



Structure-function relationships in stressosome
complexes from *Listeria monocytogenes* and
Bacillus subtilis.

Sema Ejder

Doctor of Philosophy

Newcastle University Bioscience Institute

School of Natural and Environmental Sciences

February 2021

Abstract

Listeria monocytogenes is a foodborne bacterial pathogen that can resist and overcome extreme environmental conditions, such as the extremes of temperature, salinity and pH that are encountered during food processing. Stress resistance is regulated by a supramolecular protein complex called the stressosome, which detects and integrates environmental stress signals that induce a partner-switching and phosphorylation cascade leading to the activation of an alternative RNA polymerase sigma factor, σ^B , which controls a regulon of ~200 genes involved in the general stress response. The stressosome comprises three main proteins, RsbR (which has four paralogues), RsbS and RsbT. The N-terminal domains of RsbR proteins have been proposed to act as stress sensors and they project from the core of the stressosome as ‘turrets’. However, the mechanism by which signals are perceived and transmitted is still unknown.

Structural studies of the stressosome’s sensory domains resulted in the successful determination of the crystal structures of N-RsbR1, N-RsbR2, and N-RsbR3. Ligand binding pockets were identified in N-RsbR3 that yield insight into signal perception and transduction mechanisms. The interaction of the Prli42 miniprotein with N-RsbR proteins was also assessed and shown not to occur at biologically-relevant concentrations. Common ligands and drug-like fragments were screened against binding to the putative ligand binding pocket and candidate interacting molecules were identified.

Native stressosomes pulled-down from *B. subtilis* cell lysates by affinity purification contained all four RsbR paralogues, along with RsbS and RsbT. Initial electron microscopy of the purified native stressosomes were consistent with the formation of a highly symmetric structure. Purification and EM studies of stressosome variants revealed that stressosomes can be formed by any of the RsbR paralogues. Stressosome variants were analysed by cryo-EM single particle analysis, which showed that the RsbR-RsbS complex displays similar features to the known stressosome complex of *B. subtilis*, albeit with markedly different stoichiometries.

Funding Acknowledgement



“This project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 721456”.

Acknowledgements

First and foremost, I would like to express my deep and sincere gratitude to Prof. Rick Lewis for considering me for this project. Thanks to him, I joined Newcastle University and the PATHSENSE Network funded by the MCSA. I'd like to also thank Dr. Jon Marles-Wright to co-supervise my project. As a pioneer in this project, I've learned a lot from him especially in Cryo-EM. And I'd like to thank Prof. Dr. Tracy Palmer for agreeing to be my third co-supervisor.

I'd like to thank the MSCA for funding my studies and the whole PATHSENSE network, Dr. Conor O'Byrne and Dr. Nina Tuite especially, for everything that happened to me during this project.

I would like to thank my secondments labs in Griesfald, Germany with Dr. Jan Pané-Farré and Maria Conway and in Madrid, Spain with Prof. Francisco Garcia-del Portillo and Charlotte Dessaux. And I'd like to thank Prof. Nicola Stanley-Wall, Manisha Pandey and Diana Gudynaite for the help and close collaborations on this project.

To the people of the Lewis lab, I will be forever thankful to Dr. Arnaud Basle, my French pair, who teach me a lot of the crystallography, but also a lot of computer knowledge. To Simon Booth who brought a lot of joy in the lab but also helped me a lot with data processing and using several machines around the institute. To Zoe Rutter, for our great 20th Gram positive conference trip to the U.S and Dr. Vince Rao, for his insightful advice for my project.

I'd like to thank the Astbury Biostructure Laboratory for their help in the screening and cryo-EM data collection.

I am forever grateful for Paola and Joao Victor, with whom I enjoyed the stay in Newcastle and for all the support during throughout my PhD.

Dedication

I am dedicating my PhD. to my family.

To my beloved father Mehmet Ejder (12/12/2011),
who was always proud and supportive of me throughout my studies. He always gave me insightful advice during my studies to keep on going. I wish he was there to see my achievements. Rest in peace, I am sure you are proud of me!

To my mom Nurten Taşçı Ejder,
who was always there for me and supported me throughout my studies. She allowed me to go through my studies without having to worry about anything else. Studying and getting educated was more important than anything else for me and for her. And she is proud.

To my sisters, Melek, Saibe, Fatma and Seval,
without whom I won't be myself. We are like the 5 fingers of a hand, inseparable. And to my brothers-in-law, Özkan and Bahadır, to be there whenever I needed them.

To the Umaç family,
who were truly supportive of me during my studies. Even though, we are not connected by blood, we are more than family friends, we are a family.

To my close friends, Elif, Hatice, Neslihan, Sarah,
who supported me a lot during the tough times but also cheered me with throughout my life.

And finally, to my nephews and niece, Ata, Kadir, Mahir, Arya, Mete and to the future ones. They are the best things that can happen to me and another reason to keep on going.

Table of Contents

ABSTRACT	I
FUNDING ACKNOWLEDGEMENT	II
ACKNOWLEDGEMENTS.....	III
DEDICATION	IV
TABLE OF CONTENTS	V
LIST OF FIGURES	IX
LIST OF TABLES	XII
ABBREVIATIONS LIST	XIV
CHAPTER I: INTRODUCTION	1
1.1 STUDIED SPECIES	2
1.1.1 <i>Listeria monocytogenes</i>	2
1.1.2 <i>Bacillus subtilis</i>	4
1.2. THE ALTERNATIVE SIGMA FACTOR SIGB.....	5
1.3. THE Σ^B SIGNALLING HUB	7
1.3.1. <i>The stressosome-dependent pathway</i>	9
1.3.2. <i>The stressosome-independent pathway</i>	12
1.3.3. <i>The common pathway</i>	12
1.3.4. <i>Resetting the sigB pathway</i>	14
1.4. BIOINFORMATICS AND STRUCTURAL ANALYSIS OF THE STRESSOSOME MODULE	15
1.4.1. <i>The stressosome scaffolding protein RsbS</i>	15
1.4.2. <i>RsbR, the presumed stress sensor</i>	17
1.5. THE STRESSOSOME COMPLEX	21
1.5.1. <i>The stressosome complex in B. subtilis</i>	21
1.5.2. <i>The stressosome complex in L. monocytogenes</i>	23
1.6. POTENTIAL STRESSOSOME INTERACTION PARTNER AND MEMBRANE ANCHOR: THE PrLI42 MINIPROTEIN	25
1.7. PROJECT OBJECTIVES	27
CHAPTER II: MATERIALS AND METHODS.....	29
2.1. MOLECULAR BIOLOGY.....	30
2.1.1 CIDAR MoClo Assembly.....	32
2.1.2. <i>Production of protein sequence variants</i>	35
2.1.3. <i>Fusion proteins</i>	35
Maltose Binding Protein fusions.....	35
GFP-fusion	36
2.1.4. <i>Co-expression vector</i>	37
2.2. PROTEIN EXPRESSION	38

2.2.1.	<i>Protein Expression tests</i>	39
2.2.2.	<i>Large-scale recombinant protein expression</i>	39
2.2.3.	<i>Selenomethionine labelled protein expression</i>	40
2.2.4.	<i>Large-scale B. subtilis growth for native stressosome purification</i>	41
2.3.	PROTEIN PURIFICATION	42
2.3.1.	<i>Purification of protein by affinity chromatography</i>	42
	Ion-Exchange chromatography	42
	Immobilised metal ion affinity chromatography	43
	Streptactin-desthiobiotin affinity chromatography	43
	Purification of protein by high-resolution ion-exchange chromatography	43
2.3.2.	<i>Purification of proteins by size-exclusion chromatography</i>	44
2.4.	IN VITRO COMPLEX RECONSTITUTION	44
2.5.	CIRCULAR DICHROISM	44
2.6.	SMALL ANGLE X-RAY SCATTERING	46
2.6.1.	<i>Sample preparation</i>	46
2.6.2.	<i>Data processing</i>	46
2.7.	THERMOFLUOR ASSAY	47
2.8.	FLUORESCENCE POLARISATION.....	48
2.9.	X-RAY CRYSTALLOGRAPHY.....	50
2.9.1.	<i>Crystallization screening and optimisation</i>	50
2.9.2.	<i>Crystal diffraction and data collection</i>	51
2.9.3.	<i>X-ray data processing and structure determination</i>	53
	Initial data processing.....	53
	Molecular replacement	53
	Experimental phasing	53
	Model building and refinement.....	54
	Model validation.....	54
	Structure comparisons	54
2.10.	ELECTRON CRYO-MICROSCOPY AND SINGLE PARTICLE ANALYSIS OF <i>L. MONOCYTOGENES</i> COMPLEXES	55
2.10.1.	<i>Negative stain control</i>	55
2.10.2.	<i>Negative stain 2D classes</i>	55
2.10.3.	<i>Cryo-EM grid preparation and data collection</i>	55
2.10.4.	<i>Data processing</i>	56
2.10.5.	<i>Model building</i>	58

CHAPTER III: STRUCTURAL STUDIES OF THE STRESS SENSORY DOMAINS OF THE *L. MONOCYTOGENES*

STRESSOSOME.....		59
3.1.	INTRODUCTION	60
3.2.	PURIFICATION OF RECOMBINANT <i>L. MONOCYTOGENES</i> N-RsBR PROTEINS	61
3.2.1.	<i>N-RsbR1 purification</i>	62
3.2.2.	<i>N-RsbR2 purification</i>	63

3.2.3.	<i>N-RsbR3 purification</i>	63
3.2.4.	<i>N-RsbR4 purification</i>	64
3.3.	FOLDING OF THE PURIFIED PROTEINS IN SOLUTION	65
3.4.	CRYSTALLISATION OF N-RsbR PROTEINS	66
3.5.	X-RAY DATA COLLECTION AND PROCESSING	67
3.6.	SELENOMETHIONINE-LABELLED PROTEIN PURIFICATION	69
3.6.1.	<i>Se-Met N-RsbR1 purification</i>	70
3.6.2.	<i>Se-Met N-RsbR2 purification</i>	71
3.6.3.	<i>Se-Met N-RsbR3 purification</i>	71
3.6.4.	<i>Se-Met N-RsbR4 purification</i>	72
3.7.	SE-MET LABELLED N-RsbR CRYSTALLISATION SCREENING AND DATA COLLECTION	74
3.8.	SULPHUR-SAD EXPERIMENTAL PHASING FOR N-RsbR1	76
3.9.	N-RsbR1 STRUCTURE	79
3.10.	N-RsbR2 STRUCTURE	81
3.11.	N-RsbR3 STRUCTURE	85
3.12.	STRUCTURE COMPARISON	94
3.13.	CONCLUSION	96
CHAPTER IV: INTERACTION STUDY OF RSBR WITH POTENTIAL BINDING PARTNERS		97
4.1.	INTRODUCTION	98
4.2.	INTERACTION STUDY OF Prli42 BY CO-ELUTION	99
4.3.	INTERACTION STUDY OF Prli42 AND N-RsbRs BY FLUORESCENCE POLARIZATION	100
4.3.1.	<i>Bsu RsbR NTDs purification</i>	100
	N-RsbRA purification	100
	N-RsbRB purification	101
	N-RsbRC purification	102
	N-RsbRD purification	102
4.3.2.	<i>Determination of Prli42 interaction with N-RsbR proteins by fluorescence polarization</i>	103
4.4.	STUDY OF THE LIGAND BINDING POCKET OF N-RsbRs	107
4.4.1.	<i>Study of the thermal stability of the N-RsbR proteins</i>	108
4.4.2.	<i>Effect of the nucleic acid on the N-RsbRs' thermostability</i>	111
4.4.3.	<i>Effect of amino acids on the N-RsbRs' thermostability</i>	112
4.4.4.	<i>Effect of sugars on the N-RsbRs' thermostability</i>	113
4.4.5.	<i>Effect of chemical compounds on N-RsbRs' thermostability</i>	114
4.5.	N-RsbR3 CRYSTAL SOAKING WITH FRAGLITES	116
4.6.	CONCLUSIONS	119
CHAPTER V: NATIVE STRESSOSOME AND VARIANT STRESSOSOME STUDY		121
5.1.	INTRODUCTION	122
PART A: <i>B. SUBTILIS</i> NATIVE STRESSOSOME COMPLEX COMPOSITION		122
5.2.	PURIFICATION AND VISUALISATION OF THE NATIVE STRESSOSOME COMPLEX FROM <i>B. SUBTILIS</i>	123

5.3.	INVESTIGATION OF THE NATIVE STRESSOSOME COMPOSITION FROM <i>B. SUBTILIS</i>	126
5.4.	CONCLUSION	132
PART B: STUDY OF STRESSOSOMES VARIANTS WITH RsbR PARALOGUES		132
5.5.	RsbR PARALOGUES AND RsbS PURIFICATION.....	133
5.5.1.	<i>RsbR1 recombinant protein purification</i>	134
5.5.2.	<i>RsbR2 recombinant protein purification</i>	134
5.5.3.	<i>RsbR3 recombinant protein purification</i>	135
5.5.4.	<i>RsbR4 recombinant protein purification</i>	135
5.5.5.	<i>Purification of recombinant RsbS</i>	136
5.6.	STRESSOSOME COMPLEX FORMATION BY RsbR PARALOGUES.....	137
5.6.1.	<i>RsbR1-RsbS complex</i>	137
5.6.2.	<i>RsbR2-RsbS complex</i>	138
5.6.3.	<i>RsbR3-RsbS complex</i>	140
5.6.4.	<i>RsbR4-RsbS complex</i>	142
5.7.	CONCLUSION	143
PART C: STUDY OF THE TERNARY STRESSOSOME COMPLEX INVOLVING RsbT AND RsbX.....		143
5.8.	PURIFICATION OF THE RsbT KINASE	144
5.9.	THE RsbX PHOSPHATASE PURIFICATION.....	144
5.10.	RECONSTITUTION OF THE RST AND RSX TERNARY STRESSOSOME COMPLEXES.....	145
5.11.	CONCLUSION	149
CHAPTER VI: CRYO-EM SINGLE PARTICLE ANALYSIS OF THE <i>L. MONOCYTOGENES</i> RsbR1-RsbS AND RsbR1-RsbS-RsbX STRESSOSOME COMPLEXES		150
6.1.	INTRODUCTION	151
6.2.	CRYO-EM DATA COLLECTION OF THE <i>LmORS</i> AND <i>LmORSX</i> STRESSOSOME COMPLEX.	151
6.3.	CRYO-EM DATA PROCESSING	152
6.4.	ATOMIC MODEL OF THE <i>LmORS</i> RECONSTRUCTION.....	160
6.4.1.	<i>LmoRS complex model in C1 symmetry</i>	160
6.4.2.	<i>LmoRS complex model in D2 symmetry</i>	166
6.4.3.	<i>LmoRS complex assembly</i>	170
6.4.4.	<i>Phosphorylation site dispositions</i>	173
6.5.	CONCLUSIONS.....	175
CONCLUSION AND FUTURE PERSPECTIVES.....		177
BIBLIOGRAPHY		181
APPENDICES		191

List of Figures

CHAPTER I: INTRODUCTION	1
FIGURE 1.1. 16S PHYLOGENETIC TREE FOR KEY <i>BACILLUS</i> AND <i>LISTERIA</i> SPECIES	3
FIGURE 1.2. <i>E. COLI</i> RNAP HOLOENZYME STRUCTURE	6
FIGURE 1.3. ORGANISATION OF THE <i>RSB</i> OPERON.....	8
FIGURE 1.4. SCHEMATIC REPRESENTATION OF THE SIGNALISATION PATHWAY IN UNSTRESSED AND STRESSED CELLS.....	10
FIGURE 1.5. STRUCTURE OF SPOIIAA:SPOIIB COMPLEX FROM <i>B. SUBTILIS</i>	13
FIGURE 1.6 RSB/RsBS STAS DOMAIN ALIGNMENT	16
FIGURE 1.7. CRYSTAL STRUCTURES OF THE RsBS STAS DOMAIN FROM DIFFERENT SPECIES	17
FIGURE 1.8. N-RsBR PARALOGUES AND HOMOLOGUES SEQUENCE ALIGNMENT.....	18
FIGURE 1.9. CRYSTAL STRUCTURE OF THE NON-HEME GLOBIN-LIKE N-RsBR DOMAIN	19
FIGURE 1.10. NMR STRUCTURE OF THE <i>B. SUBTILIS</i> YtvA PROTEIN.....	21
FIGURE 1.11. MODELS OF THE <i>B. SUBTILIS</i> STRESSOSOME COMPLEX BASED ON CRYO-EM RECONSTRUCTIONS.	22
FIGURE 1.12. CRYO-EM MODEL OF <i>B. SUBTILIS</i> STRESSOSOME COMPLEX IN C1 SYMMETRY	23
FIGURE 1.13. CRYO-EM MODEL OF <i>L. MONOCYTOGENES</i> STRESSOSOME COMPLEX IN C1 SYMMETRY.....	24
FIGURE 1.14. MODELS OF PrLI42 MINIPROTEIN AND RSB RNTD INTERACTIONS	26
CHAPTER II: MATERIALS AND METHODS.....	29
FIGURE 2.1. SCHEMATIC REPRESENTATION OF THE CIDAR MoClo DNA ASSEMBLY METHOD	33
FIGURE 2.2. pMAT10 PLASMID MAP.	36
FIGURE 2.3. SCHEMATIC REPRESENTATION OF THE MULTI ASSEMBLY USING CIDAR MoClo CLONING.....	37
FIGURE 2.4. SCHEMATIC REPRESENTATION OF THE CO-EXPRESSION VECTOR ASSEMBLY USING CIDAR MoClo CLONING.	38
FIGURE 2.5. MODEL OF <i>B. SUBTILIS</i> N-RsBRA STRUCTURE WITH SITES OF HIS-TAGS.	41
FIGURE 2.6. CD SPECTRA CHARACTERISTICS OF DIFFERENT PROTEIN SECONDARY STRUCTURES	45
FIGURE 2.7. TYPICAL TSA RESULTS.....	47
FIGURE 2.8. FLUORESCENCE POLARIZATION METHOD	49
FIGURE 2.9. THE VAPOUR DIFFUSION CRYSTALLISATION METHOD	51
FIGURE 2.10: THEORETICAL SELENIUM AND SULPHUR X-RAY ABSORPTION EDGE.....	52
CHAPTER III: STRUCTURAL STUDIES OF THE STRESS SENSORY DOMAINS OF THE <i>L. MONOCYTOGENES</i>	
STRESSOSOME.....	59
FIGURE 3.1. NON-HEME GLOBIN FOLD OF N-RsBR AND GLOBIN FOLD HEMAT STRUCTURE COMPARISON	61
FIGURE 3.2. N-RsBR1 PROTEIN PURIFICATION	62
FIGURE 3.3. N-RsBR2 PROTEIN PURIFICATION	63
FIGURE 3.4. N-RsBR3 PROTEIN PURIFICATION	64
FIGURE 3.5. N-RsBR4 PROTEIN PURIFICATIONS	65
FIGURE 3.6. CD SPECTRA OF <i>L. MONOCYTOGENES</i> RSB RNTDs	66
FIGURE 3.7. CRYSTALLOGRAPHIC DATA PROCESSING FLOW-CHART.	69
FIGURE 3.8. SELENOMETHIONINE-LABELLED N-RsBR1 PROTEIN PURIFICATION	70

FIGURE 3.9. SELENOMETHIONINE-LABELLED N-RsBR2 PROTEIN PURIFICATION	71
FIGURE 3.10. SELENOMETHIONINE-LABELLED N-RsBR3 PROTEIN PURIFICATION	72
FIGURE 3.11. N-RsBR4 Se-MET LABELLED RECOMBINANT PROTEIN PURIFICATION	72
FIGURE 3.12. His6-TAGGED N-RsBR4 PROTEIN PURIFICATION	73
FIGURE 3.13. X-RAY CRYSTAL STRUCTURE OF <i>L. MONOCYTOGENES</i> N-RsBR1	79
FIGURE 3.14. SEQUENCE ALIGNMENT OF BACTERIAL PROTEINS ADOPTING A GLOBIN-LIKE FOLD	80
FIGURE 3.15. X-RAY CRYSTAL STRUCTURE OF <i>L. MONOCYTOGENES</i> N-RsBR2	82
FIGURE 3.16. SAXS DERIVED MOLECULAR ENVELOPE OF N-RsBR2 IN SOLUTION.....	84
FIGURE 3.17. X-RAY CRYSTAL STRUCTURE OF <i>L. MONOCYTOGENES</i> N-RsBR3	86
FIGURE 3.18. MULTIPLE SEQUENCE ALIGNMENT OF N-RsBRs	88
FIGURE 3.19. N-RsBR3 LIGANDS IN THE PUTATIVE LIGAND BINDING POCKET	89
FIGURE 3.20. N-RsBR3 ELECTRON DENSITY AT THE DIMERIZATION INTERFACE	91
FIGURE 3.21. N-RsBR3 CONFORMATIONAL CHANGES WITH AND WITHOUT LIGAND.....	93
FIGURE 3.22. STRUCTURAL ALIGNMENT OF RsBR N-TERMINAL DOMAINS.....	95
CHAPTER IV: INTERACTION STUDY OF RsBR WITH POTENTIAL BINDING PARTNERS	97
FIGURE 4.1. N-RsBR1 AND PrLI42 CO-ELUTION SEC CHROMATOGRAMS.....	99
FIGURE 4.2. PURIFICATION OF RECOMBINANT <i>B. SUBTILIS</i> N-RsBRA	101
FIGURE 4.3. PURIFICATION OF RECOMBINANT <i>B. SUBTILIS</i> N-RsBRB	101
FIGURE 4.4. PURIFICATION OF RECOMBINANT <i>B. SUBTILIS</i> N-RsBRC	102
FIGURE 4.5. PURIFICATION OF RECOMBINANT <i>B. SUBTILIS</i> N-RsBRD	103
FIGURE 4.6. FLUORESCENCE POLARISATION EXPERIMENT TO QUANTIFY THE INTERACTION BETWEEN <i>LMO</i> PrLI42 AND <i>LMO</i> N-RsBRs.	104
FIGURE 4.7. FLUORESCENCE POLARISATION EXPERIMENT TO QUANTIFY THE INTERACTION BETWEEN <i>BSU</i> PrLI42 AND <i>BSU</i> N-RsBRs.	105
FIGURE 4.8. INTERACTION OF THE PrLI42 MINIPROTEIN WITH THE N-RsBR PROTEINS BY FLUORESCENCE POLARISATION	106
FIGURE 4.9. THERMOSTABILITY OF <i>LMO</i> N-RsBRs USING 2 DIFFERENT METHODS	109
FIGURE 4.10. THERMOSTABILITY OF <i>BSU</i> N-RsBRs USING 2 DIFFERENT METHODS.....	110
FIGURE 4.11. EFFECT OF THE ADENINE-CONTAINING COMPOUNDS ON N-RsBR THERMOSTABILITY.	111
FIGURE 4.12. EFFECT OF AMINO ACIDS ON N-RsBRs THERMOSTABILITY.	113
FIGURE 4.13. THERMOSTABILITY SHIFT ASSAY OF THE N-RsBRs WITH CHEMICAL COMPOUNDS.	115
FIGURE 4.14. ELEMENTAL ABSORPTION EDGES OF BROMINE AND IODINE	116
FIGURE 4.15. ELECTRON DENSITY MAPS OF N-RsBR3 DIMER INTERFACE WITH FG8.....	118
CHAPTER V: NATIVE STRESSOSOME AND VARIANT STRESSOSOME STUDY	121
FIGURE 5.1. <i>BSU</i> NR5623 NATIVE STRESSOSOME PURIFICATION.....	124
FIGURE 5.2. <i>BSU</i> NR5623 NATIVE STRESSOSOME VISUALIZATION BY NEGATIVE STAIN TEM	125
FIGURE 5.3. <i>BSU</i> NR5621 NATIVE STRESSOSOME VISUALIZATION	126
FIGURE 5.4. SCHEMATIC REPRESENTATION OF THE NATIVE STRESSOSOME PURIFICATION USING N-RsBRA ANTIBODIES	129
FIGURE 5.5. NATIVE STRESSOSOME PURIFICATION USING N-RsBRA ANTIBODIES	130

FIGURE 5.6. <i>BSU</i> STRAIN 168 NATIVE STRESSOSOME VISUALIZATION	131
FIGURE 5.7. PURIFICATION OF RsbR1-RsbS RECOMBINANT STRESSOSOME COMPLEX USING A CO-EXPRESSION VECTOR	133
FIGURE 5.8. S200 16/60 SEC OF RsbR1 RECOMBINANT PROTEIN	134
FIGURE 5.9. S200 16/60 SEC OF RsbR2 RECOMBINANT PROTEIN PURIFICATION	134
FIGURE 5.10. S200 16/60 SEC OF RsbR3 RECOMBINANT PROTEIN	135
FIGURE 5.11. S200 16/60 SEC OF RsbR4 RECOMBINANT PROTEIN	136
FIGURE 5.12. S200 16/60 SEC OF RsbS RECOMBINANT PROTEIN	136
FIGURE 5.13. RECONSTITUTION AND VISUALIZATION OF THE RsbR1-RsbS STRESSOSOME COMPLEX.....	138
FIGURE 5.14. RECONSTITUTION AND VISUALIZATION OF THE RsbR2-RsbS STRESSOSOME COMPLEX.....	139
FIGURE 5.15. RECONSTITUTION AND VISUALIZATION OF THE RsbR3-RsbS STRESSOSOME COMPLEX.....	140
FIGURE 5.16. CRYO-EM GRID SCREENING OF THE RsbR3-RsbS STRESSOSOME COMPLEX	141
FIGURE 5.17. RECONSTITUTION AND VISUALIZATION OF THE RsbR4-RsbS STRESSOSOME COMPLEX.....	142
FIGURE 5.18. MBP FUSED RsbT PROTEIN PURIFICATION	144
FIGURE 5.19. S200 16/60 SEC OF RsbX PHOSPHATASE PROTEIN	145
FIGURE 5.20. RECONSTITUTION OF THE RsbR1-RsbS-RsbT RECOMBINANT STRESSOSOME COMPLEX.....	146
FIGURE 5.21. RECONSTITUTION OF THE RsbR1-RsbS-RsbX STRESSOSOME COMPLEX	147
FIGURE 5.22. VISUALISATION OF THE RST AND RSX COMPLEXES BY NEGATIVE STAIN	148

CHAPTER VI: CRYO-EM SINGLE PARTICLE ANALYSIS OF THE *L. MONOCYTOGENES* RSBR1-RSBS AND RSBR1-RSBS-RSBX STRESSOSOME COMPLEXES

FIGURE 6.1. CRYO-EM MICROGRAPHS OF THE <i>LmORS</i> AND <i>LmORSX</i> STRESSOSOME COMPLEXES	152
FIGURE 6.2. CRYO-EM DATA PROCESSING FLOW-CHART FOR THE <i>LmORS</i> COMPLEX	153
FIGURE 6.3. CRYO-EM DATA PROCESSING FLOW-CHART FOR THE <i>LmORSX</i> COMPLEX	154
FIGURE 6.4. EXAMPLES OF 2D CLASSES FOR <i>LmORS</i> AND <i>LmORSX</i> PARTICLES.....	155
FIGURE 6.5. 3D CLASSIFICATION OF <i>LmORS</i> AND <i>LmORSX</i> PARTICLES	156
FIGURE 6.6. SINGLE PARTICLE RECONSTRUCTION OF THE <i>LmORS</i> COMPLEX	158
FIGURE 6.7. SINGLE PARTICLE RECONSTRUCTION OF THE <i>LmORSX</i> COMPLEX	159
FIGURE 6.8. FSC CURVES AND MASKS FOR THE <i>LmORS</i> C1 RECONSTRUCTION	160
FIGURE 6.9. LOCAL RESOLUTION MAP OF THE <i>LmORS</i> C1 MAP.....	161
FIGURE 6.10. HOMOLGY MODELS FOR RsbR AND RsbS STAS DOMAINS	162
FIGURE 6.11. J-HELIX CONFORMATION IN THE <i>LmORS</i> C1 MAP	162
FIGURE 6.12. ATOMIC MODEL OF THE C1 <i>LmORS</i> RECONSTRUCTION SHOWING THE DISTRIBUTION OF THE RsbR1 TURRETS. ...	164
FIGURE 6.13. ELECTRONIC POTENTIAL MAP OF THE C1 <i>LmORS</i> STRESSOSOME RECONSTRUCTION	165
FIGURE 6.14. FSC CURVES FOR THE <i>LmORS</i> RECONSTRUCTION IN D2 SYMMETRY	166
FIGURE 6.15. LOCAL RESOLUTION MAP OF THE <i>LmORS</i> D2 MAP	167
FIGURE 6.16. ATOMIC MODEL OF THE D2 <i>LmORS</i> RECONSTRUCTION	168
FIGURE 6.17. SCHEMATIC REPRESENTATION OF THE <i>LmORS</i> COMPLEX ASSEMBLY IN C1 SYMMETRY.....	171
FIGURE 6.18. SCHEMATIC REPRESENTATION OF THE <i>LmORS</i> COMPLEX ASSEMBLY IN D2 SYMMETRY	172
FIGURE 6.19. RsbT PHOSPHORYLATION TARGETS ON RsbR1 AND RsbS	173
FIGURE 6.20. PHOSPHORYLATION SITES ON THE <i>LmORS</i> STRESSOSOME MODEL	174

List of Tables

CHAPTER II: MATERIALS AND METHODS	29
TABLE 2.1. BACTERIAL STRAINS USED IN THIS PROJECT	30
TABLE 2.2. MoCLO THERMAL CYCLING CONDITIONS	34
TABLE 2.3. PRIMERS USED TO AMPLIFY THE NTD OF RsbR GENES.	34
TABLE 2.4. PCR CYCLES USED FOR DNA AMPLIFICATION	34
TABLE 2.5. PRIMERS USED FOR SITE-DIRECTED MUTAGENESIS	35
TABLE 2.6. PRIMERS USED FOR MBP-FUSION PROTEINS	36
TABLE 2.7: RECIPE USED FOR 15 % SDS-PAGE GELS	39
TABLE 2.8. LIGAND SCREEN FOR THE TSA AGAINST THE RsbR N-TERMINAL DOMAINS	48
TABLE 2.9. PrLI42 CYTOSOLIC TAIL SEQUENCE.....	49
TABLE 2.10: DATA COLLECTION PARAMETERS OF <i>LmORS</i> SAMPLE	56
TABLE 2.11: DATA COLLECTION PARAMETERS OF <i>LmORSX</i> SAMPLE	57
CHAPTER III: STRUCTURAL STUDIES OF THE STRESS SENSORY DOMAINS OF THE <i>L. MONOCYTOGENES</i>	
STRESSOSOME.....	59
TABLE 3.1. NATIVE PROTEIN CRYSTALS OF <i>Lmo</i> N-RsbRs	67
TABLE 3.2. DATA COLLECTION STATISTICS FOR N-RsbR PROTEINS.....	68
TABLE 3.3. SEQUENCE IDENTITY MATRIX BETWEEN N-RsbR PROTEINS.....	69
TABLE 3.4. Se-MET PROTEIN CRYSTALS OF <i>Lmo</i> N-RsbRs	74
TABLE 3.5. DATA COLLECTION STATISTICS FOR SAD EXPERIMENTS	75
TABLE 3.6. N-RsbR1 S-SAD DATA COLLECTION STATISTICS	78
TABLE 3.7. REFINEMENT STATISTICS FOR Se-SAD N-RsbR2	81
TABLE 3.8. SAXS DATA STATISTICS OF N-RsbR1 AND N-RsbR2.....	85
TABLE 3.9. REFINEMENT STATISTICS FOR Se-SAD N-RsbR2	87
TABLE 3.10: RMSD AND SEQUENCE IDENTITY VALUES FOR MONOMER STRUCTURE COMPARISON.....	95
CHAPTER IV: INTERACTION STUDY OF RsbR WITH POTENTIAL BINDING PARTNERS	97
TABLE 4.1. CALCULATED K _D OF THE FLUORESCENCE POLARIZATION DATA FOR <i>Lmo</i> N-RsbR WITH <i>Lmo</i> PrLI42.....	105
TABLE 4.2. CALCULATED K _D VALUES OF THE FLUORESCENCE POLARIZATION DATA FOR <i>Bsu</i> N-RsbR WITH <i>Bsu</i> PrLI42.....	106
TABLE 4.3. K _D VALUES OBTAINED FOR THE FP EXPERIMENTS AT 40 μM.....	107
TABLE 4.5. DATA COLLECTION STATISTICS FOR N-RsbR3 WITH FRAGLITES	117
TABLE 4.4. POTENTIAL BINDING FRAGLITES.	117
CHAPTER V: NATIVE STRESSOSOME AND VARIANT STRESSOSOME STUDY	121
TABLE 5.1. NRS5623 STRAIN PEPTIDE MASS FINGERPRINTING.	127
TABLE 5.2. NRS5621 STRAIN PEPTIDE MASS FINGERPRINTING.	127
CHAPTER VI: CRYO-EM SINGLE PARTICLE ANALYSIS OF THE <i>L. MONOCYTOGENES</i> RsbR1-RSBS AND RsbR1-RSBS-RSBX STRESSOSOME COMPLEXES	150

TABLE 6.1. C1 MAP PROCESSING AND MODEL REFINEMENT/VALIDATION STATISTICS	163
TABLE 6.2. D2 MAP DATA PROCESSING AND MODEL REFINEMENT STATISTICS.....	169

Abbreviations list

2D	Two-dimensional
³²P	Phosphorus-32
3D	Three-dimensional
A₂₈₀	Absorbance at 280 nm
ADP	Adenosine diphosphate
AKR	Acrylic acid
APS	Ammonium persulfate
ASU	Asymmetric Unit
ATP	Adenosine triphosphate
<i>B. subtilis</i>	<i>Bacillus subtilis</i>
bp	base pair
Br	Bromine
<i>Bsu</i>	<i>Bacillus subtilis</i>
CaCl₂	Calcium chloride
CC	Correlation coefficient
CCP4	Collaborative Computational Project No. 4
CD	Circular dichroism
CMOS	Complementary metal-oxide semiconductor
COFRADIC	Combined fractional diagonal chromatography
COOT	Crystallographic Object-Oriented Toolkit
CRB	Cambridge Research Biochemicals
Cryo-EM	Electron cryo-microscopy
ctc	Catabolite controlled
CTD	C-terminal domain
CTF	Contrast transfer function
CV	Column volume
DE	Germany
DLS	Diamond Light Source
DMSO	Dimethyl sulfoxide
DTT	Dithiothreitol
DVA	Destination Vector Ampicillin
DVK	Destination Vector Kanamycin
<i>E. coli</i>	<i>Escherichia coli</i>
EDTA	Ethylenediaminetetraacetic acid
EGD	Everitt George Dunne
EM	Electron microscopy
EMDB	Electron microscopy database
emPAI	Exponentially modified protein abundance index
ES	Spain
FF	Fast Flow
FFT	Fast Fourier Transform
FG	Fraglite
FMN	Flavin mononucleotide

FMT	Formic acid
FOM	Figure of merit
FP	Fluorescence polarisation
FSC	Fourier shell correlation
FT	Flow-through
GFP	Green fluorescence protein
GIT	Gastrointestinal tract
GOI	Gene of interest
His-tag	histidine tag
IN	Input
IPTG	Isopropyl β -D-1-thiogalactopyranoside
Kb	Kilobase
Kd	Dissociation constant
keV	kilo electron volt
Kpsi	Kilo-pounds per-square-inch
<i>L. ivanovii</i>	<i>Listeria ivanovii</i>
<i>L. monocytogenes</i>	<i>Listeria monocytogenes</i>
LB	Lysogeny broth
LC-MS	Liquid chromatography mass spectrometry
<i>Lmo</i>	<i>Listeria monocytogenes</i>
LOV	Light-oxygen-voltage-sensing domain
<i>M. thermoacetica</i>	<i>Moorella thermoacetica</i>
Mbp	Mega base pair
MBP	Maltose binding protein
MDa	Megadalton
Mg	Magnesium
MgCl₂	Magnesium chloride
MMT	DL-Malic acid, MES (4-Morpholineethanesulfonic acid) monohydrate, Tris
mP	millipolarisation
MR	Molecular replacement
mRNA	messenger Ribonucleic acid
MS	Mass spectrometry
MSE	Selenomethionine
MW	Molecular weight
N-RsbR	N-terminal domain of RsbR
N/A	Non applicable
NaOH	Sodium hydroxide
NCBI	National Centre for Biotechnology Information
Ni-NTA	Nickel-Nitriloacetic acid
NL	Netherlands
NMR	Nuclear magnetic resonance
NTD	N-terminal domain
NTP	Nucleoside triphosphate
OD₆₀₀	Optic density at 600 nm

orf	Open reading frame
P	Pellet
PBP1	Penicillin Binding Protein 1
PBS	Phosphate-buffered saline
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
PDF	Peptide DeFormylase
pI	Isoelectric point
RMSD	Root means square deviation
RNA	Ribonucleic acid
RNAP	Ribonucleic acid polymerase
rpm	Revolutions per minute
Rsb	Regulator of sigma B
RST	RsbR-RsbS-RsbT
RSX	RsbR-RsbS-RsbX
S-SAD	Sulphur single-wavelength anomalous dispersion
SAD	Single-wavelength anomalous dispersion
SAXS	Small angle X-ray scattering
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
Se	Selenium
Se-Met	Selenomethionine
SEC	Size exclusion chromatography
sigB	sigma B factor
SPR	Surface Plasmon Resonance
STAS	Sulphate transporter and Anti sigma factor
TAE	Tris-acetate-EDTA
TEM	Transmission electron microscopy
TEMED	Tetramethylethylenediamine
TIS	Translation initiation site
Tm	Melting temperature
TSA	Thermostability shift assay
UV	Ultraviolet
<i>V. vulnificus</i>	<i>Vibrio vulnificus</i>
Ve	Elution volume
WT	Wild type
X-gal	5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside

Chapter I: Introduction

1.1 Studied species

1.1.1 *Listeria monocytogenes*

Listeria monocytogenes, a rod-shaped Gram-positive bacterium, was first characterized in 1926 by several bacteriologists including E.G.D Murray (Murray et al., 1926). This bacterium was originally named *Bacterium monocytogenes* due to a characteristic monocytosis found in infected lab rabbits and Guinea pigs. It was later classified in the *Corynebacteriaceae* phylum in Bergey's Manual of Systematic Bacteriology 7th edition (Breed et al., 1958). It is only after a major epidemic outbreak of listeriosis in Germany in 1949 in new-borns and stillborn babies that *L. monocytogenes* was classified in the *Listeria* genus and renamed by H.P.R Seeliger (Seeliger and Höhne, 1979). The origin of *L. monocytogenes* remained unknown until another listeriosis outbreak in the late 1980, this time in adults. This outbreak allowed the infection's cause to be identified, which was due to the ingestion of spoiled food, leading to the description of *L. monocytogenes* as a foodborne pathogen (Farber and Peterkin, 1991).

To date, 17 *Listeria* species have been characterized and classified in the *Listeria* genus from the *Listeriaceae* family. Within the *Listeria* genus, 2 distinct groups have been distinguished due to their phenotypic characteristics, the *Listeria sensu stricto* (restricted) and *Listeria sensu lato* groups. The *Listeria sensu stricto* group contains the species that are motile and grow at low temperatures and were all isolated from the mammalian gut while the *sensu lato* group species were isolated from food related surfaces and the environment (Schardt et al., 2017). Six *Listeria* species (**Figure 1.1**) are within the *Listeria sensu stricto* group of which two are pathogenic, *L. ivanovii* and *L. monocytogenes*. While *L. monocytogenes* infects both humans and animals, *L. ivanovii* infects mainly ruminants, and humans via zoonosis, due to a limited distribution in the environment of these species (Guillet et al., 2010).

L. monocytogenes is usually found in the soil (Deng et al., 2010) and ruminants (e.g. cows, sheep) can become contaminated by ingesting spoiled water or silage. Food contamination can occur either by using contaminated milk or meat, or by using infected materials e.g., human cross-contaminated utensils during food processing. Subsequently, *L. monocytogenes* is found in a lot of raw and processed foods, such as dairy (milk and cheese) and meat products (dried meat and pâté).

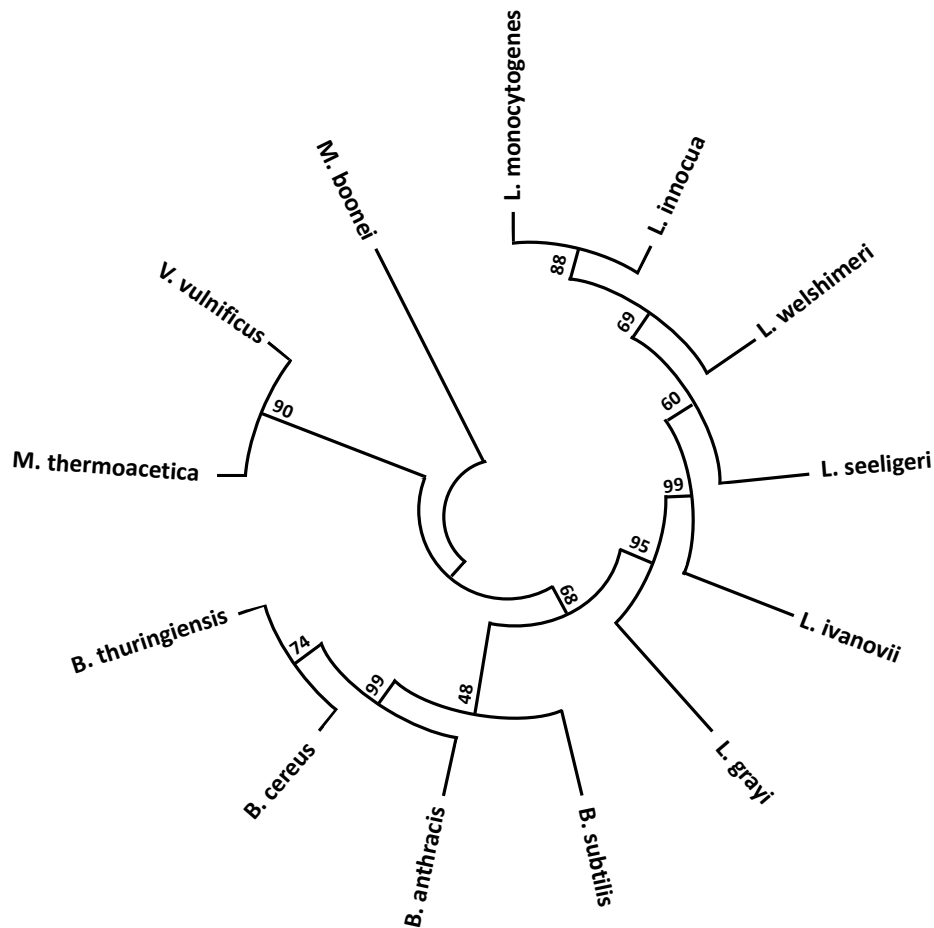


Figure 1.1. 16S phylogenetic tree for key *Bacillus* and *Listeria* species

The phylogenetic tree shows the clustering of the *Listeria* species and *Bacillus* species, while being phylogenetically distanced from *Vibrio vulnificus* and *Moorella thermoacetica*. The tree was built using the species' 16S rRNA nucleotide sequences. The sequences were aligned using the MUSCLE algorithm (Edgar, 2004). The tree was built with the Maximum Likelihood algorithm using the Tamura Nei model in the MEGA X software (Kumar et al., 2018). The values represent the percentage of the bootstrap value. The tree is rooted with a member of the Euryarchaeota, *Mehtanoregula boonei* to differentiate the species' clusters.

The *L. monocytogenes* pathogen affects mostly immunocompromised people, pregnant women, the foetus, new-born babies and the elderly (de Noordhout et al., 2014). The ingestion of this pathogen can lead to mild to severe diseases depending on the nature of the infection. *L. monocytogenes* can stay in the gastrointestinal tract (GIT) and can cause mild gastrointestinal infection without a systemic infection. However, the bacteria can cross the GIT-blood barrier to cause severe diseases such as bacteraemia, meningitis and encephalitis, which leads to death in most cases.

Despite the fact that several detection tests are performed on food before it is marketed, most reported listeriosis cases are due to the ingestion of contaminated food (EFSA, 2015). One key characteristic of *L. monocytogenes* that makes it particularly difficult to eliminate from

food is its ability to survive at the high salt concentrations and low temperatures at which food is usually processed and stored. *L. monocytogenes* can also survive in the presence of the antiseptic reagent ethanol, whereas most bacteria cannot survive or infect food under these conditions. The ability of *L. monocytogenes* to survive low temperature and high salt/ethanol conditions makes it rather difficult to eliminate from food. *L. monocytogenes* as a foodborne pathogen is thus a good subject to understand the ability of bacteria to survive the extreme conditions to which they are exposed during and after food processing.

1.1.2. *Bacillus subtilis*

B. subtilis is a Gram-positive, spore forming and motile bacterium. *B. subtilis* is widely used as a model organism for Gram-positive species, particularly to understand sporulation and biofilm formation. Compared to its close phylogenetic relatives, such as *B. anthracis* that causes anthrax and *B. cereus* that causes food poisoning (**Figure 1.1**), *B. subtilis* is non-pathogenic, which makes it a safe species for academic studies and industrial use for enzyme production and protein production due to its high-density fermentation properties (Borriss et al., 2017). It is also closely related to the *L. monocytogenes* pathogen as there are many common features in their genome (Danchin, 2001), such as the stress response genes and the regulators.

B. subtilis was first named *Vibrio subtilis* by Christian Gottfried Ehrenberg (Ehrenberg C.G., 1835) in 1835 and subsequently renamed to *B. subtilis* by Ferdinand Cohn in 1872 (Cohn, 1875). Even back in the 19th century, it was one of the first bacteria to be studied (Harwood, 1989). The *B. subtilis* genome was one of the first greater than 4 Mbp to be sequenced, and its sequence was published in 1997 by a European/Japanese consortium led by the Pasteur Institute (Kunst et al., 1997). Now, there are several websites and webserver dedicated to *B. subtilis*, such as SubtiList (Moszer et al., 1995) and SubtiWiki (Zhu and Stülke, 2018) which gather together genome information, details of pathways, known interaction partners, gene regulation, and expression under a virtual roof.

B. subtilis is a ubiquitous species found predominantly in soils but also in aquatic environments, in plant roots and in the GIT. *B. subtilis* can adapt and grow in a wide range of environments (Earl et al., 2008) because of a range of late adaptive response pathways utilised by this bacterium. These signal transduction systems underpin, for example, sporulation (Strauch and Hoch, 1993), genetic competence (Dubnau, 1991), motility (Cairns et al., 2013), biofilm formation (Mielich-Süss and Lopez, 2015), and the general stress response that is coordinated by the stressosome (Hecker and Völker, 2001).

B. subtilis and its close phylogenetic relative, *L. monocytogenes*, both have the ability to survive and overcome environmental stresses, such as the presence of ethanol and salt. This

stress-resistance ability is due to the presence of a supramolecular protein complex, called the stressosome, which activates the RNA polymerase (RNAP) sigma B (sigB, or σ^B) factor, which in turn controls a regulon encoding approximately 200 proteins that are involved in the general stress response (Price et al., 2001)

1.2. The alternative sigma factor sigB

Differential regulation is a process that allows different sets of genes to be expressed to allow cell survival under variable growth and stress conditions. While DNA replication is carried out by the DNA polymerase, the transcription of DNA into RNA is catalysed by RNA polymerase (RNAP). Only one type of RNAP is found in bacterial species compared to at least three in eukaryotic cells (Cooper, 2000). Bacterial RNAP is a multi-subunit complex composed of five catalytic subunits: two α subunits, one β , one β' , and a fifth small ω subunit (**Figure 1.2**). RNAP in its minimal form ($2\alpha\beta\beta'\omega$) catalyses the elongation and termination steps of transcription, but it is not able to initiate the transcription of specific target genes (Paget and Helmann, 2003). In order for RNAP to initiate transcription, it must be in its holoenzyme form, which is achieved by the binding of a small regulatory sigma factor, σ (Burgess et al., 1969). One of the functions of σ is to recognise specific regions within the promoters of genes about to be transcribed. Many sigma factors are encoded by bacteria, allowing each to induce the transcription of alternative subsets of genes. For example, *E. coli* σ^{70} drives transcription of genes that are involved in normal growth, while σ^{28} is required for the transcription of genes for the flagellar motor of the cell (Kazmierczak et al., 2005). In *B. subtilis* the house-keeping sigma factor is σ^A , which is similar to *E. coli* σ^{70} factor, and is actively used during growth. *B. subtilis* encodes at least 17 alternative sigma factors with diverse roles, such as the general stress response with σ^B , chemotaxis with σ^D , and competence and early sporulation with σ^H (Haldenwang, 1995). One of the first alternative sigma factors to be discovered was σ^B , previously called σ^{37} , and it plays a role in the general stress response (Haldenwang and Losick, 1980), discussed below. By binding to the RNA polymerase core complex, σ^B leads to the recognition by the holoenzyme of the promoters of the approximately 200 genes that are upregulated in response to the recognition of general stress.

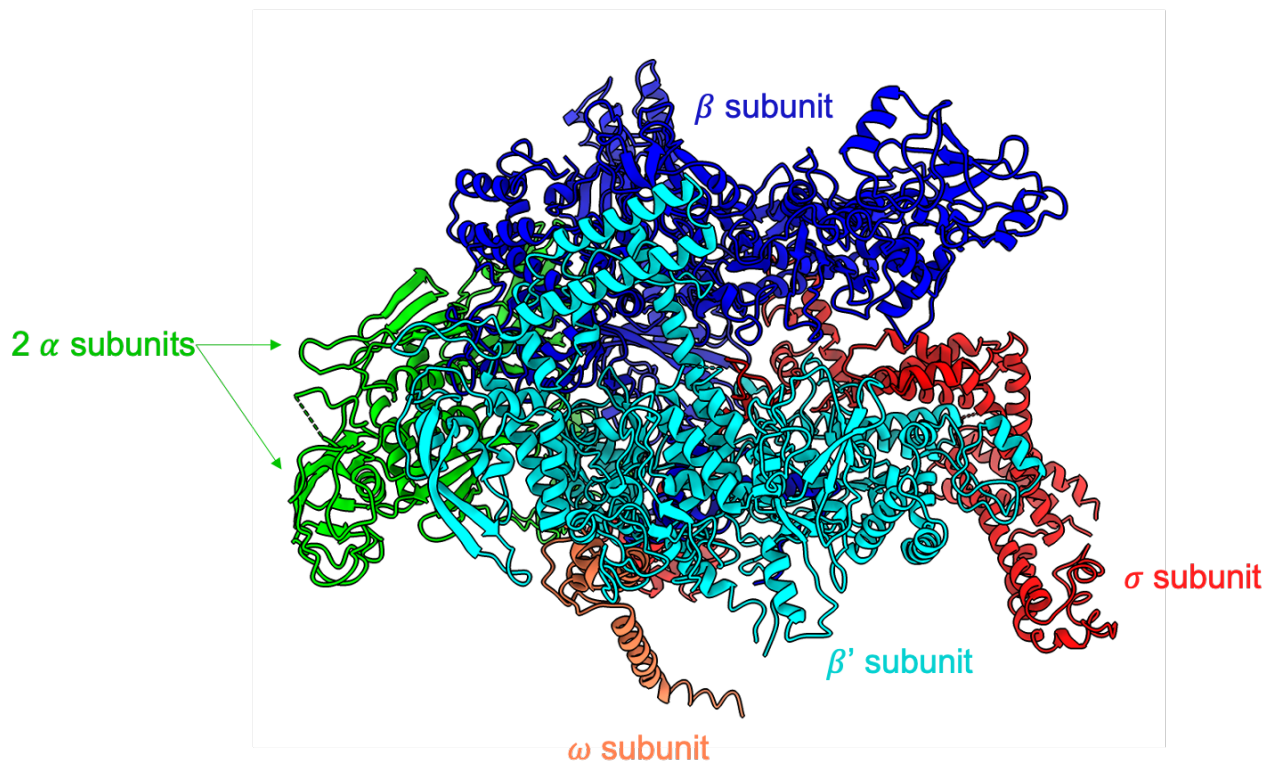


Figure 1.2. *E. coli* RNAP holoenzyme structure

The structure of the RNAP holoenzyme has been obtained by X-ray crystallography. The RNAP holoenzyme is composed of six subunits: two α subunits in green shades, one β subunit in dark blue, one β' subunit in cyan, one ω subunit in orange and one σ subunit (RpoD) in red (PDB ID 4YG2) (Murakami, 2013).

In *B. subtilis*, σ^B has been shown to play an important role during stationary phase (Haldenwang WG and Losick R, 1980) and controls the *csb* genes, which includes the *ctc* (catabolite controlled) gene, encoding for a ribosomal protein, which was shown to be activated during heat shock, oxidative stress, salt stress and glucose or oxygen limitation (Haldenwang, 1995). In one of the first σ^B gene regulation studies (Boylan et al., 1993a) the authors used blue-white screening using a *lacZ* reporter fused to genes in the *csb* region. In this study σ^B was found to regulate 11 *csb* genes and two of these genes were shown to be exclusively dependent of σ^B : *csb22* (unknown function) and *csb34* (encodes a sugar transporter). Moreover, the discovery of one *csb* gene that encodes an UDP-glucose phosphorylase showed that σ^B is involved in other environmental stresses such as high osmolarity, in addition to the one endured during stationary phase (Boylan et al., 1993b, 1993a). In a DNA microarray analysis, 125 σ^B -dependent genes were identified (Petersohn et al., 2001) of which 24 were also induced during σ^B -independent stresses; 62 of these genes were validated and were found in previous studies,

while 14 new genes were identified with the DNA microarray analysis, increasing the number of known σ^B dependent genes. Many genes are also down regulated during ethanol stress response, including genes involved in cell wall synthesis (*rodA*) and transcription (*rpoE*) (Price et al., 2001).

In *L. monocytogenes*, there are 170 genes whose expression is dependent on σ^B (Kazmierczak et al., 2003). By the application of statistics and microarray analysis, 54 genes were found to have a significant expression ratio change on the imposition of an osmotic stress (0.5 M KCl). Within these 54 *L. monocytogenes* genes, 31 have homologous gene sequences to *B. subtilis* genes and 12 of these genes are σ^B dependent. The 54 genes have been classified in ten functional groups, including transport and metabolism (20 genes), stress and virulence (13 genes) and transcriptional regulation. The transport group includes genes from the *opuC* operon, involved in the osmolyte transport system, which confers osmo-resistance when the cell is exposed to high concentrations of salt. The stress group includes enzymes such as the low-temperature-requirement C protein (*ltrC*) (or phosphatidylglycerophosphatase A), while the virulence group includes the Internalin proteins required to mediate the adhesion and internalisation of the pathogen into the mammalian host cell (Bonazzi et al., 2009). The transcriptional regulation group includes *rsbV*, *rsbW*, *sigB* and *RsbX* genes that are found in the *rsb* operon.

The phylogenetic similarity of *B. subtilis* and *L. monocytogenes* explains the similar regulation of the σ^B in both organisms, apart from the virulence genes regulated by σ^B in *L. monocytogenes*.

1.3. The σ^B signalling hub

The environmental stress sensing activity common to *B. subtilis* and *L. monocytogenes* involves several proteins that are found in the *rsb* operon. The proteins that comprise the stressosome, RsbR, RsbS and RsbT, are found in an upstream genetic module with the phosphatase RsbU (**Figure 1.3**). The upstream module is regulated by the house-keeping σ^A factor and the transcription of the module leads to a polycistronic mRNA encoding proteins comprising the entire σ^B operon (RsbR, RsbS, RsbT, RsbU, RsbV, RsbW, σ^B and RsbX) (Wise and Price, 1995).

The genes in the downstream module are involved in the partner-switch and the signalling cascade pathway that leads to the expression of genes involved in cellular stress and include the anti-anti-sigma factor RsbV, the RsbW kinase, σ^B , and the RsbX phosphatase; this downstream module is regulated by σ^B itself (Wise and Price, 1995). This *rsb* operon

organisation of both *L. monocytogenes* and *B. subtilis* (**Figure 1.3**) is conserved in many low-GC content Gram positive bacteria. However, the genetic location of *rsbX* changes in other species such as in *Moorella thermoacetica* and *Vibrio vulnificus*; the gene of the phosphatase RsbX is found in the up-stream module compared to its position in *L. monocytogenes* and *B. subtilis* (Pané-Farré et al., 2017).

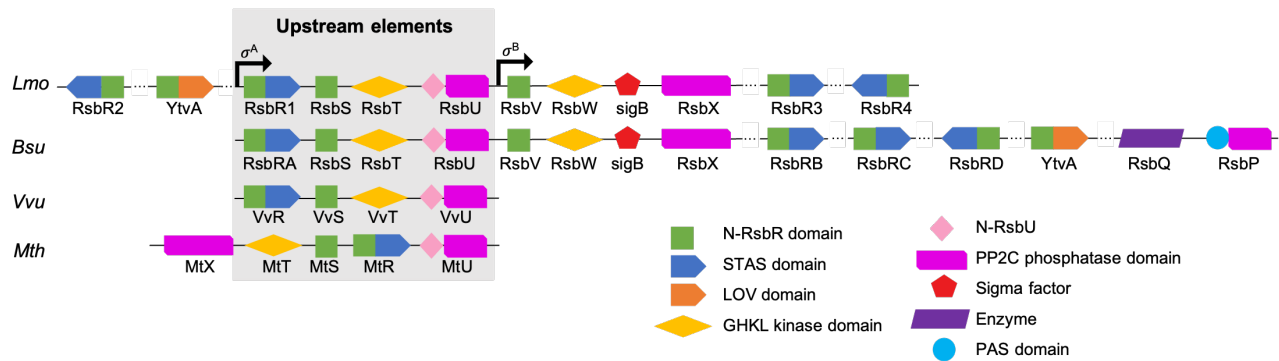


Figure 1.3. Organisation of the *rsb* operon

The genetic organisation of the *rsb* operon were taken from the KEGG genome for *V. vulnificus* (*Vvu*), from the NCBI database for *L. monocytogenes* (*Lmo*) and *M. thermoacetica* (*Mth*) and from the SubtiWiki website for *B. subtilis* (*Bsu*). The arrow representing the genes are pointing in the direction of transcription: pointing to the right shows the genes on the leading strand and pointing to the left shows the genes on the lagging strand. The genes usually cluster in close proximity to each other, except for the paralogues in *Lmo* and *Bsu* and the RsbP/Q pair in *Bsu* which are found very distant from the *rsb* operon. The upstream elements are highlighted in grey are regulated by σ^A . Note that RsbX has distinct genomic locations in *Lmo* and *Bsu* compared to *Mth*. The remaining genes are the downstream elements, except the RsbR paralogues, which includes RsbU, RsbV, RsbW, and σ^B , regulated by σ^B itself.

Transcription under the control of σ^B can be activated by two different routes, the stressosome-dependent and -independent pathways. Only the stressosome-dependent system is present in *L. monocytogenes*, while both are present in *B. subtilis*. These responses are distinct in their dependency upon the stressosome; environmental stresses are stressosome-dependent, whereas nutritional stresses are stressosome-independent and involve the RsbP/RsbQ activator/phosphatase pair (Brody et al., 2001). However, they both converge on the RNA polymerase alternative sigma factor, σ^B . The stressosome-dependent and independent pathway share common partners; therefore, each pathway is described separately below until they converge.

1.3.1. The stressosome-dependent pathway

The stressosome is a 1.8 MDa supramolecular protein complex widely distributed in the bacteria kingdom (Pané-Farré et al., 2005). It was first identified in *B. subtilis* (Dufour et al., 1996) and is currently the best-understood model of stressosome structure and function. The stressosome is composed of the three proteins found in the upstream module of the *rsb* operon: RsbRA, RsbS and RsbT. RsbRA has 4 other paralogues: RsbRB, RsbC, RsbRD and YtvA (Akbar et al., 2001), and their function and stoichiometry within the complex remains unknown. RsbS is a single STAS domain-containing protein and within the stressosome acts as a scaffolding protein in the core (Marles-Wright et al., 2008). The RsbR proteins have 2 domains: a C-terminal domain (CTD) similar to the RsbS STAS domain, also anchored in the core, and a globin-like domain that protrudes from the core to form turrets. RsbT is a serine/threonine kinase that mediates the signalling cascade pathway. The stressosome is usually activated under environmental stresses, such as high extra-cellular salt concentrations (Benson and Haldenwang, 1993) and ethanol shock (Boylan et al., 1993a). The partner-switch and signalling cascade pathway downstream of the stressosome is well known and is described in **Figure 1.4**.

In an unstressed cell (**Figure 1.4A**), the RsbT kinase is bound to the core stressosome complex, which holds it in an inactive state ready for phosphorylation and prevents activation of the downstream cascade. RsbT is a serine-threonine kinase that belongs to the GHKL kinase superfamily (Dutta and Inouye, 2000) and is a positive regulator of the σ^B signalling cascade pathway. Whenever environmental stress is “sensed”, RsbT is triggered and activated (Chen et al., 2004) (**Figure 1.4B**). This activation leads to the phosphorylation of the RsbR and RsbS STAS domains on serine and threonine residues (Gaidenko et al., 1999; Yang et al., 1996). Even though the RsbT activation mechanism remains unknown, several studies have been conducted to understand how RsbT acts in the σ^B activation pathway. RsbT was first identified as a kinase that phosphorylates RsbS (Kang et al., 1996), where it was reported that in RsbT mutant *B. subtilis* strains (Δ RsbT), σ^B activation was totally abolished after environmental stress (salt stress). It was then reported that RsbS phosphorylation was a key modification for its function. Sequence similarities with the homologues RsbW and SpoIIAB indicated that RsbT should have a kinase activity. As the SpoIIAA:SpoIIAB: σ^F model was well known, and the phosphorylated residue of SpoIIAA was identified as Ser58 (Najafi et al., 1995), sequence similarities between RsbS and SpoIIAA suggested a conserved serine residue at position 59 in *B. subtilis* RsbS was the likely site of phosphorylation.

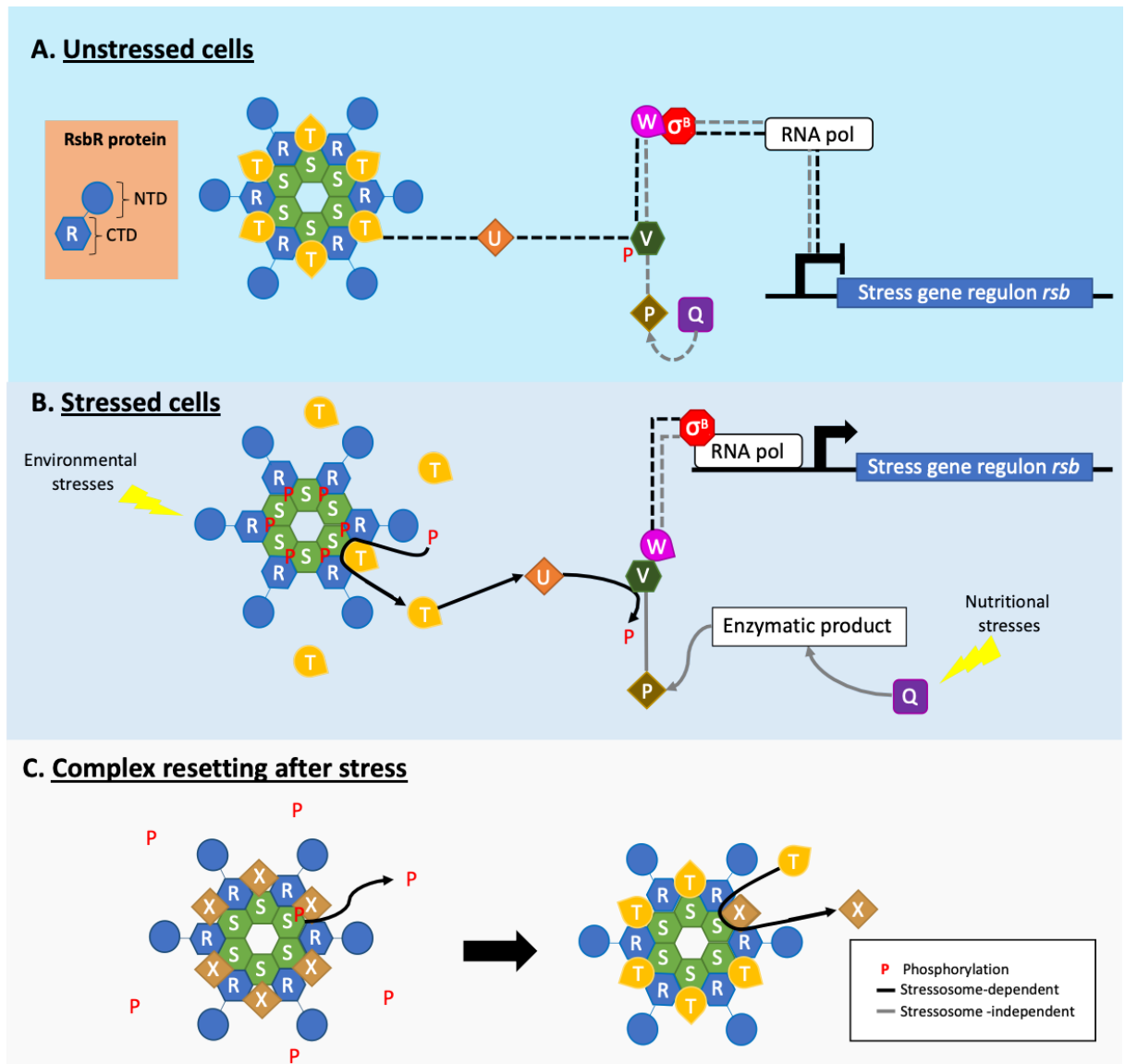


Figure 1.4. Schematic representation of the signalisation pathway in unstressed and stressed cells

The continuous lines indicate interactions between proteins and dashed lines represent where interactions are lost depending upon the stress status of the cell. **A.** In an unstressed cell, RsbT is sequestered by the stressosome and cannot activate the RsbU (U) phosphatase activity against phosphorylated anti-anti-sigma factor, RsbV-P(V). **B.** Upon stress detection by the presumed stress sensor, the NTD of RsbR (R), RsbT phosphorylates residues on the STAS domains which induces a partner-switching cascade started by the release of RsbT from the stressosome. Free RsbT activates RsbU, which in turn dephosphorylates RsbV-P. Dephosphorylated RsbV has a greater affinity to RsbW (W) and releases σ^B to interact with RNA polymerase to induce the expression of the stress regulon. **C.** The stressosome complex is reset by the phosphatase RsbX, and RsbT is sequestered.

RsbT was shown to be specific to RsbS at Ser59 by *in vitro* phosphorylation and dephosphorylation assays, where RsbT was shown to specifically phosphorylate RsbS on Ser59, as the S59A mutation abolished the RsbT kinase activity to RsbS (Yang et al., 1996). Subsequently, it was shown that RsbT also phosphorylates the RsbR STAS domain (Gaidenko et al., 1999). In fact, *in vitro* phosphorylation assays showed that RsbT phosphorylates both RsbR and RsbS (Gaidenko et al., 1999). The addition of RsbR to the RsbS-RsbT mixture showed an increase in the RsbS phosphorylation rate (Gaidenko et al., 1999). As RsbR was also phosphorylated, potential serine and threonine residues were identified using sequence comparisons to RsbS. Thr205 in the RsbR STAS domain aligns to Ser59 in RsbS and was first proposed as the target for RsbT phosphorylation. *In vivo* assays demonstrated the importance of RsbR Thr205: when mutated into an alanine, it showed an opposite effect of σ^B regulation (Akbar et al., 2001). To further analyse RsbR phosphorylation, Gaidenko *et al*, demonstrated that RsbR remained phosphorylated when Thr205 was mutated to an alanine, which suggested that RsbR had a second phosphorylation site. The second phosphorylation site was identified using ^{32}P -labelling, followed by a proteolytic digestion of the protein. The fragments were then separated and two fragments containing a labelled phosphoryl group were sequenced leading to the identification of Thr171, alongside Thr205, as substrates for the RsbT kinase (Akbar et al., 1997; Gaidenko et al., 1999). The importance of these residues was also shown *in vivo*, where Thr171/205 were mutated into an aspartic acid residue, which acts as a phosphomimic, leading to a significant decrease in σ^B activation; this decrease was even more pronounced in the single T205D variant (Akbar et al., 1997). Finally, when RsbR and RsbS are phosphorylated, RsbT is released from the complex, which then interacts with its downstream partner, RsbU.

RsbU is a protein phosphatase and is a positive regulator of σ^B activity, which is itself activated by interaction with RsbT (Kang et al., 1998), it is thus only activated in the stressed cell. RsbU is composed of two domains: the N-terminal domain has no significant sequence homology with any other protein except RsbU orthologues, and has been described as a kinase recruitment domain based upon its ability to interact with RsbT (Yang et al., 1996). By contrast, the C-terminal domain has a catalytic domain of the PP2C-type phosphatase family (Yang et al., 1996). There are no known structures of the RsbU phosphatase domain, but it shares a catalytic core with the RsbX phosphatase (PDBID: 3W43) with which it shares 24% sequence identity. As the precise mechanism of the interaction between RsbT and RsbU remains unclear, potential binding sites for RsbT interaction have been deduced based on sequence conservation (Delumeau et al., 2004), which includes residues from both monomers in the N-RsbU dimer. RsbT has been shown to bind tightly to the native RsbU protein by site-directed mutagenesis

and native gel electrophoresis, but this interaction was lost with alanine substitutions to Glu24, Tyr28, Ile74 or Ile78 (Hardwick et al., 2007). Surface Plasmon Resonance (SPR) was also used to assess this interaction quantitatively. The wild-type RsbU - RsbT interaction has an affinity of 2 μ M (Delumeau et al., 2004; Hardwick et al., 2007), while this interaction was totally abolished with the variant proteins. Once activated by the binding of RsbT, RsbU interacts with RsbV (**Figure 1.4A**), the common partner of the stressosome-dependent and -independent σ^B activation pathways.

1.3.2. The stressosome-independent pathway

Not all species that encode stressosomes have a stressosome-independent σ^B activation pathway – one such example is *L. monocytogenes*. The stressosome-independent pathway is exemplified by the system found in *B. subtilis* and in other bacilli. The stressosome-independent pathway in *B. subtilis* is activated when a nutritional/energy stress is sensed, such as a drop in cellular ATP concentration, or a lack of carbon or oxygen (Brody et al., 2001). This signalling pathway includes two Rsb proteins, RsbP and RsbQ. RsbP/Q is an activator/phosphatase pair found in the same operon, which controls σ^B activation during the response to energy stresses (Brody et al., 2001). RsbP has an N-terminal Per-ARNT-Sim (PAS) domain that can bind to a small molecule ligand or chromophore (Brody et al., 2001) and a PP2C phosphatase domain on its CTD. It has been shown with *in vivo* assays that when RsbP is deleted, σ^B cannot be activated during energy stress (Brody et al., 2001). Moreover, the interaction between RsbP and RsbQ has been demonstrated by 2-yeast hybrid experiments where RsbP interacted exclusively with RsbQ and not with RsbT or RsbU; RsbQ deletion led to an abolition of the activation of σ^B activity (Vijay et al., 2000). The RsbQ enzyme is a positive regulator, and its function is required for activating σ^B during energy stress. The mechanism of action remains unclear, but a model of the RsbP/RsbQ pair function has been proposed (**Figure 1.4B**). RsbQ has been predicted to hydrolyse a small, but as yet unknown molecule, that acts as a direct activator for the stimulation of the RsbP phosphatase activity (Brody et al., 2001) towards phosphorylated RsbV, the anti-anti-sigma factor common to both pathways.

1.3.3. The common pathway

Once the RsbU phosphatase from the environmental pathway and the RsbP phosphatase from the energy stress pathway are activated, they can both stimulate the common anti-anti-sigma factor, RsbV (**Figure 1.4**). The structure of RsbV is unknown but representative structures from the RsbV-like anti-anti-sigma factor family are known of, for example, the

sporulation related anti-anti-sigmaF, SpoIIAA (Figure 1.5) (Seavers et al., 2001), which shares around 31% amino acid sequence identity with RsbV in *B. subtilis*. In normal cell conditions, approximately 75% of RsbV is found in its phosphorylated state (RsbV-P) (Dufour and Haldenwang, 1994). The residue found to be phosphorylated is Ser57 in *S. aureus* (Pané-Farré et al., 2009) and Ser56 in *B. subtilis* (Dufour and Haldenwang, 1994). The anti-anti-sigma factor RsbV is the σ^B competitor for RsbW and, in an unstressed cell, the RsbW kinase has a greater affinity for σ^B than RsbV-P (Delumeau et al., 2002). As σ^B remains sequestered by RsbW in a complex (Benson and Haldenwang, 1993), it cannot interact with RNA polymerase and the σ^B regulon remains switched off. During a stressful condition, RsbU (Voelker et al., 1995) or RsbP (Brody et al., 2001) dephosphorylates RsbV, which then will compete with σ^B for RsbW (Dufour and Haldenwang, 1994).

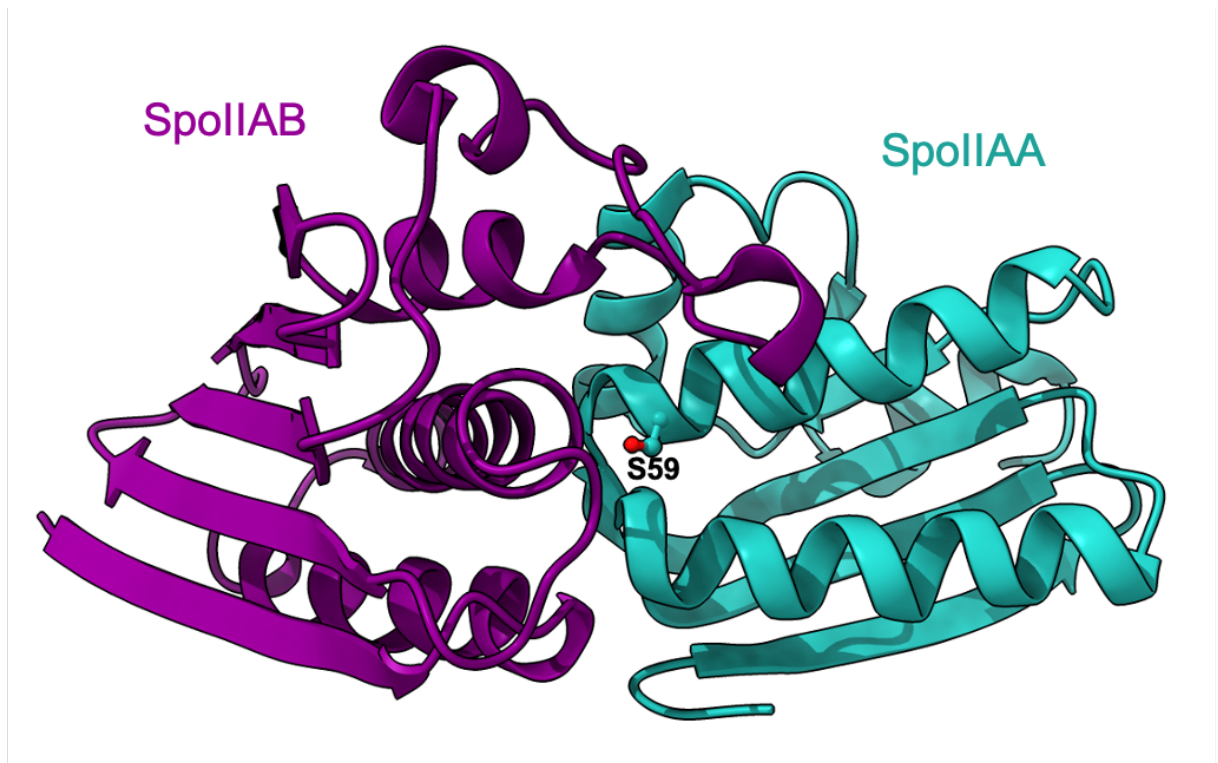


Figure 1.5. Structure of SpoIIAA:SpoIIAB complex from *B. subtilis*

The structures of SpoIIAA and SpoIIAB are displayed as cartoons. The residue phosphorylated by SpoIIAB is Ser59, shown in sticks on SpoIIAA, which can be found buried in the interaction interface. SpoIIAA forms a stable and compact structure with a central β -sheet (4 β -strands are parallel, and 1 β -strand is anti-parallel), while SpoIIAB is an alpha-beta protein and has an ATPase domain fold (PDB ID: 1TIL) (Masuda et al., 2004).

RsbW is the anti-sigma factor that binds to σ^B and prevents it from binding to RNAP to form the holoenzyme. RsbW is a serine-threonine kinase and is the RsbV antagonist (Benson and Haldenwang, 1993; Dufour and Haldenwang, 1994). It has been shown by size-exclusion chromatography that RsbW forms a complex with either RsbV or σ^B (Dufour and Haldenwang,

1994). RsbW is abundant in the cell, to ensure that all the σ^B remains sequestered in the absence of stressful conditions (Benson and Haldenwang, 1993). RsbV-P abundance is such that it also aids to keep RsbW bound to σ^B (Benson and Haldenwang, 1993). The choice of partner for RsbW is dependent on whether RsbV is phosphorylated or not: RsbW has a greater affinity to σ^B than RsbV-P (Dufour and Haldenwang, 1994). By contrast, RsbW has a greater affinity to RsbV than σ^B (Benson and Haldenwang, 1993; Dufour and Haldenwang, 1994). This partner switching interaction releases σ^B from RsbW, and RsbV will bind to RsbW, and RsbV will then be rephosphorylated (Dufour and Haldenwang, 1994). The interaction interplay between RsbV and RsbW is similar to their respective homologues SpoIIAA and SpoIIAB (**Figure 1.5**), which are involved in the sporulation pathway in *B. subtilis* involving σ^F (Schmidt et al., 1990). When released from RsbW, σ^B is free to bind to the RNAP to form the holoenzyme RNAP complex to guide it to the σ^B regulon coding for genes involved in the response to imposed stresses (Alper et al., 1996).

1.3.4. Resetting the sigB pathway

Although the resetting of the stressosome-independent pathway is not well-known, the resetting of the stressosome complex is better understood. The stressosome is reset by the RsbX phosphatase that belongs to subfamily II of the Serine/Threonine phosphatase PPM family. The structure of this family is composed of a PP2C catalytic core that binds two divalent cations (Mg^{2+}/Mn^{2+}) that are essential for the phosphatase activity (Teh et al., 2015). RsbX is a single domain protein while the homologous RsbU and RsbP proteins each possess an additional N-terminal domain that plays a regulatory role and binds to their respective partner proteins. RsbX is also a negative regulator of σ^B (Voelker et al., 1995). *In vivo* assays using *B. subtilis* cells showed that the deletion of *rsbX* ($\Delta rsbX$) led to the absence of RsbT, RsbS, RsbU and RsbV in several mutants. During the imposition of ethanol stress, the $\Delta rsbX$ cells were still able to activate the σ^B -regulated operons to a tenfold higher level compared to the wild-type (WT). The σ^B activity is transient in the WT *B. subtilis* strain, while it is persistent in the $\Delta rsbX$ strain, which demonstrates that RsbX is not required for σ^B activation, but is necessary for resetting the σ^B activity levels to a basal level after the recovery to the imposition of stress (Smirnova et al., 1998).

RsbX is known as the complex-resetting phosphatase, which dephosphorylates RsbS (Yang et al., 1996) and RsbR (Chen et al., 2004) to reset the stressosome complex and allow the stressosome to re-sequester RsbT (**Figure 1.4C**). Initially, RsbX was shown to only act on RsbS (Voelker et al., 1995). Phosphatase assays showed that the RsbU phosphatase was not active on RsbS-P, but on RsbV-P (as described previously) and, by contrast, RsbX was only

active on RsbS-P and not RsbV-P (Yang et al., 1996). It was subsequently demonstrated that RsbX was not exclusive to RsbS-P; these *in vitro* assays used isolated, purified recombinant proteins, and may explain why RsbX wasn't a phosphatase candidate for RsbR-p alone (Chen et al., 2004). In fact, RsbX is also active towards RsbR-P, but only when RsbR-P is in complex with RsbS (Chen et al., 2004). Urea-alkaline gel electrophoresis showed that the phosphorylation of RsbR Thr205 was removed by RsbX when in complex with RsbS, but no dephosphorylation was observed with RsbR-P alone (Chen et al., 2004). RsbX homologues, such as RsbU and SpoIIE, had no effect on RsbR Thr205 phosphorylation, showing that RsbX is exclusive to RsbR-P and RsbS-P (Yang et al., 1996). Compared to the Thr205 phosphorylation site on RsbR, the Thr171 phosphorylation, in complex with RsbS, was neither dephosphorylated by RsbX nor RsbU, indicating that RsbX activity is specific to RsbR-Thr205-P (Chen et al., 2004). Subsequently it was demonstrated that when RsbR and RsbS in a complex are both phosphorylated, only RsbS-P is dephosphorylated by RsbX, which suggests that the predominant phosphorylation site in RsbR is T171 and not T205 (Chen et al., 2004).

1.4. Bioinformatics and structural analysis of the stressosome module

As this project focuses on the stressosome complex and its constituent proteins, a detailed description of the structural characteristics and bioinformatics analyses is necessary to explain the state of knowledge of this system prior to the start of the project.

1.4.1. The stressosome scaffolding protein RsbS

RsbS is a single domain protein that belongs to the Sulphate Transporter and Anti-Sigma factor antagonist superfamily that was originally identified by bioinformatics analyses (Aravind and Koonin, 2000); STAS domains are usually found in the CTD region of sulphate transporters and bacterial anti-sigma factor antagonists. There is some limited evidence that this domain may have a general NTP binding function (Aravind and Koonin, 2000), Najafi et al., 1996). Whilst RsbS is a single domain protein, a STAS domain is also found at the CTD of RsbR. The STAS domain is highly conserved between RsbS and RsbR paralogues and across species' orthologues. As shown in the alignment in **Figure 1.6**, seven amino acids are strictly conserved between RsbS and RsbR proteins from *B. subtilis* and *L. monocytogenes* and these conserved amino acids are not found on the dimerization interface.

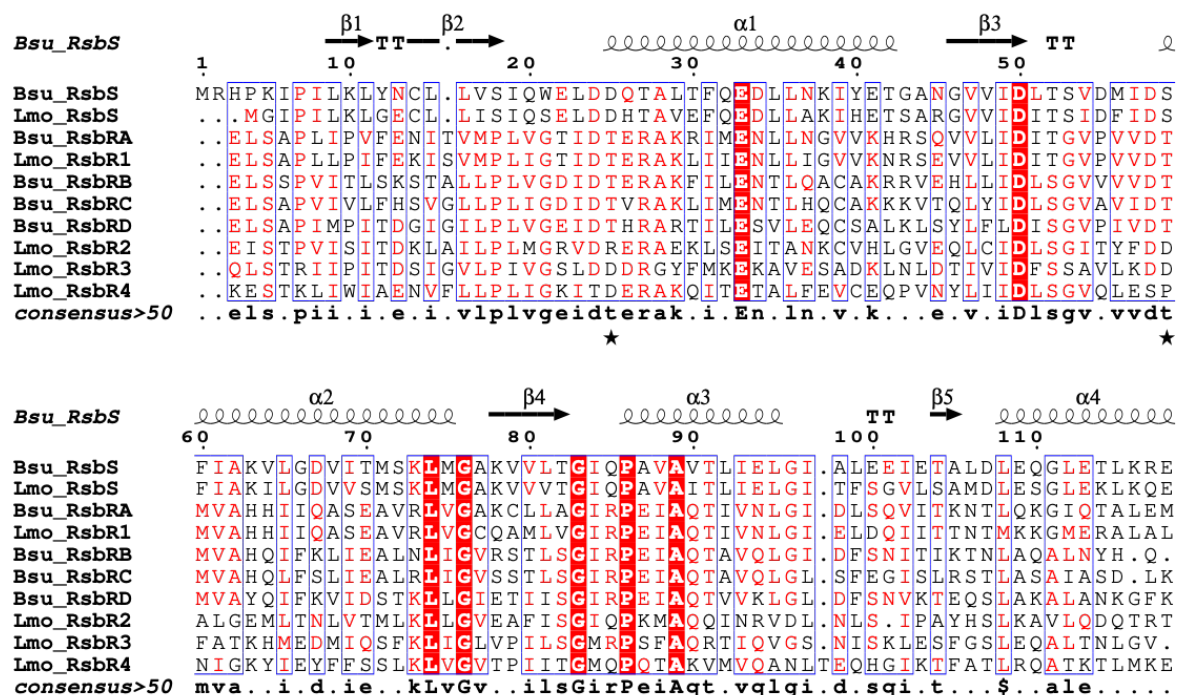


Figure 1.6 RsbR/RsbS STAS domain alignment

The STAS domain sequence alignment shows that there are seven strictly conserved residues highlighted in red. Most of the remaining residues are highly conserved in red with a consensus value greater than 50 (bottom line). The sequence alignment of the RsbS, RsbR and RsbR4 paralogues proteins from *B. subtilis* (Bsu) and *L. monocytogenes* (Lmo) was generated using the Multalin webserver (Corpet, 1988) (<http://multalin.toulouse.inra.fr/multalin/>). The residue numbering follows the sequence numbering of *B. subtilis* RsbS. The secondary structure features of *B. subtilis* RsbS (PDB ID 6JHK) are depicted above the alignment. Phosphorylated residues on RsbR T171 (position D25 on RsbS) and T205 which aligns with S59 of RsbS, are pinned with a black star below the consensus line.

The phosphorylated residues T171 and T205, the latter of which corresponds to S59 in RsbS, are well conserved; other than threonine, the residue at the equivalent to 205 is either serine, equally phosphorylatable, or the phosphomimetic aspartic acid. In one instance, Lmo_RsbR4, the non-phosphorylatable proline is found suggesting strongly the role of phosphorylation in this protein differs from other paralogues and orthologues. Alternatively, and given the presence of Glu and Ser at positions -2 and -1 from the predicted phosphosite, there is potentially a local sequence shift by one residue. Similarly, if threonine is not found at the equivalent position to 171, aspartic acid is usually found except for the sole example of Lmo_RsbR2 where it is arginine, and although this amino acid can be phosphorylated (Creixell et al., 2012) the kinase responsible is unrelated to RsbT (Akbar et al., 1997; Gaidenko et al., 1999). The wide conservation of glutamic and aspartic acids -1 to the phosphorylation sites is also evident; these residues are presumably involved in the phosphoryl transfer mechanism from ATP or invokes a pseudo-double phosphorylation mechanism for RsbT activation.

The structures of RsbS from *Moorella thermoacetica*, MtS (**Figure 1.7A**), (Quin et al., 2012) and more recently *B. subtilis* RsbS (Kwon et al., 2019) have been determined by X-ray crystallography (**Figure 1.7B**). Both structures form a dimer and share 39.8% sequence identity and have a RMSD of 1.9Å, when superimposing the mainchains, which shows that the structure is well-conserved between species. Structurally, the STAS domain is composed of a central β -sheet with flanking α -helices. The role of the STAS domain of RsbS and RsbR within the stressosome is to act as a scaffold (Chen et al., 2004) that forms the core of the complex (Kim et al., 2004). Given the STAS domain is a target of RsbT phosphorylation (Gaidenko et al., 1999; Kang et al., 1998), it remains unclear if the STAS domain has any function in the activation mechanism of the stressosome, except that RsbS is key for RsbT sequestration (Akbar et al., 2001; Woodbury et al., 2004).

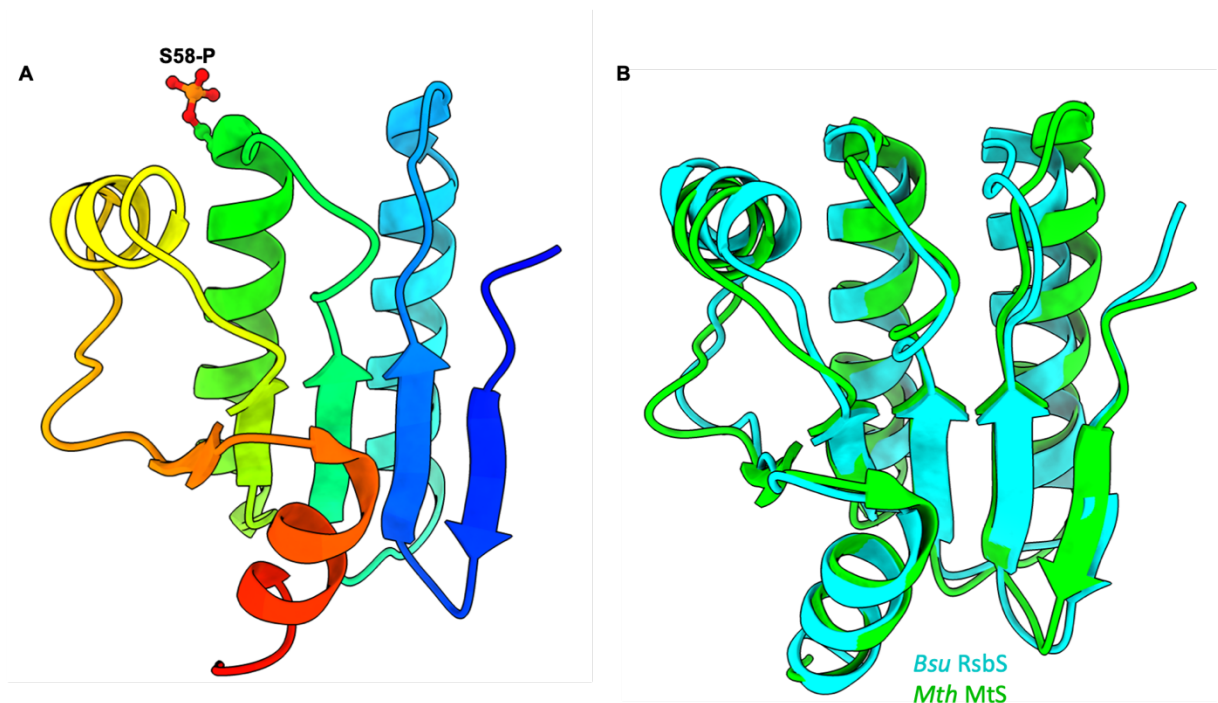


Figure 1.7. Crystal structures of the RsbS STAS domain from different species

The structures are represented in cartoon form; the RsbS STAS domain comprises five β -strands (four parallel and one anti-parallel strand) forming a central β -sheet covered with four α -helices. **A.** *M. thermoacetica* RsbS orthologue, MtS, coloured in rainbow from the N-terminus in blue to the C-terminus in red (PDBID 3ZTB). This structure shows the phosphorylated variant of S58-P (shown in sticks), the residue phosphorylated by MtRsbT. **B.** Superimposition of the MtS structure in green and *B. subtilis* RsbS in cyan (6JHK), showing the overall structure conservation.

1.4.2. RsbR, the presumed stress sensor

There are five RsbR paralogues in *B. subtilis*: RsbRA (gene name *rsbR*), RsbRB (*ykoB*), RsbRC (*vojH*), RsbD (*yqhA*) and YtvA (*ytvA*), which have been named in *L. monocytogenes* as RsbR1, RsbR2, RsbR3, RsbR4 (corresponding *L. monocytogenes* genes: *lmo0889*, *lmo0161*,

lmo1642, *lmo1842*; and YtvA/RsbL: *lmo0799*). The number and nature of the RsbR paralogues varies between species, with up to 14 paralogues in *Bacillus megaterium* of the Bacilli class; and, strikingly, up to 82 in *Sorangium cellulosum* found in the Deltaproteobacteria class (personal communication, Jan Pané-Farré).

The RsbR protein is the only component of the *rsb* operon to have an unknown function. RsbR proteins have two approximately equal-sized domains separated by a short region of poorly conserved sequence. The CTD is comprised of the STAS domain as described in the previous section, while the NTD function remains unknown. The NTD sequence of RsbR proteins are poorly conserved, as shown in the alignment **Figure 1.8**. It is pertinent to point out that there is not a single amino acid that is strictly conserved between the NTDs of RsbR paralogues in the two species that are the focus of the work in this PhD thesis. N-RsbR is the only part of the stressosome complex with an unknown function, this domain has been presumed to act as the “stress sensor” (Akbar et al., 1997).

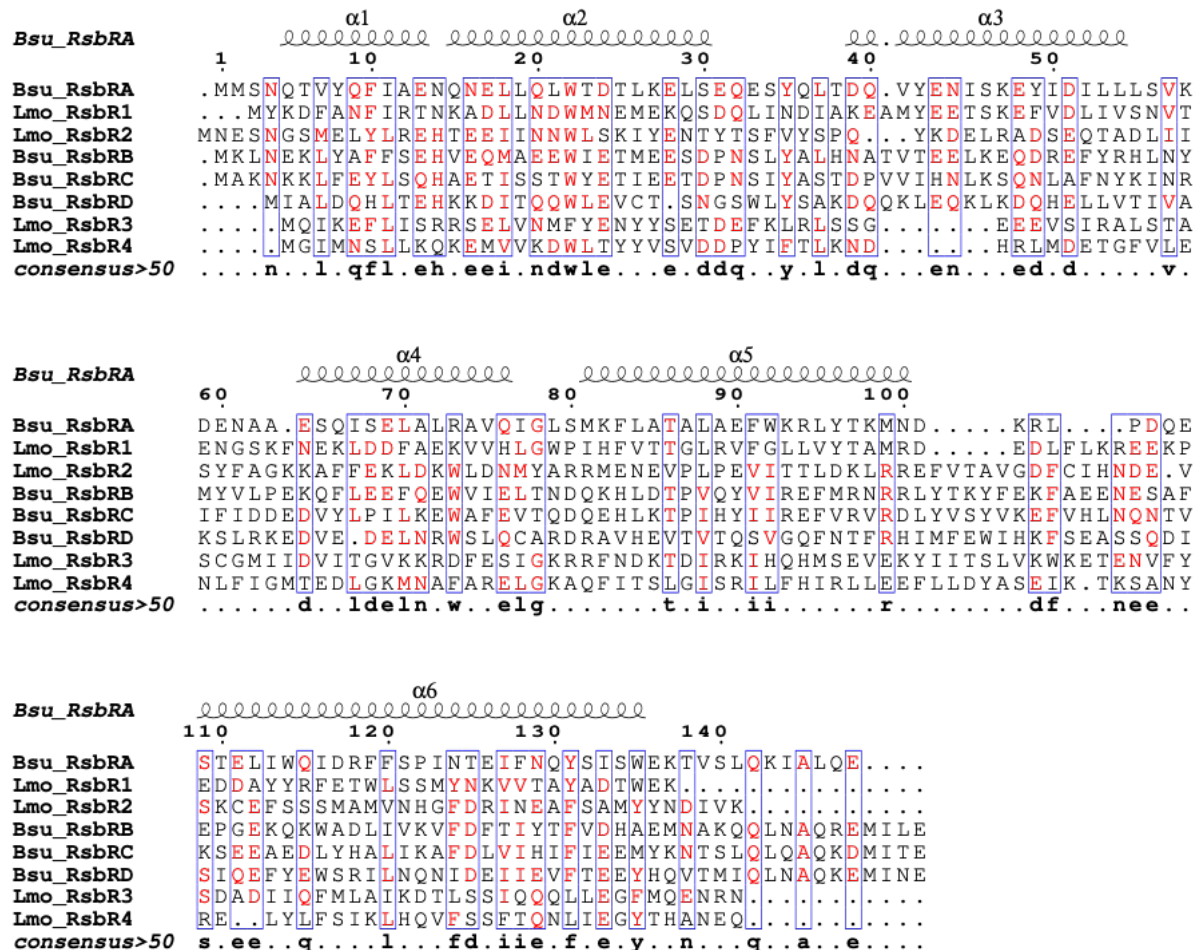


Figure 1.8. N-RsbR paralogues and homologues sequence alignment

The sequence alignment shows the poorly conserved sequence between paralogues and orthologues. A sequence alignment of the NTDs of RsbR paralogues from *B. subtilis* (*Bsu*) and *L. monocytogenes* (*Lmo*) was generated using the Multalin webserver (Corpet, 1988) (<http://multalin.toulouse.inra.fr/multalin/>). The residue numbering follows the sequence numbering of *B. subtilis* RsbRA. The secondary structure features of *B. subtilis* RsbRA (PDB ID 2BNL) is depicted above the alignment.

Structures of the NTDs of RsbRA from *B. subtilis* (Murray et al., 2005) and MtR from *M. thermoacetica* (Quin et al., 2012) have been determined by X-ray crystallography. Both are composed of five α -helices and associate into dimers (**Figure 1.9**); the N-terminal α -helix of the NTD of RsbRA is bent by about 90°. Both adopt a non-heme globin-like fold, as neither are able to bind heme (Murray et al., 2005). This is mainly due to the lack of a proximal histidine ligand in the binding pocket which is necessary to co-ordinate the Fe ion at the centre of the co-factor, and for the infilling of the space for heme by hydrophobic amino acids. There is currently no known function that can be assessed in activity assays and the structure of the domain provides no further clues to potential functions. A likely contact site for RsbT when bound to the stressosome has been mapped by site-directed mutagenesis and the subsequent failure of RsbR variant stressosomes to retain RsbT during gel filtration (Murray et al., 2005). However, there was no quantification of these binding studies and subsequent studies by the Price laboratory failed to identify a phenotype of *B. subtilis* strains carrying the same RsbRA mutations in a genetic background lacking the other paralogues (Gaidenko et al., 2011).

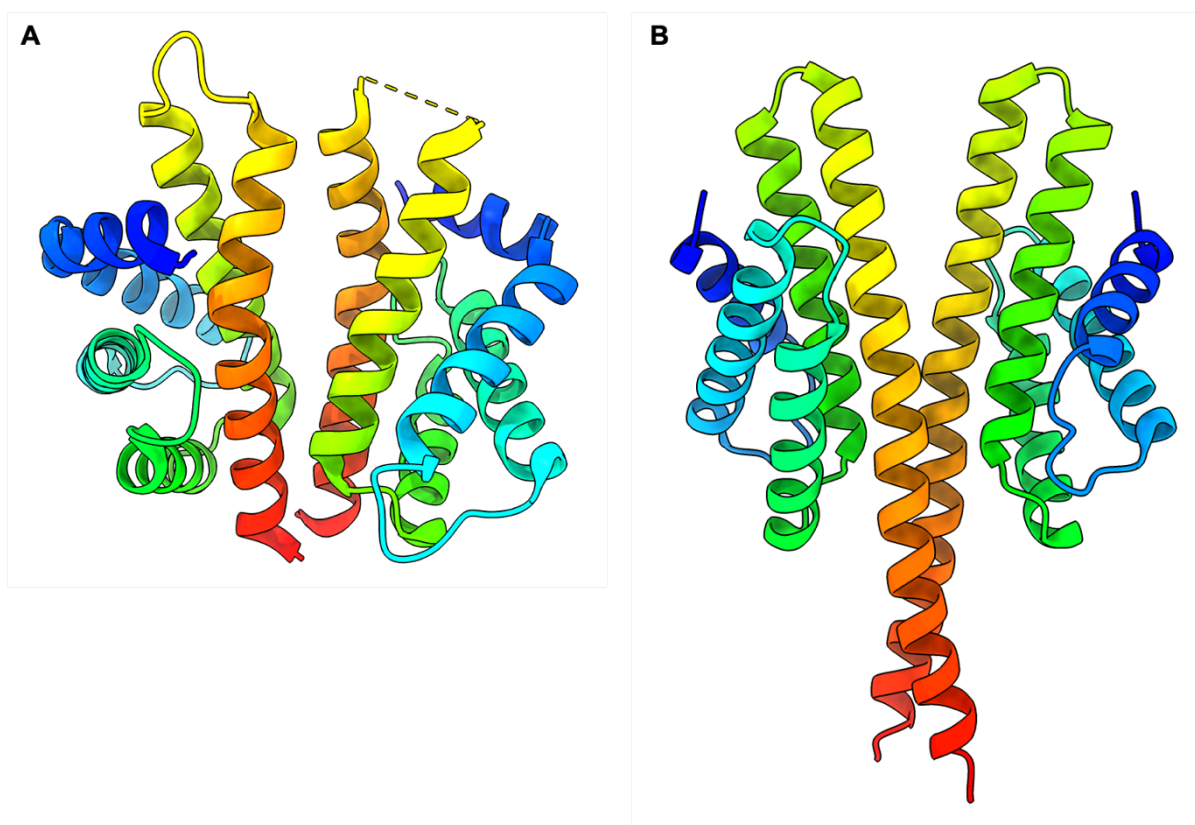


Figure 1.9. Crystal structure of the non-heme globin-like N-RsbR domain

A. *B. subtilis* N-RsbRA (PDB ID 2BNL). **B.** *M. thermoacetica* N-MtR (PDB ID: 3ZTA); both form a dimeric structure, and each adopt a non-heme globin like fold. The structures are represented in cartoon fashion showing the secondary structure features. Each monomer is coloured from the NTD in blue to the CTD in red and are formed by five α -helices.

The sensory function of N-RsbR and its paralogues, and whether they only form homodimers or can form heterodimers *in vivo*, also remains unknown. Heterodimerisation of the NTD between paralogues has not been studied but given the lack of sequence homology heterodimerisation seems unlikely. By contrast, heterodimerisation of *Cso* scaffolding paralogues has been reported in another supramolecular protein complex, the carboxysome (Ryan et al., 2019), and here the sequence identity of the interacting paralogs is just 17%, which is similar to the sequence conservation of the NTD RsbR paralogues. Consequently, heterodimerisation of the paralogues *in vivo* should not be excluded.

The YtvA paralogue, common to *L. monocytogenes* and *B. subtilis*, is reasonably well-characterised in comparison to the other paralogues. It is a blue-light receptor and has a flavin mononucleotide (FMN)-containing light-oxygen voltage (LOV) domain instead of the non-heme globin at its N-terminus. This paralogue acts as a positive σ^B regulator compared to the other paralogues that are all negative σ^B regulators (Akbar et al., 2001). YtvA is involved in the stressosome-dependent pathway, the environmental stress sensing pathway, as it has been shown that σ^B activation was abolished when RsbT and RsbU were deleted in a *B. subtilis* strain lacking all RsbR paralogues (Gaidenko et al., 2006). Gaidenko *et al* also showed the importance of Cys62 in σ^B activation as this residue, conserved in YtvA and related plant phototropins, forms an adduct with the chromophore in response to blue-light. Two variants, C62A and C62S, were obtained by mutagenesis, and σ^B activity was measured by β -galactosidase accumulation assays. The two mutants were unable to activate σ^B , which defines Cys62 as essential for σ^B activation. In site-directed mutational analyses and β -galactosidase assays (Marcela et al., 2009) using dark and light-states of YtvA, the C62A variant abolished the light activation of σ^B . Other mutated residues, including E105L, increased σ^B activity in the dark (Gaidenko et al., 2006).

An X-ray crystal structure of the RsbR protein full-length protein has yet to be determined, which may be a consequence of the fact that the two domains are related to each other by a flexible region. Therefore, the mechanism of signal transduction from the NTD to the CTD is difficult to understand on a molecular level. Nevertheless, YtvA is the only RsbR paralogue to have a full-length protein structure (Jurk et al., 2014) (**Figure 1.10**) determined by Nuclear Magnetic Resonance (NMR) in solution. The structure was solved with an FMN in each NTD monomer where it binds covalently to the Cys62, the critical residue for the positive regulation of σ^B . Glu105 is found on the NTD of YtvA and is close to the J-helix, which is the long helix that joins together the two domains, suggesting the involvement of this residue in signal transduction from the NTD to the CTD, but the exact mechanism remains unclear.

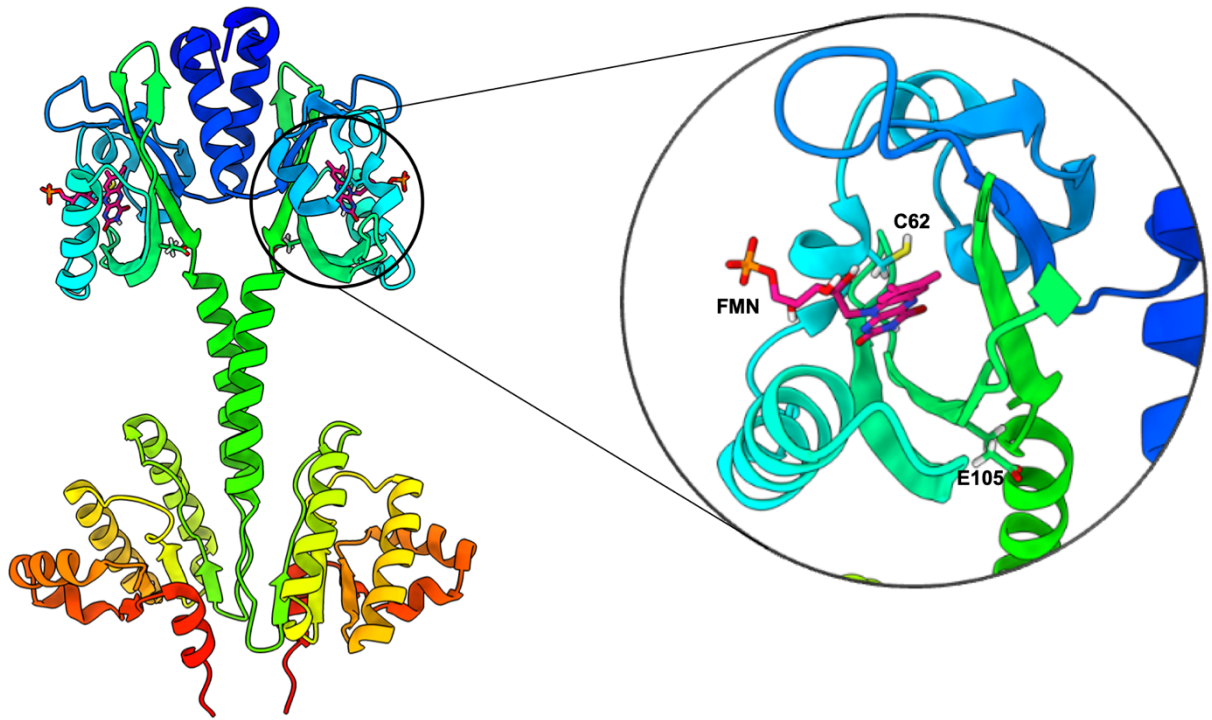


Figure 1.10. NMR structure of the *B. subtilis* YtvA protein

The YtvA structure (PDB ID 2MWG) is shown in cartoon with each monomer colored in rainbow from the NTD in blue to the CTD in red. The full-length structure shows clearly the two domains: the STAS domain on the CTD anchored to the core of the stressosome complex and the LOV domain on the NTD which protrudes from the complex. A zoom on the NTD shows the Cys62, which is in close proximity to the FMN molecule, essential for σ^B activation, and Glu105, which is essential for the dark-state activity of YtvA, is close to the J-helix.

1.5. The stressosome complex

The stressosome has been studied extensively *in vitro* from recombinant proteins originating from the upstream module of the *rsb* operon, including RsbR (RsbRA in *B. subtilis* and RsbR1 in *L. monocytogenes*), RsbS and RsbT in *B. subtilis* and *L. monocytogenes*. As the stressosome studies were carried on recombinant complexes, their stoichiometries within the native complex remains unknown.

1.5.1. The stressosome complex in *B. subtilis*

The *B. subtilis* stressosome structure was first determined by cryo-EM in 2008 by Marles-Wright *et al* (Marles-Wright *et al.*, 2008). Three forms of the complex were studied: RsbRA-CTD:RsbS, which forms the core; the binary complex of RsbS:RsbRA; and the ternary RsbS:RsbRA:RsbT complex. The cryo-EM reconstruction of the stressosome core complex was determined to 6.5 Å resolution. The reconstruction of the core complex showed an internal icosahedral symmetry of the STAS domains, with an overall core radius of 90 Å.

The RsbS:RsbRA complex was reconstructed at a final resolution of 8 Å (Figure 1.11A). The NTD of RsbRA appears as turrets protruding from the core complex and increases the complex size to a radius of 150 Å. The RsbS:RsbRA complex was consistent with D2 point-group symmetry; with the symmetry classification, the turrets projecting from the core allowed the stoichiometry of the complex to be determined and there were 20 copies of RsbS and 40 copies of RsbRA.

Finally, the structure of the RsbS:RsbRA:RsbT ternary complex (Figure 1.11B) was determined at 8.3 Å resolution. The addition of RsbT did not change the complex diameter as RsbT bound to the core complex between RsbRA turrets. The presence of RsbT induced only minor conformational changes, most likely due to the high flexibility of N-RsbRA in comparison to the stressosome core. The stoichiometry of the ternary complex is 2:1:1 for RsbRA:RsbS:RsbT, which is in accordance with previous densitometry analysis of SDS-PAGE of purified stressosome complexes (Chen et al., 2003).

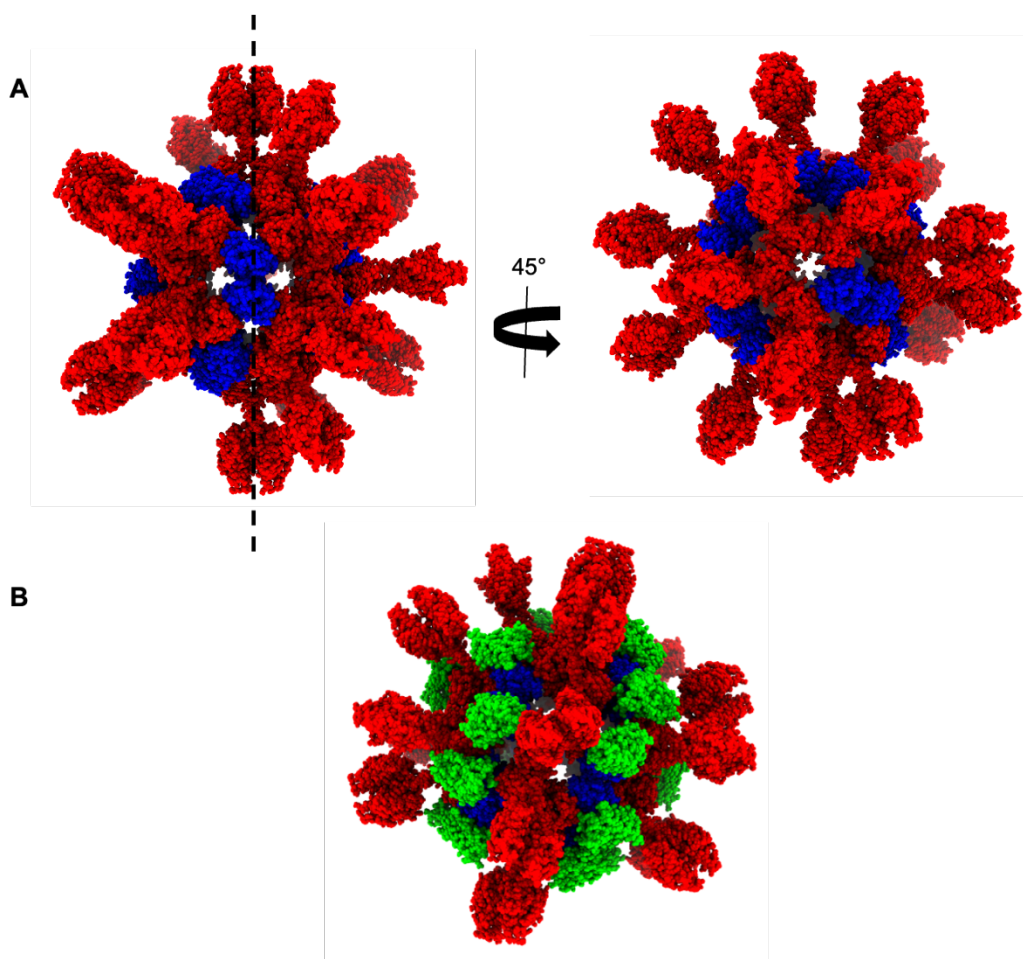


Figure 1.11. Models of the *B. subtilis* stressosome complex based on Cryo-EM reconstructions.

A. Molecular model of the *Bsu* RsbRA-RsbS complex composed of RsbS (blue) and RsbRA (red), the two views are displayed at 45° apart. The dotted line shows the D2 symmetry of the particle **B.** RsbRA-RsbS and RsbT (green) tertiary complex. The RsbT binding does not increase the size of the complex as it binds between the turrets (Marles-Wright *et al.*, 2008).

More recently, a new single-particle cryo-EM reconstruction of the same complex from *B. subtilis* was published, alongside a new RsbS crystal structure (Kwon et al., 2019). The authors applied D2, C1 and icosahedral (for the core) symmetries to the reconstruction to probe the underlying symmetry of the complex. With the D2 symmetry imposed they obtained the same complex published previously (Marles-Wright et al., 2008). With C1 symmetry, which has no symmetry constraints, the complex showed a different assembly (**Figure 1.12**). Twenty-two turrets were present in C1 symmetry, and the subsequent model therefore has 44 RsbRA and 16 RsbS protein subunits. This asymmetric complex did not provide additional detail to its biological function as the main structural features of the complex remained the same. Therefore, the exact stoichiometry cannot be determined by cryo-EM without introducing symmetry bias, as more turrets were obtained in the C1 symmetrised reconstruction compared to that in D2. Alternatively, the stressosome might be heterogeneously distributed in the sample.

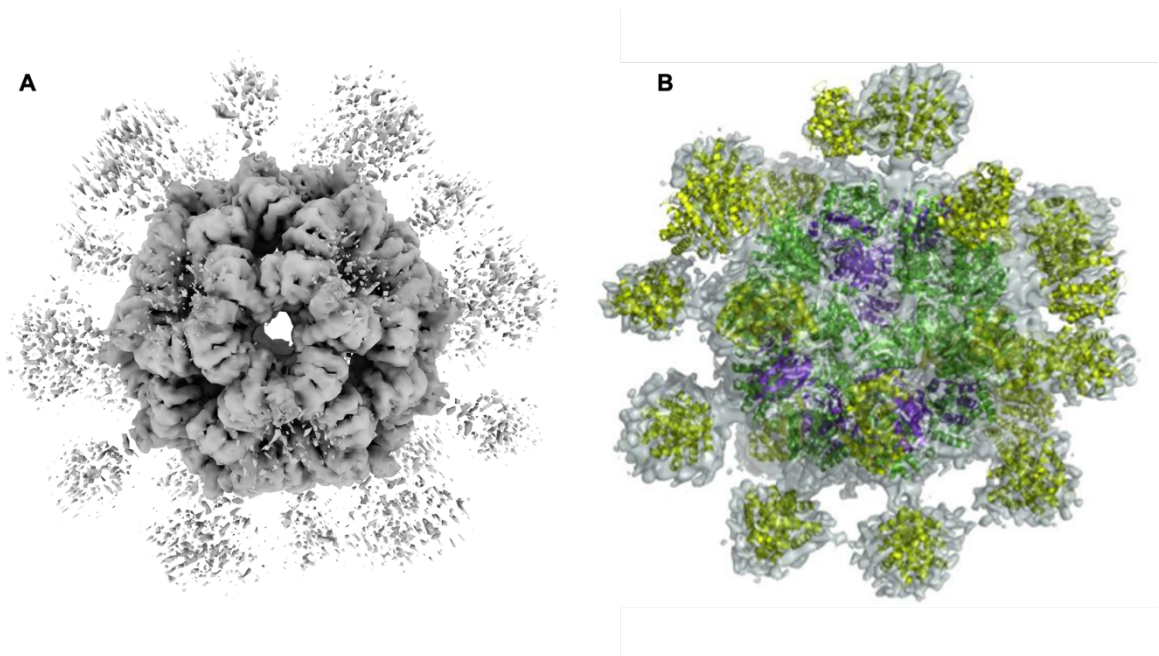


Figure 1.12. Cryo-EM model of *B. subtilis* stressosome complex in C1 symmetry
A. Map obtained in C1 symmetry at 4.1 Å resolution (EMDB ID: EMD-9924) in which the turrets have poor densities. **B.** The model of the RsbRA-RsbS complex in the transparently-rendered map with RsbS in purple, RsbR STAS domain in green and the RsbR turrets in yellow (Kwon et al., 2019).

1.5.2. The stressosome complex in *L. monocytogenes*

The structure of the RsbR1:RsbS:RsbT stressosome complex of *L. monocytogenes* has also been determined recently by cryo-EM (Williams et al., 2019). The *L. monocytogenes* stressosome structure displays the same overall architecture as that from *B. subtilis*

(**Figure 1.13**). However, density for all copies of RsbT was too weak to model it bound to the complex, which might be explained by the dissociation of RsbT from the complex during sample preparation. Missing densities were also observed for the majority of the RsbR1 NTD, which might also be explained by the high flexibility of this domain, or by mis-masked maps, which would exclude the turrets.

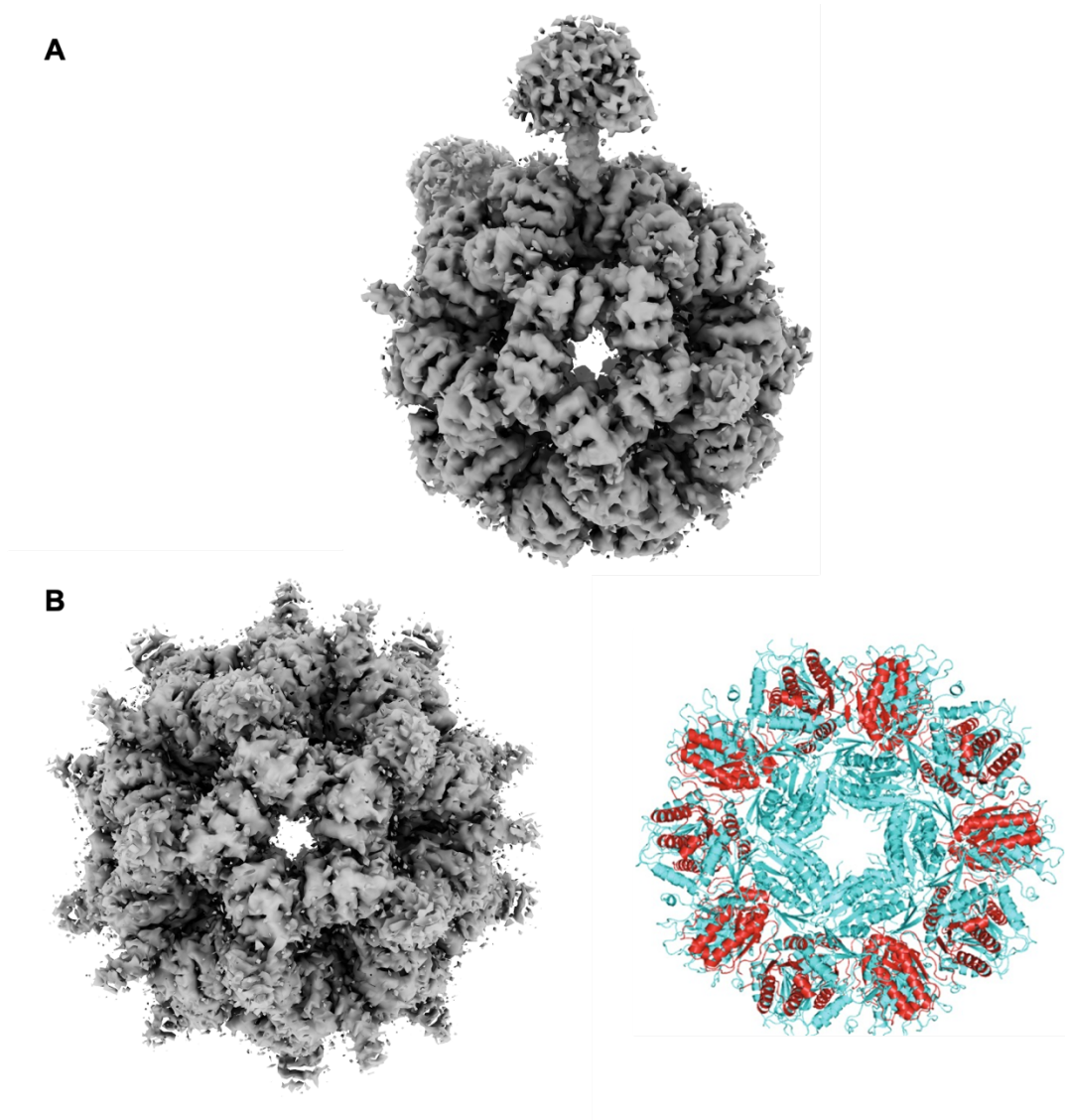


Figure 1.13. Cryo-EM model of *L. monocytogenes* stressosome complex in C1 symmetry.

A. The C1 symmetry map obtained of the *Lmo* RsbR1-RsbS-RsbT complex contoured at $\sigma = 0.0181$ at 4.21 Å (EMDB ID: EMD-4508). **B.** The map of the core ($\sigma = 0.0164$) of the stressosome at a resolution of 3.37 Å (EMDB ID: EMD-4510), and the molecular model with the RsbR STAS domain in cyan and the RsbS STAS domain in red (PDB ID: 6QCM) (Williams et al., 2019).

The stoichiometry of the *L. monocytogenes* stressosome is structurally similar to that of *B. subtilis*. However, SDS-PAGE gel densitometry suggests a stoichiometry for *Lmo*RsbR1:RsbS:RsbT as 1:1:0.5, which differs to that published previously by the Lewis

group at 2:1:1 (Marles-Wright et al., 2008; Quin et al., 2012). The absence of RsbT might be due to its dissociation while manipulating stressosome samples. Indeed, Williams *et al* also concluded that the RsbR1:RsbS complex was unstable without RsbT (Williams et al., 2019). RsbR1:RsbS and RsbR1:RsbS:RsbT complexes were recombinantly purified and visualised by negative stain; in the latter case, a ring-shaped stressosome complex was observed whereas the former complexes were not fully complete and disassembled *in vitro* EM (Williams et al., 2019). As for the reported RsbR1:RsbS ratio, it is not in accordance with the structure nor the 2:1 stoichiometry reported by others.

1.6. Potential stressosome interaction partner and membrane anchor: the Prli42 miniprotein

Recently, a membrane-anchored miniprotein with a short cytosolic tail was identified as a potential stressosome interaction partner in *L. monocytogenes* (Impens et al., 2017). This miniprotein, called Prli42, is conserved among the Firmicute phyla and was identified by a newly developed method called N-terminomics COFRADIC (Combined Fractional Diagonal Chromatography), which isolates N-terminal peptides by 2 chromatographic separations. In order to do so, the authors first mapped the translational map of *L. monocytogenes* EGD-e strain genome in all six reading frames. *L. monocytogenes* was then grown at 20°C and 37°C in exponential phase and at 37°C for stationary phase. A PDF (Peptide DeFormylase) inhibitor was added to increase the Translation Initiation Sites (TIS) identification. Prokaryotic N-termini are often formylated, where formylation of the N-termini in prokaryote is a translational modification, which is the addition of a formyl group (CHO) to the N-termini to produce an active protein. After blocking the free NH₂ groups in the samples, the proteins were extracted and digested with either trypsin or endoproteinase GluC, which allowed the generation of more N-terminal peptides. Peptides were separated using COFRADIC and sequenced by liquid chromatography-tandem mass spectrometry (LC-MS/MS). This led to the identification of 12,000 unique peptides, which were then mapped onto the genome and most of these N-terminal peptides corresponded to previously annotated TISs.

As the initial genome annotation for *L. monocytogenes* excluded ORFs with fewer than 40 amino acids, it explicitly excluded the miniproteins. Six miniproteins were identified within this study that were annotated previously as small RNAs such as *rli24*, *rli41* and *rli42*. The product of the *rli42* RNA, Prli42, was identified as a 35 amino acid long membrane-anchored miniprotein.

Interaction partners for these miniproteins were identified by co-immunoprecipitation and LC-MS/MS analysis (Impens et al., 2017). Five Prli42-binding partners were identified: RsbR1, RsbS, RsbR2, RsbR3 and YtvA. The reported interaction of Prli42 and N-RsbRA was used to generate a model of the predicted interaction by molecular docking procedures (**Figure 1.14**). The resultant model describes a potential interaction between the positively charged amino acids Lys4, Lys5 and Arg8, all found on the cytosolic N-terminal end of the Prli42 peptide, with an acidic patch in the N-RsbR1 dimer interface (**Figure 1.14A**). To test the interaction between RsbR1 and Prli42, *Listeria* strains were constructed expressing Prli42-FLAG variants designed to abrogate these interactions: K5L, K5F and R8A. Immunoblots revealed an interaction between the WT Prli42 *L. monocytogenes* strain with RsbR1 while the interaction was reduced with the K5L and K5F variants and totally abolished with R8A variant (Impens et al., 2017).

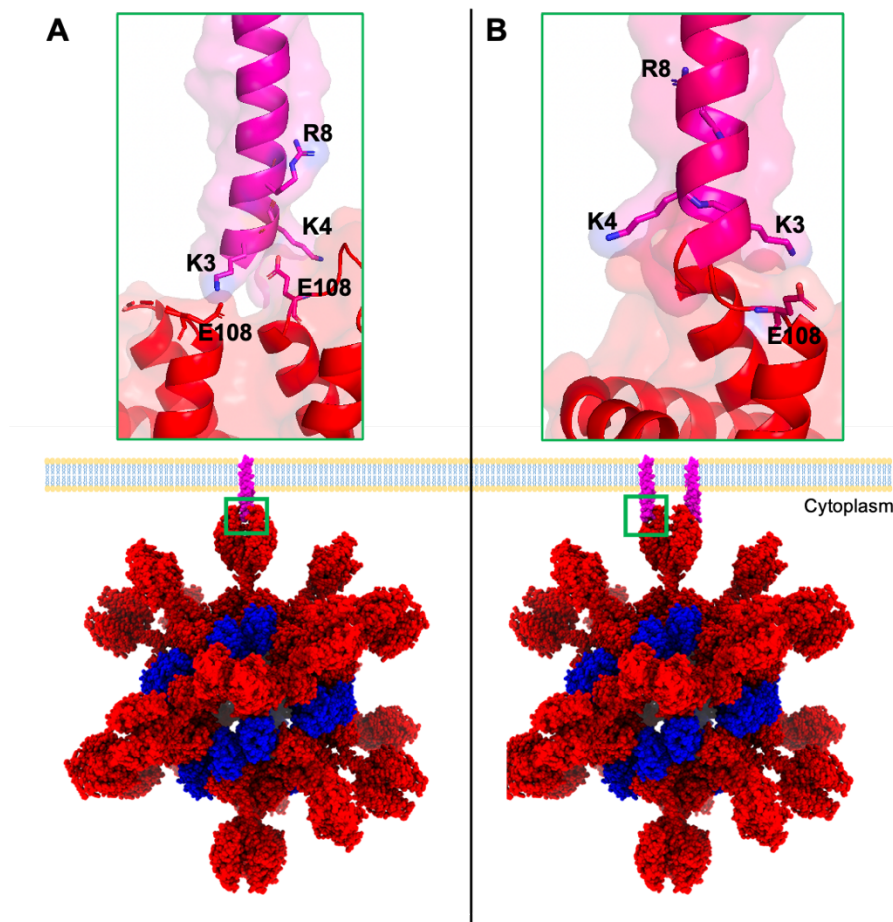


Figure 1.14. Models of Prli42 miniprotein and RsbR NTD interactions

A. First interaction model: The initial interaction model included one Prli42 miniprotein shown in pink interacting with a dimer of N-RsbRA protein (PDB ID 2BNL) shown in red (Impens et al., 2018). The relevant residues on Prli42 involves K3, K4 and R8 probably interacting with the surface E108 on N-RsbRA. **B.** Second interaction model: the revised model of interaction is described as one Prli42 molecule interacting with one N-RsbRA monomer involving the same residues as the first model (Williams *et al.* 2019). The interaction implies that the stressosome is maintained close to the membrane while interacting with RsbR in complex.

In vivo stress assays using oxidative stress (hydrogen peroxide) with *L. monocytogenes* Prli42 mutant strains ($\Delta prli42$) showed a reduced cell-growth; however, this was not significantly different to WT while cell growth was not affected by several other common stresses (Impens et al., 2017). The $\Delta prli42$ strain displayed several phenotypes compared to the wild-type, such as lower survival in macrophages and lower expression of virulence factors, which in turn could also be explained by the lower cell survival (Impens et al., 2017).

A subsequent study from the same lab suggested a new model for the interaction: one Prli42 molecule interacted with one RsbR1 monomer (Williams *et al.*, 2019), as shown in **Figure 1.14B**, instead of one Prli42 interaction per dimer (Impens *et al* 2018). In this later study, Williams *et al* also confirmed the interaction with pull-down assays. RsbR1 variants were made with changes to the new predicted site of interaction, E109Q, E110Q and E113Q in *L. monocytogenes*, which correspond to D106, E108 and E111 in *B. subtilis*, respectively. Prli42 WT was then used to pull-down WT RsbR1 and the RsbR1 variants, with a negative control of the $\Delta prli42$ strain. A significant decrease of the interaction between Prli42 and RsbR1 triple mutant was observed, but not a total abolition of this interaction. Cell survival assays were performed to show the functional effect of the interaction between RsbR1 and Prli42, still using oxidative stress (Williams et al., 2019). $\Delta rsbR1$ and $\Delta rsbS$ strains were more sensitive to this stress than the WT and this sensitivity was restored after transformation with plasmids expressing WT RsbR1 and RsbS. However, restoration was not observed in the triple RsbR1 mutant and Williams *et al* concluded that the RsbR1-Prli42 interaction might be responsible for the stress signalling of RsbR1 (Williams et al., 2019).

The authors identified Prli42 as critical for the stressosome activation; however, the evidence presented was based on cell-growth assays and showed only minor growth defects using oxidative stress only. Moreover, negative controls were absent for the pull-down assays and, most importantly, the interaction of Prli42 and RsbR1 was not quantified using any biophysical method such as Surface Plasmon Resonance or Fluorescence Polarisation. These results do not conclusively prove a role for Prli42 in the stressosome activation and the reported Prli42 interaction was poorly controlled.

1.7. Project objectives

This project aimed to determine the role and the activity of the NTD of RsbR proteins in order to solve unanswered questions about stress sensing such as: What is the trigger for the stressosome activation and how does the stressosome transmit the stress sensing signal from

the NTD of RsbR to the core of the complex? To be able to answer these questions, three main objectives were identified. The first objective was to determine the X-ray crystal structures of the amino terminal domains of *L. monocytogenes* RsbR and its paralogues. This domain is not well understood, but recently it was reported to function as a sensor by interacting with a membrane miniprotein, Prli42 (Impens *et al.*, 2017). This hypothesis was tested by assessing the binding of the RsbR NTD to the miniprotein by quantitative procedures such as fluorescence polarisation (FP). The effect on the stress response with Prli42 was determined in live bacteria, by our PATHSENSE network partners in Ede (NL), Madrid (ES) and Marburg (DE).

With the crystal structures, the critical role of amino acids in the interaction was identified in order to allow our PATHSENSE network partners in Dundee to test potentially-important residues by mutagenesis on *B. subtilis* *in vivo*. Moreover, an initial ligand screen on the NTD of RsbR alone was done in order to identify possible ligand binding to the sensing activity of RsbR.

The stressosome complex structure from *L. monocytogenes* was also studied in two forms: the minimum complex comprising RsbR1 and RsbS, and the reset complex with RsbR1, RsbS and the RsbX phosphatase, to understand how RsbX binds to and acts on the complex.

Finally, the native stressosome complex from *B. subtilis* was the subject of keen activity. The native composition of the complex was assessed by proteomics with the ultimate goal of obtaining a single particle reconstruction of the native complex.

This study will also allow introducing novel sensing capabilities to GRAS bacteria (such as *Bacillus subtilis*) with synthetic biology potential for real-time detection of low levels of toxic gases or other environmental pollutants.

Chapter II: Materials and Methods

2.1. Molecular biology

Buffers and solutions:

- 1x TAE buffer: 40 mM Tris pH 7.6, 20 mM acetic acid and 1 mM EDTA
- 1% agarose gel: 1 g of agarose in 100 mL of 1x TAE buffer
- 6x loading dye (ThermoFisher Scientific™): 10 mM Tris-HCl pH7.6, 0.03% bromophenol blue, 0.03% xylene cyanol FF, 60% glycerol and 60mM EDTA.
- Ultrapure milliQ water
- 1X ligase buffer: 50 mM Tris-HCl pH7.5, 10 mM MgCl₂, 10mM dithiothreitol and 1mM ATP.

Growth media: all media were autoclaved before use.

- Lysogeny Broth Miller (LB) from Formedium™: 0.5% (w/v) yeast extract, 1% (w/v) tryptone and 1% (w/v) NaCl, with appropriate antibiotics (ampicillin 100 µg/mL; kanamycin 50 µg/mL)
- LB broth agar: LB media + 1.5% agar, with appropriate antibiotics (ampicillin 100 µg/mL; kanamycin 35 µg/mL) and 20 µg/mL XGal and 1 mM IPTG when required.

E. coli strains: The strains used throughout this study are listed in **Table 1**.

Table 2.1. Bacterial strains used in this project

<i>E. coli</i> strain	Genotype	Description
Top10	$\Delta(ara-leu)$ 7697 <i>araD139 fhuA</i> <i>\Delta lacX74 galK16 galE15 e14-</i> <i>\phi80dlacZ\Delta M15 recA1 relA1 endA1</i> <i>nupG rpsL (StrR) rph spoT1 \Delta(mrr-</i> <i>hsdRMS-mcrBC)</i>	High efficiency transformation used for plasmid cloning and amplification.
BL21(DE3)	<i>F- ompT hsdS_B (r_B- m_B-) gal dcm lon</i> <i>(DE3)</i>	Used in over-expression of recombinant protein. Protease deficient strain and suitable for transformation. B834 derivative.
B834(DE3)	<i>F- ompT hsdS_B (r_B- m_B-) gal dcm met</i> <i>(DE3)</i>	Used for Seleno-methionine labelling of protein. Methionine auxotroph. Parent of BL21.
Rosetta(DE3)	<i>F- ompT hsdS_B (r_B- m_B-) gal dcm met</i> <i>(DE3) pRARE (Cam^R)</i>	Used to enhance eukaryotic protein expression. BL21 derivative.
T7 express	<i>fhuA2 lacZ:T7 gene1 [lon] ompT gal</i> <i>sulA11 R(mcr-73:miniTn10--TetS)2</i> <i>[dcm] R(zgb-210:Tn10--TetS) endA1</i> <i>\Delta(mcrC-mrr)114:IS10</i>	Used for high efficiency transformations and protein expression.

Chemically competent cells:

- Solutions:
 - Solution 1: 0.1 M MgCl_2 autoclaved and chilled at 4°C
 - Solution 2: 0.1 M CaCl_2 autoclaved and chilled at 4°C
 - Solution 3: 0.1 M CaCl_2 + 15% sterile glycerol

An aliquot of cells was plated on a LB agar plate and incubated overnight at 37°C. A single colony was picked from the plate for overnight growth in 10 mL of fresh LB media. One mL of the overnight growth culture was used to inoculate 100 mL of fresh LB media and grown at 37°C with shaking at 200 rpm until the cells reached an OD_{600} of 0.3-0.4; the cells were then cooled on ice for 10 minutes and pelleted by centrifugation at 3,000g for 5 minutes at 4°C. The cells were kept on ice during the following steps. The pellet was suspended gently in 30 mL of Solution 1 and incubated on ice for 30 minutes. After incubation, the suspended cells were pelleted and resuspended gently in 30 mL of Solution 2 and incubated on ice for 30 minutes. The cells were pelleted for the last time and resuspended in 2 mL Solution 3. Competent cells were flash cooled as 50 μL aliquots in liquid nitrogen and kept at -80°C until use.

Transformation by heat-shock:

Prior to transformation, aliquots of competent cells were thawed slowly on ice. Five μL of newly cloned plasmid or 1 μL of purified plasmid was added to the aliquot of competent cells and incubated on ice for 5 minutes. The transformation was performed by heat shock at 42°C for 45 seconds. Immediately after heat shock, the cells were cooled on ice for 2 minutes before the addition of 700 μL of LB media and incubated for 1 hour at 37°C. The cells were then pelleted and resuspended in 20-30 μL media and plated on LB agar plates with appropriate antibiotics. The LB-agar plates were incubated overnight at 37 °C.

QIAprep Spin miniprep kit (QIAGEN):

A 15 mL overnight culture from a single colony of the transformed Top10 cells was grown at 37°C with shaking at 200 rpm from which plasmids were extracted following the manufacturer's instructions. The concentration of purified DNA was determined with a NanodropTM system (ThermoFisher Scientific) at an absorbance of $A_{260\text{nm}}$ and the purity checked with the A_{260}/A_{280} ratio, where a ratio of 1.8 indicates the sample contains pure DNA.

DNA visualisation by electrophoresis:

PCR products and DNA digestions were visualised by electrophoresis on 1% agarose gel with 1x TAE buffer. The gels were made with the SYBR safe DNA gel stain (ThermoFisher

Scientific). The GeneRuler 1kb DNA ladder (ThermoFisher) was used as a marker reference and the gel imaging was performed using the Gel Doc EZ viewer system (BioRad).

QIAquick PCR purification kit (QIAGEN)

The PCR purification kit operates under the same principles as the miniprep kit. The PCR products were purified either after running the amplicon on an agarose gel or right after the PCR reaction, using the standard protocol from the QIAquick PCR purification kit. The PCR product was eluted with milliQ water and the concentration was assessed by Nanodrop™.

Sequencing:

To confirm successful cloning, plasmids were analysed by DNA sequencing. All the sequencing reactions were performed by Eurofins Genomics services using the Sanger sequencing method with 10 µL of plasmid at 50-100 ng/µL and 5 µL of the appropriate primer at 10 pmol/µL.

2.1.1 CIDAR MoClo Assembly

The CIDAR MoClo assembly cloning method (Iverson et al., 2016) was mainly used during this project. This assembly method allows a one-pot digestion/ligation reaction using the Type IIS restriction enzymes *BbsI* and *BsaI*, and T4 DNA ligase enzyme. These Type IIS restriction endonucleases recognise 6 bp non-palindromic sequences and cut outside their recognition sequence to generate overhangs of 4 bp. The CIDAR MoClo method uses two steps (**Figure 2.1**): first, the Gene of Interest (GOI) is cloned into a Destination Vector with Ampicillin resistance (DVA); second, the GOI is transferred from the DVA to an expression vector with kanamycin resistance. The expression vector used herein is pET28gg (Abil et al., 2014), derived from pET28a, which contains the *lacZ* selection gene in the cloning site that can be expressed with the T7 expression system. The GOI sequences were obtained from the *L. monocytogenes* EGD entry on the NCBI website. All codons were optimized for expression in *E. coli* using the GeneOptimizer webserver (ThermoFisher Scientific). The DVA used here to clone the main genes of interest possesses the C- (-AATG-) and D-type (-ATTG-) overhangs and *BbsI* restriction sites. In order to clone the GOI into the CD_DVA vector, the extremities of the GOI were designed to complement the C and D overhangs, which contain the *BbsI* recognition sites. The GOI were synthesized by IDT (Integrated DNA Technologies). The cloning reaction contained 1 µL of CD_GOI, 1 µL of the CD_DVA *lacZ* plasmid, 1 µL of T4 ligase enzyme, 1 µL of 10x ligase buffer, 1 µL of *BbsI* enzyme and the reaction was made up to 10 µL with milliQ water.

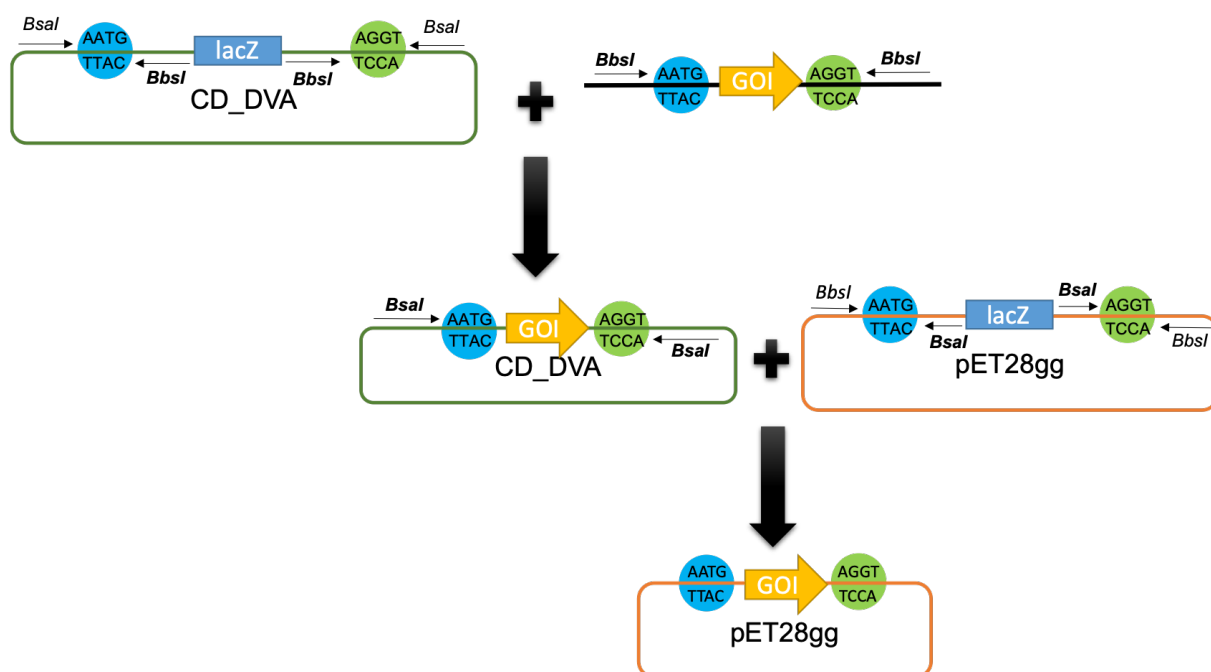


Figure 2.1. Schematic representation of the CIDAR MoClo DNA assembly method

The green plasmids represent ampicillin resistance and orange plasmids have kanamycin resistance. The *BbsI* and *BsaI* restriction sites are represented by arrows. The C-type overhang (-AATG-) is drawn as a blue circle and the D-type overhang (-AGGT-) as a green circle. From the cloning of the CD_DVA *lacZ*-containing plasmid with the GOI results in the CD_DVA containing the GOI and the replacement of the *BbsI* restriction site with that of *BsaI*. The GOI is transferred to pET28gg and loses the restriction site.

The cloning reaction was set up with a thermal cycler (**Table 2.2**) and the cloning product was used to transform *E. coli* Top10 cells by heat shock before plating on ampicillin/IPTG/X-Gal LB agar plates for selection. The positive colonies were selected by screening for LacZ complementation; the resulting white colonies replace the LacZ α fragment expression cassette with the GOI, while blue colonies retain the LacZ α fragment and are negative for the GOI insert. Single positive colonies were picked and inoculated into fresh LB media with ampicillin and incubated overnight at 37°C. Plasmids were isolated from positive colonies with a miniprep kit (QIAGEN) and subjected to Sanger sequencing by Eurofins to confirm the insert. Successful cloning also resulted in the replacement of the *BbsI* restriction sites with those of *BsaI*, which are also present in pET28gg. The GOI cloned in the CD_DVA was then transferred into the pET28gg expression vector using the same protocol as the DVA cloning but using *BsaI* restriction instead of *BbsI*. The positive colonies were selected by LacZ α complementation screening and plasmids isolated from positive colonies by miniprep were subjected to sequencing to confirm the insert.

Table 2.2. MoClo thermal cycling conditions

Step	Temperature (°C)	Time (min)	No of cycles
<i>BbsI</i> / <i>BsaI</i> activity	37	1:30	25
T4 Ligase activity	16	3:00	
Ligation	50	6:00	1
Enzyme deactivation	80	10:00	1

The RsbR domains of interest were amplified using the Polymerase Chain Reaction (PCR) with forward and reverse primers designed to clone into the DVA_CD plasmid using the gBlocks as template (**Table 2.3**). The PCRs were performed using ThermoFisher 2x Master Mix containing Phusion polymerase with the recommended PCR cycles (**Table 2.4**). The amplicons were separated by electrophoresis on a 1% agarose gel and purified by excision of the bands corresponding to the amplicon using the PCR purification kit. The amplified fragments were then used to clone into the CD_DVA vector and in the pET28gg as described previously.

Table 2.3. Primers used to amplify the NTD of RsbR genes.

The C site for the forward (FWD) primer and the D site for the reverse (RVS) primer are underlined. The *BbsI* restriction sites are highlighted in bold.

Primer	Sequence
N-RsbR1 FWD	5'-GAT GAAGAC ATAATGTACAAAGACTTCGCCAAC-3'
N-RsbR1 RVS	5'-GAT GAAGAC ATACCTTTATTCTTGCAGTGCG-3'
N-RsbR2 FWD	5'-GAT GAAGAC ATAATGAATGAAAGCAATGGTAGCATG-3'
N-RsbR2 RVS	5'-GAC GAAGAC TAACCTTTATTCTTCAATCAGACGATG-3'
N-RsbR3 FWD	5'-GAT GAAGAC ATAATGCAGATCAAAGAATTTCTGATTAG-3'
N-RsbR3 RVS	5'-GAC GAAGAC TAACCTTTACTGAATAATTTCTTTGCG-3'
N-RsbR4 FWD	5'-GAT GAAGAC ATAATGGGCATTATGAATAGCCTGC-3'
N-RsbR4 RVS	5'-GAC GAAGAC TAACCTTTATTTGATGATCTGGTTCTCTTTC-3'

Table 2.4. PCR cycles used for DNA amplification

Step	Temperature (°C)	Time (min)	No of cycles
Initial denaturation	95	1:30	1
Denaturation	95	0:15	25
Annealing	50-68	0:20	
Elongation	72	1min/kb	
Final Elongation	72	5:00	1

2.1.2. Production of protein sequence variants

Several His-tagged versions of the N-RsbR4 protein were generated using appropriate reverse and forward primers (**Table 2.5**) by PCR as previously described using the gBlocks as templates. Following the PCR, the methylated parental template DNA was digested by *DpnI* treatment. The cloning product was then used to transform *E. coli* Top10 cells. After overnight incubation on LB agar, several colonies were selected and grown in 10 mL of LB media supplemented with appropriate antibiotics overnight at 37°C. Plasmids were purified using a miniprep kit (QIAGEN) and verified by Sanger sequencing.

Table 2.5. Primers used for site-directed mutagenesis

The changed codons are in small caps, the *BsaI* restriction site is in bold, the added His-tag is in orange and the C3S TEV protease cleavage site is in purple.

Primer	Sequence
H6-C3S-N-RsbR4 FWD	5'-GGTCTCAAATGcaccaccaccaccac TTGGAAGTTCTGTTTCAGGGTCCGATGGGCATTATGAATAGCCT G-3'
IVQKE-STOP-N-RsbR4 RVS	5'-CAGATGATTGTTTCAGAAAGAGTGATAGAGACCtactagta-3'
IIK-STOP-N-RsbR4 RVS	5'-CTACTAGTAGGTCTCTACCTTTATTTGATGATCTGGTT-3'
IVQKE-NO_STOP-N- RsbR4 RVS	5'-CAGATGATTGTTTCAGAAAGAGAGGTAGAGACCtactagta-3'

2.1.3. Fusion proteins

Maltose Binding Protein fusions

A Maltose Binding Protein (MBP) fusion was made for the RsbT kinase as various members of the Lewis lab had found its homologues were unstable in isolation. The MBP fusion was cloned at the RsbT N-terminus using pMAT11 (**Figure 2.2**) containing the MBP coding sequence in frame with the multiple-cloning site, and an ampicillin resistance cassette. The GOI was amplified by PCR with accordingly designed reverse and forward primers (**Table 2.6**). A double digestion of the amplicon and the plasmids was performed at 37 °C for 1 hour using *NcoI* and *EcoRI* restriction enzymes. The digestion product was separated by electrophoresis on a 1% agarose gel and the bands for the GOI and the linear plasmid were excised and purified before ligating with T4 DNA ligase at 22 °C for 1 hour. The ligation product was used to transform *E. coli* Top10 competent cells and plasmids were isolated by miniprep from resulting colonies. The isolated plasmids were subjected to Sanger sequencing (Eurofins) to confirm the presence of the GOI.

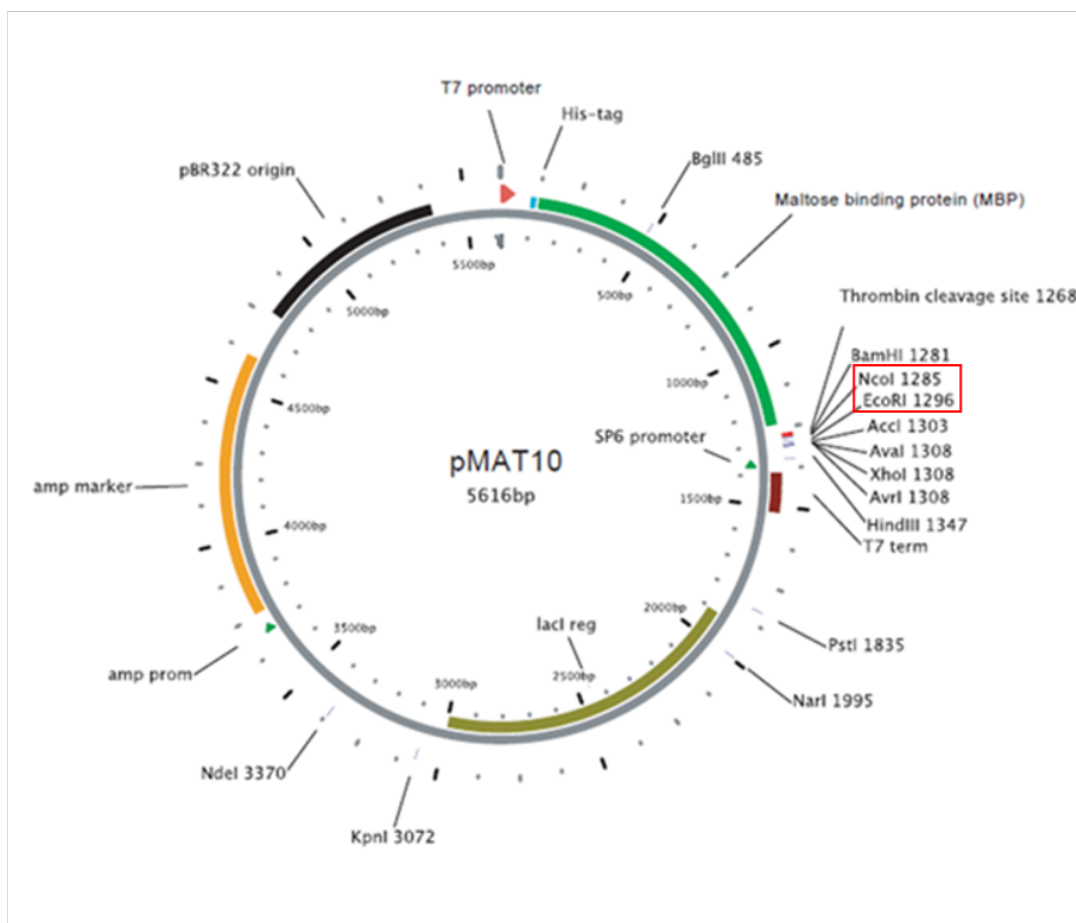


Figure 2.2. pMAT10 plasmid map.

pMAT11 is similar to pMAT10, with only minor modifications (Littlefield K and Owen D., unpublished). The GOI were inserted between *NcoI* and *EcoRI* restriction sites, in red.

Table 2.6. Primers used for MBP-fusion proteins

The restriction sites are highlighted in bold.

Primer	Sequence
RsbT-NcoI FWD	5'-GAC CCATGGG CACCTTTCAGAGCTGCG-3'
RsbT-EcoRI RVS	5'-GGTTATTACAACCACCAAATGGGTTTCGTTAAG AATTC GATC-3'

GFP-fusion

A GFP-fusion of Prli42 miniprotein was performed for interaction studies. A double stranded gBlock for Prli42 was synthesized by IDT using the A site on its 5' (-GGAG-) and the C (-AATG-) site on its 3', with *BbsI* restriction sites for the MoClo cloning system. The synthesized A-C T7 promoter with Prli42 fragment, the DVA_CD containing the GFP gene, the DVA_DE containing a StrepII tag and the DVA_EF containing the terminator were used to clone the Prli42-GFP-StrepII tag fusion in a DVK_AF (**Figure 2.3**). The cloning product was used to transform *E. coli* Top10 cells by heat shock before plating on kanamycin/IPTG/X-Gal

LB agar plates. Positive colonies were selected by screening for LacZ complementation. The plasmid was isolated from positive colonies and sequenced to verify the inserts.

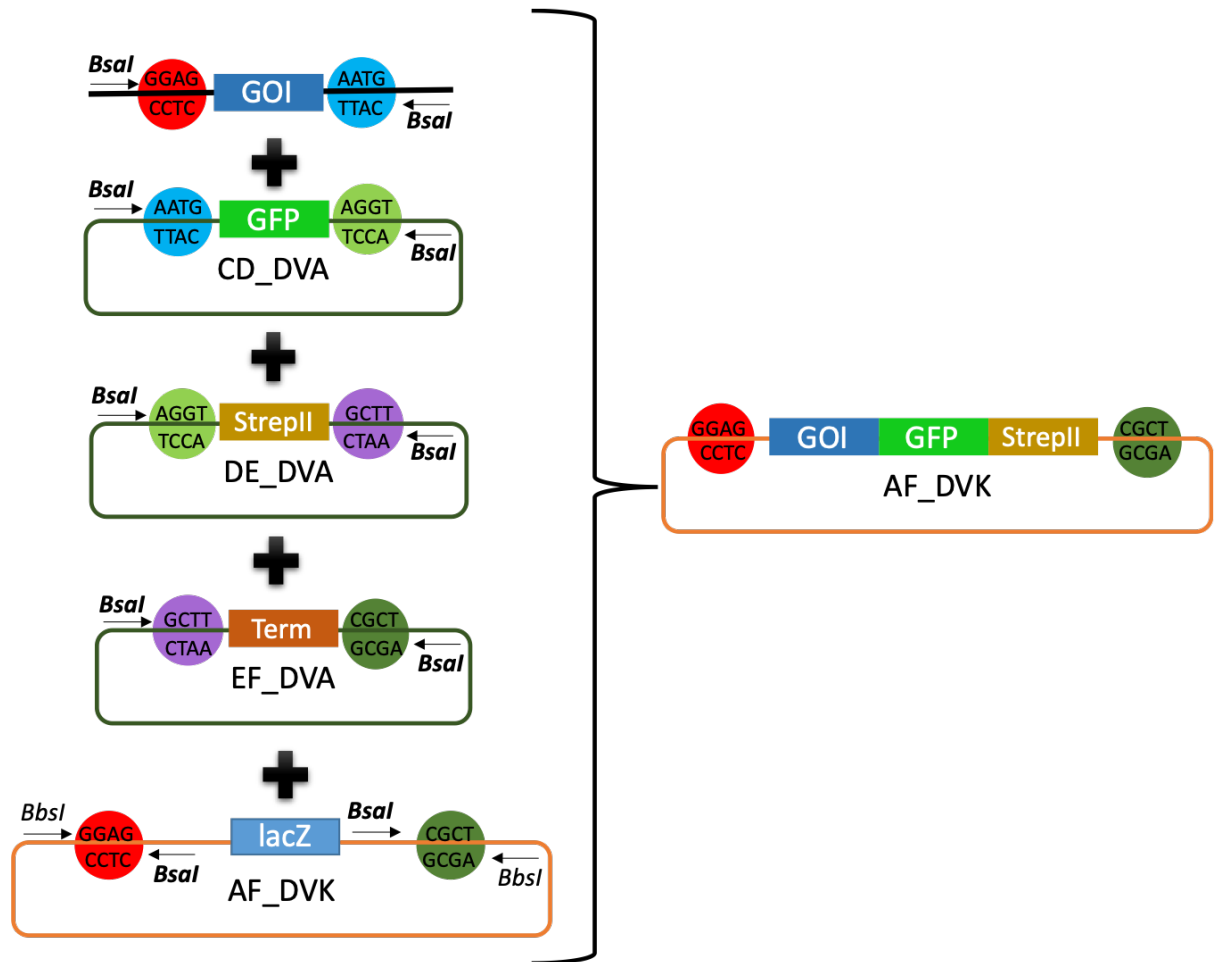


Figure 2.3. Schematic representation of the multi assembly using CIDAR MoClo cloning.

The green plasmids represent ampicillin resistance and orange plasmids have kanamycin resistance. The *BbsI* and *BsaI* restriction sites are represented by arrows. The A, C, D, E and F overhangs are coloured as shown in the legend of the figure. All the matching overhangs will anneal together when digested with *BbsI*, into the AF_DVK, by replacing the *lacZ* fragment.

2.1.4. Co-expression vector

A co-expression vector with a double expression cassette for the *L. monocytogenes* RsbR-RsbS stressosome complex was produced as follows. Using the same principle as the GFP fusion (**Figure 2.4**), the RsbR gene from CD_DVA vector have been transferred into an AE_DVK plasmid with the T7 promoter and terminator using *BbsI* restriction sites. The same was performed in parallel for *rsbS* in CD_DVA, but this time it was transferred into an EF_DVK plasmid. The two genes cloned into the DVK plasmids were transferred to an AF_DVA plasmid that has the same overhang extremities. The selection of the positive clones was performed as previously described.

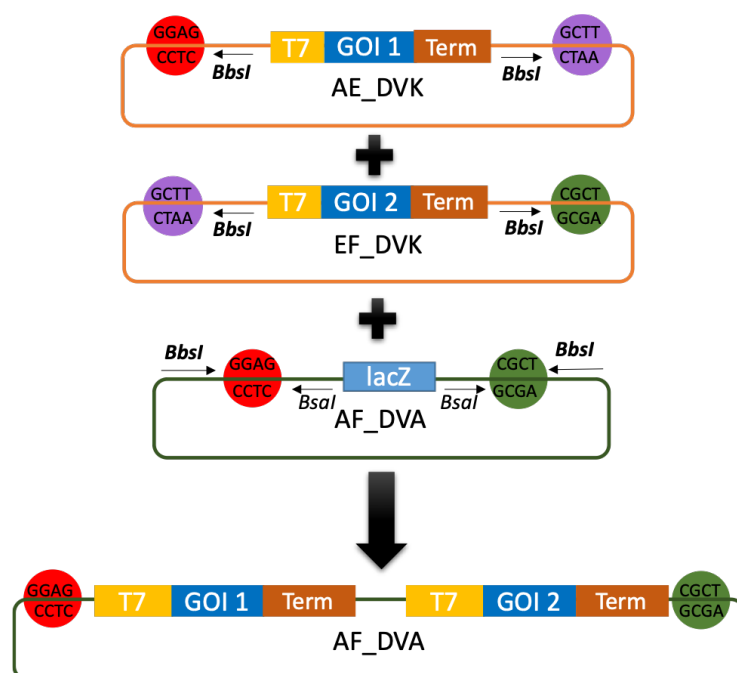


Figure 2.4. Schematic representation of the co-expression vector assembly using CIDAR MoClo cloning.

The green plasmid represents an ampicillin resistant backbone and orange plasmids have kanamycin resistance. The *BbsI* and *BsaI* restriction sites are represented by arrows. The A, E and F overhangs are coloured in red, purple and dark green respectively. When the 2 DVK plasmid with *BsaI* enzyme and the destination vector AF_DVA are mixed in a single tube reaction, it allows the cloning of the 2 genes in the same plasmid, therefore creating a co-expression vector.

2.2. Protein expression

General buffers and solutions:

- 5x SDS loading buffer: 250 mM Tris-HCl pH 6.8, 10% SDS, 30% glycerol, 0.02% Bromophenol blue, 2 mM DTT
- 10x SDS running buffer: 250 mM Tris-HCl pH 6.8, 1.92 M glycine and 1% SDS
- Coomassie staining solution: 40% ethanol, 10% acetic acid, 0.1% Coomassie (R250/G250)
- Destaining solution: 40% ethanol and 10% acetic acid

Protein analysis by SDS-PAGE:

Protein samples were analysed by 15% SDS-PAGE (gel recipe **Table 2.7**) using the Biorad ProteanTetra cell gel casting and running equipment. The SDS-PAGE gels were loaded with a molecular weight marker (PageRuler Prestained Protein Ladder, range 10-180 kDa, ThermoFisher) and 10µL of protein sample mixed with 1x loading buffer and denatured for 10 minutes at 100°C. Electrophoresis was performed using 1x SDS-PAGE running buffer at 200 V for 50 minutes. After electrophoresis, the gels were stained with Coomassie Blue staining solution for 30 minutes. The gels were rinsed with water and incubated with destaining solution

until the excess of Coomassie stain was removed from the gel, which were then visualised on a lightbox and scanned with a desktop scanner, or the BioRad gel Doc Imaging system.

Table 2.7: Recipe used for 15 % SDS-PAGE gels

4 gels, 0.75mm thickness	Stacking	Resolving
%	5	8
30% Bis-acrylamide (mL)	1	7.5
H ₂ O (mL)	5.84	7.1
1.5 M Tris pH8.8 (mL)	0	5
1 M Tris pH6.8 (mL)	1	0
10% SDS (μL)	80	200
10% APS (μL)	80	200
TEMED (μL)	8	8
Total (mL)	8	20

2.2.1. Protein Expression tests

E. coli protein production specific strains (mainly BL21, rarely T7 express and Rosetta and B834 for selenomethionine labelling) were transformed with the genes cloned in pET28gg. The resulting colonies were used in small-scale overnight cultures in liquid LB media with appropriate antibiotics. The cell culture was divided into three batches: one batch was treated as a non-induced control; the second batch was induced with 1 mM IPTG at 18 °C for 3h, and the third batch was induced with 1 mM at 37 °C for 3 hours. The cells were pelleted and lysed with 1x SDS-PAGE loading buffer. The lysate samples were heated at 100 °C and subjected to SDS-PAGE with reference to the PAGE-Ruler protein ladder (ThermoFisher).

2.2.2. Large-scale recombinant protein expression

E. coli BL21(DE3) competent cells were transformed with the cloned expression vectors. One colony was picked to inoculate 10 mL LB media as a starter culture, supplemented with the appropriate antibiotic, and incubated at 37 °C with shaking at 200 rpm until the cultures reached visual turbidity (≈3h). One litre of LB medium supplemented with the appropriate antibiotic was inoculated with the starter culture and grown at 37°C with shaking 200 rpm. Cultures were induced with 1 mM IPTG when the OD_{600nm} reached 0.6 and incubated overnight at 20 °C with shaking. The cells were pelleted by centrifugation 5,000 rpm for 30 minutes, washed with 1x PBS and kept at -20 °C until commencing the purification.

2.2.3. Selenomethionine labelled protein expression

Solution and buffers:

- Vitamin mix: niacinamide, pyroxidine monochloride, riboflavin and thiamine at 1 mg/mL each
- Amino acid mix I: 0.4% (w/v) of each non aromatic amino acid residue: alanine, arginine, asparagine, aspartic acid, cysteine, glutamine, glycine, histidine, isoleucine, leucine, lysine, proline, serine, threonine, and valine.
- Amino acid mix II: 0.4% (w/v) of each aromatic residue: phenylalanine, tryptophan, and tyrosine.
- 40% (w/v) glucose, filter sterilized
- 1 M magnesium sulphate, separately autoclaved

Solution made before use:

- Iron sulphate at 12.5 mg/mL, filter sterilised.
- L-Se-Met at 10 mg/mL, filter sterilised.

Media:

- 20x M9 media: 2% (w/v) ammonium chloride, 6% (w/v) potassium dihydrogen phosphate and 12% (w/v) disodium phosphate
- Se-Met pre-mix media for 1 L: 1x M9 media, 10 mL of amino acid mix I, 10 mL of amino acids mix II and 1 mL of vitamin mix were autoclaved together. Two mL of magnesium sulphate, 25 mg of iron sulphate, 4% (w/v) glucose solution were added after autoclave.
- Se-Met final media for 1 L: Se-Met pre-mix media and 40 mg of Se-Met.

To selenomethionine label proteins for experimental phasing of their crystals, the methionine auxotroph *E. coli* B834 competent cells was transformed with the NTD-RsbR expression vectors. An overnight cell culture was made from a single colony in 100 mL of LB media with appropriate antibiotics and incubated at 37°C with shaking at 200 rpm. A 100 mL starter culture was inoculated with 1 mL of the overnight culture and incubated until the OD_{600nm} reached ~0.2. The cells were then pelleted and washed with the premixed Se-Met media. The pre-mixed media was used to wash the cell pellets by resuspension twice to remove traces of LB media. The cells were then inoculated into the Se-Met media and a small aliquot of inoculated cell culture was removed to use as a negative control. L-Se-Met was added to the Se-Met media flask and incubated at 37°C. The OD_{600nm} of the Se-Met and the control cultures were checked hourly. The control aliquots stopped growing and remained at a low OD_{600nm}, while the cells in the L-Se-Met supplemented growth medium the OD_{600nm} increased. The cell culture was induced overnight at 20°C with 1 mM IPTG when the culture reached an

OD_{600nm} of 0.5-0.6. The cells were pelleted by centrifugation at 5,000 rpm for 30 minutes, washed with 1x PBS and kept at -20 °C until protein purification.

2.2.4. Large-scale *B. subtilis* growth for native stressosome purification

The WT and mutants of *B. subtilis* strain 168 were obtained in collaboration with Prof. Nicola Stanley-Wall in Dundee. The mutants each have a poly-histidine tag on the RsbRA NTD for pull-downing the complex by nickel affinity chromatography. The sequences were designed to introduce the His-tag on Loop 4 of the *B. subtilis* RsbR structure (PDB ID: 2BNL) (**Figure 2.5**).

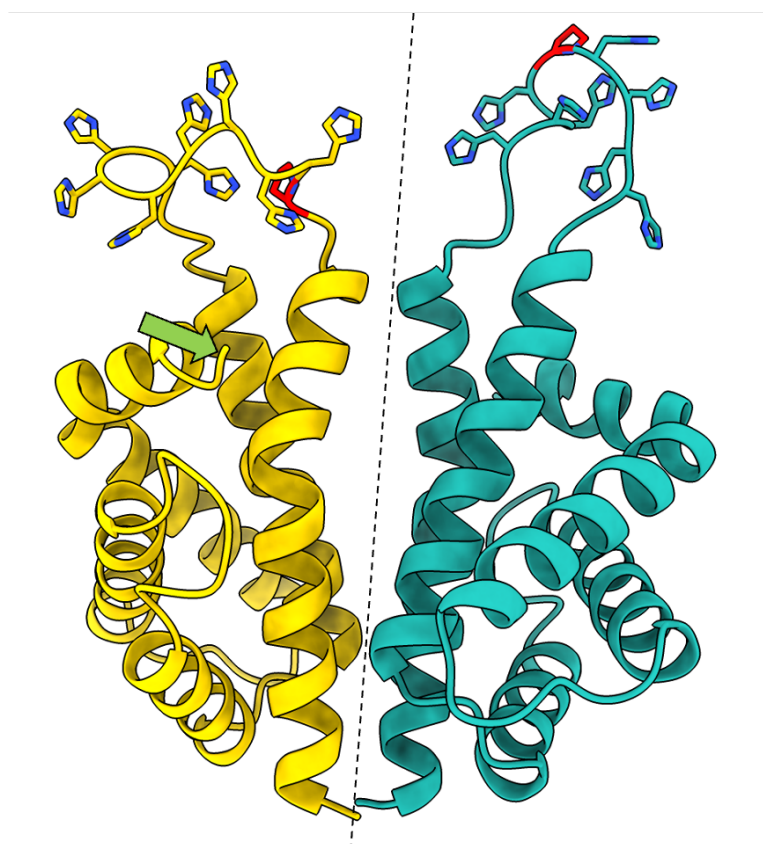


Figure 2.5. Model of *B. subtilis* N-RsbRA structure with sites of His-tags.

RsbRA was modelled for histidine-nickel affinity purification. The His₈-tag insertions were modelled. The NRS5621 (yellow) strain has a His₈-tag before the helix-breaking proline 113 in red, on loop 4. The NRS5623 (cyan) strain has four histidine residues on either side of the breaking proline shown in red on loop 4. The N-terminal tag insertion (green arrow) was lethal.

Three types of variants were proposed, one variant with an N-terminal His₈-tag and two with the His₈-tag on loop 4. For the loop variants two constructs were designed: one with a continuous His₈-tag (strain NRS5621) and one His₈-tag with 4 histidine residues on each side of the already present proline 113 (strain NRS5623). The His₈-tag on the N-terminus of RsbRA

was lethal, perhaps because of a downstream polar effect, and consequently no clones were obtained.

Large-scale *B. subtilis* growth was performed using LB media without antibiotics. The media were inoculated with an overnight starter culture and incubated at 37°C with shaking at 200 rpm. The cell growth was monitored hourly for the OD_{600nm} and the cells were harvested when cell growth reached stationary phase (8-10 hours). The cells were kept at -20°C until purification.

2.3. Protein purification

Buffers and solutions:

Ion exchange chromatography purification

- QA buffer: 25 mM Tris-HCl pH8.0 (+2 mM DTT for Se-Met preps)
- QB buffer: 25 mM Tris-HCl pH8.0 and 1 M NaCl (+2 mM DTT for Se-Met preps)
- SA buffer: 0.1M sodium acetate pH 5 + 2 mM MgCl₂
- SB buffer: SA buffer + 500 mM NaCl

Nickel Affinity purification

- HisA buffer: 25 mM Tris-HCl pH8, 500 mM NaCl and 50 mM Imidazole pH8
- HisB buffer: HisA buffer + 500 mM Imidazole

Streptavidin affinity chromatography

- Buffer W: 100 mM Tris-HCl pH8, 150 mM NaCl and 1 mM EDTA
- Buffer E: Buffer W + 2.5 mM desthiobiotin

Size-exclusion purification

- GF buffer: 25 mM HEPES-NaOH pH8 and 150 mM NaCl (+1 mM DTT for Se-Met preps)

2.3.1. Purification of protein by affinity chromatography

Ion-Exchange chromatography

Untagged proteins were purified by anion exchange chromatography at room temperature using the ÄKTA protein purification system and Q-Sepharose HiLoad 16/10 HP and/or MonoQ 5/50 GL columns (GE Healthcare).

Cell pellets were resuspended in (10x v/w) QA buffer and were lysed by sonication (pulse on/off 1 sec at 40% amplitude for 5 minutes) on ice. The lysates were clarified by centrifugation (19,000 × g, 30 min) and the supernatant was filtered with 0.45 µm and 0.22 µm filters (Millipore). The filtered supernatant was loaded onto the Q-Sepharose column pre-equilibrated with QA buffer. The column was washed with QA buffer, to remove unbound proteins, until the absorbance A_{280nm} decreased and stabilised. Bound proteins were eluted using a linear gradient (0-100%) of QB buffer over 10 Column Volumes (CV). Collected fractions were analysed by 15% SDS-PAGE and the fractions containing the protein of interest were pooled. The pooled fractions were concentrated to 2-4mL using a centrifugal concentrator

(Amicon, Millipore) with the appropriate molecular mass cut-off prior to size-exclusion chromatography.

Untagged RsbX was purified by cation exchange chromatography using the same protocol as the anion exchange chromatography with the appropriate buffers (SA and SB) and an FF S Sepharose 5mL column.

Immobilised metal ion affinity chromatography

His-tagged proteins were purified by immobilized metal affinity chromatography (IMAC) using nickel bound to nitrilotriacetic acid (Ni-NTA) agarose using nickel-poly-histidine affinity using HisTrap Fast Flow columns (GE Healthcare). Bound proteins were eluted using imidazole.

Similar to the ion-exchange affinity purification, the cell pellets were suspended in HisA buffer and were lysed by sonication on ice. The lysates were clarified by centrifugation as before and the supernatant was filtered before loading onto the HisTrap FF column pre-equilibrated with HisA buffer. The column was washed with HisA buffer, to remove unbound proteins, until the absorbance $A_{280\text{nm}}$ decreased and stabilised. Bound proteins were eluted using a linear gradient (0-100%) of HisB buffer over 10 CV. Collected fractions were analysed by 15% SDS-PAGE and the fractions containing the protein of interest were pooled and concentrated to 2-4 mL using a centrifugal concentrator (Amicon, Millipore) with the appropriate molecular mass cut-off prior to size-exclusion chromatography.

Streptactin-desthiobiotin affinity chromatography

StreptII-tagged proteins were purified using the Streptactin-biotin affinity. As the binding of biotin is irreversible, desthiobiotin was used as the competing ligand instead.

The cell pellet was resuspended into buffer W and lysed by sonication. After lysate clarification performed as above, the filtered supernatant was loaded into a home poured (10 ml) column with Streptavidin Sepharose resin (GE Healthcare). The column was washed with 5 CV of buffer W, and the protein was eluted using a desthiobiotin gradient of 5 CV with buffer E. The fractions were analysed by 15% SDS-PAGE and the fractions containing the protein of interest were pooled and concentrated with a centrifugal concentrator (Amicon, Millipore) for further size-exclusion chromatography.

Purification of protein by high-resolution ion-exchange chromatography

MonoQ HR 5/5 column (GE Healthcare) with a higher separation resolution than Q-sepharose was used. Proteins were diluted to reduce the salt concentration to below its elution point and loaded on a pre-equilibrated MonoQ HR 5/5 column. The proteins were eluted using

a salt gradient with QB over 10 CVs and the collected fractions were analysed by 15% SDS-PAGE. The fractions containing pure protein were concentrated and subjected to a further round of size-exclusion chromatography. After analysing the fractions by SDS-PAGE, the concentration of purified recombinant proteins was assessed using the NanodropTM system at A_{280nm} and calculated extinction coefficients based on their amino acid sequences using the ProtParam webserver (Wilkins et al., 1999); the purified proteins were used immediately for crystallization screening or other downstream applications. Excess purified protein was flash cooled in liquid nitrogen and stored at -80°C.

2.3.2. Purification of proteins by size-exclusion chromatography

Concentrated protein samples were further purified by size-exclusion chromatography using Superdex (S75 or S200) 16/60 columns for the proteins and a Superose 6 10/300 GL for the stressosome complex (GE Healthcare). The columns were pre-equilibrated with GF buffer and the sample was loaded into the column. The proteins were separated according to their size, with aggregates and high-molecular weight proteins eluting first, followed by smaller size proteins. The fractions were analysed, and their purity was assessed by 15% SDS-PAGE. If the protein sample had a purity level of >90%, the proteins were concentrated and used immediately for crystallography. Excess protein was flash cooled in liquid nitrogen and stored at -80 °C, until further use. If the protein sample was not pure enough for downstream experiments an additional step of high-resolution anion-exchange chromatography was performed using the MonoQ column.

2.4. *In vitro* complex reconstitution

The stressosome complex was reconstituted from purified proteins by mixing RsbRs:RsbS at a 2:1 ratio, and either the RsbT kinase or the RsbX phosphatase was added at a 2:1:1 ratio. The mix was incubated at room temperature for 1 hour before purification by SEC using a Superose 6 10/300 GL as described in **Section 2.3.2**.

2.5. Circular Dichroism

Circular Dichroism (CD) was used to assess whether the protein was folded or not. CD is a method that uses circularly polarised light to evaluate the secondary structure and folding properties of proteins by exploiting the amides of the polypeptide backbone of proteins (Greenfield, 2006). Electromagnetic radiation (ER) consists of an oscillating electric field and a magnetic field that are perpendicular to each other and to the direction of the propagation. ER

can be polarised in 2 directions, left and right. The electric field vector rotates around the propagation axis which allows it to maintain a constant magnitude. CD is based on the absorption difference between left- and right-circularly polarized light using the Beer-Lambert Law, where the molar absorptivity of a medium will be different for the left- and right-circularly polarised light. This difference is known as the molar circular dichroism. The measured difference can be quantified and traced, either in units of ΔE , the difference in absorbance of E_R and E_L by an asymmetric molecule, or in degrees ellipticity, which is defined as the angle whose tangent is the ratio of the minor to the major axis of the ellipse, in θ . The results are affected by different secondary structural elements in proteins, which have characteristic CD spectra based on the organisation of the peptide backbone (**Figure 2.6**).

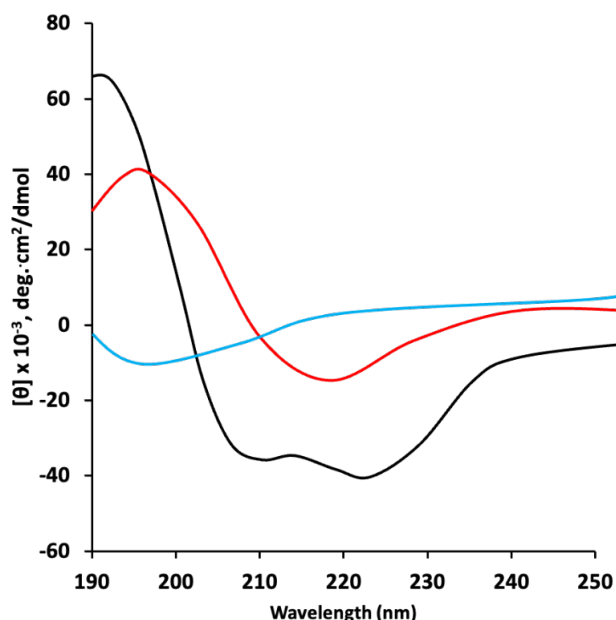


Figure 2.6. CD spectra characteristics of different protein secondary structures

Proteins with α -helices have negative values at 208 and 222nm and a positive value at 193 nm shown in black. Proteins with well-defined antiparallel β -pleated sheets (β -helices) have negatives at 218 nm and positives at 195 nm as shown in red. Disordered regions have very low ellipticity above 210 nm and negative values near 195 nm in cyan. Adapted from Greenfield, 2016.

CD experiments were run for each protein of interest at far UV (190-260nm) for secondary/tertiary structure and at near UV (260-320nm) for the quaternary structure. Compared to the far-UV range analysis, the near-UV range analysis is less well used. At the far-UV, a blank control run was performed using water. The spectrum was read 10 times and averaged. The protein CD spectra were obtained using protein diluted in water at a final concentration of 5 μ M and the run was performed at 20°C at a scan speed of 20 nm per minute. The resulting blank spectrum was subtracted from the protein spectrum to remove any non-specific signal. For the near-UV experiment, the blank control was performed with the protein

buffer (GF Buffer) and the protein spectra was obtained with a higher protein concentration (between 1000 and 1500 μM) to get better signal as the near-UV range is less sensitive to noise compared to the far-UV range. The same parameters as the far-UV range has been used, except the wavelength range. Finally, the far UV range data were subjected to the Bestsel webserver (<http://bestsel.elte.hu/index.php>) to optimize the spectra and the data were plotted in GraphPad Prism software.

Melting curves were obtained using CD to assess protein thermostability. The melting curves were measured using a temperature ramp from 20-90°C using a fixed wavelength of 222nm. The same protein concentration as the far-UV run was used, with a ramp speed of 1°C per minute.

2.6. Small Angle X-ray scattering

2.6.1. Sample preparation

Small Angle X-ray Scattering (SAXS) was used to further assess the quality of purified recombinant proteins and to confirm CD data. SAXS is a method which uses soft X-rays and provides low resolution information on the overall shape, conformation and assembly state of macromolecules in solution (Kikhney and Svergun, 2015).

Protein samples were purified and concentrated up to 20mg/ml and sent to the Dimond Light Source B21 beamline where samples were run by the beamline scientist Nikul Khunti.

2.6.2. Data processing

Primary data reduction was performed using the Scatter software (Förster et al., 2010). The data were analysed for monodispersity and homogeneity in the sample, then the signal corresponding to the sample was selected and the background noise subtracted. The data obtained were then submitted to an evaluation pipeline to analyse the radius of gyration (R_G), the molecular mass and the Porod volume. Finally, the distance distribution function $p(r)$ was computed to convert the data from reciprocal space to real space.

Using the real space information, modelling of the SAXS data was performed using the primus software suite (Konarev et al., 2003) with known structural models (2BNL/3ZTA) and the *L. monocytogenes* N-RsbR structures determined in this project. The initial modelling was done using P1 symmetry, to avoid introducing symmetry bias, but also in P2 symmetry to see if there are any improvements in model quality and fit. The discrepancy calculation, which shows the difference between the model and the experimental data, was performed using the Crysol software. The quality of fit is given by the χ^2 value, the smaller value indicates a better

fit of the model to the data. The models were then manually fitted using the volume fit tool in Chimera (Pettersen et al., 2004).

2.7. Thermofluor Assay

A thermofluor-based thermal shift assay was used to quickly assess protein:ligand binding to identify potential small molecule ligands of the RsbR N-terminal domains. A small molecule library was formulated with a range of potential ligands (**Table 2.9**). The thermal shift assay uses a fluorophore, SYPRO Orange, that binds to hydrophobic patches on the protein. The temperature is slowly ramped up to 98°C; as the protein denatures, it exposes normally buried hydrophobic residues, thus allowing the SYPRO dye to bind to the protein. The fluorescence of SYPRO is quenched in water, and as it binds to hydrophobic amino acids the water is excluded, which results in a fluorescence increase, which is usually measured in a plate reader (**Figure 2.7**).

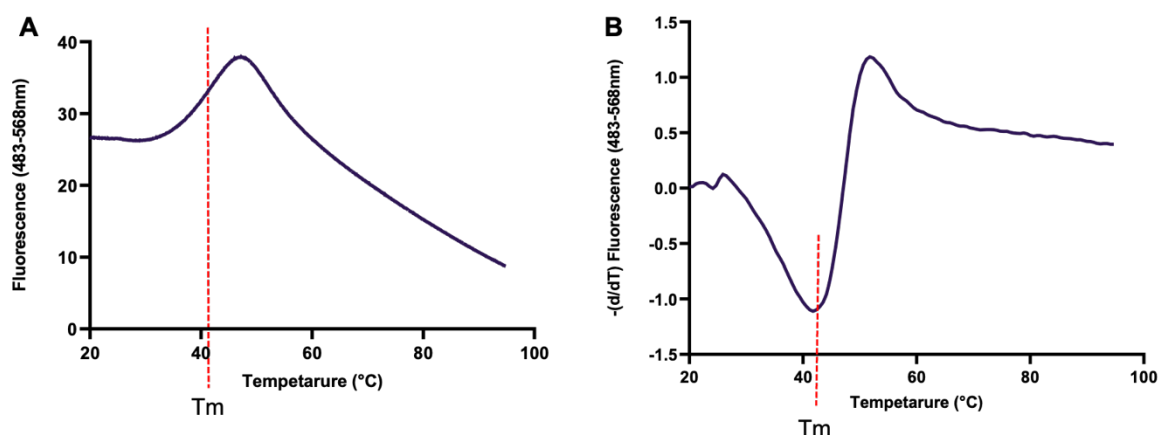


Figure 2.7. Typical TSA results

The primary data obtained are fluorescence readings over the temperature scan (**A**). The mid-point of the first slope represents the T_m of the protein, shown as a dotted red line. To aid interpretation, the negative derivative of the data (**B**) is usually plotted.

SYPRO Orange® was used in the TSA with an excitation wavelength of 300nm and emission between 470-510nm. An initial run was performed to optimize the signal; different protein concentrations were used, ranging from 2 to 7 uM and dye concentration from 2 to 5x (stock solution at 5000x), all diluted in water and aliquoted in a 96 well white and round bottom plate, sealed with UV-transparent adhesive tape. The plates were prepared using the ligand screen in **Table 2.8**, with the optimal protein:dye ratio obtained for each protein in the initial screen. Using the Roche LightCycler 480®, a temperature ramp was set up from 20 to 95°C at

a rate of 0.06°C/sec and the fluorescence was read every second. The resulting fluorescence curves were used to calculate and plot derivatives to obtain estimated T_m values and to calculate the ΔT_m between protein in buffer and with the various ligands. These data were displayed in a histogram grouped by category of molecule (e.g., amino acids, nucleotides, compounds, sugars and buffers).

Table 2.8. Ligand screen for the TSA against the RsbR N-terminal domains

The table below shows the plate setup of the ligand screen at 5x concentration. The compounds were used with a final concentration of 10 mM and the buffers at a final concentration of 25 mM with 150 mM of NaCl.

	1	2	3	4	5	6	7	8	9	10	11	12
A	Glycine	L-Alanine	D-Alanine	Valine	Proline	Leucine	Isoleucine	Phenylalanine	Methionine	Aspartate	Glutamate	Lysine
B	Arginine	Histidine	Serine	O-phospho-L-serine	Threonine	Tyrosine	Asparagine	Glutamine	Tryptophane	Cysteine	Cystine	Ascorbic acid
C	L-Argininamide	Adenine	Guanine	Cytosine	Thymine	Uracil	ADP	dATP	dGTP	dCTP	dTTP	dUTP
D	p-Anisaldehyde	α -lactose	D-(+)-glucose	D-(+)-Maltose	D-fructose 6-phosphate 2-K	Sucrose	Xylitol	L-(+)-arabinose	Octyl-D-glucopyranoside	Glycerol	Ammonium acetate	Calcium acetate
E	Potassium acetate	Sodium acetate	Zinc acetate dihydrate	Ammonium bicarbonate	Sodium bicarbonate	Ammonium carbonate	Potassium carbonate	Sodium hydrogen carbonate	Ammonium citrate dibasic	Ammonium iron (III) citrate	Tri-potassium citrate	Tri-sodium citrate
F	Tri-lithium citrate	Sodium citrate dihydrate	DL tartaric acid	DL-malic acid	Maleic acid	Malonic acid	Succinic acid	Formic acid	Fumaric acid	Ammonium formate	Magnesium formate	Potassium formate
G	Sodium formate	Ammonium metavanadate	Ammonium tartrate	Diethanolamine	Pyridoxine	Sodium pyruvate	Ethylene-diamine	Magnesium sulfate	Potassium oxalate	Potassium phosphate	NADPH	Thiamine
H	PTCP buffer pH4, NaCl	PTCP buffer pH4	Citric acid pH5.5, NaCl	Citric acid pH5.5	Hepes pH7, NaCl	Hepes pH7	Bis-Tris Propane pH8, NaCl	Bis-Tris Propane	Tris pH8.5, NaCl	Tris pH8.5	CAPS pH10, NaCl	CAPS pH10

2.8. Fluorescence polarisation

Fluorescence Polarization (FP), or fluorescence anisotropy, is a method that allows a quick, quantitative and accurate analysis of molecular interactions and this method was used to quantify the interaction between the Prli42 miniprotein and the NTDs of RsbR proteins. The principle of the method derives from the fact that the degree of polarization of a fluorophore is inversely related to its molecular rotation, itself being largely driven by Brownian motion (Lea and Simeonov, 2011). Quantitatively, FP is defined as the difference of the emission light intensity parallel and perpendicular to the excitation light plane, normalized by the total fluorescence emission intensity (**Figure 2.8**). The polarization relates a fluorophore's lifetime with its rotational relaxation time; the latter is defined as the time it takes for a molecule to rotate through approximately 68.5° angle after excitation. The unit is recorded in milli polarization units (mP).

To assess the interaction between the Prli42 miniprotein and the NTD of RsbR proteins, the cytosolic tail of *L. monocytogenes* and *B. subtilis* Prli42 (**Table 2.9**) were synthesized by Severn Biotech™, with the TAMRA fluorophore ligated to the carboxy-terminus of the peptide. All of the proteins used in these experiments, *Lmo*RsbR full-length, *Lmo*N-RsbR and *Bsu*N-RsbR were purified as described above.

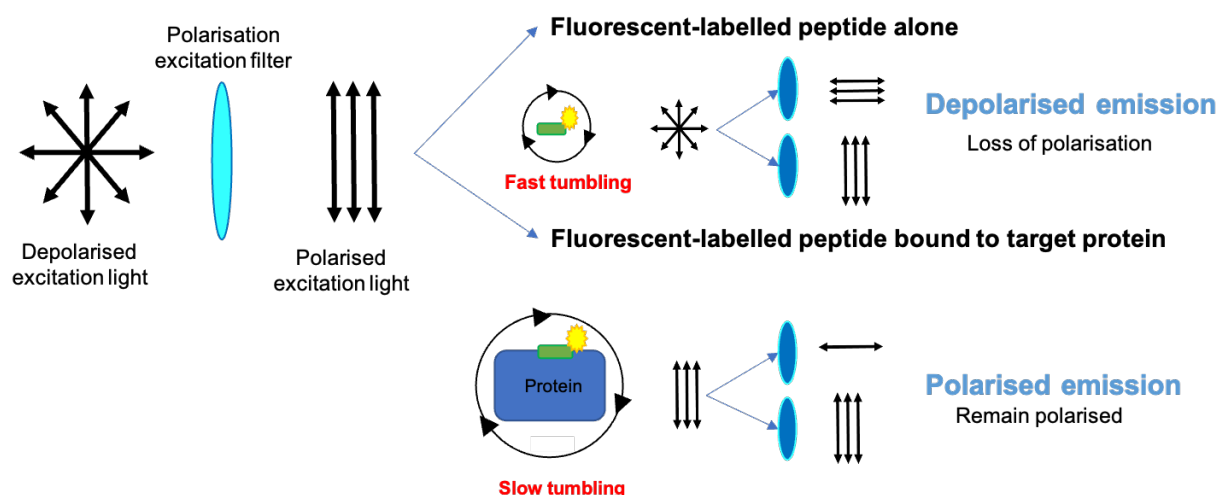


Figure 2.8. Fluorescence polarization method

Depolarized light is excited and passes through a polarization excitation filter; this polarized light will be depolarized by small, fluorescently-labelled peptides when they are free in solution due to fast Brownian movement. By contrast, when there is an interaction between the fluorescently-labelled peptide with a larger entity, such as a protein, the fluorophore follows a slower Brownian movement, which does not depolarize light.

Table 2.9. Prli42 cytosolic tail sequence

<i>Lmo</i> Prli42	MTNKKVV-TAMRA
<i>Bsu</i> Prli42	MMSQKLMK-TAMRA

In these experiments, the initial concentration of protein was 1000 μ M for the N-RsbR proteins and 300 μ M for the RsbR FL proteins, which was the highest concentration used. The protein was diluted by half, with a serial dilution up to eight times, keeping a constant concentration of the fluorolabelled peptide (40 nM and 40 μ M) in a final buffer of 25 mM Hepes-NaOH pH8, 150mM NaCl and 0.1% Triton-100X. A positive control experiment was performed using GpsB with its known interacting peptide, PBP1 (Cleverley et al., 2019) modified with the TAMRA fluorophore. The requirement for a validated positive control is because there are no known interaction partners for RsbRs or Prli42. Negative controls were also performed using the Prli42 peptide with the unrelated protein GpsB. The solutions were dispensed in triplicate into a 384-well black opaque and round bottom microplate which was

read with a PHERAstar (BMG Labtech) plate reader using the 540 nm excitation and 590 nm emission filters for TAMRA. The fluorophore labelled peptide alone was used as the background control. The readings were performed with 200 flashes per cell at a focal height of 11.3 mm. The data were plotted using the average FP in mP of the triplicates for each condition as a function of protein concentration in Graphpad. The polarization measurements were corrected by subtracting the background contribution for each measurement. The data points were fitted using a non-linear regression hyperbola type equation $y = \frac{B_{\max} \times X}{K_d + X}$, where B_{\max} is the maximum specific binding, X the protein concentration and K_d the equilibrium dissociation constant. the fitted value follows a hyperbolic curve reaching saturation, then the experiment shows an interaction of the peptide with the protein, but if the fitted data remains a straight line or a flat line, then the experiments suggest that there is no interaction between the peptide and the protein.

2.9. X-ray crystallography

2.9.1. Crystallization screening and optimisation

The purified recombinant protein samples were concentrated with a centrifugal concentrator with the appropriate nominal molecular weight cut-off value to a minimum protein concentration of 10 mg/mL. Proteins were screened for crystallisation with commercial crystallization screens: Morpheus I, PACT, Index, Structure (Molecular Dimension) and JCSG+ (Hampton Research). Sitting-drop vapour diffusion was used for the initial crystallisation screening (**Figure 2.9A**).

Eighty microliters of the screen conditions were dispensed in the reservoirs of MRC 2-well plates using a multi-channel pipette (StarLab). The MRC 2-well plates allow the screening of 96 conditions with two crystallisation experiments in parallel using different protein:crystallant ratios: 1:1 and 2:1. The crystallization drops were set up using a Mosquito pipetting robot (TTP Labtech) using a volume of 100 nl. The plates were then sealed with ClearVue Sheets (Molecular Dimension) and stored at a constant 20 °C. Crystallisation conditions yielding high quality crystals in the MRC plates were optimised using hanging-drop vapour diffusion in 24-well Linbro plates with a reservoir volume of 500 µL. The hanging-drop vapour diffusion method was used for optimisation screens (**Figure 2.9B**). The different component parameters of the initial crystallization condition were systematically modified, such as the pH, the salt, or precipitant concentration. The drops were set up manually on glass coverslips with the same protein:crystallant ratios used for crystallisation screening, with a

volume of 1 μL . The cover slips were then sealed with silicone vacuum grease and the plates were stored at 20 $^{\circ}\text{C}$.

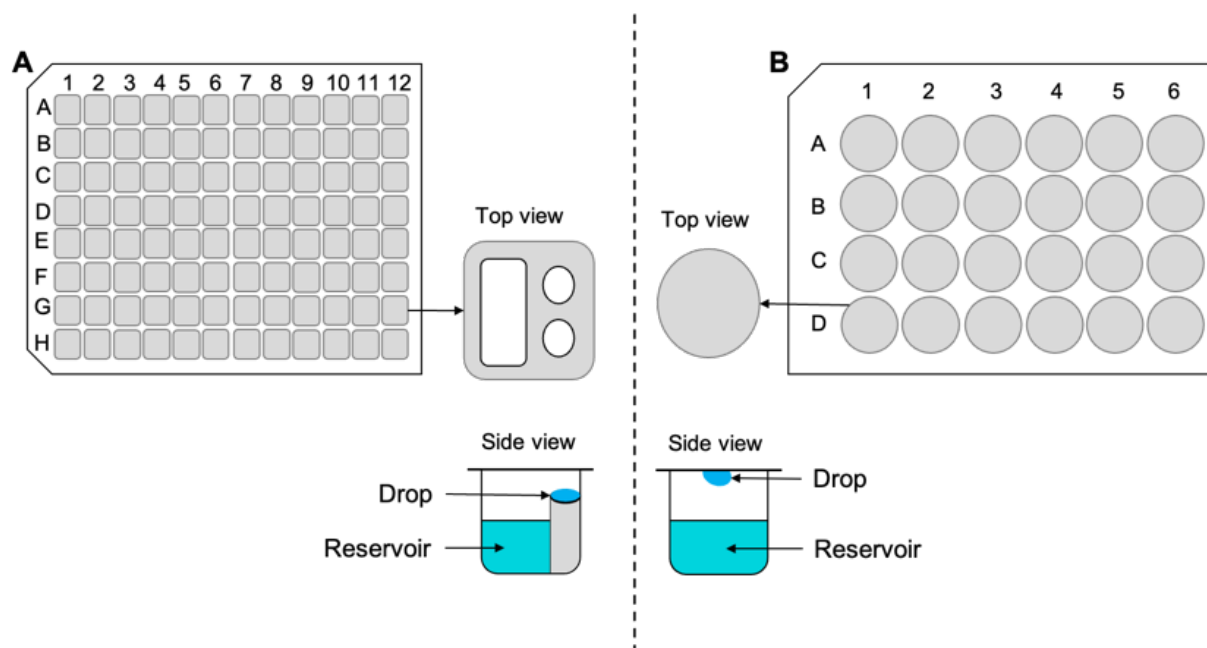


Figure 2.9. The vapour diffusion crystallisation method

A. Sitting drop method was used with MRC-2 96 well plates which allowed to set up two crystallisation experiment in parallel. In this case, the drop sits on the well in the plate. **B.** Hanging drop method was used on 24-well plates, where the drop hangs upside down on the cover slip.

The crystallisation drops were regularly checked using light microscopy (Leica) for up to 2 months, and resulting crystals were harvested for diffraction experiments. Crystals were harvested using a nylon cryoloop (Hampton Research) with a size (0.05-1mm) matched to the crystal dimensions. The harvested crystals were soaked in cryoprotectant solution to avoid the formation of hexagonal water ice crystals and crystalline ice. Crystals from conditions containing PEGs were usually cryoprotected in 20% (v/v) PEG400. Crystals obtained in conditions containing sodium formate were soaked in 6 M sodium formate. The cryoprotected crystals were picked up with loops and quickly flash-cooled in liquid nitrogen to avoid icing. The mounted crystals were stored in cryo-vials in liquid nitrogen until data collection.

2.9.2. Crystal diffraction and data collection

Harvested crystals were screened for diffraction at 100 K on the in-house X-ray generator, the MetalJet (Bruker), equipped with a CMOS detector (PHOTON 100 and PHOTON 3). The

crystals were exposed to the X-ray beam at two phi angles separated by 90° for between 2 to 60 seconds. Crystals diffracting to at least 4 Å were retained for data collection at the Diamond Light Source (DLS), using Pilatus detectors (beamline I03, Pilatus3 6M; beamline I04, Pilatus 6M-F). For molecular replacement, 200° of diffraction data were collected at 100 K, with an exposure time of 0.1 seconds at 10% transmission of the X-ray beam. For experimental phasing using selenium as the anomalous scatterer, the data were also collected at 100 K, but the X-ray wavelength for data collection was set to experimentally determined anomalous scatterer absorption edges (**Figure 2.10**) and 999° degrees of data were collected using the tuneable beamlines at the DLS synchrotron. Selenomethionine is a derivative of methionine with a selenium atom instead of a sulphur atom. Selenium has 34 electrons, which makes it a strong scatterer, this ‘anomalous’ scattering can be used to solve the phase problem in *de novo* structure solution. When the natively present sulphur atoms were exploited as anomalous scatterers, the data collection was performed *in vacuo* at the long wavelength beamline I23 at the DLS synchrotron. The datasets were collected by beamline scientists Armin Wagner, Ramona Duman, Ksamel El Omari and Vitaliy Mykhaylyk and the dataset obtained is discussed in **Chapter III**.

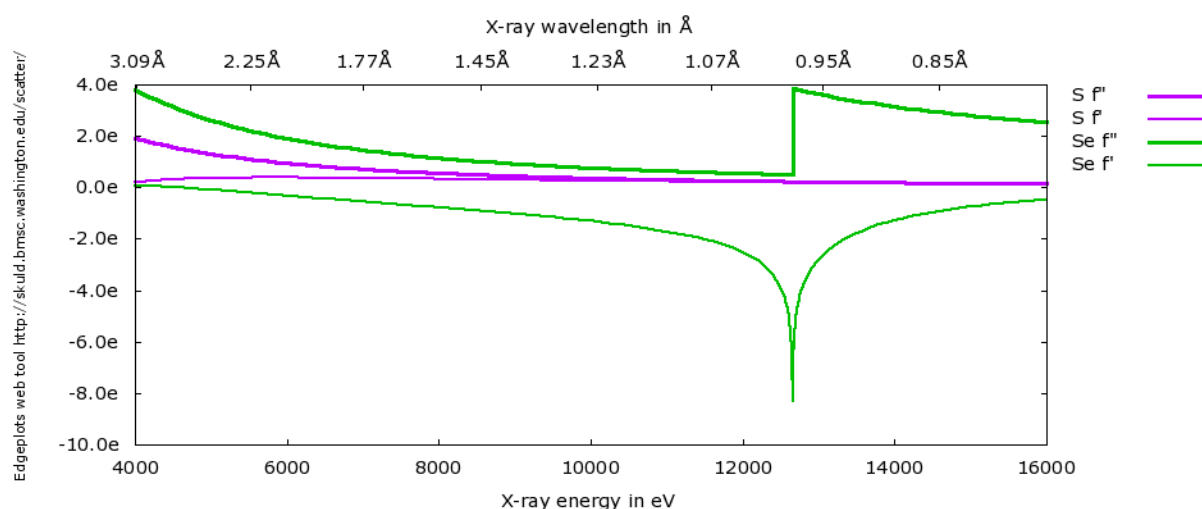


Figure 2.10: Theoretical selenium and sulphur X-ray absorption edge

For Se-SAD experimental phasing, the X-ray beam used was 0.98 Å which is close to the peak of the selenium X-ray absorption edge (green). For native S-SAD, the X-ray beam used was 2.75 Å. The plot was generated using the Biomolecular Structure centre webserver (http://www.bmsc.washington.edu/scatter/AS_form.html).

2.9.3. X-ray data processing and structure determination

Initial data processing

Diffraction data were initially automatically processed (integrated and scaled) at the DLS synchrotron using different processing pipelines in Xia2 (Winter, 2010). Two other programs were also used to integrate and scale the diffraction data, DIALS (Beilsten-Edmands et al., 2020) and XDS (Kabsch, 2010a). The output from the different pipelines were compared to determine the optimal data processing strategy. The integrated and scaled X-ray datasets were merged using Aimless (Evans, 2011). The quality and the resolution of the datasets were assessed with reference to the completeness ($>95\%$), I/σ (>1.5) and the half-set correlation coefficient $CC_{1/2}$ (>0.5) (Evans and Murshudov, 2013). The best datasets were taken forward for structure determination and model building using CCP4i2 (Potterton et al., 2018).

Molecular replacement

Molecular replacement permits structure determination using homologous models or models sharing at least 25% sequence identity. Two homologous models were used for molecular replacement: *B. subtilis* N-RsbRA (PDB ID: 2BNL (Murray et al., 2005)) and *M. thermoacetica* N-MtR (PDB ID: 3ZTA (Quin et al., 2012)). The structures were solved using the Phaser molecular replacement pipeline, and the resultant models were refined using REFMAC5 (Murshudov et al., 2011).

Experimental phasing

In contrast to molecular replacement where initial phases are obtained from the model used to solve the structure experimental phasing uses heavy atoms to determine the dataset's correct hand of phase set. The images were scaled and integrated as above and an initial heavy atom position search was performed using the SHELX CDE suite (Sheldrick, 2008). Input to this suite includes the number of monomers in the unit cell, the number of heavy atoms and the heavy atom used. SHELX C sets up the files required for the following steps: SHELX D (Schneider and Sheldrick, 2002) was used for heavy atom location, in order to obtain the substructure positions, and SHELX E (Sheldrick, 2002) was used for phasing and density modification to determine the correct hand of the dataset. The processed data were then used with the substructure positions for *de novo* structure determination.

Model building and refinement

The processed datasets were used for model building initially with the Buccaneer automated model building software (Cowtan, 2006). Buccaneer allows the protein backbone to be traced within the electron density by applying a likelihood-based density target. Once the mainchain is built, the side chains are added with the provided protein sequence for guidance. The initial model was then inspected manually in COOT (Emsley and Cowtan, 2004), residue by residue, making sure that each fitted the electron density and were in their most probable conformations. The built models were refined using REFMAC5 in an iterative process with COOT until convergence of the refinement parameters.

Model validation

The structural models were subjected to the validation pipeline in COOT, which includes (i) the Ramachandran plot, which plots the peptide backbone torsional angles phi and psi for each residue and, therefore, the outliers can be identified easily and corrected based on electron density fit; (ii) the unmodelled blob validation tool ensures the electron density is filled with appropriate molecules; (iii) the added water molecules are checked to correct any automatically-built but misplaced waters; (iv) the stereochemical and geometry parameters of the built model was checked, including sidechain rotamer analysis; (v) B-factor analysis was also performed to make sure that the model made chemical sense. Finally, the Molprobity webserver (Chen et al., 2010; Davis et al., 2007; Williams et al., 2018) was also used to check and validate dihedral-angles, atom contacts, clashes, rotamer analyses. All these validation and model quality controls are required in any event for submitting to the PDB and to produce the protein model that best describes the collected data.

Structure comparisons

The newly determined structures were compared with each other and with models deposited in the PDB. Structure superimpositions were performed using the SSM superimpose tool and the RMSD C α (Root Mean Square Deviation) values were recorded as a measure of structural similarity in COOT. PYMOL was used for structure visualisation and the preparation of figures.

2.10. Electron cryo-microscopy and single particle analysis of *L. monocytogenes* complexes

Electron cryo-microscopy (cryo-EM) can be used in concert with single particle analysis to determine the structures of macromolecular complexes. In our case, we used cryo-EM single particle analysis for the *L. monocytogenes* RsbR-RsbS (*LmoRS*) and RsbR-RsbS-RsbX (*LmoRSX*) complexes.

2.10.1. Negative stain control

Purified (*LmoRS* and *LmoRSX*) complexes were assessed by negative stain TEM. Carbon coated 300 mesh copper grids (Gilder Grids) were glow-discharged using the easiGlow glow-discharger (PELCO) for 40 seconds. Ten microliters of each sample at final concentrations of 0.3 and 0.5 mg / mL were used for negative stain, followed by 3 water washes, with excess water blotted between each wash (Whatman – 50 grade paper). The grid was stained using 1% uranyl acetate; excess stain was blotted, and the grid was dried under a desk lamp. The grids were visualized using a Hitachi HT7800 120 kV 100 kV, equipped with an EMSIS CMOS Xarosa camera. Several images were recorded to obtain initial 2D classes from negative stains grids.

2.10.2. Negative stain 2D classes

The negative stain images were processed using the EMAN software suite (Tang et al., 2007). One micrograph was used to pick a background reference, and good and bad particles for the automated picking. All automatically picked particles were manually checked and classed into 50 2D classes.

2.10.3. Cryo-EM grid preparation and data collection

Cryo-EM grid screening and data collection was performed at the Leeds Astbury Biostructure Laboratory facility. Three microliters of sample at final concentrations of 3.2 and 5 mg/mL were used to set up 1.2/1.3 Quantifoil grids. The grids were set at 95% humidity and blotted for 6 seconds with a blot force of 6 in the Vitrobot plunger (FEI). The grids were screened on a TITAN KRIOS (ThermoFisher) microscope using the Falcon 3EC (FEI) direct electron detector. The grids were stored on site for data collection of 72 hours for *LmoRS* and for 48 hours for *LmoRSX*.

2.10.4. Data processing

Cryo-EM data processing was performed using the Relion 3.1 software pipeline (Fernandez-Leiro and Scheres, 2017; Scheres, 2016; Zivanov et al., 2018, 2019a) using computer capabilities within the Newcastle SBL and the Rocket HPC. The data collection parameters can be found in **Table 2.10** for the *LmoRS* and **Table 2.11** for the *LmoRSX* complex.

Table 2.10: Data collection parameters of *LmoRS* sample

Hardware	
Microscope	Krios 1
Detector (mode)	F3 (linear)
Accelerating voltage (keV)	300
Pixel size (Å)	1.065
Data acquisition parameters	
Nominal magnification	75000x
Spot size	4
Illuminated area	1.2 µm
Total no. of collected micrographs	
Dose	
Square pixel (Å ²)	1.13
Dose per physical pixel per second	57.5
Dose per Å ² /sec	50.7
Exposure time (seconds)	1.2
Total dose (e/Å ²)	60.8
Number of fractions	47
Dose per fraction (e/Å ²)	1.3
EPU parameters	
Defocus range (-µm)	-0.7 to -3.1, 0.3 increment
Autofocus	Every 10 µm using objective
Drift measurement	0.05 nm/s (once a grid square)
Delay after stage shift	5s
Delay after image shift	5s
Apertures (size in microns)	
C1	2000
C2	70
C3	2000
Objective	100

Table 2.11: Data collection parameters of *Lmo*RSX sample

Hardware	
Microscope	Krios 1
Detector (mode)	F3 (linear)
Accelerating voltage (keV)	300
Pixel size (Å)	1.065
Data acquisition parameters	
Nominal magnification	75000x
Spot size	4
Illuminated area	1.33 μm
Total no. of collected micrographs	
Dose	
Square pixel (\AA^2)	1.13
Dose per physical pixel per second	48.3
Dose per $\text{\AA}^2/\text{sec}$	42.7
Exposure time (seconds)	1.4
Total dose ($\text{e}/\text{\AA}^2$)	60
Number of fractions	45
Dose per fraction ($\text{e}/\text{\AA}^2$)	1.33
EPU parameters	
Defocus range ($-\mu\text{m}$)	-0.7 to -3.1, 0.3 increment
Autofocus	Every 10 μm using objective
Drift measurement	0.05 nm/s (once a grid square)
Delay after stage shift	5s
Delay after image shift	5s
Apertures (size in microns)	
C1	2000
C2	70
C3	2000
Objective	100

The collected micrographs were pre-gain corrected and were initially imported and corrected for beam-induced motion using the MotionCorr task with the Falcon III MTF file in electron-counting mode at 300kV. The Contrast Transfer Function (CTF) was determined using the ctffind plugin (Rohou and Grigorieff, 2015). The corrected micrographs were used to manually pick ~1000 stressosome particles with a box size of 480 Å. An initial set of 50 2D classes were generated with the manually picked particles. The 2D classification was obtained in twenty-five iterations using a mask diameter size of 384 Å, with an in-plane angular rotation

of the psi angle sampling of 6°. Good 2D classes, with well-defined protein features, were selected and used as a reference for automatic particle picking.

The automatically picked particles were extracted and scaled to 196 Å, for faster computation with a new pixel size of 2.17 Å. The particles were classed in 100 2D classes. To exclude bad particles, five iterations of 2D classification were performed using the same parameters as the initial 2D classification. The final 2D classes were selected and used to determine an initial 3D model. The initial model was built with one class in C1 symmetry, to avoid symmetry bias, with an angular sampling of 15°. Fifty initial, 200 in-between, and 50 final iterations were performed within the initial model building task. For faster computation, the particles were pooled in batches of 40.

To assess sample heterogeneity five 3D classes were calculated using the initial 3D model as reference. The 3D classification was done with 25 iterations with an angular sampling of 7.5°, using an initial low-pass filter of 60 Å. The 3D class containing the most particles with well-defined protein features was selected for further processing.

The particles of the selected 3D class were then re-extracted in their full-size to achieve the highest possible resolution for the dataset. In the first instance, the 3D class was refined in C1 symmetry using the initial 3D model as a reference map. Subsequently, the data were processed in D2 symmetry using the *B. subtilis* stressosome map (EMD-1555) as a reference map. Masks were created for each map using a threshold value determined in Chimera to include all the protein density, using a lowpass map at 15 Å. Each map was post-processed to mask, sharpen the map and to calculate gold-standard FSC curves using these masks.

2.10.5. Model building

The maps obtained were visualised and studied in Chimera (Pettersen et al., 2004). Model building was done using homology models for RsbR and RsbS STAS domains that were generated on the Phyre2 web server (Kelley et al., 2015). The N-RsbR1 crystal structure solved during this project was used for modelling the RsbR1 NTD. A full length RsbR1 model was generated using the *LmoRS* map. The RsbS and RsbR1 models were fitted manually and placed in the density with the Fit in Map tool in Chimera. Further model building was done by manually refining the built model using Coot (Emsley and Cowtan, 2004) and real-space refined using the Phenix pipeline (Liebschner et al., 2019).

Chapter III: Structural studies
of the stress sensory domains
of the *L. monocytogenes*
stressosome

3.1. Introduction

The N-terminal domains (NTDs) of RsbR (N-RsbR) proteins have been suggested to act as the stress sensors that activate the stressosome (Akbar et al., 1997; Murray et al., 2005). The signalling cascade pathway begins with the RsbT-catalysed phosphorylation of Ser59 on RsbS, and Thr171 and Thr205 on RsbR. RsbT is released from the stressosome once these phosphorylation reactions have been carried out to trigger the partner switching cascade that leads to the activation of σ^B (Boylan et al., 1993a). However, the sensing mechanism of the stressosome - leading to the activation of RsbT - remains unclear.

Before starting this project, two N-RsbR structures had been determined: N-RsbRA from *B. subtilis* and N-MtR from *M. thermoacetica* (**Chapter I, Figure 1.9**). Even with the molecular knowledge gleaned from these structures, the mechanism(s) of the sensory activity could not be determined. As described previously (**Chapter I, section 1.4.2**), the pairwise and overall sequence identities of the RsbR NTDs are very low between homologues and paralogues, yet the two N-RsbR models available share similar structural features: they both adopt a globin-like fold composed of 5 α -helices and 4 loops, and they dimerise with the primary dimer interface formed along the extended α -helix at the C-terminus of the domain. The structures show similarity to the classical heme-binding globin fold, but they lack the proximal histidine residue, which coordinates the essential iron within the heme cofactor in true haemoglobins (Tejero and Gladwin, 2014) (**Figure 3.1**).

To investigate the structural features of the N-RsbR domain further, this study focused on the RsbR proteins from *L. monocytogenes*. At the time of writing, no crystal structures of any of the four RsbR proteins from *L. monocytogenes* structures had been determined. The interest in RsbR proteins from this organism is also motivated by a desire to interrogate any structural differences in the stress sensor domain in a human pathogen, compared to the model organism *B. subtilis*, as part of the overall consortium's research activities.

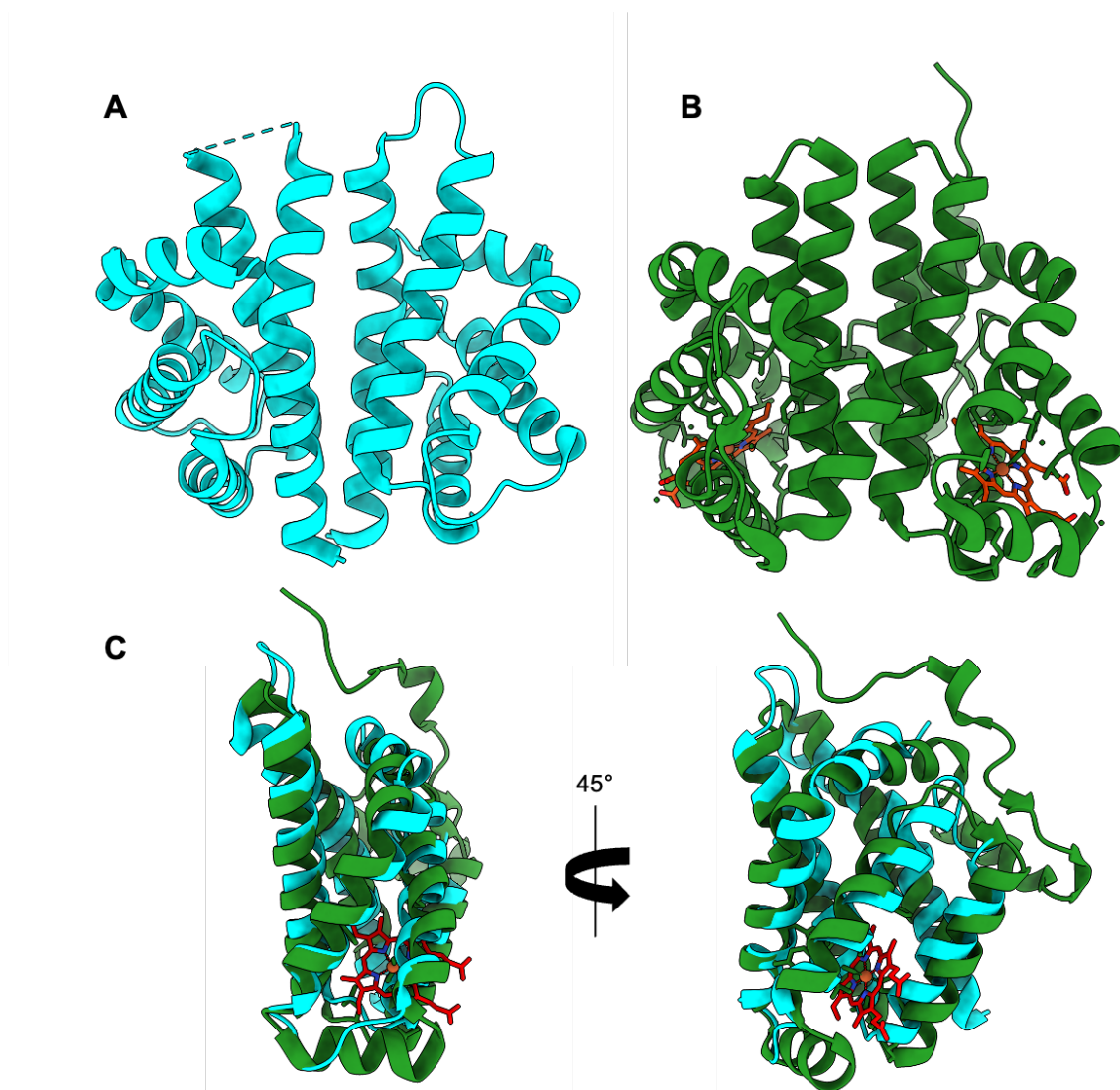


Figure 3.1. Non-heme globin fold of N-RsbR and globin fold HemAT structure comparison

The model in cyan represents N-RsbRA (**A**: PDB ID 2BNL) and in dark green the HemAT globin domain (**B**: PDB ID 1OR4), with the heme shown in red. Both structures share a similar structure fold (**C**) with an RMSD of 3.06Å and 9.4% sequence identity, with the exception that N-RsbRA cannot bind a heme.

3.2. Purification of recombinant *L. monocytogenes* N-RsbR proteins

The reading frames for all the *L. monocytogenes* RsbR NTDs proteins were cloned into the pET28gg expression vector using the CIDAR MoClo cloning method (**Chapter II, section 2.1.1**). After their successful cloning, verified by DNA sequencing of the inserted *orf*, the proteins were expressed and purified for further structural and biophysical studies. The proteins were not expressed as genetic fusions to affinity-tags and were purified by anion exchange chromatography using a Q Sepharose column, as the calculated pI of the RsbR NTDs ranges between 4.6 and 5.8 (**Chapter II, section 2.3.1**). Following anion exchange chromatography,

size-exclusion chromatography (SEC) was performed using 2 different columns, S75 or S200 (depending on the sample size and availability of the columns), in order to further separate the protein of interest from contaminating proteins after the anion exchange step (**Chapter II, section 2.3.2**). Most of these proteins underwent a second anion exchange chromatography run, using a MonoQ column, to obtain the purest sample possible. Following the second anion-exchange step, a final SEC was performed for each of the proteins. With each purification step an SDS-PAGE was run to estimate the quality and the quantity of the purified sample. For clarity, only the final purification step will be displayed for each protein.

3.2.1. N-RsbR1 purification

RsbR1, or *lmo0799*, is the paralogue that is found on the large σ^B operon, and its NTD was the first of the paralogues to be purified. After anion exchange chromatography (**Supplemental figures S1**), the last purification step of N-RsbR1 was the SEC shown in **Figure 3.2**. The N-RsbR1 SEC chromatogram yields a symmetric elution peak with an elution volume of 71.9 mL from an S75 16/60 column. The SDS-PAGE of the N-RsbR1 size-exclusion fractions show a highly pure and concentrated sample.

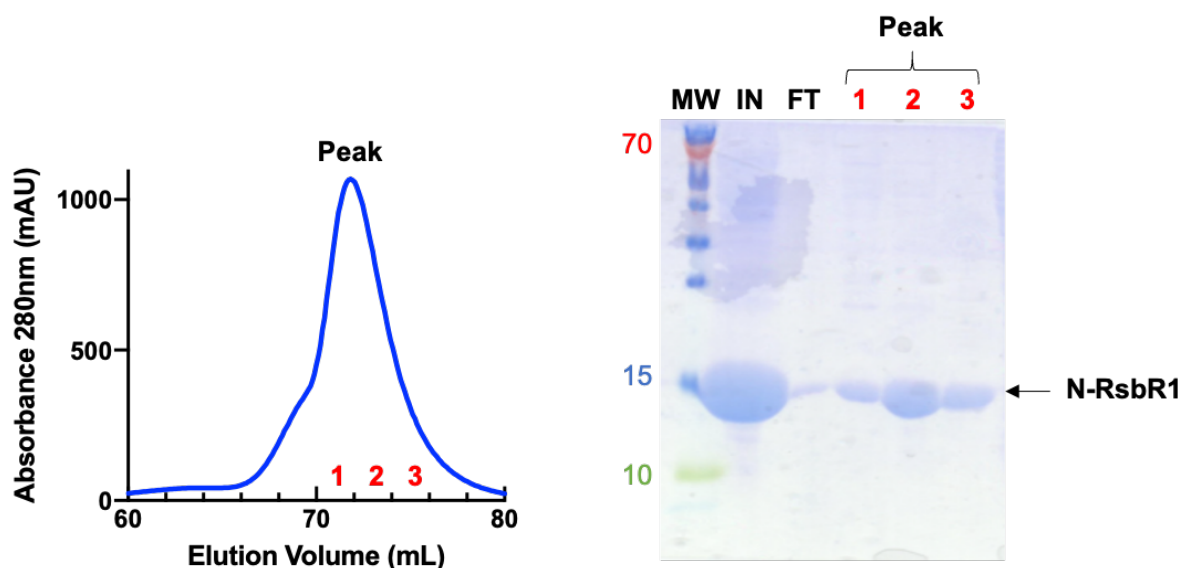


Figure 3.2. N-RsbR1 protein purification

S75 size-exclusion chromatogram of N-RsbR1 with the collected fraction in red (fractions volume: 2 mL). The 15% SDS-PAGE with a stained protein ladder (MW, in kDa) displays the input of the SEC (IN) and the flow-through (FT) during the sample concentration. Fractions 1-3 of the elution peak were loaded onto the gel and the peak fractions of N-RsbR1 (MW ~15 kDa) show an electrophoretic purity in excess of 95 %.

3.2.2. N-RsbR2 purification

The second RsbR paralogue to be purified was RsbR2 or *lmo0161*. After the anion affinity exchange chromatography (**Supplemental figures S2**), the SEC of the N-RsbR2 recombinant protein purification, shown in **Figure 3.3**, showed a more contaminated sample from the purification protocol in comparison to that of N-RsbR1. The first peak from the SEC is consistent with relatively high-molecular weight proteins, which represent the main contaminants in this protein preparation as seen on the SDS-PAGE (**Figure 3.3**). The second peak is N-RsbR2, which eluted at 89.2mL from the S200 16/60 column. The N-RsbR2 purification allowed a clean protein sample to be obtained when the first fraction of the second peak was excluded from the final pool as this fraction contains some contaminants.

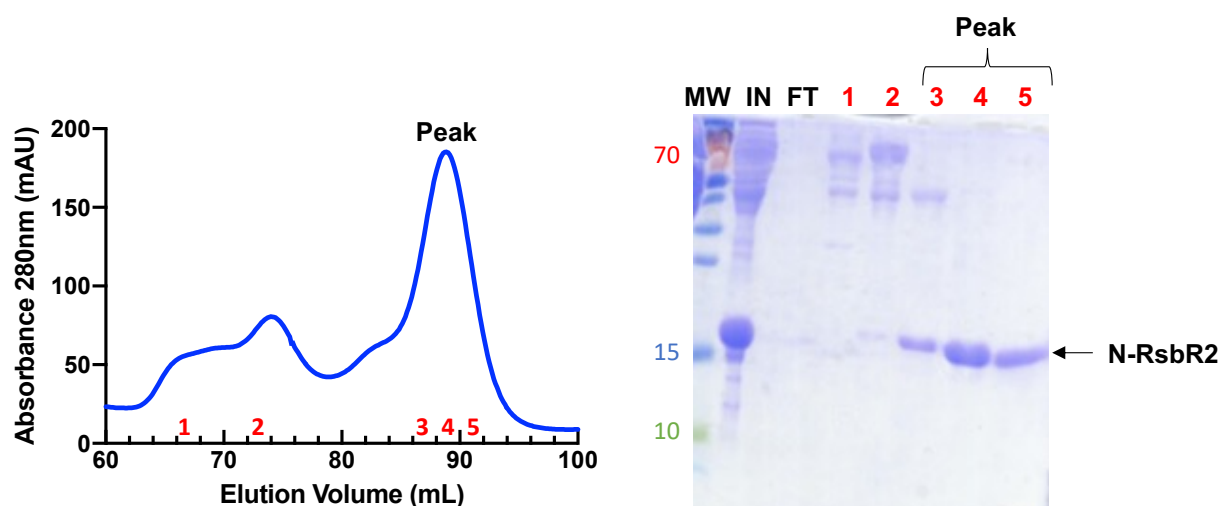


Figure 3.3. N-RsbR2 protein purification

The S200 size-exclusion chromatogram of N-RsbR2 shows 2 peaks: the first peak has high MW contaminants and N-RsbR2 is found predominantly in the second peak, fractions 3-5. The 15% SDS-PAGE with a stained protein ladder (MW) displays the input of the SEC (IN) and the flow-through (FT) during the sample concentration. The fractions of the elution peaks were loaded onto the gel and peak fractions of N-RsbR2 (MW ~15 kDa) has electrophoretic purity of ~95%.

3.2.3. N-RsbR3 purification

The third *L. monocytogenes* RsbR paralogue is RsbR3 or *lmo1642*. Following anion exchange affinity chromatography (**Supplemental figures S3**), the N-RsbR3 purification worked well as can be seen on the last SEC chromatogram (**Figure 3.4**) where the N-RsbR3 protein eluted with a clearly symmetric peak. The elution volume of the sample is 72.07 mL using the S75 16/60 column. The protein sample was highly pure on the SDS-PAGE, of which the first fraction from the SEC was excluded.

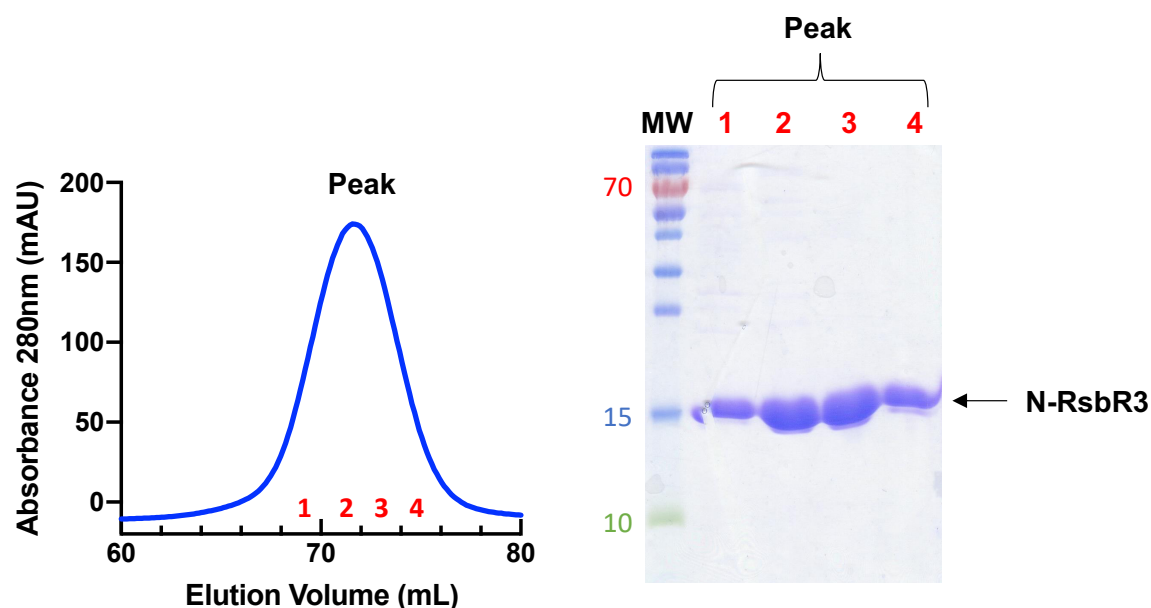


Figure 3.4. N-RsbR3 protein purification

Final size-exclusion chromatogram of N-RsbR3 using a S75 16/60 column, with the collected fractions annotated in red (2 mL volume). The 15% SDS-PAGE with a stained protein ladder (MW in kDa) displays the fractions of the elution peak in which N-RsbR3 protein migrates to approximately of 15 kDa.

3.2.4. N-RsbR4 purification

The last RsbR paralogue to be studied was RsbR4 or *lmo1842*. N-RsbR4 was the most difficult to purify to the quality and yields demanded by crystallography. The N-RsbR4 purification required additional purification steps as the sample obtained after anion exchange (**Supplemental figures S4**) and SEC was too contaminated to be useful. In a first attempt to purify N-RsbR4, the final protein sample was very contaminated with a low yield of N-RsbR4 (**Figure 3.5A**). The chromatogram looked highly heterogeneous and a clear peak representing the N-RsbR4 protein was not obtained. Therefore, a second attempt at the purification using a bigger volume of cell culture was made using the same purification protocol, in order to obtain sufficient sample for downstream processes. With the second attempt, a better N-RsbR4 protein sample was obtained (**Figure 3.5B**), and from the final SEC step had a V_e of 85.4 mL. Other expression strains such as Rosetta and T7 express were tested, however they did not result in better expression. The yield of the N-RsbR4 protein sample obtained was just enough for protein crystallisation and biophysical experiments.

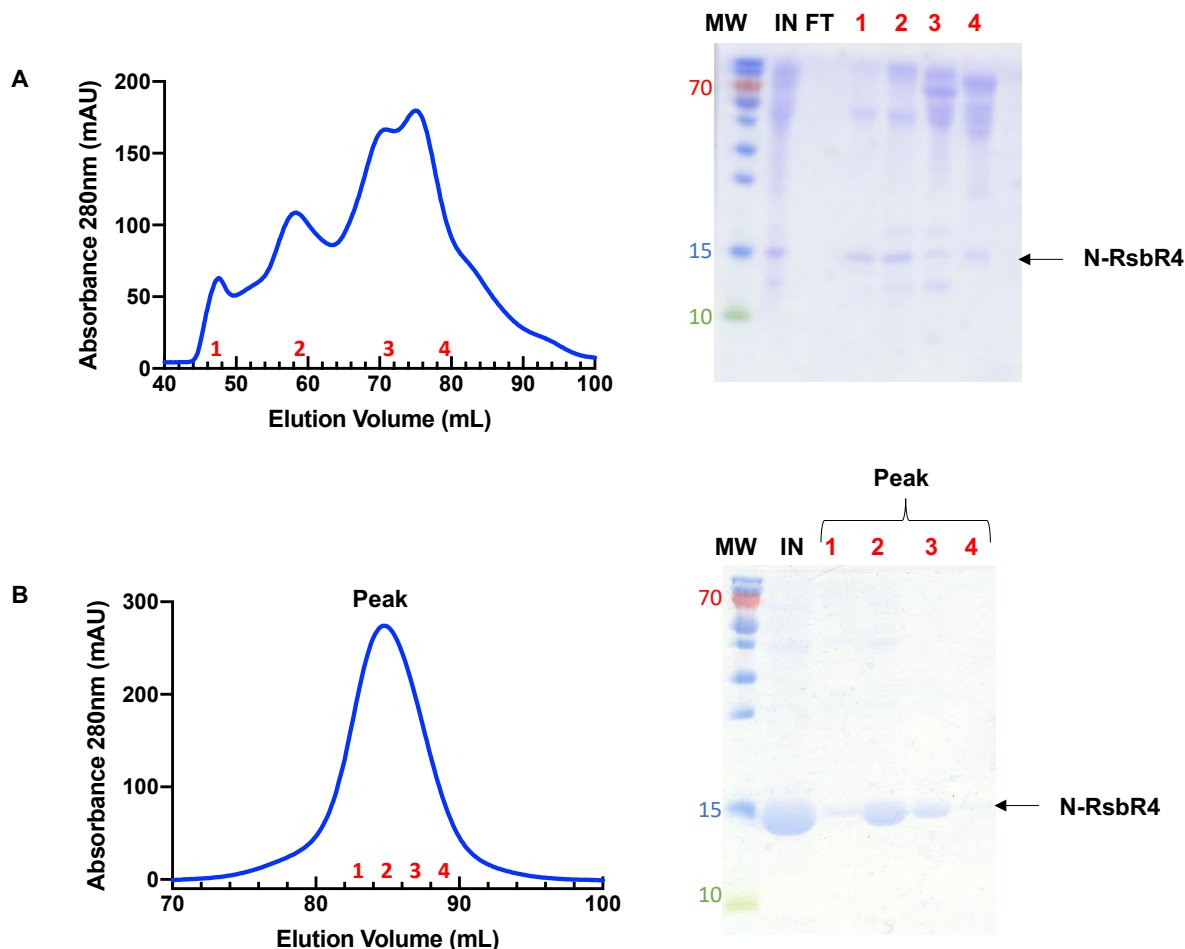


Figure 3.5. N-RsbR4 protein purifications

A. Left: last step S200 size-exclusion chromatogram for N-RsbR4 shows several peaks suggesting that the sample is highly contaminated. Right: the 15% SDS-PAGE of N-RsbR4 size exclusion peaks confirms the N-RsbR4 sample was highly contaminated and present only at low yield. **B.** After using a bigger cell culture volume for N-RsbR4 expression, the SEC suggests a better yield (left), displaying a single mostly symmetric peak but a low yield was retrieved as shown in the 15% SDS-PAGE (right).

3.3. Folding of the purified proteins in solution

Following the purification of the recombinant N-RsbRs, the protein samples were subjected to Circular Dichroism (CD) spectroscopy to analyse the secondary structure of the proteins to assess the stable folding of the proteins (**Chapter II, section 2.5**). The N-RsbR structures solved to date are composed only of α -helices, and the CD spectra from proteins dominated by α -helices usually have local minima at 208 and 222nm and a local maximum at 195nm. The CD spectra obtained for the N-RsbRs proteins suggested that all the proteins were folded and displayed features typical of proteins with a high proportion of α -helices (**Figure 3.6**) and were thus suitable for subsequent crystallisation screening and other

biochemical/biophysical analysis methods. Thermal melts using CD was also performed on each protein and the results are discussed in **Chapter 4, section 4.4.1**.

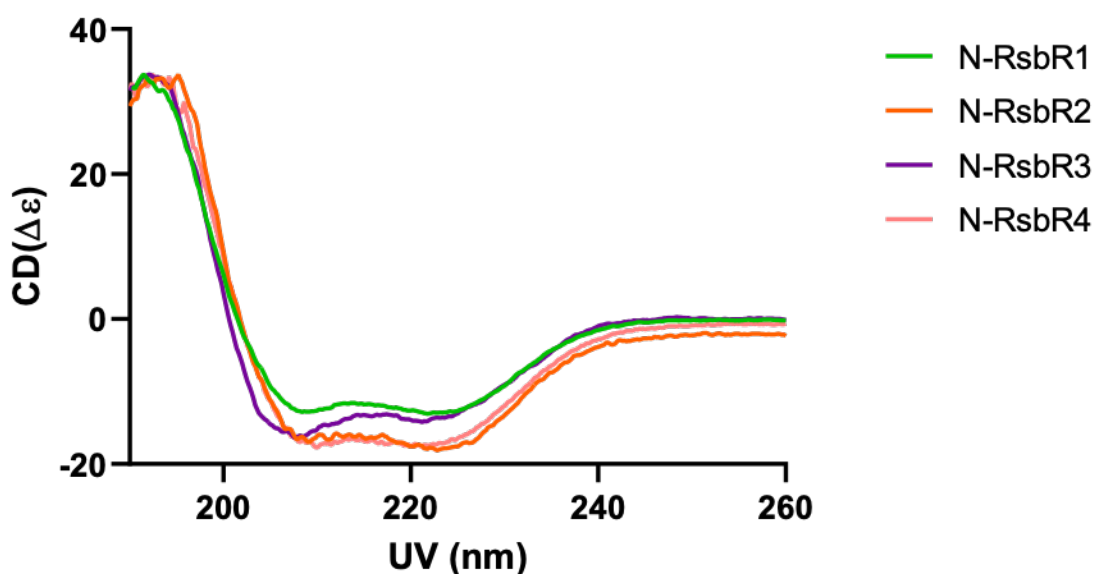


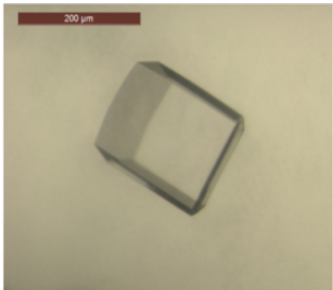
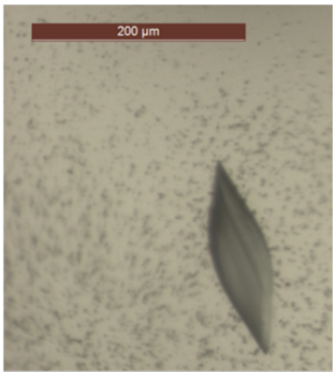
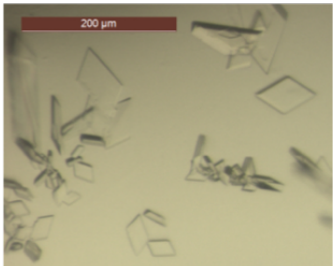
Figure 3.6. CD spectra of *L. monocytogenes* RsbR NTDs

The CD spectra shows that all the purified N-RsbR proteins are folded in solution and are mainly composed of α -helices

3.4. Crystallisation of N-RsbR proteins

All of the purified recombinant N-RsbR proteins were concentrated using Amicon filters with a MWCO of 3 kDa when using a fixed angle rotor as some of the proteins leaked to the flow-through when using a 10 kDa MWCO; a 10 kDa was used with a swinging bucket rotor. Crystallisation screening was set up at protein concentrations of 20.4 mg/mL for N-RsbR1, 13.5 mg/mL for N-RsbR2, 11.6 mg/mL for N-RsbR3 and at 11 mg/mL for N-RsbR4. In total, 480 crystallisation conditions were tested for each protein at 20°C (**Chapter II, section 2.9.1**). The initial crystallisation screening was set with the sitting drop method on MRC2 96 well plate, which allowed 2 crystallisations to be set up in parallel: on each plate the protein was crystallised at 1:1 and 2:1 protein:condition. The plates were incubated at 20°C and checked daily with a zoom stereo light microscope for a week. Crystals were obtained (**Table 3.1**) for N-RsbR1 in the Structure condition A6 (0.1M sodium acetate pH 4.6 and 8% w/v PEG 4000), N-RsbR2 in the PACT condition G6 (0.2M sodium formate, 0.1 Bis Tris propane pH 7.5 and 20% w/v PEG 3350) and N-RsbR3 in the PACT condition D9 (0.2M lithium chloride, 0.1M Tris pH8.0 and 20% w/v PEG 6000); by contrast no crystals were obtained for N-RsbR4, the screens for which led only to aggregates and heavy precipitates, which may be due to the low purity of the sample.

Table 3.1. Native protein crystals of *Lmo* N-RsbRs

Protein	Crystals	Condition
N-RsbR1		0.1M sodium acetate pH 4.6 8% w/v PEG4000
N-RsbR2		0.2M sodium formate 0.1M Bis-Tris-HCl Propane pH7.5 20% w/v PEG3350
N-RsbR3		0.2M lithium chloride 0.1M Tris-HCl pH8.0 20% w/v PEG6000

3.5. X-ray data collection and processing

Crystals for the three N-RsbR proteins were harvested using the appropriate loop size and 20% (v/v) PEG400 as a cryo-protectant, and flash-cooled in liquid nitrogen. Diffraction datasets from several crystals of each protein were collected at the Diamond Light Source Synchrotron (DLS) (**Chapter II, section 2.9.2**). The datasets were integrated using either the XDS or Dials software, followed by scaling and symmetry determination with the Aimless/Pointless programs and the data processing statistics for the best data sets are shown in **Table 3.2**.

The quality of a dataset is defined by parameters including but not limited to its resolution, the signal to noise ratio $I/(\sigma)I$, the correlation coefficient for a randomly selected half dataset $CC_{1/2}$ and its completeness (Evans and Murshudov, 2013). To build atomic models the resolution needs to be at least $<2.5\text{-}3\text{\AA}$, the $I/(\sigma)I$ for data in the highest resolution shell ≥ 1.5 , $CC_{1/2}$ for data in the highest resolution shell ≥ 0.5 and the completeness needs to be close

to 100% (Evans and Murshudov, 2013). The quality of the datasets summarised in **Table 3.2** have statistics good enough to be used for further analysis and structure determination by molecular replacement (MR).

Table 3.2. Data collection statistics for N-RsbR proteins

The values in parentheses correspond to the highest resolution outer shell

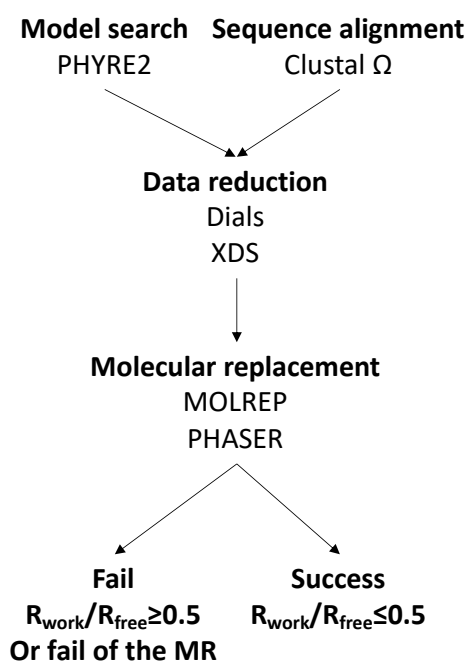
Data collection	N-RsbR1	N-RsbR2	N-RsbR3
Date	15/04/2018	15/02/2019	15/04/2018
Type of data collection	MR	MR	MR
Source	I04-1	I04	I04-1
Wavelength (Å)	0.911Å	0.916Å	0.916Å
Space group	P2 ₁ 2 ₁	P3 ₁ 21	P2 ₁ 2 ₁ 2 ₁
Cell dimensions			
a, b, c (Å)	31.92, 59.62, 141.37	55.39, 55.39, 98.71	49.80, 81.51, 82.78
α, β, γ (°)	90, 90, 90	90, 90, 120	90, 90, 90
Resolution (Å)	59.62 – 2.10 (2.16 - 2.10)	19.74 – 2.71 (2.84 - 2.71)	42.68 – 1.93 (1.98 - 1.93)
No. of measured reflections	106,122 (6,462)	52,660 (7,415)	186,088 (12,317)
No. of unique reflections	16,519 (1,317)	5,106 (671)	26,059 (1,720)
Multiplicity	6.4 (4.9)	10.3 (11.1)	7.1 (7.2)
CC _{1/2}	0.999 (0.87)	1.0 (0.81)	1.0 (0.66)
I/(σ)I	12.5 (2.0)	17.8 (1.5)	16.4 (1.7)
R _p im (all I+ and I-)	0.011 (0.318)	0.010 (0.446)	0.009 (0.507)
Completeness (%)	99.8 (99.9)	99.9 (100.0)	100.0 (100.0)
Anomalous multiplicity	3.3 (2.5)	5.6 (5.8)	3.7 (3.7)

Atomic models with a sequence identity of $\geq 25\%$, and sometimes as low as 20%, to the target structure are widely considered as the minimum for molecular replacement to be successful (Simpkin et al., 2019). There are two candidate models for solving the *L. monocytogenes* N-RsbR structures: 2BNL, the N-RsbRA from *B. subtilis* (Murray et al., 2005) and 3ZTA, the N-MtR from *M. thermoacetica* (*Mth*) (Quin et al., 2012). However, the sequence identity is lower than 25% in all cases (**Table 3.3**) with 2BNL and N-RsbR1 having the highest with 23.1%. Despite using a variety of molecular replacement software packages (PHASER, Molrep) and model preparation protocols (CHAINSAW, ensembles, poly-ala traces, Phyre2), the low sequence identity between the homologues and paralogues most likely led directly to the failure of the molecular replacement method (**Figure 3.7**).

Table 3.3. Sequence identity matrix between N-RsbR proteins

	N-RsbR3	N-RsbR4	3ZTA	N-RsbR2	2BNL	N-RsbR1	HemAT*
N-RsbR3							
N-RsbR4	21.09						
MtRsbRA	12.28	14.29					
N-RsbR2	18.94	10.00	16.13				
Bs-RsbRA	9.68	11.57	9.26	14.50			
N-RsbR1	13.14	14.07	10.66	20.28	23.13		
HemAT*	16.67	19.01	15.38	12.50	12.93	16.04	

*HemAT included as an exemplar bacterial globin

**Figure 3.7. Crystallographic data processing flow-chart.**

3.6. Selenomethionine-labelled protein purification

As the unmodified N-RsbR protein crystals did not support structure determination by molecular replacement, experimental phasing using selenomethionine (Se-Met) labelled proteins were used to determine their crystal structures *de novo*. The Se-Met labelled proteins were expressed using the methionine auxotrophic *E. coli* strain B834 (DE3) that was grown in a minimal medium supplemented with Se-Met (**Chapter II, section 2.2.3**). The Se-Met-labelled recombinant proteins were purified using the same methods as described previously

(**Chapter II, section 2.3**). Similar purification results were obtained with the Se-Met labelled proteins and only the final step of the protein purification by SEC results are displayed and commented upon for clarity's sake.

3.6.1. Se-Met N-RsbR1 purification

Following the anion exchange chromatography (**Supplemental figures S5**), the Se-Met N-RsbR1 protein purification (**Figure 3.8**) had a good elution peak with a SEC elution volume of 85.6 mL using the S200 16/60 column. The SDS-PAGE of the Se-Met N-RsbR1 SEC fractions show a highly pure and concentrated sample (**Figure 3.8**).

The incorporation of the Se-Met was verified by Liquid Chromatography Mass Spectrometry (LC-MS) (**Supplemental figures S6**). N-RsbR1 has 6 methionines (including the initiating methionine at position 1) and has a MW of 17922.3 Da. The Se-Met labelled N-RsbR1 has a MW of 18203.3 Da, as determined by LC-MS, a difference in mass of 281Da. Sulphur has an atomic mass of 32.1 Da and that of selenium is 78.96 Da and therefore there is a mass difference between them of 46.86 amu. The difference in observed mass by LC-MS for the proteins divided by the difference in mass between Se and S atoms yields an estimate of how successful Se-Met has replaced the natural sulphur-containing methionine. The answer in this instance is 5.997, and therefore it is safe to conclude that all the methionine residues were replaced by Se-Met.

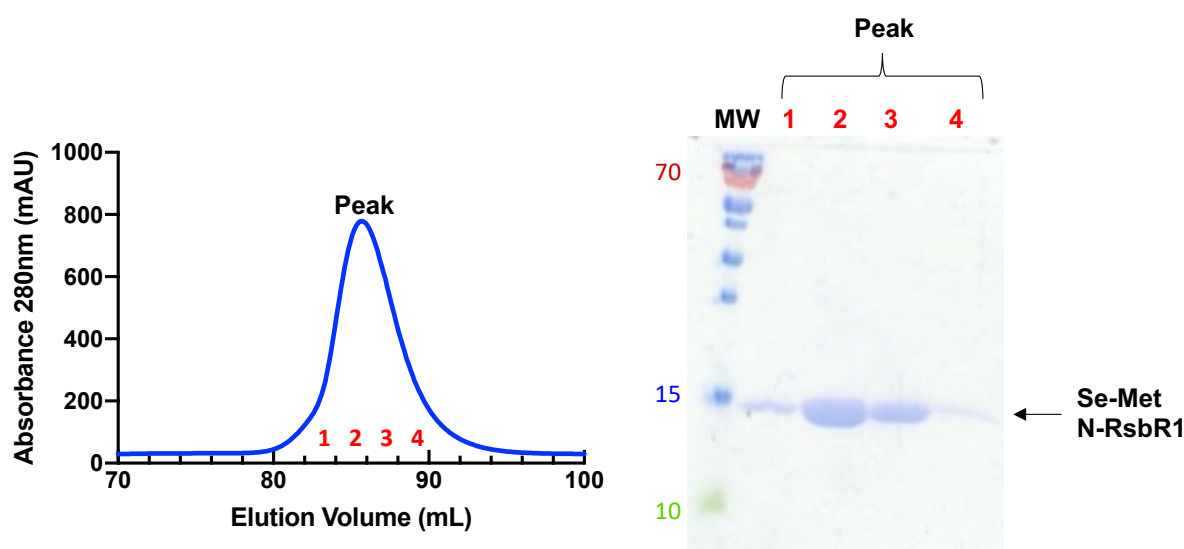


Figure 3.8. Selenomethionine-labelled N-RsbR1 protein purification

Size-exclusion chromatogram of the Se-Met N-RsbR1 recombinant protein, using a S200 16/60 column, with the collected fractions (2 mL volume) annotated in red. The 15% SDS-PAGE with a stained protein ladder (MW in kDa) displays fractions 1-4 of the elution peak.

3.6.2. Se-Met N-RsbR2 purification

The Se-Met labelled N-RsbR2 protein purification (**Figure 3.9, Supplemental figures S7**) presented a single peak compared to the unlabelled N-RsbR2 purification. The elution peak of N-RsbR2 was highly symmetric and has a V_e of 89.2mL using the S200 column. The purification of the Se-Met labelled N-RsbR2 yielded a highly pure sample, excluding the first 2 fractions for higher purity.

The incorporation of the Se-Met was also verified by LC-MS (**Supplemental figures S8**). In the case of N-RsbR2, which has 7 methionines (including the initiating methionine at position 1), the measured unlabelled protein MW is 18961.53Da, and that of the Se-Met labelled protein was 19290.81Da: the corresponding MW difference is 329.28Da. Following the same principle as above, the degree of Se-Met incorporation is 7.027 and, therefore, all 7 methionines were replaced by Se-Met.

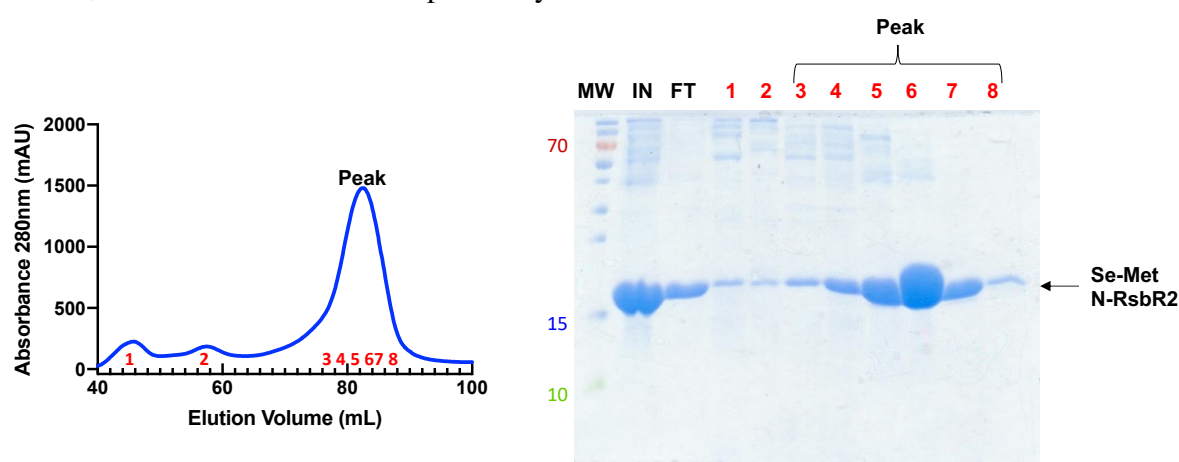


Figure 3.9. Selenomethionine-labelled N-RsbR2 protein purification

Size-exclusion chromatogram of Se-Met labelled N-RsbR2 using a S200 16/60 column. There are three peaks and N-RsbR2 is found in peak 3, with the fractions numbered in red. The 15% SDS-PAGE with a stained protein ladder (MW) displays the input of the SEC (IN) and the flow-through (FT) during the sample concentration. One fraction from each of the first 2 peaks and all fractions of the elution peak were visualised on the SDS-PAGE. The N-RsbR2 protein displays a MW at approximately of 17 KDa which is around the MW of N-RsbR2.

3.6.3. Se-Met N-RsbR3 purification

Se-Met labelled N-RsbR3 protein purification (**Figure 3.10, Supplemental figures S9**) worked as well as the purification of the unlabelled version of N-RsbR3.

The incorporation of Se-met was also assessed using LC-MS (**Supplemental figures S10**). N-RsbR3 has 6 methionine (including the initiating methionine) and a MW of 17901.5Da. The Se-met labelled N-RsbR3 has a MW of 18183.6 Da by MS and the difference is 282.1 Da, corresponding to the complete replacement of 6.020 methionines by Se-Met.

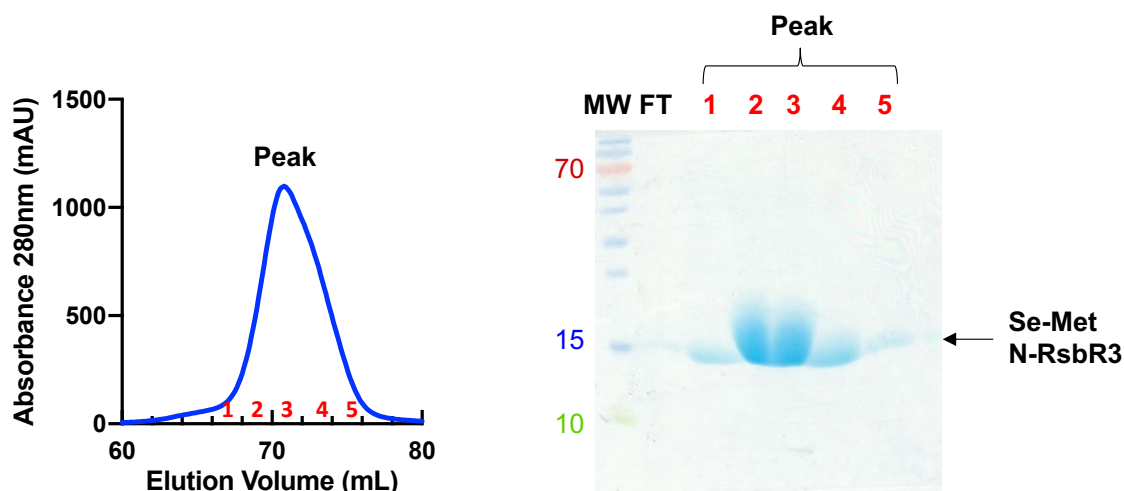


Figure 3.10. Selenomethionine-labelled N-RsbR3 protein purification

Size-exclusion chromatogram of Se-Met N-RsbR3, using a S75 16/60 column. The 15% SDS-PAGE with a stained protein ladder (MW) displays the flow-through (FT) during the sample concentration and the peak fractions (1-5). A highly pure Se-Met N-RsbR3 protein sample has been purified, with an apparent MW of approximately 15 kDa.

3.6.4. Se-Met N-RsbR4 purification

The Se-Met labelled N-RsbR4 protein purification was, again, the most difficult to obtain at the levels demanded for X-ray crystallography (**Supplemental figures S11**). The Se-Met N-RsbR4 purification yielded a highly contaminated final sample with a low N-RsbR4 content as seen on the SDS-PAGE (**Figure 3.11**). Having a heavily contaminated protein sample, such as Se-Met labelled N-RsbR4, is simply inappropriate for crystallography.

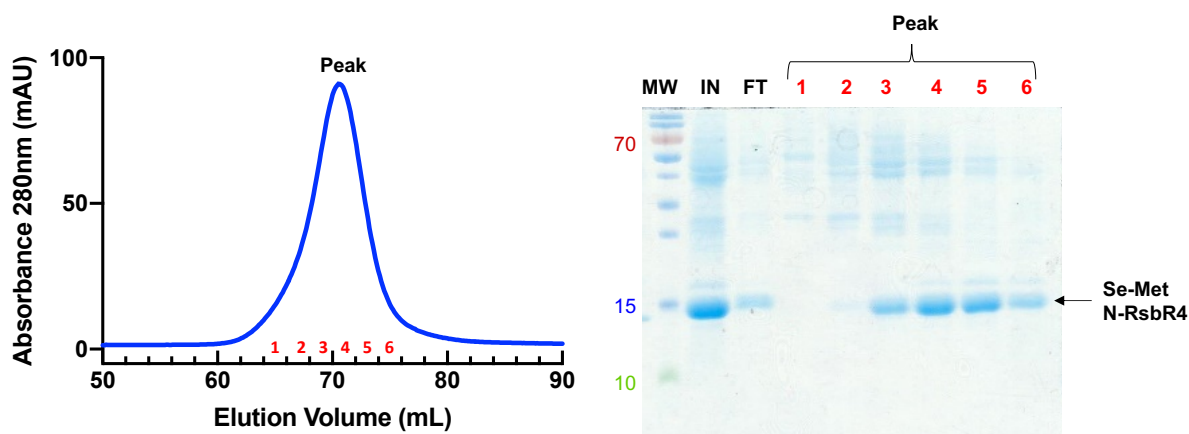


Figure 3.11. N-RsbR4 Se-Met labelled recombinant protein purification

Size-exclusion chromatogram of Se-Met labelled N-RsbR4 using a S75 16/60 column. The fractions collected (2 mL volume) are annotated in red. The 15% SDS-PAGE with a stained protein ladder (MW) displays the input of the SEC (IN), the flow-through (FT) during the sample concentration and the fractions of the elution peak.

To obtain a better yield and pure protein sample, a hexa-histidine tag (His₆-tag) was added to the CTD of N-RsbR4 (**Chapter II, section 2.1.2**). The His₆-tagged protein was first expressed without Se-Met labelling (**Figure 3.12**), and purified (**Chapter II, section 2.3.1**) but the addition of the tag led the protein to form inclusion bodies and all the expressed protein was found in the pellet when clarifying the cell lysates. At this point the challenges of preparing N-RsbR4 meant that this protein was put to one side and the focus was placed on N-RsbR1, 2 and 3.

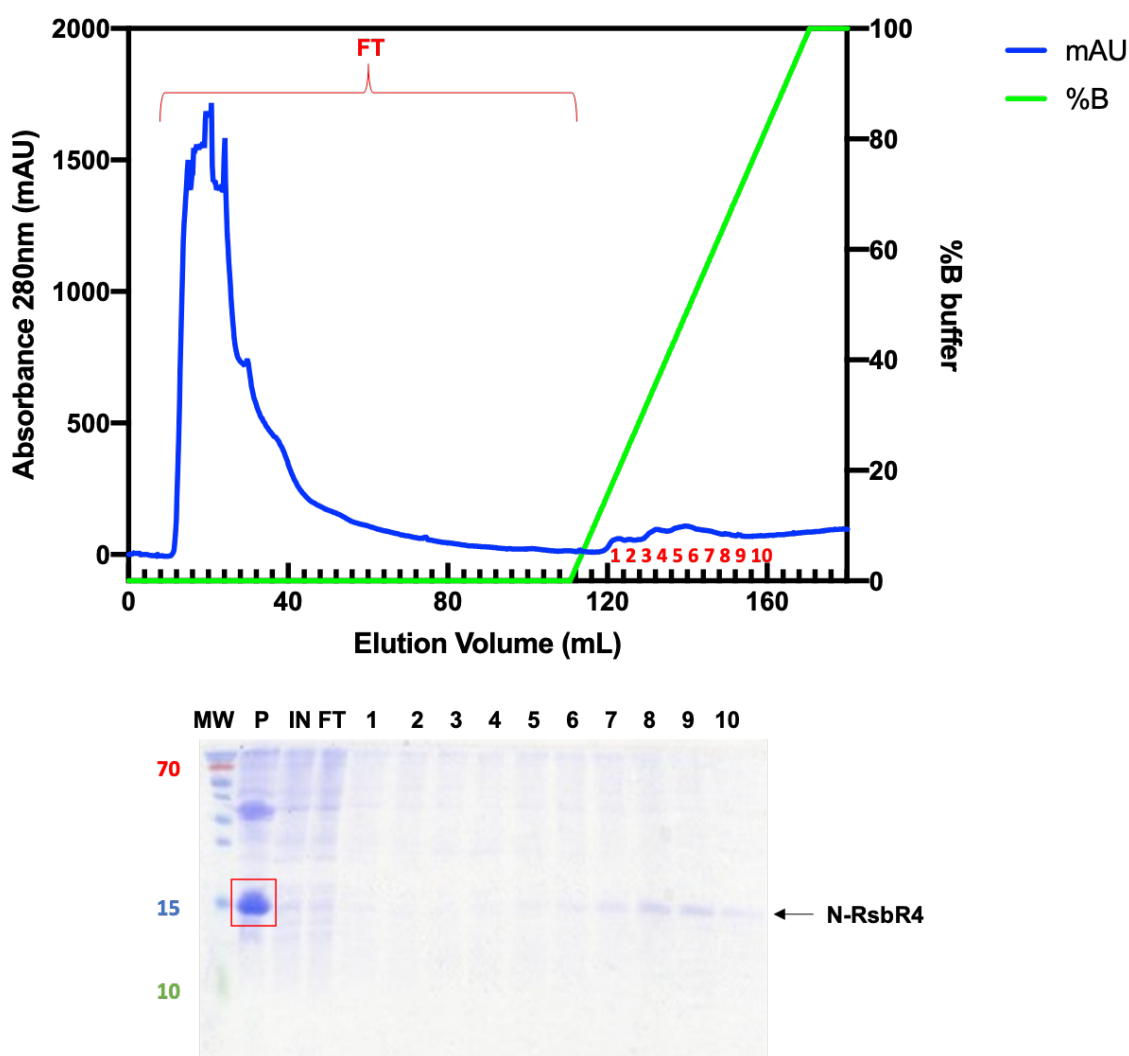


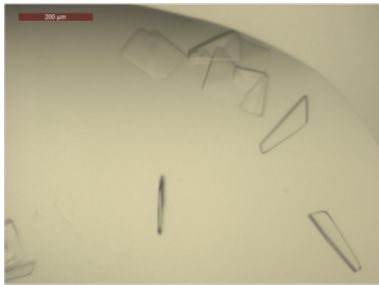
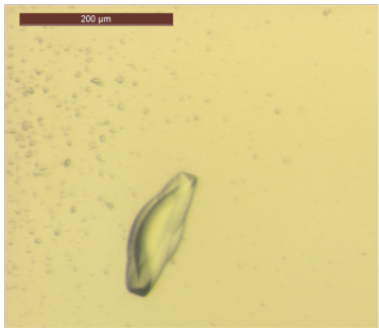
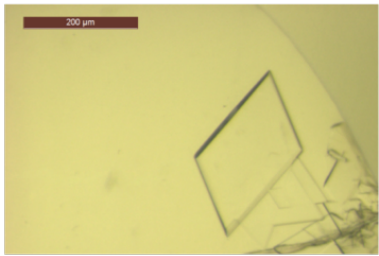
Figure 3.12. His₆-tagged N-RsbR4 protein purification

IMAC of His₆-tagged N-RsbR4 protein purification. The sample was eluted from the NTA column with an imidazole gradient (buffer B). The 15% SDS-PAGE with a stained protein ladder (MW, in kDa) displays the pellet (P), the input of the sample (IN) and the flow-through (FT) of the column, which is composed of the unbound proteins. Most of the His₆-tagged N-RsbR4 protein is found in the pellet, and only a little amount was found in the eluted fractions.

3.7. Se-Met labelled N-RsbR crystallisation screening and data collection

As the protein had been modified by the addition of the Se-Met, new crystallisation screening was set up. The subsequent Se-Met protein samples were concentrated to approximately the same concentrations as the unlabelled proteins for crystallisation screening. Se-N-RsbR1 was used at a concentration of 20 mg/mL, Se-N-RsbR2 at 13mg/mL and Se-N-RsbR3 at 12.3 mg/mL. Diffracting crystals for all of the 3 proteins were obtained in various conditions (**Table 3.4**). Crystals were obtained in the PACT condition D2 (0.1 M MMT buffer pH 5.0 and 25% w/v PEG 1500) for Se-N-RsbR1; in PACT condition F4 (0.2 M potassium thiocyanate, 0.1M Bis Tris propane pH 6.5 and 20% w/v PEG 3350) after a longer incubation period for Se-N-RsbR2 compared to the other proteins; and finally, in the Index condition E11 (0.2M magnesium chloride, 0.1M HEPES pH 7.5 and 22% w/v poly(acrylic acid sodium salt)) for Se-N-RsbR3. The crystals were harvested using appropriate loop size and cryo-protected with 20% PEG400 for N-Rsb1 and N-RsbR2 and 6M Na formate for N-RsbR3.

Table 3.4. Se-Met protein crystals of *Lmo* N-RsbRs

Protein	Crystals	Condition
Se-N-RsbR1		0.1M MMT buffer pH5.0 25% w/v PEG1500
Se-N-RsbR2		0.2M potassium thiocyanate 0.1M Bis-Tris propane HCl pH6.5 20% w/v PEG3350
Se-N-RsbR3		0.2M magnesium chloride 0.1M Hepes-OH pH7.5 22% w/v poly(acrylic acid sodium salt

Diffraction datasets from these crystals were collected and processed as for the native crystals (**Table 3.5**). The datasets contain both I+ (hkl) and I- (-h-k-l) observations, in the case of a non-experimental phasing dataset no significant differences are usually observed between I+ and I-; by contrast, a difference between I+ and I- can be observed when an anomalous scatterer is present and if the data were collected at an appropriate wavelength. This difference is given in an anomalous multiplicity value with the $CC_{1/2anom}$ value, with values of 0.8 and above considered as good anomalous signal (Karplus and Diederichs, 2015). These datasets have good statistics as described above, in addition to a good anomalous signal for experimental phasing: 19.7 for N-RsbR1, 10.2 for N-RsbR2 and 15.8 for N-RsbR3.

Table 3.5. Data collection statistics for SAD experiments

The values in parentheses correspond to the highest resolution outer shell

Data collection	Se-N-RsbR1	Se-N-RsbR2	Se-N-RsbR3
Date	21/07/2018	13/05/2019	08/07/2018
Type of data collection	Se-SAD	Se-SAD	Se-SAD
Source	I04	I24	I24
Wavelength (Å)	0.976Å	0.979Å	0.979Å
Space group	P2 ₁ 2 ₁ 2 ₁	P3 ₁ 21	P2 ₁ 2 ₁ 2 ₁
Cell dimensions			
a, b, c (Å)	30.93 58.24 142.38	54.89 54.89 94.81	50.0 81.0 85.0
α, β, γ (°)	90 90 90	90 90 120	90 90 90
Resolution (Å)	47.46 – 2.65 (2.78 - 2.65)	94.81 – 2.80 (2.95 - 2.80)	43.1 – 1.66 (1.69 - 1.66)
No. of measured reflections	283050 (38229)	74421 (11769)	1457881 (6569)
No. of unique reflections	8033 (1024)	3979 (619)	41594 (2019)
Multiplicity	35.2 (37.3)	144.8 (111.2)	35.1 (32.5)
CC _{1/2}	1.0 (0.77)	0.999 (0.75)	1.0 (0.70)
I/(σ)I	17.7 (1.5)	13.2 (1.6)	18.4 (1.5)
R _{pim} (all I+ and I-)	0.010 (0.462)	0.014 (1.03)	0.012 (0.565)
Completeness (%)	100.0 (100.0)	100.0 (100.0)	100.0 (100.0)
CC _{1/2} anom	0.639 (0.033)	0.804 (0.019)	0.867 (-0.021)
Anomalous multiplicity	19.2 (19.7)	10.4 (10.2)	17.4 (15.8)

Two crystal structures were determined using Se-Met experimental phasing, Se-N-RsbR2 and Se-N-RsbR3 which will be discussed later in this Chapter. Se-SAD phasing for Se-N-RsbR1 failed as the substructure solution could not be found. A good anomalous signal

($d''/\sigma = 0.85$) was found in SHELX C up to the 2.8 Å resolution shell but no strong anomalous scatterer positions were found among 10000 tries in SHELX D. Even with a subset of the best weak positions, the substructure search with SHELX E failed to generate a solution with a good FOM and CC.

Therefore, another experimental phasing approach was necessary for N-RsbR1. Since the native N-RsbR1 protein crystals diffract well (beyond 2.1 Å resolution), it is possible to exploit the weak, but nevertheless present and measurable, anomalous scattering from the native sulphur atoms for experimental phasing as used and described 40 years' ago for crambin (Hendrickson and Teeter, 1981). The N-RsbR1 sulphur SAD experiments were conducted using the long wavelength I23 beamline at the DLS synchrotron, a brand-new facility with design features not found on traditional synchrotron beamlines, where X-ray flux is usually the primary consideration, to mitigate the particular challenges of working at longer X-ray wavelengths (Wagner et al., 2018). The novel I23 beamline features include a Pilatus 12M detector arranged in a cylindrical geometry and with a specifically built goniometer with an inverse κ angle which allows to measure a wide range of diffraction angles. The beam path, goniometer and detector are maintained *in vacuo* to reduce radiation damage and to avoid air scattering. The specific challenges posed include radiation damage and the requirement of absorption correction to correct the different absorption length of individual reflections, which are a function of the sample size and morphology. The I23 beamline can use energies corresponding to X-ray wavelengths between 1.5 Å and 4 Å whereas most other beamlines for X-ray crystallography can only access between approximately 0.6 Å and 2.5 Å. The N-RsbR1 sulphur SAD experiment collected at I23 was one of the first successful structure determination from the Diamond users at Newcastle University.

3.8. Sulphur-SAD experimental phasing for N-RsbR1

The use of sulphur as an anomalous scatterer is still new and not widely used. With the help of the I23 beamline scientist Armin Wagner and his team, the anomalous signal from sulphur atoms within the protein was exploited. The anomalous signal has to be measured not too close to the absorption edge of the anomalous scatterer in long wavelength experiments as photon absorption increases (Wagner et al., 2016) and it has been previously suggested that the optimal wavelength for native SAD is around 2.1 Å based on radiation damage and data collection statistics (Mueller-Dieckmann et al., 2005). But the optimal wavelength is still dependant on each sample and varies between samples. Therefore, in the case of the N-RsbR1 crystals, the incident beamline energy was tuned to a wavelength equivalent of 2.75 Å.

Seventeen datasets were collected from two crystals of N-RsbR1 from the Structure A6 condition at different kappa angle sets (5°, 10°, 15° and 20°). Following the data processing protocol described in **Chapter II, section 2.9.3**, each individual dataset was processed (indexed, integrated and scaled) using XDS (Kabsch, 2010b) and their statistics, including the anomalous multiplicity, was checked carefully. Each dataset had an approximate anomalous signal of 3, which is weak for experimental phasing (Terwilliger et al., 2016). The datasets were scaled together using XSCALE (Kabsch, 2010a) and each was merged pairwise to see how the anomalous signal changed with each dataset, bearing in mind that radiation damage was likely to reduce the anomalous signal. Finally, all 17 datasets were merged into one to obtain a strong anomalous signal as both crystals were closely isomorphous ($a, b, c = 31.25\text{\AA}, 59.45\text{\AA}, 142.11\text{\AA}$ versus $31.33\text{\AA}, 59.65\text{\AA}, 141.98\text{\AA}$). The anomalous signal has a $CC_{1/2\text{anom}}$ of 0.609 and an anomalous multiplicity of 56.1, which suggests a strong anomalous signal should be present within the data.

Following scaling, the dataset was reduced with the Aimless/Pointless (Evans and Murshudov, 2013) programs found in the CCP4i2 suite (Potterton et al., 2018). These programs determine the correct space group and generate statistics on the overall data quality that needs to be inspected. The Matthews coefficient (Matthews, 1968) enabled the solvent content (37%) and the number of monomers (2) in the asymmetric unit (ASU) to be calculated. The scaled dataset was then used as input to the SHELX CDE suite (Sheldrick, 2010) to find the substructure of the anomalous scatterer, trace an initial backbone and to generate a map. In parallel, the substructure and the initial backbone trace was imported to the CCP4i2 suite (Potterton et al., 2018) and used for automated model building using the Buccaneer program (Cowtan, 2006) with the protein sequence as a guide. Two copies of the built model were placed in the ASU using MOLREP (Vagin and Teplyakov, 2010) and refined using Refcma5 (Murshudov et al., 2011). The model was then inspected manually using COOT (Emsley and Cowtan, 2004) to correct the path of the backbone, adjust misplaced sidechains and add waters. The modifications to the model were saved, and the updated model was refined against the data to see how the modification affected global and local refinement parameters. Cycles of building and refinement were interspersed until the refinement statistics converged. Using this protocol, the structure of N-RsbR1 was solved and refined at a resolution of 2.14\AA , with R_{work} and R_{free} values of 0.233 and 0.297, respectively. A summary of the data collection and model refinement statistics can be found in **Tables 3.6**.

Table 3.6. N-RsbR1 S-SAD data collection statistics

The values in parentheses correspond to the highest resolution outer shell

Data collection	N-RsbR1
Date	15/02/2019
Type of data collection	S-SAD
Source	I23
Wavelength (Å)	2.755Å
Space group	P2 ₁ 2 ₁ 2
ASU content	2
Cell dimensions	
a, b, c (Å)	31.33 59.65 141.98
α, β, γ (°)	90 90 90
Resolution (Å)	70.99 – 2.14 (2.20 - 2.14)
No. of measured reflections	2234206 (138822)
No. of unique reflections	15431 (1248)
Multiplicity	144.8 (111.2)
CC_{1/2}	1.0 (0.64)
I/σI	33.9 (1.6)
R_{pim}	0.006 (0.837)
Completeness (%)	100.0 (100.0)
CC_{1/2} anom	0.609 (0.045)
Anomalous multiplicity	74.8 (56.1)
Refinement	N-RsbR1
Resolution	71.09-2.10
No. of reflection all/free	16276/802
R_{work}/R_{free}	0.233/0.297
No. atoms	
Protein	2176
Water	2
R.M.S.D	
Bond length (Å)	0.0071
Bond angles (°)	1.408
B factor	73.87
Main chain	68.11
Side chain	79.22
Water	68.86
Ramachandran outliers	
Favoured	243 (97.20%)
Allowed	4 (1.60%)
High-energy	3 (1.20%)
Molprobability score	2.32

3.9. N-RsbR1 structure

The N-RsbR1 structure is a dimer (**Figure 3.13**) and overall is similar to those of *B. subtilis* N-RsbRA (2BNL) and *M. thermoacetica* N-MtR (3ZTA) with RMSD values of 1.94Å and 2.62Å, respectively, based on a monomer:monomer comparison. When dimers are superimposed, the RMSD values increase slightly to 2.78Å and 3.41Å, which indicates that the dimer assemblies are also highly comparable.

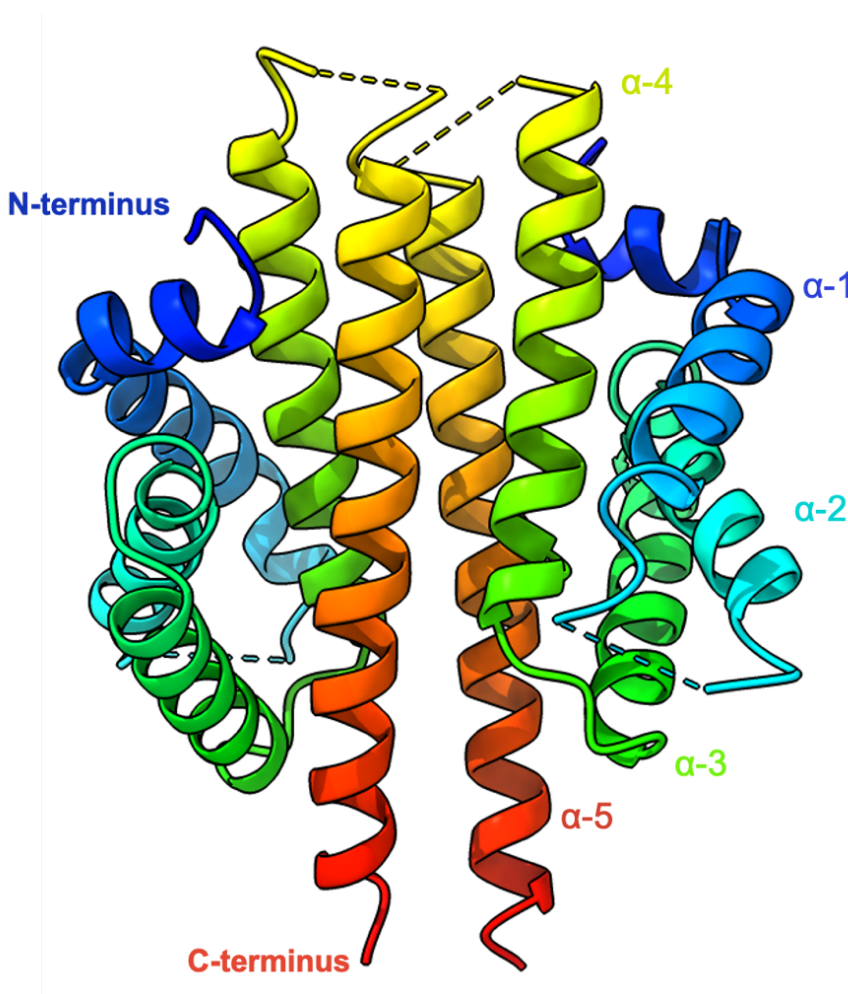


Figure 3.13. X-ray crystal structure of *L. monocytogenes* N-RsbR1

The X-ray crystal structure of *L. monocytogenes* N-RsbR1 was determined by sulphur SAD. Each monomer is represented in cartoon depiction and coloured in a rainbow fashion from the N- to the C-terminus from blue to red. All α -helices are noted in their respective colours. The discontinuous lines represent unmodelled loops: loop 1 from residues Q30 to F38 and loop 4 from residues D103 to D115.

As with the other N-RsbR structures, N-RsbR1 adopts a globin-like fold but without the capacity of binding heme and present 2 possible waters (**Supplemental figures S12**). The area in which heme would be found is filled in N-RsbR1 by the following amino acids: W20, M24, F48, W79 and F83. Two loops were not modelled in each monomer because of the absence of

interpretable electron density: loop 1, between α -1 and α -2 (from Q30 to F38) and loop 4, between α -4 and α -5 (from D103 to D115). Compared to N-MtR, the C-terminal J-helices at the dimer interface (implicated in the signalling pathway in YtvA and other LOV-sensors (Gaidenko et al., 2006; Quin et al., 2012)) are parallel and do not cross one another (**Chapter I, Figure 1.9**).

The structural similarity of N-RsbR1 with N-RsbRA (2BNL) and N-MtR (3ZTA), the genes for which are all found in sig^B-like operons, might indicate they have a similar sensing function or sense similar types of stress such as high salt or low temperature. Though there is not the capacity to bind a ligand as large as heme, as in the HemAT globin-coupled sensor (1OR4), there remains the potential to bind ligands in a similar location between helices 1, 2 and 4. The sequence alignment of these three protein sequences (**Figure 3.14**) reveals that just five residues are strictly conserved, with overall sequence conservation between N-RsbR1 with 2BNL and 1OR4 of 23.1% and 17.1%, respectively. Nevertheless, the overall conservation of residue properties – if not absolute sequence identity – allows the proteins to adopt the same globin fold.

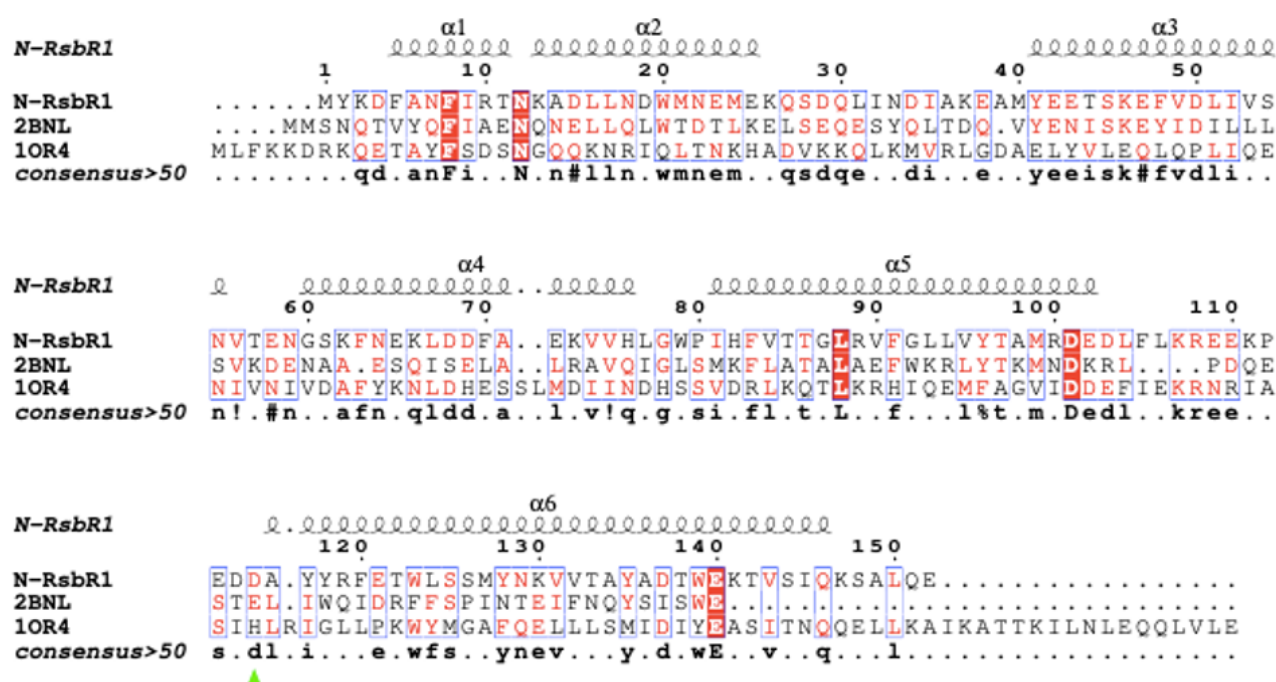


Figure 3.14. Sequence alignment of bacterial proteins adopting a globin-like fold
The sequence alignment of the NTDs of RsbR paralogues from *B. subtilis* (PDB ID: 2BNL), *L. monocytogenes* (N-RsbR1) and HemAT (PDB ID: 1OR4) reveals five strictly conserved residues highlighted in red (F8, N12, L88, D101 and E140). The green triangle shows the position of the proximal histidine on 1OR4. The alignment was generated with the Multalin webserver (Corpet, 1988) (<http://multalin.toulouse.inra.fr/multalin/>). The residue numbering follows that of *L. monocytogenes* RsbR1 and its secondary structure features are depicted above the alignment.

3.10. N-RsbR2 structure

The structure of N-RsbR2 was solved by SAD from crystals of selenomethionine-labelled protein using the same pipeline for the S-SAD analysis of N-RsbR1, except that only one dataset was available from a single crystal (**Table 3.5**). The solvent content is 45.0% and there is 1 molecule of N-RsbR2 in the asymmetric unit. The model was built and refined to convergence with R_{work} and R_{free} values of 0.207 and 0.268, respectively (**Table 3.7**).

Table 3.7. Refinement statistics for Se-SAD N-RsbR2

Refinement	Se-N-RsbR2
Resolution	47.85-2.90
No of reflection all/free	4053/231
$R_{\text{work}}/R_{\text{free}}$	0.207/0.268
Space group	P3 ₁ 21
ASU content	1
No. atoms	
Protein	1113
Water	2
R.M.S.D	
Bond length (Å)	0.0065
Bond angles (°)	1.547
B factor	104.58
Main chain	99.75
Side chain	109.13
Water	70.9
Ramachandran outliers	
Favoured	113 (92.62%)
Allowed	4 (3.28%)
High-energy	5 (4.10%)
Molprobity score	2.93

Though completely α -helical, the structure of N-RsbR2 is radically different from the structures of the other N-RsbR paralogues and orthologues solved thus far (**Figure 3.15A**) and does not adopt the globin-like fold maintained in the other structures. When aligned with the N-RsbR1 globin-like fold, the J-helix of N-RsbR1 align to the 2nd α -helix of N-RsbR2 (**Figure 3.15B**) and the RMSD of the superimposition between the 2 models is 4.3Å which indicates that the structures are not homologous to each other.

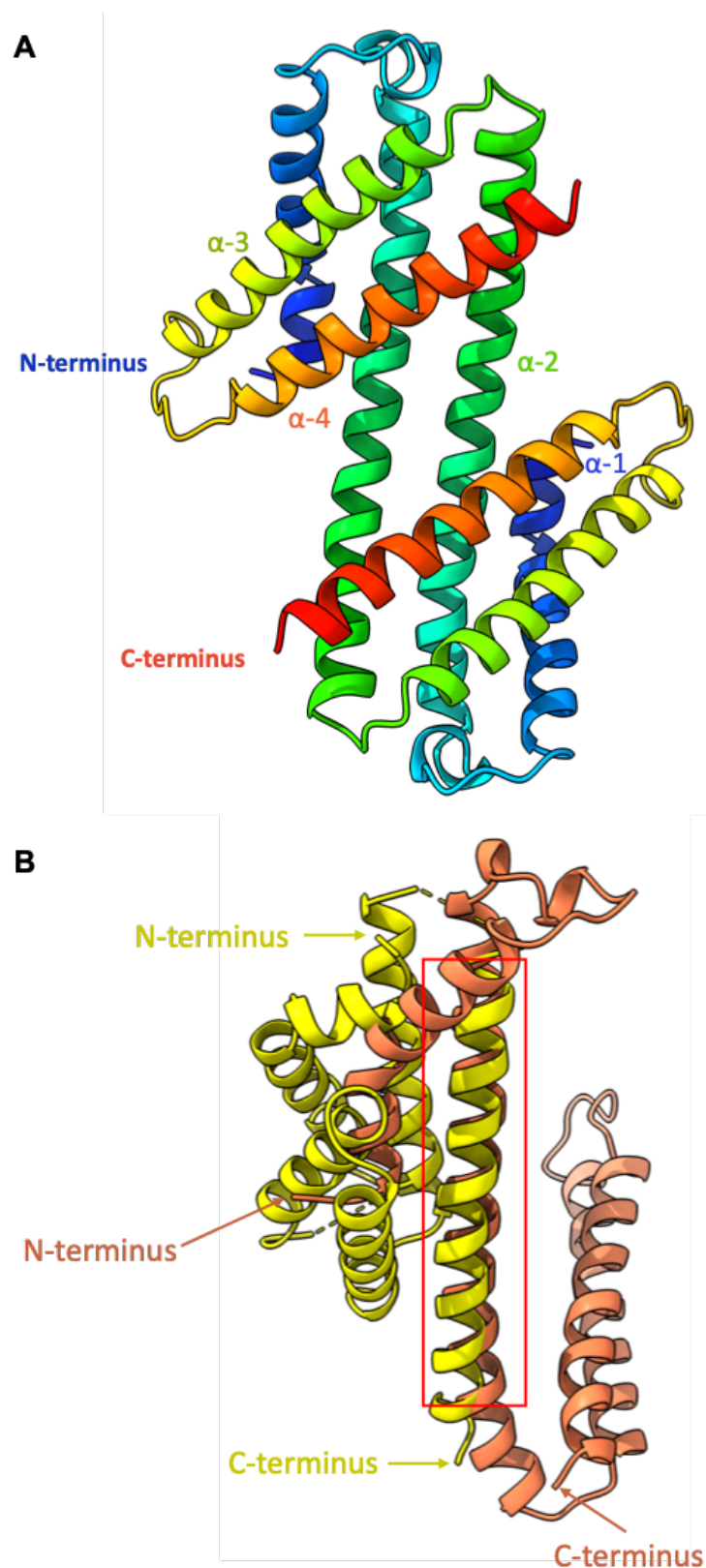


Figure 3.15. X-ray crystal structure of *L. monocytogenes* N-RsbR2

A. The X-ray crystal structure of *L. monocytogenes* N-RsbR2 was determined by Se-Met SAD. Each monomer is represented in cartoon depiction and coloured in a rainbow fashion from the N- to the C-terminus from blue to red. All α -helices are noted with their respective colours. **B.** The structure alignment shows only the superimposition of the N-RsbR1 J-helix in yellow to the 2nd α -helix of the N-RsbR2 model in orange, in red.

It appears that the crystallised N-RsbR2 protein went through a 2-step modification of its structure: local unfolding between $\alpha 2$ -L2- $\alpha 3$ followed by its refolding into a single, long α -helix that forms the $\alpha 2$ helix in this model and which aligns to the J-helix of N-RsbR1 (**Figure 3.15B**). The source of this unfolding and refolding is unknown. The N-RsbR proteins studied in this thesis were expressed, purified and handled in equivalent manners yet only N-RsbR2 has produced a structure unique from the others. The local un- and re-folding observed is similar to a well-known phenomenon called ‘domain-swapping’ (Lafita et al., 2019), which can be promoted by extremes of pH, high protein concentration and repeated freeze/thaw cycles (Zhang and Ertbjerg, 2019). In the case of N-RsbR2, which was purified and crystallised at pH values of 8.0 and 6.5, respectively, the protein was neither handled at extreme pH nor exposed to multiple rounds of freezing and thawing.

This Se-N-RsbR2 model was subsequently used as the search model to solve the structure of native N-RsbR2 by molecular replacement. The structure solved of N-RsbR2 was, to all intents and purposes, identical to that of Se-N-RsbR2 with $R_{\text{work}}/R_{\text{free}}$ values of 0.242/0.292 for the final model and RMSD values of 0.346 Å on 141 overall superimposed C α atoms (representing 100% of the total chain length of N-RsbR2). Therefore, the crystallised native and SeMet-labelled N-RsbR2 proteins adopt the same structure.

Since there was only a single molecule in the asymmetric unit for either N-RsbR2 or Se-N-RsbR2, a dimer was displayed by applying the unique crystallographic two-fold axis found in space group P3₁21. In so doing, however, the parallel nature of the J-helices of the usual globin-fold become anti-parallel in this newly determined N-RsbR2 structure dimer that is not compatible with the formation of the stressosome complex, with its protruding ‘turrets’ of N-RsbR domains, which requires the J-helices to be parallel on entry to the stressosome core. To understand if the local refolding of N-RsbR2 observed in the crystal structure was a function of the crystallisation process, Small Angle X-ray Scattering (SAXS) was used to determine the overall shape of the protein envelope in solution. The N-RsbR2 protein was concentrated to 27 mg/mL in the GF Buffer and sent to the BL21 beamline at the DLS synchrotron. With the local beamline scientist, Nikul Khunti, SAXS datasets from N-RsbR1 and N-RsbR2 were collected from a Superdex S200 Increase 3.2 size exclusion column at a flow rate of 0.075 ml/min. The data were processed as described in **Chapter II, section 2.6**. The SAXS data statistics and molecular envelope fitting (**Figure 3.16, Table 3.8**), processed in P1 and P2, indicates that a protein with a MW of approximately 34 kDa would fit in the molecular envelope. Moreover, the volume obtained for N-RsbR2 (60,439 Å³) is bigger than that obtained for N-RsbR1

(55,024Å³) from SAXS data, suggesting that the N-RsbR2 dimer is more elongated than the N-RsbR1 dimer. The molecular envelope derived from the SAXS data suggests that the N-RsbR2 protein in solution has the same conformation as the crystal structure as there is a better fit, with a smaller Chi square value, of the crystal structure to the SAXS envelope in comparison to the other models tested, including the N-RsbR2 monomer.

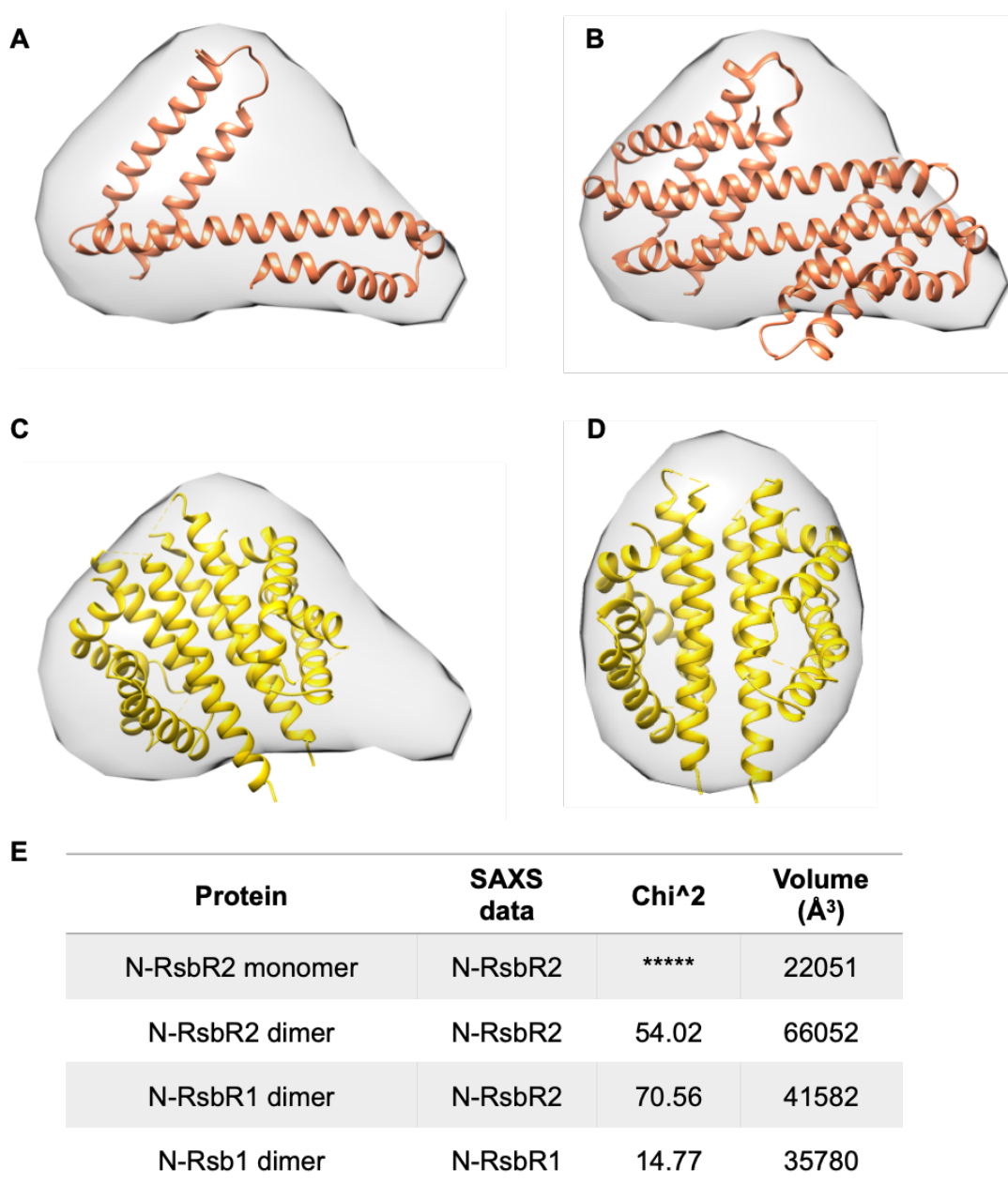


Figure 3.16. SAXS derived molecular envelope of N-RsbR2 in solution

Fitting of N-RsbR2 (**A**) monomer and (**B**) dimer in orange and (**C**) N-RsbR1 dimer in yellow in the molecular envelope. The chi-square values and the volumes are displayed in the table.

Table 3.8. SAXS data statistics of N-RsbR1 and N-RsbR2

Protein	Space group	Rg (Å)	Rmax (Å)	Mw (Da)	Final Chi	Max Radius (Å)	Volume (Å ³)
N-RsbR1	P1				1.966	38.31	
		20.85	62.67	34,258			55,024
	P2				1.976	38.29	
N-RsbR2	P1				1.388	61.2	
		24.86	83.93	34,272			60,439
	P2				1.379	49.4	

The unusual folding of N-RsbR2 in comparison to all other N-RsbR paralogues, and the incompatibility of the N-RsbR2 quaternary structure with the formation of the stressosome complex, raises an important question about whether RsbR2 is a functional paralogue. This question has been investigated by our network partners in the wider PATHSENSE consortium. First, *L. monocytogenes* strains were generated by Duarte Gueirero in Galway in which all *rsbR* paralogues – including *ytvA* – were deleted except for *rsbR2*. These strains have no obvious phenotype when grown under normal conditions and display a wildtype σ^B activation in response to acid stress, measured by cell survival. Furthermore, the RsbR2 paralogue was also found, by Western blot and peptide mass fingerprinting, to be incorporated in the native stressosome complex of *L. monocytogenes* in work conducted by our network partner in Madrid, Charlotte Dessaux. Therefore, RsbR2 is a functional paralogue and it would appear that the proper folding of N-RsbR2 may not occur in the absence of its C-terminal STAS domain.

3.11. N-RsbR3 structure

The structure of N-RsbR3 was determined by Se-SAD to a resolution of 1.66Å (**Figure 3.17**) using the same data processing pipeline as previously described (data collection statistics in **Table 3.5**). The model was refined to convergence with R_{work} and R_{free} values of 0.208 and 0.242, respectively; the final refinement statistics are displayed in **Table 3.9**. Structurally, N-RsbR3 is similar to *B. subtilis* N-RsbRA (2BNL) and *M. thermoacetica* N-MtR (3ZTA) models, with RMSD values of 3.39 and 3.75 Å, respectively, after superposition of the dimer.

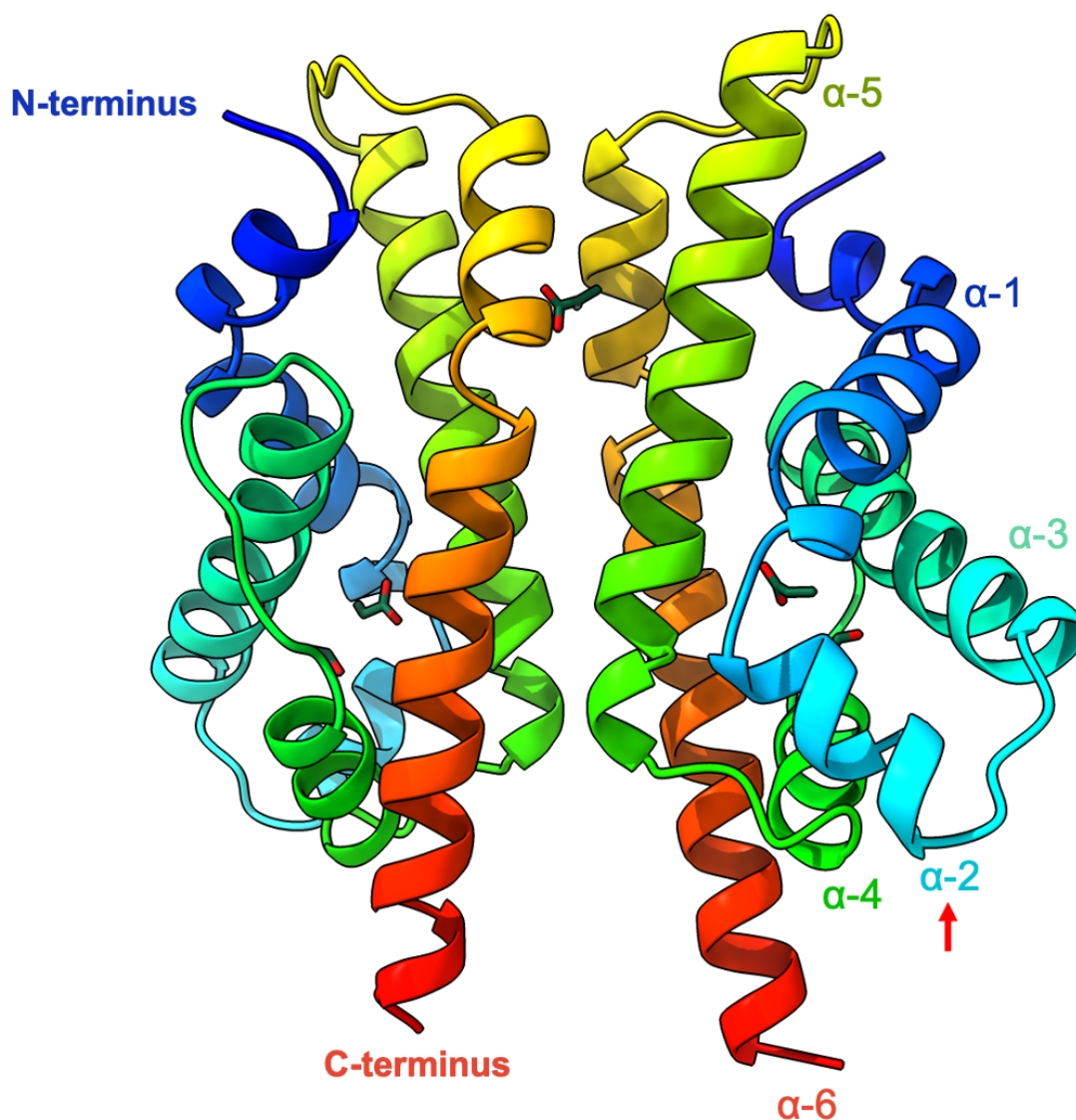


Figure 3.17. X-ray crystal structure of *L. monocytogenes* N-RsbR3

The X-ray crystal structure of *L. monocytogenes* N-RsbR3 was determined by Se-SAD. Each monomer is represented in cartoon depiction and coloured in a rainbow fashion from the N- to the C-terminus from blue to red. The N-RsbR3 structure displays an additional alpha-helix ($\alpha-2$, red arrow) compared to N-RsbR structures solved previously. Five small molecule ligands were discovered bound to this structure (shown in green sticks), 1 acrylic acid monomer at the top of the dimer interface and 1 acrylic acid in each monomer's ligand binding pocket in close proximity to a formic acid molecule.

Table 3.9. Refinement statistics for Se-SAD N-RsbR2

Refinement	Se-N-RsbR3
Resolution	42.54-1.66
No of reflection all/free	41526/2050
R_{work}/R_{free}	0.208/0.242
Space group	P3 ₁ 21
ASU content	1
No. atoms	
Protein	2296
Ligand	123
Water	122
R.M.S.D	
Bond length (Å)	0.0129
Bond angles (°)	1.808
B factor	29.69
Main chain	26.22
Side chain	32.88
Ligands	33.11
Water	33.92
Ramachandran outliers	
Favoured	250 (99.21%)
Allowed	2 (0.79%)
High-energy	0 (0%)
Molprobity score	1.43

However, there are some key structural differences observed, which includes the presence of an additional α -helix in N-RsbR3, between α -1 and α -2 and extending from Asp26 to Ser35. A sequence alignment of all the known N-RsbR structures (**Figure 3.18**) reveals the sequence insertion in N-RsbR3 and 3ZTA after the α -2 helix; in N-RsbR3 this helix has a length of 9 amino acids and the sequence insertion is 10 amino acids long. This insertion can be found in 2BNL and N-RsbR1 on the alignment in **Figure 3.18**.

Another structural difference is the conformation of the J-helices, which are different from what has been observed previously. The N-RsbR1 and N-RsbRA crystal structure have J-helices that are more or less parallel to each other and the N-MtR crystal structure has J-helices that are crossing like a pair of scissors. Here, the newly determined N-RsbR3 structure, the J-helices are wider and do not cross each other. The area in which heme would be found is filled in N-RsbR3 by the following amino acids: W22, 23, E28, R71, 72 and H85.

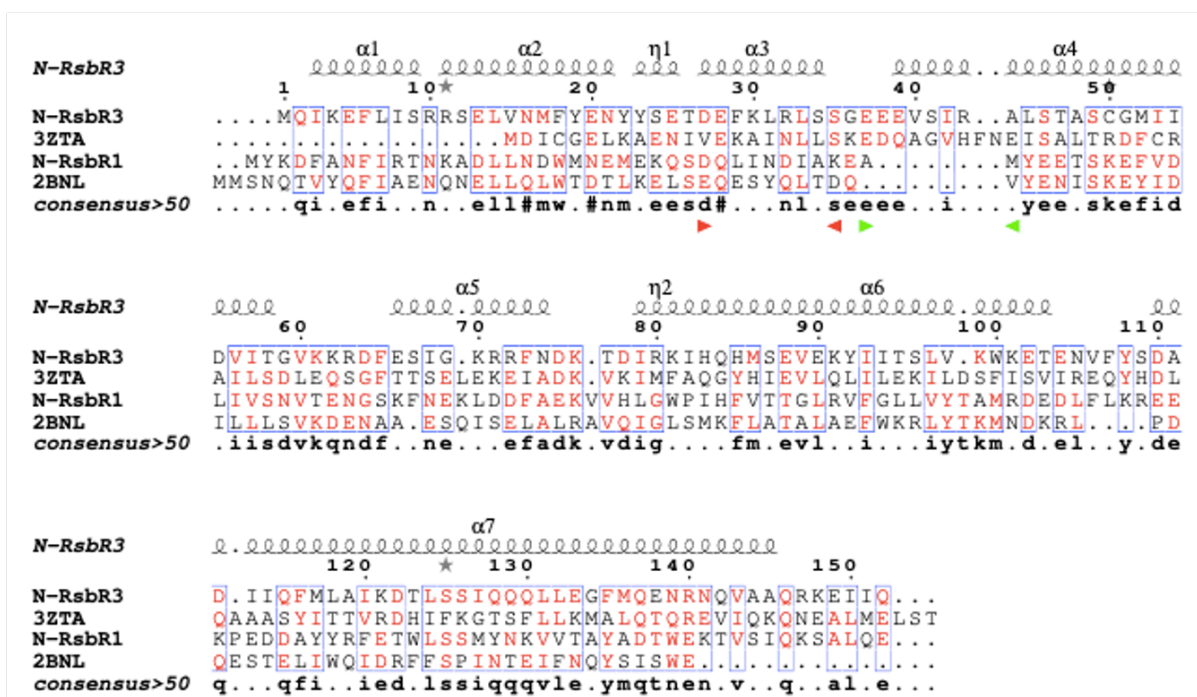


Figure 3.18. Multiple sequence alignment of N-RsbRs

The sequence alignment of the NTDs of RsbR paralogues from *B. subtilis* (PDB ID: 2BNL), *L. monocytogenes* (N-RsbR1 and N-RsbR3) and *M. thermoacetica* (PDB ID: 3ZTA) was generated using the Multalin webserver (Corpet, 1988) (<http://multalin.toulouse.inra.fr/multalin/>). The residue numbering follows that of *L. monocytogenes* RsbR3 and displays its secondary structure features above the alignment. The red triangles show the limits of the additional α-2 on the N-RsbR3 structure and the green triangles delimits the sequence insertion observed.

Finally, several ligands were found in the N-RsbR3 crystal structure. Two distinct electron density ‘blobs’ were left after the refinement of the protein had been completed that could not be unaccounted for by any proteinaceous atoms. These were located in a cleft between helices 2, 3 and 5, and are thus in a position that would be occupied by the co-factor in any standard heme-containing globin. Neither blob in isolation nor in combination is big enough to be a heme; each is much smaller and likely contain no more than about half a dozen non-hydrogen atoms. Several candidate molecules were built into the electron density, including single amino acids and dipeptides, but none could be refined satisfactorily. However, these protein crystals grew from solutions containing poly-acrylic acid as the precipitant and when an acrylic acid monomer was built into the larger of the two blobs, it refined satisfactorily with no significant change in the global refinement parameters R_{work} and R_{free} , but also without leaving any positive or negative residual electron density peaks in the immediate vicinity of the built acrylic acid (**Figure 3.19A**).

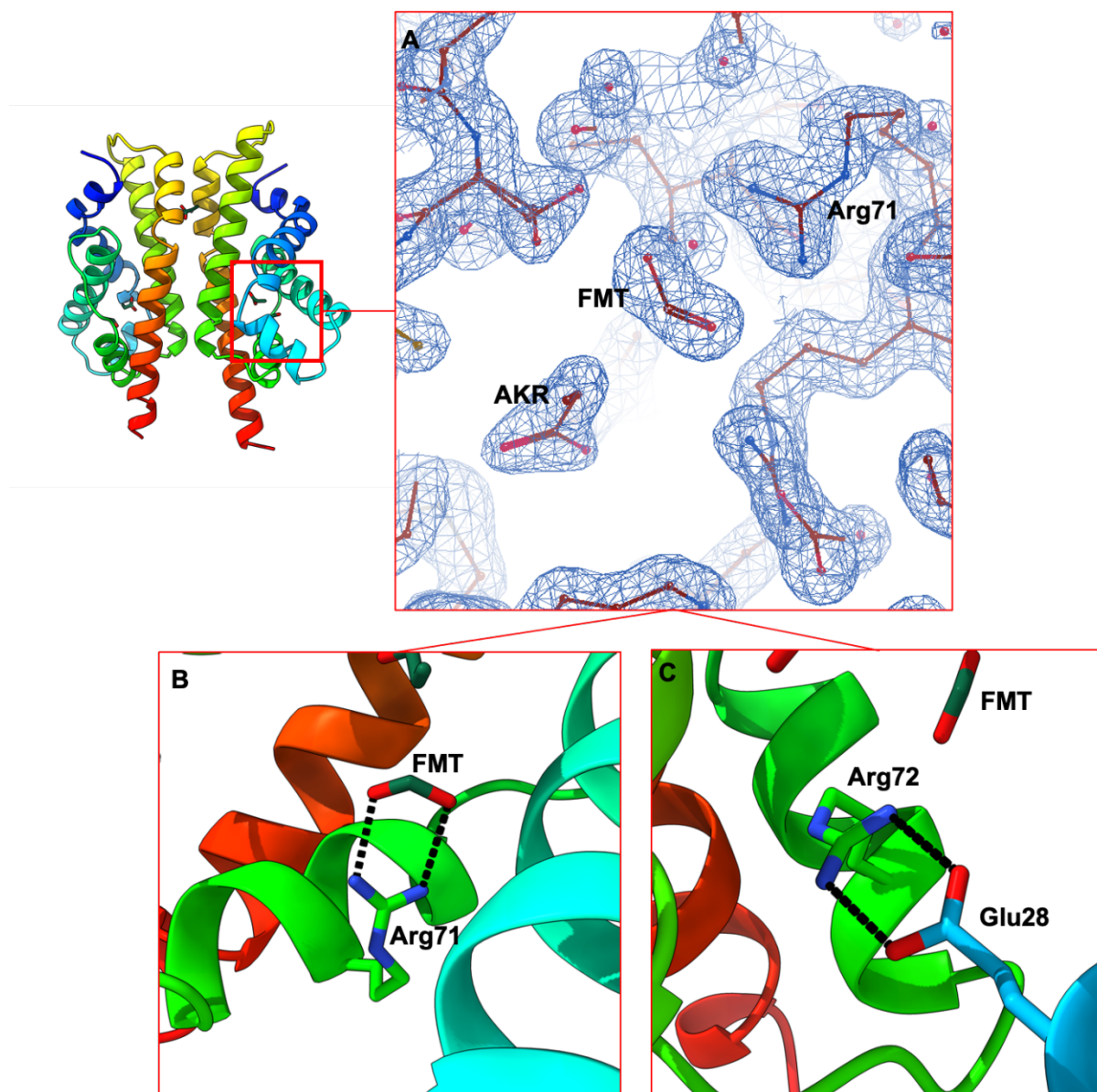


Figure 3.19. N-RsbR3 ligands in the putative ligand binding pocket

A. Final, Refmac-weighted $2F_{\text{obs}} - F_{\text{calc}}$ electron density in the vicinity of acrylic acid (AKR) and the formic acid (FMT) in the ligand binding pocket. **B.** The AKR ligand is held by weak interactions while the carboxylic acid group of FMT is held by H-bonds to the guanidinium moiety of Arg71. **C.** The extra α -helix in N-RsbR3 is stabilized with hydrogen bonds between the guanidinium group of Arg72 (α -helix 4) and the carboxylate group of Glu28.

Just as polyethylene glycol is not a single polymer but is in fact a range of polymers of ethylene glycol of average mass, e.g. 400 Da, polyacrylic acid is also a heterogeneous mixture of polymers and as such acrylic acid monomers are probably a contaminant. Alternatively, the single acrylic acid observed is the terminal moiety in a longer chain, but this would require a tunnel to be available through the protein and the protein surface does not display such a feature. The acrylic acid (AKR) seems to be held in place by a number of weak and non-specific van der Waals' interactions from N-RsbR3 residues including Ile53, Met86 and Cys50. The second

feature in this location could be modelled satisfactorily by formic acid, a derivative of acrylic acid, and this is mostly held in place by hydrogen bonds between the carboxylic acid group of the formate to the guanidinium of the sidechain of Arg71 (**Figure 3.19B**). The presence of formic acid is likely a consequence of the use of 6 M sodium formate as the cryoprotectant.

The observed extra α -helix in N-Rsb3 may play a role in binding potential ligands as part of a sensing mechanism, potentially uncovered by the binding of these ligands to carboxylates in the protein. The binding site covers a distinct cavity in the protein and the helix position is maintained by hydrogen bonds between the side chains of Glu28 in helix 2 and Arg72 in helix 4 (**Figure 3.19C**).

There is an additional unexplained electron density feature at the dimerization interface, at close proximity to MSE117 in both chains (**Figure 3.20A**). This electron density could be explained either by an acrylic acid (**Figure 3.20B**) or by an alternative MSE conformation (**Figure 3.20C**) as the density of MSE117 is discontinuous up to the selenium atom. An anomalous difference map was generated (**Figure 3.20C**) to determine if an alternative conformation of MSE117 could explain the feature. The anomalous difference map showed a strong signal for the position of the selenium in the two MSE117 residues, by contrast the anomalous difference signal for the putative alternative conformation of MSE117 was weak to non-existent, in essence no greater than the background noise in the anomalous difference density map. Even if the alternative conformation of MSE117 had a low occupancy, clear anomalous difference density should still have been apparent. Without such independent support for the identification of the unmodelled electron density as an alternative conformation of MSE117, an acrylic acid monomer was built instead in 2 different position at half occupancy each and refined satisfactorily.

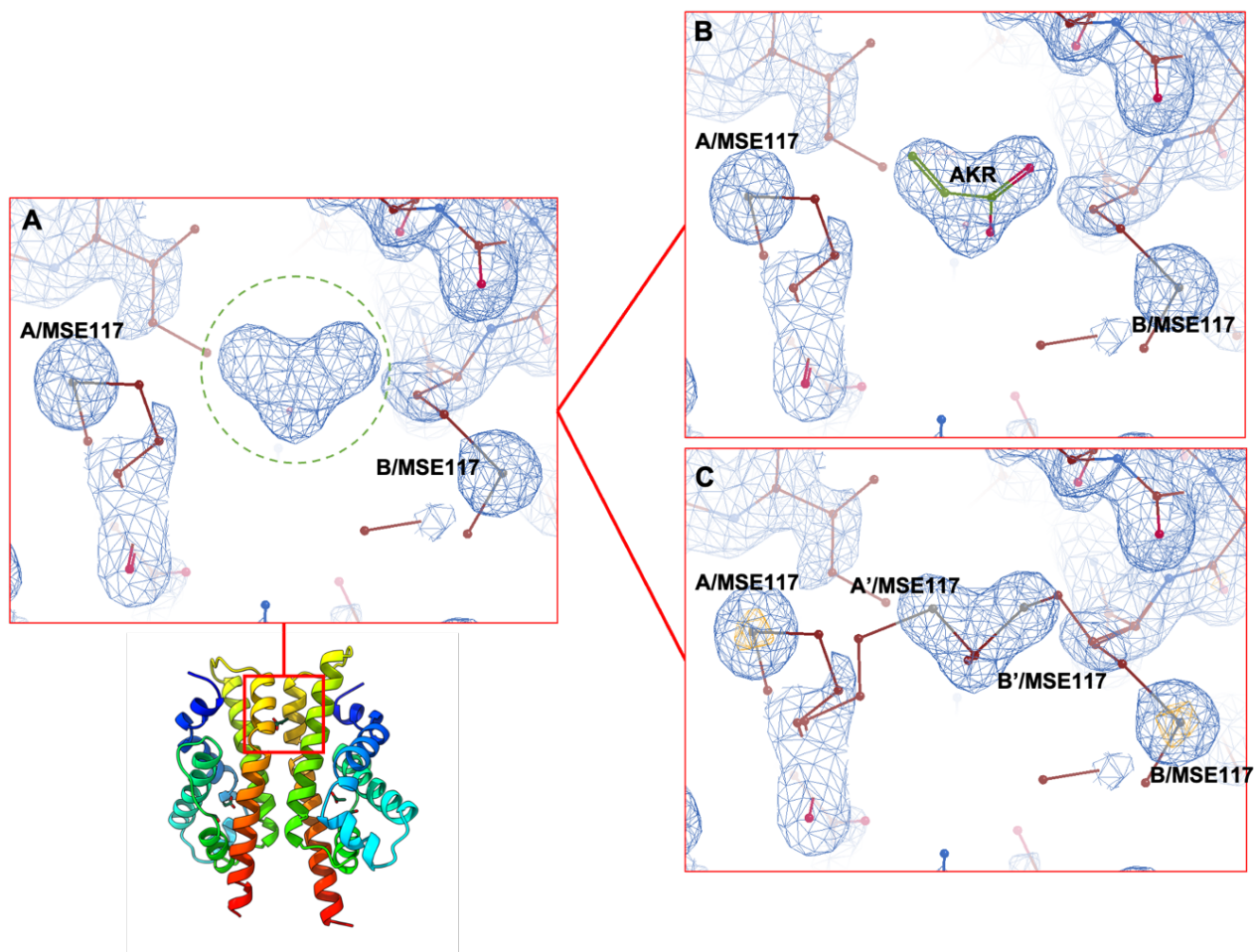


Figure 3.20. N-RsbR3 electron density at the dimerization interface

A. Unmodelled electron density map in the final, Refmac-weighted $2F_{\text{obs}} - F_{\text{calc}}$ electron density map (circled by a dotted green line) at the dimerization interface in close proximity to MSE117 from chains A and B. Two possibilities were tested: **B.** an acrylic acid (AKR) held by weak hydrophobic and ionic interactions or **C.** alternative conformations of both MSE117 residues. The anomalous difference in orange shows strong peaks at the difference map level corresponding to the MSE117 A and B positions, and weak or non-apparent for the alternative conformations of MSE117 A' and B'.

With knowledge of the ligand-binding sites in N-RsbR3, several residues that interact with either acrylic or formic acids could be identified and their role – if any – in stress sensing tested by mutagenesis. In the current absence of any *in vitro* assay for ligand sensing / a sensory mechanism, the effect of mutations will have to be tested *in vivo* in appropriate mutated *B. subtilis* strains and a σ^B -dependent reporter, such as *lacZ* at the *ctc* locus (Boylan et al., 1991, 1993a). For instance, Arg71 interacts directly with the modelled FMT ligand and thus it is not inconceivable to think that a natural ligand sensed by N-RsbR3 would also have a carboxylic

acid group in order to exploit the guanidinium moiety of Arg71. Moreover, the H-bonds between Arg72 and Glu28 might act as gatekeepers to the pocket for larger ligands with movement of α -2 helix linked to the opening and closing of the gate. At the time of writing a mutational analysis of N-RsbRA of *B. subtilis* was still being concluded by our consortium partners at the University of Dundee, Prof. Nicola Stanley-Wall and Diana Gudynaite. Key residues mutated to alanine included Arg71, Glu28 and Arg73, and σ^B activity was measured using β -galactosidase as the reporter from a σ^B -dependent gene reporter in the mutated *B. subtilis* strains. The outcome of these *in vivo* experiments, however, showed that the targeted single point mutations in RsbRA did not lead to any defect in σ^B activation. Therefore, pairs of mutations were combined, but the results of these experiments have yet to reach conclusion.

A second structure of N-RsbR3 was solved that was not crystallised in the presence of polyacrylic acid and that was not cryoprotected with sodium formate – these particular crystals were grown from and were cryoprotected in polyethylene glycol. This second structure does not have any serendipitous ligands bound. A structural superimposition (**Figure 3.21A**) of the two N-RsbR3 crystal structures has an overall RMSD of 0.93 Å. These changes can also be seen in the Kleywegt plot of chain B but not in chain A (**Figure 3.21B**). The differences between the bound and apo structures is a little higher in the vicinity of α -2 (Arg11 to Thr58) and the ligand binding site, with RMSD values exceeding 0.37 Å. The mean B factors for the bound and apo structures are 41.65 Å² and 59.09 Å², respectively, which suggest a more stable region for the bound structure. The ligand-bound form of N-RsbR3 would seem to be more stable than the apo structure. Indeed, the overall flexibility of the ligand binding site is underlined by the observation that the two chains in the asymmetric unit of the unbound N-RsbR3 structure have different conformations in the vicinity of the ligand-binding pocket (**Figure 3.22**). The flexibility in this region may well play a role in ligand binding during sensing and dissociation during the resetting of the signal transduction switch.

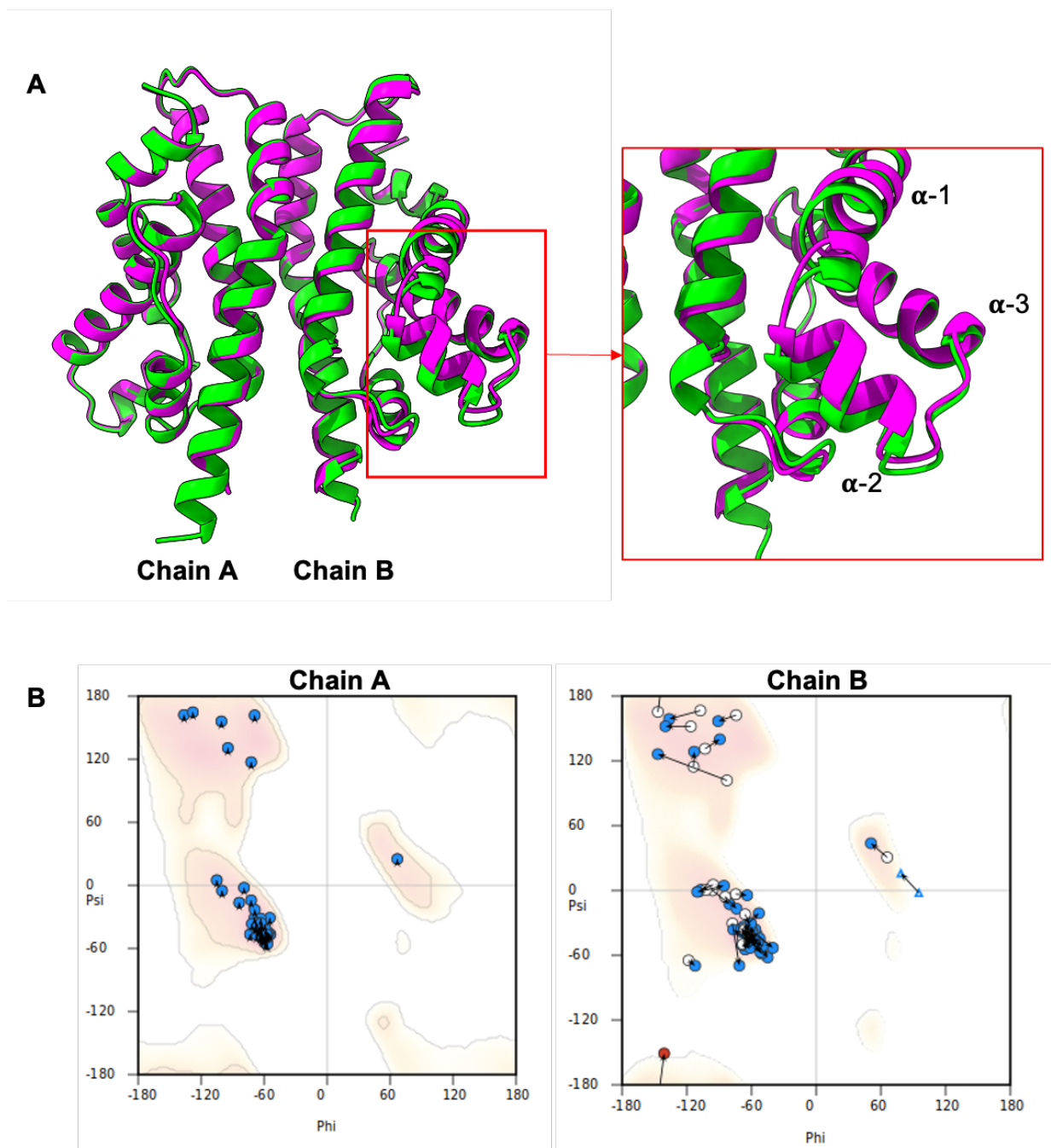


Figure 3.21. N-RsbR3 conformational changes with and without ligand

A. Superposition of the N-RsbR3 crystal structure solved in the absence (pink) and presence (green) of ligand. The structural differences are limited to the region around helices 1 to 3 as seen on the close-up view. **B.** Kleywegt plot for each chain showing the displacement of the residues from both ligand-bound and apo N-RsbR3 structures. Significant displacement can be seen for chain B compared to chain A.

3.12. Structure comparison

There are now four N-RsbR models that share similar structural features (**Figure 3.22**) in addition to the unique N-RsbR2 structure that probably represents a mis-folded outlier. When monomers of each N-RsbR structure are superimposed, the structural features line up well (**Figure 3.22**), and there are pairwise RMSD values in the range of 1.94 to 3.88 Å (**Table 3.10**). In this superimposition, the extra α -helix of N-RsbR3 stands out in comparison to the other structures. The length of the J-helix shows the greatest structural variation observed. *MtR*, 3ZTA, has the longest J-helix containing 12 turns and they cross like a pair of scissors. *L. monocytogenes* N-RsbR1 (8 turns) and *B. subtilis* N-RsbRA 2BNL (7 turns) share similar J-helices, which are more or less parallel to each other, and which are approximately the same length. N-RsbR3 shows a wider J-helix disposition in comparison to the other models but is approximately the same length as N-RsbR1, with 9 turns.

Perhaps unsurprisingly, the most closely related structures are *B. subtilis* N-RsbRA and *L. monocytogenes* N-RsbR1, with a RMSD value of 1.9Å for the superposition of the monomer. These two structures also share the highest sequence identity of 17.12%. These similarities probably arise from the fact that these are taxonomically closer than any of the other relationships between RsbR orthologues. The most divergent structure among the RsbR NTD, is N-*MtR*, which shares one of the lowest sequence identities, 7.3% with N-RsbRA and the highest structural deviation with 3.8 Å RMSD after global superposition at the monomeric level. This high deviation might reflect the fact that N-*MtR* has longer J-helices compared to the other N-RsbR structures. As discussed previously, the N-RsbRs share low sequence identities, nevertheless, the structures remain globally conserved. The fact that all the N-RsbRs adopt the same fold might imply that they are all involved in the same activity. If the NTD of RsbR proteins is confirmed to act as a stress sensor in future studies, it is likely that the stress sensing ability varies between paralogues and/or homologues of the RsbR proteins. The J-helix may also play a role in signal transduction from the NTD to the CTD of the RsbR protein. Indeed, the J-helix has been observed in several conformations, either crossed, parallel or wide open. This J-helix conformation could lead to the exposure of the phosphorylatable residues to the RsbT kinase. Nevertheless, to confirm these conclusions, further studies are required that are, due to time constraints, beyond the scope of this thesis.

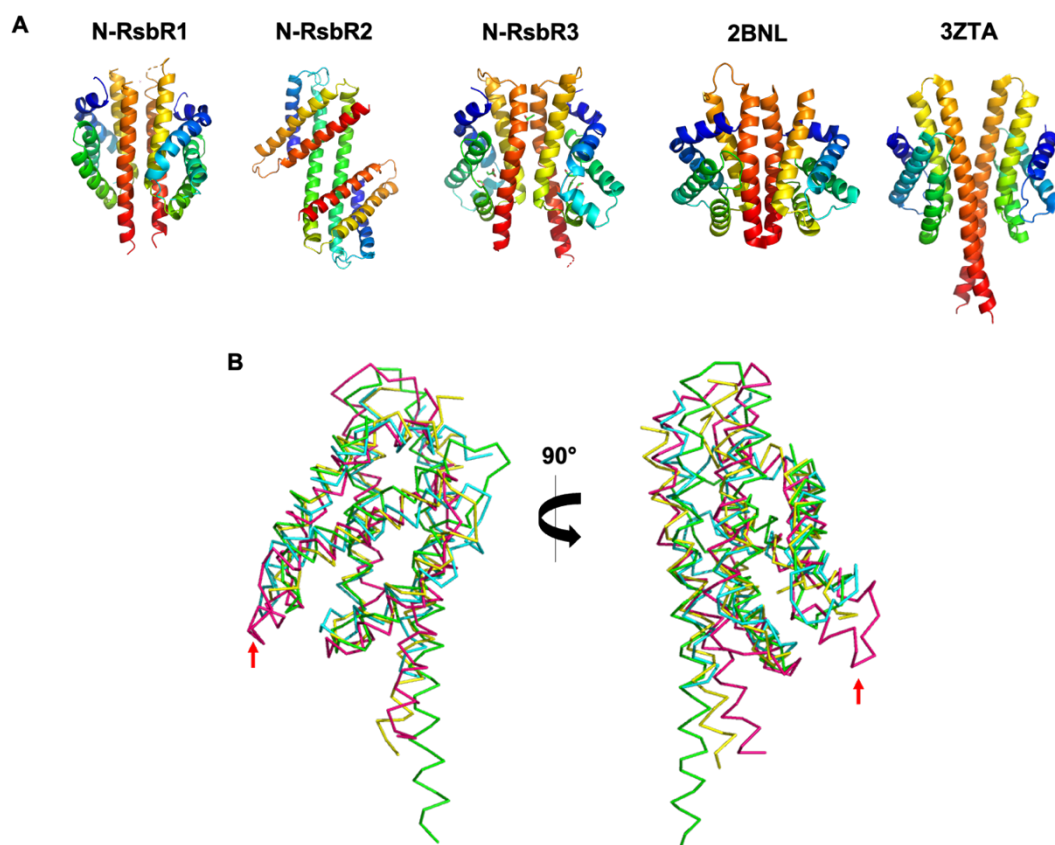


Figure 3.22. Structural alignment of RsbR N-terminal domains

A. All of the known crystal structures of RsbR homologues (N-RsbR1, 2BNL and 3ZTA) and paralogues (N-RsbR2 and N-RsbR3). **B.** A structure superimposition of all the N-RsbRs except for N-RsbR2. The structures are displayed in cartoon with each N-RsbR monomer colored separately (3ZTA in green, 2BNL in cyan, N-RsbR3 in pink and N-RsbR1 in yellow). The red arrow points to the extra helix observed in the N-RsbR3 structure.

Table 3.10: RMSD and sequence identity values for monomer structure comparison

RMSD values in Å (in bold) and the sequence identity in percentage (below) using the SSM superimpose tool in COOT. HemAT (PDB ID: 1OR4) added for comparison of the globin fold.

RMSD (Å) Seq ID (%)	N-RsbR1	N-RsbR3	N-MtR 3ZTA	N-RsbRA 2BNL	HemAT 1OR4
N-RsbR1		2.55 9.17	2.62 9.91	1.94 17.12	3.15 7.69
N-RsbR3			3.45 11.11	2.65 8.85	2.45 9.01
N-MtR 3ZTA				3.88 5.71	2.60 0.0
N-RsbRA 2BNL					3.00 10.08
HemAT 1OR4					

3.13. Conclusion

Solving three new structures of the N-RsbR domain has yielded new insights to a possible sensing mechanism. Structural differences between the paralogues include the presence of an extra α -helix and the more open conformation of the J-helix in N-RsbR3. These differences might imply signal transduction from the N-terminal globin-like domain to the C-terminal STAS domain by conformational changes after ligand binding by the former domain. It is unlikely that acrylic acid is a true ligand for any RsbR paralogue as it is not a natural metabolite in bacteria and the question now is what is or are the ligands that binds to this putative ligand binding pocket in RsbR paralogues? This question will be addressed in **Chapter IV**. Finally, with the N-RsbR3 structure in particular, ligand binding pockets have been identified and the role of amino acids important for ligand binding were assessed *in vivo* by our Network Partners in Dundee. However, to date the mutagenesis analysis has not led to the identification of residues that might impact the stress sensing activity of RsbRA in *B. subtilis*. The mutagenesis study might have to be done in pairs instead of single point mutations to identify important residues in signal transduction.

Chapter IV: Interaction study of RsbR with potential binding partners

4.1. Introduction

The N-terminal domain of the RsbR proteins (N-RsbRs) found in stressosome complexes have been proposed to act as stress sensors (Murray et al., 2005); however, the nature of the stresses sensed and any structural responses to these are not currently known. In **Chapter III**, the structures of three *L. monocytogenes* N-RsbR domains were determined. Although the sequences of paralogous N-RsbR domains are divergent, their structures have remained closely aligned during evolution and all adopt a globin-like fold. The divergent sequences but constant domain architecture could point to a role in stress sensing in which the sensory mechanism has been conserved whilst allowing each paralogue to have different stress sensing abilities.

A membrane-anchored miniprotein, Prli42, was identified recently as a putative interaction partner of N-RsbR1 in *L. monocytogenes* (Impens et al., 2017). The sequence of Prli42 is conserved within the firmicutes phylum. As discussed in **Chapter I, section 1.6**, the interaction between N-RsbR and Prli42 was modelled initially as one molecule of Prli42 interacting with a dimer of N-RsbR1, at the top of the dimer interface (Impens et al., 2017). The interaction was subsequently modelled as one molecule of Prli42 interacting with one monomer of N-RsbR1, also at the top of the N-RsbR1 monomer (Williams et al., 2019). The importance of Prili42 in stress sensing was also studied *in vivo* (**Chapter I, section 1.6**), however, missing controls and relevant data failed to describe the binding affinity between N-RsbR and Prli42. Therefore, the interaction affinity between all the *L. monocytogenes* and *B. subtilis* N-RsbRs against their respective Prli42 miniproteins was studied in this chapter, using a combination of different approaches, to better understand this interaction.

Moreover, putative ligand binding pockets in the crystal structure of *L. monocytogenes* N-RsbR3 were identified in **Chapter III**. With these observations, new tools were developed to further understand the stress sensing mechanism of the N-RsbRs. The role of residues that might be involved in the stress sensing activity of RsbR3, such as Arg71, was probed by mutagenesis and the impact on *Listeria* strains harbouring these mutations was assessed by consortium partners. In recognition that a ligand triggering the σ^B general stress response might have a carboxylate to interact with the guanidinium group of Arg71, several *in vitro* methods were also developed, including Thermal Shift Assays (TSAs), co-crystallisation with a home-made ligand screen and crystal soaking with a halogenated compound library called FragLites (Wood et al., 2019).

4.2. Interaction study of Prli42 by co-elution

A simple co-elution experiment was performed in the first instance to understand better the proposed interaction between N-RsbR1 and Prli42. Since Prli42 is a small protein of 32 amino acids, Prli42 was fused to a GFP-Strep-tag-II to aid purification. Prli42-GFP was purified by PATHSENSE consortium member Algirdas Miskys during his secondment in Newcastle from a streptavidin affinity column and bound proteins were eluted with desthiobiotin. Purified N-RsbR1 and Prli42-GFP proteins were mixed at a molar ratio of 1:1 and incubated at RT for 1 hr before being applied to a Superose 6 size-exclusion chromatography (SEC) column. The individual N-RsbR1 and Prli42-GFP proteins were also applied to the same SEC column using the same buffer (**Figure 4.1**).

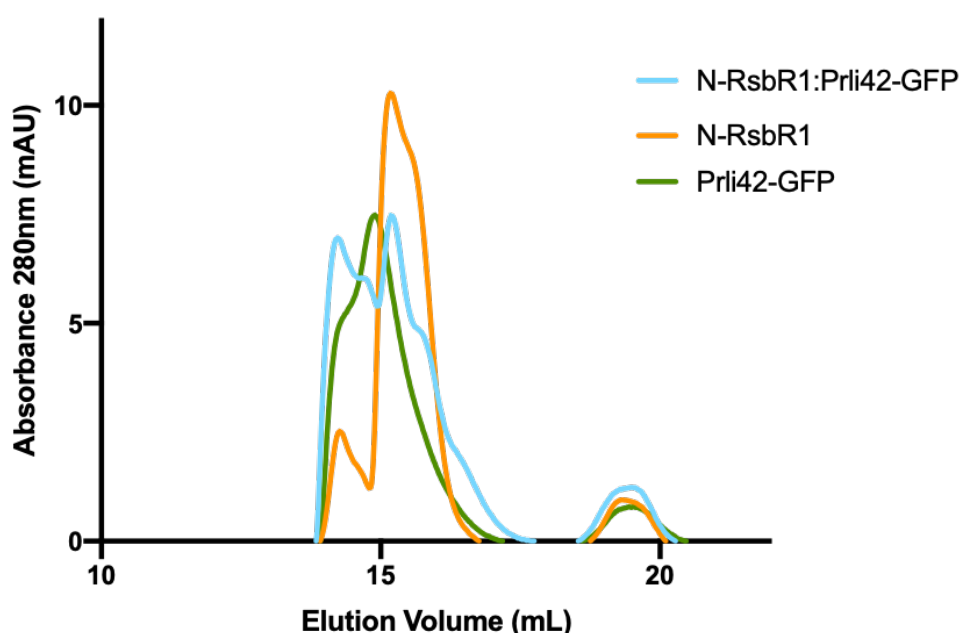


Figure 4.1. N-RsbR1 and Prli42 co-elution SEC chromatograms

Prli42-GFP (green) and N-RsbR1 (orange) were each run in isolation on a Superose 6 column. A mixture of N-RsbR1 and Prli42-GFP at a ratio of 1:1 was applied to the same column (cyan), however, there was little indication of a shift of the peak for either individual protein.

There were 2 peaks for N-RsbR1 and the majority of Prli42-GFP eluted between the two N-RsbR1 peaks. The N-RsbR1:Prli42-GFP sample had 3 peaks, and all the samples eluted around the same V_e . Monomeric Prli42-GFP has a MW of 32.9 kDa and dimeric N-RsbR1 has a MW_{app} of 35.8 kDa, hence a complex of Prli42-GFP and N-RsbR1 should have a MW of at least 68.7 kDa. However, no such peak was observed on the SEC, and since Superose 6 operates well as a separating resin between 70 and 30 kDa, it can be concluded that there is no interaction between the proteins under these experimental conditions. That said, the low protein

concentrations used might preclude the observation of a stable interaction and furthermore the N-RsbR1 binding site on Prli42 could be blocked by the presence of the GFP. Given these caveats and the lack of any rigorous quantification with this method, a fluorescence polarization assay was developed using a synthetic Prli42 peptide tagged with a small chromophore.

4.3. Interaction study of Prli42 and N-RsbRs by Fluorescence Polarization

4.3.1. *Bsu* RsbR NTDs purification

For this part of the project, expression plasmids for *B. subtilis* N-RsbRA, B and D were already available in the Lewis lab archives, with one missing expression plasmid for N-RsbRC. To complete the sets of expression plasmids, the reading frame of *B. subtilis* N-RsbRC was cloned into the pET28gg expression vector using the CIDAR MoClo cloning method (**Chapter II, section 2.1.1**). After the successful cloning of N-RsbRC, the integrity of all whole plasmids was verified by DNA sequencing. The proteins were not expressed as genetic fusions to affinity-tags and were purified by anion exchange chromatography using a Q Sepharose column (**Chapter II, section 2.3.1**) as for the purification of *L. monocytogenes* N-RsbR proteins. Following anion exchange chromatography, the protein of interest was separated further from contaminating proteins by SEC using either S75 or S200 depending on the sample MW and the column availability (**Chapter II, section 2.3.2**). Most proteins underwent a second anion exchange chromatography run, using a MonoQ column, to obtain the purest sample possible. Following the second anion-exchange step, a final SEC was performed for each of the proteins. With each purification step an SDS-PAGE was run to estimate the quality and the quantity of the purified sample. For clarity, only the final purification step will be displayed for each protein.

N-RsbRA purification

After the anion exchange (**Supplemental figures S13**), the gel-filtration chromatogram for N-RsbRA from an S75 16/60 column (**Figure 4.2**) showed two main peaks. The first peak at 63 mL corresponds to high molecular weight contaminants, as visualised on the SDS-PAGE gel; the second peak at 72 mL corresponds to pure N-RsbRA as seen on the SDS-PAGE gel. The sample obtained is approximately 98% pure and was snap-frozen in liquid nitrogen in small aliquots ready for use.

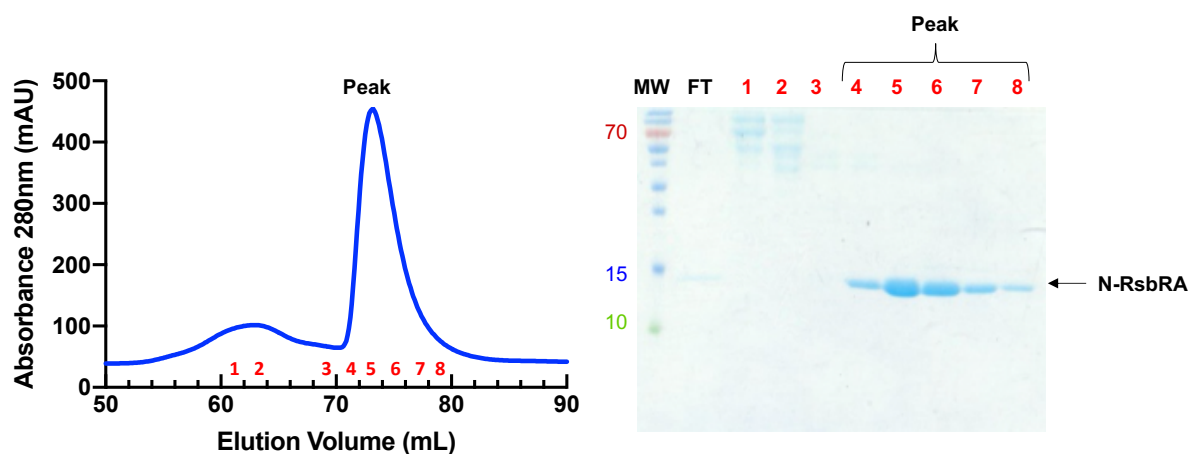


Figure 4.2. Purification of recombinant *B. subtilis* N-RsbRA

N-RsbRA SEC (left) and the corresponding 15% SDS-PAGE (right) with a stained protein ladder (MW), which displays the flow-through (FT) during the sample concentration and 2 fractions of a small elution peak at a V_e of 63 mL. Fractions 4-8 of peak 2 shows N-RsbRA that electrophoreses to approximately 15 kDa, the expected MW of N-RsbRA.

N-RsbRB purification

The N-RsbRB purification (**Figure 2.3**) shows a more contaminated sample from the anion exchange purification (**Supplemental figures S14**). N-RsbRB was subsequently purified further by size exclusion chromatography using an S75 16/60 column. Three peaks were observed; SDS-PAGE of appropriate fractions revealed that the first two peaks were aggregates and high molecular mass contaminants, and the third peak contained N-RsbRB, which eluted at an elution volume of 67.1 mL. The final pooled sample of N-RsbRB was approximately 95% pure, and this was also snap frozen for later use (**Figure 2.3**).

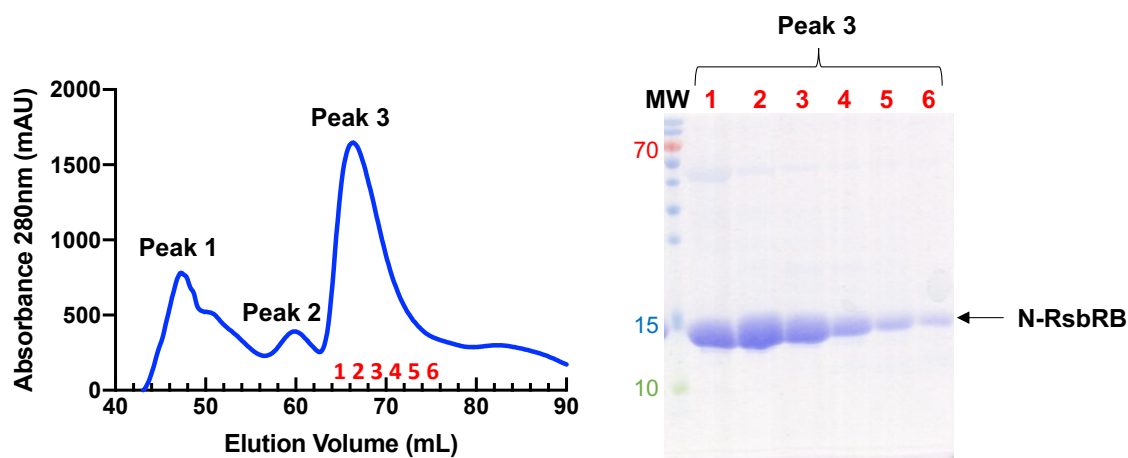


Figure 4.3. Purification of recombinant *B. subtilis* N-RsbRB

SEC of N-RsbRB using a S75 16/60 SEC column (left), that has at least 3 peaks. The 15% SDS-PAGE (right) with a stained protein ladder (MW) displays the fractions 1-6 of the third elution peak and N-RsbRB migrates to approximately 15 kDa.

N-RsbRC purification

The N-RsbRC purification (**Figure 2.4**) worked similarly to the purifications of other N-RsbR paralogues. Following anion exchange chromatography (**Supplemental figures S15**), the SEC elution volume of N-RsbRC was 73.3 mL from an S75 16/60 column. However, even at this point the electrophoretic purity of N-RsbRC was low, even after a high-resolution ion exchange chromatography, the downside of purifying untagged recombinant proteins.

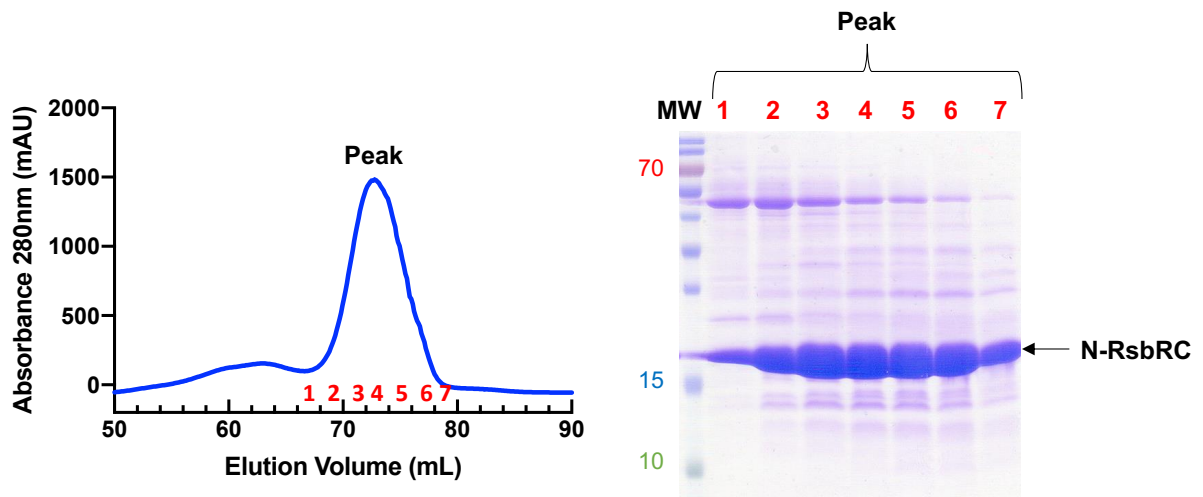


Figure 4.4. Purification of recombinant *B. subtilis* N-RsbRC

SEC of N-RsbRC using a S75 16/60 SEC column (left). Though the majority of the protein eluted in the main peak with an elution volume of 73.3 mL, the 15% SDS-PAGE (right) with a stained protein ladder (MW) reveals the main N-RsbRC fractions are mildly contaminated.

N-RsbRD purification

The N-RsbRD purification yielded a high quantity of recombinant protein at high purity after the first anion exchange chromatography (**Supplemental figures S16**). The table-top nature of the SEC of N-RsbRD (**Figure 4.5**) is consistent with a saturated detector, due to the high amount of N-RsbRD loaded. The V_e of N-RsbRD is thus not estimable but it is within the V_e range of the previously purified N-RsbR proteins.

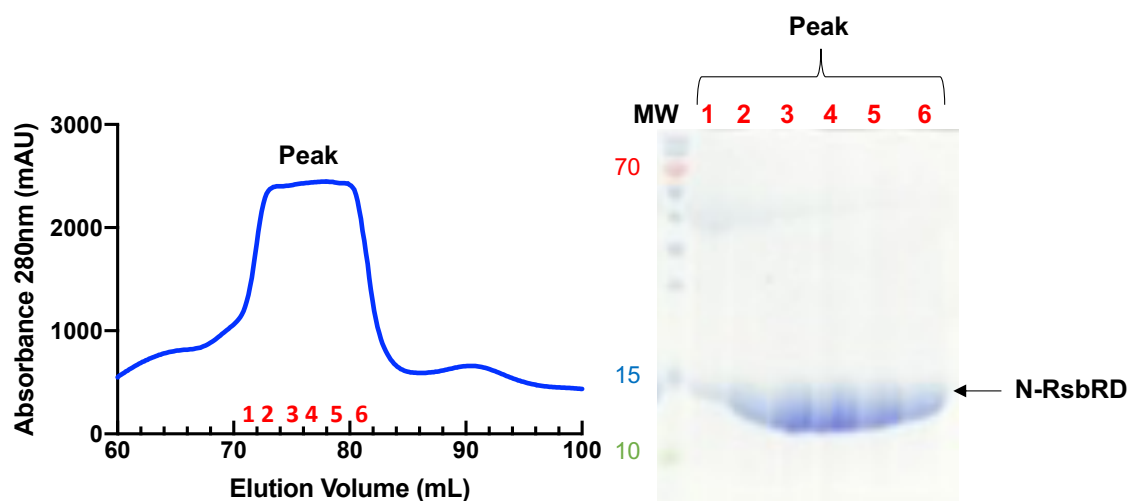


Figure 4.5. Purification of recombinant *B. subtilis* N-RsbRD

Size-exclusion chromatogram of N-RsbRD using a S75 SEC column (left) and the 15% SDS-PAGE (right) with a stained protein ladder (MW) displays the fractions of the elution peak and N-RsbRD protein migrates to approximately 15 kDa.

4.3.2. Determination of Prli42 interaction with N-RsbR proteins by fluorescence polarization

The principle of Fluorescence Polarization (FP) relies on Brownian movement of entities in solution (**Chapter II, section 2.8**). In this case there are two entities, the target protein (N-RsbRs) and their corresponding Prli42 peptides, each tagged with a TAMRA fluorophore. Alone, the protein has a slow Brownian movement due to its size while the peptide tumbles much faster as it is a smaller entity. The fluorophore does not disturb polarized light when rotating quickly, but when the peptide binds to the protein the fluorophore follows the Brownian movement of the protein which is much slower. Consequently, there is a loss of polarization. Using this principle, the interaction of the *L. monocytogenes* (*Lmo*) Prli42 against the N-RsbRs of *Lmo* (**Figure 4.6**) was measured and the data statistics are displayed in **Table 4.1**.

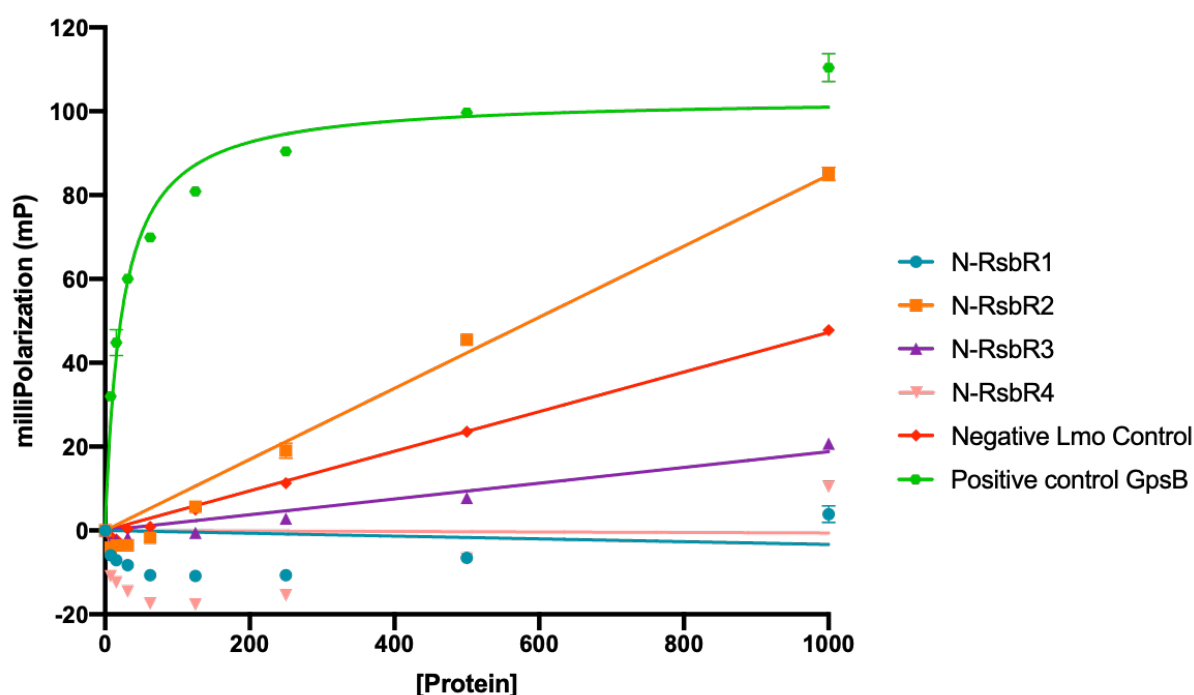


Figure 4.6. Fluorescence polarisation experiment to quantify the interaction between *Lmo* Prli42 and *Lmo* N-RsbRs.

The FP assay used a constant peptide concentration of 40 nm. Each data point is the mean of triplicate independent experiments and the error bars represent the standard deviation.

For this experiment, there was no positive control available for *Lmo* N-RsbRs or for Prli42. Therefore, a validated protein:ligand interaction partner was used as a positive control. GpsB, a divisome component, interacts with PBP2 (Cleverley et al., 2019) and this interaction has been studied extensively in the Lewis lab using bacterial genetics, two-hybrid, X-ray crystallography, isothermal titration calorimetry and fluorescence polarization (Cleverley et al., 2019). The GpsB:PBP2 positive control resulted in a clear sigmoidal and saturable binding event, with a K_d value of $27.88 \pm 2.27 \mu\text{M}$ (Figure 4.6), in the range of that measured ($20\mu\text{M}$) and published previously (Cleverley et al., 2019). The negative control, which is the GpsB protein with the *Lmo* Prli42 peptide, showed no indication of a sigmoidal binding curve, rather a straight line that indicates the data do not conform to a 1:1 binding model resulting in elevated K_d values. Therefore, the Prli42 peptide does not bind to GpsB. The data of the *Lmo* N-RsbRs against Prli42 share the same tendency as the negative control, displaying straight lines where there is no dose-dependent and saturable binding and have aberrant best-fit and K_d values (Table 4.1).

Table 4.1. Calculated Kd of the fluorescence polarization data for *Lmo* N-RsbR with *Lmo*Prli42.

Sample	Kd values (μM)
N-RsbR1	> 10,000
N-RsbR2	> 10,000
N-RsbR3	> 10,000
N-RsbR4	> 10,000
Negative control	> 10,000
Positive control	27.88 ± 2.27

To complement and extend the interaction studies of the *Lmo* N-RsbR proteins, the same experiment was performed using N-RsbR proteins and a Prli42 peptide from *B. subtilis*. The same positive control was used as above, and GpsB was also used for the negative control with the *Bsu*Prli42 peptide (**Figure 4.7** and **Table 4.2**). The experiments were run on the same day using the same plate; therefore, the positive control is the same as the one used for the *Bsu* proteins. The negative control showed no indication of saturable binding, just a linear response to increased protein concentration. The *Bsu* N-RsbR:Prli42 data also mostly show a linear response, and though they fit marginally better to a 1:1 binding model equation the calculated Kd values were all far in excess, for N-RsbRB and N-RsbRC, of what would be considered the normal range for protein:protein interactions (Kastritis et al., 2011), while no Kd values could be obtained for N-RsbRA and N-RsbRD.

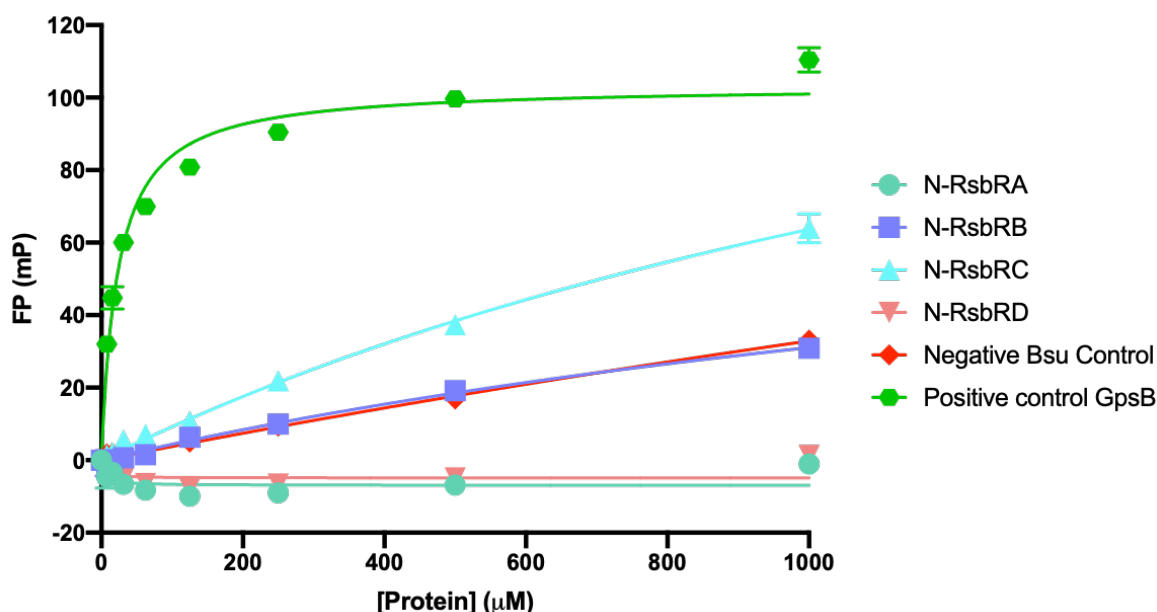


Figure 4.7. Fluorescence polarisation experiment to quantify the interaction between *Bsu* Prli42 and *Bsu* N-RsbRs.

The FP assay used a constant peptide concentration of 40nm. Each data point is the mean of triplicate independent experiments and the error bars represent the standard deviation.

Table 4.2. Calculated Kd values of the fluorescence polarization data for *Bsu* N-RsbR with *Bsu*Prli42.

Sample	Kd values (μM)
N-RsbRA	N/A
N-RsbRB	2199 ± 784.6
N-RsbRC	1943 ± 326.3
N-RsbRD	N/A
Negative control	6488 ± 1904
Positive control	27.88 ± 2.27

In order to eliminate the possibility that the low concentration of the peptide used in these experiments was insufficient to monitor binding to the target proteins, the same experiments were run with peptide at 40 μM , however, similar results were observed to that collected with peptides at 40 nM (**Figure 4.8 and Table 4.3**).

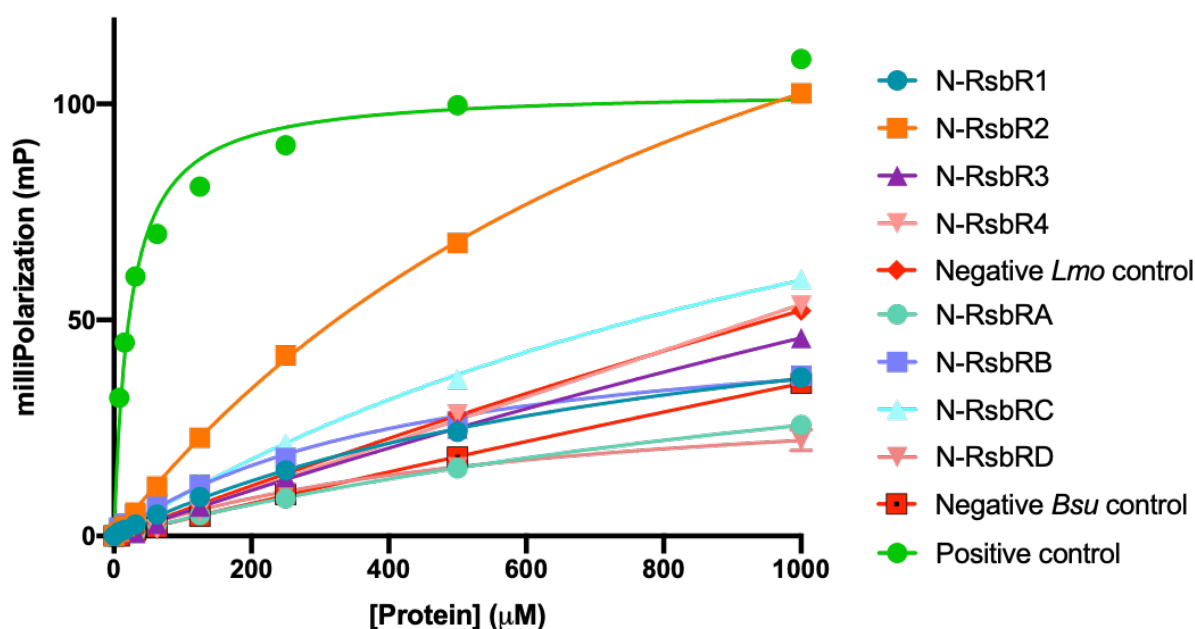


Figure 4.8. Interaction of the Prli42 miniprotein with the N-RsbR proteins by fluorescence polarisation

The FP assay used a constant peptide concentration of 40 μM . Each data point is the mean of the triplicate independent experiments and is fitted by non-linear regression.

Table 4.3. Kd values obtained for the FP experiments at 40 μ M

Sample	Kd values (μ M)
N-RsbR1	877.3 \pm 117.1
N-RsbR2	1012 \pm 67.98
N-RsbR3	5414 \pm 1675
N-RsbR4	45870 \pm 46535
Negative <i>Lmo</i> control	7025 \pm 605.9
N-RsbRA	1805 \pm 574.9
N-RsbRB	445 \pm 61.78
N-RsbRC	1408 \pm 15.70
N-RsbRD	651.7 \pm 241.6
Negative <i>Bsu</i> control	11808 \pm 6230
Positive control	23.07 \pm 1.28

The FP results clearly show that the interaction between N-RsbRs proteins for either *Lmo* or *Bsu* with their respective Prli42 peptides is non-existent. Indeed, the results obtained are clearly similar to the negative controls. The obtained Kd values are unrealistically high and the datasets do not fit a 1:1 binding model. Therefore, the Prli42 constructs used in these experiments do not interact *in vitro* with the N-RsbRs of either *Lmo* or *Bsu*.

4.4. Study of the ligand binding pocket of N-RsbRs

In the absence of evidence for the binding of Prli42 miniproteins to N-RsbR proteins *in vitro* the possibility that the N-RsbR domains bind instead to small molecule signal inducers is explored here. As demonstrated in **Chapter III**, the N-RsbR3 structure has three potential ligand binding pockets: one at the dimer interface and one in each monomer. Finding small molecule ligands within the structure raised questions as to whether these were physiologically relevant for stressosome function. Given the non-physiological nature of the crystallographic ligands (acrylic and formic acids), a set of common, biologically relevant small molecules were chosen for initial biophysical studies of ligand binding (**Table 2.8**). This set encompassed a set of ligands commonly found in the laboratory, such as amino acids, nucleotides, organic acids and cations.

An initial high-throughput ligand binding screen was developed using a plate reader based thermostability shift assay (TSA) (**Chapter II, section 2.7**). The thermostability of each N-

RsbR protein was tested against the ligand screen and changes to the observed melting temperature with the ligand compared to the protein alone were calculated. The results of these experiments are discussed below.

4.4.1. Study of the thermal stability of the N-RsbR proteins

Data on the thermostability of each of the apo N-RsbR proteins were obtained using two different methods. First, the melting temperature of each of the *Lmo* N-RsbR proteins was determined using CD spectroscopy (**Chapter II, section 2.5**), which can be estimated from the centre of the sigmoidal slope as the T_m represents the temperature at which equal amounts of the folded and unfolded states are present (Greenfield, 2006). T_m values of 77.6°C for N-Rsb1, 49.6°C for N-RsbR2, 49.7°C for N-Rsb3 and 48.5°C for N-RsbR4 were obtained from the CD thermal shift data (**Figure 4.9A**): N-RsbR1 is thus substantially more stable than its paralogues. These values were then compared to those obtained using a fluorescence-based TSA method (**Chapter II, section 9**) using SYPRO Orange® as the fluorophore. Here, the fluorescence of SYPRO Orange® increases with the hydrophobicity of the sample – as would occur when the hydrophobic core of proteins becomes exposed as a function of denaturation. The derivative of the fluorescence curves obtained by TSA were used to obtain the T_m (Grøftehaug et al., 2015). T_m values for N-RsbR1, N-Rsb3 and N-RsbR4 were obtained with the TSA method of 59.3°C, 60.6°C and 61.2°C, respectively (**Figure 4.9B**). For N-RsbR2, the derivative plots revealed a sharp increase in $-d/dT$ at 40°C, followed by a sustained increase in $-d/dT$, which is consistent with protein stability being markedly affected by the presence of the SYPRO Orange® dye; an initial rapid denaturation of N-RsbR2 is followed by its aggregation as a function of temperature (Kroeger et al., 2017). Though the absolute T_m values for the N-RsbR proteins are dissimilar between methods, despite the use of the same buffer for all experiments, the values obtained for each paralogue (except N-RsbR1 by CD and N-RsbR2 by TSA) by each method are quite similar, around 50°C by CD and about 60°C by TSA. This 10°C differential most likely reflects the use of different methodologies.

The same experiments were performed on the *Bsu* N-RsbR domains using CD spectroscopy (**Figure 4.10A**). N-RsbRA has a highly stable domain, with an estimated melting temperature of 74.2°C, similar to its *Lmo* paralogue, N-RsbR1 (77.6°C). N-RsbRB and N-RsbRD also share similar thermostabilities, 47.2 °C and 45.6°C respectively, which are also similar to their *Lmo* paralogues' stability N-RsbR2 (49.6°C) and N-RsbR4 (48.5°C). The *Bsu* paralogue that stands out is N-RsbRC, which has a thermostability of 63.6°C, some 14°C higher than that of its *Lmo* paralogue, N-RsbR3 (49.7°C).

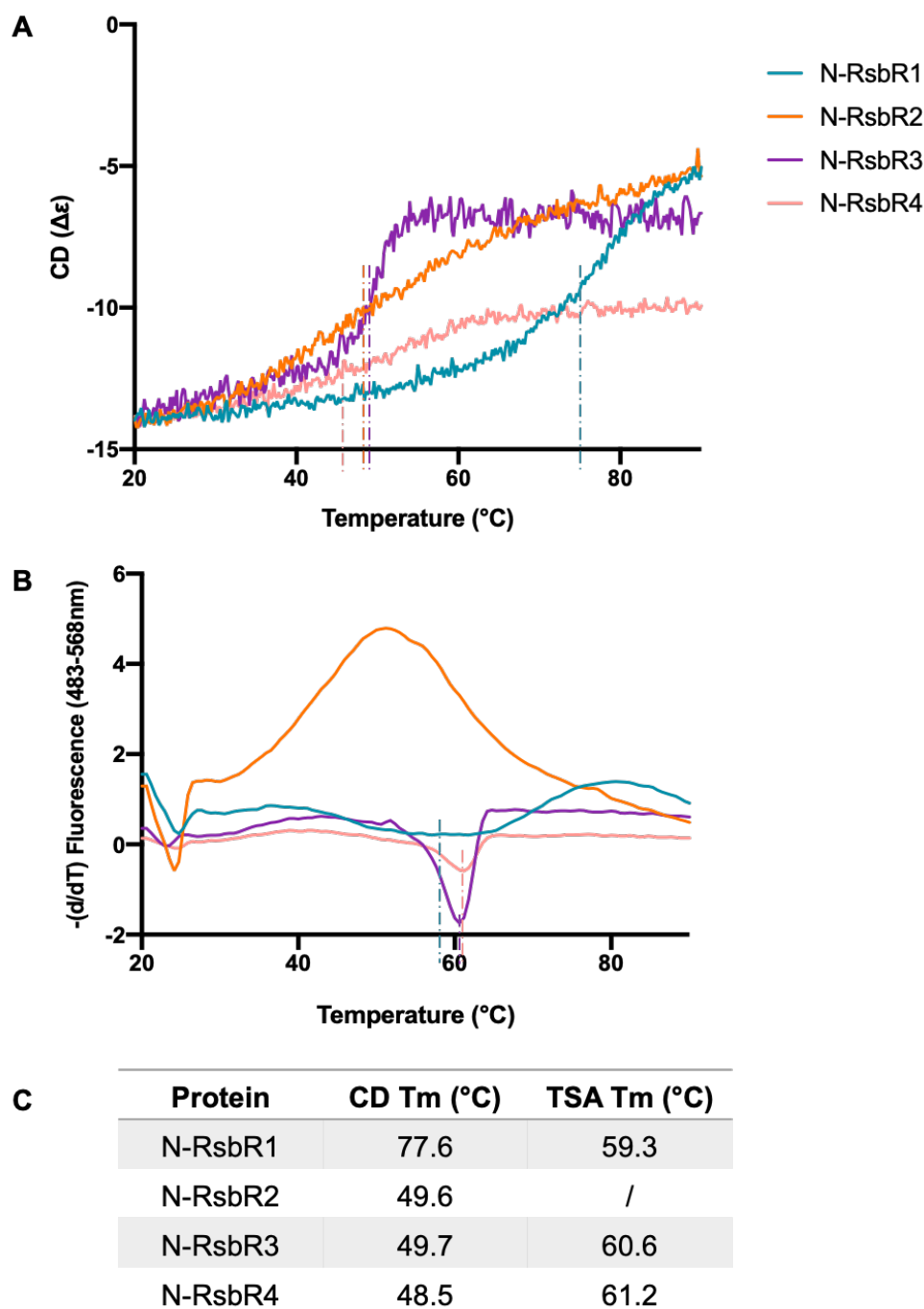


Figure 4.9. Thermostability of *Lmo* N-RsbRs using 2 different methods

A. Melt curves obtained using CD spectroscopy. The Tm is read at the middle of the sigmoidal slope (dotted line). **B.** Derivative curves from the fluorescence data obtained by TSA. The Tm is read at the local minimum (dotted line). **C.** The Tm ($^{\circ}\text{C}$) values obtained by each method are summarised in the table.

The Tm values of the *Bsu* N-RsbRs obtained by CD were then compared to those obtained by TSA (**Figure 4.10B**). Tm values of 40 $^{\circ}\text{C}$ for N-RsbRA and N-RsbRB, 64.7 $^{\circ}\text{C}$ for N-RsbRC and 41.2 $^{\circ}\text{C}$ for N-RsbRD were obtained. In the case of N-RsbRB, N-RsbRC and N-RsbRD similar values were obtained between the two methodologies, but in the case of N-RsbRA there was a substantial difference between the two methods, 74.2 $^{\circ}\text{C}$ vs 40 $^{\circ}\text{C}$. However, neither N-

RsbRA and N-RsbR2 data set could be described as ideal as there is no sharp peak in the TSA derivative curve and similarly no obvious sigmoid with clear initial and final plateaus for the CD data.

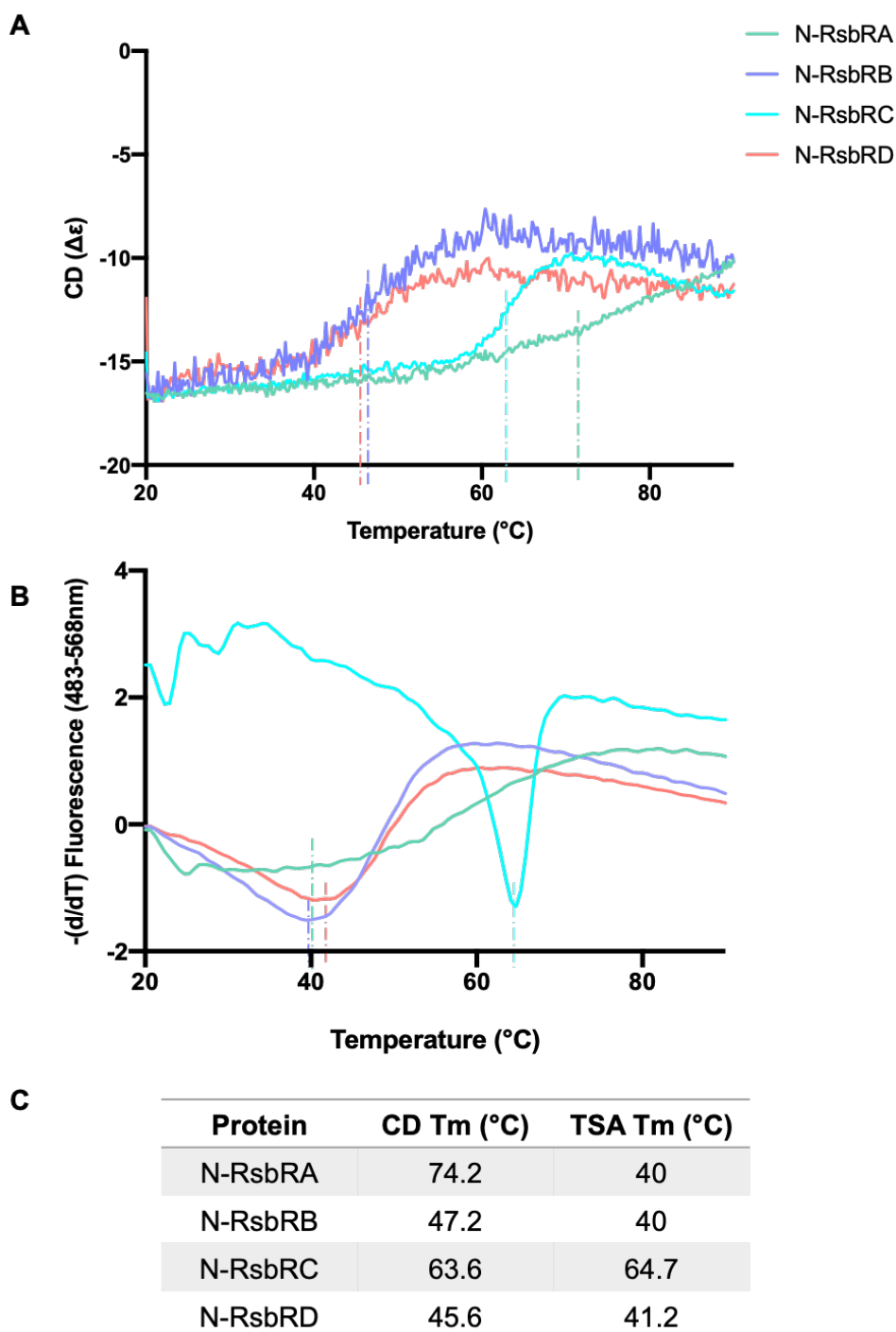


Figure 4.10. Thermostability of *Bsu* N-RsbRs using 2 different methods

A. Melt curves obtained using CD spectroscopy. The Tm is read at the middle of the sigmoidal slope (dotted line). **B.** Derivative curves for the fluorescence data obtained by TSA. The Tm is read at the local minimum (dotted line). **C.** The Tm (°C) values obtained by each method are summarised in the table.

One clear advantage the TSA method has over that of CD, given the constraints of the equipment available during the course of these studies, is that the TSA method is amenable to high-throughput as it can be conducted in a plate reader capable of reading a 384-well plate. By contrast, the Jasco 1100 CD spectrometer in the Faculty of Medical Sciences is at least 20 years' old and cannot be run in a high-throughput manner; there is a single cuvette holder within the instrument that means only one experiment can be conducted at a time. Therefore, the N-RsbRs' ability to interact with a set of ligands was screened using the TSA method with the control values obtained by TSA as baselines.

4.4.2. Effect of the nucleic acid on the N-RsbRs' thermostability

The impact of a range of nucleotides and nucleoside acids on the stability of the N-RsbR proteins was screened using TSA (**Figure 4.11**). For clarity, the compounds with little or no effect have been removed but the data can be found in **Supplemental figures S17**. The raw data of compounds with an effect can be also found in the **Supplemental figures S18**. Among the eleven nucleotides and nucleoside acids tested, the stability of the N-RsbRs was impacted only by dATP, ADP and adenine.

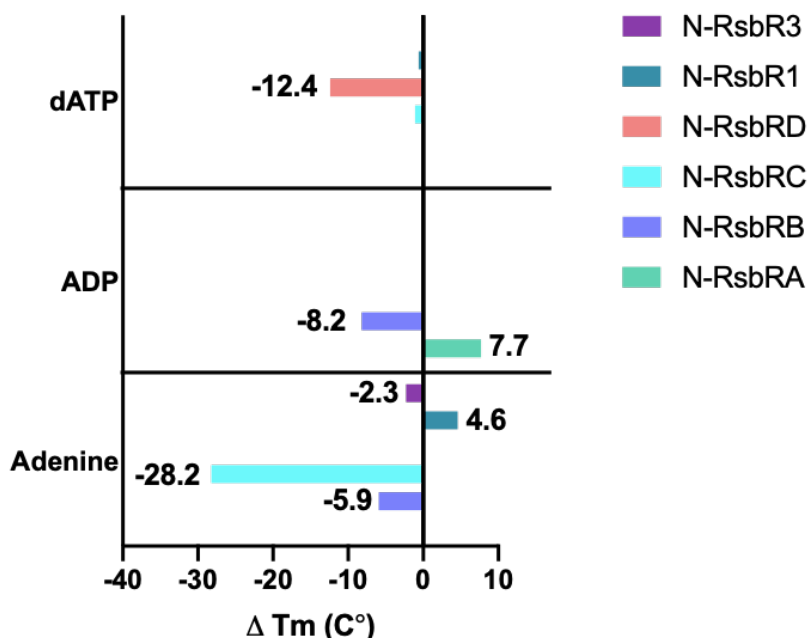


Figure 4.11. Effect of the adenine-containing compounds on N-RsbR thermostability.

The T_m obtained with compounds containing adenine were subtracted from the control T_m of the protein alone. The resulting T_m difference (ΔT_m) was plotted and the ΔT_m values are displayed next to the bars.

However, the effect of these compounds is not the same on different N-RsbRs: dATP decreased the N-RsbRD T_m by 12.4°C and showed no effect on the remaining N-RsbRs. ADP increased the N-RsbRA T_m by 7.7°C while it caused a decrease of the N-RsbRB T_m by 8.2°C. Adenine had a substantial negative impact on N-RsbRC thermostability, with a difference of T_m of -28.2°C, and also negatively affected – though not to the same extent as N-RsbRC – the T_m of N-RsbRB and N-RsbR3 whereas it stabilized N-RsbR1 by 4.6°C.

4.4.3. Effect of amino acids on the N-RsbRs' thermostability

The effect of amino acids on the N-RsbR domains was also assessed (**Figure 4.12**). Among the 23 amino acids tested (the 20 natural amino acid plus L-argininamide, cystine and D-alanine), tryptophan, histidine, glutamic acid and aspartic acid have either a positive or a negative effect on the N-RsbRs; for instance, tryptophan stabilized N-RsbRA by over 18°C whereas histidine destabilized N-RsbR3 and N-RsbRC each by over 20°C. L-Argininamide is a variant of L-arginine that has been condensed with ammonia and has a terminal amide group instead of a guanidyl group. Arginine had little or no effect on the thermostability of the N-RsbRs (**Supplemental figures S19 and 20**). By contrast, L-argininamide increased the T_m of N-RsbR1 by 25.9°C and decreased the T_m of N-RsbRC and N-RsbR3 by 14.7 and 10.6°C, respectively. Histidine had a similar effect as L-argininamide on the N-RsbRs, by increasing the T_m of N-RsbR1 by 18.3°C but decreasing the T_m for N-RsbC and N-RsbR3 by 22.9°C and 24.7°C, respectively. L-argininamide has an amide group and histidine has an imidazole group, and they both have H-bond donors and acceptors. Aspartic acid had opposite effects only on N-RsbRB and N-RsbRC, by increasing the stability of N-RsbRB of 4°C, while decreasing the stability of N-RsbRC by 15.7°C. Tryptophan only enhances the T_m of N-RsbRA, by 18.3°C, and glutamic acid only increased the T_m of N-RsbR1 by 5.2°C. In the case of N-RsbRC, six other amino acids had small negative effects on its stability, including cysteine, glutamine, asparagine, serine, methionine and phenylalanine.

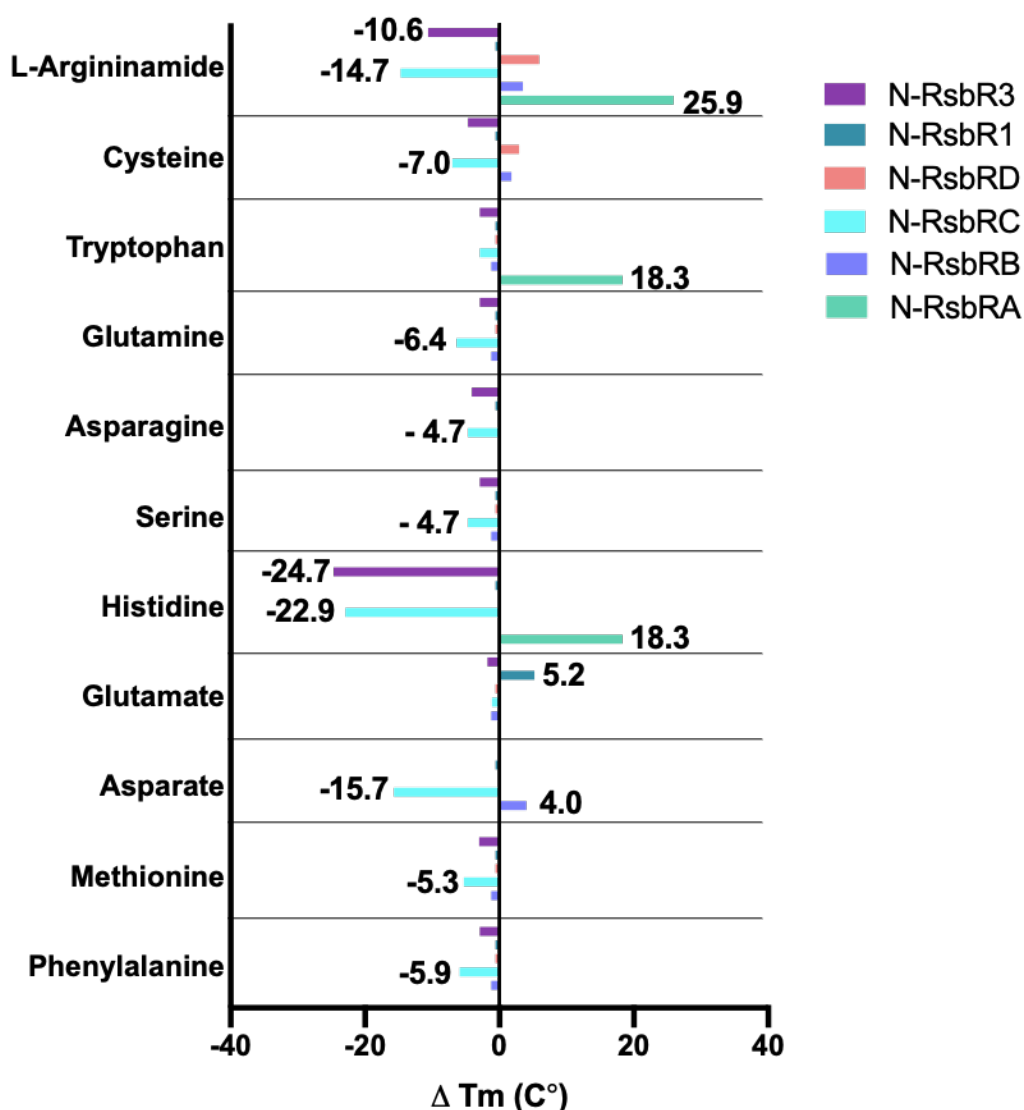


Figure 4.12. Effect of amino acids on N-RsbRs thermostability.

The Tm obtained with amino acids were subtracted from the control Tm of the protein alone. The resulting Tm difference (ΔT_m) was plotted and the ΔT_m values are displayed next to the bars.

4.4.4. Effect of sugars on the N-RsbRs' thermostability

The effect of some sugars on the N-RsbR domains' stability was also assessed, including sucrose, glucose, α -lactose and xylitol. No effects on N-RsbRA, N-RsbRB, N-RsbRD and N-RsbR1 were observed (**Supplemental figures S21**). The most impacted protein in this series of experiments was N-RsbR3 in the presence of D-fructose 6-phosphate, where a Tm decrease of 25.9°C was recorded. Secondary fluorescence peaks for N-RsbR3 were observed in presence of α -lactose and xylitol. The second peak was obtained at -15.9°C with α -lactose and at -25.9°C with xylitol.

4.4.5. Effect of chemical compounds on N-RsbRs' thermostability

Finally, the effect of some compounds found commonly in labs were tested (**Figure 2.13**). The T_m of N-RsbRA increased by 14.1°C in the presence of potassium phosphate and 15.9°C in the presence of ammonium citrate. By contrast, N-RsbRA lost stability of approximately 14°C in the presence of ammonium tartrate, ammonium metavanadate and ammonium formate. With N-RsbRB, 2 compounds had a negative effect on T_m : -8.8°C with diethanolamine and -15.9 with zinc acetate. There were 3 compounds with a positive effect on N-RsbRB's stability, which are magnesium sulphate and magnesium formate both by 10.6°C , and calcium acetate by 8.9°C . N-RsbRC is the protein with the greatest number of negative impacts on its stability; thiamine, pyrodoxine, ammonium formate, malonic acid, DL-malic acid and DL tartaric acid all reduced the T_m by between -26.4 and -39.4°C . No compounds in the screen increased the stability of N-RsbRC. The stability of N-RsbRD increased with magnesium sulphate, magnesium formate and calcium acetate, while it decreases with diethanolamine and zinc acetate, which is similar to the effects seen on N-RsbRB. For N-RsbR1, there were only positive impacts on its thermostability, when in the presence of ammonium citrate (10.5°C), magnesium formate (10.0°C) and potassium phosphate (10.0). N-RsbR3, in contrast to N-RsbR1, only had negative thermostability effects when in the presence of ammonium citrate (-17.0) and potassium phosphate (-5.3°C).

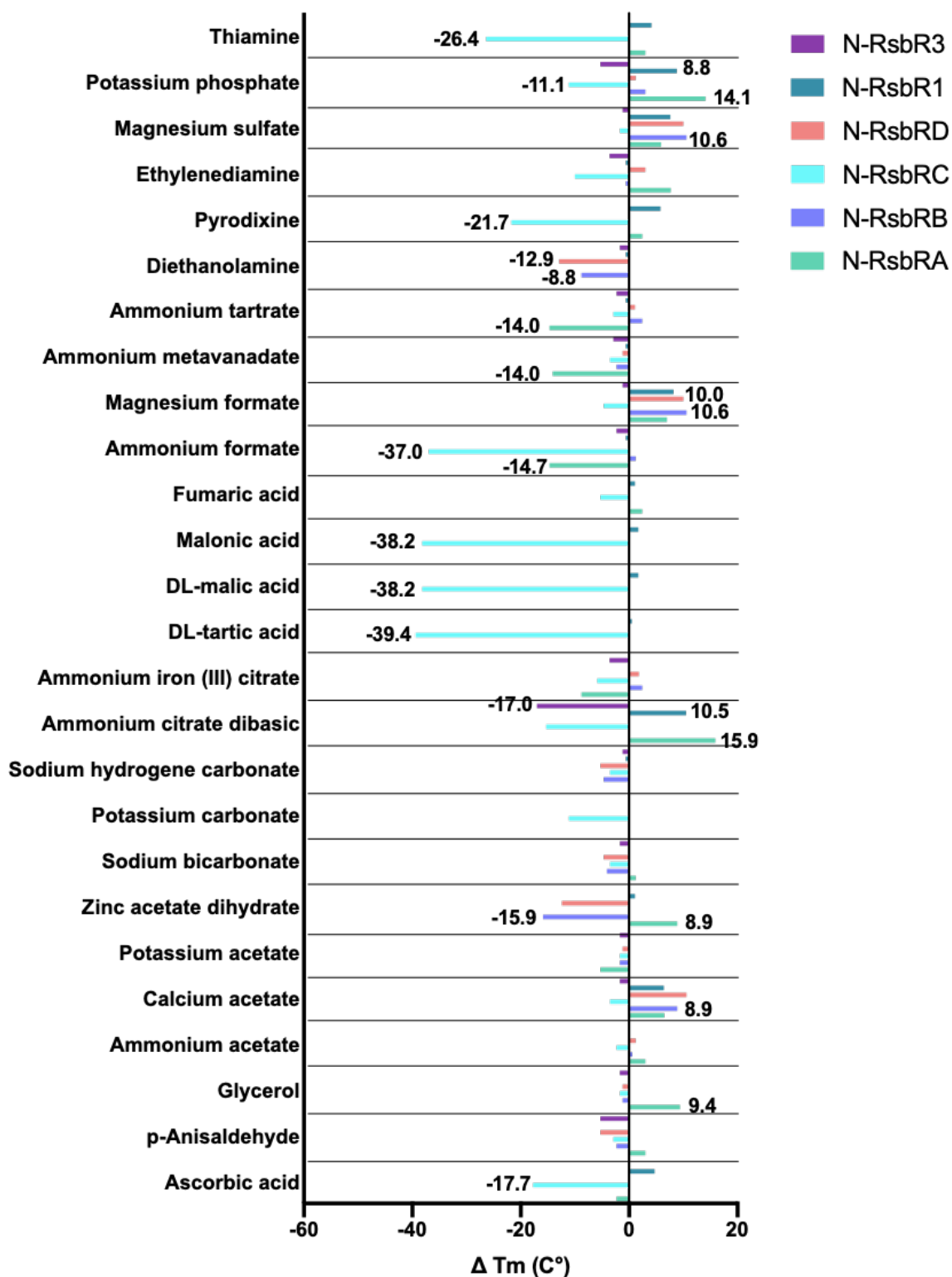


Figure 4.13. Thermostability shift assay of the N-RsbRs with chemical compounds.

The T_m obtained with various chemical compounds were subtracted from the control T_m of the protein alone. The resulting T_m difference (ΔT_m) was plotted and the ΔT_m values are displayed next to the bars.

4.5. N-RsbR3 crystal soaking with FragLites

As the previous assays did not identify unique ligands for the N-RsbR proteins, a new set of compounds were explored for binding to give insight into potential ligands for these proteins and to identify potential inhibitory compounds. The compounds used were a screen of halogenated drug-like fragment compounds used in drug discovery studies (Wood et al., 2019). As acrylic acid and formate ligands from the crystallisation conditions and cryoprotectant, respectively, were found in the N-RsbR3 crystal structure, N-RsbR3 crystals from a different crystallisation condition (0.2M lithium chloride, 0.1 M Tris pH 8.0 and 20% w/v PEG6000) were used to avoid filling the ligand binding pockets. These crystals were then used for soaking with the FragLite library of 31 compounds, which were each soaked into different crystals; from the 31 soaked crystals, seventeen datasets were obtained. This work was done in collaboration with Dr Arnaud Basle, the NSBL facility manager. The missing coverage is due to damage to crystals and the loss of diffraction due to the soaking with the FragLite compounds. The datasets were collected at the Diamond Light Source at 0.898Å, a wavelength that is suitable for obtaining anomalous scattering signals from both bromine and iodine (Figure 4.14).

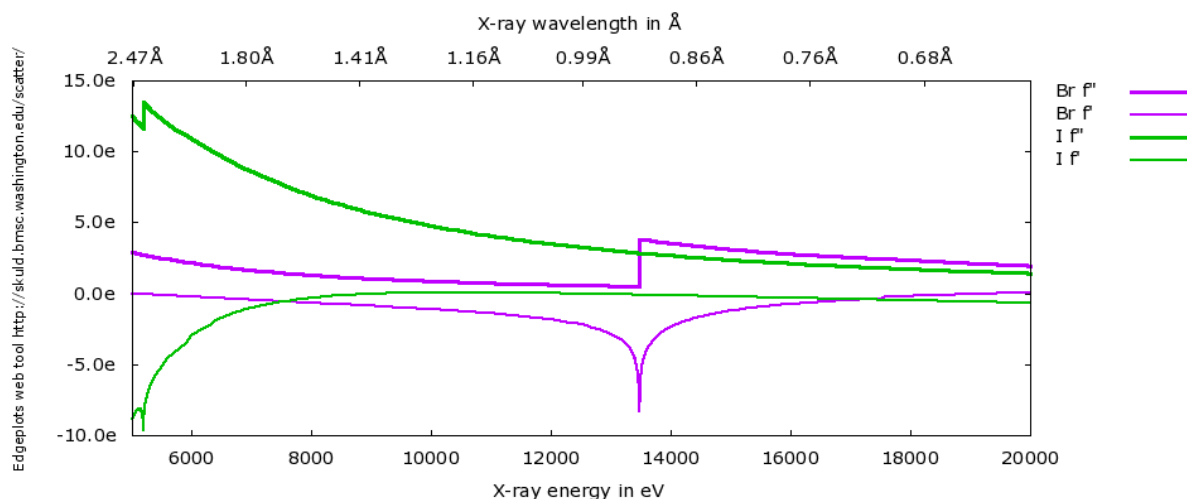


Figure 4.14. Elemental absorption edges of bromine and iodine

The plot shows the X-ray energy (eV) and X-ray wavelength (Å) for bromine in pink and iodine in green. The data were collected at 0.898Å which is suitable for both anomalous scatterers (f'' in bold). The plot was generated using the Biomolecular Structure center webserver (http://www.bmsc.washington.edu/scatter/AS_form.html).

The datasets were processed using standard procedures, their structures solved by molecular replacement and the presence of a FragLite in the solved crystal structures was first checked with the Find Ligand tool in Coot (Emsley and Cowtan, 2004) then manually checked.

Out of the seventeen datasets, three have potential binding candidates (**Table 4.4**). The data statistics for these data can be found in **Table 4.5**.

Table 4.4. Potential binding FragLites.

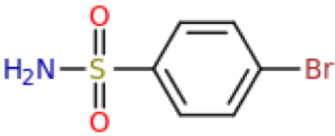
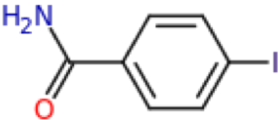
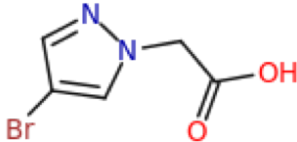
FragLite	2D structure	SMILES code
8		<chem>BrC1=CC=C(S(N)(=O)=O)C=C1</chem>
12		<chem>IC1=CC=C(C(N)=O)C=C1</chem>
27		<chem>BrC1=CN(CC(O)=O)N=C1</chem>

Table 4.5. Data collection statistics for N-RsbR3 with FragLites

The values in parentheses correspond to the highest resolution outer shell

Data collection	N-RsbR3/FG8	N-RsbR3/FG12	N-RsbR3/FG27
Date	08/10/2020		
Type of data collection	MR	MR	MR
Source	I03	I03	I03
Wavelength (Å)	0.898Å	0.898Å	0.898Å
Space group	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁
Cell dimensions			
a, b, c (Å)	50.73, 82.21, 82.70	50.01, 81.81, 82.85	50.85, 81.61, 82.79
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90
Resolution (Å)	43.17 – 2.38 (2.30 - 2.30)	41.42 – 2.10 (2.16 - 2.10)	41.40 – 2.0 (2.05 – 2.00)
No. of measured reflections	210,770 (21,056)	273,924 (22,062)	319,846 (22,885)
No. of unique reflections	15,975 (1,524)	20,525 (1,660)	23,994 (1,755)
Multiplicity	13.2 (13.8)	13.3 (13.3)	13.3 (13.0)
CC _{1/2}	0.999 (0.59)	0.999 (0.89)	0.999 (0.617)
I/(σ)I	10.3 (1.4)	14.2 (2.3)	12.8 (1.5)
R _p im (all I+ and I-)	0.044 (0.60)	0.030 (0.353)	0.034 (0.531)
Completeness (%)	100.0 (100.0)	100.0 (100.0)	100.0 (100.0)
CC _{1/2} anom	0.110 (0.010)	-0.016 (-0.001)	0.176 (-0.005)
Anomalous multiplicity	7.1 (7.2)	7.1 (6.9)	7.1 (6.8)

For the three datasets, partial electron density consistent with the bound FragLite was found. As the ligands each contain a halogen, an anomalous difference map was generated using CCP4i (Winn et al., 2011). Each of the anomalous difference maps were superimposed with the 2mFo-DFc electron density maps. Partial electron densities were observed in the intramolecular ligand binding site for each of the three FragLites, however, the anomalous difference maps only showed an anomalous signal for FG8 which was also bound at the dimer interface (**Figure 4.14**). The FG8 is found at the same place as the acrylic acid that was found within the N-RsbR3 crystal structure. This observation indicates that the dimer interface is potentially a ligand binding pocket relevant for the stress sensing activity, and ought to be the focus of future studies.

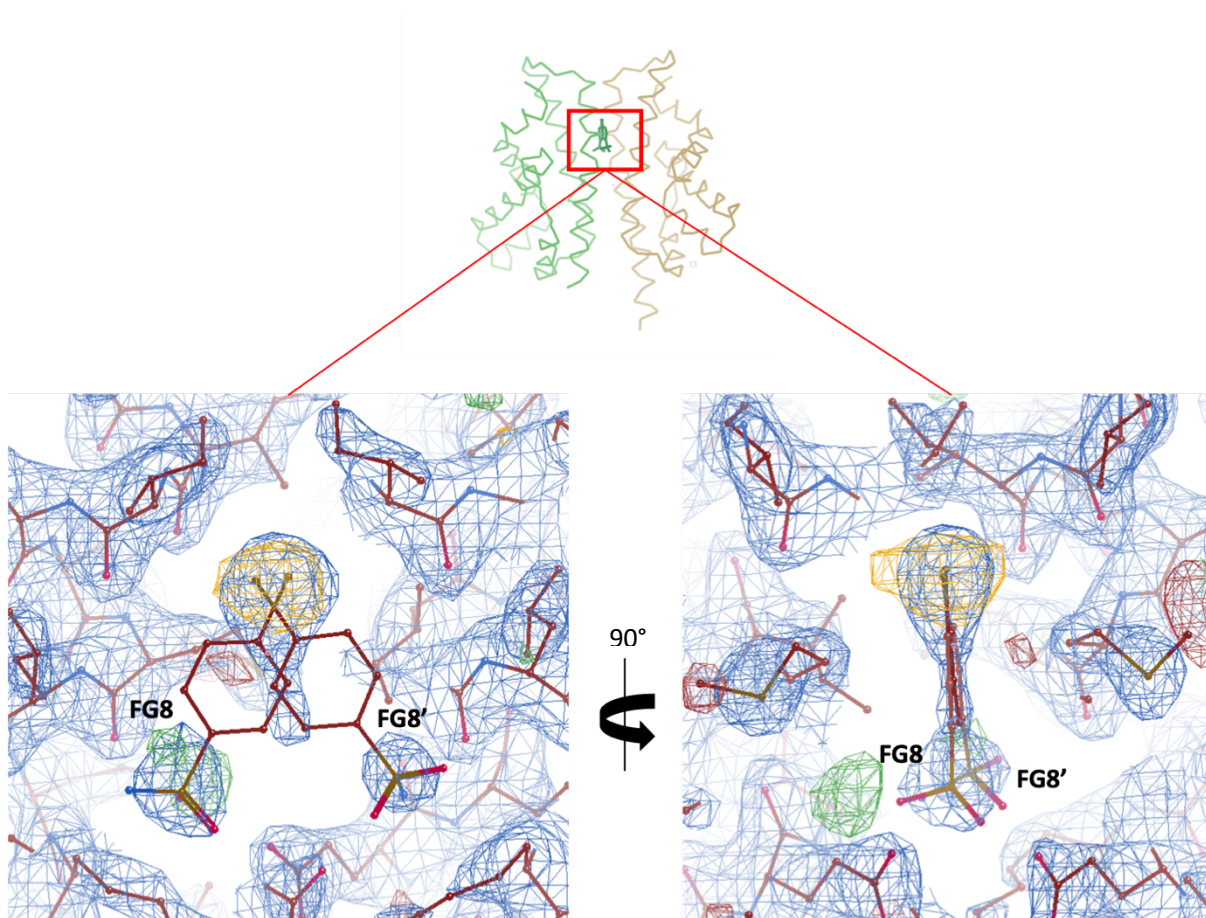


Figure 4.15. Electron density maps of N-RsbR3 dimer interface with FG8

The experimental 2mFo-dFc electron density map shown in blue, the difference map Fo-Fc shown in green/red and the anomalous difference map coloured in orange of the N-RsbR3 soaked with FG8. The FG8 ligand was modelled at the dimer interface in alternative conformations (FG8/FG8') as two density blobs where observed.

4.6. Conclusions

Prli42 and N-RsbRs interaction

The interaction of Prli42 with RsbR1 has been proposed and modelled previously (Impens et al., 2017; Williams et al., 2019) and is discussed in depth in **Chapter I, section 1.6**. No *in vitro* binding studies or biophysical analyses had been carried out by Impens or Williams. Fluorescence polarization experiments were used to quantify the affinity of the interaction between Prli42 and N-RsbR proteins. However, no interaction was observed between Prli42 and any N-RsbR protein from either *L. monocytogenes* or *B. subtilis* with these *in vitro* experiments. The missing roof of Prli42 binding might also be explained by the fact that several Prli42 miniproteins might bind to the stressosome in cooperation. Indeed, the stressosome might be engulfed in the membrane such as viruses. And the engulfment of the stressosome might involve other partner binding. Therefore single protein-protein interaction might not be significative. Nevertheless, our consortium partners, Prof. Dr. Francisco Garcia del Portillo and Charlotte Dessaux, in Madrid, Spain, studied the impact of Prli42 on the stress sensing activity of the *L. monocytogenes* stressosomes, but their results showed that Prli42 mutations had no effect on σ^B activation. To extend this interaction study, our partners in Ede, Netherlands, Dr. Marjon Wells-Bennik and Claire Yeak, studied the effect of Prli42 on *B. subtilis* σ^B activation, and also noted little or no effect of Prli42 mutations. More recently, our partners studied the subcellular localization of the stressosome (Dessaux et al., 2020), which revealed that the stressosome complex are mostly found in the cytosolic fraction, with or without stress condition, and is not membrane-associated as implied by the Prli42 interaction reported by Impens and Williams. These recent observations are consistent with those of Marles-Wright, who also showed that stressosomes in *B. subtilis* are no membrane-associated (Marles-Wright et al., 2008). These results all complement our observation that the stressosome does not interact *in vitro* with the membrane-anchored miniprotein, Prli42. Therefore, Prli42 does not interact with N-RsbR proteins, does not sequester stressosomes at the membrane and does not have any effect on the stress sensing activity of the stressosome.

Ligand binding study

The thermostability shift assay was used to develop a high-throughput assay to screen a wide range of compounds to identify ligands with common effects on the N-RsbR proteins. However, similar effects on N-RsbR proteins' thermostability were not observed in the presence of a range of ligands. This observation might be consistent with each N-RsbR having a unique sensory activity, linked to their divergent sequences even though they share the same

structural features. Due to time constraints and the Covid-19 impact, the ligand binding study could not be concluded but it does open up new avenues of research, including *in silico* drug design to determine what type, size and shape of ligand can be bound.

FragLites binding investigation

The FragLites identified as potential binding partners for N-RsbR3 should be followed up with further experiments. Unfortunately, due to time constraints and the Covid-19 impact, it was not possible to investigate the FragLites leads any further. A high throughput screen such as X-Chem might help to further explore potential ligands (Krojer et al., 2017). Finally, a mutational analysis of the N-RsbR3 domain might also identify residues important for ligand binding and signal transduction such that a molecular mechanism for stress perception of the stressosome can be described.

Chapter V: Native stressosome and variant stressosome study

5.1. Introduction

The stressosome is the supramolecular machinery comprised of 2 main proteins, RsbR and RsbS, which together sequester the RsbT kinase (**Chapter I, section 1.3.1**). Upon environmental stress, the stressosome releases RsbT, which in turn activates the signalling cascade and partner-switching pathway that leads to the release of σ^B (Akbar et al., 1997; Benson and Haldenwang, 1993; Boylan et al., 1993a). Both the *B. subtilis* and the *L. monocytogenes* stressosomes have been studied *in vitro* using reconstituted recombinant protein complexes, using mainly RsbS and RsbR1/A (Kwon et al., 2019; Marles-Wright et al., 2008; Williams et al., 2019). An overall description of the structure and genomics of the stressosome is reviewed in **Chapter I, section 1.4.2**. All *in vitro* studies of the stressosome reported to date have focussed on the complex including only the main RsbR paralogue, which is found in the σ^B operon alongside RsbS and RsbT (**Chapter I, section 1.3**). This RsbR paralogue is RsbR1 in *L. monocytogenes* and RsbRA in *B. subtilis*. The role of the remaining RsbR paralogues in the stress sensing activity of the complex remains unclear. The first goal of this Chapter is to understand the composition of the native complex as purified from *B. subtilis* cells. The second goal is to understand if all the isolated *L. monocytogenes* paralogues can reconstitute the complex in the presence of the scaffolding protein RsbS as has already been demonstrated in *B. subtilis* (Delumeau et al., 2006).

Part A: *B. subtilis* native stressosome complex composition

The native stressosome complex from *B. subtilis* was studied to better understand its native composition *in vivo* and for possible structural and functional studies. This part of the study was performed in collaboration with our PATHSENSE consortium partners at Dundee University, Scotland, with Prof. Dr. Nicola Stanley-Wall, Manisha Pandey and Diana Gudynaite. A pull-down approach was first tried, using a modified N-RsbRA on which one of three different His-tags designs was added to the chromosomal *rsbRA* coding sequence (**Chapter II, section 2.2.4., Figure 2.5**). One His-tag was incorporated at the N-terminus of RsbRA, immediately after the initial methionine, such that the sequence of RsbRA started as: M-**HHHHHH**-MSNQTV...; the two other His-tags were incorporated into loop 4 – which is invariably disordered in the crystal structures of N-RsbRA paralogues – using two different tag designs. The first of these internal His-tags was a continuous His₈ tag after Pro105, in the centre of loop 4, to result in ...KRLP-**HHHHHHHH**-DQES...

The second internal His-tag was designed to span Pro105 to result in a sequence ...KRL-**HHHH**-P-**HHHH**-DQES... The tag designs were cloned by our partners in the WT *B. subtilis* 168 strain using pminiMAD with appropriate flanking sequences and transformed into the *B. subtilis* strain. Successful clones for two out of three designs were obtained: the two internal loop 4 His-tags (the continuous His-tag, strain NRS5621 and the two part His-tag strain, NRS5623). By contrast, the N-terminal modification was lethal for reasons that are not understood. To confirm these changes to the RsbRA sequence did not impact the stress-sensing activity of the stressosome, the effect of the variant RsbRAs on σ^B activity was also investigated by our partners in Dundee. A *lacZ* reporter at the σ^B -dependent *ctc* gene was used to monitor σ^B activity in the variant *B. subtilis* strains generated. According to the resultant β -galactosidase assays the activity of σ^B was not impacted by the intramolecular His-tags in RsbRA and similar σ^B activation levels were observed in these two mutants in comparison to the unmodified RsbRA-carrying *B. subtilis* strain (personal communication, N. Stanley-Wall).

5.2. Purification and visualisation of the native stressosome complex from *B. subtilis*

First, strain NSR5623 was used for a large-scale cell culture growth of 12 L. The cells were harvested at stationary phase and lysed using a French press at 30 kpsi. The lysate was clarified by centrifugation and the lysate supernatant was passed over a Ni-NTA column (**Chapter II, section 2.3.1**) and the bound proteins were eluted using a buffer containing 500 mM imidazole (**Figure 5.1A**). As imidazole absorbs light at 280 nm, the UV chromatogram rises with increasing imidazole concentration and therefore elution peaks in this instance were hidden by the imidazole absorption. Nonetheless, a small elution bump can be seen when the chromatogram is zoomed-in (**Figure 5.1B**). As the UV chromatogram shows, the concentration and amount of the complex is low, so the collected fractions were each concentrated to 100 μ L before analysis by SDS-PAGE (**Figure 5.1C**). The complex cannot be visualised in the elution fraction, but faint bands can be observed in the concentrated samples. Bands at the MWs corresponding to RsbS and the RsbRs can be seen on the SDS-PAGE.

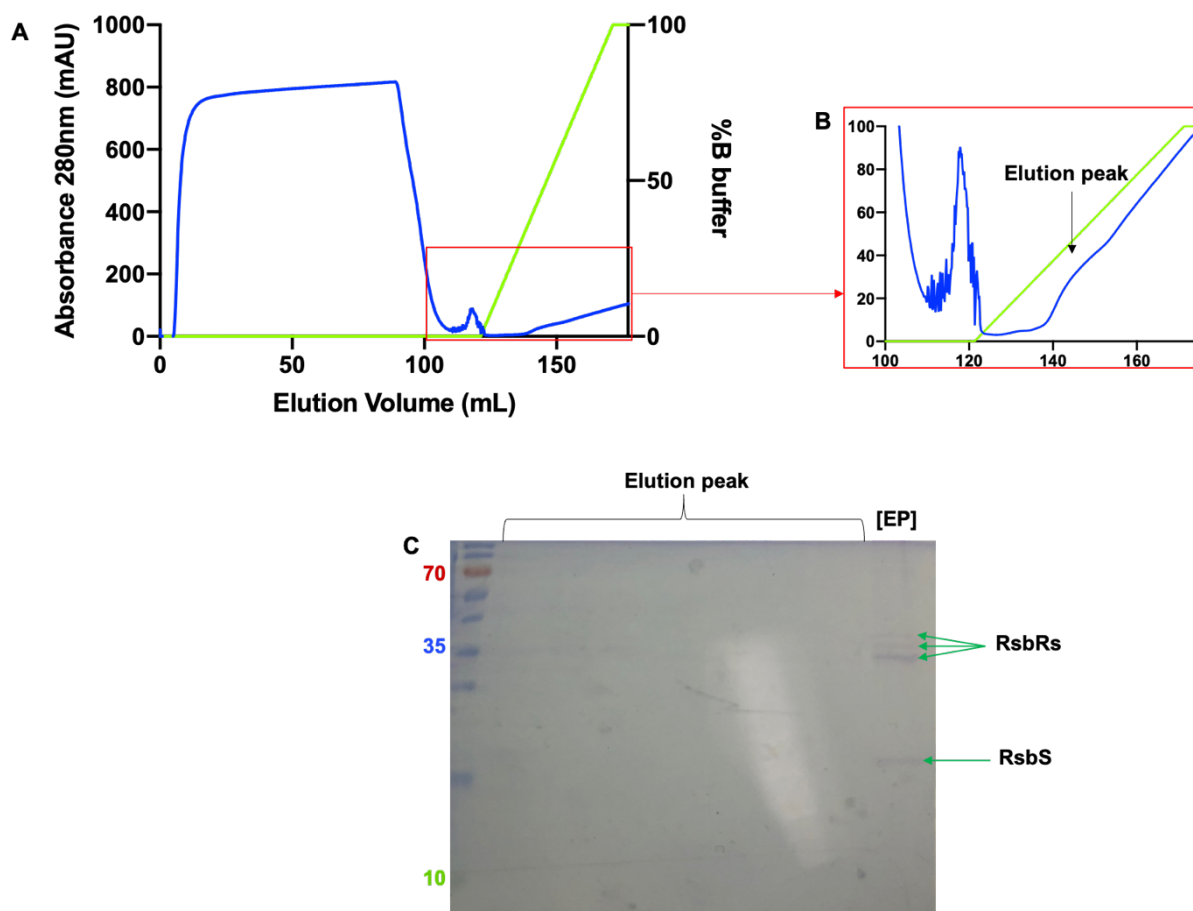


Figure 5.1. *Bsu* NR5623 native stressosome purification

A. Ni-NTA affinity purification chromatogram with the UV absorption at 280nm in blue and the percentage of buffer B in green. **B.** Zoom-in of the chromatogram at the elution peak for better visualization. **C.** 15% SDS-PAGE of the elution peak and the concentrated elution peak [EP] of the fractions with the protein MW ladder. The visible bands are picked out with green arrows with a cluster of bands corresponding to the expected MW of RsbR-like proteins.

The purified proteins were assumed to form a protein complex, and this was visualised by negative stain (**Chapter II, section 2.10.1. and Figure 5.2**) using a Hitachi HT7800 120 kV Transmission Electron Microscope (TEM). Some stressosome complexes can be observed on the micrographs obtained, but there are also many contaminants. The stressosome complex has a ring shape of approx. 20 nm, which corresponds to the diameter of the core of the complex, and the N-RsbR turrets are not visible at this resolution as they are too small (approx. 30 kDa).

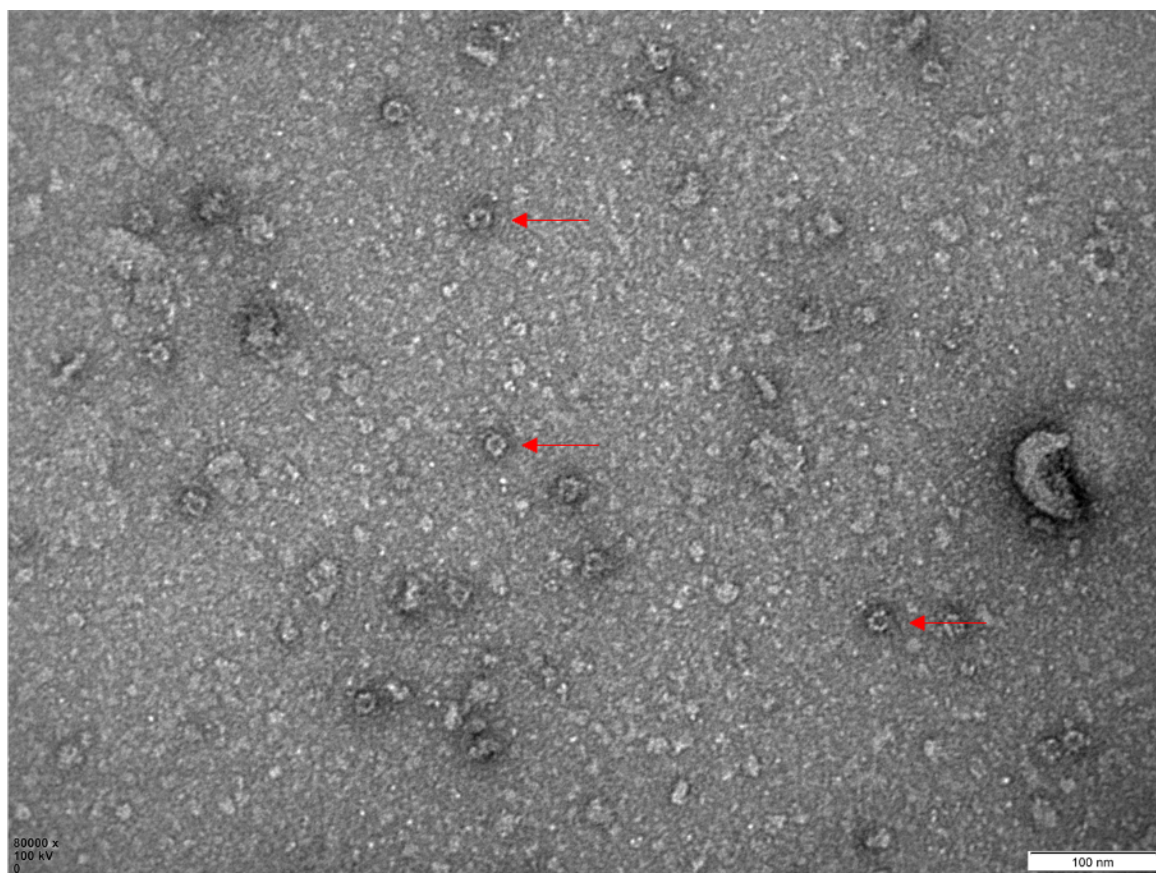


Figure 5.2. *Bsu* NR5623 native stressosome visualization by negative stain TEM
 Negative stain TEM of the concentrated sample from the NTA affinity purification of stressosomes from strain NR5623. Stressosome complexes are picked out with red arrows. Magnification 80k at 100kV. Scale bar 100nm.

Using the same cell growth, lysis and purification protocol, the native complex from strain NRS5621 was also purified, concentrated and visualised by negative stain (**Figure 5.3**). The stressosome purification from strain NRS5621 was more successful than that from strain NRS5623 as more stressosome particles and fewer contaminants were observed. Both samples were subsequently analysed by peptide mass finger printing to investigate the composition of the purified stressosome sample (**Section 5.3**).

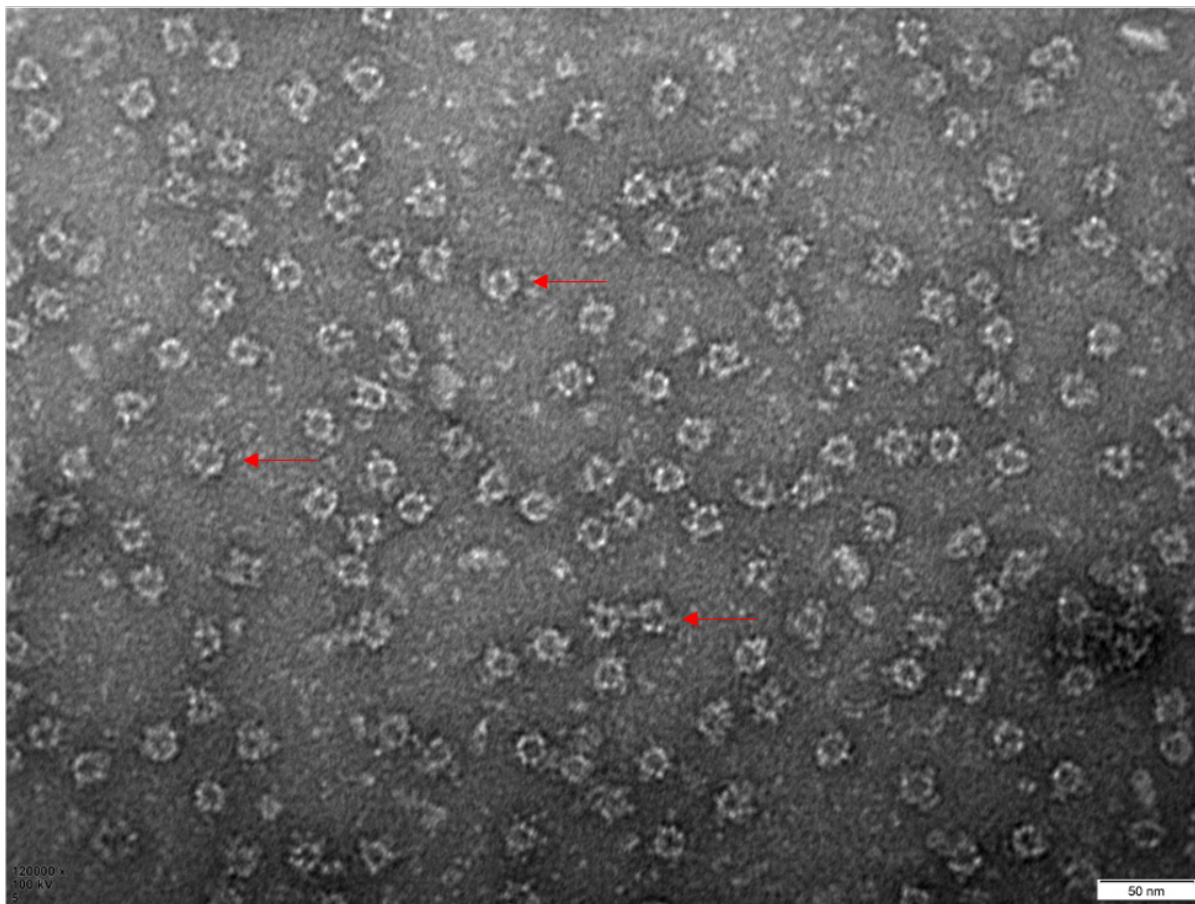


Figure 5.3. *Bsu* NR5621 native stressosome visualization

Negative stain TEM of the concentrated sample from the NTA affinity purification of stressosomes from strain NR5621. Stressosome complexes are picked out with red arrows. Magnification 120k at 100kV; Scale bar 50nm.

5.3. Investigation of the native stressosome composition from *B. subtilis*

The NR5623 sample was analysed by the Dundee proteomics facility (in collaboration with Prof. Dr. Nicola Stanley-Wall). The NR5621 sample was analysed by the Newcastle University Proteomic services during the PATHSENSE summer school in Newcastle, 2019. The results were analysed, and a subset of these results are displayed in **Table 5.1** and **Table 5.2**.

Table 5.1. NRS5623 strain peptide mass fingerprinting.

Accession	Score	Mass	emPAI	Description
sp O34860 RSBRB_BACSU	4611	32424	78.71	RsbT co-antagonist protein RsbRB
sp P28598 CH60_BACSU	3030	57389	10,55	60 kDa chaperonin
sp P02968 FLA_BACSU	2387	32607	24,51	Flagellin
sp P42409 RSBRA_BACSU	2304	31087	14,14	RsbT co-antagonist protein RsbRA
sp O31856 RSBRC_BACSU	1288	32146	4,94	RsbT co-antagonist protein RsbRC
sp P54504 RSBRD_BACSU	136	32043	0,46	RsbT co-antagonist protein RsbRD
sp P42410 RSBS_BACSU	68	13359	0,24	RsbT antagonist protein RsbS
sp P42411 RSBT_BACSU	63	14397	0,5	Serine/threonine-protein kinase RsbT

Table 5.2. NRS5621 strain peptide mass fingerprinting.

Accession	Score	Mass	iBAQ	Description
sp O34860 RSBRB_BACSU	323,31	32387	3701400	RsbT co-antagonist protein RsbRB
sp P02968 FLA_BACSU	323,31	32626	2007900	Flagellin
sp O31856 RSBRC_BACSU	323,31	32109	1614300	RsbT co-antagonist protein RsbRC
sp P42409 RSBRA_BACSU	323,31	31050	1215600	RsbT co-antagonist protein RsbRA
sp O34627 PHOT_BACSU	323,31	29194	640680	Blue-light photoreceptor pfyP
sp P28598 CH60_BACSU	323,31	57424	345490	60 kDa chaperonin
sp P54504 RSBRD_BACSU	133,89	31834	179010	RsbT co-antagonist protein RsbRD
sp P42410 RSBS_BACSU	17,807	13310	13381	RsbT antagonist protein RsbS

The NRS5623 sample was analysed by LC-MS/MS from which the score of the hit, the mass of the protein and the emPAI score was obtained (**Table 5.1**). The most abundant, and those proteins related to the stressosome, were filtered for *rsb* proteins in **Table 5.1** and **Table 5.2**. The emPAI score is the Exponentially Modified Protein Abundance Index. It is an estimate of protein abundance in the sample from peptide counts in a single LC-MS/MS experiment. Taking that value into account, RsbRB is the most abundant protein in the sample with an emPAI score of 78.71. Following RsbRB in abundance is a 60 kDa chaperonin protein and flagellin. The next most abundant proteins were the RsbR paralogues RsbRA and RsbRC. RsbRD, RsbS and RsbT were found with very low scores and emPAI values below 1, which

might reflect a lack of natural abundance or that these proteins do not yield many peptides because of the small size of the proteins from which the peptides are generated.

Similar distributions of the stressosome proteins were found when the purification product from strain NRS5621 was analysed (**Figure 5.2**). Like the emPAI score, the iBAQ value is an intensity-based absolute quantification and is thus a measure of protein abundance. The RsbRB paralogue is also the most abundant protein in this native stressosome complex sample, also followed by the flagellin contaminant. The next most abundant proteins were RsbRC and RsbRA. Interestingly, the blue-light photoreceptor YtvA, also called *pfyP*, was found only in this sample. The second major contaminant, the chaperonin 60 kDa, was also found. RsbRD and RsbS were identified but only with very low scores and iBAQ values, similar to the NRS5623 sample. Finally, the kinase RsbT was not identified.

RsbS was present in both samples in addition to the four RsbR paralogues. The blue-light photoreceptor YtvA was only found in the NR5621 sample while RsbT was only found in the NR5623 sample. Finding YtvA in one sample and not in the other one could suggest that the native complex is dynamic *in vivo* or might only be found in one sample because of sampling bias, even though the same process for each sample was used. The presence of RsbT in one of the samples could suggest that the cell was not stressed, whereas RsbT could not be found in the second sample for the reasons cited above for YtvA. The presence of stressosome-related proteins in a pull-down against His-tagged RsbRA suggests that the identified proteins interact into a complex.

The two major contaminants present in both samples are flagellin and the 60 kDa chaperonin; flagellin was previously reported to be a major contaminant in pull-downs of native stressosomes (Delumeau et al., 2006). The contaminants from the His-tag purification procedure could not be eliminated using size-exclusion chromatography as the total yield of purified stressosomes was too low (less than 15 µg). An alternative strategy to purify the native complex was by using antibodies raised against N-RsbRA by Cambridge Research Biochemicals Ltd (CRB). Purified samples of the N-RsbRA protein (**Chapter IV, section 4.3.1**) were sent to CRB to produce antibodies in rabbits. The α -N-RsbRA antibodies were purified by CRB using AffiGel-immobilised N-RsbRA. The purified α -N-RsbRA antibodies were subsequently also immobilised onto an AffiGel® Hz Immunoaffinity resin and this affinity column was used to purify unmodified native stressosomes from the WT *B. subtilis* 168 strain. 12 L of *B. subtilis* were grown to stationary phase and lysed by sonication following the same procedures used for the purification of His-tagged stressosomes (**Figure 5.4**). The lysate was clarified by centrifugation before being passed twice over the α -N-RsbRA column with a bed volume of 2 mL. After washing with 10 CV of 1x PBS buffer, the sample was eluted first

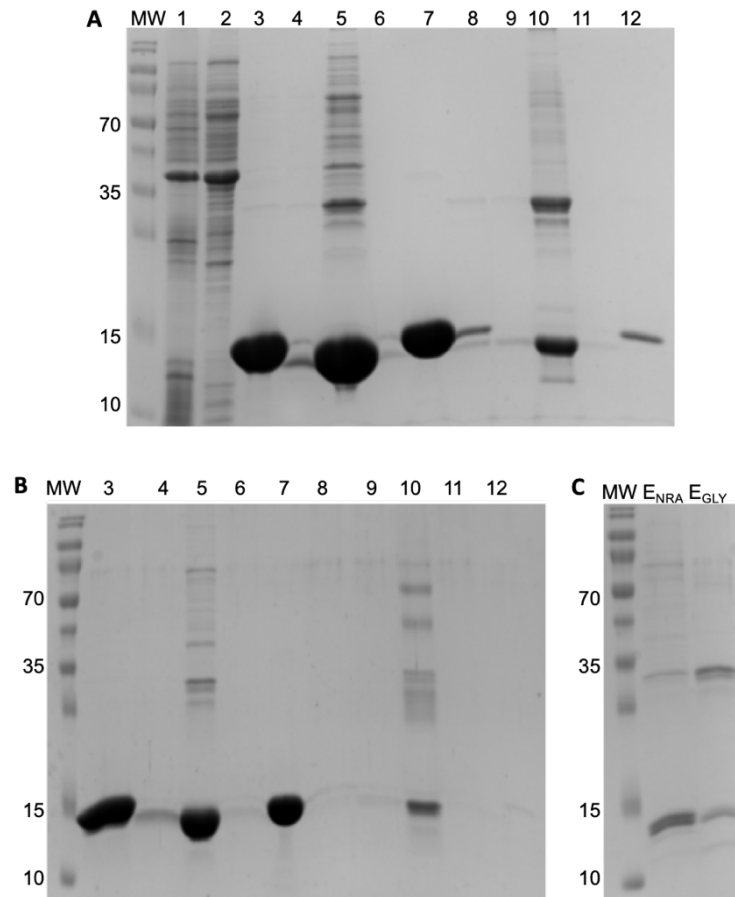


Figure 5.5. Native stressosome purification using N-RsbRA antibodies

A. 15% SDS-PAGE of the native stressosome purification. The following samples were loaded: pellet (1); antibody column flow-through (FT) after passing the input twice (2); competitive elution using N-RsbRA (3); FT of the Ni-NTA column using sample 3 (4); concentrated sample 4 (5); FT on concentrating sample 4 (6); elution of the Ni-NTA column with 100% buffer B (7); acidic elution (8); FT of the Ni-NTA column from sample 8 (9); concentrated sample 9 (10); FT on concentrating sample 8 (11); 100% buffer B elution from the Ni-NTA column (12). **B.** The reverse Ni-NTA affinity column to remove remaining His-tagged N-RsbRA using the same samples as described in panel **A**. **C.** Final concentrated sample for the competitive elution (E_{NRA}) and the acidic elution (E_{GLY})

The final samples were then visualised by negative stain TEM (**Figure 5.6**). The stressosome sample purified by competition led to low stressosome yields (**Figure 5.6A**) with a lot of background noise, which might have arisen from the His-tagged N-RsbRA used to compete stressosomes off the affinity column, as it could not be removed fully even after two cycles of reverse Ni-NTA purification. Only a few stressosome complexes can be seen on the micrograph, and it is difficult to ascertain their quality, homogeneity and suitability for 3D reconstruction by EM.

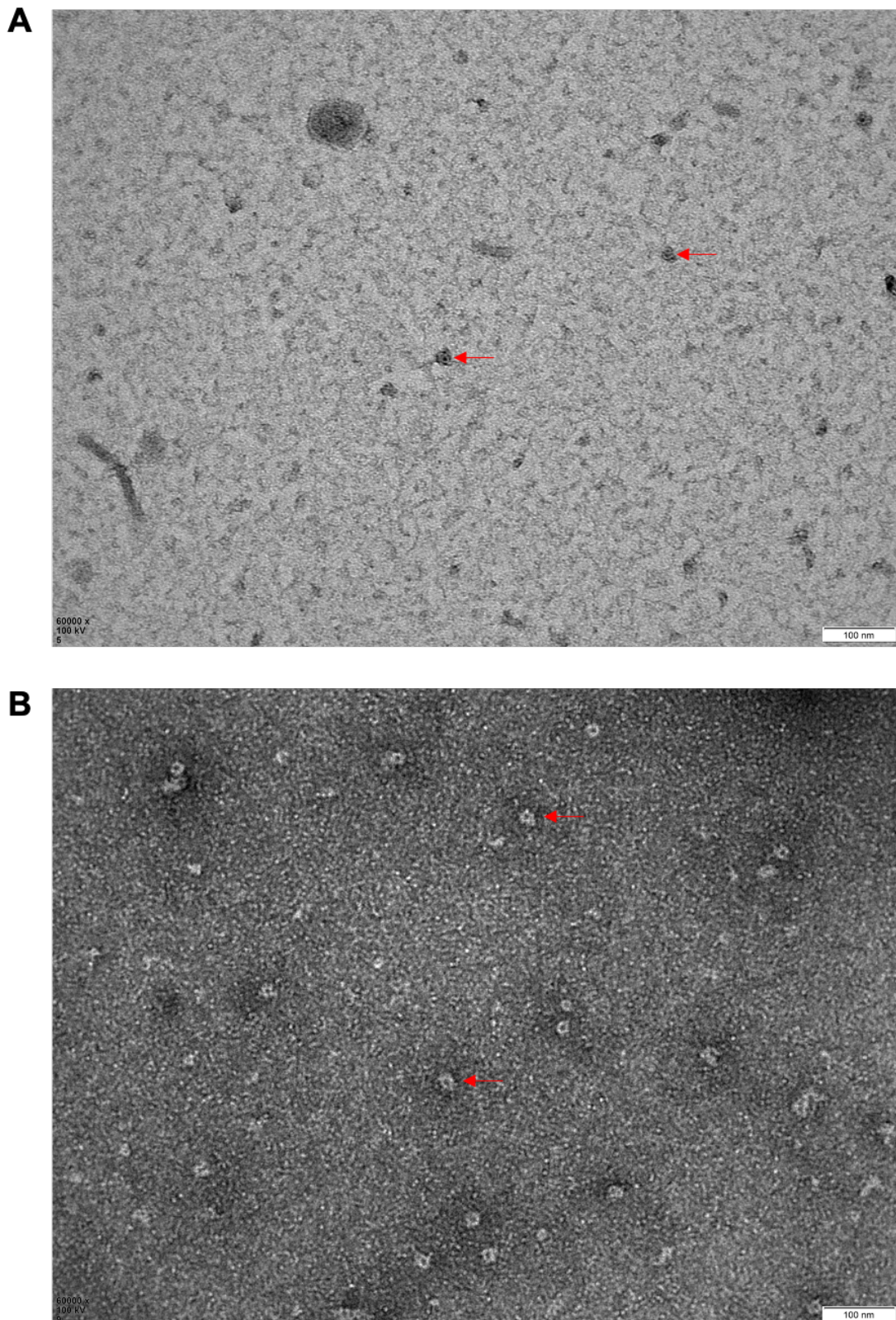


Figure 5.6. *Bsu* strain 168 native stressosome visualization

A. Negative stain TEM of the concentrated sample from the competitive elution. **B.** Negative stain TEM of the concentrated sample from the acidic elution. Scale bar 100nm. In both images the magnification is 60K, a field strength of 100 kV and the scale bars are 100 nm. The stressosome complexes in both micrographs are picked out in red.

By contrast, the stressosome complexes purified by acidic elution produced high quality micrographs with little background noise (**Figure 5.6B**). The approximate dimensions of these ‘native’ stressosomes are the same as those of recombinant minimalist stressosomes and are homogeneous across the micrograph. Therefore, the acidic elution gave better purification results for the native stressosome complex than by competitive elution, but the yield obtained is still low, approximately 5 µg from 12 L of cells, and therefore not enough for single particle analysis. In order to obtain a larger total yield, optimisation of the protocol is required such as to increase the initial volume of cell culture and use only acidic elution. Due to the COVID-19 pandemic, the purification of the native complex couldn’t be optimised.

5.4. Conclusion

The native composition of *B. subtilis* stressosomes has been reported previously (Delumeau et al., 2006). Two LC-MS data sets in this part of the thesis confirmed that all four RsbR paralogues co-purified when RsbRA was the target of independent affinity purification procedures. The ambition of this part of the study was to understand the stoichiometry of the paralogues in the stressosome complex, or complexes. Unfortunately, the Covid-19 pandemic introduced significant restrictions in accessing the laboratory and consequently this part of the project could not be completed. If time had permitted, several batches of the native complex would have been purified to produce a bigger input sample for quantitative mass spectrometry to determine the stoichiometry of the RsbR paralogues in native stressosomes. Furthermore, a greater yield of purified complex would have enabled single particle cryo-EM for native stressosome 3D structure solution. Nevertheless, the native complex was prepared successfully using different purification strategies that lays the groundwork for future studies of the native *B. subtilis* stressosome complex, and potentially those from different organisms.

Part B: Study of stressosomes variants with RsbR paralogues

To date, all studies of the stressosome *in vitro* have used the RsbR1/A paralogue as this is the only one within the *sigB* operon. In **Part A**, all four of the paralogues and the blue-light receptor YtvA were found in the native *B. subtilis* complex with an RsbRA pull-down. These results were consistent with and extended those obtained previously by Delumeau *et al*, 2006. In this part of the project, the ability of the RsbR paralogues of *L. monocytogenes* to form complexes with RsbS was investigated. An initial co-expression vector was produced to express *Lmo* RsbR1-RsbS following the methods described in **Chapter II, section 2.1.4**. The

stressosome was expressed and purified as described in **Chapter II, section 2.3**, however, the sample, even after many purification steps, only yielded a sample heavily contaminated with high molecular weight species contaminants. Moreover, the *Lmo* RsbR1-RsbS complex eluted both in the void volume of this column (approximately 8 mL: fraction 1 **Figure 5.7**) as well as across a large elution volume indicating a severe degree of heterogeneity in the final purified sample (**Figure 5.7**).

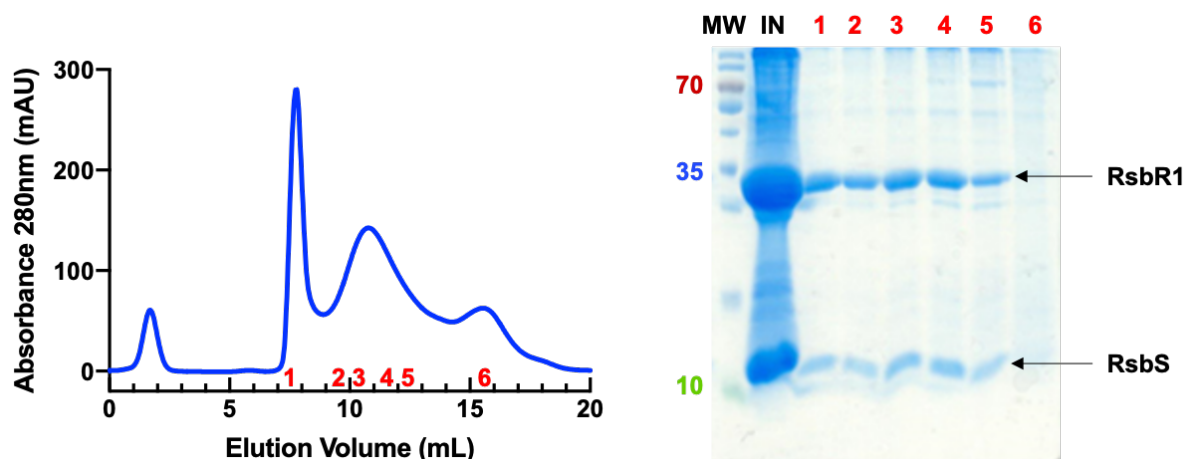


Figure 5.7. Purification of RsbR1-RsbS recombinant stressosome complex using a co-expression vector

Size-exclusion chromatogram with a Superose 6 S10/300 GL column of the RsbR1-RsbS protein with, in red numbers the fraction collected (1 mL). The 15% SDS-PAGE with a stained protein ladder (MW) displays the fractions of the SEC run

Another strategy was used to obtain the purest and most homogeneous recombinant stressosome sample, by mixing RsbR and RsbS. Each RsbR paralogue and RsbS from *L. monocytogenes* were pre-purified and the complex was reconstructed *in vitro*.

5.5. RsbR paralogues and RsbS purification

The full-length reading frames for all the *L. monocytogenes* RsbR and RsbS proteins were cloned into pET28gg (**Chapter II, section 2.1.1**). After successful cloning, verified by DNA sequencing of the inserted *orf*, the proteins were expressed and purified. The proteins were not expressed as genetic fusions to affinity-tags and were purified by anion exchange chromatography (**Chapter II, section 2.3.1**; **Supplemental figures S22-26**) followed by SEC (**Chapter II, section 2.3.2**). With each purification step SDS-PAGE was run to estimate the quality and the quantity of the purified sample. For clarity, only the final purification step will be displayed for each protein.

5.5.1. RsbR1 recombinant protein purification

RsbR1 eluted from a Superdex S200 HR16/60 size exclusion column at a volume of 71.5 mL (**Figure 5.8**) and could be prepared on a large scale to a high degree of final purity, at least 90%.

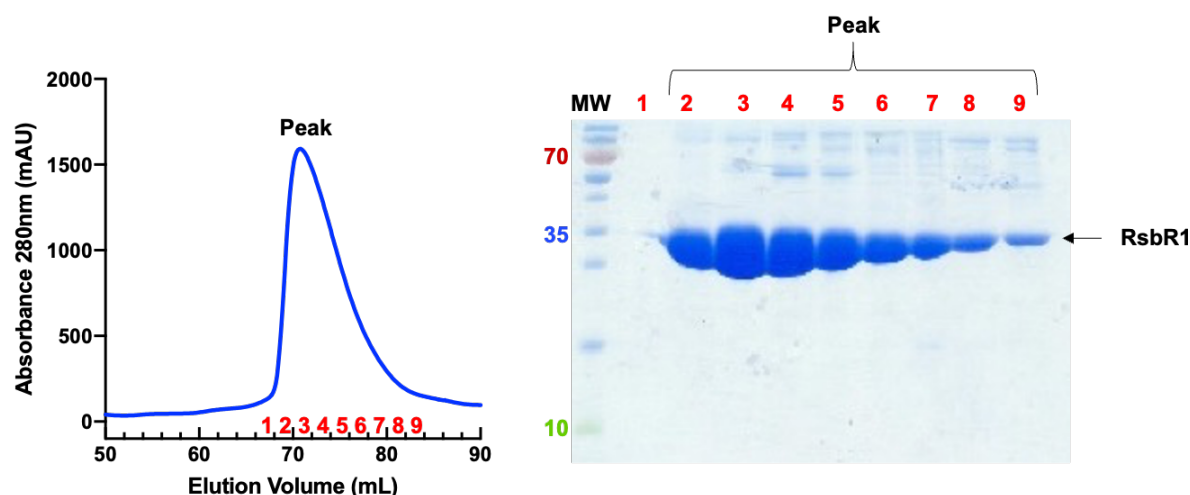


Figure 5.8. S200 16/60 SEC of RsbR1 recombinant protein

Final S200 16/60 size-exclusion chromatogram of RsbR1 (left), with a peak at a V_e of 71.5 mL. The collected fractions (2 mL) are numbered in red. 15% SDS-PAGE (right) with a stained protein ladder (MW) displays the fractions of the elution peak that electrophorese to approximately 35 kDa, which is the expected MW of RsbR1.

5.5.2. RsbR2 recombinant protein purification

RsbR2 eluted from a Superdex S200 HR16/60 size exclusion column at a volume of 79 mL (**Figure 5.9**). The fractions of the elution peak were analysed by SDS-PAGE (**Figure 5.9**), which indicates that RsbR2 is abundant but still contains some high MW contaminants.

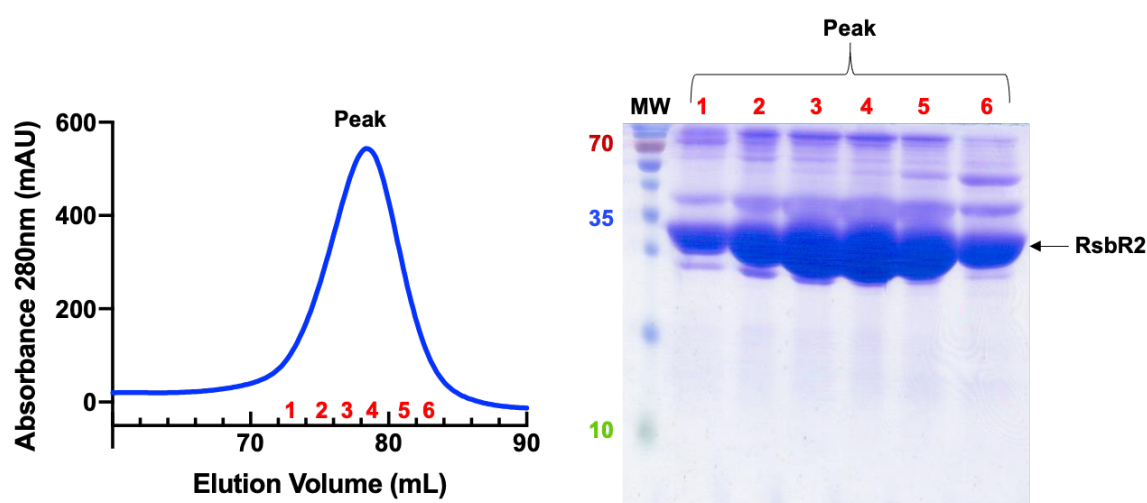


Figure 5.9. S200 16/60 SEC of RsbR2 recombinant protein purification

Final S200 16/60 size-exclusion chromatogram of RsbR2 (left), with a V_e peak of 79 mL and the collected fractions (2mL) are numbered in red. Right: 15% SDS-PAGE with a stained protein ladder (MW) and the corresponding fractions of the elution peak.

As this protein will be re-purified in complex with RsbS, it is possible that the high MW contaminants will be removed in subsequent purification steps and therefore no further purification steps on isolated RsbR2 were performed.

5.5.3. RbsR3 recombinant protein purification

RsbR3 eluted from a Superdex S200 HR16/60 size exclusion column at a volume of 78.2 mL (**Figure 5.10**). The fractions were analysed by 15% SDS-PAGE (**Figure 5.10**), which showed a high abundance of RsbR3, but also with some high MW contaminants. As this protein will also be re-purified in complex with RsbS, it is possible that the high MW contaminants will be removed in subsequent purification steps, hence no further purification procedures were carried out.

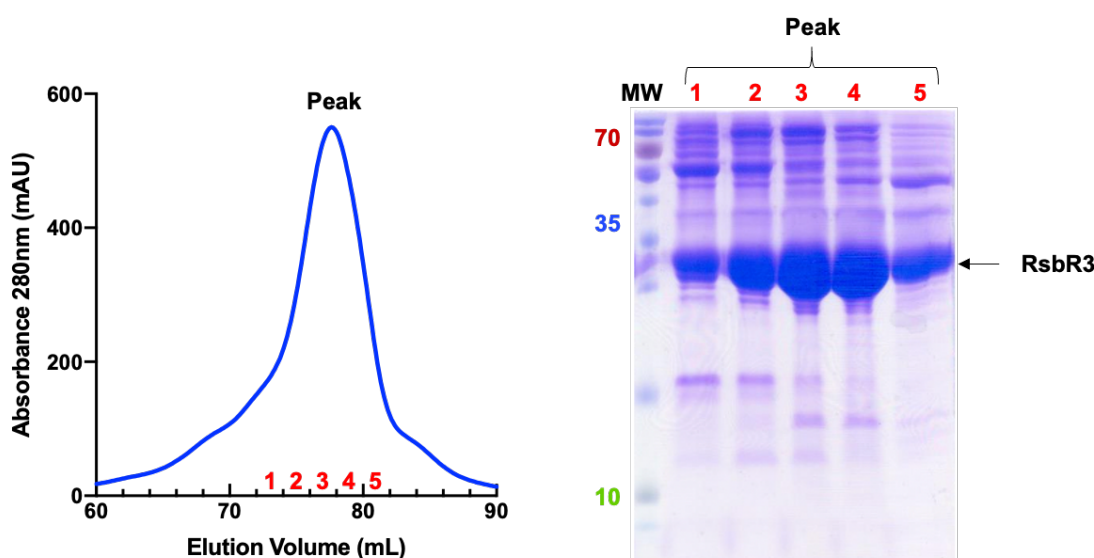


Figure 5.10. S200 16/60 SEC of RsbR3 recombinant protein

Final S200 16/60 size-exclusion chromatogram of RsbR3 (left), with a V_e peak of 78.2 mL. The collected fractions (2mL) are numbered in red. 15% SDS-PAGE (right) with a stained protein ladder (MW) and the corresponding fractions of the elution peak.

5.5.4. RsbR4 recombinant protein purification

RsbR4 precipitated after the anion exchange step, probably due to its high abundance. The sample was centrifuged and the supernatant was loaded onto a 16/60 S200 SEC column and RsbR4 eluted in a symmetrical peak at a volume of 78.8 mL (**Figure 5.11**). The pellets of the precipitates from the anion exchange, the supernatant of the sample - the input of the SEC - and the elution fractions of the peak were loaded onto a 15% SDS-PAGE (**Figure 5.11**). Most of the RsbR4 protein precipitated and was found in the pellet. Some RsbR4 was also found in the supernatant and eluted in the major elution peak. This sample is also highly contaminated, but as this protein will be re-purified in complex with RsbS, it is possible that the high MW contaminants will be removed in subsequent purification steps.

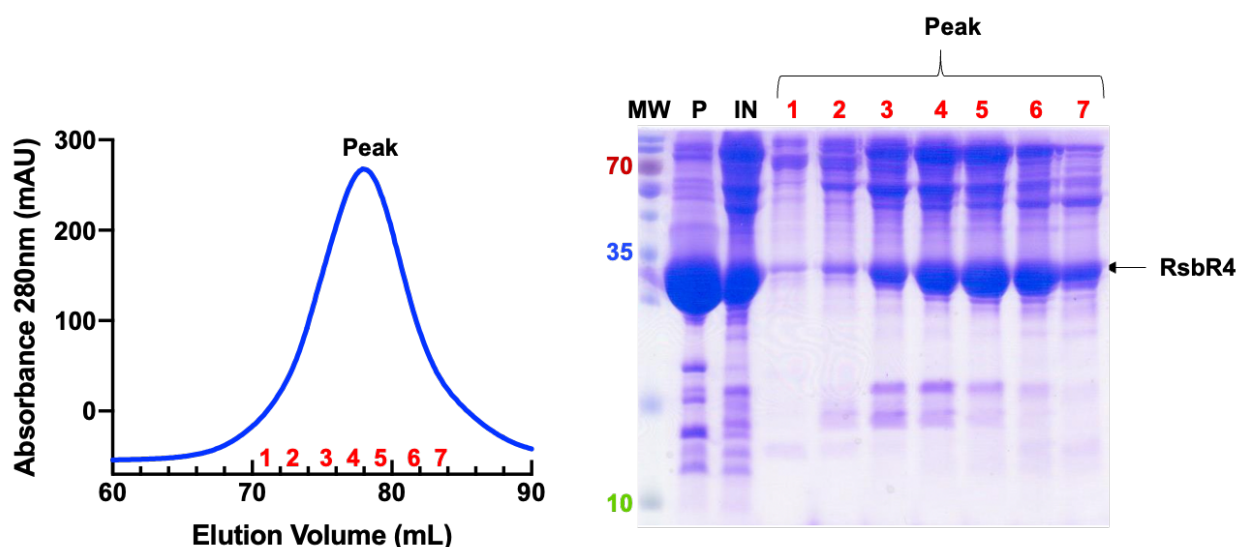


Figure 5.11. S200 16/60 SEC of RsbR4 recombinant protein

S200 16/60 size exclusion chromatogram of RsbR4 (left), with a peak a V_e of 78.8 mL. The collected fractions (2mL) are numbered in red. 15% SDS-PAGE (right) includes a stained protein ladder (MW), the pellet of the RsbR4 sample precipitation (P), the supernatant used as the input (IN) for the SEC and the corresponding fractions of the major elution peak.

5.5.5. Purification of recombinant RsbS

The RsbS recombinant protein was purified using the same protocol as the RsbR paralogues. RsbS size-exclusion chromatography with the S200 16/60 column yielded a single symmetrical elution peak (**Figure 5.12**) with a V_e of 85.4 mL. The purification of RsbS was quite successful as the RsbS sample obtained is highly pure according to SDS-PAGE analysis (**Figure 5.12**).

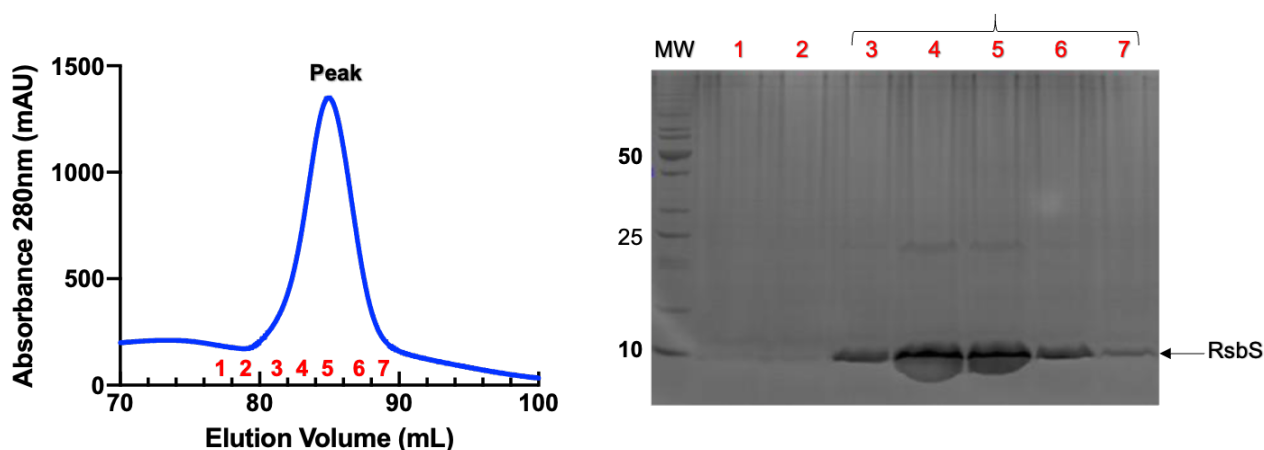


Figure 5.12. S200 16/60 SEC of RsbS recombinant protein

S200 16/60 size exclusion chromatogram of RsbS (left), with a V_e peak of 85.4 mL with the collected fractions (2mL) numbered in red. 15% SDS-PAGE (right) with a stained protein ladder (MW) of the corresponding fractions of the elution peak.

5.6. Stressosome complex formation by RsbR paralogues

All the RsbR paralogues and RsbS were pre-purified and ready to use for complex reconstitution. Each RsbR paralogue was mixed and incubated with purified RsbS with details in **Chapter II, section 2.4**. The protein mixtures were purified using a Superose 6 10/300GL column which has the capacity of separating proteins in the 5 kDa to 5 MDa range.

5.6.1. RsbR1-RsbS complex

The SEC of the RsbR1-RsbS mixture (**Chapter II, section 2.4**), with RsbS in excess, (**Figure 5.13A**) shows two main elution peaks and fractions from the two main elution peaks were loaded onto the SDS-PAGE (**Figure 5.13B**). RsbR1 and RsbS co-eluted in peak 1 and the excess of RsbS was found in the second elution peak. To determine if this RsbR1-RsbS co-elution led to the formation of the stressosome complex, fractions from peak 1 were pooled and concentrated to 3.5 mg / mL. This sample was then visualised by negative stain TEM (**Figure 5.13C**). RsbR1 alone was also investigated by negative stain, which did not form stressosome complexes under these experimental conditions (**Supplemental figures S27**). The stressosome is a globular macromolecular complex, the core of which is a hollow sphere. The core has a diameter of approximately 20 nm and is highly visible compared to the turrets that project from core. On the negative stain micrograph obtained for this sample, it is evident that RsbR1 and RsbS form a stressosome complex; at this low resolution, the turrets formed by the N-RsbR domain are not well resolved, a trait commonly observed in TEM studies of negatively stained stressosome samples (Chen et al., 2003; Delumeau et al., 2006). This sample was used for high-resolution single-particle cryo-EM, the results of which are described and discussed in **Chapter IV**.

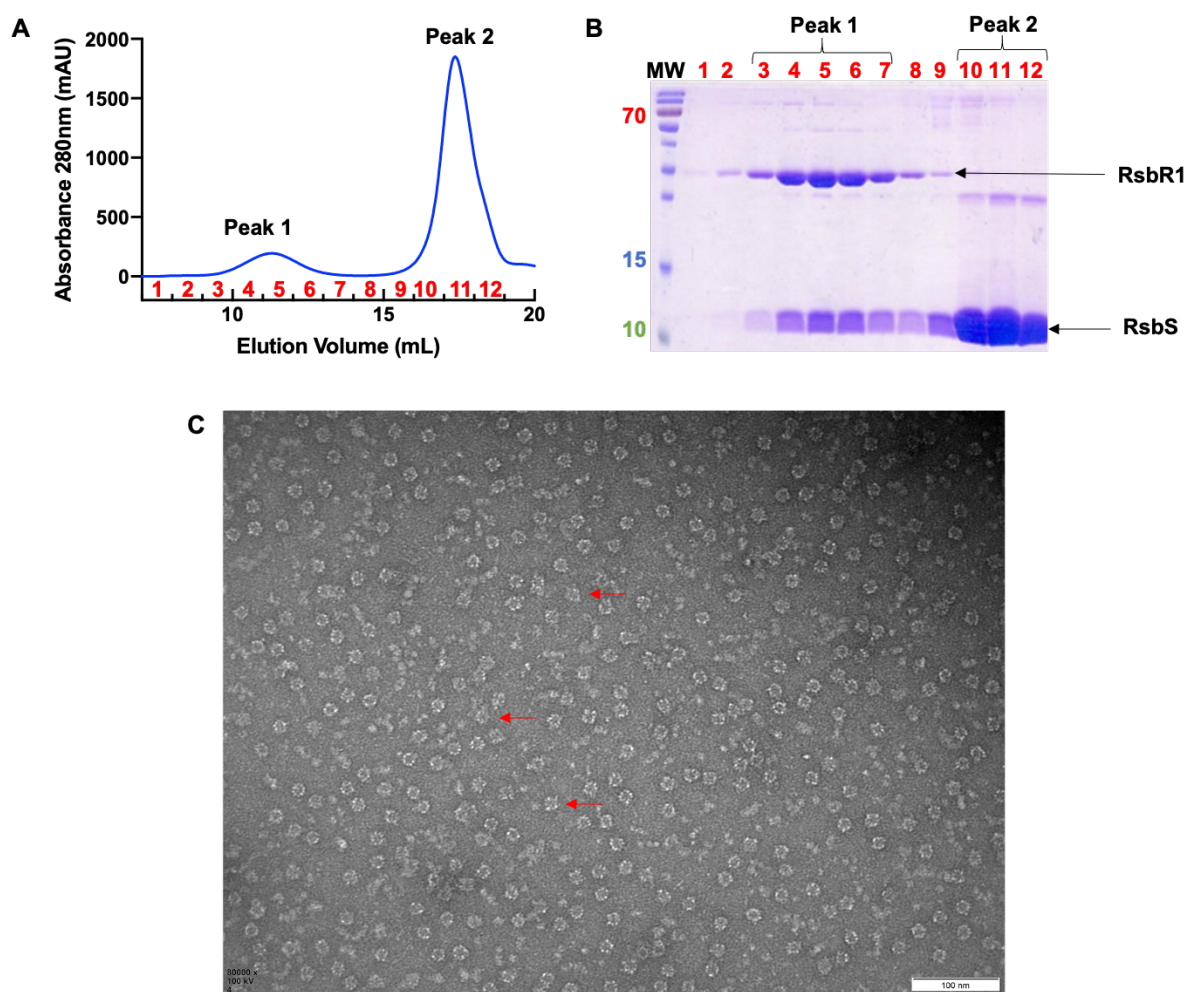


Figure 5.13. Reconstitution and visualization of the RsbR1-RsbS stressosome complex

A. Size-exclusion chromatogram with a Superose 6 S10/300 GL column of the RsbR1-RsbS protein mixture, with numbers in red correlating to the fraction collected (1 mL). RsbR1 and RsbS co-eluted in a peak with a V_e of 11.4 mL. Excess RsbS eluted in a second peak with a V_e of 17 mL. **B.** The 15% SDS-PAGE with a stained protein ladder (MW) of fractions of the SEC covering both elution peaks. **C.** Negative stain TEM of peak 1, containing the RsbR1-RsbS stressosome complex (red arrow). 80K magnification, 100 nm scale bar.

5.6.2. RsbR2-RsbS complex

RsbR2 is intriguing because results described in **Chapter III, section 3.10** revealed that N-RsbR2 did not adopt a globin-like fold in isolation. Nevertheless, our consortium partners have shown that RsbR2 is found in the native complex by LC-MS and that RsbR2 alone allows cell survival after an acidic stress in a *L. monocytogenes* strain lacking all of the RsbR paralogues except RsbR2. The potential of RsbR2 to form a complex with RsbS was analysed using the Superose 6 column (**Figure 5.14A**) in essence as described for RsbR1:RsbS above. There are several elution peaks on the SEC profile, and all were analysed by SDS-PAGE (**Figure 5.14B**). Very few proteins could be seen on the gel for peak 1, which presumably corresponds to some protein aggregates. The fractions from the second peak were pooled and concentrated, to

improve contrast on the gel, before analysis by SDS-PAGE. RsbR2 co-eluted with RsbS in this pooled and concentrated fraction, albeit with some minor contaminants present. Peak 3 contained uncomplexed RsbR2, and the fourth peak contained contaminants. The second elution peak was visualised by negative stain on a TEM (**Figure 5.14C**). RsbR2 also forms a complex with RsbS that resembles closely all other minimalist stressosome complexes. Since full-length RsbR2 forms stressosome complexes with RsbS it implies that RsbR2 is properly folded. By contrast, the crystal structure of N-RsbR2 is not compatible with the formation of stressosome complexes, (**Chapter III, section 3.10**), consistent with the conclusion the N-RsbR2 crystal structure is artefactual. The micrograph also revealed high background protein levels that might reflect instability of the RsbR2-RsbS complex or that the complex requires further purification to remove contaminants visible by SDS-PAGE.

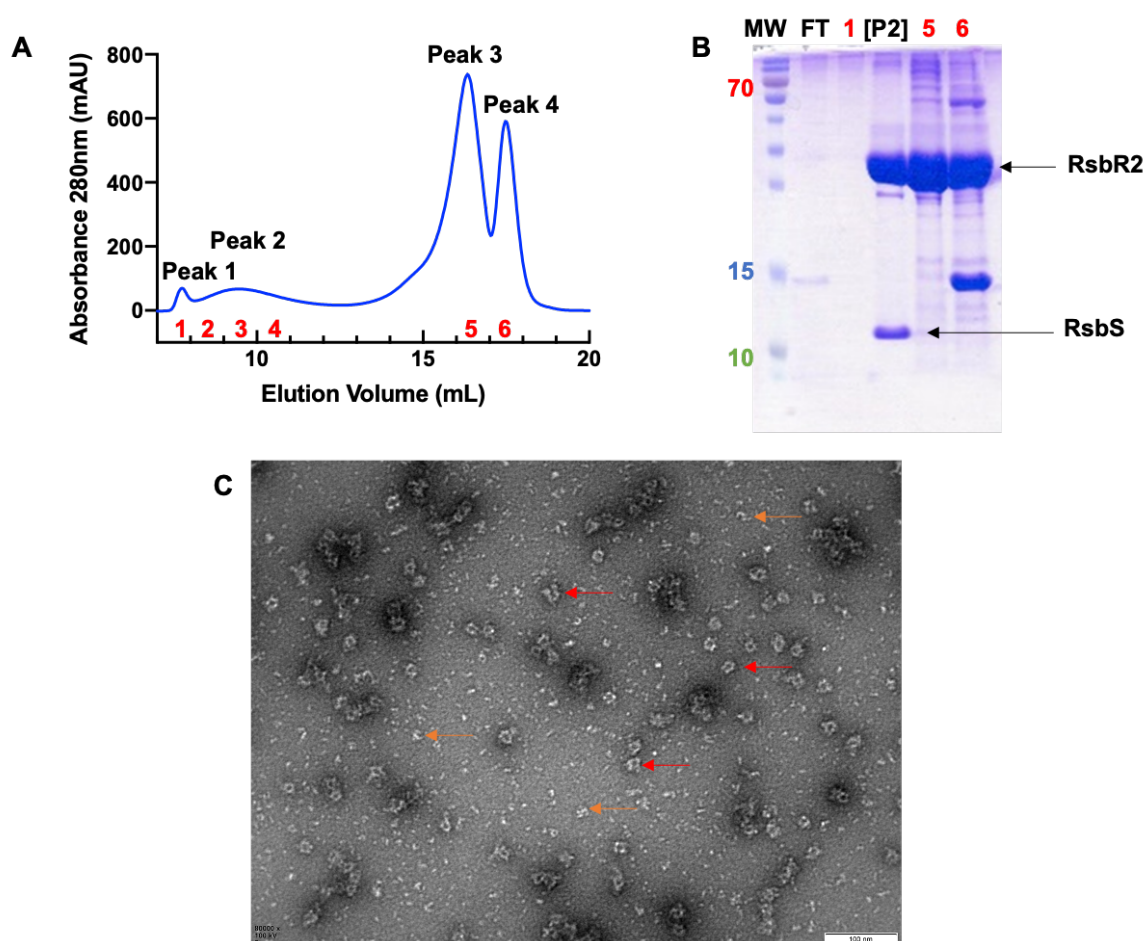


Figure 5.14. Reconstitution and visualization of the RsbR2-RsbS stressosome complex

A. Size-exclusion chromatogram with a Superose 6 S10/300 GL column of the RsbR2-RsbS protein mixture with numbers in red corresponding to fractions of 1 mL. RsbR2 and RsbS co-eluted in peak 2, with an approximate V_e of 9.5 mL. Fractions 2, 3 and 4 from peak 2 were pooled and concentrated to form sample P2. **B.** 15% SDS-PAGE with a stained protein ladder (MW); the flow-through (FT) from concentrating sample P2; sample P2 and fractions 1, 5 and 6 (corresponding to peaks 1, 3 and 4). **C.** Negative stain TEM of sample P2 where RsbR2-RsbS stressosome complexes are picked out with red arrows and some traces of incomplete complexes can be seen (orange arrows). 80K magnification, 100 nm scale bar.

5.6.3. RsbR3-RsbS complex

The ability of isolated RsbR3 to form a complex with RsbS was also tested. A mixture of RsbR3-RsbS with RsbS in excess was analysed by SEC using a Superose 6 column (**Figure 5.15A**). The SEC profile shows several elution peaks that were analysed by SDS-PAGE (**Figure 5.15B**). The first elution peak did not have any apparent proteins on the SDS-PAGE, and it most likely contains some protein aggregates or non-proteinaceous entities that absorb at 280 nm. RsbR3 co-eluted with RsbS in the second elution peak, likely to correspond to the elution of the stressosome complex. The thirds and fourth elution peaks correspond to contaminants not bound to the stressosome and the fifth elution peak corresponds mainly to excess RsbS. To confirm that RsbR3 and RsbS form stressosome complexes, the sample was visualised by negative stain on the TEM (**Figure 5.15C**).

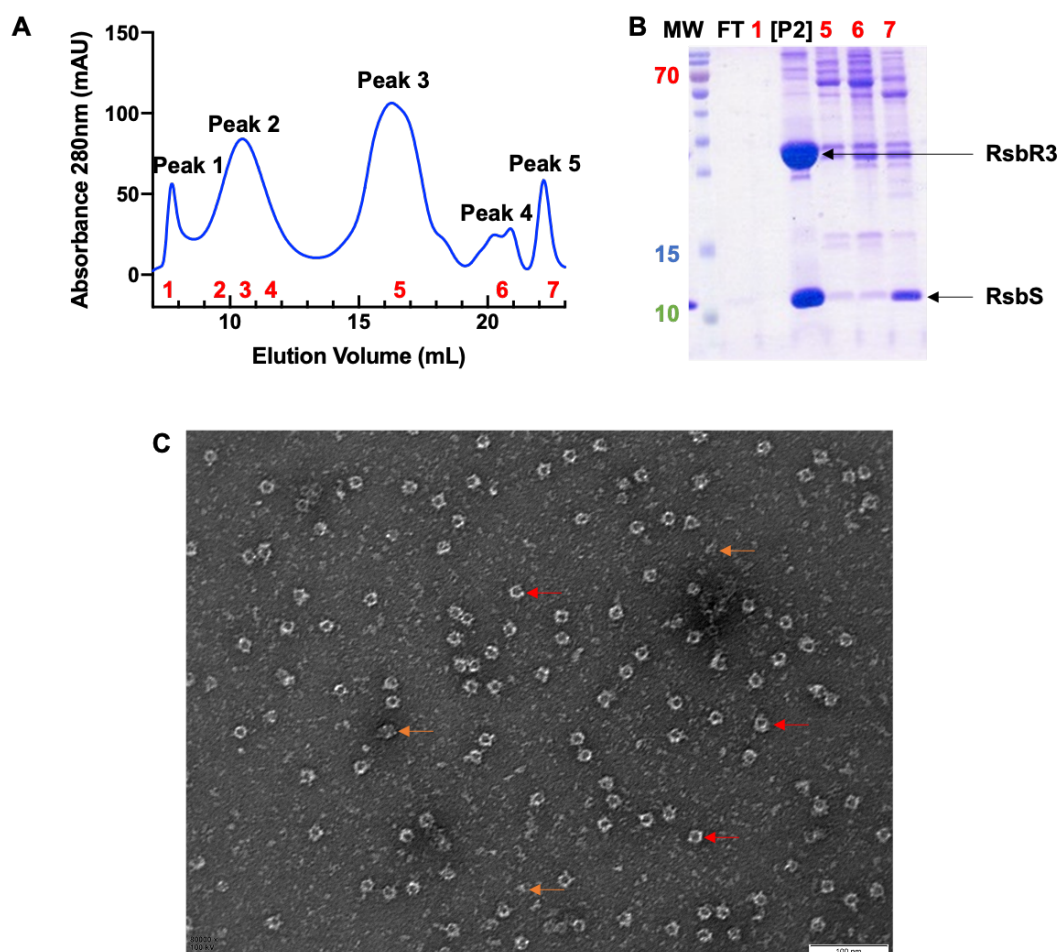


Figure 5.15. Reconstitution and visualization of the RsbR3-RsbS stressosome complex

A. Size-exclusion chromatogram of the RsbR3-RsbS protein mixture with a Superose 6 S10/300 GL column. The co-elution peak of RsbR3 and RsbS (peak 2) has a V_e of 10.6 mL, with red numbers corresponding to the collected fraction of 1 mL. **B.** The 15% SDS-PAGE has a stained protein ladder (MW); the flow-through (FT) from concentrating fractions 2, 3 and 4 from peak 2 [P2]; and single fractions from peaks 3, 4 and 5. **C.** Negative stain micrograph of peak 2 containing the RsbR3-RsbS stressosome complex (red arrow) with some traces of incomplete complexes (orange arrow). 80K magnification, 100 nm scale bar.

As for RsbR1 and RsbR2, RsbR3 can also form stable stressosome complexes with RsbS; there are many round-shaped stressosome complexes on the micrograph with diameters of approximately 20 nm. This complex might be an important target of further study, as a ligand binding pocket was identified in **Chapter III, Section 3.11**. This complex might yield insight to the signal transduction mechanism from the NTD to the CTD of RsbR3 when a ligand becomes bound, however, target ligands must be identified to go further with this study. This RsbR3-RsbS stressosome complex was also a candidate for single particle analysis using cryo-EM. Grids using a range of RsbR3-RsbS concentrations were prepared and screened at the Astbury Bioimaging centre at the University of Leeds. However, none provided a homogeneous stressosome spread (**Figure 5.16**) and only aggregates were observed. Due to the Covid-19 pandemic, there was no opportunity to pursue this project further, such as optimising the flash-cooling process or the use of different grid types.

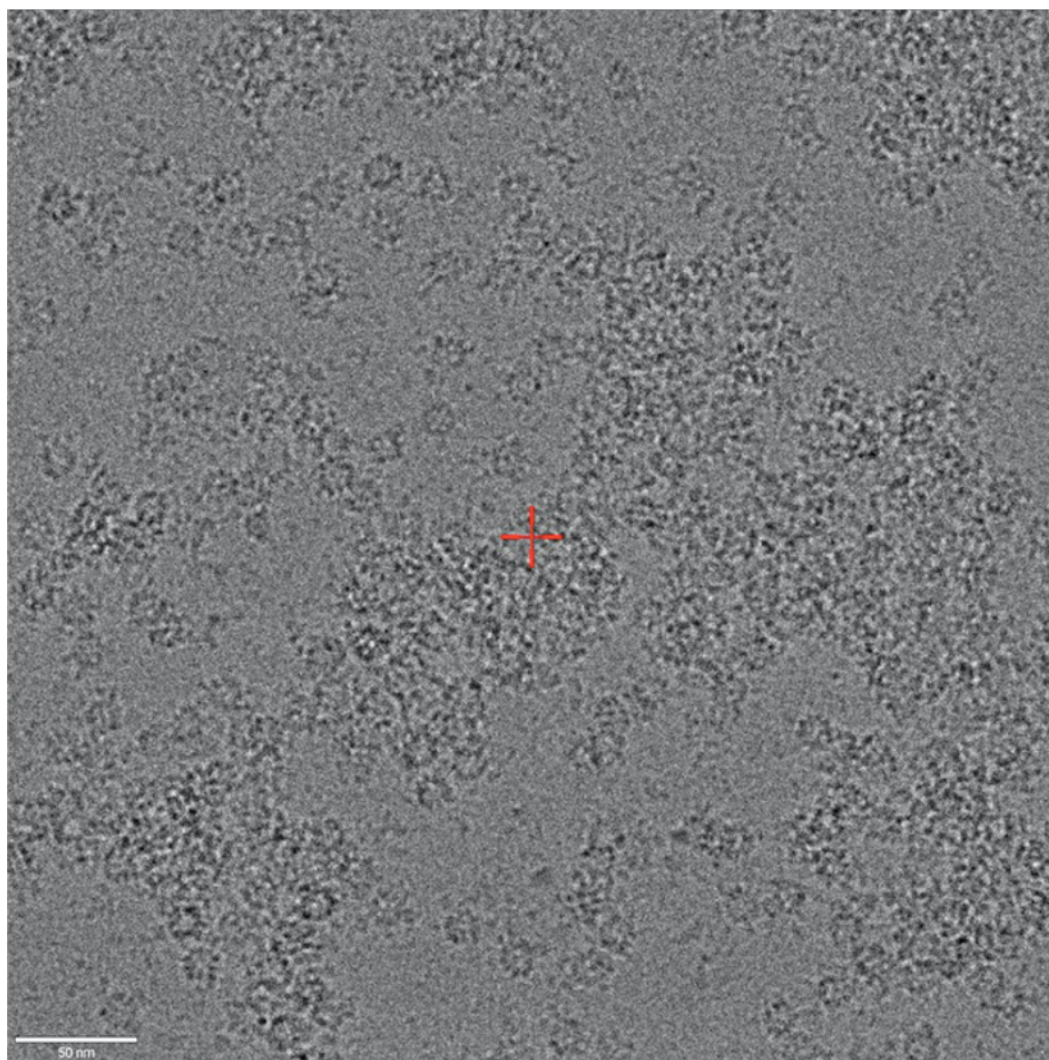


Figure 5.16. Cryo-EM grid screening of the RsbR3-RsbS stressosome complex
The RsbR3-RsbS stressosome complex flash cooled on Quantifoil 1.2/1.3. Only stressosome-like aggregates can be seen on this micrograph. 70K magnification, 50 nm scale bar.

5.6.4. RsbR4-RsbS complex

Finally, the ability of RsbR4 to form stressosome complexes with RsbS was analysed. The SEC profile of the RsbR4/RsbS protein mixture, with RsbS in excess, also displayed several elution peaks (**Figure 5.17A**). As previously observed, the first elution peak did not show much protein and may include only a small amount of protein aggregates (**Figure 5.17B**). The second peak corresponded to the co-elution of RsbR4 and RsbS. The third and the fourth elution peaks contained remaining contaminants and excess RsbS.

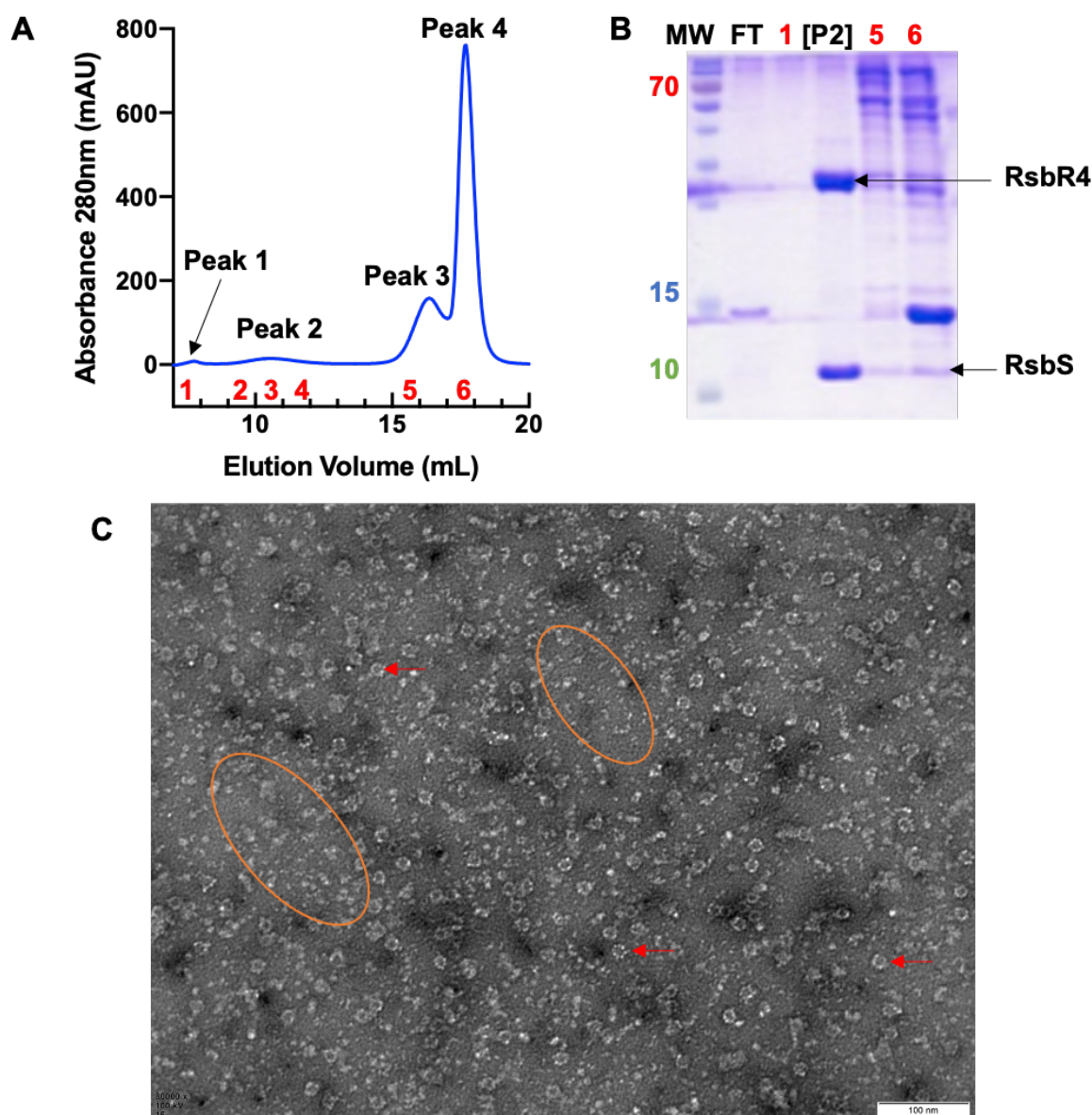


Figure 5.17. Reconstitution and visualization of the RsbR4-RsbS stressosome complex

A. Size-exclusion chromatogram of the RsbR4-RsbS protein mixture using a Superose 6 S10/300 GL column. Peak 2, in which RsbR4 and RsbS co-eluted, has a V_e of 10.6 mL. Fractions analysed by SDS-PAGE are depicted in red. **B.** 15% SDS-PAGE with a stained protein ladder (MW); the flow-through (FT); concentrated peak 2 [P2] and one fraction from each of the remaining peaks. **C.** Negative stain micrograph of P2 with RsbR4-RsbS stressosomes picked out by red arrows and big patches of incomplete or degraded complexes are circled in orange. 80K magnification, 100 nm scale bar.

To determine if the co-eluted RsbR4 and RsbS form stressosome complexes, this sample was visualised by negative stain on the TEM (**Figure 5.17C**). The micrographs showed some full stressosome complexes, but a lot of background staining is visible, which may correspond to broken or partially-formed stressosomes that accrued during sample preparation, or high MW aggregates.

5.7. Conclusion

The results above show that each isolated *L. monocytogenes* RsbR paralogue can form a stable stressosome complex with RsbS, data that are fully consistent with what has been described before for their orthologues in *B. subtilis* (Delumeau et al., 2006). Consequently, also as in *B. subtilis*, ‘native’ stressosomes from *L. monocytogenes* are also likely to be comprised of all RsbR paralogue, though the stoichiometry of them within the complex remains unknown in any bacterial species. Indeed, it remains to be determined whether there is a single composition in all stressosomes or whether stressosomes are dynamic and are comprised of varying levels of each RsbR paralogue. Time constraints has meant that this question became out of the scope of these studies. The RsbR1-RsbS stressosome yielded the highest quality stressosome complexes on negatively-stained electron micrographs and these were taken forward for further study.

Part C: Study of the ternary stressosome complex involving RsbT and RsbX

In an unstressed cell, the stressosome sequesters the kinase RsbT to form the ternary complex of RsbR/RsbS/RsbT. In a stressed cell, however, RsbT is released from the stressosome after phosphorylating the STAS domains of RsbR and RsbS. At the end of the response to stress, the phosphatase RsbX binds to the phosphorylated stressosome to dephosphorylate the STAS domains and to reset the signal transduction switch (Voelker et al., 1995; Yang et al., 1996). In order to prepare ternary complexes for 3D reconstruction by single particle analysis, RsbT and RsbX were purified, their potential for binding stressosomes was assessed by co-elution by SEC and finally were visualised by negative stain TEM, in order to see if the complex maintains its integrity in the presence of these enzymes.

5.8. Purification of the RsbT kinase

Recombinant RsbT is unstable when purified in isolation (Delumeau et al., 2006). Therefore, an MBP-fusion was used (**Chapter II, section 2.1.3**) that allowed the expression and solubilisation of RsbT on a scale to support structural studies. The MBP-fusion also had a His-tag for protein purification.

The RsbT-MBP fusion was expressed in *E. coli* and purified by Ni-NTA affinity (**Chapter II, section 2.3.1**). The sample was analysed by SDS-PAGE (**Figure 5.18A**) where a high yield of the RsbT-MBP fusion was evident. The fractions were pooled and the MBP tag was cleaved overnight using the C3S protease; the sample was dialysed at the same time to remove excess imidazole. The sample was then subjected to a reverse Ni-NTA purification in which RsbT would be found in the flow-through while MBP and any remaining MBP-RsbT fusion would be bound by the Ni-NTA column and eluted with imidazole. The obtained fractions were analysed by SDS-PAGE (**Figure 5.18B**). RsbT released from the MBP fusion was found in the flow-through and MBP (and/or MBP-RsbT) eluted later with imidazole. The purified RsbT was stabilised by the addition of 0.1 mM ADP and 2 mM MgCl₂. This RsbT was then used to reconstitute the ternary RsbR-RsbS-RsbT stressosome complex.

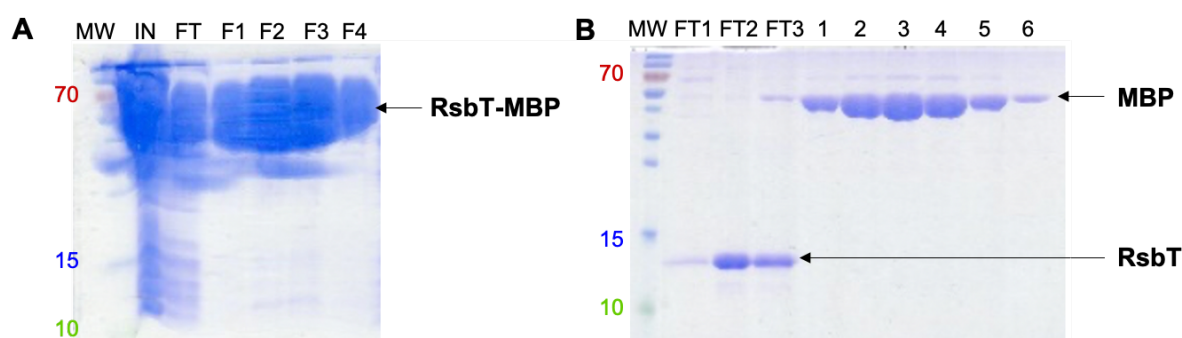


Figure 5.18. MBP fused RsbT protein purification

A. 15% SDS-PAGE of the elution fractions of the Ni-NTA purification with MW representing the protein MW ladder; the sample input (IN); the column flow-through (FT); and the four fractions from the imidazole gradient elution (F1, F2, F3, F4). **B.** 15% SDS-PAGE of the reverse Ni-NTA purification with the MW ladder (MW); the column flow-through (FT1, FT2, FT3); and fractions from the imidazole gradient elution (1 to 6).

5.9. The RsbX phosphatase purification

RsbX has a pI of 8.41, which is well-suited for purification by cation exchange (**Chapter II, section 2.3.1**). After the cation exchange step (**Supplemental figures S28**), RsbX-containing fractions were concentrated by ultrafiltration and purified further by SEC using an S200 16/60 column (**Figure 5.19**). The SEC profile shows a good symmetrical elution peak

with a V_e at 82.23 mL. Some minor contaminants were also evident on the SDS-PAGE but, as with the purification of the isolated RsbR paralogues, this sample will be purified further when the RsbR-RsbS-RsbX complex is reconstituted.

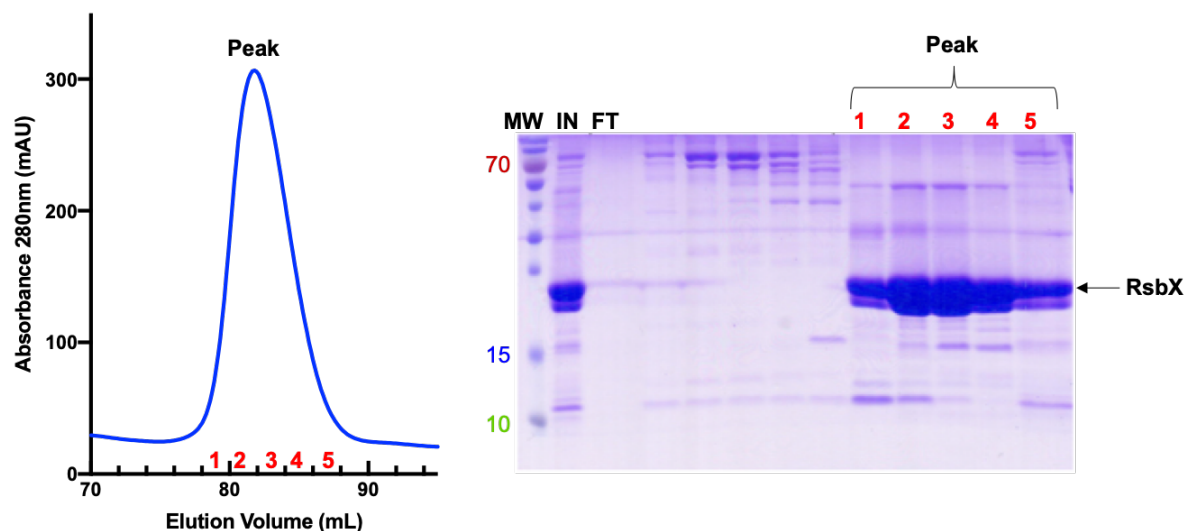


Figure 5.19. S200 16/60 SEC of RsbX phosphatase protein

Size-exclusion chromatogram of RsbX with a S200 16/60 column (left) in which RsbX elutes at a V_e of 82.23 mL with fractions in red denoting those analysed by SDS-PAGE. 15% SDS-PAGE (right) with a stained protein ladder (MW); the flow-through (FT) of the concentrated input (IN); and the fractions corresponding to the RsbX elution peak.

5.10. Reconstitution of the RST and RSX ternary stressosome complexes

The purified recombinant RsbR1-RsbS complex was used to reconstitute ternary complexes containing RsbT (RST complex) and RsbX (RSX complex). RsbT and RsbX were each added to separate pools of the RsbR1-RsbS complex and left to incubate for 1 hour. Both samples were then analysed by SEC to assess whether there was any co-elution of either RST or RSX complexes.

The SEC profile of the RST sample yielded several elution peaks in which one (peak 1) corresponding to the elution behaviour of the stressosome has a V_e of 10.9 mL (**Figure 5.20**). The fractions of peak 1 contained RsbR1, RsbT and RsbS when analysed by SDS-PAGE (**Figure 5.20**). The addition of RbsT did not impact the elution volume as presumably RsbT binds between N-RsbR1 turrets and, therefore, the complex maintains a similar Stokes radius to the R:S complex. The approximate stoichiometry on the gel for RsbR1:RsbS:RsbT is 2:1:1, which is in accordance with previously published data (Marles-Wright et al., 2008; Williams et al., 2019). The second major peak contains RsbS, presumably indicating that some of the

RsbR1:RsbS complex has disassociated and the absence of corresponding RsbR1 fractions suggests that some of this protein has degraded.

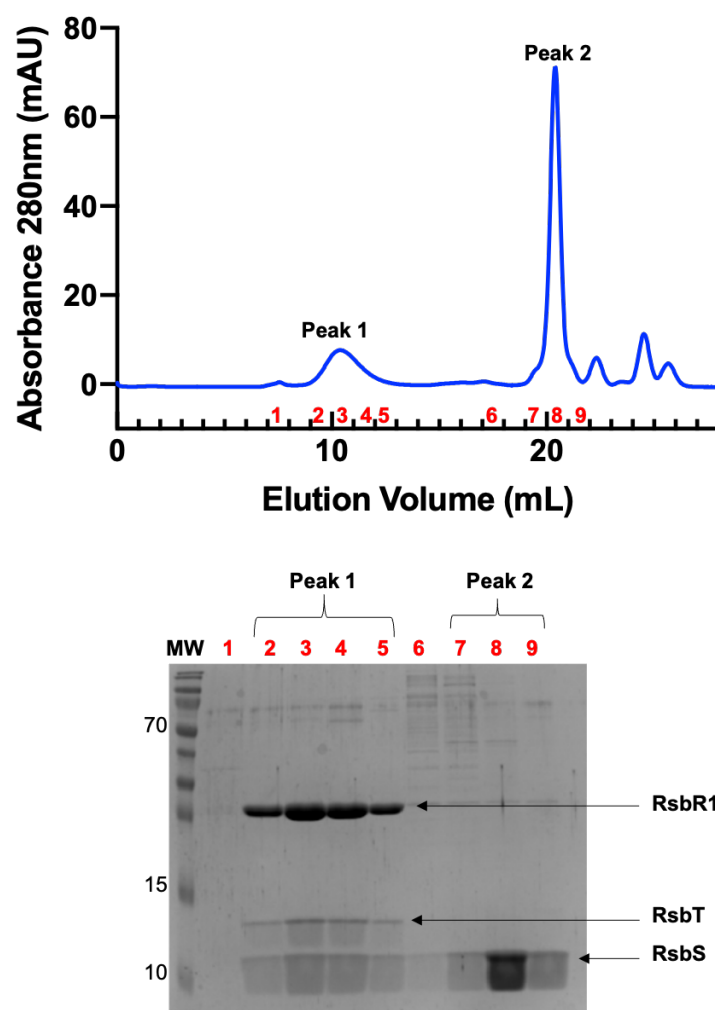


Figure 5.20. Reconstitution of the RsbR1-RsbS-RsbT recombinant stressosome complex

Top: Size-exclusion chromatogram of the RST sample with a Superose 6 S10/300 GL column with numbers in red denoting the fraction collected (1 mL). Bottom: The 15% SDS-PAGE with a stained protein ladder (MW) and fractions across peaks 1 and 2 from the SEC.

The RSX sample SEC profile also yielded several elution peaks; peak 1, corresponding to the stressosome elution, elutes at 11.7 mL (**Figure 5.21**). When analysed by SDS-PAGE RsbX was seen to co-elute with the RsbR1-RsbS complex in peak 1. Presumably, and similarly to RsbT, RsbX is likely to bind between the N-RsbR1 turrets and consequently RsbX binding does not affect the Stokes radius.

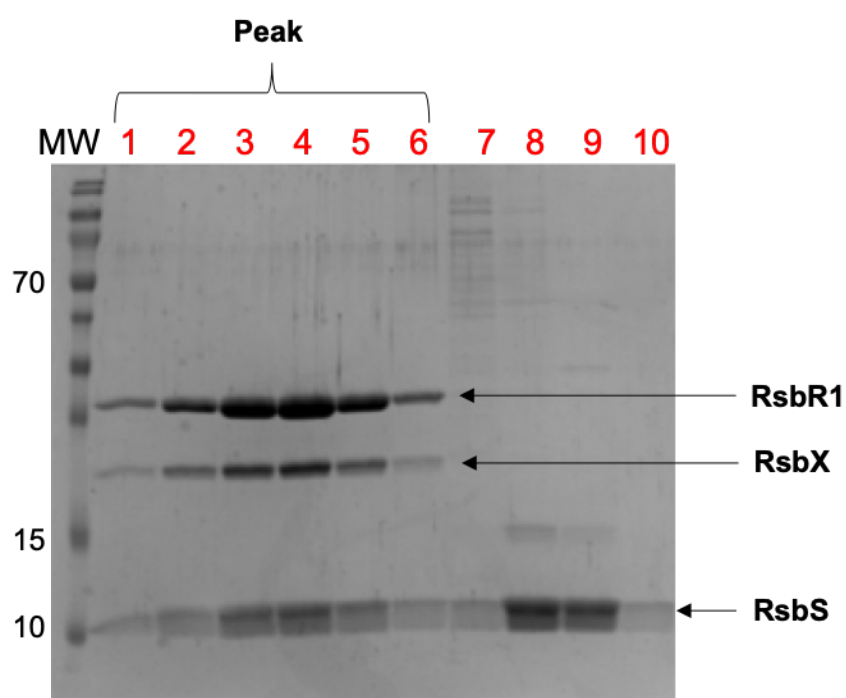
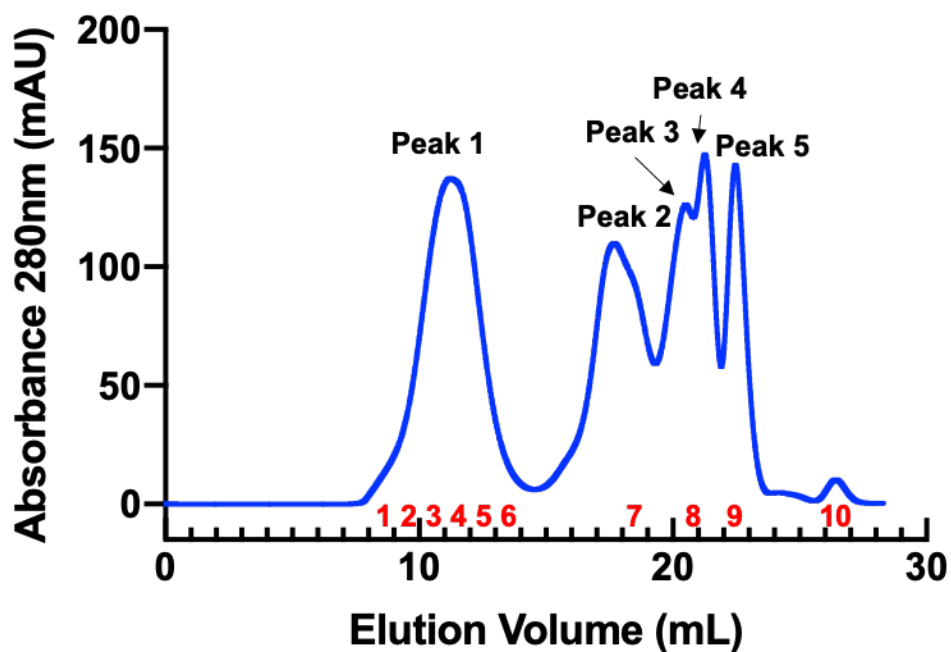


Figure 5.21. Reconstitution of the RsbR1-RsbS-RsbX stressosome complex

Top: Size-exclusion chromatogram of the RsbR1-RsbS-RsbX mixture with a Superose 6 S10/300 GL column; red numbers indicate the fraction collected (1 mL). RsbR1-RsbS-RsbX co-elute in peak one with a V_e of 11.7 mL. Bottom: 15% SDS-PAGE with a stained protein ladder (MW) of the fractions of SEC peaks 1-6.

For both samples, the stoichiometry cannot be determined definitively as RsbS appears to dissociate from the binary complex in forming the respective ternary complexes. Both the

RST and the RSX complexes were visualised by negative stain TEM (**Figure 5.22**). In both cases stressosome complexes are apparent that are similar in appearance to the R1:S complex, but at this resolution neither RsbT nor RsbX is visible since both are likely to bind close to the core surface, between the N-RsbR1 turrets, and neither protein is sufficiently large to be observable by negative stain TEM in the first place (Orlova and Saibil, 2011). It was also not possible to discern differences between the RsbRA:RsbS and RsbRA:RsbS:RsbT complexes from *B. subtilis* when analysed by the same TEM procedure (Chen et al., 2003).

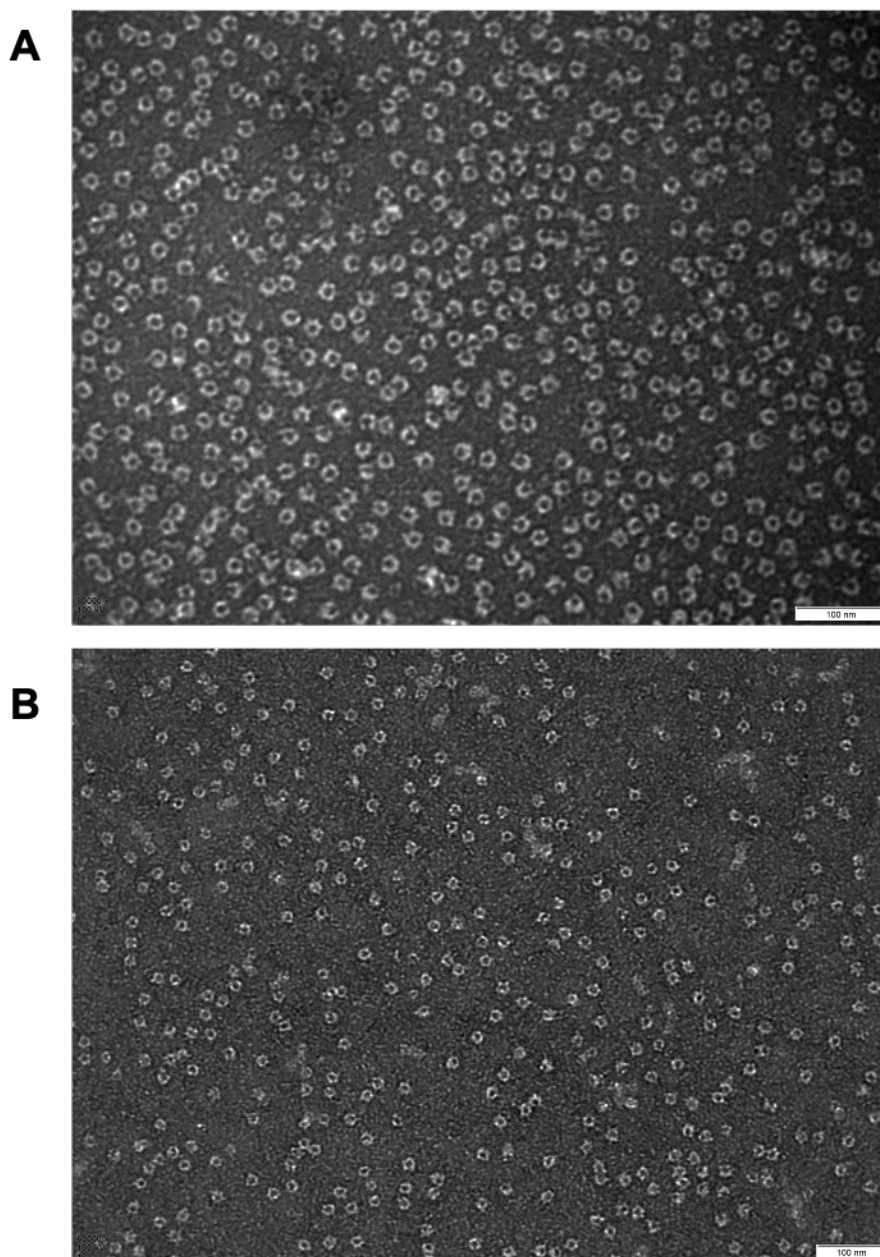


Figure 5.22. Visualisation of the RST and RSX complexes by negative stain
A. RST stressosome visualization by negative stain collected at 80K magnification. **B.** RSX stressosome complex visualization at 60K magnification. Scale bars are 100 nm in both cases.

5.11. Conclusion

In both cases, using the same RsbR1-RsbS stressosome sample batch, RsbT and RsbX were able to bind to the RsbR1-RsbS complex *in vitro*. RsbT was prepared in the presence of ADP in order to stabilise it; a better ATP mimic would have been a non-hydrolysable analogue such as ADP-N-P or ADP-C-P, but these molecules are not cost-effective for use during protein preparation. Nonetheless, the ADP-loaded RsbT is still capable of interacting with the *L. monocytogenes* stressosome, as is the case for the *B. subtilis* equivalent (Chen et al., 2003; Marles-Wright et al., 2008). This is the first time, however, that an interaction has been described for the stressosome and its phosphatase, RsbX, and this did not require any phosphoryl group mimic to assist binding, except for the loading of RsbX with magnesium ions, which are required for catalysis (Shi, 2009). Both the RST and RSX complexes appear suitable from these micrographs for quasi-atomic resolution reconstruction by single particle analysis and cryo-EM. However, the Covid-19 pandemic has meant that it has not been possible to pursue these studies any further for the RST complex, and these will form the basis of future work but outside the scope of this thesis.

Chapter VI: Cryo-EM single
particle analysis of the
L. monocytogenes RsbR1-RsbS
and RsbR1-RsbS-RsbX
stressosome complexes

6.1. Introduction

Structurally, the stressosome complex has been studied most widely in *B. subtilis* (Kwon et al., 2019; Marles-Wright et al., 2008) but more recently also in *L. monocytogenes* (Williams et al., 2019). As described previously, the stressosome in both organisms has a stoichiometry of RsbR:RsbS of 2:1; in total, 20 dimers of RsbR and 10 dimers of RsbS were found. However, the stress-sensing mechanism of these stressosomes could not be determined after their structures were determined. In fact, only the positions of the phosphorylated residues in RsbR and RsbS in the whole complex, which cluster at the dimerization interface of RsbR-RsbS, were identified and which are close to the RsbT binding site (Marles-Wright et al., 2008; Williams et al., 2019). As described in **Chapter I, section 1.3.1**, RsbT is the kinase that is sequestered by the stressosome in the absence of perceived environmental stress. When stress is detected, RsbT is activated by an unknown mechanism and phosphorylates the STAS domains of RsbR and RsbS and RsbT is then released from the complex. At the end of a stress response, the stressosome complex is reset, this time by the RsbX phosphatase. In contrast to RsbT, the binding of RsbX remains unknown, except that logically it must bind to the same site as RsbT to effectively dephosphorylate the phosphoserine and phosphothreonine residues on the RsbR and RsbS STAS domains. Therefore, to investigate any RsbX-induced structural changes to the stressosome, the RsbR1-RsbS (*LmoRS*) and the RsbR1-RsbS-RsbX (*LmoRSX*) complexes were subjected to cryo-EM single particle analysis.

6.2. Cryo-EM data collection of the *LmoRS* and *LmoRSX* stressosome complex.

The *LmoRS* and *LmoRSX* complexes were recombinantly expressed, purified and visualized by negative stain (**Chapter V, section 5.5.1 and 5.10**). These samples were subsequently sent to the Astbury Biostructure Laboratory at Leeds University, in a collaboration with Rebecca Thompson, Charlotte Scraff, Dan Maskell and Neil Ranson that commenced shortly before the covid-19 pandemic. Cryo-EM grids were prepared and screened by the Leeds team from both samples (**Chapter II, section 2.9.3**). The grids were screened for suitability for data collection using Titan Krios (ThermoFisher Scientific) transmission electron microscope operating at 300 KeV; the grid screening revealed thin amorphous ice thickness, with well spread particles (**Figure 6.1**), grids that were suitable for full data collection.

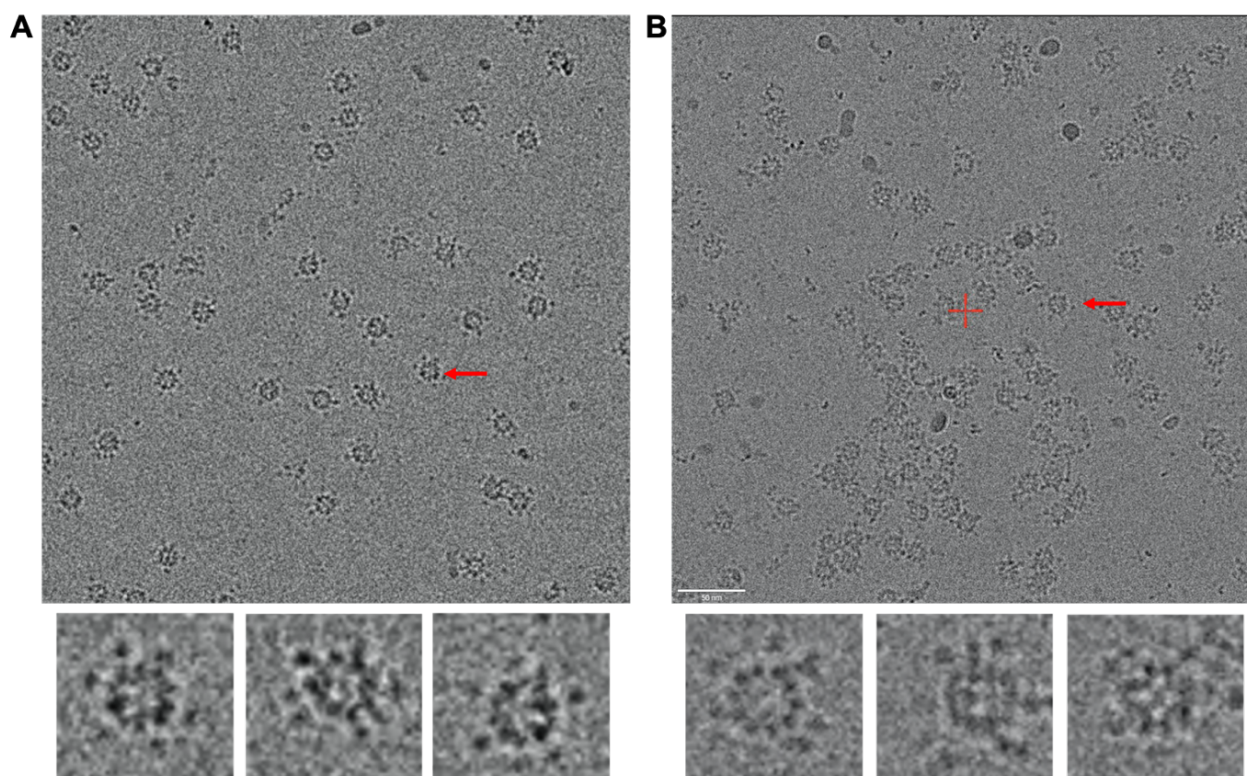


Figure 6.1. Cryo-EM micrographs of the *LmoRS* and *LmoRSX* stressosome complexes

Screening of the cryo-EM grids for *LmoRS* (A) and *LmoRSX* (B). The micrographs were taken on 1.2/1.3 Quantifoil grids. The stressosome complexes are picked out by red arrows. Scale bar 50 nm.

6.3. Cryo-EM data processing

The Leeds team collected approximately 10,000 micrographs in total for the two stressosome samples (Table 2.10 for *LmoRS* and Table 2.11 for *LmoRSX*) and the data were processed using the Relion 3.1 pipeline (Fernandez-Leiro and Scheres, 2017; Scheres, 2016; Zivanov et al., 2018) as detailed in the flow-chart in Figure 6.2 for *LmoRS* and Figure 6.3 for *LmoRSX*.

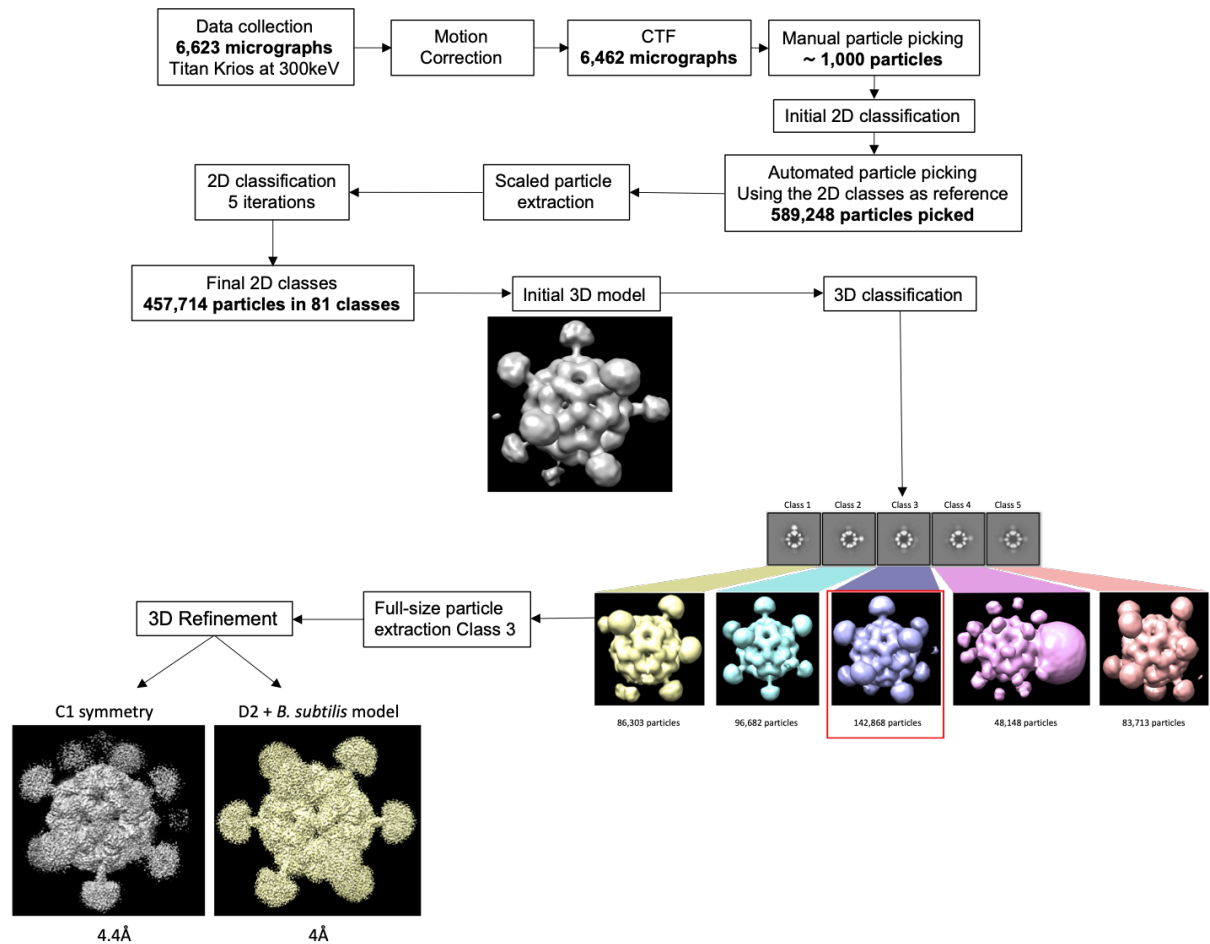


Figure 6.2. Cryo-EM data processing flow-chart for the *LmoRS* complex

The data were processed using the Relion3.1 pipeline. The final data set was processed in C1 and D2 symmetries at final resolutions of 4.4 Å and 4 Å respectively.

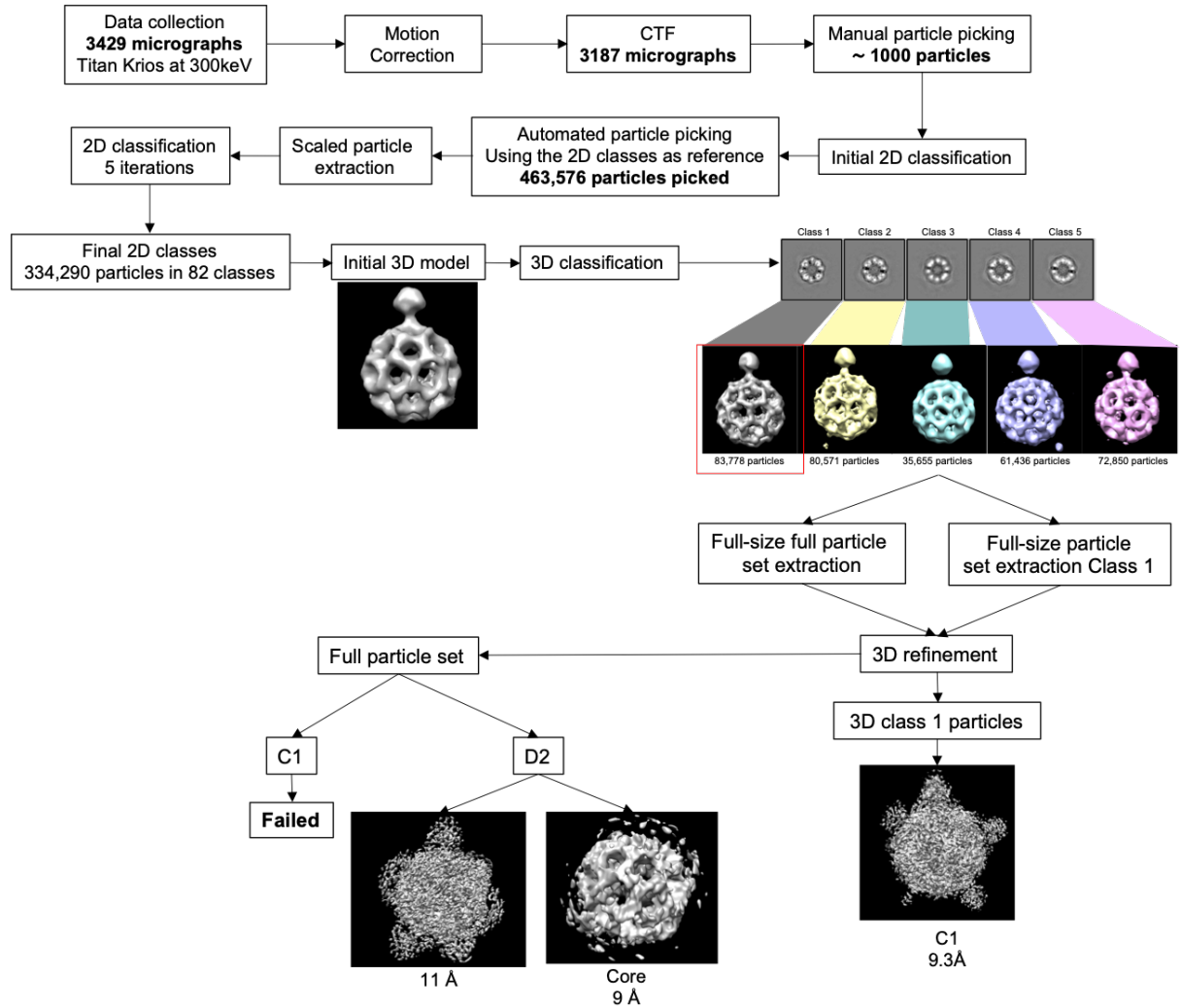


Figure 6.3. Cryo-EM data processing flow-chart for the *LmoRSX* complex

The data were processed using the Relion pipeline. The data were processed in C1 and D2 symmetries using different subsets of particles and with a focus on the core of the complex. The obtained maps were of insufficient quality to continue with analysis.

The micrograph movies were motion corrected using Motioncorr (Zheng et al., 2017), which allows the correction of whole frame image motions that occurred during data acquisition. Following the motion correction, the micrographs were also CTF corrected using CTFFIND4.1 (Rohou and Grigorieff, 2015) to correct aberrations from the microscope optics and the CTF of the microscope. The power spectra obtained from the CTF were inspected manually, and micrographs with poor power spectra, or crystalline ice contamination, were removed from further data processing: from the 6623 micrographs for *LmoRS* and 3429 micrographs collected for *LmoRSX*, 161 and 242 micrographs, respectively, were removed due to poor quality. From the remaining micrographs, an initial manual particle picking of approximately 1000 particles was performed. The manually picked particles were extracted and

reference-free 2D class averages were obtained, using a particle-edge size of 380 pixels and a box size of 400 pixels. Ten initial 2D classes were used as a reference for the automated particle picking on the whole datasets. The automated particle picking job selected 589,248 particles for *LmoRS* and 463,576 particles for *LmoRSX*. These particles were extracted by first scaling down the particle size to 196 pixels and second subjected to five iterations of 2D classification. With the 2D classification iterations, most of the incorrectly picked particles were removed. The final 2D classification gave 81 classes for *LmoRS* and 82 for *LmoRSX* with 457,714 and 334,290 particles in total, respectively (**Figure 6.4**).

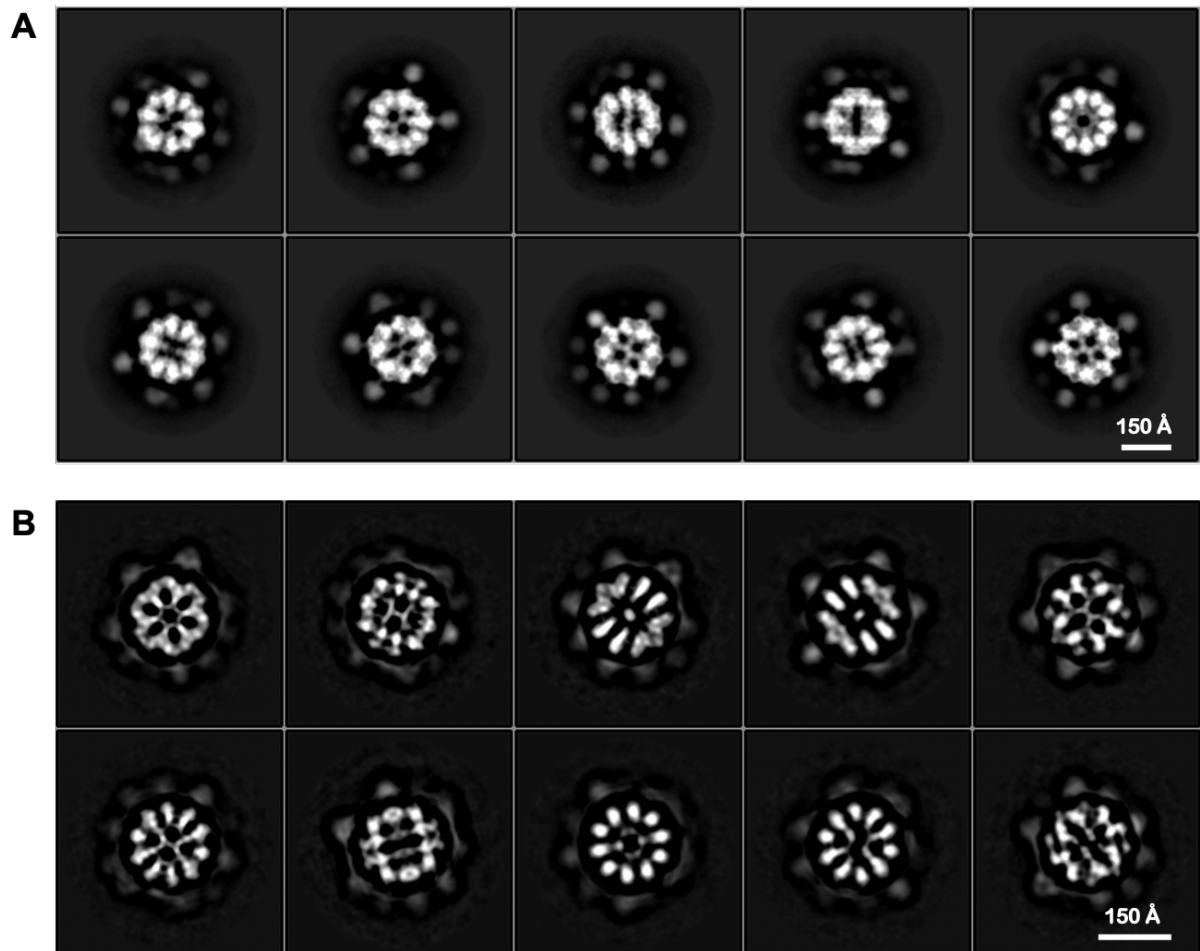


Figure 6.4. Examples of 2D classes for *LmoRS* and *LmoRSX* particles

Ten 2D classes for *LmoRS* (**A**) and *LmoRSX* (**B**) are shown. The 2D classes are displayed at a sigma contrast of 5 and the background was darkened for better contrast with the particle.

An initial 3D model was generated for each dataset using the sets of particles selected from the previous 2D classification. The initial models were generated without any symmetry constraints and were used as a reference for the 3D classification. The particles were split into five 3D classes (**Figure 6.5**). On the central slices of the 3D classes, the core of the stressosome is well defined for both datasets.

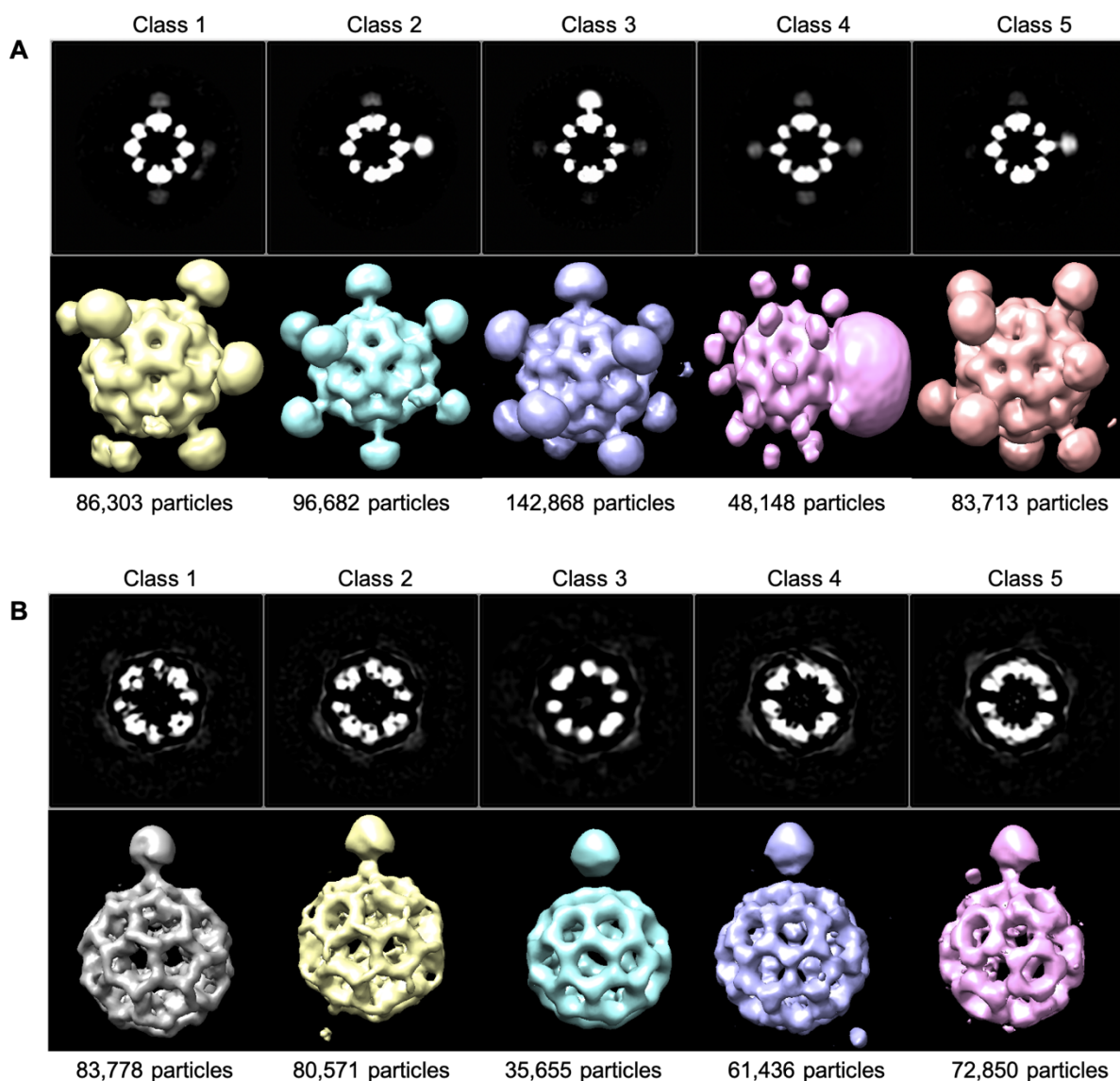


Figure 6.5. 3D classification of *LmoRS* and *LmoRSX* particles

Obtained 3D classes for *LmoRS* (**A**) and *LmoRSX* (**B**) particles with respective 2D sections of the 3D classes at a sigma contrast of 5 with a corrected background (above) and 3D maps of each 3D class (below). Class 3 was chosen for further processing of the *LmoRS* data (see below). For the *LmoRSX* data, due to high similarity between all 5 classes, all the particles were used for further processing.

For *LmoRS*, turrets corresponding to the N-terminal domains of RsbR1 can clearly be seen in classes 2, 3 and 5, however, just one turret in each case is well defined. By contrast, the turrets in *LmoRSX* are barely visible. The turrets have been found to be highly flexible in previously published reconstructions (Marles-Wright et al., 2008; Williams et al., 2019) and, therefore, it is quite difficult to observe them at high resolution. *LmoRS* class 4 (**Figure 6.5A**) displays an additional large density feature attached to the core and the distribution and volume of the turrets are dissimilar to the other four classes, so class 4 was not carried forward for further analysis. In the first instance, class 3 of the *LmoRS* data was selected for 3D refinement. Compared to the other classes, class 3 has many more particles; since all classes possess different turrets disposition there is probably a significant degree of heterogeneity in the sample. Therefore, the remaining particles in classes 1, 2 and 5 were also not used for further processing.

All the five 3D classes from the *LmoRSX* data look similar (**Figure 6.5B**), therefore, all the particles were used, in the first instance, in further data processing. In a second instance, the particles of the 3D class 1 were used for comparison between the whole set of particles and the subset of class 1 particles, as it is the class with the most particles. The particles of these classes were first re-extracted at their full size, which were then used to refine the 3D structure (3D refine). The refinement was first done without imposing any symmetry, i.e. in point group C1 (**Figure 6.6A**). A mask for each reconstruction used a threshold value determined in Chimera to include all the protein density, using a lowpass map at 15 Å. Using these masks, each map was post-processed to mask, sharpen, and to calculate gold-standard FSC curves.

Following mask creation and post-processing, the *LmoRS* data yielded a reconstruction at an overall resolution of 4.4 Å by the FSC (0.5) criterion. To compare with the reported D2 symmetry reconstruction for the *B. subtilis* stressosome (Marles-Wright et al., 2008), the *LmoRS* dataset were also refined against the *B. subtilis* map as a reference and with the imposition of D2 symmetry. An improvement of the *LmoRS* map was observed with an increase in resolution to 4 Å and better-defined turrets (**Figure 6.6B**).

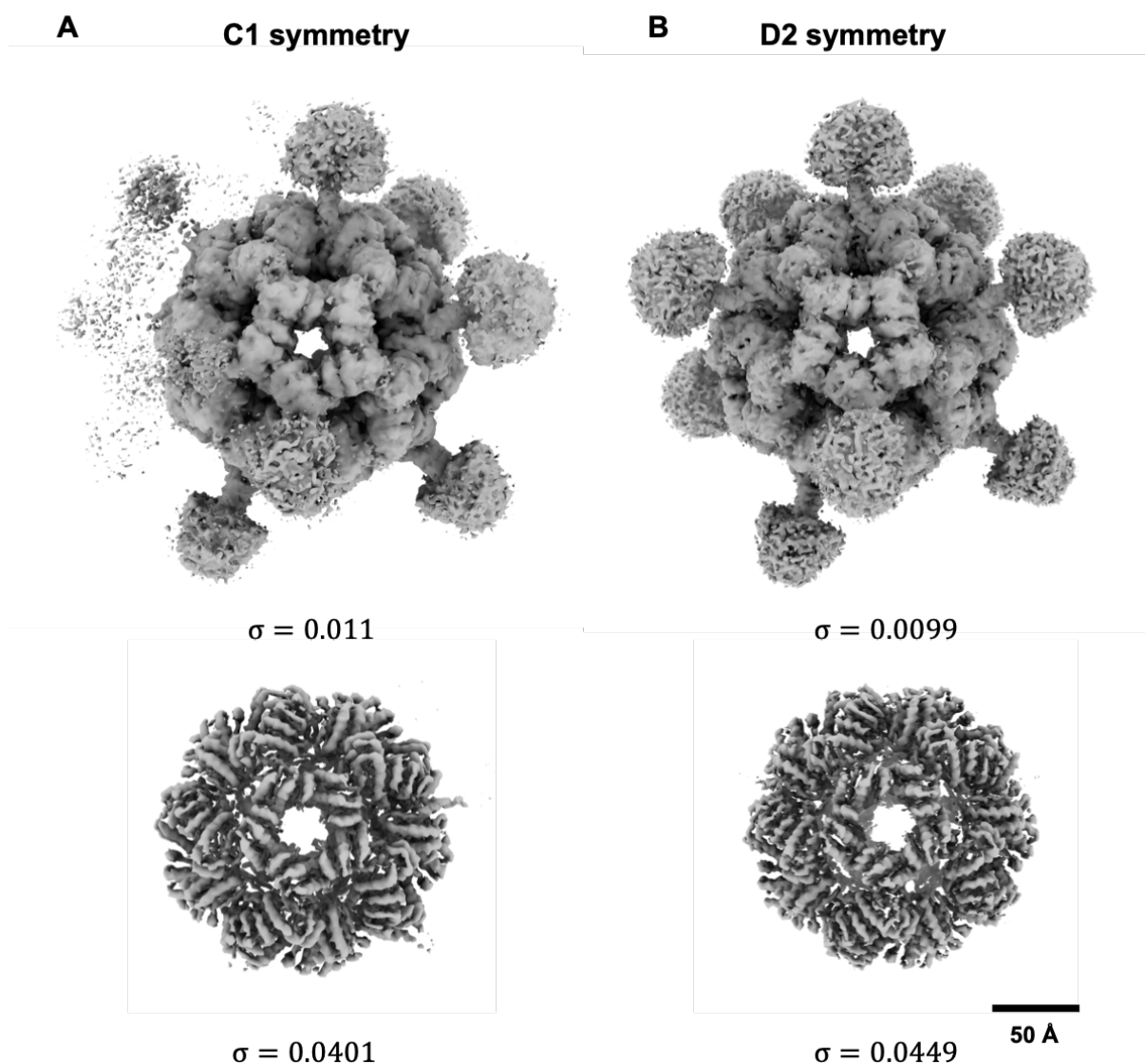


Figure 6.6. Single particle reconstruction of the *LmoRS* complex

The maps obtained in C1 (A) and D2 (B) symmetries display some structural differences. In the C1 map, half of the map has well-defined turrets while noise-like turrets are observed on the other side of the molecule. By contrast, all the turrets are well defined in the D2 map. When the sigma level is increased, the core displays similar arrangements between the C1 and D2 symmetries, in which secondary structural features are clearly visible.

Whereas the C1 symmetry 3D refinement for *LmoRSX* failed to converge on a map, which is probably due to the particles distribution on the grid, the D2 symmetry 3D refinement yielded the first map of this complex (Figure 6.7A), albeit at the low resolution of 11 Å by the FSC (0.5) criterion. Some turrets can be seen in this map, but no model can be built because of the low resolution. To have better insight of the *LmoRSX* core, a spherical mask was created and used during the 3D refinement job, which also yielded a low-resolution map - to 9 Å - and the RsbX density cannot be differentiated from the core of the complex (Figure 6.7B). Finally,

another 3D refinement job was performed using only the 3D class 1 particles in point group C1, yielding a rather featureless map at 9.3 Å (**Figure 6.7C**).

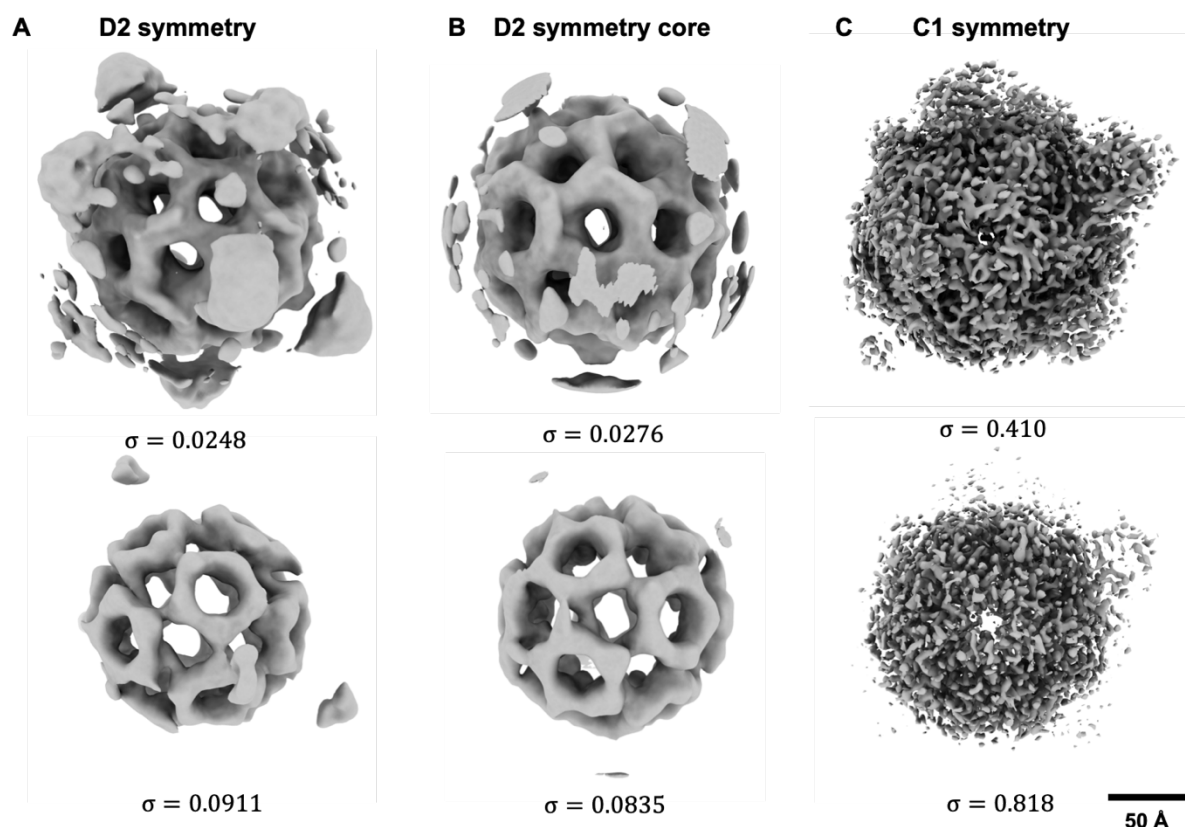


Figure 6.7. Single particle reconstruction of the *LmoRSX* complex

The maps obtained for the full set of particles in the D2 symmetry (**A**), with a spherical mask (**B**) and in the C1 symmetry from the 3D class (**C**) displays low resolution densities. When increasing the sigma level, the cores do not display any visible secondary features due to the low resolution of the maps.

Therefore, the reconstruction of the RSX complex could not be obtained from the current *LmoRSX* data. This might be because the particles are too close to each other, as stressosome complexes tend to stick together by their turrets (see **Figure 6.1**). Therefore, a new automated particle picking routine was performed with an inter-particle distance of 480 Å, but this yielded only 9,839 particles, which is insufficient for high-resolution reconstruction. Therefore, the *LmoRSX* data were not used further.

The *LmoRS* data yielded two sharpened and masked maps with a resolution of 4.4 Å in C1 and 4.1 Å in D2 symmetries. Using the Relion CTF refinement pipeline (Scheres, 2016; Zivanov et al., 2019a), the particles were subjected to a CTF and a per-particle defocus fitting, followed by a 3D refinement (Scheres, 2012). The output particles were then subjected to Bayesian polishing (Zivanov et al., 2019b), but due to the low resolution of the initial

reconstructions these CTF and polishing steps did not result in a higher resolution reconstruction. Indeed, the resolution increased to 6.4 Å in C1 and 7 Å in D2 symmetries. Judging by this increase of resolution and a poor map reconstruction, the CTF refinement and the Bayesian polishing steps did not improve the quality of the resultant maps. Therefore, the final map used for model building were the maps obtained after the first refinement in both symmetries (**Figure 6.6**).

6.4. Atomic model of the *LmoRS* reconstruction

6.4.1. *LmoRS* complex model in C1 symmetry

The stressosome complex in C1 symmetry has a heterogeneous distribution of turrets (**Figure 6.6A**). On one side of the complex, there are well defined turrets arranged in a manner that follows D2 symmetry when compared to the D2 map (**Figure 6.6B**). By contrast, heterogeneous and less well-defined turrets are found on the other side of the molecule (**Figure 6.6A**). The FSC was calculated using 2 types of mask for solvent flattening and map sharpening: a non-spherical mask (**Figure 6.8A**) and a spherical mask without the middle part as the stressosome core is empty. The FSC-determined resolutions obtained for the spherical mask and the non-spherical mask are of 4.7 Å and 4.4 Å, respectively (**Figure 6.8B and 6.8C**).

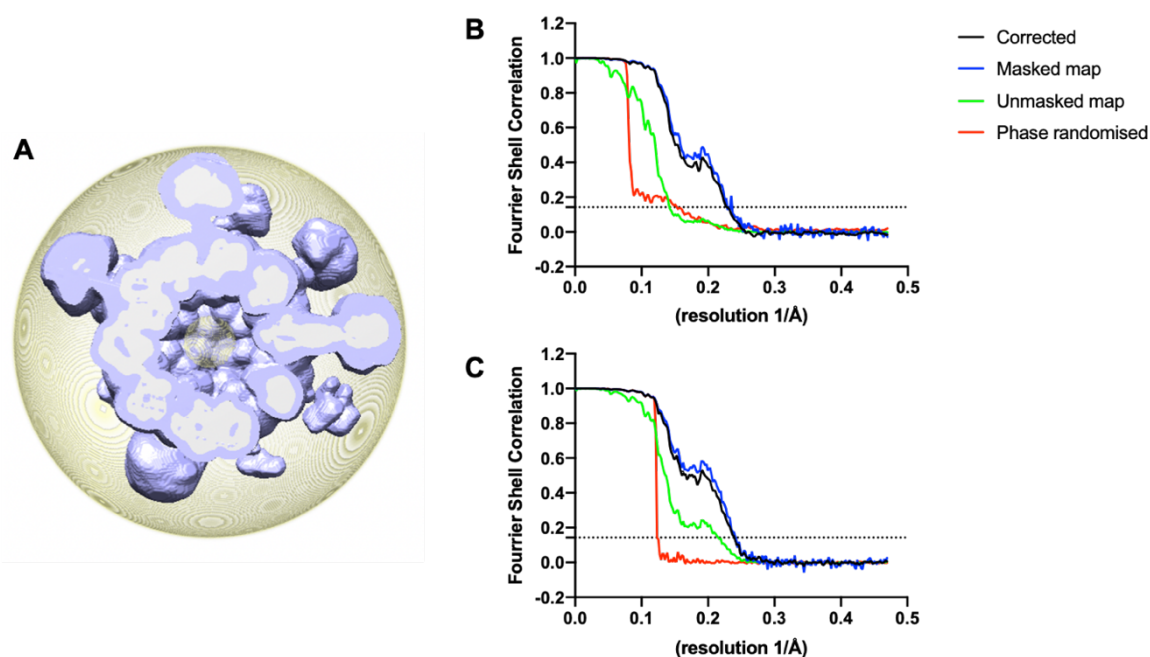


Figure 6.8. FSC curves and masks for the *LmoRS* C1 reconstruction

Two types of masks were used and superimposed (**A**): a spherical mask with an outer radius of 150 Å and an inner radius of 25 Å which excludes the core's void (in yellow mesh); and a non-spherical mask around the particle (in purple) at a 0.0016 threshold. The FSCs in **B** and **C** are: the masked map (green); the unmasked map (blue); the corrected FSC with randomised phases from the masked map (red); and the corrected FSC (black). The FSCs obtained with the purple mask (**B**) and the spherical mask (**C**) have resolutions of 4.4 Å and 4.7 Å, respectively.

A local resolution map was calculated in Relion3.1 (**Figure 6.9**) using the non-spherical masks. The maximum resolution achieved was 19.37 Å, with a mean of 7.33 Å and a median of 5.57 Å. The resolution of the whole map varies hugely, and the flexibility of the turrets in comparison to the core can be seen with a large resolution decrease from 4.42 Å to 9.12 Å.

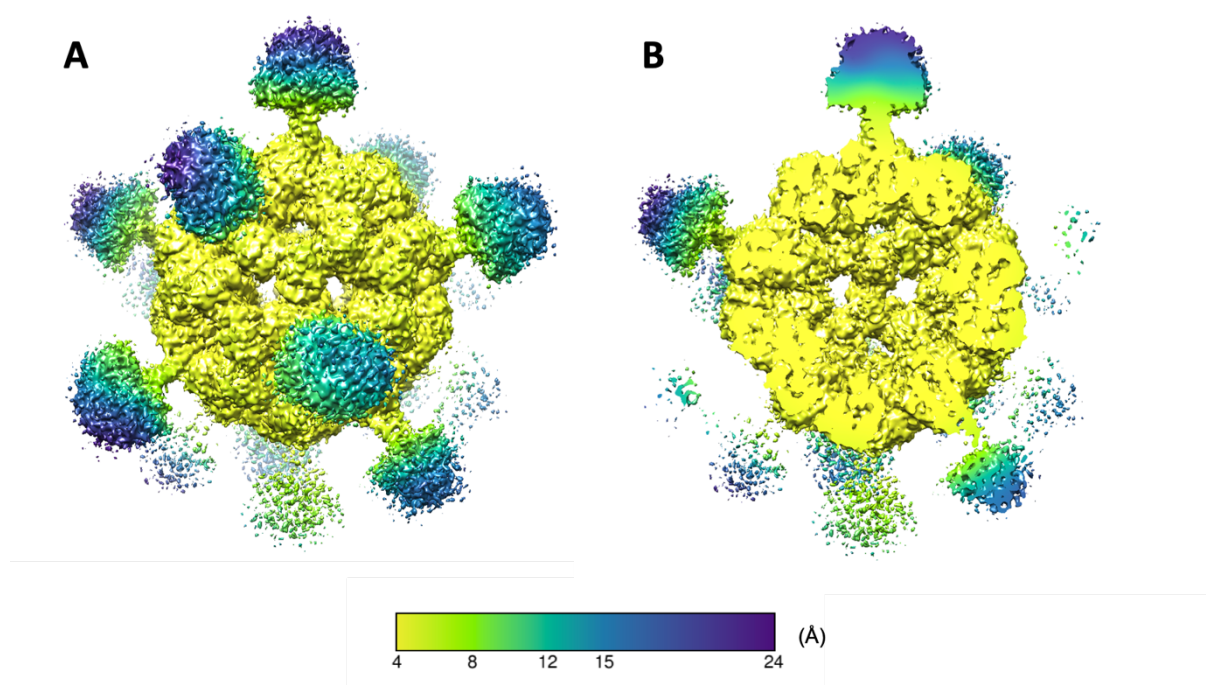


Figure 6.9. Local Resolution map of the *LmoRS* C1 map

The whole (A) and sliced view (B) of the reconstructed map was colored using the viridis coloring palette according to resolution, from yellow at 4 Å to purple at 24 Å.

Homology models were generated for the RsbR1 and RsbS STAS domains using the Phyre2 webserver (**Figure 6.10**) (Kelley et al., 2015). The RsbS model was built with 99.9% confidence and covers 97% of the sequence based on the crystal structure of *B. subtilis* RsbS (PDB ID:6JHK, (Kwon et al., 2019)). The RsbR1 STAS domain model was also built with 99.9% confidence and covers 91% of the sequence based on the crystal structure of *M. thermoacetica* RsbS (PDB ID: 2VY9, (Marles-Wright et al., 2008)). The newly determined N-RsbR1 crystal structure was used for modelling the N-RsbR turrets, (**Chapter III, section 3.9**).

These atomic models were placed in the map manually using Chimera (Pettersen et al., 2004). The C1 map had to be flipped on the z-axis to obtain the correct hand. The fitted models were real space refined manually in Coot (Emsley and Cowtan, 2004) and refined using real-space refinement (Afonine et al., 2018; Liebschner et al., 2019) in the Phenix pipeline. First, the core of the complex was built as the core has better resolution than the turrets. Second, the turrets were built using the N-RsbR1 model.

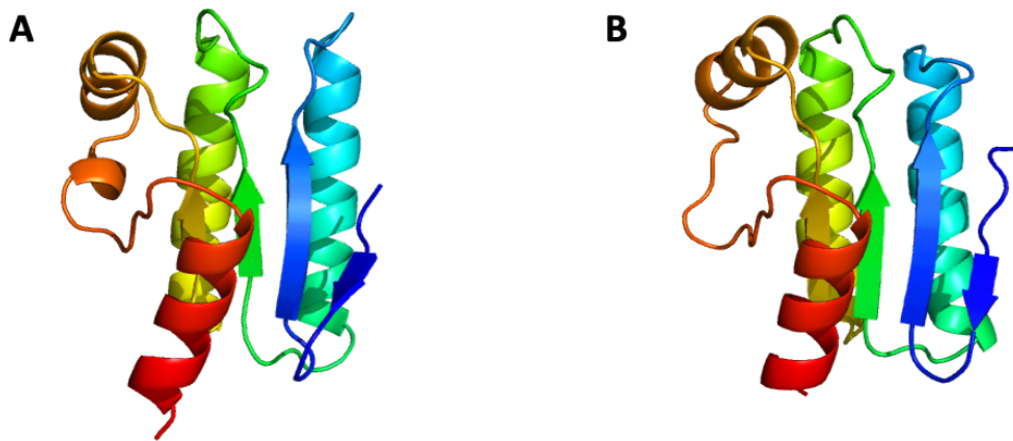


Figure 6.10. Homology models for RsbR and RsbS STAS domains

The RsbR (A) and the RsbS (B) STAS domain models were generated using Phyre2. The models are displayed in a cartoon fashion and colored in rainbow from N-terminus (blue) to the C-terminus (red).

The low resolution of the turrets made their model building difficult; therefore, the turret building was guided by the well-defined J-helices that link the N-terminal turrets to the C-terminal STAS domain. The crystal structure of the N-RsbR1 dimer did not fit properly to the turrets in the EM map as the J-helices are crossed in the EM map but were not crossed in the crystal structure, possibly an artefact of working with the isolated domain (**Figure 6.11**).

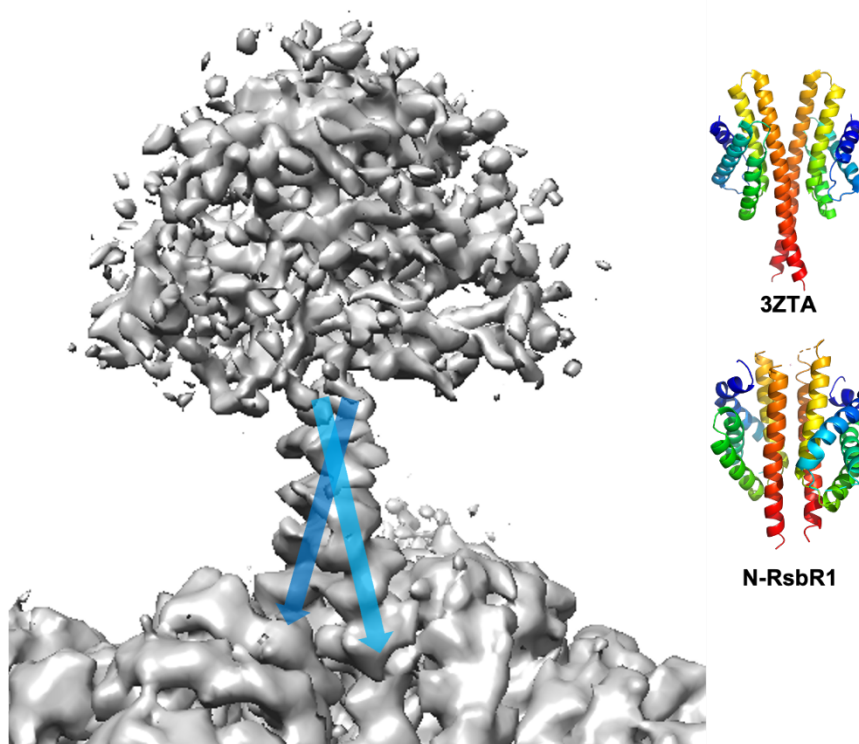


Figure 6.11. J-helix conformation in the *LmoRS* C1 map

The turret of the *LmoRS* map shows the crossed J-helices which are indicated with blue arrows. The crossed J-helix conformation is similar to that of the 3ZTA crystal structure in contrast to the newly solved N-RsbR1 crystal structure.

Therefore, the N-RsbR1 crystal structure was used as a monomer and was placed by superimposition to 3ZTA (Quin et al., 2012), which maintains the crossed pair of J helices. The final model was subjected to the comprehensive validation tool for Cryo-EM in phenix (Afonine et al., 2018) with validation statistics summarized in **Table 6.1**.

Table 6.1. C1 map processing and model refinement/validation statistics

<i>Data processing</i>	
Symmetry imposed	C1
Initial particle images	589,248
Final particle images	142,868
Map resolution (masked A)	4.01
Map resolution (unmasked A)	4.06
FSC threshold	0.143
<i>Refinement</i>	
Model composition	
Atoms	90580
Residues	11609
R.M.S.D	
Bond length (Å)	0.009
Bond angles (°)	1.102
B-factors (Å ²)	
Protein (min/max/mean)	34.19/627.23/149.32
Validation	
Molprobit score	2.99
Clash score	30.41
Poor rotamers (%)	4.36
Ramachandran plot (%)	
Outliers	0.10
Allowed	9.42
Favoured	90.49
Model vs. Data	
CC (mask)	0.60
CC (box)	0.76
CC (main chain)	0.62
CC (side chain)	0.65

The C1 map and model obtained is close to those already published. 22 RsbRA turrets and 8 RsbS dimers were observed in D2 symmetry by Kwon *et al* (Kwon et al., 2019); by contrast Marles-Wright *et al* reported 20 RsbRA turrets and 10 dimers of RsbS in the D2 symmetrised reconstruction of the *B. subtilis* RsbRA:RsbS complex (Marles-Wright et al., 2008). This stoichiometry was also seen in the C1 symmetry reconstruction of *Lmo*RsbR1:RsbS (Williams et al., 2019). The new C1 map supports a model with 17 turrets, of which 8 have strong electron density and 9 has somewhat poorer density (**Figure 6.12**). The turrets with good

density are found on the D2-like side of the particle while the less well-defined turrets are found on the other side of the particle. Having 17 turrets makes a total of 34 RsbR1 molecules and 26 RsbS molecules in the overall quasi-icosahedral complex.

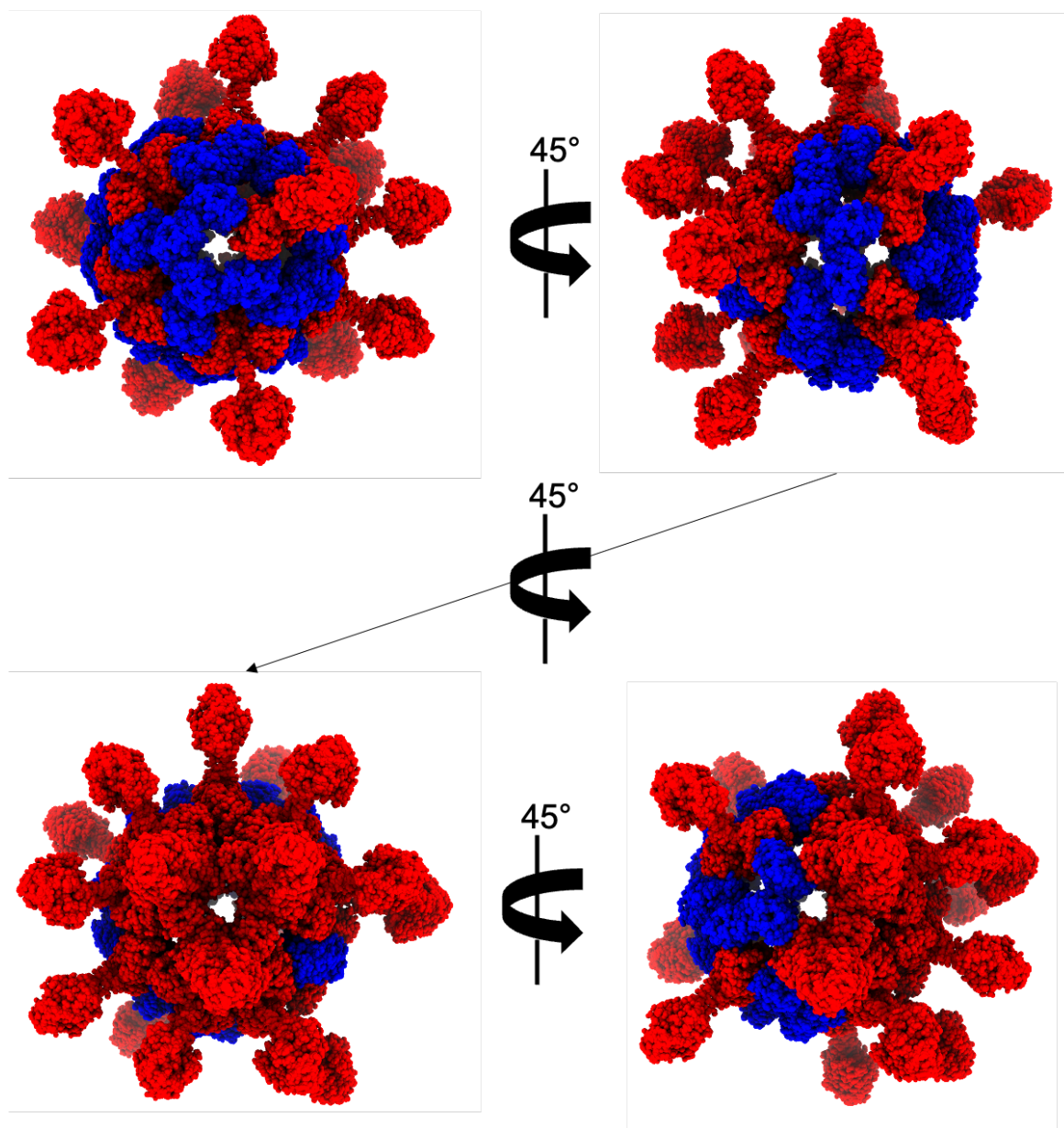


Figure 6.12. Atomic model of the C1 *LmoRS* reconstruction showing the distribution of the RsbR1 turrets.

The stressosome model was built using homology models for RsbS and RsbR1 STAS domains and the N-RsbR1 crystal structure for the turrets. The models are shown in different views 45° apart, and are represented by atomic spheres, in which RsbR1 is coloured red and RsbS is coloured blue.

However, this observation is not in accordance with the ratio of the complex determined by SDS-PAGE at 2:1 for RsbR1:RsbS (**Chapter V, section 5.6**). Indeed, the ratio from that reconstruction is approximately 3:1 ratio for RsbR1:RsbS is also not consistent with the 2:1 ratio for RsbRA:RsbS described by Marles-Wright *et al* in both the EM reconstruction and from

SDS-PAGE densitometry, and which is consistent with the 1 MDa mass of the particle from analytical ultracentrifugation (Chen et al., 2003). In **Figure 6.12**, it can also be seen that the RsbS proteins are all concentrated on one side of the core, which is the D2-like side of the particle.

The final resolution achieved is 4 Å, yielding clear secondary structure elements in the maps to allow the overall fold of the STAS domains to be traced (**Figures 6.12 and 6.13**). However, the final model obtained (**Figure 6.12**) does not allow us to infer further conclusions on the structure and properties of the stressosome, including its activation mechanism, because the refinement failed due to the poor electronic potential map in the vicinity of the turrets.

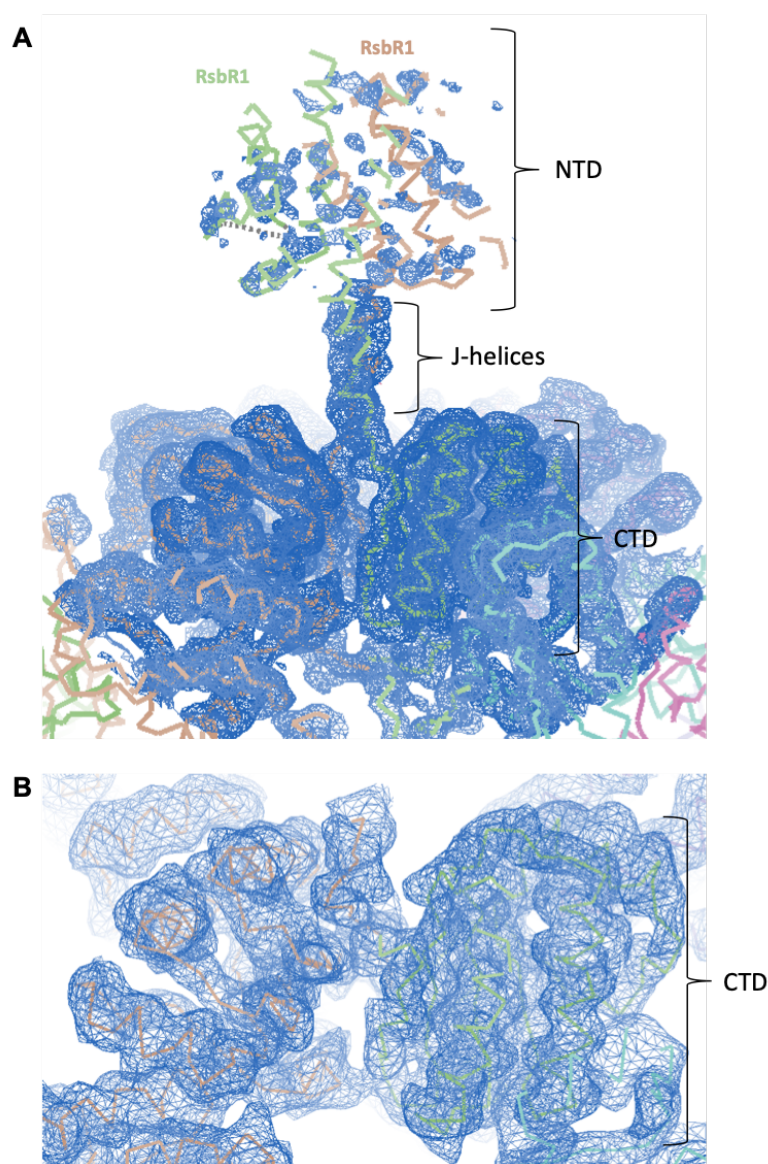


Figure 6.13. Electronic potential map of the C1 *LmoRS* stressosome reconstruction

The electronic potential map for the C1 *LmoRS* stressosome (blue mesh) shows tubes of the secondary structures with the models' C α backbone and the dashed lines the unmodelled loops in **A**. The turrets formed by the RsbR1 dimers (NTD) (**A**) have weaker density while the RsbR1 STAS domains have stronger density (CTD) (**B**).

The electronic potential map is sufficient to describe the stressosome assembly in C1 symmetry (discussed in **section 6.4.3**), but overall, the validation statistics of the model suggest that it is no better than average (**Table 6.1**).

6.4.2. *LmoRS* complex model in D2 symmetry

The stressosome map in D2 symmetry shows a homogeneous turret distribution (**Figure 6.6B**) compared to the C1 map. The turrets are better defined than in the C1 symmetry reconstruction probably due to the signal to noise gain from averaging the particles.

The FSC resolution was also calculated using two types of masks, as already described for the C1 map: the same spherical mask as the C1 map, with the mask threshold set at 0.0116,

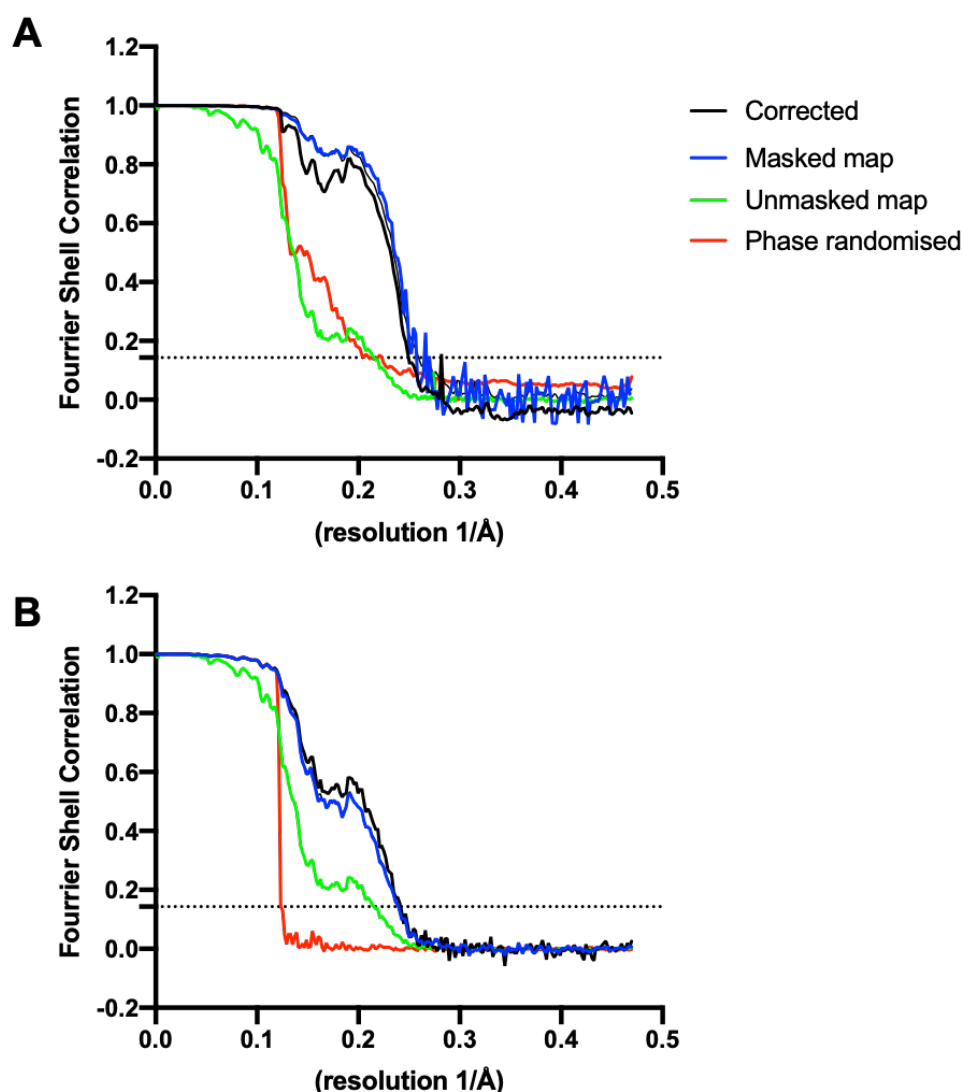


Figure 6.14. FSC curves for the *LmoRS* reconstruction in D2 symmetry

The FSC curves obtained with the particle mask (**A**) have an estimated resolution of 4.0 Å and that for the spherical mask (**B**) results in a resolution of 4.2 Å. The dotted line is the FSC at 0.143. The graph shows FSCs for the masked map (green); the unmasked map (blue); the corrected FSC with randomised phases of the masked map (red); and the corrected FSC (black).

and a mask based on the particle itself. The FSC curves obtained are displayed in **Figure 6.14A** and **6.14B**. The resolution obtained for the spherical and the particle mask is 4.2 Å and 4.0 Å, respectively.

The local resolution of the D2 map reconstruction was generated in Relion3.1, with a B-factor of 201 Å, and can be seen in **Figure 6.15**. The maximum resolution achieved was 10.60 Å, with a mean of 4.93 Å and a median of 4.10 Å. The core has an overall resolution of 3.68 Å, which is better than that of the C1 map with a core resolution of 4.04 Å. The turrets remain poorly defined with a resolution of 10.46 Å but they are better defined than the C1 map, in which the turrets' resolution was 24.35 Å.

Using the same model building process as the C1 map, the D2 map was fitted with the homology models for the STAS domains of RsbS and RsbR1 and the N-RsbR1 crystal structure. As in the C1 map, the J-helices are also crossed in the D2 map and the model building of the turrets was performed using 3ZTA as a template.

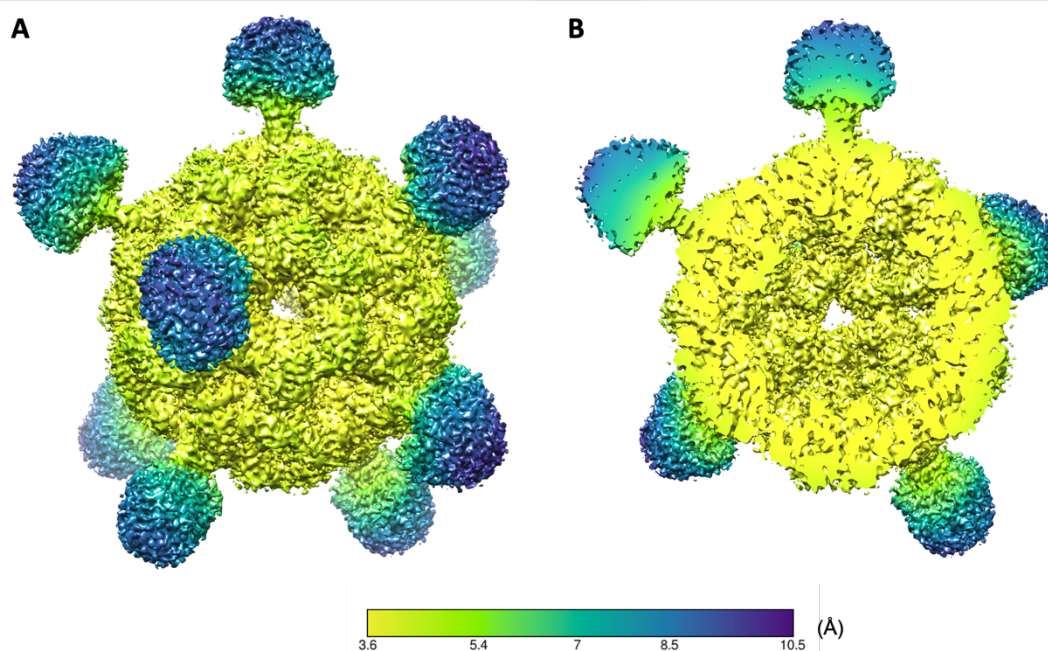


Figure 6.15. Local Resolution map of the *LmoRS* D2 map

The whole (A) and sliced view (B) of the D2 reconstructed map was colored using the viridis coloring palette according to resolution from yellow at 3.6 Å to purple at 10.5 Å.

The D2 map allowed the building of 10 turrets of RsbR1 (hence 20 molecules) and 40 molecules of RsbS (**Figure 6.16**). This observation reverses the RsbR1:RsbS ratio to 1:2, which is also in contrast to previous observations, where 40 RsbRAs (or RsbR1s) were found for 20

RsbS (Kwon et al., 2019; Marles-Wright et al., 2008; Williams et al., 2019). The final model was validated as described below and the validation statistics are summarized in **Table 6.2**.

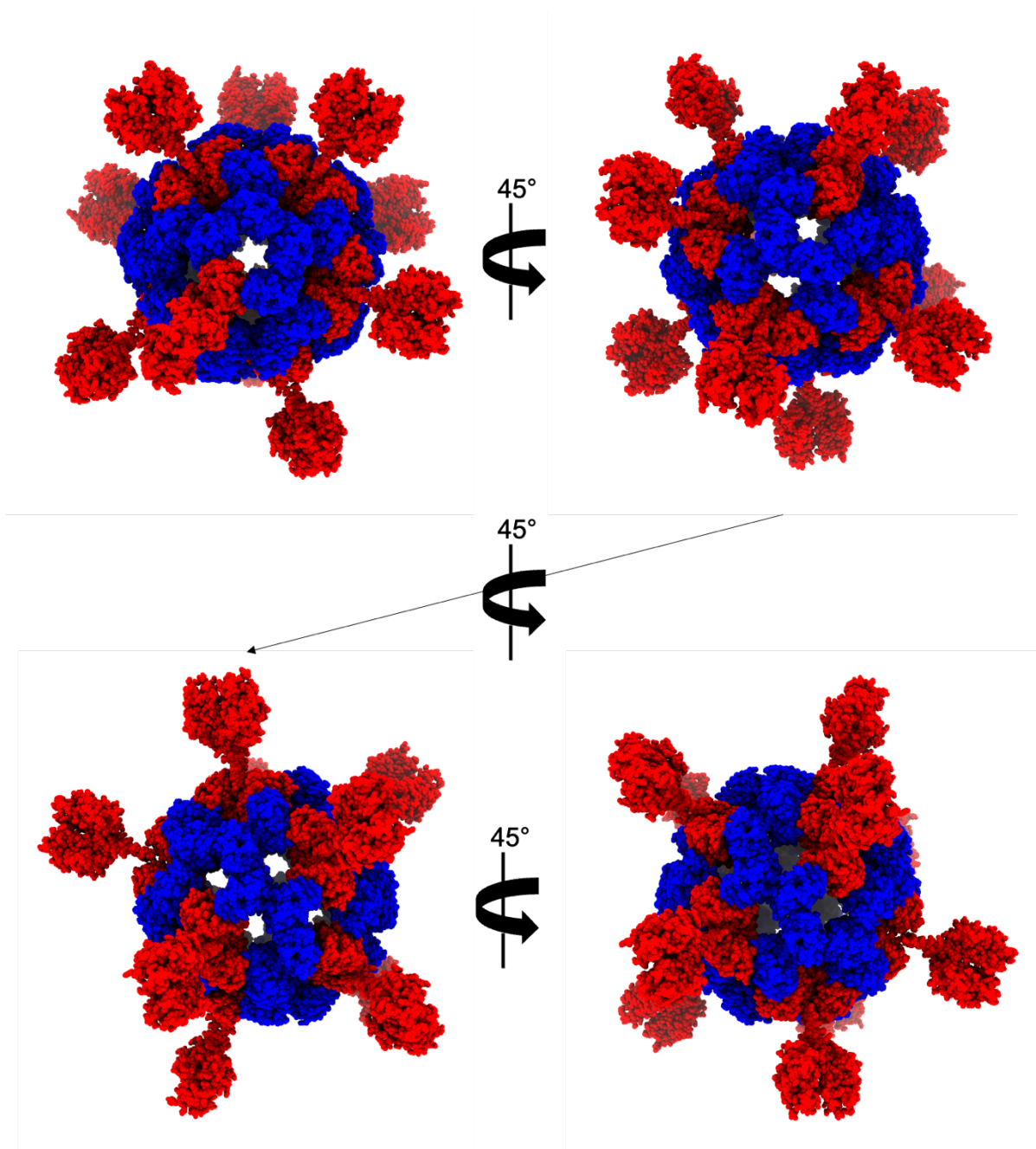


Figure 6.16. Atomic model of the D2 *LmoRS* reconstruction

The stressosome model was built using homology models for RsbS and RsbR1 STAS domains for the core and the N-RsbR1 crystal structure for the turrets. The models are shown in different views 45° apart and are represented by atomic spheres in which RsbR1 is coloured red and RsbS is colored blue.

Table 6.2. D2 map data processing and model refinement statistics

<i>Data processing</i>	
Symmetry imposed	D2
Initial particle images	589,248
Final particle images	142,868
Map resolution (masked A)	3.64
Map resolution (unmasked A)	3.68
FSC threshold	0.143
<i>Refinement</i>	
Model composition	
Atoms	74500
Residues	9670
R.M.S.D	
Bond length (Å)	0.009
Bond angles (°)	1.380
B-factors (Å ²)	
Protein (min/max/mean)	9.33/514.17/153.29
Validation	
Molprobity score	3.19
Clash score	28.81
Poor rotamers (%)	10.49
Ramachandran plot (%)	
Outliers	0.21
Allowed	7.39
Favoured	92.40
Model vs. Data	
CC (mask)	0.79
CC (box)	0.85
CC (main chain)	0.78
CC (side chain)	0.78

As with the C1 map, the low resolution of the D2 map does not allow side chains to be visualised, which excludes any significant study of intramolecular interactions. Compared to the C1 model, the D2 model has a lower outlier score and lower bond length/angles R.M.S.Ds. The medium Molprobity score suggests that this model is as good as any other structure at this resolution (Chen et al., 2010).

The high correlation coefficients ($CC \leq 0.5$) also suggest that the model fits the data properly after a refinement cycle. Therefore, the model obtained for the *LmoRS* complex in D2 symmetry is the most reliable of all the reconstructions of this thesis, however, the side chains and turrets still cannot be built confidently due to a lack of density.

6.4.3. *LmoRS* complex assembly

Although the resolution of the maps does not allow proper model building, the assembly of the *LmoRS* stressosome complex can still be studied. The C1 and D2 map symmetries gave two different assemblies of the *LmoRS* complex. Both models have 20 hexagons and 12 pentagons as previously reported (Marles-Wright et al., 2008; Williams et al., 2019) and both display a truncated icosahedral core symmetry.

As seen previously in the C1 symmetry model (**Figure 6.17A/B**), there is a heterogeneous distribution of the RsbR1 and RsbS proteins: the RsbR1 and RsbS molecules are each concentrated on opposing sides of the complex. The C1 symmetry comprises six different types of faces: three pentamers (**Figure 6.17C**) and three hexamers (**Figure 6.17D**). The pentamers set comprise either: three monomers of RsbR1 and two monomers of RsbS; two RsbS dimers and one RsbR1 monomer; or five RsbR1 monomers (**Figure 6.17C**). The hexamers are composed of either: one RsbS monomer, two RsbS dimers and one RsbR1 dimer; one RsbS dimer and two RsbR1 dimers; or two RsbS dimers and one RsbR1 dimer (**Figure 6.17D**). The heterogeneity of the *LmoRS* complex in C1 symmetry is underlined by the make-up of these heterogeneous faces.

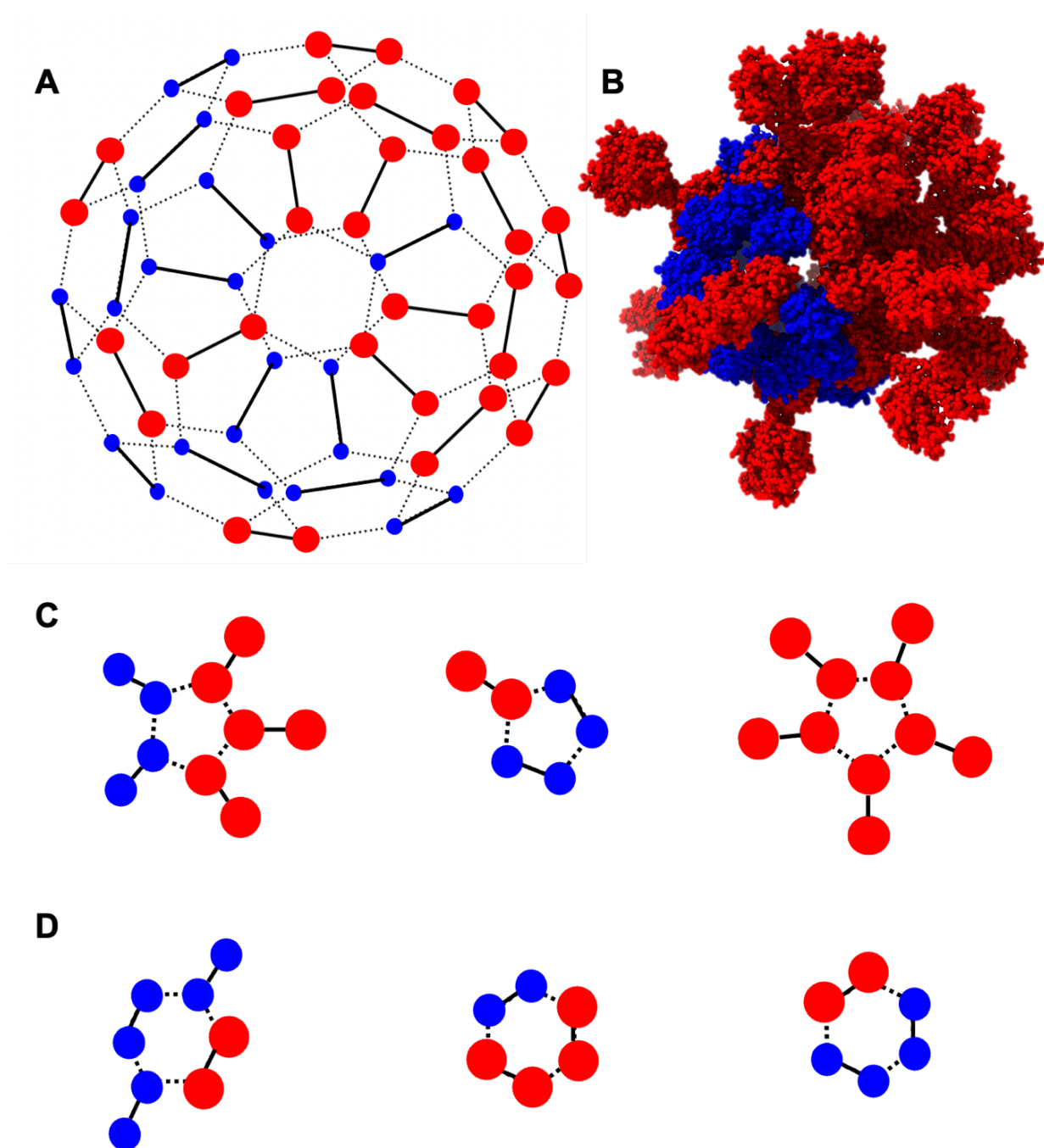


Figure 6.17. Schematic representation of the *LmoRS* complex assembly in C1 symmetry

The *LmoRS* complex assembly in the C1 symmetry model (A) and the respective molecular model (B) with RsbR1 monomers in red and RsbS monomers in blue. The homodimers are linked by a solid line and the inter-protein interactions are represented with dotted lines. The core of the complex displays 12 pentamers (C) and 20 hexamers (D), with three different combinations of each.

Compared to the C1 symmetry model, the D2 model displays a more homogeneous distribution between RsbR1 and RsbS proteins within the *LmoRS* complex (**Figure 6.18**): only three repetitive units of hexagons and pentagons were obtained. The sole hexagonal repetitive unit is formed by one RsbR1 dimer and 2 RsbS dimers (**Figure 6.18C**). The two pentagonal types are composed of either three RsbS monomers and two RsbR1 monomers or four RsbS monomers and one RsbR1 monomer (**Figure 6.18D**).

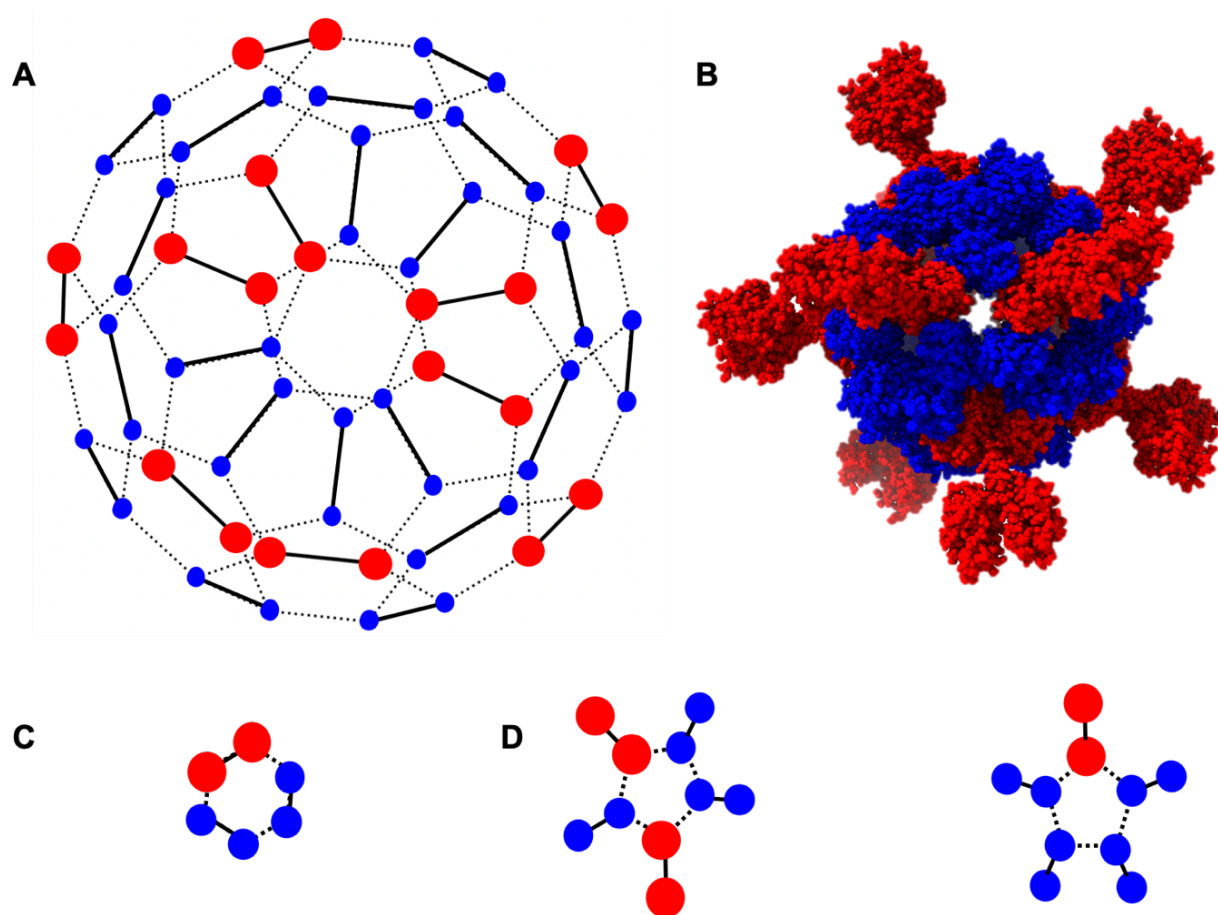


Figure 6.18. Schematic representation of the *LmoRS* complex assembly in D2 symmetry

The schematic of the *LmoRS* complex assembly for the D2 symmetry model (**A**) and the associated molecular model representation (**B**) with RsbR in red and RsbS in blue. The homodimers are linked by a solid line and the inter-protein interactions are represented with dotted lines. The whole complex is formed of 20 hexagons and 12 pentagons. One type of hexagon composition (**C**) and two types of pentagons (**D**) are the repetitive units of the whole complex.

6.4.4. Phosphorylation site dispositions

The modest resolution of the maps does not permit a study of the local conformation of the phosphorylation sites in RsbR1 and RsbS, but it still does allow their overall positions to be determined. As discussed in the Introduction (**Chapter I, section 1.3.1**), the stressosome complex is phosphorylated by RsbT. The phosphorylated residues are found in the STAS domains of RsbR1 and RsbS, in the stressosome core. *B. subtilis* RsbT phosphorylates exclusively residues T171 and T205 in RsbRA and S59 in RsbS (Gaidenko et al., 1999). The equivalent residues in *L. monocytogenes* RsbR1 and RsbS are T175, T209 and S56, respectively (**Figure 1.6**). T175 is found on a flexible loop immediately preceding the first α -helix of the RsbR1 STAS domain, whereas T209 and S56 are both at the N-terminal end of the second α -helix of their respective STAS domains (**Figure 6.19**).

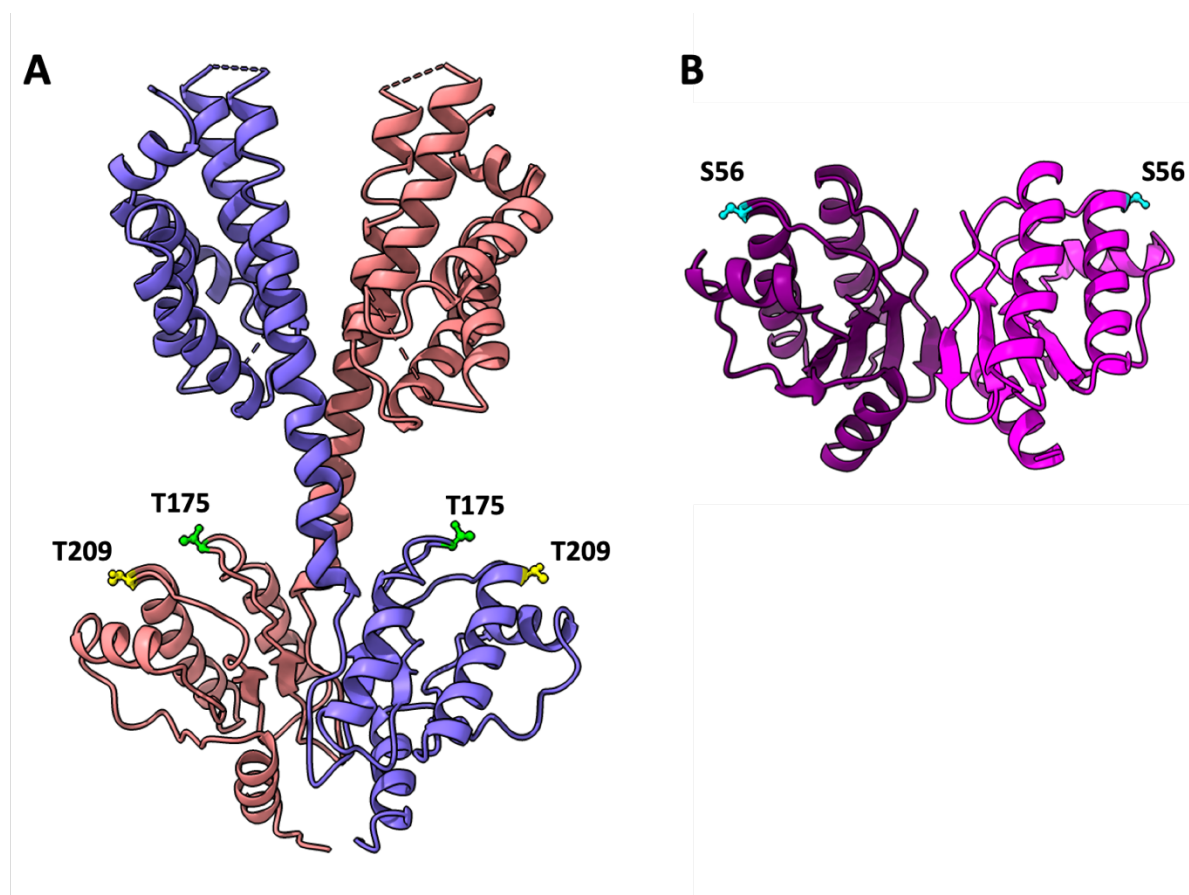


Figure 6.19. RsbT phosphorylation targets on RsbR1 and RsbS

The full-length RsbR1 dimer model (**A**) reconstructed in the D2 map shows the location of the phosphorylated residues on the surface of the STAS domain: T175 is coloured green and T209 is in yellow. The position of phosphorylated S56 (cyan) in RsbS (**B**) is also found on the surface of the STAS domain.

These residues are all exposed on the surface of the core, where RsbT is expected to bind (**Figure 6.20**). When the potential for RsbT binding is studied using the C1 symmetry model, the close clustering of homo-pentamer of RsbR1s (**Figure 6.20A** Line 1, 3rd view) creates a significant steric hindrance to the binding of RsbT to this part of the core. Therefore, the obtained map and consequent model of the *Lmo*RS complex in C1 could be viewed as an artefactual complex assembly.

By comparison, the model obtained in D2 symmetry (**Figure 6.20B**) shows a better all-round accessibility of the RsbR1 and RsbS phosphorylable residues to RsbT and to the RsbX phosphatase. Since the *Lmo*RSX data were too poor to permit the reconstruction of a ternary complex model, little further can be concluded about the RsbX binding site.

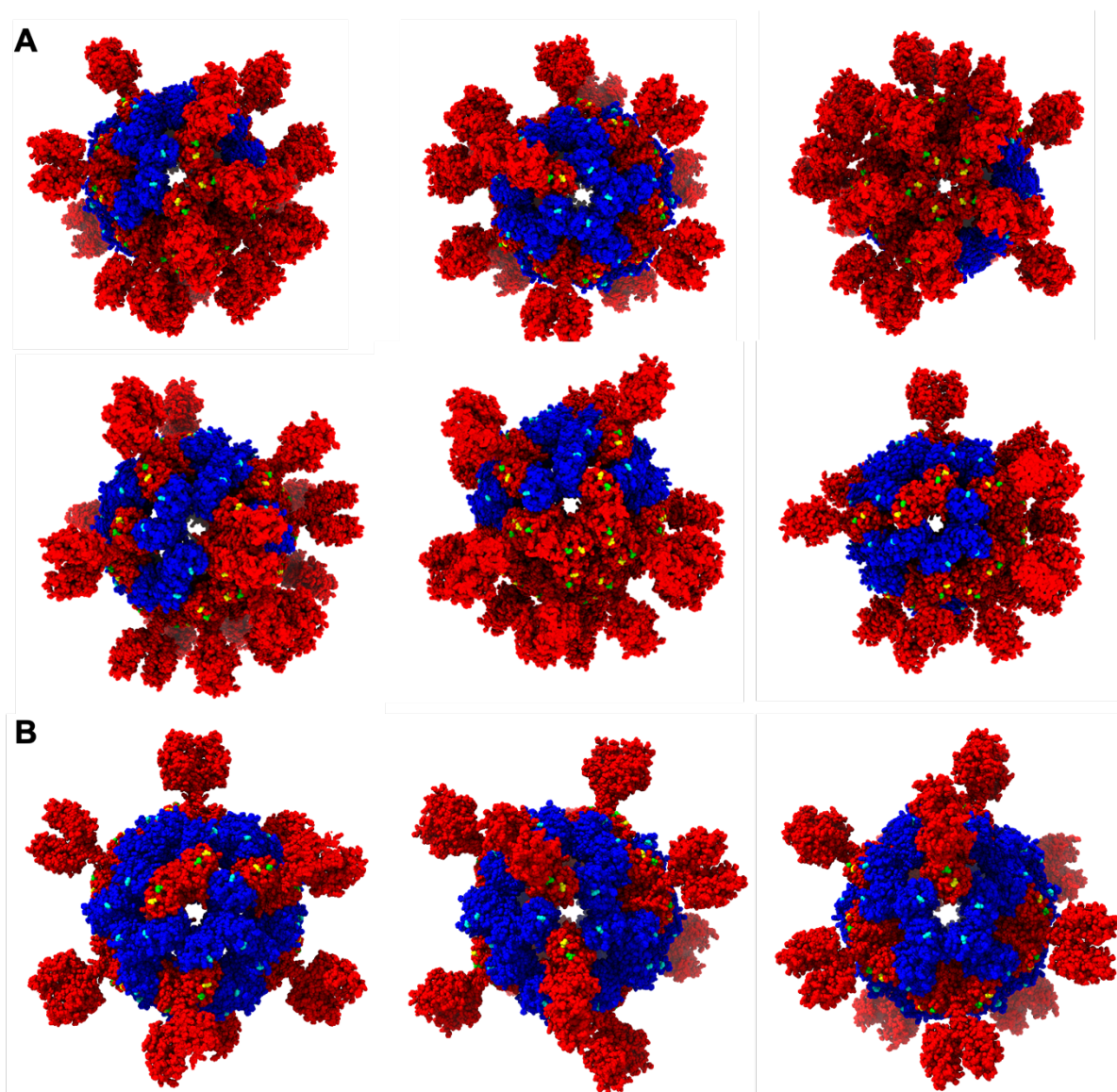


Figure 6.20. Phosphorylation sites on the *Lmo*RS stressosome model

Different views of the pentamers and hexamers of the C1 map (**A**) and the D2 map (**B**) with RsbR1 in red and RsbS in blue. The position of T175 and T209 in RsbR1 is shown in green and yellow, respectively; S56 in RsbS is shown in cyan.

6.5. Conclusions

To conclude this chapter, the single particle analysis did not support the molecular reconstruction of the *Lmo*RSX stressosome particle. Therefore, the mode of binding of RsbX to the stressosome remains unclear. By contrast, the *Lmo*RS dataset allowed the successful reconstruction of two new cryo-EM maps: one in C1 and one in D2 symmetry. These reconstructions indicate that the *Lmo* RsbR:RsbS stressosome is stable without RsbT, which is in contrast to the conclusions drawn previously by Williams *et al* (Williams et al., 2019). The new maps also show clear differences to other work published previously (Kwon et al., 2019; Marles-Wright et al., 2008; Williams et al., 2019).

First, the RsbR:RsbS stoichiometry is distinct to that observed by Williams, Kwon and Marles-Wright. In the C1 reconstruction, 34 RsbR and 26 RsbS proteins were observed while in the D2 reconstruction, 20 RsbR and 40 RsbS proteins were overserved. These values contrast with those of Williams *et al*, who reported 40 RsbR1 and 20 RsbS for the C1 reconstruction. The equivalent numbers of Kwon *et al* are 44 & 20 RsbRA and 16 & 20 RsbS, in the C1 and D2 reconstruction, respectively, whereas Marles-Wright reported 40 RsbRA and 20 RsbS in D2 but did not report these details in a C1 reconstruction. Consequently, the true stoichiometry of the complex is still a matter of debate *in vitro* and *in vivo*, and the influence of experimental conditions presumably has a significant effect on the assembly of stressosomes. Marles-Wright used a bicistronic operon expression system, which replicates the genetic organisation of the σ^B operon, whereas Kwon and Williams mixed separately purified RsbR and RsbS at a 2:1 ratio *in vitro*. In this thesis, the *Lmo*RS complex was also reconstituted by mixing the purified RsbR1 and RsbS proteins in isolation, but with an excess of RsbS.

Second, the C1 model also shows differences in terms of the distribution of RsbR1 and RsbS within the complex. The C1 model would preclude RsbT binding on some faces of the complex due to an over-population of RsbR turrets, which would also block access to the phosphorylation sites in RsbR1. In the case of the C1 reconstruction of the *B. subtilis* stressosome (Kwon et al., 2019), the stoichiometry also varies. Indeed, 44 RsbR and 16 RsbS have been modelled, compared to our C1 model, in which 34 RsbR and 26 RsbS were modelled. These results show how the C1 reconstruction of the *B. subtilis* stressosome model is over-populated with RsbR dimers compared to the newly determined *L. monocytogenes* stressosome complex (*Lmo*RS). The turrets distribution remains more homogeneous in the *B. subtilis* model compared to *Lmo*RS model. In the case of the previously determined *L. monocytogenes* complex in C1 symmetry (Williams et al., 2019), 40 RsbR and 20 RsbS proteins were modelled and using D2 symmetry didn't improve the turrets densities. Therefore, all the phosphorylation sites are more accessible for RsbT in the D2 model in comparison to the C1 model, lending

credence to the conclusion that the D2 model has more biological significance than the C1 model.

Finally, the validation and further analysis of these reconstructions is limited by the poor density of the turrets. Further experiments could lock these turrets in position to reduce their flexibility, either through disulphide crosslinking or by using antibody fragments directed to the N-terminal domains of RsbR1.

Conclusion and future perspectives

To conclude, the different results obtained in this thesis has provided new structural insight into stress sensing by *Listeria monocytogenes*. In **Chapter III**, three new structures of the N-terminal domains of RsbR - N-RsbR1, 2 and 3 - were determined by X-ray crystallography. Although N-RsbR2 had a different conformation to already known N-RsbR structures, the newly determined N-RsbR1 and 3 structures appeared as variations on the dimeric, non-heme globin fold. N-RsbR1 is most similar to N-RsbRA from *B. subtilis*, whereas structural differences were observed between N-RsbR3 and other known N-RsbR structures (N-RsbRA, N-MtR and N-RsbR1): these differences include the presence of an extra α -helix and a wider conformation of the J-helix in N-RsbR3. Moreover, small molecule ligands, including acrylic acids from the crystallisation conditions, were also found in the N-RsbR3 crystal structure. Three ligand binding sites were observed: one in a pocket formed by the main helices of each monomer, and one at the dimerization interface. As the N-RsbR domains adopt a globin-like fold, it was not surprising to observe ligand binding sites within the monomers, especially as the intramolecular site overlaps with where the heme would be found in canonical globins, but a possible ligand binding site at the dimerization interface was not expected. Therefore, a number of ligands may be able to interact with each RsbR protein to activate the stressosome. Taken together, these ligand binding sites, and conformational differences seen when comparing the structures of the N-RsbR domains, imply that stress signals are transduced from the N-terminal globin-like domain to the C-terminal STAS domain by conformational changes after ligand binding. These structural observations led to the further study and understanding of the different types of ligands that can bind to N-RsbR1, presumably at the same site, the results of which are summarised in **Chapter IV**.

At the inception of this project, only a single potential interacting partner of the RsbR proteins had been identified, the miniprotein Prli42, and its interaction with stressosomes was modelled previously (Impens et al., 2017; Williams et al., 2019). To properly assess the basis for this interaction, biophysical analyses of the binding between Prli42 and all the N-RsbR paralogues from *B. subtilis* and *L. monocytogenes* were performed *in vitro* in **Chapter IV**. However, no interaction was observed between Prli42 and any N-RsbR protein from either *L. monocytogenes* or *B. subtilis*. If there is an interaction, the stressosome would be most likely to be engulfed in the membrane which will create cooperative binding. Therefore the interaction at the single protein level might not reflect the interaction *in vivo*. But the putative interaction was studied in depth amongst partners within the PATHSENSE consortium. Appropriate mutant *L. monocytogenes* and *B. subtilis* strains were subjected to a wide range of stresses

including salt, pH and oxidative stresses. These experiments failed to show any influence of Prli42 on σ^B activation *in vivo*. By contrast, Impens *et al* formed the opposite conclusion after stressing *L. monocytogenes* only by oxidative-stress (Impens et al., 2017), which is not a widely-studied stress inducer. Fluorescence microscopy experiments and cellular fractionation also showed that the stressosome complex is mostly found in the cytosol and is not demonstrably membrane associated (Dessaux et al., 2020; Marles-Wright et al., 2008), which would be the case if it interacted with the membrane anchored Prli42 miniprotein. Therefore, the miniprotein Prli42 neither interacts with nor has any effect on the stress sensing activity of the stressosome.

As Prli42 did not interact with the stressosome, new strategies to understand the ligand binding and activation of N-RsbR were used. First, a thermostability shift assay was used to screen a wide range of compounds to potentially identify a ligand that had common effects on the N-RsbR proteins (**Chapter IV**). However, there were no clearly identified ligands or ligand families for any of the N-RsbR proteins tested. While this screen did not identify any clear candidate ligands further work should investigate broader panels of small molecule targets and fragment compounds. To further explore the ligand binding of the N-RsbRs, a new library of halogenated compounds was used for ligand screening (Wood et al., 2019). Using crystal soaking methods, N-RsbR3 crystals were grown in a condition without acrylic acid, to avoid competitive binding of this molecule in the potential ligand binding site. Three compounds from the library showed potential binding to N-RsbR3, but while the anomalous signals from the halides were clear, the binding poses for the fragments were not. Due to time constraints and Covid-19 impact this avenue of investigation could not be brought to conclusion.

The native composition of the stressosome *in vivo* was reported previously (Delumeau et al., 2006) and was confirmed in this study in **Chapter V**. All four RsbR paralogues were found to co-purify when RsbRA was the target of independent affinity purification procedures. The original ambition of this part of this thesis was to understand the stoichiometry of the paralogues in the stressosome complex, or complexes. Unfortunately, the Covid-19 pandemic introduced significant restrictions in accessing the laboratory and as a consequence this part of the project could not be explored fully. Nevertheless, the native *B. subtilis* stressosome complex was prepared using different affinity purification strategies, to lay the groundwork for future studies. This work would also address the question of whether there is a single composition in all stressosomes, or whether stressosomes are dynamic and are comprised of varying levels of each RsbR paralogue.

In this thesis (**Chapter V, part A**) each *L. monocytogenes* RsbR paralogue was demonstrated to have the ability to form stable stressosome complexes, as shown previously

for *B. subtilis* RsbR paralogues (Delumeau et al., 2006). The *L. monocytogenes* RsbR-RsbS stressosome complex forms a stable complex in the absence of RsbT, an observation that is in marked contrast to that published previously (Williams et al., 2019). In **Chapter V, part B**, it was shown that RsbT and RsbX can bind to purified recombinant RsbR1:RsbS stressosome complexes. To further study the complex with and without RsbX (*LmoRS* and *LmoRSX*), single particle analysis of both complexes was initiated. Unfortunately, the cryo-EM of *LmoRSX* did not permit reconstruction. By contrast, the *LmoRS* cryo-EM dataset supported two new reconstructions, one in C1 with no symmetry applied and one in D2 symmetry. The stoichiometry of RsbR and RsbS in the new reconstructions was distinct to previously published models (Kwon et al., 2019; Marles-Wright et al., 2008; Williams et al., 2019). Indeed, in the C1 reconstruction, 34 RsbR and 26 RsbS proteins were observed; while in the D2 reconstruction, 20 RsbR and 40 RsbS proteins were overserved. The observed stoichiometries are different from what have been observed previously which might be due to orientation bias. In the case of the *B. subtilis* model, 40 RsbR and 20 RsbS proteins were found in the D2 reconstruction (Kwon et al., 2019; Marles-Wright et al., 2008), while 44 RsbR and 16 RsbS were modelled in the C1 reconstruction (Kwon et al., 2019). In the case of the *L. monocytogenes* complex (Williams et al., 2019), only the C1 symmetry model was used as the D2 symmetry imposition didn't improve the density of the map, which allowed modelling 40 RsbR and 20 RsbS proteins. These heterogeneous results raise questions about the physiological stoichiometry of the complex, as they may differ due to experimental and environmental conditions.

The D2 reconstruction had a resolution of 3.68 Å in the core and 10.46 Å in the turrets, but at this resolution, it was not possible to build a full atomic model of the RsbR proteins. Therefore, the turret densities were fit with the crystallographic model of N-RsbR1 (solved in **Chapter III**) using the STAS-domains and J-helices as a guide. The overall reconstruction showed different assemblies in C1 and D2 symmetries and the true assembly of the stressosome *in vivo* remains open to further investigation.

In conclusion, new insights of signal perception and transduction in the stressosome have been obtained. First of all, the crystal structures of the NTDs of the RsbR proteins have revealed the presence and location of several putative ligand binding pockets. The corresponding ligand binding activity might play a role in signal perception even though, at this stage, the ligand in question remains unknown. Upon ligand binding, which can also be described as signal perception, the signal is transduced from the NTD to the CTD of the RsbR protein by structural movements. As has also been shown by the crystallography of the N-RsbR

proteins in this project, the J-helices that link the NTD to the CTD have markedly different conformations. In the case of the crystal structures of 2BNL and N-RsbR1, the J-helices are more or less parallel, while they are flared apart and hence are wider in N-RsbR3, the structure in which ligands from the crystallisation solution were found to have been bound. Finally, the J-helices are crossed in the 3ZTA crystal structure and in the cryo-EM reconstruction of the *LmoRS* complex. Structural flexibility of the J-helix that connects the NTD of the RsbR protein could explain these different observations. The crossing of the J-helices can thus be likened to the movements observed in a pair of scissors. Nevertheless, further investigations are required to understand the ligand binding activity, the chemical identity of the molecule(s) recognised by the RsbR paralogues and the magnitude and dynamics of the structural movements associated with signal transduction.

Bibliography

- Abil, Z., Denard, C.A., and Zhao, H. (2014). Modular assembly of designer PUF proteins for specific post-transcriptional regulation of endogenous RNA. *Journal of Biology Engineering*, 8:7.
- Afonine, P.V., Klaholz, B.P., Moriarty, N.W., Poon, B.K., Sobolev, O.V., Terwilliger, T.C., Adams, P.D., and Urzhumtsev, A. (2018). New tools for the analysis and validation of cryo-EM maps and atomic models. *Acta Cryst D Structural Biology* 74, 814–840.
- Akbar, S., Kang, C.M., Gaidenko, T.A., and Price, C.W. (1997). Modulator protein RsbR regulates environmental signalling in the general stress pathway of *Bacillus subtilis*. *Molecular Microbiology* 24:3, 567–578.
- Akbar, S., Gaidenko, T.A., Kang, C.M., O'Reilly, M., Devine, K.M., and Price, C.W. (2001). New family of regulators in the environmental signalling pathway which activates the general stress transcription factor σ B of *Bacillus subtilis*. *Journal of Bacteriology*, 183:4, 1329–1338.
- Alper, S., Dufour, A., Garsin, D.A., Duncan, L., and Losick, R. (1996). Role of adenosine nucleotides in the regulation of a stress-response transcription factor in *Bacillus subtilis*. *Journal of Molecular Biology*, 260:2, 165–177.
- Aravind, L., and Koonin, E.V. (2000). The STAS domain - a link between anion transporters and antisigma-factor antagonists. *Current Biology*, 10, R53-55.
- Beilsten-Edmands, J., Winter, G., Gildea, R., Parkhurst, J., Waterman, D., and Evans, G. (2020). Scaling diffraction data in the DIALS software package: algorithms and new approaches for multi-crystal scaling. *Acta Cryst D Structural Biology* 76, 385–399.
- Benson, A.K., and Haldenwang, W.G. (1993). *Bacillus subtilis* sigma B is regulated by a binding protein (RsbW) that blocks its association with core RNA polymerase. *Proc. Natl. Acad. Sci. U.S.A.* 90, 2330–2334.
- Borriss, R., Danchin, A., Harwood, C.R., Médigue, C., Rocha, E.P.C., Sekowska, A., and Vallenet, D. (2017). *Bacillus subtilis*, the model Gram-positive bacterium: 20 years of annotation refinement. *Microb Biotechnol* 11, 3–17.
- Boylan, S.A., Thomas, M.D., and Price, C.W. (1991). Genetic method to identify regulons controlled by nonessential elements: isolation of a gene dependent on alternate transcription factor sigma B of *Bacillus subtilis*. *Journal of Bacteriology*, 173, 7856–7866.
- Boylan, S.A., Redfield, A.R., and Price, C.W. (1993a). Transcription factor sigma B of *Bacillus subtilis* controls a large stationary-phase regulon. *Journal of Bacteriology* 175, 3957–3963.
- Boylan, S.A., Redfield, A.R., Brody, M.S., and Price, C.W. (1993b). Stress-induced activation of the sigma B transcription factor of *Bacillus subtilis*. *Journal of Bacteriology* 175, 7931–7937.
- Breed, R., Murray, E.G.D., and Smith, N. (1958). Bergey's Manual of Determinative Bacteriology. *Journal of the American Pharmaceutical Association (Scientific Ed.)* 47, 156.

- Brody, M.S., Vijay, K., and Price, C.W. (2001). Catalytic function of an α/β hydrolase is required for energy stress activation of the σ B transcription factor in *Bacillus subtilis*. *Journal of Bacteriology* 183, 6422–6428.
- Burgess, R.R., Travers, A.A., Dunn, J.J., and Bautz, E.K. (1969). Factor stimulating transcription by RNA polymerase. *Nature* 221, 43–46.
- Cairns, L.S., Marlow, V.L., Bissett, E., Ostrowski, A., and Stanley-Wall, N.R. (2013). A mechanical signal transmitted by the flagellum controls signalling in *Bacillus subtilis*. *Molecular Microbiology*, 90, 6–21.
- Chen, C., Lewis, R.J., Harris, R., Yudkin, M.D., and Delumeau, O. (2003). A supramolecular complex in the environmental stress signalling pathway of *Bacillus Subtilis*. *Molecular Microbiology*, 49:9, 1657-1669.
- Chen, C.-C., Yudkin, M.D., and Delumeau, O. (2004). Phosphorylation and RsbX-dependent dephosphorylation of RsbR in the RsbR-RsbS complex of *Bacillus subtilis*. *Journal of Bacteriology* 186, 6830–6836.
- Chen, V.B., Arendall, W.B., Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Cryst D Biological Crystallography* 66, 12–21.
- Cleverley, R.M., Rutter, Z.J., Rismondo, J., Corona, F., Tsui, H.-C.T., Alatawi, F.A., Daniel, R.A., Halbedel, S., Massidda, O., Winkler, M.E., et al. (2019). The cell cycle regulator GpsB functions as cytosolic adaptor for multiple cell wall enzymes. *Nat Communications* 10.
- Cohn, F. (1875). *Untersuchungen über Bakterien: I* (J. U. Kern).
- Cooper, G.M. (2000). *Eukaryotic RNA polymerases and general transcription factors. The Cell: A Molecular Approach*. 2nd Edition.
- Cowtan, K. (2006). The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Cryst D Biological Crystallography*, 62, 1002–1011.
- Creixell, P., Schoof, E.M., Tan, C.S.H., and Linding, R. (2012). Mutational properties of amino acid residues: implications for evolvability of phosphorylatable residues. *Philosophical Transactions of the Royal Society B*, 367, 2584–2593.
- Danchin, A. (2001). *Bacillus Subtilis*. In *Encyclopedia of Genetics*, S. Brenner, and J.H. Miller, eds. (New York: Academic Press), pp. 135–144.
- Davis, I.W., Leaver-Fay, A., Chen, V.B., Block, J.N., Kapral, G.J., Wang, X., Murray, L.W., Arendall, W.B., Snoeyink, J., Richardson, J.S., et al. (2007). MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Research*, 35, W375–W383.
- Delumeau, O., Lewis, R.J., and Yudkin, M.D. (2002). Protein-protein interactions that regulate the energy stress activation of sigma(B) in *Bacillus subtilis*. *Journal of Bacteriology*, 184, 5583–5589.

- Delumeau, O., Dutta, S., Brigulla, M., Kuhnke, G., Hardwick, S.W., Völker, U., Yudkin, M.D., and Lewis, R.J. (2004). Functional and structural characterization of RsbU, a stress signalling protein phosphatase 2C*. *Journal of Biological Chemistry*, 279, 40927–40937.
- Delumeau, O., Chen, C.-C., Murray, J.W., Yudkin, M.D., and Lewis, R.J. (2006). High-molecular-weight complexes of RsbR and paralogues in the environmental signalling pathway of *Bacillus subtilis*. *Journal of Bacteriology*, 188, 7885–7892.
- Deng, X., Phillippy, A.M., Li, Z., Salzberg, S.L., and Zhang, W. (2010). Probing the pan-genome of *Listeria monocytogenes*: new insights into intraspecific niche expansion and genomic diversification. *BMC Genomics* 11, 500.
- Dessaux, C., Guerreiro, D.N., Pucciarelli, M.G., O’Byrne, C.P., and García-del Portillo, F. (2020). Impact of osmotic stress on the phosphorylation and subcellular location of *Listeria monocytogenes* stressosome proteins. *Scientific Reports* 10.
- Dubnau, D. (1991). Genetic competence in *Bacillus subtilis*. *Microbiology Reviews*, 55, 395–424.
- Dufour, A., and Haldenwang, W.G. (1994). Interactions between a *Bacillus subtilis* anti-sigma factor (RsbW) and its antagonist (RsbV). *Journal of Bacteriology*, 176, 1813–1820.
- Dufour, A., Voelker, U., Voelker, A., and Haldenwang, W.G. (1996). Relative levels and fractionation properties of *Bacillus subtilis* sigma(B) and its regulators during balanced growth and stress. *Journal of Bacteriology* 178, 3701-9.
- Dutta, R., and Inouye, M. (2000). GHKL, an Emergent ATPase/kinase Superfamily. *Trends in Biochemical Sciences*, 25, 24-28.
- Earl, A.M., Losick, R., and Kolter, R. (2008). Ecology and genomics of *Bacillus subtilis*. *Trends in Microbiology*, 16, 269.
- EFSA (2015). *Campylobacter* and *Listeria* infections still rising in the EU - say EFSA and ECDC.
- Ehrenberg, C.G. "Dritter Beitrag zur Erkenntniss grosser Organisation in der Richtung des kleinsten Raumes." *Physikalische Abhandlungen der Koeniglichen Akademie der Wissenschaften zu Berlin aus den Jahren 1833-1835* (1835). pp.143-336.
- Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Cryst D Biological Crystallography*, 60, 2126–2132.
- Evans, P.R. (2011). An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Cryst D Biological Crystallography*. 67, 282–292.
- Evans, P.R., and Murshudov, G.N. (2013). How good are my data and what is the resolution? *Acta Cryst D Biological Crystallography*, 69, 1204–1214.
- Farber, J.M., and Peterkin, P.I. (1991). *Listeria monocytogenes*, a food-borne pathogen. *Microbiology Reviews*, 55, 476–511.
- Fernandez-Leiro, R., and Scheres, S.H.W. (2017). A pipeline approach to single-particle processing in RELION. *Acta Cryst D Structural Biology*, 73, 496–502.

- Förster, S., Apostol, L., and Bras, W. (2010). Scatter: software for the analysis of nano- and mesoscale small-angle scattering. *Journal of Applied Crystallography* 43, 639–646.
- Gaidenko, T.A., Yang, X., Lee, Y., and Price, C.W. (1999). Threonine phosphorylation of modulator protein RsbR governs its ability to regulate a serine kinase in the environmental stress signalling pathway of *Bacillus Subtilis*. *Journal of Molecular Biology*, 288, 29-39.
- Gaidenko, T.A., Kim, T.-J., Weigel, A.L., Brody, M.S., and Price, C.W. (2006). The blue-light receptor YtvA acts in the environmental stress signalling pathway of *Bacillus subtilis*. *Journal of Bacteriology*. 188, 6387–6395.
- Gaidenko, T.A., Bie, X., Baldwin, E.P., and Price, C.W. (2011). Substitutions in the presumed sensing domain of the *Bacillus subtilis* stressosome affect its basal output but not response to environmental signals. *Journal of Bacteriology* 193, 3588–3597.
- Greenfield, N.J. (2006). Using circular dichroism spectra to estimate protein secondary structure. *Nature Protocols* 1, 2876–2890.
- Grøftehaug, M.K., Hajizadeh, N.R., Swann, M.J., and Pohl, E. (2015). Protein–ligand interactions investigated by thermal shift assays (TSA) and dual polarization interferometry (DPI). *Acta Crystallogr D Biological Crystallography*, 71, 36–44.
- Guillet, C., Join-Lambert, O., Le Monnier, A., Leclercq, A., Mechaï, F., Mamzer-Bruneel, M.-F., Bielecka, M.K., Scortti, M., Disson, O., Berche, P., et al. (2010). Human Listeriosis caused by *Listeria ivanovii*. *Emerg Infect Dis* 16, 136–138.
- Haldenwang, W.G. (1995). The sigma factors of *Bacillus subtilis*. *Microbiol Rev* 59, 1–30.
- Haldenwang, W.G., and Losick, R. (1980). Novel RNA polymerase sigma factor from *Bacillus subtilis*. *Proc. Natl. Acad. Sci. U.S.A.* 77, 7000–7004.
- Hardwick, S.W., Pané-Farré, J., Delumeau, O., Marles-Wright, J., Murray, J.W., Hecker, M., and Lewis, R.J. (2007). Structural and functional characterization of partner switching regulating the environmental stress response in *Bacillus subtilis*. *Journal of Biological Chemistry*, 282, 11562–11572.
- Harwood, C.R. (1989). Introduction to the Biotechnology of *Bacillus*. In *Bacillus*, C.R. Harwood, ed. (Boston, MA: Springer US), pp. 1–4.
- Hecker, M., and Völker, U. (2001). General stress response of *Bacillus subtilis* and other bacteria. *Advance in Microbial Physiology, Science Direct*, 44, 35–91.
- Hendrickson, W.A., and Teeter, M.M. (1981). Structure of the hydrophobic protein crambin determined directly from the anomalous scattering of sulphur. *Nature* 290, 107–113.
- Impens, F., Rolhion, N., Radoshevich, L., Bécavin, C., Duval, M., Mellin, J., García Del Portillo, F., Pucciarelli, M.G., Williams, A.H., and Cossart, P. (2017). N-terminomics identifies Prli42 as a membrane miniprotein conserved in Firmicutes and critical for stressosome activation in *Listeria monocytogenes*. *Nature Microbiology*, 2, 17005.
- Iverson, S., Haddock, T., Beal, J., and Densmore, D. (2016). CIDAR MoClo: Improved MoClo assembly standard and new *E. coli* part library enable rapid combinatorial design for synthetic and traditional biology. *ACS Synth Biol*, 5, 99-103.

- Jurk, M., Dorn, M., Reichenwallner, J., Bardiaux, B., Hinderberger, D., and Schmieder, P. (2014). Solution Structure of YtvA from *Bacillus subtilis* Provides Insight into Activation Mechanism and Regulation of Bacterial Stress Response. Not Published Yet.
- Kabsch, W. (2010a). XDS. *Acta Cryst D Biological Crystallography*, *66*, 125–132.
- Kabsch, W. (2010b). Integration, scaling, space-group assignment and post-refinement. *Acta Cryst D Biological Crystallography*, *66*, 133–144.
- Kang, C.M., Brody, M.S., Akbar, S., Yang, X., and Price, C.W. (1996). Homologous pairs of regulatory proteins control activity of *Bacillus subtilis* transcription factor sigma(b) in response to environmental stress. *Journal of Bacteriology*, *178*, 3846–3853.
- Kang, C.M., Vijay, K., and Price, C.W. (1998). Serine kinase activity of a *Bacillus subtilis* switch protein is required to transduce environmental stress signals but not to activate its target PP2C phosphatase. *Molecular Microbiology*, *30*, 189–196.
- Karplus, P.A., and Diederichs, K. (2015). Assessing and maximizing data quality in macromolecular crystallography. *Curr Opin Struct Biol* *34*, 60–68.
- Kastritis, P.L., Moal, I.H., Hwang, H., Weng, Z., Bates, P.A., Bonvin, A.M.J.J., and Janin, J. (2011). A structure-based benchmark for protein-protein binding affinity. *Protein Sci* *20*, 482–491.
- Kazmierczak, M.J., Mithoe, S.C., Boor, K.J., and Wiedmann, M. (2003). *Listeria monocytogenes* sigma B regulates stress response and virulence functions. *Journal of Bacteriology*, *185*, 5722–5734.
- Kazmierczak, M.J., Wiedmann, M., and Boor, K.J. (2005). Alternative sigma factors and their roles in bacterial virulence. *Microbiology and Molecular Biology Reviews*, *69*, 527–543.
- Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., and Sternberg, M.J.E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols* *10*, 845–858.
- Kikhney, A., and Svergun, D. (2015). A Practical Guide to Small Angle X-ray Scattering (SAXS) of Flexible and Intrinsically Disordered Proteins. *FEBS Letters*, *589*, 2570–2577.
- Kim, T.-J., Gaidenko, T.A., and Price, C.W. (2004). A Multicomponent Protein Complex Mediates Environmental Stress Signaling in *Bacillus Subtilis*. *Journal of Molecular Biology*, *341*, 135–150.
- Konarev, P.V., Volkov, V.V., Sokolova, A.V., Koch, M.H.J., and Svergun, D.I. (2003). PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J Appl Cryst* *36*, 1277–1282.
- Kroeger, T., Frieg, B., Zhang, T., Hansen, F.K., Marmann, A., Proksch, P., Nagel-Steger, L., Groth, G., Smits, S.H.J., and Gohlke, H. (2017). EDTA aggregates induce SYPRO orange-based fluorescence in thermal shift assay. *PLoS One* *12*.
- Krojer, T., Talon, R., Pearce, N., Collins, P., Douangamath, A., Brandao-Neto, J., Dias, A., Marsden, B., and von Delft, F. (2017). The XChemExplorer graphical workflow tool

- for routine or large-scale protein–ligand structure determination. *Acta Cryst D* 73, 267–278.
- Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G., Azevedo, V., Bertero, M.G., Bessières, P., Bolotin, A., Borchert, S., et al. (1997). The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. *Nature* 390, 249–256.
- Kwon, E., Pathak, D., Kim, H.-U., Dahal, P., Ha, S.C., Lee, S.S., Jeong, H., Jeoung, D., Chang, H.W., Jung, H.S., et al. (2019). Structural insights into stressosome assembly. *IUCrJ* 6, 938–947.
- Lafita, A., Tian, P., Best, R.B., and Bateman, A. (2019). Tandem domain swapping: determinants of multidomain protein misfolding. *Curr Opin Struct Biol* 58, 97–104.
- Lea, W.A., and Simeonov, A. (2011). Fluorescence polarization assays in small molecule screening. *Expert Opin Drug Discov* 6, 17–32.
- Liebschner, D., Afonine, P.V., Baker, M.L., Bunkóczi, G., Chen, V.B., Croll, T.I., Hintze, B., Hung, L.-W., Jain, S., McCoy, A.J., et al. (2019). Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Cryst D* 75, 861–877.
- Marcela, A.-P., Vreede, J., Tang, Y., Bende, O., Losi, A., Gärtner, W., and Hellingwerf, K. (2009). In Vivo Mutational Analysis of YtvA from *Bacillus subtilis* mechanism of light activation of the general stress response. *J. Biol. Chem.* 284, 24958–24964.
- Marles-Wright, J., Grant, T., Delumeau, O., Duinen, G. van, Firbank, S.J., Lewis, P.J., Murray, J.W., Newman, J.A., Quin, M.B., Race, P.R., et al. (2008). Molecular Architecture of the “Stressosome,” a Signal Integration and Transduction Hub. *Science* 322, 92–96.
- Matthews, B.W. (1968). Solvent content of protein crystals. *Journal of Molecular Biology* 33, 491–497.
- Mielich-Süss, B., and Lopez, D. (2015). Molecular mechanisms involved in *Bacillus subtilis* biofilm formation. *Environ. Microbiol.* 17, 555–565.
- Moszer, I., Glaser, P., and Danchin, A. (1995). SubtiList: a relational database for the *Bacillus subtilis* genome. *Microbiology (Reading, Engl.)* 141 (Pt 2), 261–268.
- Mueller-Dieckmann, C., Panjikar, S., Tucker, P.A., and Weiss, M.S. (2005). On the routine use of soft X-rays in macromolecular crystallography. Part III. The optimal data-collection wavelength. *Acta Crystallogr D Biol Crystallogr* 61, 1263–1272.
- Murray, E.G.D., Webb, R.A., and Swann, M.B.R. (1926). A disease of rabbits characterised by a large mononuclear leucocytosis, caused by a hitherto undescribed bacillus *Bacterium monocytogenes* (n.sp.). *The Journal of Pathology and Bacteriology* 29, 407–439.
- Murray, J.W., Delumeau, O., and Lewis, R.J. (2005). Structure of a nonheme globin in environmental stress signalling. *PNAS* 102, 17320–17325.
- Murshudov, G.N., Skubák, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., Winn, M.D., Long, F., and Vagin, A.A. (2011). REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr* 67, 355–367.

- Najafi, S.M., Willis, A.C., and Yudkin, M.D. (1995). Site of phosphorylation of SpoIIAA, the anti-anti-sigma factor for sporulation-specific sigma F of *Bacillus subtilis*. *Journal of Bacteriology* *177*, 2912–2913.
- Najafi, S.M., Harris, D.A., and Yudkin, M.D. (1996). The SpoIIAA protein of *Bacillus subtilis* has GTP-binding properties. *J. Bacteriol.* *178*, 6632–6634.
- de Noordhout, C.M., Devleeschauwer, B., Angulo, F.J., Verbeke, G., Haagsma, J., Kirk, M., Havelaar, A., and Speybroeck, N. (2014). The global burden of listeriosis: a systematic review and meta-analysis. *Lancet Infect Dis* *14*, 1073–1082.
- Orlova, E.V., and Saibil, H.R. (2011). Structural Analysis of Macromolecular Assemblies by Electron Microscopy. *Chem Rev* *111*, 7710–7748.
- Paget, M.S., and Helmann, J.D. (2003). The $\sigma 70$ family of sigma factors. *Genome Biol* *4*, 203.
- Pané-Farré, J., Lewis, R.J., and Stülke, J. (2005). The RsbRST stress module in bacteria: a signalling system that may interact with different output modules. *J Mol Microbiol Biotechnol* *9*, 65–76.
- Pané-Farré, J., Jonas, B., Hardwick, S.W., Gronau, K., Lewis, R.J., Hecker, M., and Engelmann, S. (2009). Role of RsbU in Controlling SigB Activity in *Staphylococcus aureus* following Alkaline Stress. *J Bacteriol* *191*, 2561–2573.
- Pané-Farré, J., Quin, M.B., Lewis, R.J., and Marles-Wright, J. (2017). Structure and Function of the Stressosome Signalling Hub. In *Macromolecular Protein Complexes: Structure and Function*, J.R. Harris, and J. Marles-Wright, eds. (Cham: Springer International Publishing), pp. 1–41.
- Petersohn, A., Brigulla, M., Haas, S., Hoheisel, J.D., Völker, U., and Hecker, M. (2001). Global Analysis of the General Stress Response of *Bacillus subtilis*. *J Bacteriol* *183*, 5617–5631.
- Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem* *25*, 1605–1612.
- Potterton, L., Agirre, J., Ballard, C., Cowtan, K., Dodson, E., Evans, P.R., Jenkins, H.T., Keegan, R., Krissinel, E., Stevenson, K., et al. (2018). CCP4i2: the new graphical user interface to the CCP4 program suite. *Acta Cryst D* *74*, 68–84.
- Price, C.W., Fawcett, P., Cérémonie, H., Su, N., Murphy, C.K., and Youngman, P. (2001). Genome-wide analysis of the general stress response in *Bacillus subtilis*. *Mol. Microbiol.* *41*, 757–774.
- Quin, M.B., Berrisford, J.M., Newman, J.A., Baslé, A., Lewis, R.J., and Marles-Wright, J. (2012). The Bacterial Stressosome: A Modular System that Has Been Adapted to Control Secondary Messenger Signaling. *Structure* *20*, 350–363.
- Rohou, A., and Grigorieff, N. (2015). CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J Struct Biol* *192*, 216–221.
- Ryan, P., Forrester, T.J.B., Wroblewski, C., Kenney, T.M.G., Kitova, E.N., Klassen, J.S., and Kimber, M.S. (2019). The small RbcS-like domains of the β -carboxysome structural

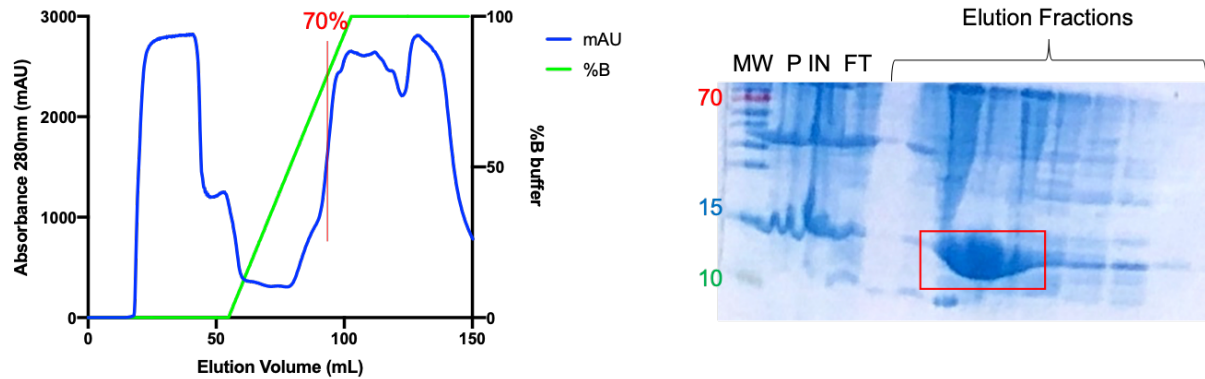
- protein CcmM bind RubisCO at a site distinct from that binding the RbcS subunit. *J Biol Chem* 294, 2593–2603.
- Schardt, J., Jones, G., Müller-Herbst, S., Schauer, K., D’Orazio, S.E.F., and Fuchs, T.M. (2017). Comparison between *Listeria sensu stricto* and *Listeria sensu lato* strains identifies novel determinants involved in infection. *Sci Rep* 7, 17821.
- Scheres, S.H.W. (2012). RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *J Struct Biol* 180, 519–530.
- Scheres, S.H.W. (2016). Processing of Structurally Heterogeneous Cryo-EM Data in RELION. *Methods Enzymol* 579, 125–157.
- Schmidt, R., Margolis, P., Duncan, L., Coppolecchia, R., Moran, C., and Losick, R. (1990). Control of Developmental Transcription Factor Sigma F by Sporulation Regulatory Proteins SpoIIAA and SpoIIAB in *Bacillus Subtilis*. *Proc. Natl. Acad. Sci. USA*, 87, 9221–9225.
- Schneider, T.R., and Sheldrick, G.M. (2002). Substructure solution with SHELXD. *Acta Crystallogr. D Biol. Crystallogr.* 58, 1772–1779.
- Seavers, P.R., Lewis, R.J., Brannigan, J.A., Verschueren, K.H., Murshudov, G.N., and Wilkinson, A.J. (2001). Structure of the *Bacillus* cell fate determinant SpoIIAA in phosphorylated and unphosphorylated forms. *Structure* 9, 605–614.
- Seeliger HPR, Höhne K 1979. *Listeria monocytogenes* and related species. In T Bergan & JR Norris (eds), *Methods in Microbiology*, Academic Press, London, p. 31–49.
- Sheldrick, G.M. (2002). Macromolecular phasing with SHELXE. *Zeitschrift Fur Kristallographie* 217, 644–650.
- Sheldrick, G.M. (2008). A short history of SHELX. *Acta Cryst A* 64, 112–122.
- Sheldrick, G.M. (2010). Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr D Biol Crystallogr* 66, 479–485.
- Shi, Y. (2009). Serine/Threonine Phosphatases: Mechanism through Structure. *Cell* 139, 468–484.
- Simpkin, A.J., Thomas, J.M.H., Simkovic, F., Keegan, R.M., and Rigden, D.J. (2019). Molecular replacement using structure predictions from databases. *Acta Crystallogr D Struct Biol* 75, 1051–1062.
- Smirnova, N., Scott, J., Voelker, U., and Haldenwang, W.G. (1998). Isolation and Characterization of *Bacillus subtilis* sigB Operon Mutations That Suppress the Loss of the Negative Regulator RsbX. *Journal of Bacteriology* 180, 3671–3680.
- Strauch, M.A., and Hoch, J.A. (1993). Signal transduction in *Bacillus subtilis* sporulation. *Curr. Opin. Genet. Dev.* 3, 203–212.
- Tang, G., Peng, L., Baldwin, P.R., Mann, D.S., Jiang, W., Rees, I., and Ludtke, S.J. (2007). EMAN2: an extensible image processing suite for electron microscopy. *J Struct Biol* 157, 38–46.

- Teh, A.H., Makino, M., Hoshino, T., Baba, S., Shimizu, N., Yamamoto, M., and Kumasaka, T. (2015). Structure of the RsbX phosphatase involved in the general stress response of *Bacillus subtilis*. *Acta Crystallogr. D Biol. Crystallogr.* *71*, 1392–1399.
- Tejero, J., and Gladwin, M.T. (2014). The globin superfamily: functions in nitric oxide formation and decay. *Biol Chem* *395*, 631–639.
- Terwilliger, T.C., Bunkóczi, G., Hung, L.-W., Zwart, P.H., Smith, J.L., Akey, D.L., and Adams, P.D. (2016). Can I solve my structure by SAD phasing? Anomalous signal in SAD phasing. *Acta Crystallogr D Struct Biol* *72*, 346–358.
- Vagin, A., and Teplyakov, A. (2010). Molecular replacement with MOLREP. *Acta Crystallogr D Biol Crystallogr* *66*, 22–25.
- Vijay, K., Brody, M.S., Fredlund, E., and Price, C.W. (2000). A PP2C phosphatase containing a PAS domain is required to convey signals of energy stress to the σ B transcription factor of *Bacillus subtilis*. *Molecular Microbiology* *35*, 180–188.
- Voelker, U., Dufour, A., and Haldenwang, W.G. (1995). The *Bacillus subtilis* *rsbU* gene product is necessary for RsbX-dependent regulation of sigma B. *J. Bacteriol.* *177*, 114–122.
- Wagner, A., Duman, R., Henderson, K., and Mykhaylyk, V. (2016). In-vacuum long-wavelength macromolecular crystallography. *Acta Crystallogr D Struct Biol* *72*, 430–439.
- Wagner, A., Duman, R., Omari, K. el, Grama, V., and Mykhaylykh, V. (2018). The Long-wavelength Macromolecular Crystallography I23 at Diamond Light Source. *Nihon Kessho Gakkaishi* *60*, 233–239.
- Wilkins, M.R., Gasteiger, E., Bairoch, A., Sanchez, J.-C., Williams, K.L., Appel, R.D., and Hochstrasser, D.F. (1999). Protein Identification and Analysis Tools in the ExPASy Server. In *2-D Proteome Analysis Protocols*, A.J. Link, ed. (Totowa, NJ: Humana Press), pp. 531–552.
- Williams, A.H., Redzej, A., Rolhion, N., Costa, T.R.D., Rifflet, A., Waksman, G., and Cossart, P. (2019). The cryo-electron microscopy supramolecular structure of the bacterial stressosome unveils its mechanism of activation. *Nat Commun* *10*, 3005.
- Williams, C.J., Headd, J.J., Moriarty, N.W., Prisant, M.G., Videau, L.L., Deis, L.N., Verma, V., Keedy, D.A., Hintze, B.J., Chen, V.B., et al. (2018). MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci* *27*, 293–315.
- Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G.W., McCoy, A., et al. (2011). Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* *67*, 235–242.
- Winter, G. (2010). xia2: an expert system for macromolecular crystallography data reduction. *J Appl Cryst* *43*, 186–190.
- Wise, A.A., and Price, C.W. (1995). Four additional genes in the sigB operon of *Bacillus subtilis* that control activity of the general stress factor sigma B in response to environmental signals. *J Bacteriol* *177*, 123–133.

- Wood, D.J., Lopez-Fernandez, J.D., Knight, L.E., Al-Khawaldeh, I., Gai, C., Lin, S., Martin, M.P., Miller, D.C., Cano, C., Endicott, J.A., et al. (2019). FragLites-Minimal, Halogenated Fragments Displaying Pharmacophore Doublets. An Efficient Approach to Druggability Assessment and Hit Generation. *J Med Chem* 62, 3741–3752.
- Woodbury, R.L., Luo, T., Grant, L., and Haldenwang, W.G. (2004). Mutational Analysis of RsbT, an Activator of the *Bacillus subtilis* Stress Response Transcription Factor, σ B. *J Bacteriol* 186, 2789–2797.
- Yang, X., Kang, C.M., Brody, M.S., and Price, C.W. (1996). Opposing pairs of serine protein kinases and phosphatases transmit signals of environmental stress to activate a bacterial transcription factor. *Genes Dev.* 10, 2265–2275.
- Zhang, Y., and Ertbjerg, P. (2019). On the origin of thaw loss: Relationship between freezing rate and protein denaturation. *Food Chem* 299, 125104.
- Zheng, S.Q., Palovcak, E., Armache, J.-P., Verba, K.A., Cheng, Y., and Agard, D.A. (2017). MotionCor2 - anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat Methods* 14, 331–332.
- Zhu, B., and Stülke, J. (2018). SubtiWiki in 2018: from genes and proteins to functional network annotation of the model organism *Bacillus subtilis*. *Nucleic Acids Res* 46, D743–D748.
- Zivanov, J., Nakane, T., Forsberg, B.O., Kimanius, D., Hagen, W.J., Lindahl, E., and Scheres, S.H. (2018). New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* 7, e42166.
- Zivanov, J., Nakane, T., and Scheres, S.H.W. (2019a). Estimation of High-Order Aberrations and Anisotropic Magnification from Cryo-EM Datasets in RELION-3.1. *BioRxiv* 798066.
- Zivanov, J., Nakane, T., and Scheres, S.H.W. (2019b). A Bayesian approach to beam-induced motion correction in cryo-EM single-particle analysis. *IUCrJ* 6, 5–17.

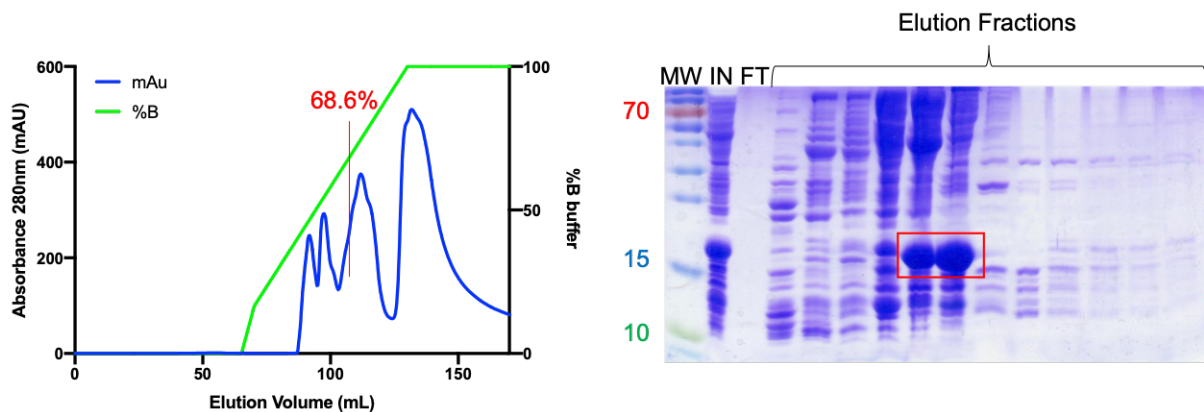
Appendices

Supplemental figures



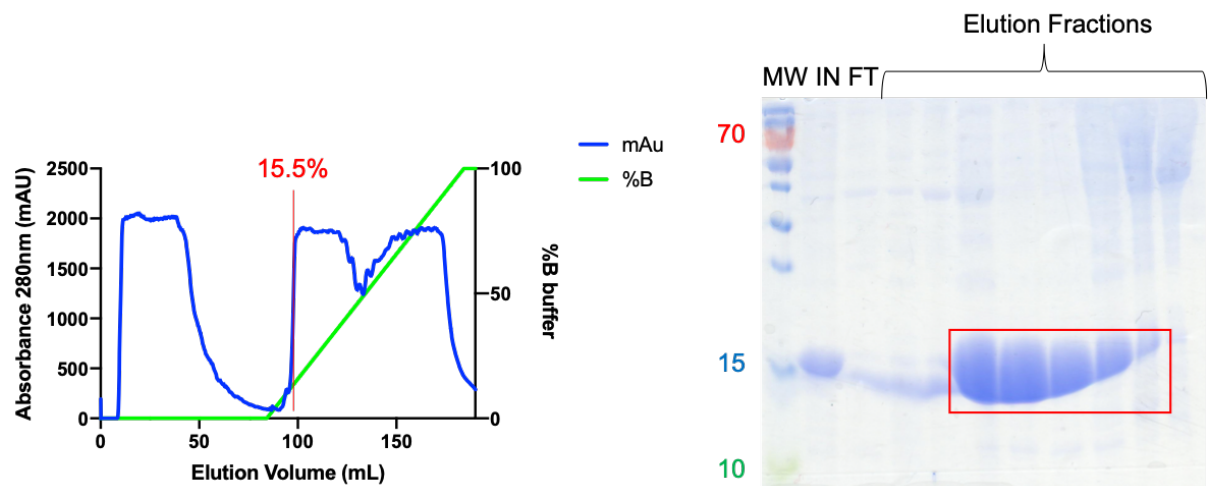
S1. Anion exchange affinity step of N-RsbR1 protein purification

Q Sepharose anion exchange chromatogram (left) for N-RsbR1 with absorbance at 280nm in blue and the elution buffer percentage in green. N-RsbR1 eluted at approximately 70% of buffer B. 15% SDS-PAGE (right) with the molecular mass marker (MW), the pellet (P), the input (IN) of the column, the flow-through (FT) from the column and the fractions from the elution buffer gradient. The fractions containing the protein of interest are boxed in red.



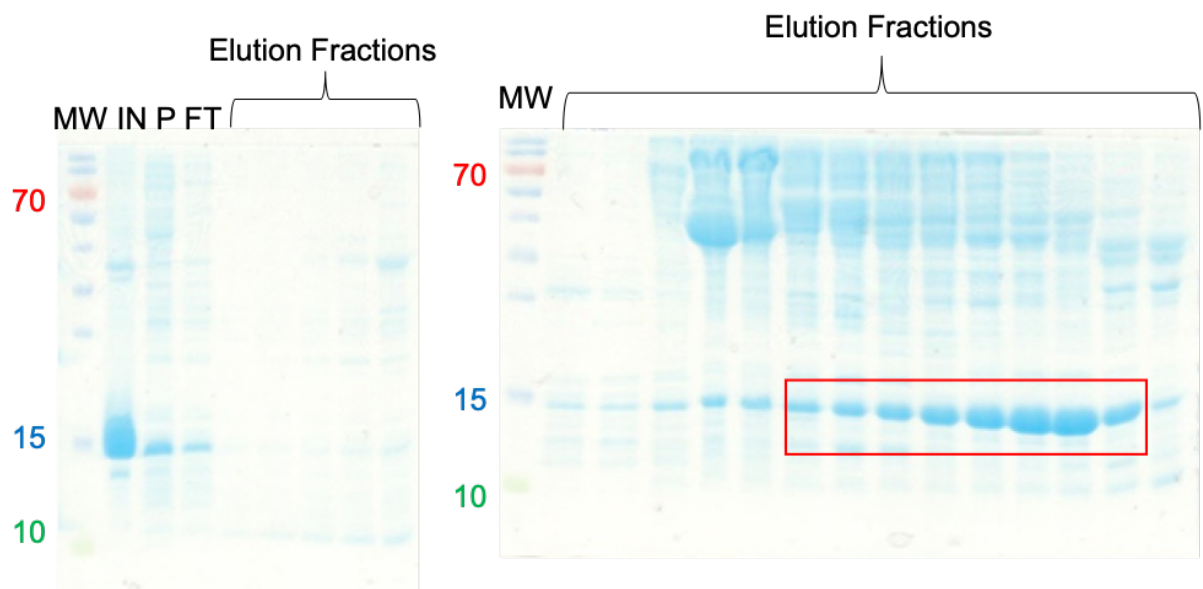
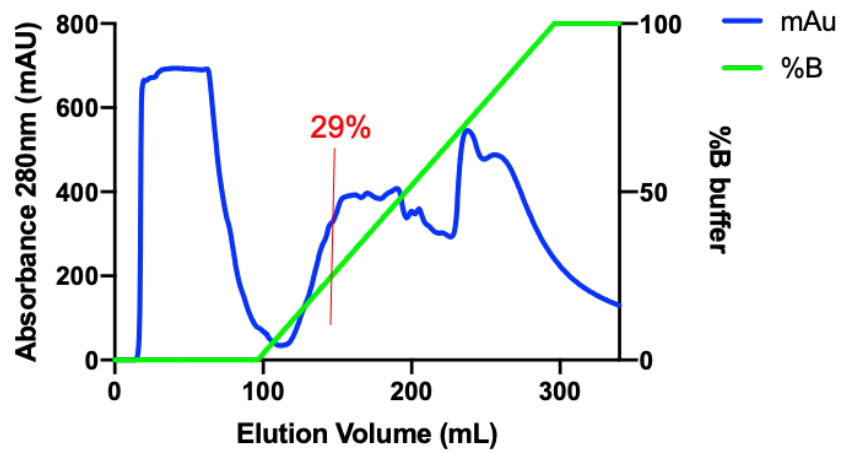
S2. Anion exchange affinity step of N-RsbR2 protein purification

Q Sepharose anion exchange chromatogram for N-RsbR2 (left) with absorbance at 280nm in blue and the elution buffer percentage in green. N-RsbR2 eluted at 68.6% of buffer B. 15% SDS-PAGE (right) with the molecular mass marker (MW), the input (IN), the flow-through (FT) from the column and the fractions from the elution buffer gradient. Fractions containing the protein of interest are boxed in red.



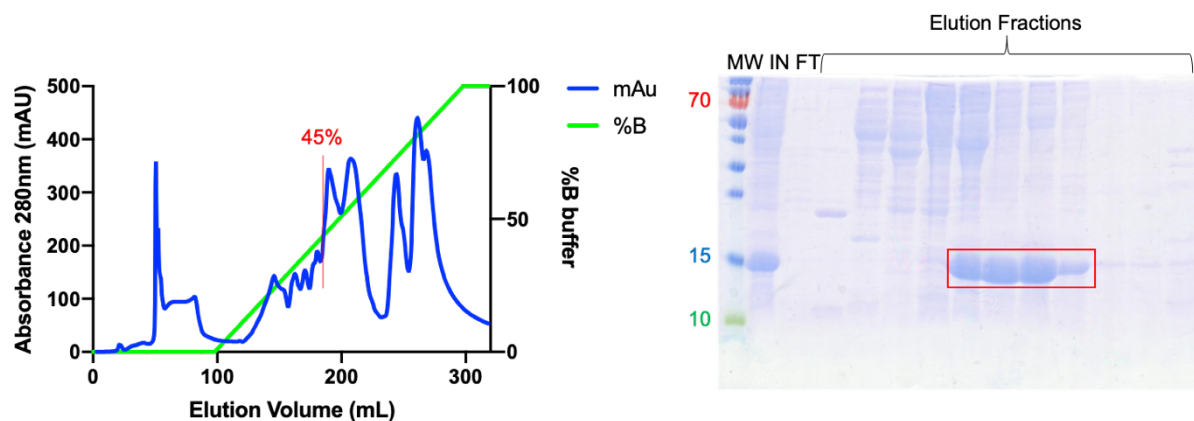
S3. Anion exchange affinity step of N-RsbR3 protein purification

Q Sepharose anion exchange chromatogram for N-RsbR3 (left) with absorbance at 280nm in blue and elution buffer percentage in green. N-RsbR3 eluted at 15.5% of buffer B. 15% SDS-PAGE (right) with the molecular mass marker (MW), the input (IN), the flow-through (FT) from the column and the fractions from the elution buffer gradient. Fractions containing the protein of interest are boxed in red.



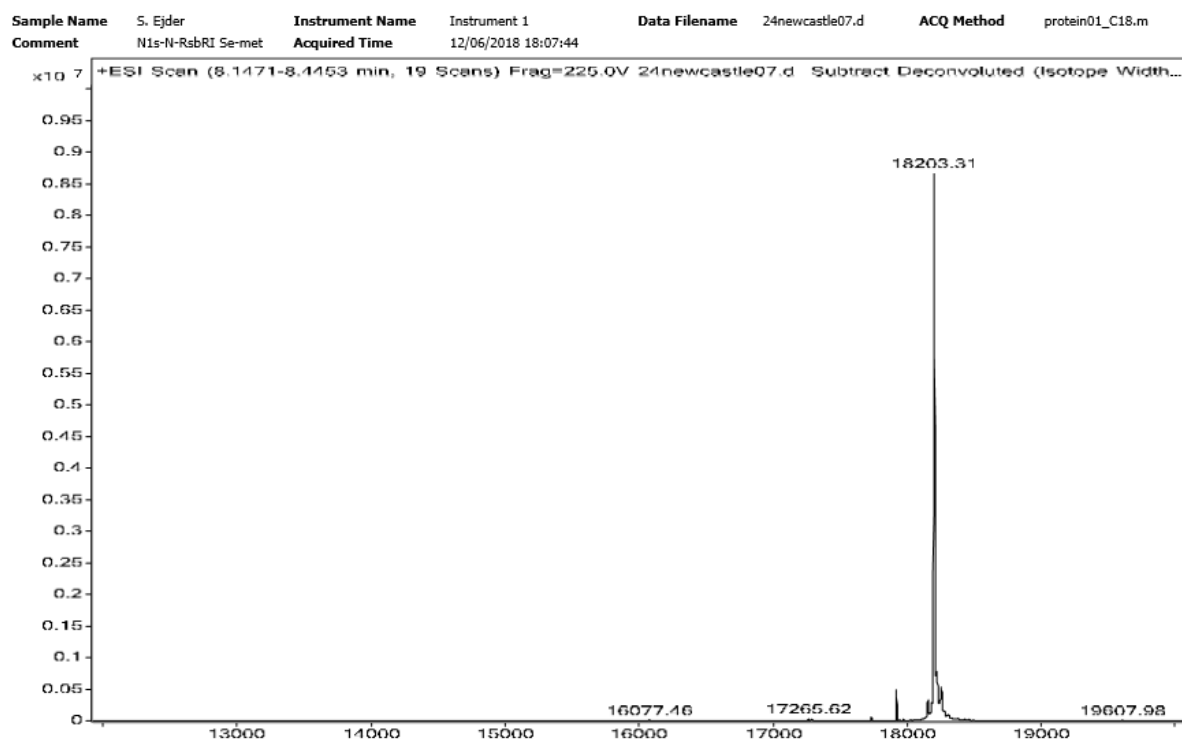
S4. Anion exchange affinity step of N-RsbR4 protein purification

Q Sepharose anion exchange chromatogram for N-RsbR4 (top) with absorbance at 280nm in blue and the elution buffer percentage in green. N-RsbR4 eluted at 29% of buffer B. 15% SDS-PAGE (bottom) with the molecular mass marker (MW), the input (IN), the pellet (P), the flow-through (FT) from the column and the fractions from the elution buffer gradient. Fractions containing the protein of interest are boxed in red.



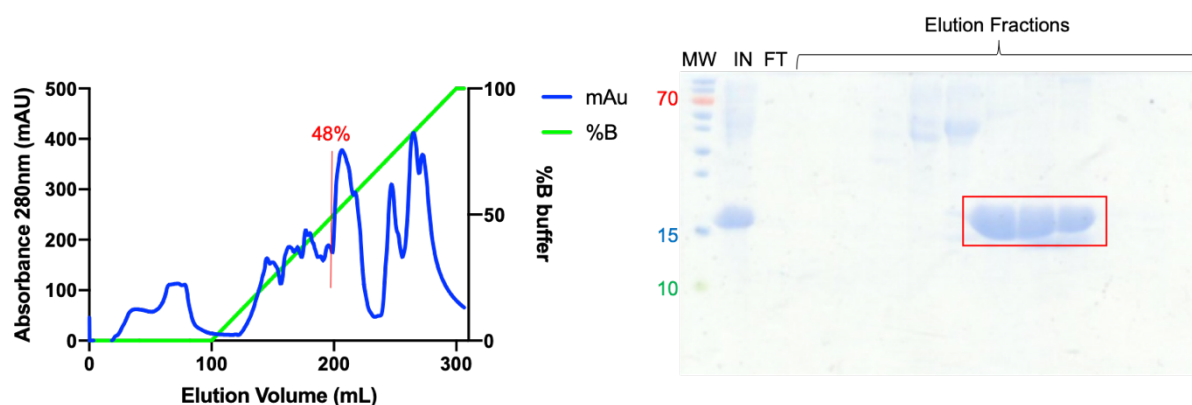
S5. Anion exchange affinity step of Se-Met labelled N-RsbR1 protein purification

Q Sepharose anion exchange chromatogram of Se-Met N-RsbR1 (left) with absorbance at 280nm in blue and elution buffer percentage in green. Se-Met N-RsbR1 eluted at approximately 45% of buffer B. 15% SDS-PAGE (right) with the molecular mass marker (MW), the input (IN), the flow-through (FT) from the column and the fractions from the elution buffer gradient. Fractions containing the protein of interest are boxed in red.



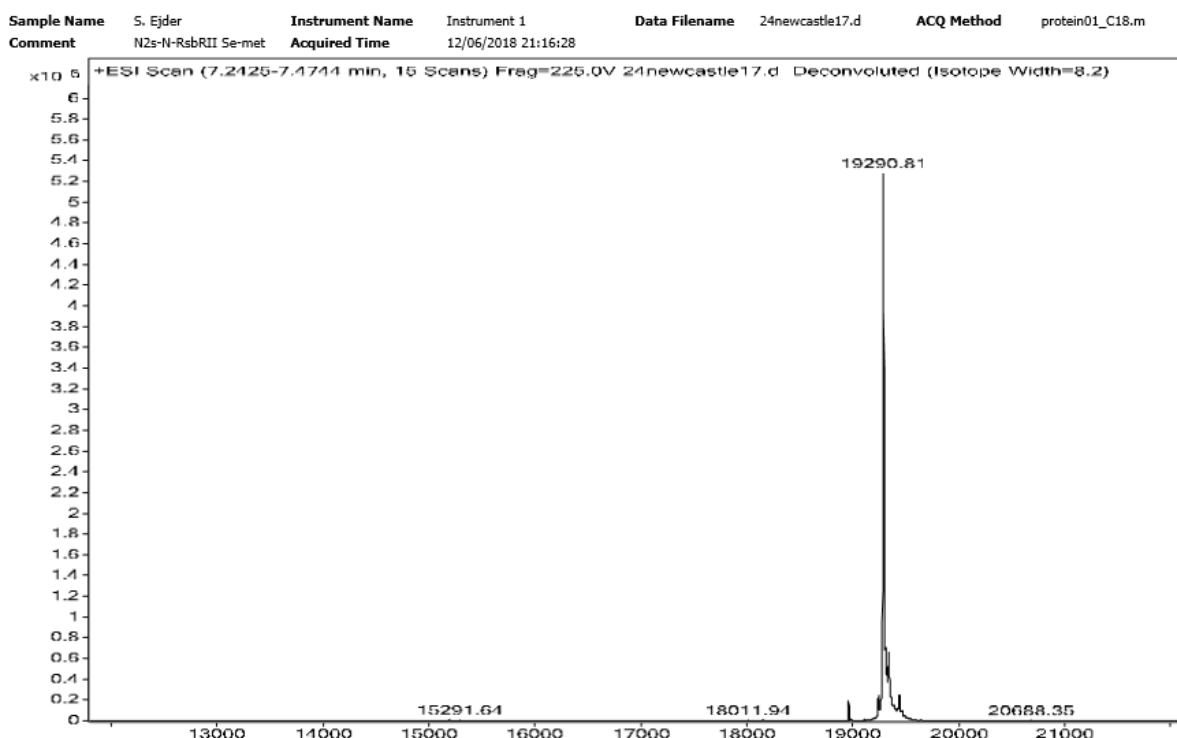
S6. LC-MS of the purified Se-Met labelled N-RsbR1

The LC-MS spectrum shows that most the purified Se-Met labelled N-RsbR1 has a MW of 18203.31 Da.



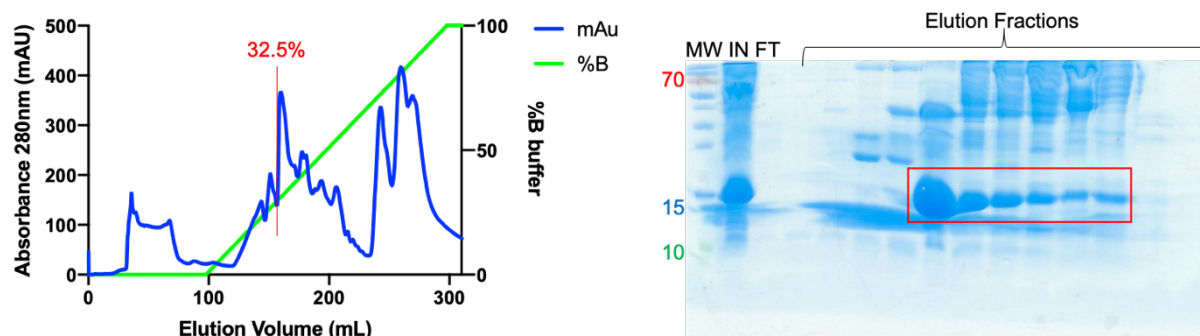
S7. Anion exchange affinity step of Se-Met labelled N-RsbR2 protein purification

Q Sepharose anion exchange chromatogram (left) with absorbance at 280nm in blue and elution buffer percentage in green. Se-Met N-RsbR2 eluted at approximately 48% of buffer B. 15% SDS-PAGE (right) with the molecular mass marker (MW), the input (IN), the flow-through (FT) from the column and the fractions from the elution buffer gradient. Fractions containing the protein of interest are boxed in red.



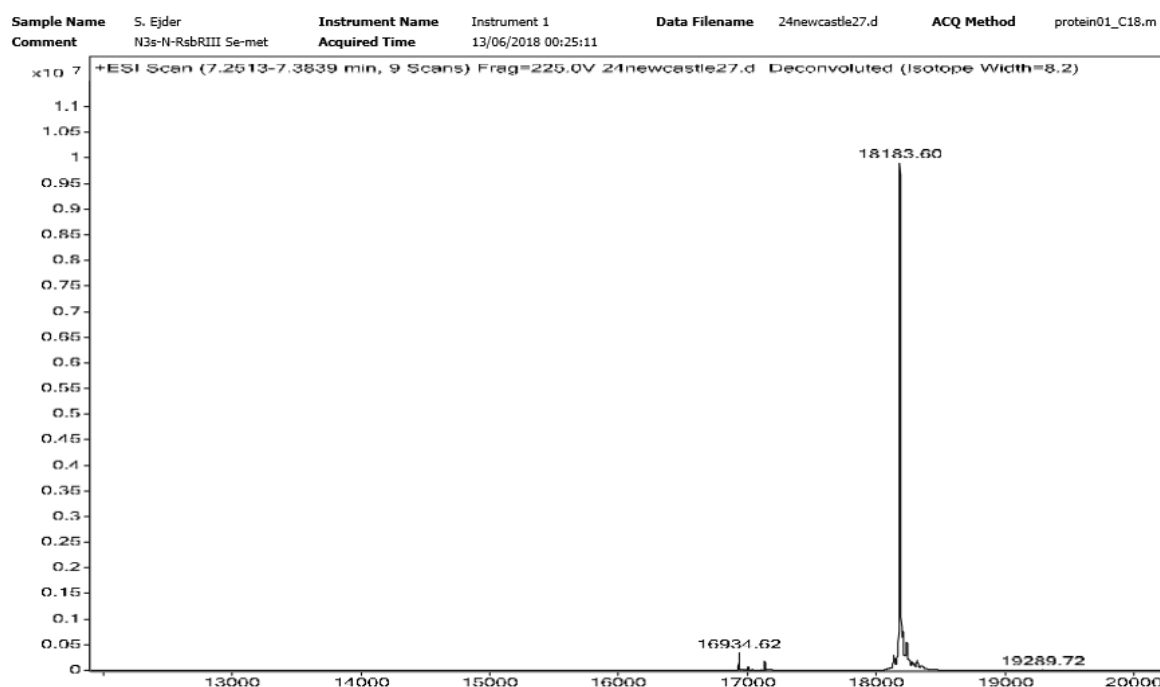
S8. LC-MS of the purified Se-Met labelled N-RsbR2

The LC-MS spectrum shows most the purified Se-Met labelled N-RsbR2 has a MW of 19290.81 Da.



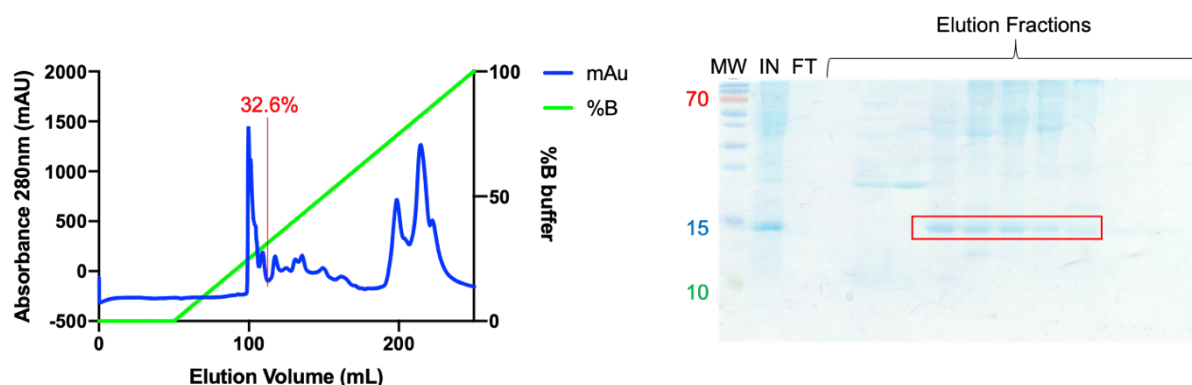
S9. Anion exchange affinity step of Se-Met labelled N-RsbR3 protein purification

Q Sepharose anion exchange chromatogram (left) with absorbance at 280nm in blue and elution buffer percentage in green. Se-Met N-RsbR3 eluted at 32.5% of buffer B. 15% SDS-PAGE (right) with the molecular mass marker (MW), the input (IN), the flow-through (FT) from the column and the fractions from the elution buffer gradient. Fractions containing the protein of interest are boxed in red.



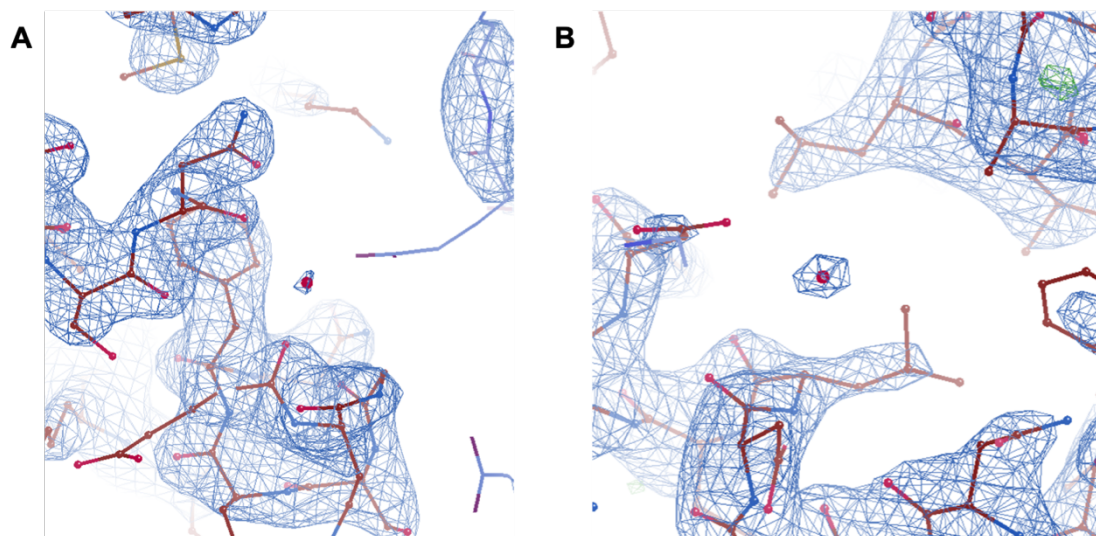
S10. LC-MS of the purified Se-Met labelled N-RsbR3

The LC-MS spectrum shows most the purified Se-Met labelled N-RsbR3 has a MW of 18183.60 Da.



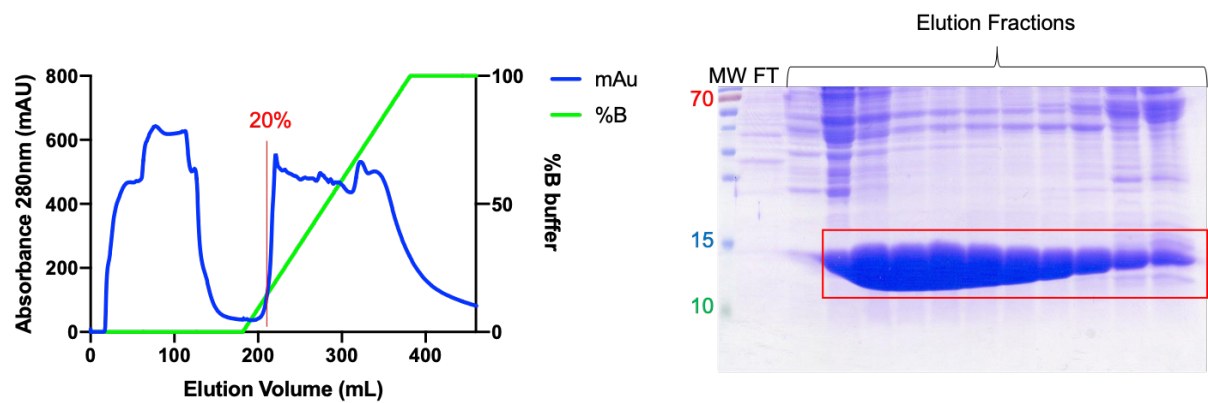
S11. Anion exchange affinity step of Se-Met labelled N-RsbR4 protein purification

Q Sepharose anion exchange chromatogram (left) with absorbance at 280nm in blue and the elution buffer percentage in green. Se-Met N-RsbR4 eluted at 32.6% of buffer B. 15% SDS-PAGE (right) with the molecular mass marker (MW), the input (IN), the flow-through (FT) from the column and the fractions from the elution buffer gradient. Fractions containing the protein of interest are boxed in red.



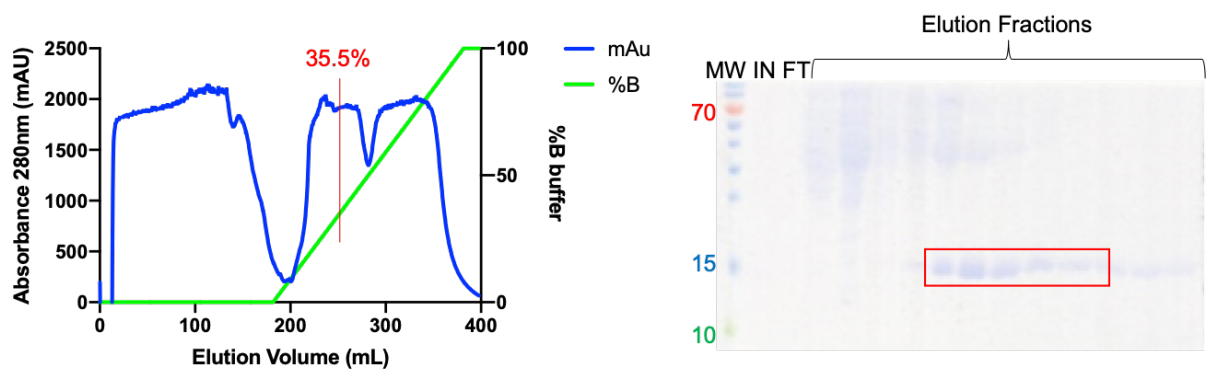
S12. Electron density maps of N-RsbR1 waters

The experimental 2mFo-dFc electron density map shown in blue, the difference map Fo-Fc shown in green/red of the N-RsbR1 crystal structure.



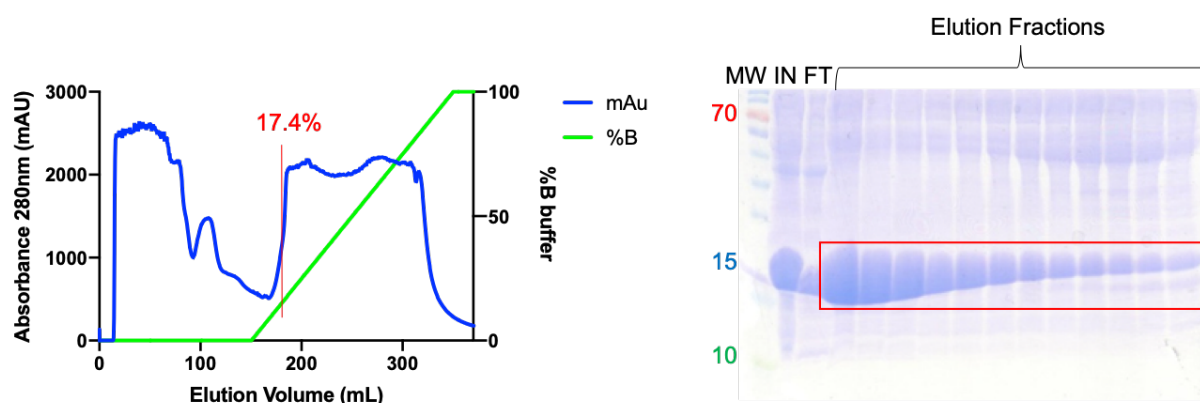
S13. Anion exchange affinity step of N-RsbRA protein purification

Q Sepharose anion exchange chromatogram (left) with absorbance at 280nm in blue and the elution buffer percentage in green. N-RsbRA eluted at 20% of buffer B. 15% SDS-PAGE (right) with the molecular mass marker (MW), the flow-through (FT) from the column and the fractions from the elution buffer gradient. Fractions containing the protein of interest are boxed in red.



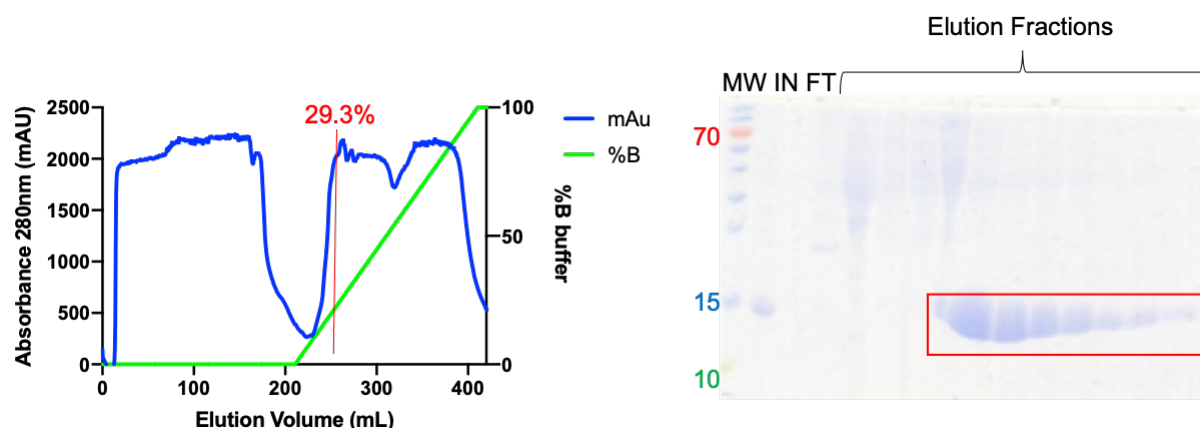
S14. Anion exchange affinity step of N-RsbRB protein purification

Q Sepharose anion exchange chromatogram (left) with absorbance at 280nm in blue and elution buffer percentage in green. N-RsbRB eluted at 35.5% of buffer B. 15% SDS-PAGE (right) with the molecular mass marker (MW), the input (IN), the flow-through (FT) from the column and the fractions from the elution buffer gradient. Fractions containing the protein of interest are boxed in red.



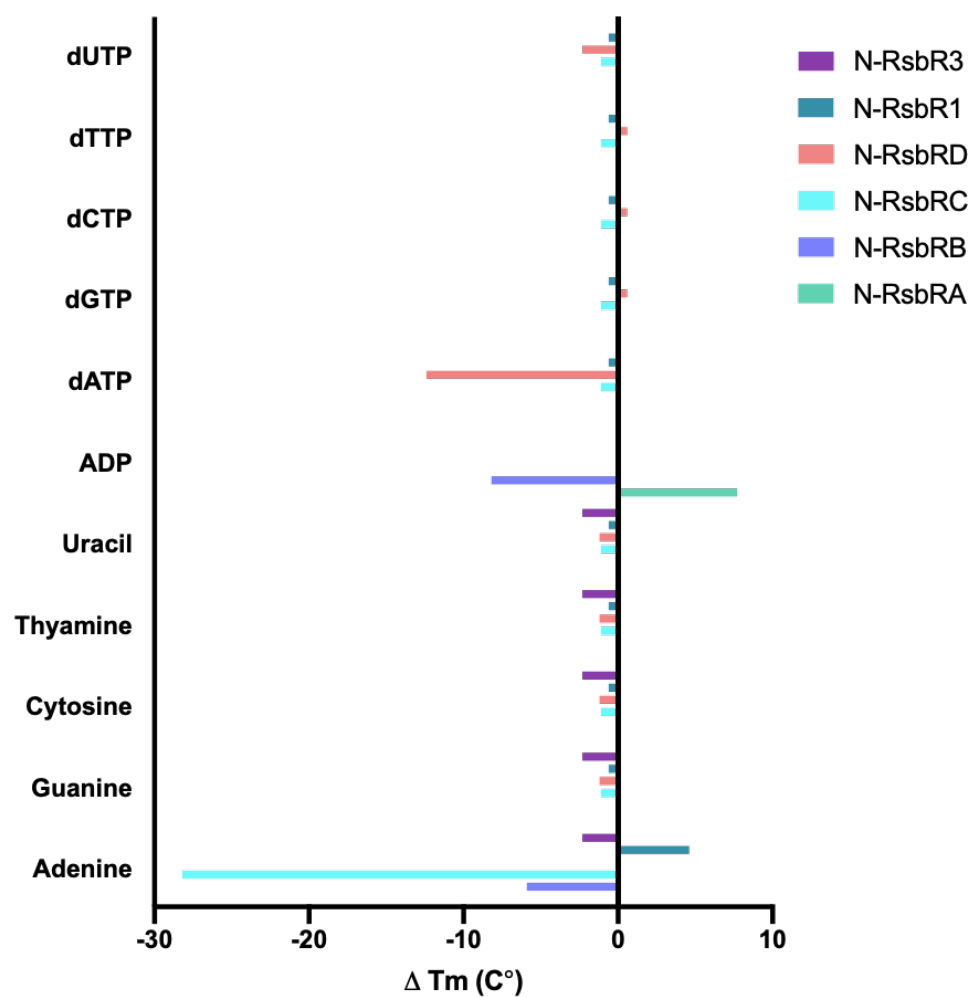
S15. Anion exchange affinity step of N-RsbRC protein purification

Q Sepharose anion exchange chromatogram (left) with absorbance at 280nm in blue and elution buffer percentage in green. N-RsbRC eluted at 17.4% of buffer B. 15% SDS-PAGE (right) with the molecular mass marker (MW), the input (IN), the flow-through (FT) from the column and the fractions from the elution buffer gradient. Fractions containing the protein of interest are boxed in red.



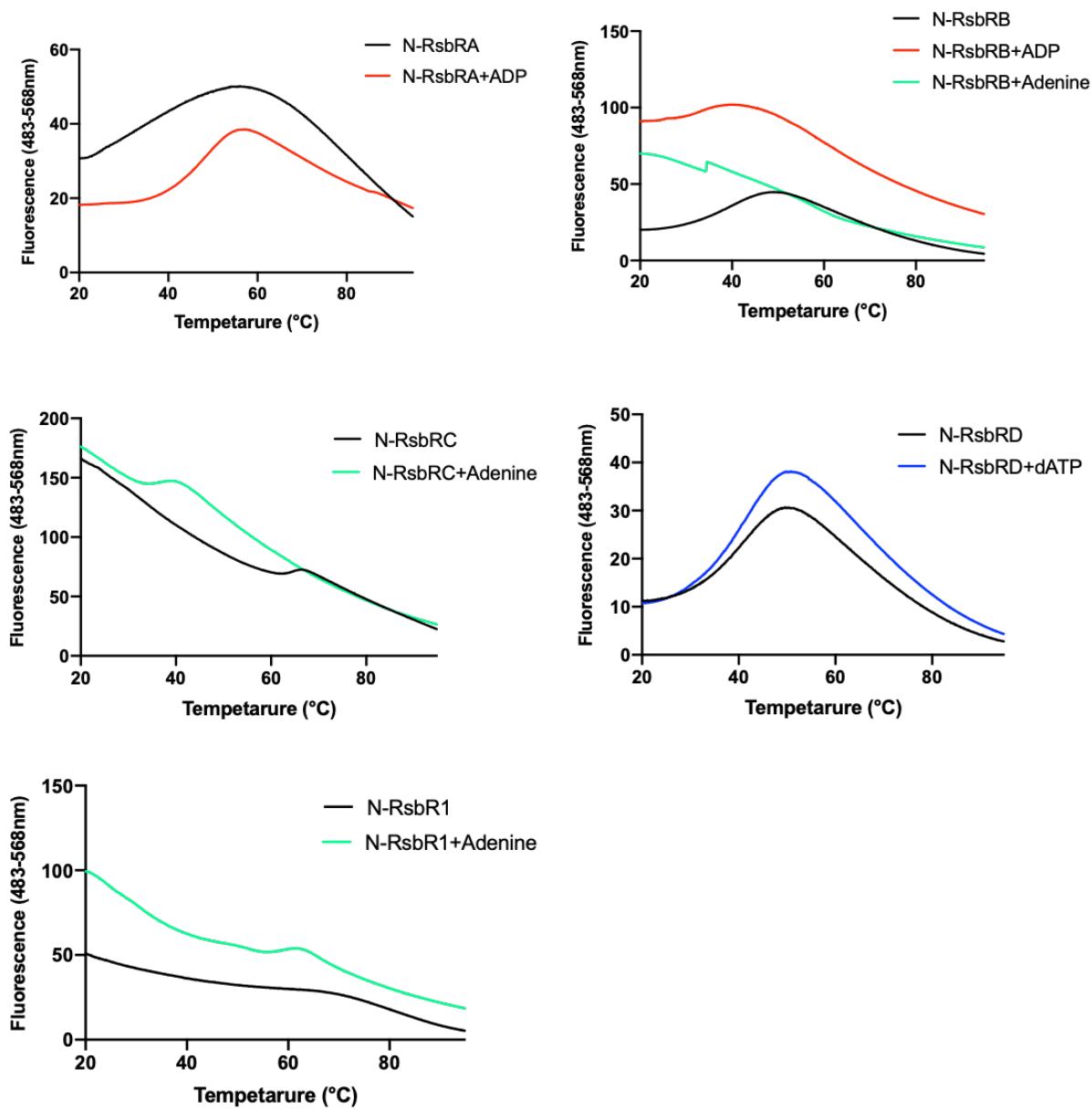
S16. Anion exchange affinity step of N-RsbRD protein purification

Q Sepharose anion exchange chromatogram (left) with absorbance at 280nm in blue and elution buffer percentage in green. N-RsbRD eluted at 29.3% of buffer B. 15% SDS-PAGE (right) with the molecular mass marker (MW), the input (IN), the flow-through (FT) from the column and the fractions from the elution buffer gradient. Fractions containing the protein of interest are boxed in red.



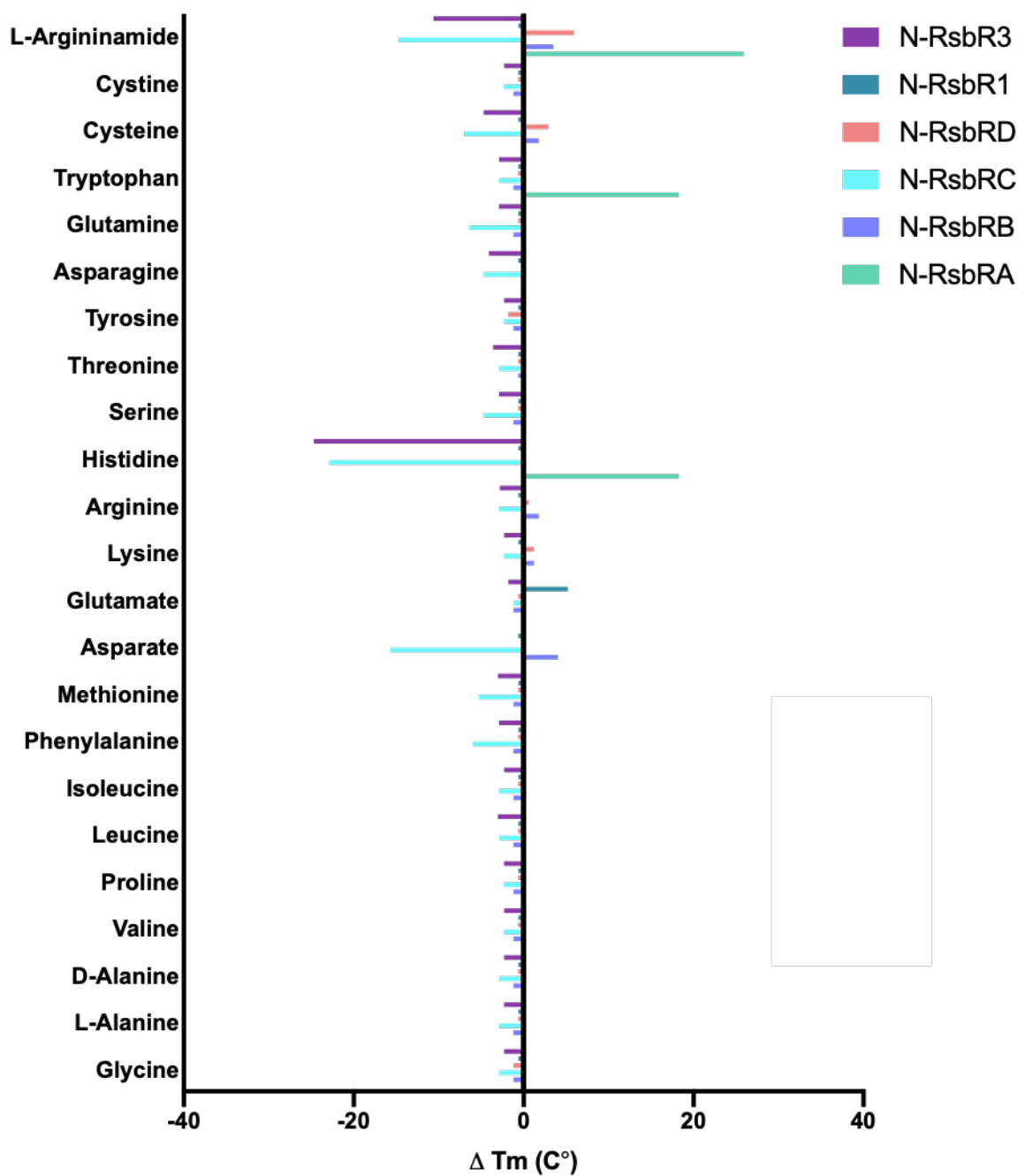
S17. Effect of all the nucleic acid compounds on N-RsbR thermostability.

T_m values obtained with compounds were subtracted from T_m values of the protein alone. The resulting T_m differences (ΔT_m) were plotted.



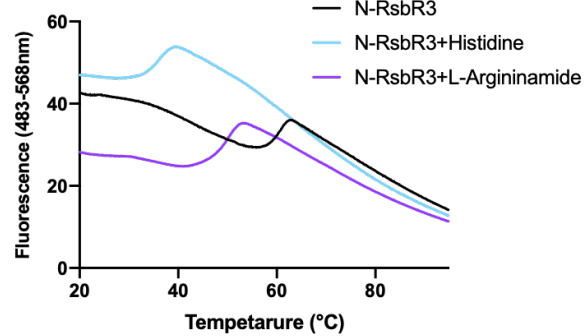
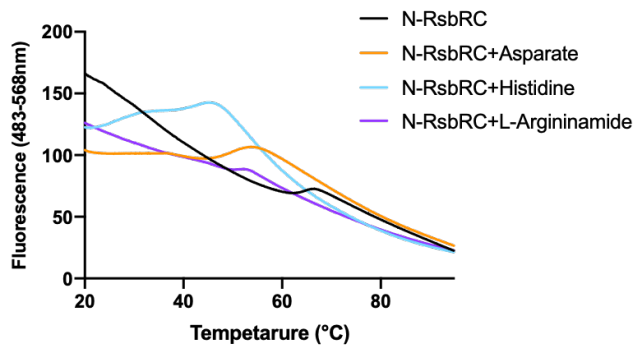
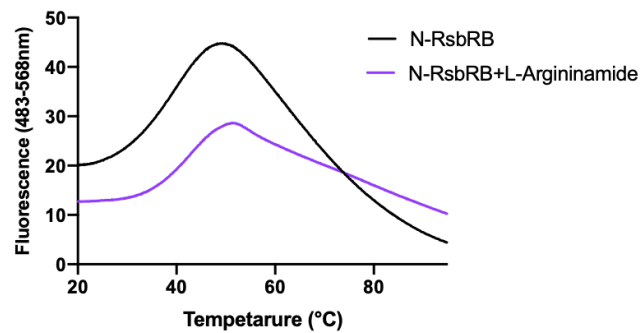
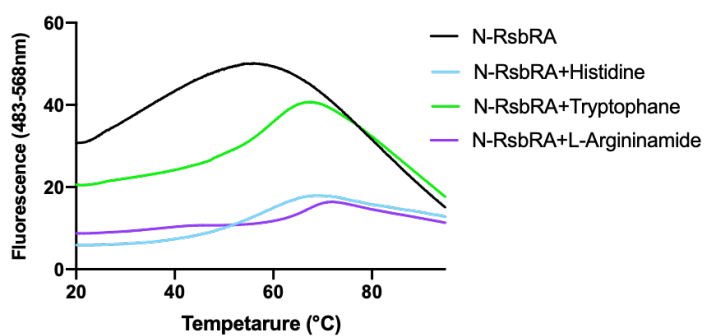
S18. TSA raw data with nucleic acid component.

Fluorescence curves of the protein without any compounds (black) and with compounds (ADP in red, adenine in green and dATP in blue) are shown.



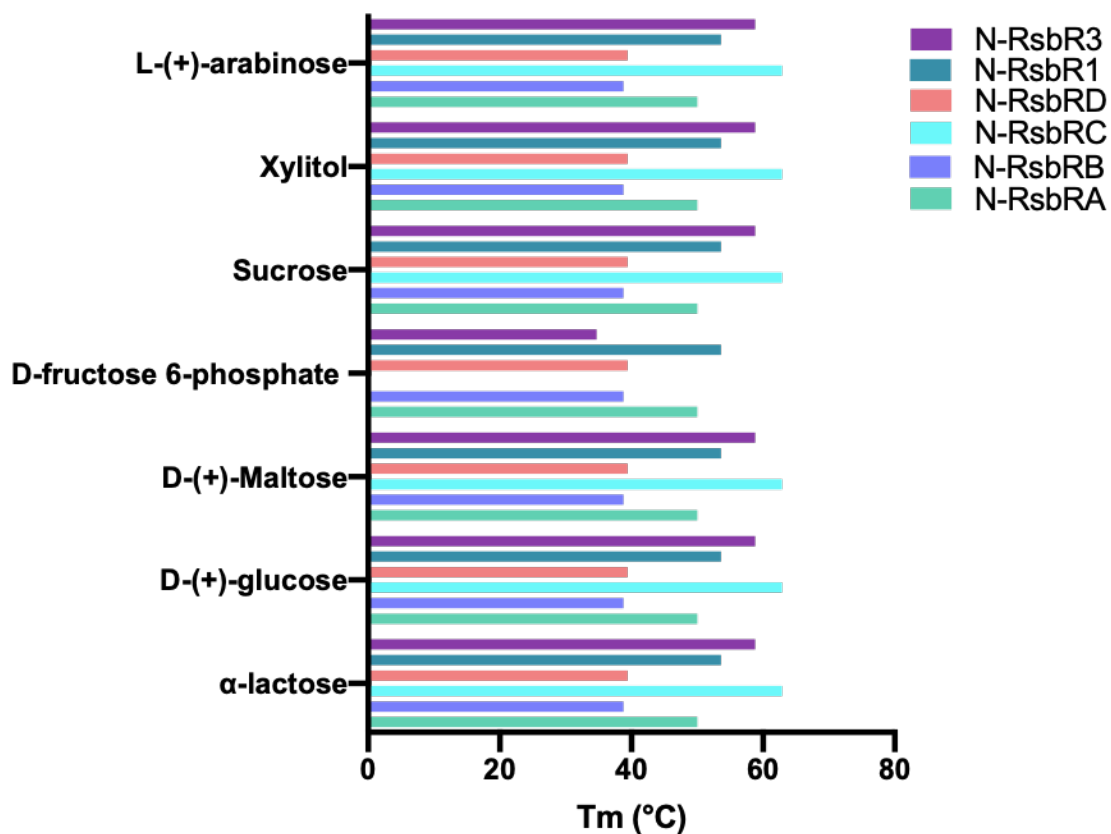
S19. Effect of all the amino acid compounds on N-RsbR thermostability.

T_m values obtained with compounds were subtracted from the T_m values of the protein alone. The resulting T_m differences (ΔT_m) were plotted.



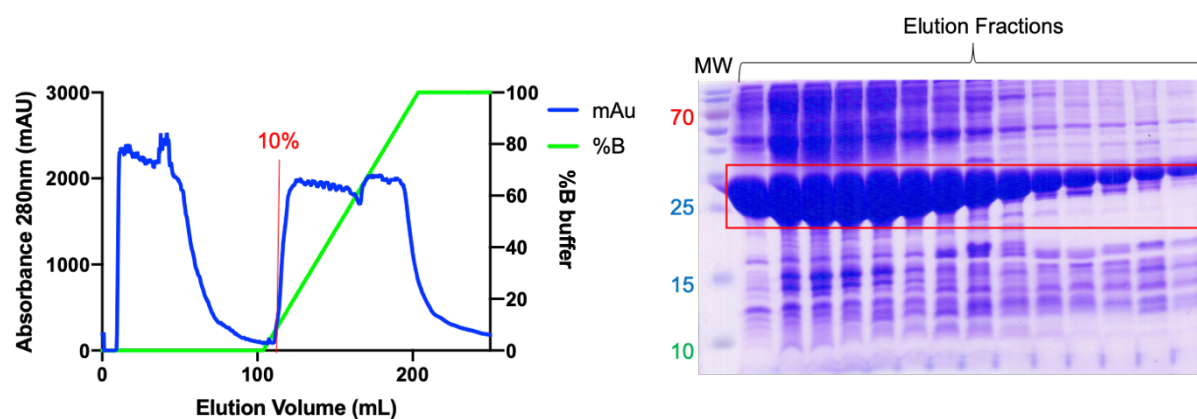
S20. TSA raw data with amino acid component.

Fluorescence curves of the protein without any compounds (black) and with compounds (histidine in light blue, tryptophan in green, L-argininamide in purple and aspartic acid in orange) are shown.



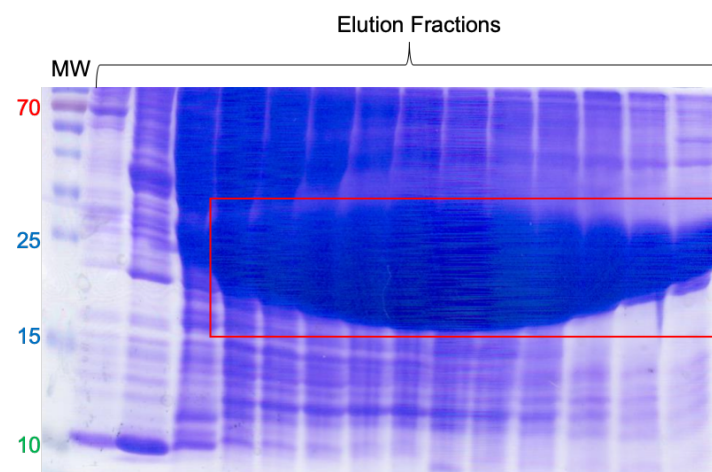
S21. Melting temperatures of then-RsbRs protein in presence of sugars.

T_m values obtained with sugar compounds are plotted on the graph.



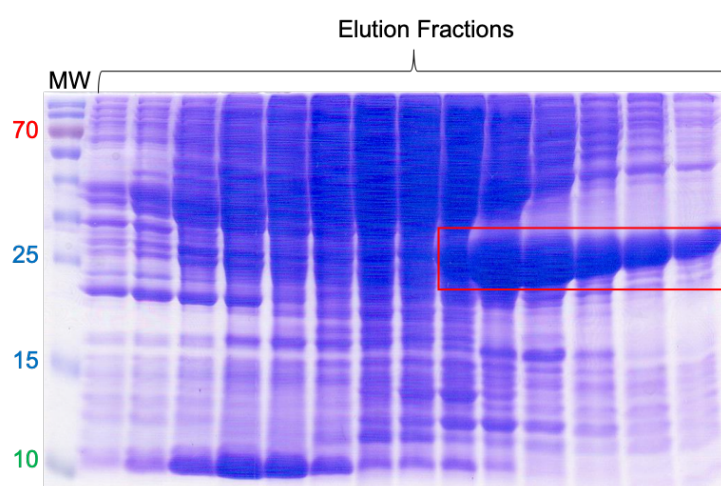
S22. Anion exchange affinity step of RsbR1 protein purification

Q Sepharose anion exchange chromatogram (left) with absorbance at 280nm in blue and elution buffer percentage in green. RsbR1 eluted at 10% of buffer B. 15% SDS-PAGE (right) with the molecular mass marker (MW) and the fractions from the elution buffer gradient. Fractions containing the protein of interest are boxed in red.



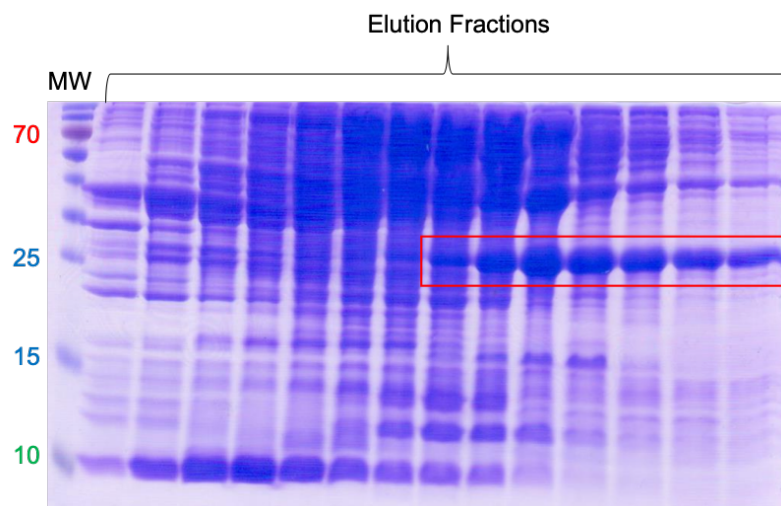
S23. SDS-PAGE of the anion exchange affinity step of RsbR2 protein purification

15% SDS-PAGE of the anion exchange chromatogram on which the molecular mass marker (MW) and the fractions from the elution buffer gradient were loaded. Fractions containing the protein of interest are boxed in red.



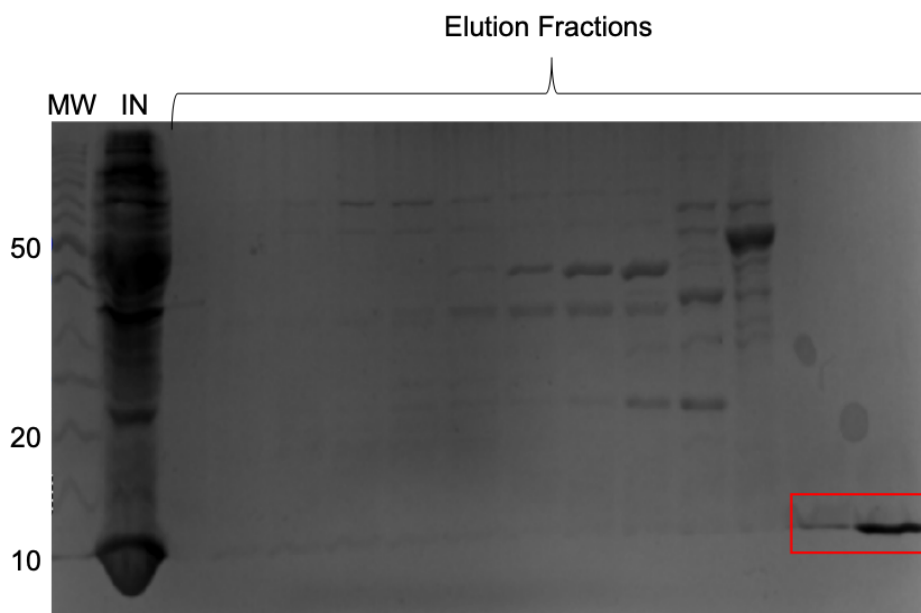
S24. SDS-PAGE of the anion exchange affinity step of RsbR3 protein purification

15% SDS-PAGE of the anion exchange chromatogram on which the molecular mass marker (MW) and the fractions from the elution buffer gradient were loaded. Fractions containing the protein of interest are boxed in red.



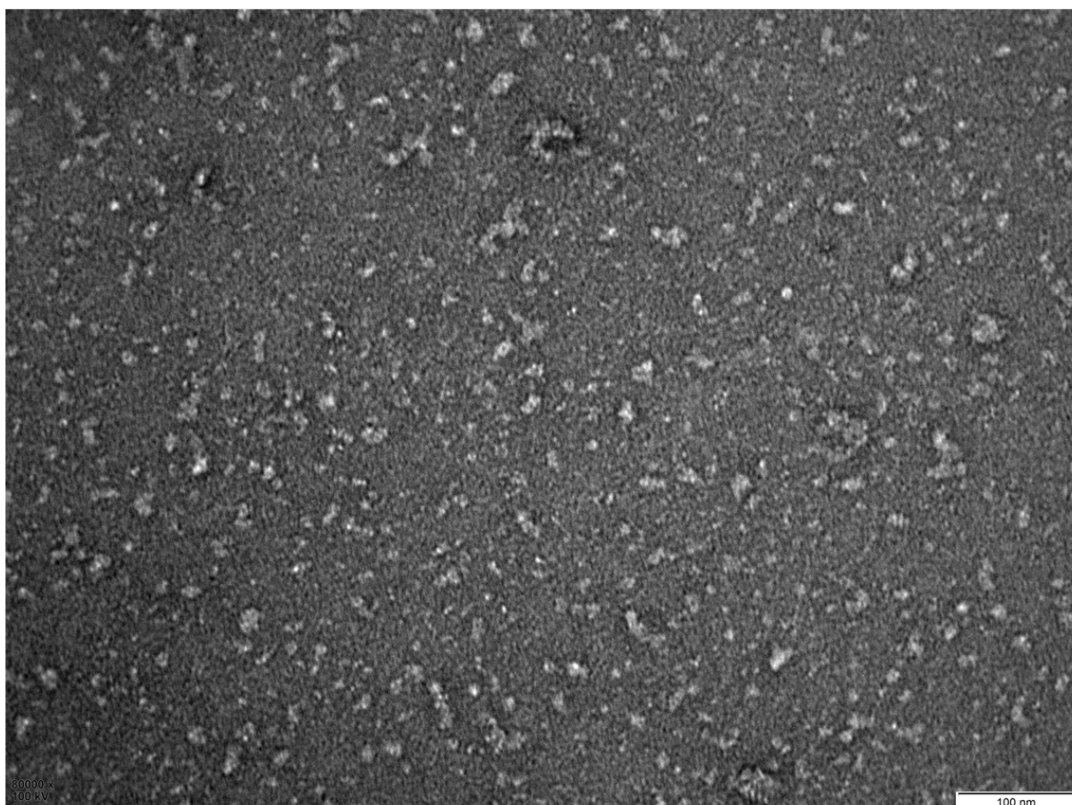
S25. SDS-PAGE of the anion exchange affinity step of RsbR4 protein purification

15% SDS-PAGE of the anion exchange chromatogram on which the molecular marker (MW) and the fractions from the elution buffer gradient were loaded. Fractions containing the protein of interest are boxed in red.



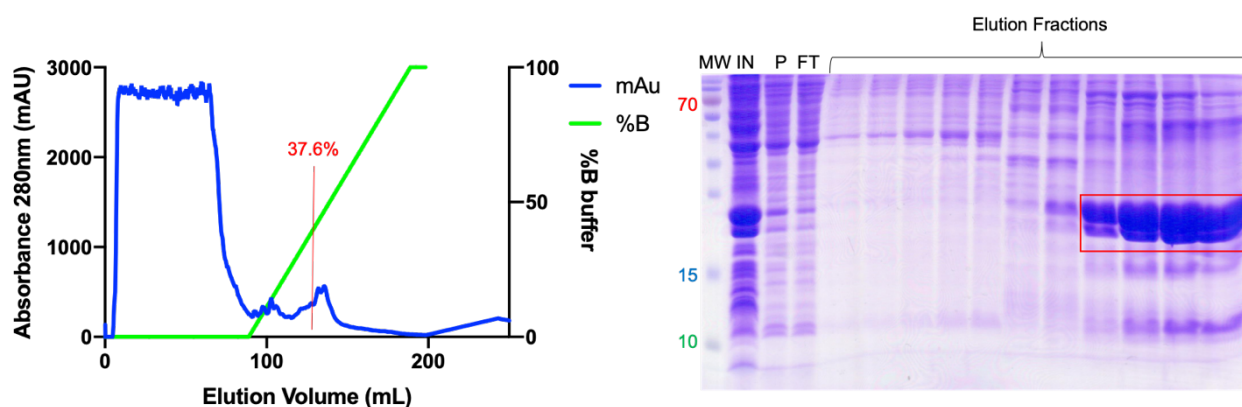
S26. SDS-PAGE of the anion exchange affinity step of RsbR4 protein purification

15% SDS-PAGE of the anion exchange chromatogram on which the molecular mass marker (MW) and the fractions from the elution buffer gradient were loaded. Fractions containing the protein of interest are boxed in red.



S27. Negative stain micrograph of RsbR1

Negative stain micrograph of purified RsbR1 that does not form stable stressosome complexes on its own. 80K magnification, 100 nm scale bar.



S28. Cation exchange affinity step of RsbX protein purification

S Sepharose anion exchange chromatogram (left) with absorbance at 280nm in blue and elution buffer percentage in green. RsbX eluted at 37.6% of buffer B. 15% SDS-PAGE (right) with the molecular mass marker (MW), the input (IN), the flow-through (FT) and the fractions from the elution buffer gradient. Fractions containing the protein of interest are boxed in red.