic Acids Res*. 2000;28(1):235-42.
236. Gardner KH, Kay LE. The use of ²H, ¹³C, ¹⁵N multidimensional NMR to study the structure and dynamics of proteins. *Annual review of biophysics and biomolecular structure*. 1998;27:357-406.
237. Kanelis V, Forman-Kay JD, Kay LE. Multidimensional NMR methods for protein structure determination. *IUBMB life*. 2001;52(6):291-302.
238. Liu HL, Hsu JP. Recent developments in structural proteomics for protein structure determination. *Proteomics*. 2005;5(8):2056-68.
239. Sugiki T, Kobayashi N, Fujiwara T. Modern Technologies of Solution Nuclear Magnetic Resonance Spectroscopy for Three-dimensional Structure Determination of Proteins Open Avenues for Life Scientists. *Comput Struct Biotechnol J*. 2017;15:328-39.
240. Hahn EL, Maxwell DE. Spin Echo Measurements of Nuclear Spin Coupling in Molecules. *Physical Review*. 1952;88(5):1070-84.
241. Overhauser AW. Polarization of Nuclei in Metals. *Physical Review*. 1953;92(2):411-5.
242. Marion D. An Introduction to Biological NMR Spectroscopy. *Molecular & Cellular Proteomics*. 2013;12(11):3006-25.
243. Frueh DP. Practical aspects of NMR signal assignment in larger and challenging proteins. *Prog Nucl Magn Reson Spectrosc*. 2014;78:47-75.
244. Clore GM. Practical Aspects of Paramagnetic Relaxation Enhancement in Biological Macromolecules. *Methods Enzymol*. 2015;564:485-97.
245. Chen K, Tjandra N. The use of residual dipolar coupling in studying proteins by NMR. *Top Curr Chem*. 2012;326:47-67.
246. Rosato A, Tejero R, Montelione GT. Quality assessment of protein NMR structures. *Current Opinion in Structural Biology*. 2013;23(5):715-24.
247. Mao B, Guan R, Montelione GT. Improved technologies now routinely provide protein NMR structures useful for molecular replacement. *Structure*. 2011;19(6):757-66.
248. Ramelot TA, Raman S, Kuzin AP, Xiao R, Ma LC, Acton TB, et al. Improving NMR protein structure quality by Rosetta refinement: a molecular replacement study. *Proteins*. 2009;75(1):147-67.
249. Wüthrich K. NMR Studies of Structure and Function of Biological Macromolecules (Nobel Lecture). *Angewandte Chemie International Edition*. 2003;42(29):3340-63.
250. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, et al. Circos: an information aesthetic for comparative genomics. *Genome Res*. 2009;19(9):1639-45.
251. Thompson JM, Sgourakis NG, Liu G, Rossi P, Tang Y, Mills JL, et al. Accurate protein structure modeling using sparse NMR data and homologous structure information. *Proceedings of the National Academy of Sciences of the United States of America*. 2012;109(25):9875-80.
252. Mao B, Tejero R, Baker D, Montelione GT. Protein NMR structures refined with Rosetta have higher accuracy relative to corresponding X-ray crystal structures. *J Am Chem Soc*. 2014;136(5):1893-906.

253. Yin Z, Kaelber JT, Ebright RH. Structural basis of Q-dependent antitermination. *Proceedings of the National Academy of Sciences*. 2019;116(37):18384.
254. Kelley LA, Sternberg MJE. Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc*. 2009;4(3):363-71.
255. Baek M, DiMaio F, Anishchenko I, Dauparas J, Ovchinnikov S, Lee GR, et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science*. 2021;373(6557):871.
256. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021.

Every reasonable effort has been made to acknowledge the owners of copyright material. I would be pleased to hear from any copyright owner who has been omitted or incorrectly acknowledged.