



University of  
Sheffield

Faculty of  
Science

Structural studies of surface antigen proteins of *Eimeria tenella* and related  
Apicomplexan parasites

A thesis submitted to The University of Sheffield for the degree of Doctor of  
Philosophy in the Faculty of Science

Abdul Aty Dakhil

The University of Sheffield

School of Bioscience

April 2024

## Abstract

Apicomplexa comprises a diverse group of parasitic organisms responsible for severe diseases in humans and livestock. Toxoplasmosis, caused by *Toxoplasma gondii*, poses risks to pregnant women, a problem with animals such as sheep and immunocompromised individuals and Malaria, caused by *Plasmodium spp.*, remains a global health challenge with complex transmission dynamics and widespread drug resistance. Chicken coccidiosis, caused by *Eimeria spp.*, leads to substantial economic losses in the poultry industry. *Eimeria* parasites exhibit distinct preferences for infecting specific sections of the chicken gut, mediated by proteins from micronemes protein, rhoptries proteins and surface antigens (SAGs) proteins of the CAP-like superfamily that are involved in host-parasite interactions and pathogenicity. Surface antigen proteins that belong to the separate SRS family in *Toxoplasma* and *Plasmodium* parasites play similar critical roles in invasion, immune evasion, and virulence. Understanding the diversity and functions of these surface antigen proteins may pave the way for novel intervention strategies against apicomplexan diseases.

This thesis presents the design and evaluation of a construct for the efficient production of large quantities of soluble CAP-like SAG proteins, based on a cleavable thioredoxin solubility tag, optimised codon usage and structure based identification of the core SAG domain boundaries, resulting in the successful expression of twenty-two CAP-like SAG proteins. The structures of six representative SAG proteins from the three *Eimeria* SAG subfamilies were determined by X-ray crystallography, with each having the same core single domain  $\alpha\beta\alpha$  sandwich structure, containing two or three disulphide bonds and a conserved buried arginine in an NxxR motif. Despite the similarity in secondary structure, clear differences in the length, position and sequence of the connecting loops results in a wide diversity of surface shape, charge and functional groups, indicating that this family of CAP-like SAG proteins are unlikely to share a common binding partner. Genomic analysis revealed that CAP-like SAG genes are present in other apicomplexan parasite genomes that contain multiple SRS-like SAG proteins such as *Toxoplasma gondii*, *Neospora caninum*, *Besnoitia besnoiti* and *Cystoisospora suis*, with *Plasmodium spp.*, *Babesia bovis* and *Theileria orientalis* having just a single CAP-like SAG gene. The structure of the *P. vivax* CAP-like SAG was determined revealing the same core, single domain structure, with sequence analysis across these parasites indicating that some parasites have examples of genes with multiple CAP-like SAG domains. Transcriptome



analysis showed no consistent expression of this protein family in different parasite life stages, possibly suggesting that the primary function of some, or perhaps all, of these proteins is more likely to be associated with the stimulation of the immune system as a response to parasite invasion. In conclusion, this thesis contributes to our understanding of the structural and functional diversity of CAP-like SAGs across apicomplexan parasites, paving the way for future studies aimed at elucidating their roles in parasite biology and exploring their potential as therapeutic targets or vaccine candidates.

## ACKNOWLEDGMENTS

In completing this thesis, I wish to extend my deepest gratitude to my esteemed supervisors, Dr. John Rafferty and Dr. Patrick Baker. Their unwavering guidance, support, and expertise have been invaluable throughout this journey. Their insightful advice and encouragement helped shape the direction of my research and provided me with the tools to navigate through the myriad challenges encountered in the laboratory. I am particularly indebted to Dr. Patrick for his assistance in crafting this thesis and meticulously analysing the results. His dedication and attention to detail have significantly enhanced the clarity and coherence of this work. Additionally, I extend my heartfelt appreciation to Prof. David Rice for his invaluable assistance in analysing the results in a scientific manner. His expertise and constructive feedback were instrumental in refining the findings presented in this thesis.

Special thanks are due to Dr. Sam Dix for his invaluable aid at the inception of my project, providing essential guidance on all the necessary requirements. I am equally grateful to Dr. Svetlana Sedelnikova for her steadfast support during the protein purification steps, which played a pivotal role in the success of this research. I extend my gratitude to Dr. Aidan Taylor for providing me with crucial information for the ThermoFluor assay. I also wish to express my appreciation to both my advisors, Prof. Per Bullough and Dr. Rosemary Staniforth, for their unwavering support during moments of uncertainty at the project's outset. A special mention is reserved for my colleague, Dr. Hannah Fisher, whose assistance with various submission requirements and technical software proved invaluable.

My deepest gratitude is reserved for my family, whose steadfast support and understanding have sustained me throughout this endeavour. Their patience and encouragement during moments of intense experimentation have been a constant source of strength and motivation. I extend special thanks to my father and mother, whose prayers and encouragement have inspired me to strive for success.

Finally, I would like to express my gratitude to all individuals, colleagues, and friends who have contributed to this work in various capacities. Your support, encouragement, and collaboration have been truly invaluable and have made this research possible.

## ABBREVIATIONS AND SYMBOLS

AIDS	Acquired immune deficiency syndrome
AMA	Apical membrane antigen
bp	Base pair
° C	Degree Celsius/ degree centigrade
Ca <sup>2+</sup>	Ion Calcium
CaCl <sub>2</sub> .2H <sub>2</sub> O	Calcium chloride dihydrate
CAP	Cysteine-rich secretory proteins, antigen 5 protein and plant pathogenesis-related 1 protein
Da	Dalton
EDTA	Ethylene diamine tetra acetic acid
EGF	Epidermal growth factor
ER	Endoplasmic reticulum
E.t	<i>Eimeria tenella</i>
EtMIC	<i>Eimeria tenella</i> microneme protein
GAPR-1	Golgi-associated pathogenesis-related 1
GPI	Glycosylphosphatidylinositol
GRA	Dense granule
IFN- $\gamma$	Inteferon gamma
IL	Interleukin
IMC	Inner membrane complex
IPTG	Isopropy $\beta$ -D-1- thiogalactopyranoside
K <sub>2</sub> HPO <sub>4</sub>	Dipotassium phosphate
KH <sub>2</sub> PO <sub>4</sub>	Monopottasium phosphate

kDa	Kilodalton
MIC	Microneme protein
MJ	Moving Junction
MSP1	Merozoite surface protein 1
NaCl	Sodium chloride
NaOH	Sodium Hydroxide
ng	Nanogram
(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	Ammonium sulphate
NO	Nitric oxide
OD	Optical density
PM	Plasma membrane
PR-1	Pathogenesis related 1 protein
PV	Parasitophorous vacuole
PVM	Parasitophorous vacuole membrane
PAGE	Polyacrylamide gel electrophoresis
PDB	Protein Data Bank
PEG	Polyethelene glycol
rpm	Revolutions per minute
RON	Rhoptry neck protein
ROP	Rhoptry protein
SAG	Surface antigen protein
SDS	Sodium dodecyl sulphate
<i>spp</i>	Species
SRS	SAG-related sequence protein

TEMED	N,N,N',N',tetramethylethylenediamine
Tris	Tris(hydroxymethyl)methylamine
TNF- $\alpha$	Tumour necrosis factor $\alpha$
TH-1/TH-2	T-helper cells $\frac{1}{2}$
TLR	Toll-Like Receptors
UV	Ultra-Violet
v/v	Volume per volume
w/v	Weight per volume

### Crystallographic terms

Å	Angstrom
a, b, c	Real space unit cell dimensions
$\alpha, \beta, \gamma$	Real space unit cell angles
a. u.	Asymmetric unit
B	Crystallographic temperature factor
FT	Fourier Transform
$\alpha_{\text{calc}}$	Calculated phases
$F_{hkl}$	Structure factor for reflection
$F_{\text{obs}}$	Observed structure factor
$hkl$	Reciprocal lattice indices
I	Intensities
$I_{hkl}$	Intensities of the reflection $hkl$
$\rho(x, y, z)$	Fourier Transform of a structure factor to calculate the electron density in which for every x, y and z position in unit cell

MAD	Multi wavelength anomalous dispersion
R-factor	Crystallographic refinement R-factor
$R_{\text{free}}$	Free R-factor
$R_{\text{merge}}$	R-factor relating agreement between symmetry related reflections
r.m.s	Root mean square
$R_{\text{work}}$	Working R-factor
$\sigma(I)$	SigmaI
$\lambda$	X-ray wavelength
$\varphi, \Psi, \kappa$	Polar angles
V	Volume of unit cell
$V_M$	Matthew's number

## Table of contents

### Chapter1

<b>1.1 Introduction.....</b>	<b>1</b>
1.1.1 Cellular Structure of Apicomplexan Parasites .....	1
1.1.2 Lifecycle of Apicomplexan Parasites.....	5
1.1.3 Host-Pathogen Interactions .....	6
1.1.4 Disease of Apicomplexan Parasites .....	9
1.1.5 Eimeria parasite.....	13
1.1.6 Extending the exploration to other apicomplexan surface antigens related to the Eimeria tenella SAG proteins .....	27
1.1.7 Comparison of SAG proteins Within Apicomplexan parasites .....	31
<b>1.2 Aims of the Project .....</b>	<b>35</b>

### Chapter2..... 36

<b>2.1 MATERIALS AND METHODS.....</b>	<b>36</b>
2.1.1 Strains, vectors, and recombinant plasmids .....	36
<b>2.2 Growth Media and Solutions.....</b>	<b>39</b>
2.2.1 Stock Solutions.....	39
2.2.2 Luria-Bertani broth media (LB) .....	39
2.2.3 LB Agar.....	39
2.2.4 M9 Minimal Salts, 5X.....	39
2.2.5 Terrific Broth (TB) media .....	40
2.2.6 Antibiotics .....	40
2.2.7 Transformation .....	40
2.2.8 Plasmid Extraction .....	41
<b>2.3 Cell Growth .....</b>	<b>41</b>
2.3.1 Over Expression .....	41
2.3.2 Preparation of Cell Free Extracts (CFE) .....	42
2.3.3 Determination of protein concentration .....	43
2.3.4 SDS-PAGE Electrophoresis.....	43
<b>2.4 Protein purification .....</b>	<b>44</b>
o Protocols .....	44
<b>2.5 Protein crystal preparation.....</b>	<b>46</b>
2.5.1 Initial Robot Screening.....	46
2.5.2 Crystal Optimisation .....	46
2.5.3 Mounting Crystals .....	47
2.5.4 Structure Determination .....	48
2.5.5 Data collection.....	49
2.5.6 Data processing .....	49
2.5.7 Molecular replacement (MR) .....	50
2.5.8 Model Building and Refinement.....	50
2.5.9 Validation and analysis .....	51
<b>2.6 ThermoFluor assay .....</b>	<b>51</b>

### Chapter3 Cloning, Overexpression, Purification, and Crystallisation of Surface Antigen Protein (SAG)..... 52

<b>3.1 Overview .....</b>	<b>52</b>
<b>3.2 Construct Design.....</b>	<b>53</b>
3.2.1 Core SAG Length.....	54

3.2.2 Codon Optimization .....	54
3.2.3 <i>PelB</i> expression system.....	54
3.2.4 <i>Trx</i> expression system in <i>Origami</i> cells .....	57
<b>3.3 SAG Protein Overexpression.....</b>	<b>58</b>
<b>3.4 Purification of the SAG proteins .....</b>	<b>59</b>
<b>3.5 SAG-B Family .....</b>	<b>60</b>
3.5.1 Surface Antigen 13 (SAG-B13) .....	60
3.5.2 Surface Antigen 16 (SAG-B16) .....	62
3.5.3 Surface Antigen 22 (SAG-B22) .....	64
3.5.4 Surface Antigen 41 (SAG-B41) .....	66
<b>3.6 SAG-C family .....</b>	<b>68</b>
3.6.1 SAG-C family Overexpression .....	69
3.6.2 Crystallization of SAG-C Family Proteins.....	74
<b>3.7 <i>Eimeria Tenella</i> SAG A family.....</b>	<b>76</b>
3.7.1 Overexpression and Purification .....	76
3.7.2 Crystallization of SAG-A Proteins.....	85
<b>3.8 Summary .....</b>	<b>88</b>
<b><i>Chapter4 Structure determination and preliminary crystallographic analysis of E. tenella SAG proteins. ....</i></b>	<b><i>89</i></b>
<b>4.1 Overview .....</b>	<b>89</b>
<b>4.2 Data collection and processing .....</b>	<b>89</b>
<b>4.3 Structure Determination of SAG-B family Proteins .....</b>	<b>90</b>
4.3.1 SAG-B13 Data Processing .....	90
4.3.2 SAG-B16 Data Processing .....	95
4.3.3 SAG-B22 Data Processing .....	101
4.3.4 SAG-B41 Data Processing .....	107
4.3.5 SAG-A91 Data Processing.....	114
4.3.6 SAG-C75 Data Processing .....	118
<b>4.4 Summary .....</b>	<b>123</b>
<b><i>Chapter5 Structural analysis of Eimeria Tenella SAG proteins.....</i></b>	<b><i>125</i></b>
<b>5.1 Overview .....</b>	<b>125</b>
<b>5.2 Exploring the Structural Similarity in the SAG-B Family. ....</b>	<b>126</b>
5.2.1 Disulphide bonds and cis-peptides in the SAG-B members. ....	129
5.2.2 Exploring the Divergence within the SAG-B Family .....	130
5.2.3 Structure-Based Sequence Alignment of Family B SAG Proteins .....	133
5.2.4 All <i>E. tenella</i> SAG-B Family .....	137
<b>5.3 The Structure of SAG-A91 an SAG-A family member.....</b>	<b>140</b>
5.3.1 The Comparison between SAG-A91 and SAG-B family structures.....	142
5.3.2 Sequence Diversity and Similarity within the SAG-A family. ....	143
<b>5.4 The structure of SAG-C75.....</b>	<b>145</b>
<b>5.5 Comparison of the structures of the SAG-A, SAG-B and SAG-C families.....</b>	<b>147</b>
<b>5.6 <i>E. tenella</i> Binding Partner for SAGs protein .....</b>	<b>153</b>
<b><i>Chapter6 Overexpression, purification, crystallisation and structure determination of CAP-like SAG in Plasmodium Vivax (P. vivax) .....</i></b>	<b><i>155</i></b>



<b>6.1 Overview .....</b>	<b>155</b>
<b>6.2 Overexpression.....</b>	<b>156</b>
<b>6.3 Purification of <i>P. vivax</i> CAP-like SAG .....</b>	<b>158</b>
<b>6.4 Crystallization of <i>P. vivax</i> CAP-like SAG.....</b>	<b>160</b>
<b>6.5 <i>P. vivax</i> CAP-like SAG Data Collection and Processing .....</b>	<b>160</b>
6.5.1 Molecular Replacement.....	162
6.5.2 <i>P. vivax</i> CAP-like SAG Refinement .....	162
6.5.3 Structure Validation of <i>P. vivax</i> CAP-like SAG. ....	166
6.5.4 Structure Description of <i>P. vivax</i> CAP-like SAG .....	167
<b>6.6 Conservation of the <i>P. vivax</i>-like SAG across different species of <i>Plasmodia</i> .....</b>	<b>172</b>
<b><i>Chapter7 Discussion</i> .....</b>	<b>177</b>
<b>7.1 Design of a new Trx construct .....</b>	<b>177</b>
<b>7.2 CAP-like SAGs in the wider apicomplexan parasite family. ....</b>	<b>178</b>
<b>7.3 Critical conserved features of the CAP- superfamily structure. ....</b>	<b>185</b>
<b>7.4 Discovery of ETH-00002020 is the closest homologue to CAP-like SAGs (<i>Babesia bovis</i>, <i>Theileria</i>, <i>Besnoitia</i>, <i>Neospora</i>, <i>Cystoisospora</i>, <i>Toxoplasma</i> and <i>Cyclospora</i>). ....</b>	<b>195</b>
<b>7.5 The Function of Apicomplexan CAP-like SAGs.....</b>	<b>196</b>
<b>7.6 Concluding remarks .....</b>	<b>199</b>
<b>8. References.....</b>	<b>197</b>
<b>9. Appendix.....</b>	<b>211</b>

## List of Figure

Figure 1.1: Apicomplexan Morphology. ....	2
Figure 1.2: Apicomplexa gliding motility.. ....	2
Figure 1.3: The general structure for the apical complex for Apicomplexan.....	4
Figure 1.4: Apicomplexan Life Cycle.....	6
Figure 1.5: The invasion process of Apicomplexan parasites.....	9
Figure 1.6: The life cycle of <i>Toxoplasma gondii</i> .....	10
Figure 1.7: Life cycle of <i>Plasmodium</i> parasites.. ....	12
Figure 1.8: <i>Eimeria</i> Life Cycle.....	18
Figure 1.9: SAG attachment to membrane. ....	21
Figure 1.10: The surface antigen <i>Eimeria tenella</i> SAG1 with host cell.....	25
Figure 1.11: <i>Eimeria tenella</i> SAG19 Structure.. ....	26
Figure 1.12: Represents the surface antigens of <i>Toxoplasma gondii</i> .. ....	29
Figure 1.13: The s48/45 six-cysteine protein family plays a vital role in the <i>P. falciparum</i> life cycle.....	31
Figure 1.14: A comparison of the structure of <i>T. gondii</i> , <i>P. falciparum</i> and <i>E. tenella</i> SAG proteins ..	33
Figure 3.1: The full length of SAG sequence.....	53
Figure 3.2: The original Trx-construct.. ....	53
Figure 3.3: The PelB construct inserted into the pET-22b vector.. ....	55
Figure 3.4: SDS-PAGE analysis the SAG-B19 overexpression on M9 media at 16°C using the PelB expression system.....	56
Figure 3.5: Trx-construct insertion into pET32a.....	57
Figure 3.6: SDS-PAGE analysis of one example of each SAG family overexpression were tested at 17°C with 1mM IPTG induction.. ....	58
Figure 3.7: SAG-B13 purification process.....	61
Figure 3.8: Crystal of SAG-B13.....	62
Figure 3.9: SAG-B16 purification process.....	63
Figure 3.10: Crystals of SAG-B16.....	64
Figure 3.11: SAG-B22 purification process.....	65
Figure 3.12: Crystals of SAG-B22 grown in different condition from PACT suite.....	66
Figure 3.13: SAG-B41 purification process.....	67
Figure 3.14: Grown crystals of SAG-B41.....	68
Figure 3.15: SDS-PAGE analysis of SAG-C proteins overexpression which was tested at 18°C O/N with 1mM IPTG.....	69
Figure 3.16: SAG-C75 purification processes.....	70
Figure 3.17: EBC1 SAG purification processes.....	71
Figure 3.18: EBC2 SAG purification processes.....	72
Figure 3.19: EBC3 SAG purification processes.....	73
Figure 3.20: Grown crystals of SAG-C75 on PH-Clear suite.....	74
Figure 3.21: Crystal of EBC1.....	75
Figure 3.22: SAG-A31 purification process.....	77
Figure 3.23: SAG-A10 purification process.....	78
Figure 3.24: A) SAG-A7 purification process.....	79
Figure 3.25: SAG-A6 purification process.....	80
Figure 3.26: SAG-A870 purification process.....	81
Figure 3.27: SAG-A905 purification process.....	82
Figure 3.28: SAG-A1 purification process.....	83
Figure 3.29: SAG-A91 purification process.....	84
Figure 3.30: Crystals of SAG-A91.....	85
Figure 3.31: Crystal of SAG-A31.....	86
Figure 3.32: Crystal of SAG-A10.....	87

Figure 4.1: Weak electron density of SAG-B13 map.....	92
Figure 4.2: The initial electron density map for SAG-B13.. .....	93
Figure 4.3: SAG-B13 structure.....	95
Figure 4.4: The poor electron density of the SAG-B16 map.....	98
Figure 4.5: The initial electron density map for SAG-B16 in chain B.....	99
Figure 4.6: SAG-B16 structure.....	101
Figure 4.7: The poor electron density map of the SAG-B22 final model.. .....	103
Figure 4.8: The initial electron density map for SAG-B22 in chain A.....	104
Figure 4.9: SAG-B22 structure.....	106
Figure 4.10: The poor electron density map of the SAG-B41 final model.. .....	110
Figure 4.11: The initial electron density map for SAG-B41 in chain B.....	111
Figure 4.12: SAG-B41 structure.....	113
Figure 4.13: The initial electron density map for SAG-A91.....	115
Figure 4.14: SAG-A91 Structure.....	118
Figure 4.15: The initial electron density map for SAG-C75.. .....	120
Figure 4.16: SAG-C75 structure.....	123
Figure 4.17: The <i>E. tenella</i> SAG structure.. .....	124
Figure 5.1: Structure of the five SAG-B members.. .....	127
Figure 5.2: The 2D topology diagram of five SAG-B members.. .....	128
Figure 5.3: Key interactions creating the pocket for the conserved buried arginine.....	130
Figure 5.4: Multiple sequence alignment of the five SAG-B members that were successfully structures determined.. .....	133
Figure 5.5: Diagram to show the variation in the structure of the L4 loop in the five SAG-B proteins.. .....	135
Figure 5.6: The conserved 41 residues across the five SAG-B members.. <b>Error! Bookmark not defined.</b>	
Figure 5.7: Key interaction creating the salt bridges..... <b>Error! Bookmark not defined.</b>	
Figure 5.8: Electrostatic surfaces diagrams of five <i>E. tenella</i> SAG-B members.. .....	139
Figure 5.9: 3D structure of SAG-A91.. .....	140
Figure 5.10: Key interactions creating the pocket for the conserved buried arginine in SAG-A91....	141
Figure 5.11: Superposition of the SAG-A91 and SAG-B19 Structures:.....	143
Figure 5.12: Superposition of SAG-B19 and SAG-A91 conserved disulphides bonds.. .....	144
Figure 5.13: Conserved residues across the 51 members of SAG-A family.. .....	145
Figure 5.14: Schematic diagrams of the structure of SAG-C75 to show the overall fold and the positions of elements of regular secondary structure.....	146
Figure 5.15: Key interactions creating the pocket for the conserved buried arginine in SAG-C75....	147
Figure 5.16: Superposition of the SAG-A91, SAG-B19 and SAG-C75 structures:.....	149
Figure 5.17: Sequence alignment of SAG-B19, SAG-A91 and SAG-C75.. .....	150
Figure 5.18: The position of the conserved residues across the three SAG families.. .....	151
Figure 5.19: Electrostatic surfaces diagrams within the <i>E. tenella</i> family.. .....	152
Figure 6.1: The CAP-like SAG in Trx-construct.....	157
Figure 6.2: Expression of CAP like SAG proteins.. .....	158
Figure 6.3: <i>P. vivax</i> CAP-like SAG Purification process.....	159
Figure 6.4: Crystals of CAP-like SAG in <i>P. vivax</i> SAG.. .....	160
Figure 6.5: The initial electron density map for the <i>P. vivax</i> CAP-like SAG.. .....	164
Figure 6.6: The C-terminus region.. .....	165
Figure 6.7: The N-terminus region.. .....	165
Figure 6.8: <i>P. vivax</i> CAP-like SAG structure.....	168
Figure 6.9: Key interactions creating the pocket for the conserved buried arginine.....	170

Figure 6.10: Key interactions creating the pocket for the conserved buried arginine in Tablysin-15 structure (PDB 3U3U).....	171
Figure 6.11: Phylogenetic tree of <i>Plasmodium spp</i> based on the analysis of sequences of complete mitochondrial genomes.....	173
Figure 6.12: Pattern of conservation on the surface of the <i>P. vivax</i> and <i>P. knowlesi</i> CAP-like SAGs.....	175
Figure 6.13: The positions of sequence conservation and sequence difference across the CAP-like SAGs from the important human plasmodium species.....	175
Figure 6.14: Multiple sequence alignment of the CAP-like SAGs in <i>Plasmodium</i> strains that are important in human disease.. ..	176
Figure 7.1: Multiple CAP-like SAG domains in proteins from different apicomplexan parasites.. ...	180
Figure 7.2: Alignment of CAP-like SAGs from a range of Apicomplexans.....	183
Figure 7.3: Alignment of CAP-like SAGs from a range of Apicomplexans including SAG-B19.....	184
Figure 7.4: The conserved residues between the <i>P. vivax</i> CAP-like SAG and the SAG proteins from a range of apicomplexan parasites.....	184
Figure 7.5: Schematic diagram of helix dipole.....	186
Figure 7.6: the environment of the critical arginine in the CAP fold of SAG-B41.....	189
Figure 7.7: Key interactions creating the pocket for the conserved buried arginine in SAG-B41.....	190
Figure 7.8: Interactions creating the pocket for the conserved buried arginine in the structure of the bacterial protein, BB0689.....	191

## List of Table

Table 1.1: The comparison of the seven <i>Eimeria</i> species that are known to affect chickens involves evaluating their location of infection, degree of harm caused, and the morphology of their oocysts. The table modified from (Burrell et al., 2020).	14
Table 1.2: <i>Eimeria</i> species biology and genomic sag repertoire.	23
Table 2.1: <i>Eimeria</i> SAGs Trx-construct.	38
Table 2.2: CAP-like SAG from different Apicomplexan parasites in trx-construct.	38
Table 3.1: SAG proteins of <i>Eimeria tenella</i> and <i>Eimeria brunetti</i> of different family.	52
Table 3.2: Four SAG-B family proteins were chosen for overexpression.	60
Table 3.3: Details fo333r SAG-C family proteins used in this study.	68
Table 3.4: Overexpression and Purification Summary for Sag-C family proteins.	70
Table 3.5: Shows the similarity of SAG-A members to SAG-B19 with those having the RRL motifs.	76
Table 3.6: Overexpression and Purification Summary for SAG-A family proteins.	77
Table 4.1: Crystallization and data collection overview of successful structure determinations of SAG proteins.	89
Table 4.2: SAG-B13 Data processing statistics.	90
Table 4.3: Final refinement statistics for SAG-B13.	92
Table 4.4: MolProbity analysis of the SAG-B13 final model including the geometry.	94
Table 4.5: SAG-B16 detailed statistics for dataset SAD.	96
Table 4.6: shows the details of last model for SAG-B16 structure.	97
Table 4.7: MolProbity analysis of the SAG-B16 final model including the geometry.	100
Table 4.8: SAG-B22 detailed statistics for dataset SAD.	102
Table 4.9: Shows the details of last good model for SAG-B22 structure.	104
Table 4.10: MolProbity analysis of the SAG-B22 final model including the geometry.	105
Table 4.11: SAG-B41 detailed statistics for dataset SAD.	107
Table 4.12: the merging-statistics data for SAG-B41 crystal.	108
Table 4.13: Final refinement statistics for SAG-B41 structure.	111
Table 4.14: MolProbity analysis of the SAG-B41 final model structure including the geometry.	112
Table 4.15: SAG-A91 detailed statistics for dataset SAD.	114
Table 4.16: Shows the details of last good model for SAG-A91 structure.	116
Table 4.17: MolProbity analysis of the SAG-A91 final model structure including the geometry.	117
Table 4.18: Table 4.12: SAG-C75 detailed statistics for dataset SAD.	119
Table 4.19: Shows the details of last good model for SAG-C75 structure.	121
Table 4.20: MolProbity analysis of the SAG-C75 final model structure including the geometry.	122
Table 4.21: Showed the $\beta$ and $\alpha$ length of residues for each SAG.	124
Table 5.1: Summary of the SAG family members for which a structure has been determined.	125
Table 5.2: Disulphide bridges and cis-peptides in the SAG-B family structures.	129
Table 5.3: Shows the length of the loop region of SAG-B group.	131
Table 5.4: Hydrogen bonds of conserved hydrophilic residues across the five SAG-B family showed in SAG-B19 structure.	135
Table 5.5: Sequence identity and structural similarity between the SAG-B members.	136
Table 5.6: The structural similarity between the structures of the SAG B proteins determined as part of this thesis against SAG-B19 generated by using GESAMT.	137
Table 5.7: Sequence identity between the <i>E. tenella</i> three SAG family.	148
Table 5.8: The melting point of <i>E. tenella</i> SAGs protein including the testing temperature (C°) to the monosaccharides.	154
Table 6.1: Data collection of Merging statistics for CAP-like SAG in <i>P. vivax</i> crystal	161

Table 6.2: Shows the final model refinement of <i>P. vivax</i> CAP-like SAG.....	163
Table 6.3: MolProbity analysis of the final refined model structure of <i>P. vivax</i> CAP-like SAG.....	166
Table 6.4: Similarity between the structure of the <i>P. vivax</i> CAP-like SAG and <i>E. tenella</i> SAG families A, B and C. ....	169
Table 6.5: Hydrogen bonding pattern of the cis-peptide N175-P176 in SAG-C75 compared to the trans-peptide D159-S160 in <i>P. vivax</i> CAP-like SAG.....	169
Table 7.1: The level of identity between an example of the sequences of the CAP-like SAGs in selected apicomplexan parasites and the <i>P. vivax</i> homologue. ....	182
Table 7.2: Frequency of helical boundary residues in 15 proteins* .....	187
Table 7.3: The summary of SAG proteins that were part of this study.....	192
Table 7.4: Key structural features found in the <i>Eimeria tenella</i> SAGs compared to other proteins of the CAP superfamily. ....	193
Table 7.5: Disruption of SAGs in selected apicomplexan parasites.....	194
Table 7.6: The identity (%) between ETH_00002020 and the SAGs studied as part of this thesis. ....	196

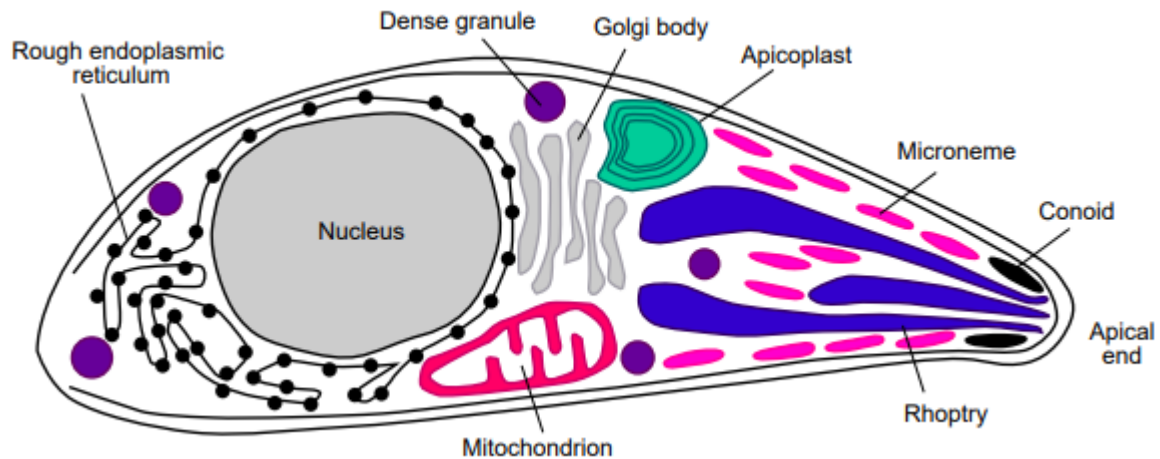
# Chapter1

## 1.1 Introduction

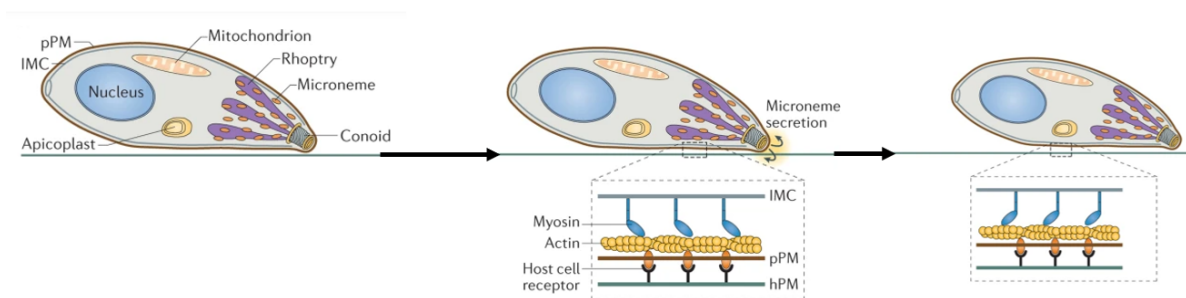
The Apicomplexa is a monophyletic group primarily consisting of parasitic organisms, comprising of 5000 identified species and 350 genera (Cavalier-Smith, 1993). These organisms are responsible for causing severe diseases in both humans and livestock, affecting millions of individuals each year (Portman and Šlapeta, 2014). Apicomplexans are well-known for being intracellular parasites of vertebrates (e.g. mosquitoes or ticks) and are responsible for causing significant human diseases including malaria caused by *Plasmodium*, cryptosporidiosis caused by *Cryptosporidium*, and toxoplasmosis caused by *Toxoplasma gondii* and also animal disease such as coccidiosis in chickens caused by *Eimeria* species. These parasites localise in the host cell through the invasion of specific cell types such as lymphocytes, erythrocytes, macrophages, or cells lining the digestive system (Sam-Yellowe, 1996). For example in the case of the *Plasmodium* the parasites are opportunistic, incapable of reproducing or existing outside of the host, but not (*Toxoplasma* and *Eimeria*) and able to evade host immunity by altering the molecular features of the parasite cell surface (Frénal and Soldati, 2009). For all these diseases treatments are suboptimal, with limited vaccines and increasing drug resistance proving problematic. Increased understanding of how the parasites invade their host cells may lead to improved treatment.

### 1.1.1 Cellular Structure of Apicomplexan Parasites

Apicomplexans share common features including a distinct cytoskeleton (Morrisette and Sibley, 2002), polarized secretory organelles, and a unique form of movement called “gliding” motility [Figure 1.1] (Tilney *et al.*, 2004). In contrast to other modes of cell mobility like cilia, flagella-based propulsion, or crawling, gliding works by translocating adhesive proteins along the parasite cell surface similar to cargo moving along a conveyor belt. Rather than relying on appendages or changes in cell shape, apicomplexans use this conveyor belt-like mechanism to propel themselves forward secretion by rearward translation of adhesion molecules from apical organelles followed by their engagement with receptors on the substrate to generate traction. This gliding motility [Figure 1.2], based on polarized secretion of adhesins sets apicomplexans apart from other motile eukaryotic cells and represents a defining biosynthetic and mechanical feature of these parasites [Figure 1.1] (Tilney *et al.*, 2004).



**Figure 1.1: Apicomplexan Morphology:** The unique morphology of apicomplexan parasites encompasses fundamental eukaryotic organelles such as the nucleus, Golgi body, endoplasmic reticulum, mitochondrion, and apicoplast. Moreover, they possess three vital secretory organelles – micronemes, rhoptries, and dense granules – crucial for accomplishing successful invasion and creating the parasitophorous vacuole. Adapted from (Ajioka, Fitzpatrick and Reitter, 2001).



**Figure 1.2: Apicomplexa gliding motility.** The gliding motility involves microneme organelles releasing at the parasite's apical end, where their transmembrane proteins merge into the parasite's plasma membrane (pPM) and bind to host cell receptors. This action drives gliding motility through the rearward movement of adhesin-receptor complexes as shows on black arrow, powered by myosin motors moving along actin filaments between the inner membrane complex (IMC) and pPM. Adapted and modified from (Frénal et al., 2017).



### **1.1.1.1 The unique Apical Complex**

The Apicomplexa phylum is characterised by a structural feature named the apical complex [Figure 1.3] (Levine, 1986). The apical complex is the key feature of apicomplexan parasites that enables their obligate intracellular infection cycle. In some of these parasites it has unique secretory compartments called rhoptries, micronemes, and dense granules that help the parasite connect to the host cell and invade it. They form the parasitophorous vacuole with unique intracellular compartment dividing the parasite from the host cell cytoplasm. For example, in *Toxoplasma* this invasion is a quick process that typically lasts less than 30 seconds (Gilson and Crabb, 2009). The development of the movable junction between the membranes of the host cell and the parasite is a crucial step in this process. This junction acts as an anchor, withstanding shear stresses during entry, and it additionally supplies the required traction for the actomyosin-driven gliding motility that propels the parasite inward (Aikawa *et al.*, 1978). The moving junction arises from the coordinated release of microneme and rhoptry proteins to assemble this intricate anchor complex. Overall, the polarized apical secretory organelles and proteins enable these parasites to actively invade host cells and establish an intracellular compartment favourable for replication (Frénal *et al.*, 2017).

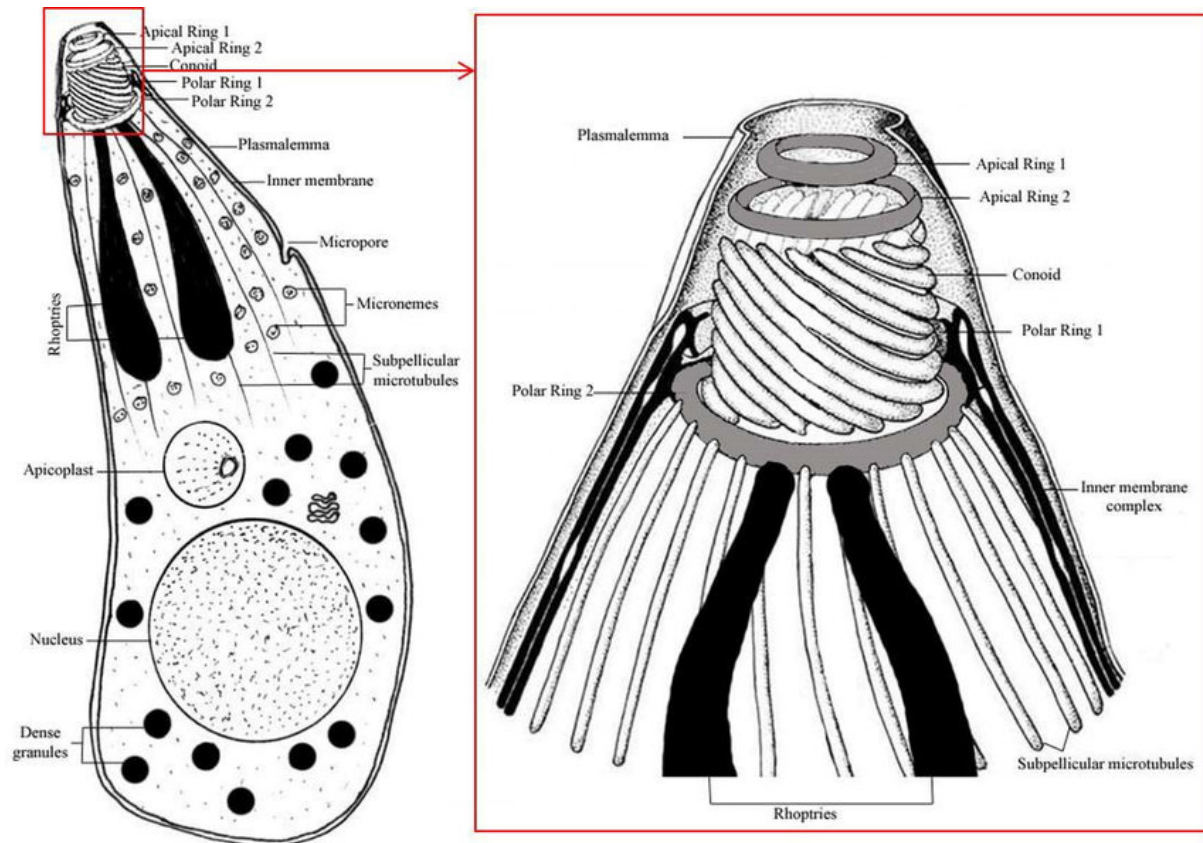


Figure 1.3: *The general structure for the apical complex for Apicomplexan. Adapted from (Burrell et al., 2022).*

### 1.1.1.2 Intricate Membrane Structure

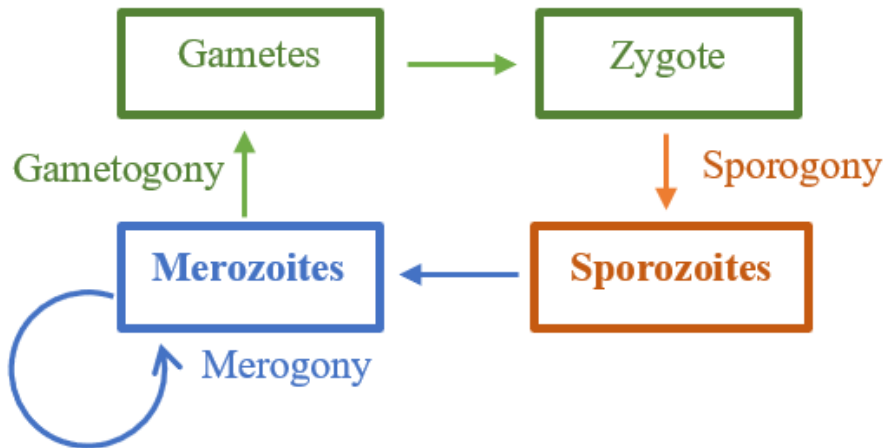
The intricate membrane structures in apicomplexan parasites play essential roles in their survival, host cell invasion, and immune evasion (Harding and Frischknecht, 2020). A distinctive pellicle structure made up of the plasma membrane and a closely apposed inner membrane complex (IMC) surrounds these parasites (Morrisette and Sibley, 2002). The IMC consists of flattened alveolar vesicles arranged in an almost continuous layer underlying the plasma membrane. This alveolar matrix is supported by a submembrane protein skeleton that maintains cell shape and strength (Koreny *et al.*, 2023). Additionally, the IMC serves as a scaffold for the parasite's actomyosin-based gliding motility apparatus, enabling tissue migration and host cell penetration (Harding and Frischknecht, 2020).

### **1.1.2 Lifecycle of Apicomplexan Parasites**

The complex lifecycle of apicomplexan parasites involves both sexual and asexual stages (Striepen *et al.*, 2007). Both vertebrates and invertebrates are susceptible to parasitic infection (West *et al.*, 2000). Some genera of apicomplexans, such as *Plasmodium*, require a change in the host species to complete their lifecycles (Young *et al.*, 2005). Upon invasion, the parasite resides within a parasitophorous vacuole inside the host cell, where replication begins. The newly replicated parasite needs to invade new host cells to survive.

#### **1.1.2.1 A complex Multi-stage Lifecycle**

The Apicomplexa have intricate life cycles involving sporogony, merogony, and gamogony phases with both sexual and asexual reproduction. This involves transformations between sporozoites, merozoites, gametes, and zygotes [Figure 1.4] (Smith *et al.*, 2002). For example, in the case of *Eimeria*, infection begins when sporozoites from an oocyst invade host cells using their apical end ingestion or insect vectors. The sporozoite becomes a trophozoite and undergoes merogony, generating merozoites which infect more host cells (Gubbels *et al.*, 2020). After successive rounds of replication, some merozoites become female macrogametocytes and most become male microgametocytes. Fusion of the motile microgamete and stationary macrogamete forms a zygote. The zygote undergoes asexual reproduction, dividing to produce new sporozoites (Guttery *et al.*, 2015). Transmitting sporozoites to another host completes the complex developmental cycle (Gubbels *et al.*, 2020).



*Figure 1.4: Some of Apicomplexan (Eimeria) Life Cycle. Apicomplexan parasites have complex life cycles that involve different stages. Gametocytes and gametes are produced by gamogony, which combine to form a zygote, which is the developmental stage of apicomplexan. Sporozoites grow inside a resistant thick-walled oocyst during sporogony. By a process called merogony or schizogony, sporozoites can grow into merozoites, giving rise to a large number of infectious merozoites.*

### 1.1.3 Host-Pathogen Interactions

These protozoan parasites have the common trait of having an apical complex made up of specialised secretory organelles. These organelles play a crucial role in the process of invading host cells (Dubremetz *et al.*, 1998). Despite infecting various types of host cells, these parasites shared a common mechanism for invading the host cell (Sibley, 2010). This mechanism involves the sequential release of proteins from micronemes and rhoptries, enabling parasite motility, promoting tight adhesion to the target cell, and facilitating the active penetration of the host cell (Sibley, 2010). The parasite replicates inside a membrane-bound compartment called the parasitophorous vacuole (PV) that is enclosed by the host cell membrane during invasion (Shen and Sibley, 2012). As the host cell invasion progresses, these parasites establish the moving junction (MJ) [Figure 1.5].

#### 1.1.3.1 Microneme Proteins: Key Players in Adhesion and Invasion

The micronemes, located at the apical end of the parasite, are the smallest secretory organelles [Figure 1.5 (2)]. Within them, proteins are stored that, once the parasites have attached to the host, release their contents onto the parasites' surface where they can bind to host cell receptors

and invade. The actinomyosin motor causes anterior to posterior movement of the ligand-receptor connections, resulting in motility that drives invasion and promotes movement through tissues and on solid substrate. The parasite ligands are shed from the surface as a result of proteases that are embedded in the basal end of the plasma membrane (O'Donnell *et al.*, 2006; Dowse *et al.*, 2005). Within these organelles, numerous adhesins that exhibit specific binding capabilities to host cell receptors are found. The evolution of these proteins has taken distinct paths in various species, driven by their co-evolution with diverse host cell receptors. For example, the apical membrane antigen 1 (AMA-1) protein (involved in the invasion process), which is thought to trigger the release of rhoptries, is one of the rare proteins that is shared among all apicomplexans. Whereas, in *Toxoplasma* perforin-like protein (TgPLP1) secreted by the micronemes plays a crucial role in facilitating egress, highlighting that these organelles are not limited to invasion but also contribute significantly to the egress process (Carruthers and Tomley, 2008; Kafsack *et al.*, 2009).

#### **1.1.3.2 Rhoptry Secretory Organelles: Delivering Effectors into Host Cells**

The rhoptries are larger than micronemes and are specialised organelles involved in secretion [Figure 1.5 (4)]. They are club-shaped, with one end linked to the parasite's apical region (Bradley and Sibley, 2007). These organelles share similarities with secretory lysosomal organelles and play a pivotal role in shaping the parasitophorous vacuolar membrane (PVM). The release of rhoptry contents occurs subsequent to microneme release and coincides with a close interaction between the parasite and host cell membranes. In certain species, rhoptries also house lamellar membranes that contribute to the formation of the PVM. Recent studies indicate that the rhoptry neck (RON) and rhoptry bulb (ROP) represent distinct compartments, each containing different sets of proteins, and they are released at different stages. For instance, the RON proteins stored in the rhoptry neck are essential for establishing a tight junction between the parasite and the host cell, and they are secreted prior to the ROP proteins, which modify both the vacuolar membrane and the host cell (Ben Chaabene *et al.*, 2021).

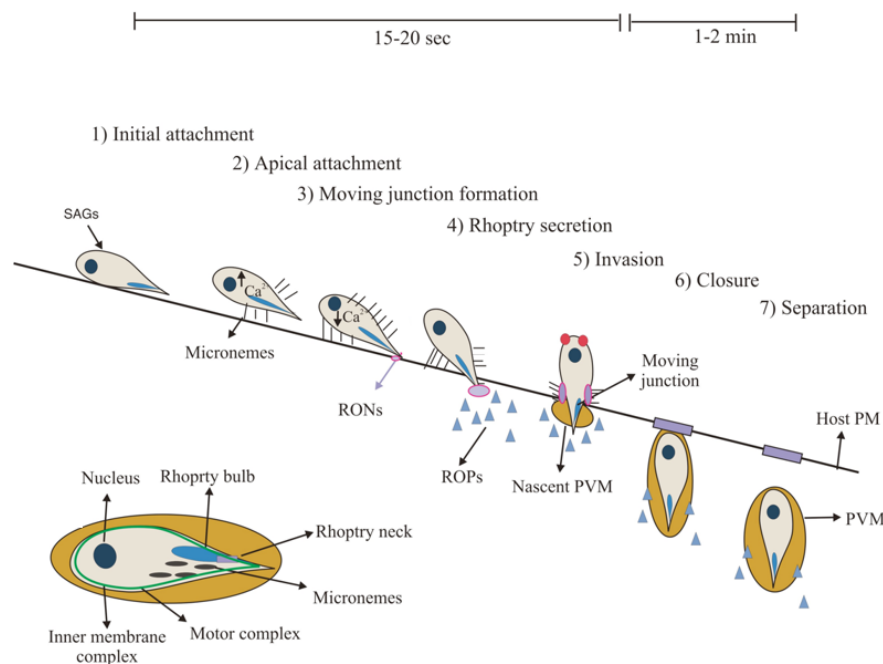
#### **1.1.3.3 Dense Granules: Role in Parasite Survival and Host Immune Evasion**

The other secretory organelles in these parasites are called dense granules or dense bodies. They are microspheres surrounded by a unique membrane and have a diameter of approximately 200 nm (Karsten *et al.*, 1998). In contrast to micronemes and rhoptries, dense granules are not concentrated at the apical end, instead, they are dispersed throughout the cell. These granules are released promptly and continuously following invasion and throughout the process of intracellular replication (Ravindran and Boothroyd, 2008). The number of dense granules varies among different parasites. For example, *Eimeria* don't have dense granules, while *Toxoplasma* and *Neospora* have around twenty (Rommereim *et al.*, 2016). Once the parasite is internalized in the vacuole, these granules are released into the vacuolar space. These granules are believed to contribute to the development of the parasitophorous vacuole, transforming it into a metabolically active compartment. Recent evidence suggests that certain effectors from these organelles are necessary for modifying the host (Ben Chaabene *et al.*, 2021; Tobin and Knoll, 2012).

#### **1.1.3.4 Surface Proteins: Modulators of Host Recognition and Immune Response**

A number of surface protein on the parasite facilitate attachment to the host and can also be recognised by the host immune system. These proteins recognize and interact with host cell ligands before the invasion takes place. Therefore, they have the potential to serve as effective targets for immunoprophylaxis. Most of these surface proteins are anchored to the parasite surface using a GPI anchor (Liu *et al.*, 2016). It is noteworthy that GPIs found in related parasites have a significant impact on modulating the host immune system (Tachado *et al.*, 1999). For example in the case of *Eimeria*, when host Toll-like receptors 2 and/or 4 recognize GPI anchored surface proteins in malaria parasites, a potent pro-inflammatory cytokine response from host cells *in vitro* is triggered (Gowda, 2007). Whilst, in the case of *Toxoplasma gondii*, activation of the same receptors by GPI anchored proteins appears to be crucial for initiating an innate immune response (Debierre-Grockiego *et al.*, 2007). Furthermore,

mounting evidence supports the role of GPI proteins in evading the host immune system in related parasites (Delbecq, 2022).



**Figure 1.5: The invasion process of Apicomplexan parasites.** (1) There's reversible attachment to host surface receptors mediated by surface antigen, (2) apical attachment involving the deployment of micronemes. (3) The formation of the moving junction (MJ). (4) Rhoptry started to release into the host cytoplasm. (5) The parasite actively penetrates by pulling transmembrane, (invaginating the host plasma membrane to create the parasitophorous vacuole). (6 and 7) Closure and separation, the final steps, are rate-limiting and involve PV and host membrane fission. The figure adapted from (Carruthers and Boothroyd, 2007).

## 1.1.4 Disease of Apicomplexan Parasites

### 1.1.4.1 Toxoplasmosis Disease: Zoonotic Infections and Human Health

The parasite responsible for causing toxoplasmosis is *Toxoplasma gondii*, a coccidian parasite that can infect various tissues in both mammalian and avian species (Lekutis *et al.*, 2001). *T. gondii* infection poses a significant risk to pregnant women, sheep and immunocompromised individuals, such as those with AIDS, organ transplant recipients, or neoplastic diseases (Jacquet *et al.*, 2001; Wu *et al.*, 2009).

Due to its flexible life cycle, *T. gondii* is a common parasite that can infect hosts through oocysts shed in cat feces or tissue cysts found in meat. Oocysts are transmitted through the ingestion of contaminated food or water containing sporulated oocysts (Grigg and Sundar, 2009). With the exception of desiccation and extreme heat, oocysts have a high percentage of environmental survival and are resistant to chemical and environmental destruction. After undergoing excystation in the gut lumen, oocysts release sporozoites, which proliferate and develop into infectious tachyzoites within the lamina propria's enterocytes and myeloid cells. [Figure 1.6] (Grigg and Sundar, 2009). The other mode of transmission involves the ingestion of tissue cysts from raw or undercooked meat, wherein bradyzoites are present. After ingestion, bradyzoites quickly differentiate into tachyzoites, enabling dissemination and subsequent production of bradyzoites before host immunity is induced (Grigg and Sundar, 2009). The production of bradyzoites at this stage is crucial for the parasite's survival within the host and its resistance to host immunity. Only bradyzoites are transmissible at this stage, highlighting the importance of interconversion between bradyzoites and tachyzoites for the parasite's survival (Grigg and Sundar, 2009).

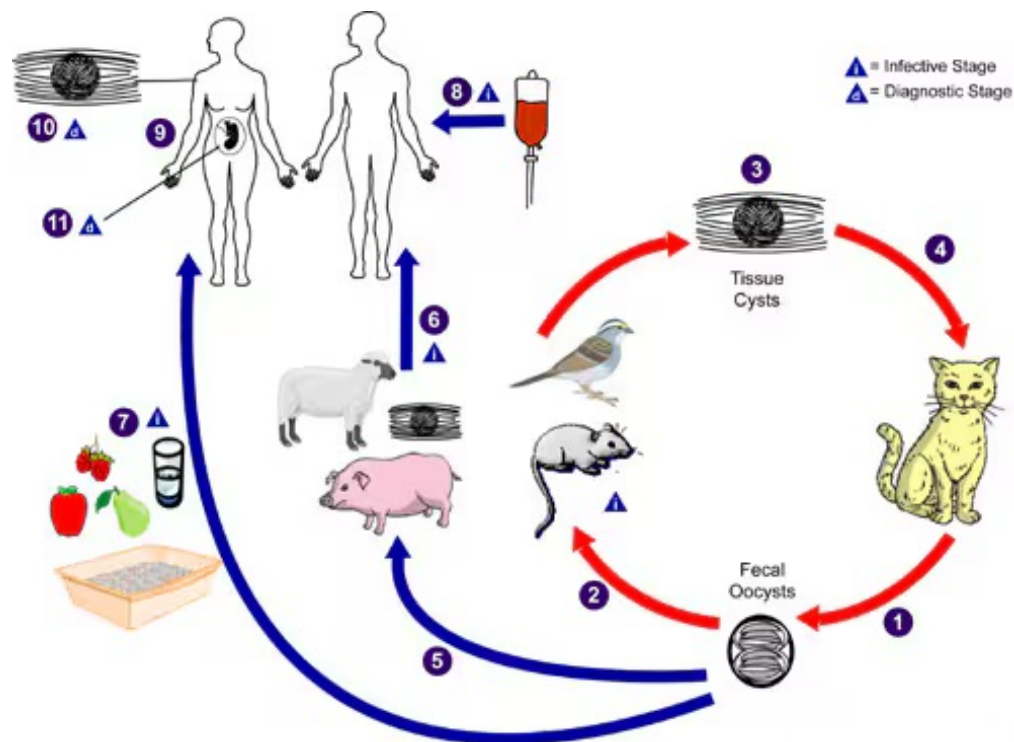


Figure 1.6: **The life cycle of *Toxoplasma gondii*.** The schematic shows the crucial involvement of bradyzoites in transmitting the parasite between intermediate and definitive hosts. Members of the

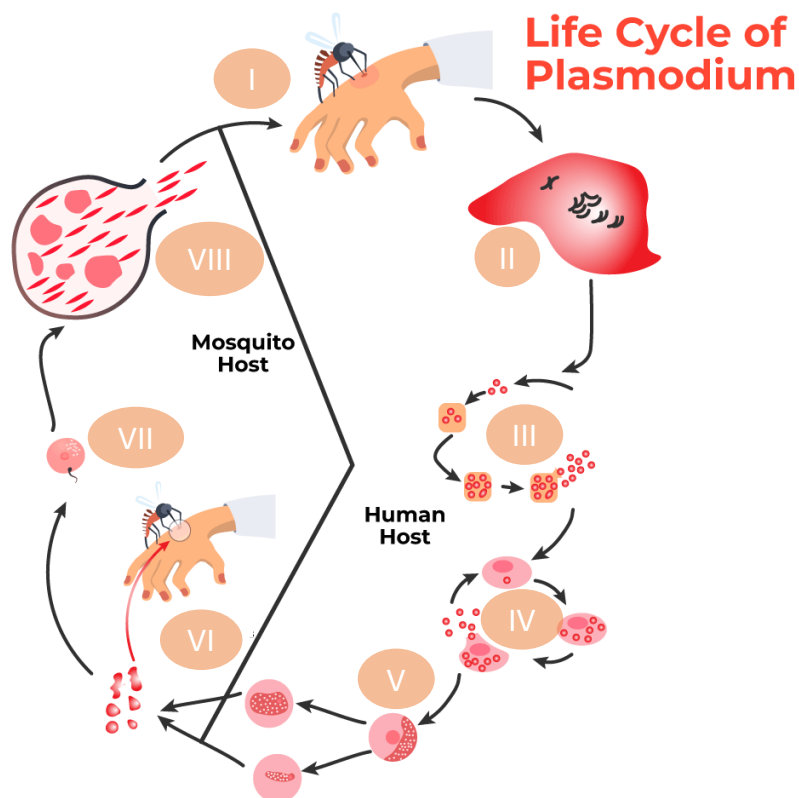


*family Felidae (domestic cats and their relatives) are the only known definitive hosts for Toxoplasma gondii. Cats shed unsporulated oocysts in their feces for 1–3 weeks, but in large quantities. These oocysts become infective after 1–5 days of sporulation in the environment. Intermediate hosts, such as birds and rodents, become infected by ingesting soil, water, or plant material contaminated with oocysts. After ingestion, oocysts transform into tachyzoites, which localize in neural and muscle tissue and develop into tissue cyst bradyzoites. Cats then become infected by consuming these intermediate hosts with tissue cysts or directly by ingesting sporulated oocysts. Animals raised for human consumption and wild game may also become infected by ingesting sporulated oocysts from the environment. In humans, T. gondii forms tissue cysts primarily in skeletal muscle, the heart, brain, and eyes, and these cysts can persist for life, (<https://www.cdc.gov/dpdx/toxoplasmosis/index.html>).*

#### **1.1.4.2 Malaria Disease: A Global Health Challenge**

*Plasmodium* is a member of the apicomplexan phylum and is the causative agent of malaria. Each year, between 300 and 500 million new cases of malaria occur, resulting in 1-3 million deaths (Mongui *et al.*, 2010). There are five species of *Plasmodium* that infect humans, namely *P. falciparum*, *P. vivax*, *P. ovale*, *P. malariae*, and *P. knowlesi* (Oddoux *et al.*, 2011). *P. falciparum* is responsible for the high mortality rates associated with malaria in sub-Saharan Africa (Doolan, 2011; Kerlin *et al.*, 2012). Conversely, *P. vivax*, although less deadly, is more prevalent in subtropical and tropical regions, with 81% of cases recorded in Central America and the East Mediterranean (Kerlin *et al.*, 2012). Numerous factors contribute to the clinical manifestation of malaria, including the genetics of the human host, the age of the host, and the transmission dynamics of the parasite (Doolan, 2011). In *Plasmodium*, a complex life cycle involving an insect vector (the female mosquito Anopheles) and a vertebrate (humans/other host such Monkey) is observed (Doolan, 2011; Moreira *et al.*, 2004). *Plasmodium* undergoes three different asexual replicative stages, namely exoerythrocytic schizogony, schizogony of the blood stage, and sporogony, which involve the merozoites and sporozoites (infected forms). Sexual reproduction occurs by switching between vertebrate and invertebrate hosts, generating

invasive ookinetes, which are embryonic forms produced from gametocytes [Figure 1.7] (Ghosh *et al.*, 2000). *Plasmodium* poses a significant global health issue, and the development of a vaccine is challenging due to the existence of widespread gene polymorphism. Additionally, the growth of insecticide resistance, inadequate health systems, and the lack of sustainable infrastructure in affected countries contribute to the impact of the disease (Ahoudi *et al.*, 2010).



**Figure 1.7: Life cycle of *Plasmodium* parasites.** I, (sporogony) when a mosquito bites a human, it injects sporozoites into the bloodstream; II, (multiplication of sporozoites) these sporozoites then travel to the liver; III, (merozoites) where they reproduce asexually within liver cells, eventually bursting the cells and entering the bloodstream again; IV, (split and producing more of merozoites) inside the bloodstream, the parasites continue to replicate asexually within red blood cells, leading to the bursting of these cells and the onset of fever and other symptoms. The released parasites then infect new red blood cells; V, (gametogony), sexual stages known as gametocytes develop within these infected red blood cells; VI, (microgametocytes and macrogametocytes) when a female mosquito feeds on the infected blood, it ingests these gametocytes; VII, (zoites) within the mosquito's intestine, fertilization occurs, and the parasites develop further; VIII, (oocytes) mature infective stages, now called sporozoites, escape from the intestine and migrate to the mosquito's salivary glands, ready to be transmitted to another human through a subsequent bite. Figure adapted from (Geeks, 2024).

#### **1.1.4.3 Coccidiosis Disease: Impact on Livestock and Poultry Industry**

Chicken coccidiosis, a prevalent enteric disease caused by several species of *Eimeria* (*more than ten species*), poses significant economic losses to the global poultry industry. The estimated global impact of this disease in 2016 was approximately €12 billion (Blake *et al.*, 2020). While there have been extensive studies on the economic impact and epidemiology of coccidiosis in intensive poultry production systems, there is limited knowledge about its effects on small-scale productions such as family household or backyard poultry (Blake *et al.*, 2020). Coccidiosis refers to a gastrointestinal disease that results from being infected with one or multiple kinds of *Eimeria* species (Chapman, 1998).








Broiler chickens that are affected by coccidiosis typically experience diarrhea that can range from mucoid and watery to haemorrhagic. This disease can also cause a reduction in weight gain and feed intake, and in severe cases, can lead to death. Although most chickens will become infected with coccidia at some point in their lives, only a small number will display clinical signs of coccidiosis (Lang *et al.*, 2009). Clinical signs of coccidiosis usually appear in young animals but can sometimes affect adult chickens as well. Vaccination strategies are based on the principle that a low-level exposure to *Eimeria* can activate the chicken's immune response and provide protection against future infections. It is generally believed that coccidiosis only occurs when the chicken is exposed to a high level of infectious agents (Soutter *et al.*, 2020).

#### **1.1.5 *Eimeria* parasite**

*Eimeria* parasites are closely related to other coccidian genera, including *Cyclospora*, *Toxoplasma*, and *Neospora*. They are more distantly related to *Cryptosporidium* and *Plasmodium*. There are more than 1500 known species of *Eimeria*, and they are considered obligatory intracellular parasites that can only infect certain hosts and tissues. *Eimeria* is generally invade and develop within epithelial cells of the intestinal tract of chicken (Dubey, 2019).

There are seven species that can affect chickens, namely *E. acervulina*, *E. brunetti*, *E. maxima*, *E. mitis*, *E. necatrix*, *E. praecox*, and *E. tenella*. The severity of infections can vary depending on factors such as the specific parasite species, infective doses, immune status of the chickens, and breed line. Among these chicken *Eimeria* species can be divided into two groups: those causing haemorrhagic disease (*E. brunetti*, *E. necatrix*, and *E. tenella*) and those causing primarily malabsorption (*E. acervulina*, *E. maxima*, *E. mitis*, and *E. praecox*) (Rochell *et al.*, 2016). Interestingly, each of these *Eimeria* species exhibits variations in terms of their preference for infecting specific sections of the chicken's gut. For example, *E. maxima* is primarily found in the jejunum and ileum, *E. acervulina* predominantly infects the duodenum, and *E. tenella* is typically found in the caeca [Table 1.1] (Macdonald *et al.*, 2017).

*Table 1.1: The comparison of the seven Eimeria species that are known to affect chickens involves evaluating their location of infection, degree of harm caused, and the morphology of their oocysts. The table modified from (Burrell et al., 2020).*

Species	Site of pathology	Lesion score	Oocyst morphology Z.range (15-30um)
<i>E. tenella</i>	Caeca 	High	Medium round
<i>E. maxima</i>	Mid small intestine 	Medium	Large ovel
<i>E. acervulina</i>	Upper small intestine 	Medium	Small ovel
<i>E. necatrix</i>	Mid small intestine 	High	Small-medium round
<i>E. brunetti</i>	Distal small intestine + colon 	High	Medium ovel
<i>E. mitis</i>	Upper small intestine 	Low	Small round
<i>E. praecox</i>	Upper small intestine 	Low	Medium round

### 1.1.5.1 *Eimeria* Genomic Structure: Insights into Parasite Evolution

The structure of *Eimeria* genomes is believed to comprise a nuclear genome made up of 15 chromosomes with total length of 53.25 Mb (Reid *et al.*, 2021), a mitochondrial genome of approximately 6200 bp (Morgan and Godwin, 2017), and a circular apicoplast genome of approximately 35 kb (Cai *et al.*, 2003). Compared to many other eukaryotes, the process of karyotyping *Eimeria* genomes has been slow due to difficulties in accessing and analysing the contents of the cell, primarily due to the extreme mechanical resistance of the oocyst wall. Breaking down this wall typically required significant mechanical force, which often resulted in the disruption of cell contents (Blake *et al.*, 2020).

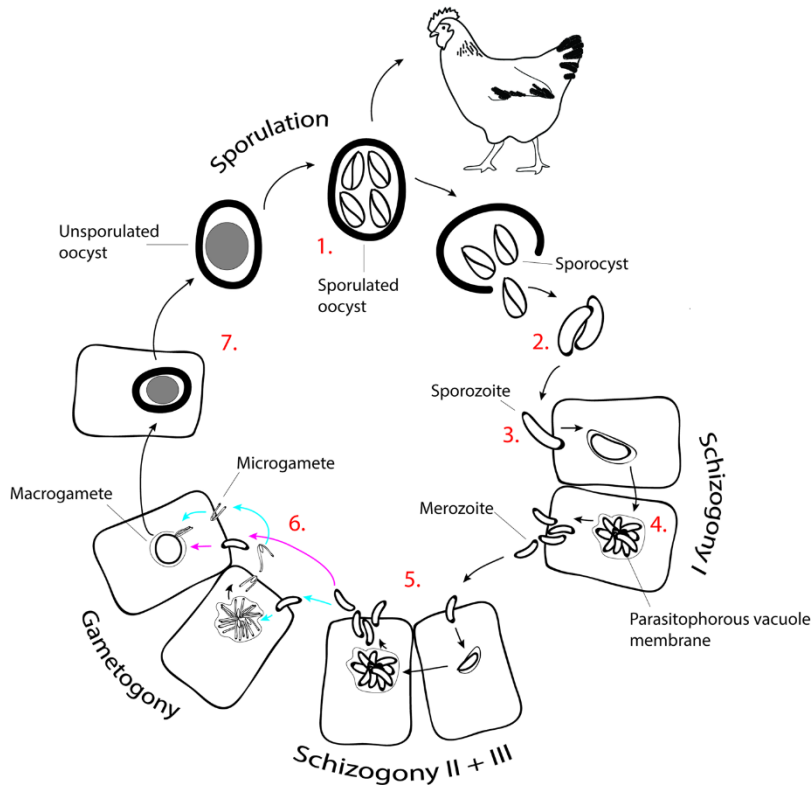
All sequenced *Eimeria* genomes share a segmented chromosome structure that includes both repeat-rich (R) and repeat-poor (P) regions (Heitlinger *et al.*, 2014; Ling *et al.*, 2007; Reid *et al.*, 2014). The most frequent repeat sequence observed in these genomes is the trinucleotide CAG, which is found throughout the genome, including protein-coding regions (Ling *et al.*, 2007; Reid *et al.*, 2014). These CAG repeats result in homopolymeric amino acid repeats (HAARs) that do not appear to have any impact on protein structure or function. Although other repeats such as the heptamer AAACCCT/AGGGTTT and fragmented retrotransposon-like elements similar to chromovirus long-terminal repeat (LTR) retrotransposons are also present, they are not located within coding sequences where they could potentially disrupt coding frames. The high degree of sequence degeneracy observed in these retrotransposons suggests that they are probably not functional. However, the analysis of these elements has been hampered by the quality of the sequence assemblies. Although repeat types are well conserved in sequenced *Eimeria* species, the frequency and location of these repeats vary among species and may even differ among strains within the same species (Shirley, 1994; Reid *et al.*, 2014). The variations in homologous chromosome sizes among different strains of *E. tenella* have been observed (Shirley, 1994), and researchers have proposed that these differences in size could be attributed to variations in the number of CAG repeats. The exact function of these repeats is still unclear, and they have not been linked to specific genes or gene families. However, it has been suggested that these repeats may influence recombination levels and contribute to genome evolution, potentially playing a role in the observed variations (Reid *et al.*, 2014; Ling *et al.*, 2007).

### 1.1.5.2 Life Cycle of *Eimeria*: Understanding Reproductive Strategies

The infection of *Eimeria* begins when sporocysts, released from oocysts, are exposed to gastrointestinal enzymes, such as trypsin, and bile salts. This releases sporozoites, which then invade epithelial cells. The oocysts have a glycoprotein and lipid wall, which in birds is broken down in the grinding process of the ventriculus (gizzard), whereas in non-avian species the wall is degraded by carbon dioxide and stomach acid (Blake *et al.*, 2020). Once the sporozoites have invaded the host cell, they undergo two to four rounds of schizogony, an asexual replication process that results in large numbers of merozoites and egress and re-invasion between each round of replication. These go on to infect other cells and further schizogonic activity, or they engage in sexual replication, producing male and female gametes (microgamonts and macrogamonts). Microgamonts contain into microgametes capable of fertilising macrogametes to produce zygotes; these develop into oocysts. Within the *Eimeria* life cycle all developmental stages are haploid, except for a diploid zygote stage [ Figure 1.8] (Reid *et al.*, 2014; Blake *et al.*, 2011). When the environmental conditions of humidity, temperature and oxygen content are conducive, haploid sporozoites are produced from oocysts typically released in faeces; the sporozoites are the products of rounds of meiosis and mitosis (sporulation) (Shirley and Harvey, 1996). Oocysts are very robust, able to survive for long durations in various environmental conditions. Research into the conditions in commercial broiler houses found *Eimeria* oocysts remain viable in the environment for several weeks between different sets of broilers, making the disease particularly difficult to control. Previous batches of chickens indirectly pass the infection on to the next batch of chickens, thereby continuing the infection in an open-ended manner (Jenkins *et al.*, 2019).

Compared to the lifecycle of other apicomplexans, those of most *Eimeria* genera are quite quick, typically less than one week (Morris and Gasser, 2006). Because there is no *in vivo* cyst phase, the parasite is unable to persist within the host, implying the parasite's life cycle adopts a "hit and run" lifecycle. It is noteworthy that migratory birds, for example, cranes (infected by *E. gruis* and *E. reichenowi*) and corncrakes (infected by *E. crecis* and *E. nenei*) can experience enduring disseminated visceral coccidiosis (Jeanes *et al.*, 2013). The species of *Eimeria* that infect these avian species seem to differ phylogenetically from the species of *Eimeria* that infect non-migratory birds and mammals (Matsubayashi *et al.*, 2005). More research and sampling of diverse *Eimeria* genomes is needed to elucidate the taxonomy relationships

between different species and genera, such as the closely related *Cyclospora* and *Cystoisospora* (Liu *et al.*, 2016; Palmieri *et al.*, 2017).



**Figure 1.8: *Eimeria* Life Cycle.** (1) Sporulation of the oocyst takes place in the environment, followed by the chicken consuming it orally. (2) During its passage through the chicken's digestive system, the sporocysts and sporozoites are released. (3) The sporozoites actively invade the epithelium of the ceca (*E. tenella*) and form the trophozoite intracellularly within the parasitophorous vacuole. (4) The first round of schizogony takes place, and the first-generation merozoites are released. (5) The second and third rounds of schizogony occur, and the second and third generation merozoites are released, respectively. (6) The microgametes and macrogametes undergo gametogony and fecundation. (7) The zygote develops into the oocyst, which is then released into the environment as an unsporulated oocyst. The figure adapted and modified from (Burrell *et al.*, 2020).

### 1.1.5.3 Chicken Immunity Against *Eimeria*: Host Defence Mechanisms

The pathogenicity of *Eimeria* species depends on their ability to successfully replicate within the host intestine. Even a single oocyst can yield millions of invasive parasites through multiple generations of schizogony (Levine, 1986). *Eimeria* invasion damages the epithelial lining of the intestine, causing inflammation, hemorrhage, and diarrhea. Younger animals are more susceptible and display more severe disease due to exposure, whereas older chickens have



some resistance as result of susceptible if not previously exposed (Dalloul *et al.*, 2007). Host genetics also influence disease severity, as evidenced by differing outcomes in inbred chicken strains infected with *E. tenella* or *E. acervulina* (Lillehoj, 1998). Infection with *Eimeria* induces long-lasting and highly specific protective immunity against the particular *Eimeria* species involved. While a substantial number of oocysts is typically needed to trigger a robust immune response, exceptions exist, such as *E. maxima*, which is highly immunogenic and requires only a small number of oocysts to induce almost complete immunity (Lillehoj, 2016). Host immunity mainly targets the early endogenous stages of the *Eimeria* life cycle, particularly the exponential growth phase, inhibiting intracellular development (Lillehoj, 2016).

*E. tenella* infection triggers both humoral and cellular immune responses in chickens, although antibody-mediated immunity appears to have a minor role in natural *Eimeria* infections (Ritzi *et al.*, 2016). Nonetheless, several studies have indicated that antibodies raised against *Eimeria* proteins, whether administered parenterally or transferred to hatching chickens through maternal immunization, can provide partial defense against coccidiosis (Ritzi *et al.*, 2016). Therefore, there is potential for using a combination of immunogens that induce high-titer antibody responses to offer maximum protection against *Eimeria* infection (Wallach, 2010).

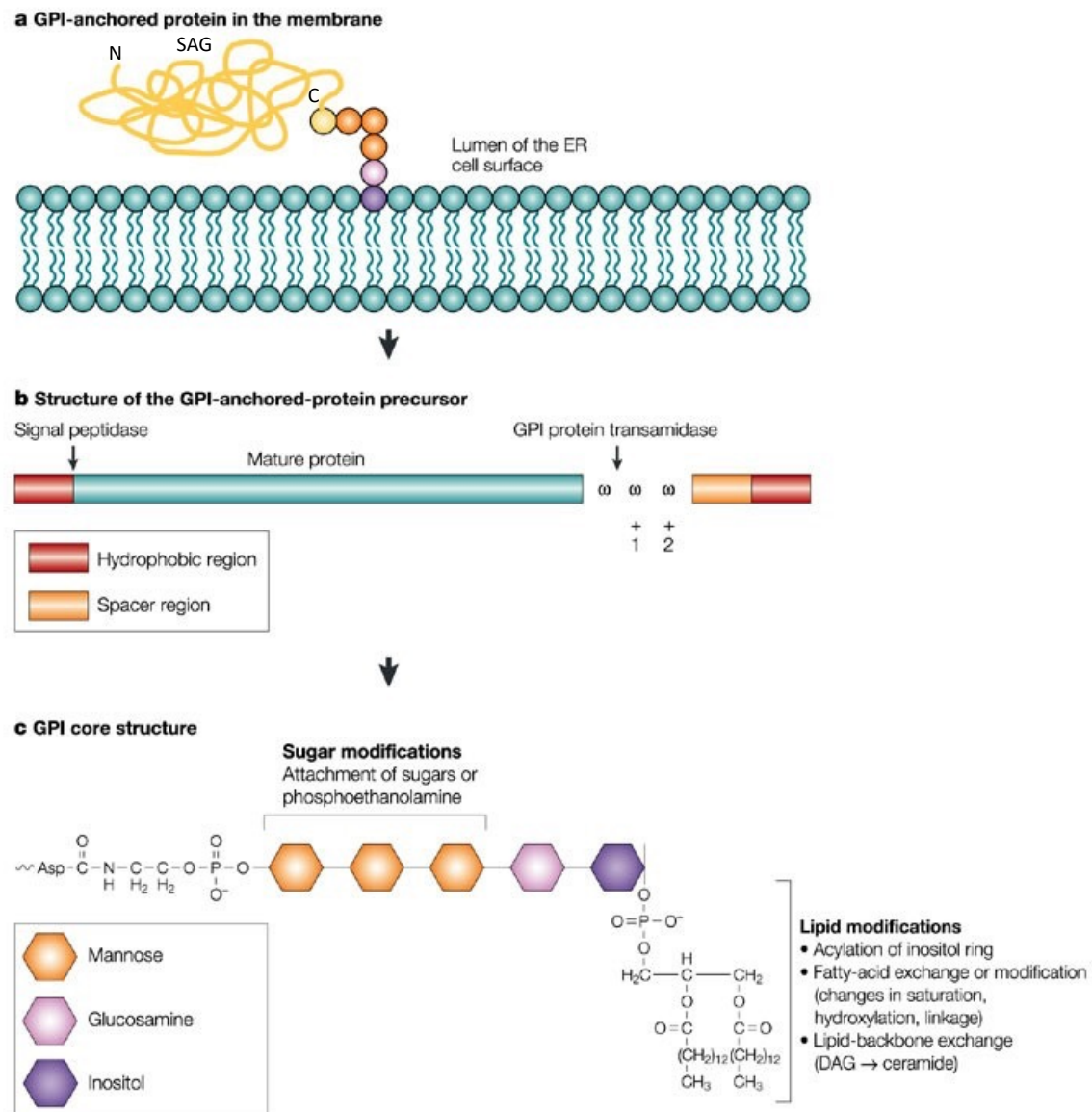
In *E. tenella* the disease primarily occurs due to the rapid proliferation of second-generation schizonts within crypt epithelial cells, which migrate deep into the lamina propria. Subsequently, the rupture of these schizont cells leads to the release of second-generation merozoites (Zhou *et al.*, 2020). Following *E. tenella* infection in chickens, there is a significant infiltration of macrophages into the caecal lamina propria on the first day, accompanied by the secretion of large quantities of cytokines (Dalloul *et al.*, 2007).

#### **1.1.5.4 *Eimeria*-Specific Surface Antigens (SAGs) Play Critical Roles in Host-Parasite Interactions and Pathogenicity**

In the *Eimeria* genus, within the invasion process, there are proteins localized on the surface of the parasite cell membrane, which are identified as GPI-anchored surface antigens (SAGs) [Figure 1.9]. These SAG proteins are found on invasive forms of the parasite such as sporozoites and merozoites, and they interact with the host prior to invasion. Although there is limited information and an unclear biological function associated with these proteins, they are believed to play a significant role in the interaction between pathogens and host cells (Spence *et al.*, 2013). These proteins are exposed to the host's immune system and have the potential to

be targeted for the development of protective immunity (Heitlinger *et al.*, 2014). There are three subfamilies within the SAG gene family (SAG-A, SAG-B, and SAG-C). The SAG-A family are present in all species of *Eimeria*, while the SAG-B family is found exclusively in *E. tenella* and *E. necatrix*. The SAG-C family is limited to the remaining species, with mostly presence in *E. brunetti* and *E. mitis* [Table 1.2] (Reid *et al.*, 2014). All families have genes that encode signal peptides and sites for GPI anchors, the extracellular domain of SAG-C only contains four conserved cysteines, whereas SAG-A and SAG-B have six. The SAG-B and SAG-C genes consist of five exons each, indicating a closer relationship between them compared to the SAG-A genes, which have four exons. The presence of SAG-A genes in all *Eimeria* species suggests they perform a fundamental function, while SAG-B and SAG-C genes may have specific functions related to different clades within the genus (Reid *et al.*, 2014).

*E. tenella*, *E. necatrix*, and *E. brunetti*, which cause deeper tissue damage, inflammation, and intestinal hemorrhage, possess 89, 119, and 105 SAG genes, respectively (Reid *et al.*, 2014). *E. mitis* stands out from this observed pattern, as it boasts the highest number of SAG genes (172), despite not inflicting considerable damage to the intestines. Nonetheless, it still negatively affects bird performance and productivity. In contrast, species such as *E. maxima*, *E. praecox*, and *E. acervulina*, which stimulate strong immunity against reinfection even after exposure to a small number of parasites, have fewer SAG genes. Conversely, species like *E. necatrix* and *E. tenella*, which provoke a weaker immune response, have a higher number of SAG genes. However, *E. brunetti* and *E. mitis*, which exhibit moderate levels of immunogenicity, do not adhere to this trend (Reid *et al.*, 2014).



**Figure 1.9: SAG attachment to membrane.** A schematic diagram of a SAG protein (yellow) attached to the cell surface (blue) through the GPI-anchor (orange and purple ovals). a) GPI-anchored proteins are situated within the extracellular or luminal leaflet of membranes via their glycolipid components and are not accessible from the cytosolic side of the membrane. b) These proteins are initially produced as precursors with a cleavable, hydrophobic N-terminal signal sequence that guides the protein to the lumen of the endoplasmic reticulum (ER) and a cleavable C-terminal signal sequence for GPI anchoring. The GPI-anchoring signal includes a hydrophobic region separated from the GPI attachment site ( $\omega$ -site) by a hydrophilic spacer. Small side chain amino acids are highly preferred at the two positions following the  $\omega$ -site. c) The GPI core structure, which is conserved, includes ethanolamine phosphate linked to the protein's C-terminus, three mannose residues, glucosamine, and phosphatidylinositol. Adapted from (Mayor and Riezman, 2004).



Table 1.2: *Eimeria* species biology and genomic sag repertoire

species	Site of development	Disease type	Pathogenicity	SAG-A	SAG-B	SAG-C	Total	Pseudogene fragments
<i>E. praecox</i> 'H'	SI (upper)	M	+	15	0	4	19	20
<i>E. maxima</i> 'H'	SI (mid)	M	+++	35	0	4	39	29
<i>E. acervulina</i>	SI (upper)	M	++	13	1	2	16	16
<i>E. brunetti</i> 'H'	SI (lower), rectum, caeca	H	++++	61	0	44	105	39
<i>E. mitis</i> 'H'	SI (lower)	M	++	145	0	27	172	128
<i>E. necatrix</i> 'H'	SI (mid), caecaa	H	+++++	86	32	1	119	102
<i>E. tenella</i> 'H'	Caeca	H	++++	60	28	1	89	23

The sag gene counts, subfamily distribution, and the presence of sag pseudogenes (in the case of *E. tenella*) or pseudogene fragments (in other species) vary across different *Eimeria* genomes. The designations (SI), (M), and ('H') refer to the small intestine, malabsorptive disease, and hemorrhagic disease, respectively. Modified and adapted from (Reid *et al.*, 2014)

#### **1.1.5.5 Molecular Characterization of *Eimeria tenella* Surface Antigen (SAGs) Reveals Unique Structural Features**

*Eimeria tenella*, contains a total of 89 SAG genes, which are classified into three subfamilies (A, B, and C). The *Eimeria tenella* SAGs are characterized by a cysteine-rich ectodomain that spans approximately 300 amino acid residues, and they exhibit less than 5% overall sequence identity (Reid *et al.*, 2014). These proteins possess an N-terminal signal sequence and a glycosylphosphatidylinositol (GPI) anchor site at the C-terminus, enabling their attachment to the cell membrane [Figure 1.9] (Tabarés *et al.*, 2004). Within *Eimeria tenella*, these proteins make up around 1% of the proteome and are differentially expressed during the invasive sporozoite and various merozoite stages of the parasite's life cycle. Some of these proteins have demonstrated the ability to induce a pro-inflammatory response when exposed to avian macrophages in laboratory culture (Chow *et al.*, 2011). *In vitro*, *Eimeria tenella* SAG1 has the capability to adhere to epithelial cells, potentially serving a crucial role in the attachment of the parasite to the host cell surface before invasion. Moreover, it can induce protective immunity against a homologous challenge (Jahn *et al.*, 2009; Song *et al.*, 2015; Reid *et al.*, 2014; Song *et al.*, 2015). Initiating the process of host invasion, it has been proposed that positively charged regions on the outer surface of *Eimeria tenella* SAG1 interact with negatively charged sulphated proteoglycans on the host cell surface [Figure 1.10] (Jahn *et al.*, 2009).

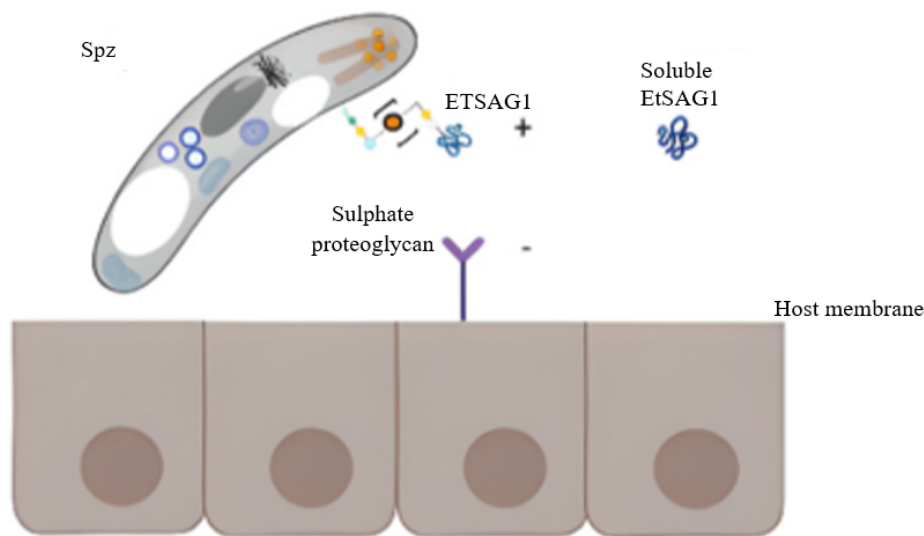


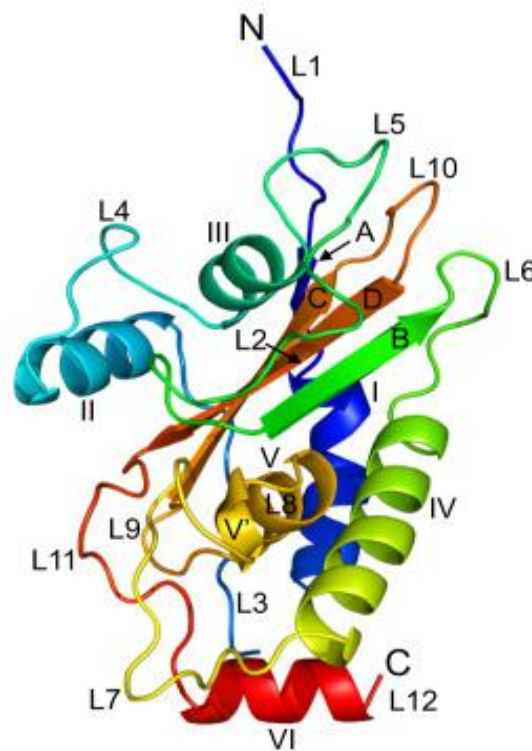
Figure 1.10: Interaction with **surface antigen *Eimeria tenella* SAG1** with **host cell of chicken**. The surface antigen *Eimeria tenella* SAG1 is displayed on the extracellular sporozoite in patches that contain positively charged regions (indicated with +) facing outward. These positively charged patches may facilitate attachment to negatively charged sulphated proteoglycans (indicated with -) on the host cell surface. *Eimeria tenella* SAG1 can also be cleaved to produce a soluble form that is shed from the parasite surface. Adapted and modified from (Britez et al., 2023).

#### 1.1.5.6 Unravelling Sequence Similarities of *Eimeria Tenella* SAGs

The determination of *Eimeria tenella* SAG19 (the only one known structure of *E. tenella*) structure at a resolution of 1.32 Å, showing that this protein folds into  $\alpha\beta\alpha$  sandwich [Figure 1.11] (Ramly et al., 2021) has significantly advanced the understanding of the structural biology of *Eimeria tenella* SAGs within the apicomplexan parasites. Unlike its counterparts in other apicomplexans such as *Toxoplasma* and *Plasmodium*, in which their SAG proteins exhibit a two-domain fold related to SRS-SAG proteins (He et al., 2002; Dietrich et al., 2022), SAG19 features a one domain fold that comprises a four-stranded, anti-parallel  $\beta$ -sheet surrounded by six  $\alpha$ -helices and connected by loops. This structure shares the same overall fold as seen in the cysteine-rich secretory proteins, antigen 5, and pathogenesis-related 1 proteins (CAP) superfamily of proteins, underscoring its distinct difference from the SRS-SAG structural fold seen in *Toxoplasma* and *Plasmodium* (Ramly et al. 2021).

The sequences alignment of the three family of *E. tenella* (SAG-A, SAG-B, and SAG-C) highlighted conserved motifs and cysteines aligning to SAG19's fold, with specific residues,

including an NxxR motif, being conserved across all subfamilies and integral to the core structure of SAG19 (Ramly *et al.* 2021). This suggests a unifying structural agreement for the *E. tenella* SAG family, bridging gaps in sequence similarity with structural congruence. Thus, the SAG19 structure not only enriches the understanding of the complex structural patterns within the *Eimeria* SAG families but also underscores the evolutionary intricacies underlying the diversification of surface antigen proteins in apicomplexan parasites, offering a new vista for exploring the mechanistic roles and potential therapeutic targets within these pathogens.



*Figure 1.11: Eimeria tenella SAG19 Structure.* The 3D model of SAG19 structure, which fold as  $\alpha\beta\alpha$  sandwich, the structure consists of four antiparallel  $\beta$ -strand surrounded by six  $\alpha$ -helix and connected with number of loops (Ramly *et al.* 2021).

#### 1.1.5.7 *Eimeria* SAG Proteins in relationship to CAP Superfamily: A Shared Structural Scaffold

Structural characterization of *Eimeria tenella* SAG19 has revealed conserved secondary structure features shared with other CAP superfamily members such as Tablysin-15 (Xu *et al.*, 2012), as well as a conserved NxxR sequence motif, together indicating an evolutionary



relationship, highlighting the high level of similarity (Ramly *et al.*, 2021). Nevertheless, the sequence identity between *Eimeria tenella* SAG19 and some of the CAP proteins (GAPR-1, Ves V 5 and Na-ASP 2) (Serrano *et al.*, 2004; Henriksen *et al.*, 2001; Asojo *et al.*, 2005) is quite low, at only 8-13% (Ramly *et al.*, 2021). While the central domain structure is conserved, there are variations in the loops, terminal extensions in some proteins, and differences in surface charge patterning and hydrophobic residues (Ramly *et al.*, 2021). These differences highlight the nuanced variations that confer unique properties and functions to each protein, underscoring the intricacies of protein structure-function relationships vital in fields like drug design.

#### **1.1.6 Extending the exploration to other apicomplexan surface antigens related to the *Eimeria tenella* SAG proteins**

Apicomplexan surface antigens genes encoding surface proteins are important virulence factors that are responsible for the pathogen's initial interactions with the host cell surface and components of the host immune response. These antigens, derived from asexual reproductive stages like sporozoites, first-generation schizonts, and merozoites have been linked to providing immunity against subsequent infections (McDonald *et al.*, 1986). Interestingly, surface antigens have also been extensively studied regarding their involvement in how parasites grow, develop, and survive (Wallach *et al.*, 2012). There is evidence that susceptibility to *T. gondii* lethal challenge is minimised by the passive transfer of both mono- and polyclonal antibodies raised to surface antigens of asexual phase of this parasite (Brinkmann *et al.*, 1993). Likewise, robust suppression and anti-infection effect are demonstrated by *Eimeria* and *Plasmodium* anti-sporozoite and anti-merozoite antibodies detecting glycosylphosphatidylinositol (GPI-) anchored antigens in *Eimeria tenella* and *Plasmodium falciparum*, respectively (Witcombe *et al.*, 2004).

##### **1.1.6.1 *Toxoplasma* SAGs: Diversity and Roles in Pathogenesis**

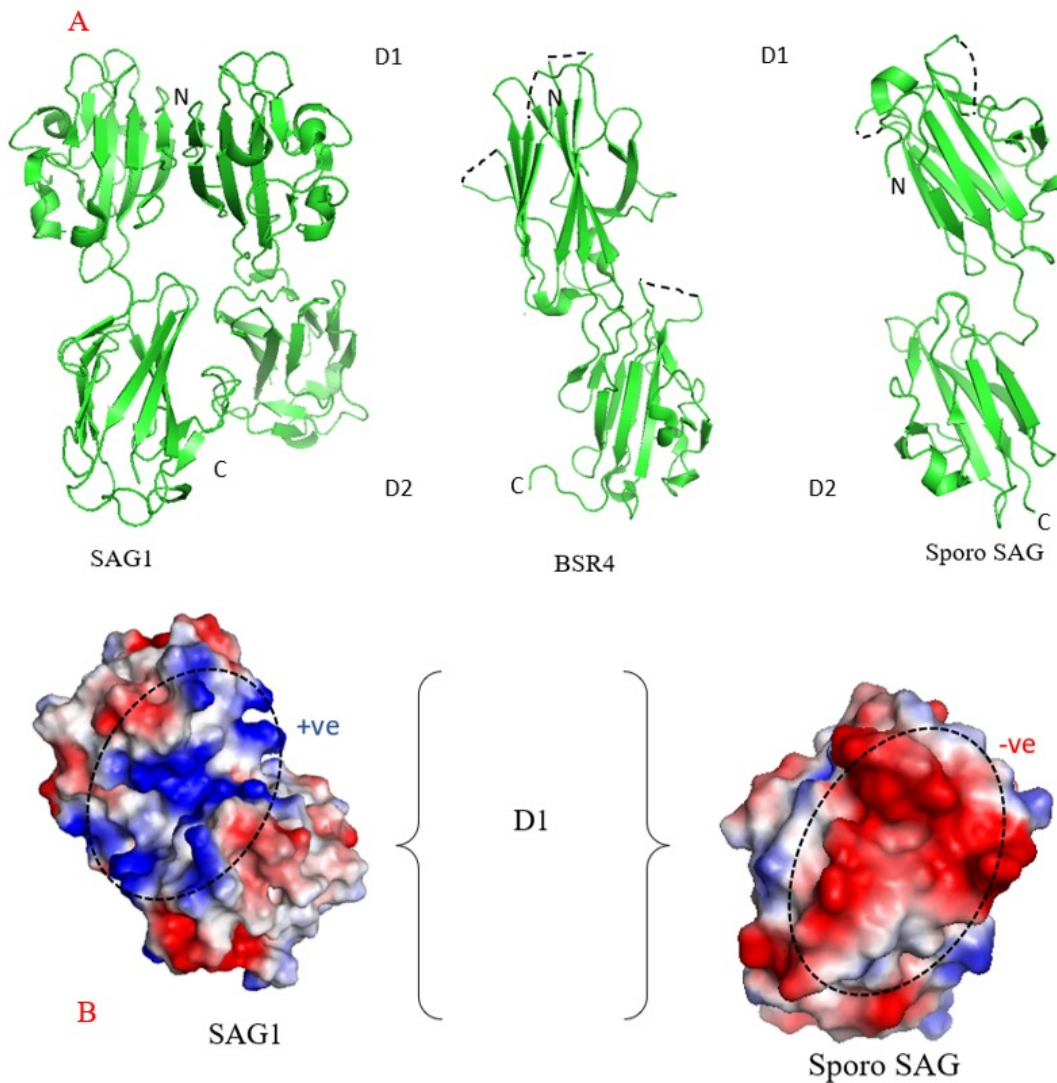
The surface antigen (SAG) proteins of the *Toxoplasma* parasite belong to a different SAG superfamily (the SRS SAGs), but do share some basic similarities, having a putative signal peptide at the N-terminus and a C-terminal hydrophobic region believed to act as a site for the glycosylphosphatidylinositol (GPI) anchor for surface attachment (Nagel and Boothroyd, 1989). The SRS-like superfamily created by these SAGs ramify into two subfamilies, one of

which is represented by surface antigen 1 (SAG1) and the other by surface antigen 2 (SAG2) (Jung *et al.*, 2004). The characteristics of the SAG superfamily include an abundance of cysteine residues, typically with 12 conserved cysteines, a signal peptide at the N-terminus, and a GPI-anchor (Crawford *et al.*, 2009).

These antigens are implicated in host cell invasion, immune regulation, and potentially reducing virulence. Additionally, the function might be aids the parasite's survival in its environment (Lekutis *et al.*, 2001). The GPI portion of these proteins can stimulate the innate immune system via Toll-Like Receptors (TLR), such as TLR-2 and TLR-4, impacting the production of nitric oxide, IL-12, and TNF- $\alpha$  (Debierre-Grockiego *et al.*, 2007). *In vitro* studies further support the importance of SAG1 and SAG2 in the invasion process, with anti-SAG1 antibodies reducing tachyzoite distribution in host cells, and anti-SAG2 antibodies influencing parasite reorientation and invasion events (Grimwood and Smith, 1996).

As revealed by expression analyses (Wang and Yin, 2014), the developmental phase determines the SAG expression pattern, for instance, SAG1 expression is associated with the tachyzoite developmental phase, while the SAG family member bradyzoite surface antigen (BSR4) is associated with the bradyzoite phase (Crawford *et al.*, 2009), whereas sporo-SAG expression is associated with the sporozoite phase (Döşkaya *et al.*, 2014). Structural studies have established that these three SAGs exhibit a fold that contains two connected  $\beta$ -sandwich domains with one domain folded from residues at the N-terminus (D1) and the other from the C-terminus (D2) portion [Figure 1.12 A] (Crawford *et al.*, 2010; He *et al.*, 2002). On the other hand, the quaternary structure shows some variation between the SAGs, with sporo SAG being monomeric (Crawford *et al.*, 2010) while SAG1 and BSR4 are dimeric [Figure 1.12 A] (He *et al.*, 2002). Furthermore, these SAGs are very much dissimilar in terms of molecular surface, even though they do not differ significantly in terms of the general structure. More specifically,

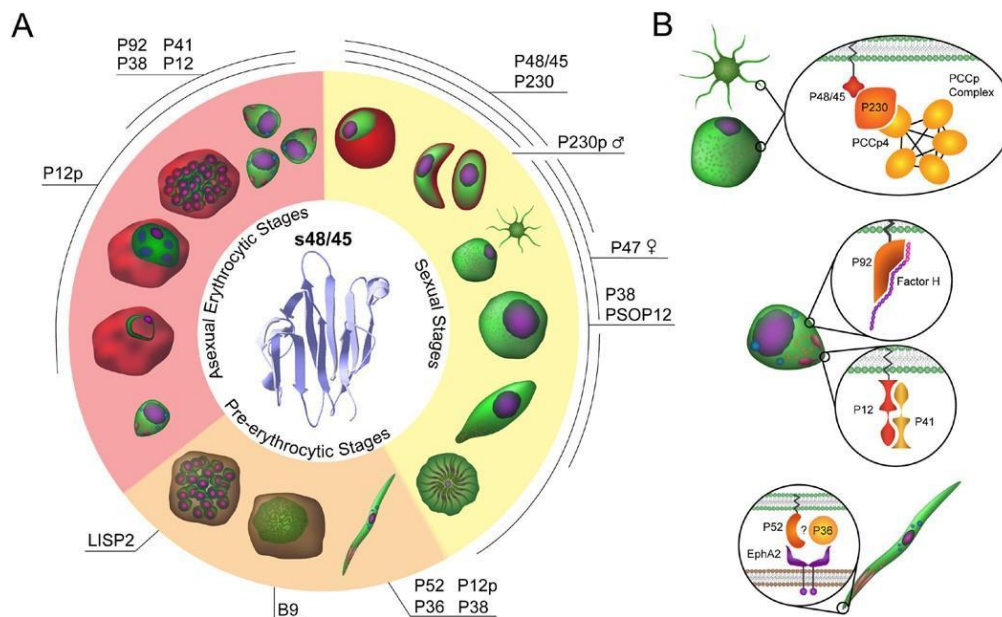
SAG1 and BSR4 surfaces display positive charge, whereas the sporo SAG surface displays primarily a negative charge [Figure 1.12 B] (Crawford *et al.*, 2009; He *et al.*, 2002).



**Figure 1.12: Represents the surface antigens of *Toxoplasma gondii*.** A) SAG1, BSR4, and Sporo SAG folded as D1 and D2 domain that comprise  $\beta$  sheet sandwiches (PDB accession code 1KZQ, 2JKS, and 2WVK respectively). B) Electrostatic representation of SAG1 D1 domain, highlights the distinct localisation of positive charge (left), while Sporo SAG D1 domain, highlights the distinct localisation of negative charge (right).

#### 1.1.6.2 Plasmodium SAGs: Implications for Invasion and Host Interaction

*Plasmodium* parasites possess numerous families of adhesion proteins. These families have either evolved alongside the apicomplexan parasite lineages since their split from free-living ancestors or were acquired through horizontal gene transfer from an ancient host (Rowe *et al.*, 2009). Many of these proteins play crucial roles in facilitating interactions between the parasite's sexual stages during mating and its extracellular invasive stages during infection and growth within host cells (Arredondo and Kappe, 2017). These pivotal interactions are among the functions of the surface protein family, which in *P. falciparum* comprises 14 members expressed across different stages of the parasite cycle (Pfs230, Pfs48/45, Pfs230p, Pfs47, and PfPSOP12 during sexual stages; Pf52, Pf36, PfLISP2, and PfB9 during pre-erythrocytic stages; and Pf12, Pf12p, Pf41, Pf38, and Pf92 during asexual erythrocytic stages) [Figure 1.13] (Arredondo and Kappe, 2017). These proteins are commonly found on the parasite's surface and are conserved across various *Plasmodium* species, with most members having counterparts in human, non-human primate, and rodent malaria parasites (Aurrecoechea *et al.*, 2009). This underscores the universal significance of the surface protein family in the survival and proliferation of *Plasmodium*.



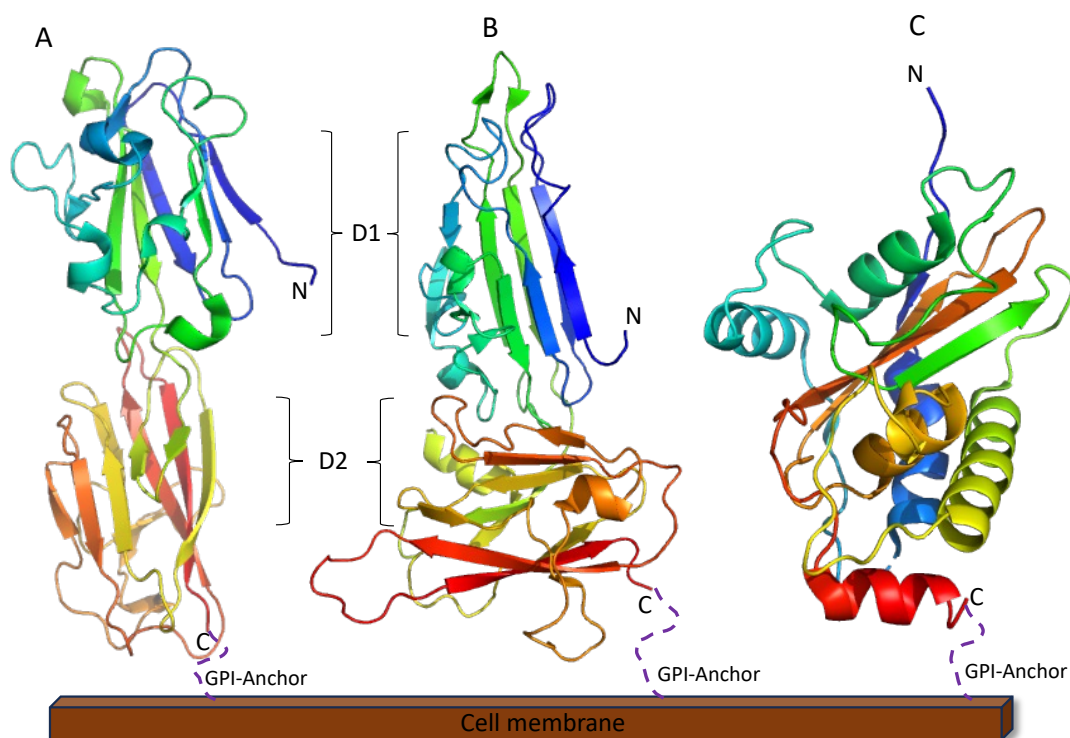
**Figure 1.13: The s48/45 six-cysteine protein family plays a vital role in the *P. falciparum* life cycle.** The life cycle can be categorized into three main stages: sexual stages, pre-erythrocytic stages, and asexual erythrocytic stages, each represented by specific shading. The s48/45 protein family is expressed at various stages, and proteins containing at least one s48/45 domain are identified. Panel B illustrates the protein interactions of the s48/45 family, with confirmed interactions such as PfCCp-Pfs230 complex and Pfs48/45 on gametes, Pf92 and Factor H, and Pf12 with Pf41 on the merozoite. Additionally, there are suspected interactions of P52 and P36 with EphA2 on the sporozoite. Adapted from (Arredondo and Kappe, 2017).

### 1.1.7 Comparison of SAG proteins Within Apicomplexan parasites

Apicomplexan parasites share common features in their surface antigens, including glycosylphosphatidylinositol (GPI) anchors, conserved cysteine motifs, and stage-specific expression (Reid *et al.*, 2012). However, there are also notable differences between genera. In *Eimeria*, the CAP-like SAG proteins likely facilitate host interaction and invasion. They comprise three subfamilies, with SAG-A conserved across species (Reid *et al.*, 2014). In *Toxoplasma* a different SAG superfamily of SRS proteins is divided into two subfamilies distinct by SAG1 and SAG2. These are involved in invasion and survival, with developmental regulation (Khanaliha *et al.*, 2014). In *Plasmodium* these SAG proteins, which also referred as adhesion proteins, mediate interactions during growth and infection. These protein members have conserved roles despite stage-specific expression.

In addition, SRS proteins are also found in *Besnoitia besnoiti* and *Neospora caninum*, with 280 and 233 members respectively identified (O'Toole and Jeffrey, 1987), these parasites are also known for causing significant diseases in livestock. The analysis of these proteins suggests a closer relationship to the *Toxoplasma gondii* SAG family rather than the *Eimeria* SAG family. As described, the sequences of these SRS SAG proteins from *B. besnoiti* and *N. caninum*, similar to those conducted on *T. gondii*, indicates that the highest variability in sequence is found in the D1 domain, away from the cell membrane (Ramly *et al.* 2021). In contrast, the D2 domain exhibits more sequence conservation, aligning with observations made in *T. gondii* SAGs/SRS proteins (Ramly *et al.* 2021). This shared structural characteristic among the SRS proteins suggests a common pattern of evolution or function across these species.

The determination structure of *Eimeria tenella* SAG 19 showed that the protein folds into one domain in a three-layer  $\alpha\beta\alpha$  sandwich (Ramly *et al.*, 2021). Interestingly, this fold is distinct and unrelated to the dimeric structures observed in the SAG1 (PDB 1KZQ) (He *et al.*, 2002) and SAG2 (PDB 2WVK) (Crawford *et al.*, 2010) families of *T. gondii*, as well as the SRS domains (PDB 7USR) found in *P. falciparum* [Figure 1.14] (Dietrich *et al.*, 2022). The SAG proteins in *Plasmodium* and *Toxoplasma* fold into two  $\beta$ -sheet sandwich domains. Consequently, the structure of *E. tenella* SAG19 stands as a unique example among SAGs from apicomplexan parasites (Ramly *et al.*, 2021).



**Figure 1.14: A comparison of the structure of *T. gondii*, *P. falciparum* and *E. tenella* SAG proteins.** A) indicates the structure of *T. gondii* SAG1 that folds into two domains (D1 and D2). B) shows the *P. falciparum* Pfs230 that also has similar fold to *T. gondii* SAG1 which fold into two domains. C) The structure of *E. tenella* SAG19 that folds into one domain with clear differences in fold compared to that seen in *Toxoplasma* (SAG1) and *Plasmodium* (Pfs230). (PDB accession code *E. tenella* SAG19 6zzb, *T. gondii* SAG1 1KZQ, and *P. falciparum* Pfs230 7USR).

#### 1.1.7.1 Unravelling Common Themes and Unique Features of Apicomplexan Surface Antigens

As the surface antigens are critical proteins expressed on the outer membrane of pathogenic apicomplexan parasites such as *Plasmodium*, *Toxoplasma*, and *Eimeria*, these molecules play indispensable roles mediating host-parasite interactions during key processes like adhesion, invasion, and immune evasion. The *Eimeria* SAG protein family comprises abundant cysteine-rich surface coat proteins with a conserved structural fold, despite sequence diversity, which assists gut invasion (Ramly *et al.*, 2021). *Plasmodium* and *Toxoplasma* also express variant surface antigens (VSAs) involved in cytoadherence and immune stimulation (Opitz *et al.*, 2002). Other well-studied antigens include AMA1, which facilitates parasite attachment to host cells, and proteins secreted during invasion that aid adhesion (Giovannini *et al.*, 2011). While sharing some architectural themes, surface antigens among apicomplexan genera exhibit great variability and adaptability (Wallach *et al.*, 2012). There was research that characterised a new

antigen, which compares conserved versus unique features between orthologs, and elucidates structure-function relationships. These efforts provide insights into parasite pathogenesis mechanisms and identify promising vaccine candidates against these debilitating diseases (Blake *et al.*, 2011). Overall, surface coat molecules are essential for parasite infectivity and survival, highlighting their importance as targets for interventions against apicomplexan infections (Blake *et al.*, 2011).

#### **1.1.7.2 Evolutionary Dynamics and Functional Consequences of Apicomplexan Surface Antigens**

The evolutionary dynamics and functional roles of surface antigens in apicomplexan parasites such as *Plasmodium*, *Toxoplasma*, and *Babesia* are important research areas providing insights into host-parasite interactions. Surface antigens exhibit great evolutionary adaptability, as evidenced by antigenic variation enabling immune evasion. *Babesia* variant antigen genes reveal ongoing innovation in parasite genomes for antigens involved in immune evasion (Jackson *et al.*, 2014). Malaria parasites have also evolved sophisticated mechanisms for gene switching and variant expression. Although the factors governing apicomplexan genome evolution are unclear, surface antigen gene expansion and contraction have influenced lineage-specific genome sizes (Swapna and Parkinson, 2017).

Functionally, surface antigens mediate critical host-parasite interactions including adhesion, invasion, and modulation of immune responses. Surface coat molecules play indispensable roles in merozoite invasion and erythrocyte targeting. A study reveals that the convergent evolution of immune evasion functions reflects conserved host immune pressures (Lim *et al.*, 2014). Lineage-specific surface antigens in apicomplexans are probable players in host-specificity and niche adaptation. Overall, surface antigens represent a dynamic protein class exhibiting great evolutionary adaptability and functional versatility at the host-parasite interface. Elucidating the molecular evolution and functional roles of these key parasite factors provides insights into host-pathogen interactions and identifies strategic targets for interventions (Kuo and Kissinger, 2008).



## 1.2 Aims of the Project

The Apicomplexan family thus clearly contains (at least) two different types of surface antigen proteins, the well-studied SRS SAGs of *Toxoplasma* and *Plasmodium* and the less understood CAP-like SAGs as seen in *Eimeria*. It was thus decided to focus this thesis on the large family of *Eimeria* CAP-like SAGs in attempts to elucidate the structure-function relationships of this important family of proteins and to attempt to understand the *Eimeria* parasite interaction within its host. The project was based on unravelling structural variation within the CAP-like surface antigen (SAG) superfamily. The first step was to establish a robust expression system for reliable production of properly folded SAG proteins. This would enable downstream structural studies using X-ray crystallography to determine the structures of representative proteins from this family.

SAGs exhibit sequence variability that likely contributes to functional differences. By analysing SAG structures, the project aimed to delineate structure-function relationships and gain insights into how sequence variation impacts protein function and parasite-host interactions. This project represented a crucial step toward unravelling *Eimeria* biology, and the findings could ultimately facilitate development of improved strategies to control *Eimeria* infections and mitigate coccidiosis impact.

The specific goals of the project included:

- Optimization of SAG protein expression for structural studies.
- Production of SAG protein crystals.
- Determination of crystal structures of SAGs.
- Analysis of SAG structural diversity and its relation to sequences.
- Identification of host binding partners.
- Expansion of the structural characterization to *Eimeria*-like SAGs in *Plasmodium* and other apicomplexan parasites.

Progress towards these aims is detailed in the following chapters.

## Chapter2

### 2.1 MATERIALS AND METHODS

The chapter has been divided into two sections. The first part discusses the molecular techniques used in these studies, while the second part provides an overview of the X-ray crystallographic techniques employed to determine the three-dimensional structure of surface antigens (SAGs).

#### Part I: Molecular biology techniques

This section describes all the molecular biology methods that were used in the research, including cloning, transformation, cell growth, protein purification, and crystallization.

##### 2.1.1 Strains, vectors, and recombinant plasmids

###### 2.1.1.1 Bacterial strains and vectors

In these studies, various *Escherichia coli* strains were utilized, including DH5 $\alpha$  for plasmid amplification, and BL21 (DE3) and *Origami* strains for overexpression. All competent cells were obtained from Invitrogen and Stratagene. Cloning and expression vectors, such as pET22 (b) and pET32 (a)+, were used and purchased from Genscript.

###### 2.1.1.2 Recombinant plasmid

Plasmids (pET32a/pET22b(+)-SAGX) containing constructs of the codon optimised genes for SAGs from *Eimeria tenella*, *Eimeria brunetti* [Table 2.1] and SAGs from other species (*Plasmodium*, *T. gondii*, *Babesia bovis* and *Besnoitia Besnoiti*) [Table 2.2] were ordered from Genscript which is a biotech company providing life sciences services and products.

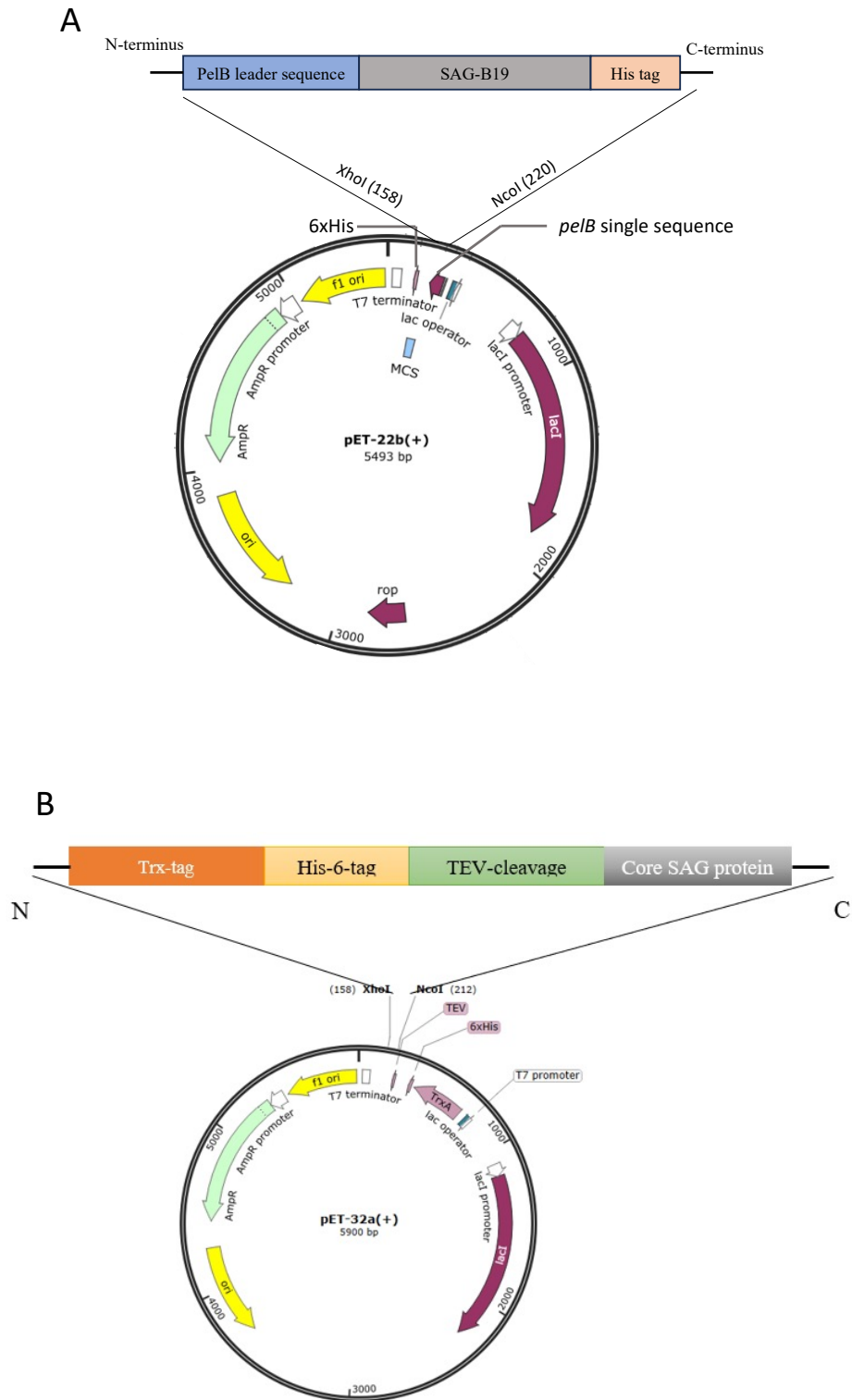


Figure 2.1: **The plasmid map.** a) The *PelB* construct inserted into the *pET-22b* vector while b) *Trx*-construct insertion into *pET32a*.

Table 2.1: *Eimeria* SAGs Trx-construct.

NO	Family	Gene ID	CONSTRUCT DESCRIPTION
1	A	SAG 10	N-terminal-Trx-6xHis-tag-TEV-SAG10-C-terminal
2	A	SAG 31	N-terminal-Trx-6xHis-tag -TEV-SAG31-C-terminal
3	A	SAG 7	N-terminal-Trx-6xHis-tag -TEV-SAG7-C-terminal
4	A	SAG6	N-terminal-Trx-6xHis-tag -TEV-SAG6-C-terminal
5	A	SAG1	N-terminal-Trx-6xHis-tag -TEV-SAG1-C-terminal
6	A	ETH_00010870	N-terminal-Trx-6xHis-tag -TEV-ETH_00010870-C-terminal
7	A	ETH_00034905	N-terminal-Trx-6xHis-tag -TEV-ETH_00034905-C-terminal
8	A	ETH_00023375 (SAG-A91)	N-terminal-Trx-6xHis-tag -TEV-ETH_00023375-C-terminal
9	B	SAG-B16	N-terminal-Trx-6xHis-tag -TEV-SAG16-C-terminal
10	B	SAG-B22	N-terminal-Trx-6xHis-tag -TEV-SAG22-C-terminal
11	B	SAG-B41	N-terminal-Trx-6xHis-tag -TEV-SAG41-C-terminal
12	B	SAG-B13	N-terminal-Trx-6xHis-tag -TEV-SAG13-C-terminal
13	C	ETH_00001975 (SAG-C)	N-terminal-Trx-6xHis-tag -TEV-ETH_00001975-C-terminal
14	C	EBH_0036680	N-terminal-Trx-6xHis-tag -TEV-EBH_0036680-C-terminal
15	C	EBH_0015570	N-terminal-Trx-6xHis-tag -TEV-EBH_0015570-C-terminal
16	C	EBH_0027980	N-terminal-Trx-6xHis-tag -TEV-EBH_0027980-C-terminal

Table 2.2: CAP-like SAG from different Apicomplexan parasites in trx-construct.

SPECIES	SAG GENE ID	CONSTRUCT DESCRIPTION
<i>Plasmodium falciparum</i>	CK202_5340	N-terminal-Trx-6xHis-tag-TEV-CK202_5340-C-terminal
<i>Plasmodium ovale</i>	PocGH01_01013600	N-terminal-Trx-6xHis-tag -TEV- PocGH01_01013600-C-terminal
<i>Plasmodium vivax</i>	PVX_087830	N-terminal-Trx-6xHis-tag -TEV- PVX_087830-C-terminal
<i>Plasmodium Malaria</i>	PmUG01_01017300	N-terminal-Trx-6xHis-tag -TEV- PmUG01_01017300-C-terminal
<i>Babesia bovis</i>	BBOV_III003800	N-terminal-Trx-6xHis-tag -TEV- BBOV_III003800-C-terminal
<i>Besnoitia besnoiti</i>	BESB_017040	N-terminal-Trx-6xHis-tag -TEV- BESB_017040-C-terminal

## **2.2 Growth Media and Solutions**

### **2.2.1 Stock Solutions**

The stock solutions used in this work were prepared using Ultra-pure water from the Barnstead EASY Pure II water purification machine (Thermo Scientific) and filtered through 0.22 $\mu$  filters.

### **2.2.2 Lysogeny broth media (LB)**

The preparation of LB Media, was prepared (Miller 1972), involved mixing 10 grams of tryptone, 10 grams of NaCl, and 5 grams of yeast extract into 1 liter of deionized water. This mixture was then autoclaved promptly at 121°C and subsequently stored at room temperature for future use.

### **2.2.3 LB Agar**

LB agar preparation involved adding 15 g of bacterial agar to 1 liter of LB media. This mixture was then sterilized in an autoclave at 121°C. The LB agar was kept at room temperature until required. When ready to use, the LB agar was dissolved in a microwave on medium setting, and 25 ml was dispensed into each petri dish. The plates were then supplemented with either 100  $\mu$ g/ml of ampicillin for plasmid propagation or with 50  $\mu$ g/ml of kanamycin and 10  $\mu$ g/ml tetracycline and, when necessary. The plates were used immediately once set.

### **2.2.4 M9 Minimal Salts, 5X**

The media was prepared by dissolving 56.4 g of minimal microbial growth medium in 1 L of distilled water. A 5x concentration of M9 salts was then prepared and autoclaved for 15 minutes at 121°C. Other reagents such as a 1M stock solution of MgSO<sub>4</sub>, a 1M stock solution of CaCl<sub>2</sub>, and a 10 mg/ml stock solution of thiamine were also prepared and autoclaved.

### 2.2.5 Terrific Broth (TB) media

TB media was prepared by adding 24 g of yeast extract, 20 g of tryptone, and 4 ml of glycerol into 900 ml of deionized water. The mixture was shaken or stirred until all the solutes were fully dissolved. Then, sterilize the solution by autoclaving it for 20 minutes at 15 *psi*. The solution was cooled down to room temperature before diluting to 1x.

10X TB salts	0.17 M $\text{KH}_2\text{PO}_4$	23.1 g/L
	0.72 M $\text{K}_2\text{HPO}_4$	125.4 g/L

Dissolved 23.1 g  $\text{KH}_2\text{PO}_4$  and 125.4 g  $\text{K}_2\text{HPO}_4$  in water to a final volume of 1 L and autoclaved for 15 min at 121°C.

### 2.2.6 Antibiotics and IPTG

The stock solutions of antibiotics (ampicillin, kanamycin and tetracycline) were prepared and added to transformants containing antibiotic resistance genes. All antibiotics were filter sterilized and stored at -20°C.

- **Ampicillin:** Stock of 100 mg/ml was prepared, kept at -20 C, and used as required at a final concentration of 100 ug/ml in growth media.
- **Kanamycin:** Stock of 50 mg/ml was prepared, kept at -20 C, and used as required at a final concentration of 50 ug/ml.
- **Tetracycline:** Stock of 10 mg/ml was prepared, kept at -20 C, and used as required at a final concentration of 10 ug/ml.

**All the antibiotics dissolved in MLQ water**

#### ○ Isopropyl-D-Thiogalactopyranoside (IPTG)

1M of stock solution was prepared and filter sterilized using a 0.22µm filter and stored at -20°C as for farther use.

### 2.2.7 Transformation

Plasmids containing surface antigens (SAGs) obtained from Genscript were transformed into either *E. coli* DH5α for increase amount of Plasmid or *Origami* cells (DE3) for overexpression. The *Origami* cell (DE3) was chosen due to its ability to facilitate the folding of proteins involving disulphide bridges, which were anticipated to be present in SAGs. The strains were purchased as chemically competent cells and stored in glycerol at -80°C. Cells were defrosted

on ice before adding 1 µl of plasmid and incubating for 5 minutes. Cells and plasmids were mixed gently and then incubated on ice for 30 minutes. Following this, cells were heat-shocked at 42°C for 30 seconds and then incubated on ice for 2 minutes to recover. 100 µl of SOC media was added to the cells, which were then incubated at 37°C with shaking at 200 rpm for 1 hour. Finally, cells were plated onto LB-agar with the appropriate antibiotics and incubated at 37°C overnight. The plate was stored in the refrigerator until needed.

### **2.2.8 Plasmid Extraction**

To propagate the plasmids, *E. coli* DH5α colonies that containing plasmids were inoculated into 50 ml of fresh LB medium supplemented with the appropriate antibiotic. The cultures were grown overnight at 37°C and 200 rpm. Subsequently, the cultures were harvested by centrifugation for 10 minutes at 3500 g and 4°C. Plasmid extraction was carried out using the BioLabo miniprep kit, following the manufacturer's instructions.

## **2.3 Cell Growth**

### **2.3.1 Over Expression**

The study employed an expression vector from the *pET* system, which utilizes the bacteriophage T7 RNA polymerase to regulate the overexpression of proteins. To initiate the overexpression of the target protein, this system relies on the addition of IPTG, a molecule similar to lactose that binds to the lac repressor, resulting in reduced affinity for DNA. When IPTG is introduced into the culture, the T7 polymerase is expressed and immediately starts transcribing the cloned gene, leading to the translation of the target protein.

Overexpression was generally tested on a small scale (50 ml) before proceeding to large-scale (500 ml) growths. The conditions that yielded the highest level of overexpressed protein (soluble protein) were used in the large-scale growths. Plasmids containing SAGs protein fused to the N-terminal Trx-His-tag-TEV were transformed into *E. coli* strain *Origami* cells (DE3), as previously described in section 2.2.7.

#### **A) Overexpression on small scale**

50 ml LB with 100 µg/ml ampicillin, 50 µg/ml kanamycin and 10 µg/ml tetracycline was inoculated with a single colony of *origami* cells that contains plasmid and then incubated

overnight at 37 °C on 220 rpm shaking. 1% v/v of the overnight suspension from the overnight culture was used to inoculate a fresh 50 ml LB with containing the same three antibiotics at the same concentration and incubated at 37 °C until an OD<sub>600</sub> of 0.6 was reached. Then cells were induced with either different concentration of IPTG at various temperatures to reach the best expression conditions. Protein expression was analysed by SDS-PAGE to optimise the best temperature for soluble protein expression.

## **B) Overexpression on Large scale**

50 ml LB with 100 ug/ml ampicillin, 50 ug/ml kanamycin and 10 ug/ml tetracycline was inoculated with a single colony of *origami* cells that contains plasmid and then incubated overnight at 37 °C on 200 rpm shaking. 1% v/v of the overnight suspension from the overnight culture was used to inoculate a 6x 500 ml LB flasks with containing the same three antibiotics at the same concentration and incubated at 37 °C until an OD<sub>600</sub> of 0.6 was reached. Then the flasks were induced with 1mM IPTG and incubated overnight at 17°C on 200 rpm shaking, which this is the best temperature and best concentration of IPTG.

All cultures were transferred to centrifugation tubes (500 ml) and harvested by centrifugation for 20 minutes at 8,000 g (8,452 rpm) at 4°C using a F500 rotor in a Beckman Avanti centrifuge. The supernatant was discarded, leaving sufficient media to resuspend the pellet. The cells were transferred into new 50 mL Falcon tubes and centrifuged at 5,400 g (5,500 rpm) for 15-20 minutes at 4°C using a Sigma 3-16K rotor. The resulting pellets (cell paste) were weighed and frozen at -20°C for further analysis.

### **2.3.2 Preparation of Cell Free Extracts (CFE)**

To examine the protein expression level, cell lysates were prepared by resuspending cells in 50 mM Tris-HCl pH 8.0 (5-10 ml per gram of cell paste) and disrupted using a sonicator (Soniprep 150 machine) at 16-micron amplitude (3 cycles for 10 seconds). The cells were cooled between cycles for 30 seconds on ice. The soluble fraction was separated from the cell debris by centrifugation for 15 minutes at 72,000 g (24,500 rpm) using a J-25.50 rotor in a Beckman Avanti centrifuge at 4°C. The resulting supernatant, referred to as the cell-free extract (CFE), was used for protein purification. To analyse the protein solubility, the insoluble fraction (cell



debris) was resuspended in water to an equal volume of the CFE. The concentration of both the soluble and insoluble fractions was estimated using the Bradford method (Bradford, 1976) as described below and further analysed by SDS-PAGE electrophoresis described in section 2.3.4.

### 2.3.3 Determination of protein concentration

Protein concentration was measured using either the Bradford assay (Bradford, 1976) or using IMPLEN P300 Nanophotometer. For the Bradford assay, 0.2 ml of Bio-Rad Bradford reagent was added to the protein sample (volume dependent on the protein being analysed), and the total volume was made up to 1 ml with Milli-Q water in a plastic cuvette. The contents of the cuvette were mixed by inversion, and the absorbance reading at a wavelength of 595 nm (OD<sub>595</sub>) was taken. Protein concentration was calculated using the following formula.

$$\text{Protein concentration } \left( \frac{mg}{ml} \right) = \frac{OD_{595} \times 15}{\text{Volume protein } (\mu L)}$$

### 2.3.4 SDS-PAGE Electrophoresis

SDS-PAGE (sodium dodecyl sulphate polyacrylamide gel electrophoresis) was used to examine the composition of different protein samples, analyse the solubility of any overexpressed protein, and determine the purity of any products. Gels were prepared using a resolving gel acrylamide (12%) and 1M Tris-HCl pH 8.8, with a stacking gel of 5% acrylamide and 1M of Tris-HCl (pH 6.8). A comb was put in the top of the gel to produce 15 wells.

Resolving gel components for one mini gel (Bio-Rad)

<b>12% <i>resolving gel</i></b>	<b>6 % <i>stacking gel</i></b>
2.5 ml 30 % acrylamide (Bio-Rad)	0.75 ml 30 % acrylamide
2.35 ml Tris pH 8.8	0.47 ml Tris pH 6.8
1.28 ml Ultra – pure water	2.46 ml Ultra – pure water
62.5 µl 10 % SDS	37.5 µl 10 % SDS
10 µl TEMED	3.75 µl TEMED
62.5 µl 10 % Ammonium persulphate (APS)	37.5 µl APS

Samples were prepared by mixing 10 – 15  $\mu$ g of protein with 7  $\mu$ l of 5x ready-made SDS loading buffer. The samples were heated at 95°C for 1 minute prior to loading into the wells. 1x running buffer was added to the gel tank, and electrophoresis was performed at 200 volts for approximately 45 minutes. The gel was then removed from the cast and transferred into a container containing (Coomassie blue) staining solution. The gel was soaked for 20 – 60 minutes and then kept overnight to allow staining.

## **2.4 Protein purification**

There were four steps involved in the purification of SAGs: optimization of this process is presented in chapter 3 for each SAG (section 3.4).

1. Affinity chromatography on a Ni-NTA column (5 ml)
2. Cleavage with Tobacco etch virus (TEV) protease
3. Second Ni-NTA column (1 ml)
4. Gel filtration

Chromatography was performed using the AKTA purifier machine (GE Healthcare, USA). For the first step and the third steps, either a 1 ml or 5 ml His-Trap <sup>TM</sup>HP cartridge (GE Healthcare) was used, while in the fourth step, gel filtration was carried out on a 16x60 Hi-Load<sup>TM</sup> Superdex<sup>TM</sup>200 column (GE Healthcare). The AKTA machine was operated using the UNICORN program.

### **○ Protocols**

To separate the SAG construct from the soluble *E. coli* proteins, the cell free extract (CFE) was prepared in buffer A (50 mM Tris-HCl, pH 8.0 + 0.5 M NaCl), following the procedure outlined in section 2.3.2. Subsequently, the cell extract was introduced to the 5 ml His-Trap HP cartridge where the unbound material was washed out with 1 column volume (CV) of buffer A. In most cases, the protein was in 3ml fractions and eluted using a gradient of 0 to 50% imidazole in buffer B (0.5 M imidazole + buffer A) (15 CV) at a flow rate of 4-5 ml/min. Specific parameters utilized in each purification process are outlined for each SAG purification Chapter 3. The protein concentrations of each fraction were estimated using the Bradford method (Bradford, 1976) as described in section 2.3.3. SDS-PAGE electrophoresis was employed to analyse the

fractions, and those containing the target protein were combined and concentrated using a VivaSpin concentrator with either a 10 kDa or 5 kDa molecular weight cut-off filter (Sartorius).

- **Cleavage with TEV followed by second Ni-NTA column (1 ml) and then followed by Gel Filtration**

The construct contained a TEV cleavage site, with the sequence (ENLYFQ\S), between the N-terminus-Trx-His-tag and the SAG core domain. The purified protein obtained from Ni-NTA column 1 (5ml His-Trap column) was incubated with 50 ug of a TEV protease (Novagen) per 1mg of protein to detach and separate the target protein from the N-terminal-Trx-His tag. The protein was cleaved overnight in cleavage buffer (Buffer A, 2-fold dilution) using a diafiltration cup and stirring overnight. Following cleavage, the protein was collected, its concentration checked and loaded onto a second Ni-NTA column (1ml His-Trap column) and washed with buffer A. Un-cleaved protein and his-tagged, TEV protease remained bound to the column, the cleaved SAG protein checked in the flow through.

The protein was concentrated by centrifugation using a VivaSpin concentrator with a 10 kDa molecular weight cut-off filter (Sartorius) to a volume of 2 ml and then loaded onto a gel filtration a Hiload™ 16/600 Superdex™ 200 pg column. The fraction that contains the SAG protein was eluted with buffer A and collected and concentrated using a VivaSpin concentrator with either 5 or 10 kDa cut-off filter. The protein was used immediately for crystallization and in some cases stored at 4°C for further use.

## **PART II: X-ray crystallographic techniques used for three-dimensional structure determination.**

### **2.5 Protein crystal preparation**

For crystallization the protein was initially concentrated, and buffer exchanged into ten-fold less of buffer A, to 18 – 23 *mg/ml* using a Vivaspin concentrator (Sartorius) with a molecular weight cutoff smaller than that of the desired protein. The protein concentration was subsequently adjusted as required based on the results of crystallization trials.

#### **2.5.1 Initial Robot Screening**

The initial screening for crystallization conditions using vapor diffusion was performed using a TTP Labtech Mosquito LCP crystallisation robot in 96 well plates. Each plate consisted of 96 individual experiments, with one large well designated for the mother liquor and three small wells that could accommodate different ratios of protein and crystallization reagent. In each experiment, one of the small wells contained a mixture of the mother liquor and protein in a 1:1 ratio. A volume of 40  $\mu$ l of each crystallization solution was pipetted into the large well, and then 150 *nl* of this solution was dispensed into the small well and 150 *nl* of protein was added to the same small well. The plate was sealed with transparent crystal-clear tape and stored at 17°C. The growth of crystals was monitored by observing each droplet under a light microscope. Crystallization trials were conducted using commercially available crystallization screens, including PACT, JCSG, pH clear, and AmSO4 screens (Molecular Dimensions, Appendix A.4).

#### **2.5.2 Crystal Optimisation**

##### **➤ Sitting drop of a 96 well plate**

Some of the crystals grown in the initial crystallization screening were further optimized to enhance their quality and size. The conditions that resulted in the formation of protein crystals were adjusted and optimized to obtain crystals with improved diffraction quality. This was achieved by varying the pH of the buffer, the concentration of salt, and the concentration of the precipitant solution as required. A Formulator robot (FORMULATRIX®) was used to prepare individual optimized 96-well plates, each with a volume of 40  $\mu$ l. Subsequently, the plates were screened using the same strategy as mentioned in section 2.5.1 for protein crystal growth.

### ➤ **Microbridge, a 24 well plate**

An alternative method for optimizing crystal conditions is the hanging drop technique. For some of the proteins a 24-well hanging drop plate was utilized to facilitate the scaling up of crystallization conditions and promote the formation of larger crystals. For each optimization, a volume of 1 ml of the crystallization solution was prepared in the large well of the hanging drop plate. A polished siliconized coverslip was then placed on top, and 1  $\mu$ l of the solution from the large well and 1  $\mu$ l of the protein were mixed by gently pipetting up and down and stirring. The coverslip was inverted and sealed above the large well using high vacuum grease. The hanging drop plates were subsequently incubated at a temperature of 17°C to allow for water-vapor equilibration and the crystallization process to take place.

### **2.5.3 Mounting Crystals**

To collect data, the crystals were taken out of the solution where they formed and mounted onto loops that matched the crystal size. Under a microscope, the covering tape on the well containing the crystals was removed to expose the drop. A small amount of cryoprotection solution was then added to the adjacent well, which is smaller in size. All crystallization solutions had the same concentration as the components from the crystallization well, with the addition of 20% cryoprotection (ethylene glycol or glycerol).

Using a magnetic wand, the loop was attached and used to carefully capture a single crystal from the drop. This was done in a side-on manner to minimize surface tension. The crystal was briefly immersed in the cryoprotection solution and placed in liquid nitrogen to cool rapidly, and then stored in the data collection pucks in liquid nitrogen. Crystals that were grown in conditions already acting as cryoprotection were directly transferred from the growth drop into liquid nitrogen. Excess cryoprotection solution on the looped crystal was blotted to remove as much as possible. The well was then resealed using more clear tape and returned to the incubation temperature.

#### 2.5.4 Structure Determination

By using X-ray crystallography an electron density distribution can be obtained within a crystal, which is directly correlated to the positions of atoms in the crystal's unit cell and thus the molecular structure of the contents of the unit cell can be determined.

The electron density  $\rho(x, y, z)$  is given by

$$\rho(x, y, z) = \frac{1}{V} \sum_{hkl} F(hkl) \cdot e^{-2\pi i(hx+ky+lz)}$$

where

- $\rho(x, y, z)$  is the electron density at a particular point in the crystal.
- $V$  is the volume of the unit cell.
- The sum is taken over all reflections  $hkl$ , where  $h, k, l$  are the Miller indices of the planes in the crystal lattice.
- $F(hkl)$  is the structure factor for each reflection, which is a complex number that represents both the amplitude and phase of the wave diffracted by the crystal planes with indices  $hkl$ .
- $e^{-2\pi i(hx+ky+lz)}$  is the exponential term representing the wave function, with  $i$  being the imaginary unit, and  $x, y, z$  are the coordinates within the unit cell.

The structure factor  $F(hkl)$  for each reflection ( $hkl$ ) is a vector and can be written as  $F(hkl) \cdot e^{i\alpha}$ .  $F(hkl)$  can be obtained experimentally as its magnitude is proportional to the square root of the intensity,  $I(hkl)$ , of each diffracted ray from the crystal. The associated phase,  $\alpha$ , cannot be measured experimentally and to address this "phase problem" and reconstruct the three-dimensional structure of the protein, various experimental techniques can be employed, such as Multiple Isomorphous Replacement (MIR), Multiple Anomalous Diffraction (MAD), and Molecular Replacement (MR). In this project MR was used because a model for the core SAG-B19 protein was known and was likely to be homologous to the other SAG proteins. By placing the SAG-B19 search model in the correct orientation in the unit cell of the unknown SAG structure, initial phases estimates could be obtained by computational calculations using only the native data set and the coordinates of the search model (Evans and McCoy, 2007).

### 2.5.5 Data collection

The X-ray experiment conducted to determine the structure of a protein necessitates the measurement of intensities for all the diffracted X-rays from the crystal. The crystals were placed in a metal puck and transported to the Diamond Light Source in Oxfordshire, UK, in a liquid nitrogen transport Dewar for subsequent data collection utilizing high-intensity X-rays. At the synchrotron, pucks were mounted on various Diamond beamlines, with all data collected at 100K to minimise radiation damage using the unattended data collection (UDC) mode. Data were collected over a 360° sweep in increments of 0.1° to maximize completeness and multiplicity, with the exposure time automatically determined from test images in the UDC mode.

### 2.5.6 Data processing

The X-ray diffraction data used in this project was automatically processed at the Diamond Light Source facility using the pipelines: Xia2 3dii, Xia2 Dials, and FastDP. Each pipeline employs different programs for the key steps of spot detection, indexing reflections, determining space group symmetry, scaling, merging, and data reduction (Winter *et al.*, 2010).

Specifically, FastDP uses XDS (Kabsch, 2010), for spot finding and indexing, POINTLESS and XDS for space group determination, and XDS, POINTLESS and Scala for scaling, merging and reduction (Evans, 2006; Kabsch, 2010). Xia2 Dials utilizes DIALS (Winter *et al.*, 2018) for spot detection and indexing and POINTLESS and Aimless for downstream processing. Xia2 3dii uses XDS (Kabsch, 2010) for indexing, POINTLESS for space group analysis, and Aimless for processing. Therefore, while the pipelines share some common programs, they employ different combinations of specialized software tools to automatically conduct the diffraction data processing, from initial spot identification through to final scaled and merged datasets. This provides complementary analysis strategies and validation for optimizing the X-ray data quality and interpretation.

For each data set of each SAG crystals, the pipeline was chosen based on the best resolution, completeness,  $R_{\text{pim}}$ ,  $I/\sigma(I)$  and CC half was chosen for subsequent analysis.

### 2.5.7 Molecular replacement (MR)

Molecular replacement is a commonly employed method for phasing when a suitable homologous structure serves as a search model. This approach leverages phase information from a known structure's structure factors to estimate initial phases for a new protein structure (McCoy *et al.*, 2007). In the first stage, the rotation function aligns the search model correctly within the unit cell of the unknown structure. In this analysis, Phaser MR was utilized for rotation function calculations (McCoy *et al.*, 2007). In the second stage, the translation function positions the oriented model correctly to generate phases, which are then used to calculate the initial electron density map of the target structure (Evans and McCoy, 2007). This two-stage method capitalizes on prior structural knowledge to orient and position search models for phasing new crystal data, offering an efficient approach for solving novel protein structures. For each protein whose structure was determined in this project, *Phaser* was used, and successful solutions were assessed using metrics such as TFZ and LLG scores, packing clashes, and resultant density quality (McCoy *et al.*, 2007).

### 2.5.8 Model Building and Refinement

Once phases are determined through molecular replacement, an initial electron density map can be calculated and interpreted to trace the polypeptide chain's backbone. In most cases, Buccaneer, which automatically builds a model of the macromolecule's atomic structure into the electron density map, was used for automated model building before proceeding to manual editing. If much of the model fits the density, refinement can proceed to improve agreement between calculated and observed structure factors. Refinement primarily involves adjusting three positional parameters (x, y, z) and isotropic temperature factors (B) for each non-hydrogen atom to minimize the crystallographic R-factor (R<sub>cryst</sub>), which compares observed and calculated structure factor amplitudes. However, over-reliance on R<sub>cryst</sub> can lead to overfitting, so R<sub>free</sub> is also calculated from a subset of 'free' data excluded from refinement as an independent measure.

In this project, all SAG structures that were solved using phasing by molecular replacement underwent iterative refinement. This refinement process involved the use of programs such as COOT (Emsley and Cowtan, 2004) for manual model building, Buccaneer for automated building, and Refmac5 (Murshudov *et al.*, 2011) for maximise like-hood minimization, which



optimized the model to align with the experimental data based on the R factors. Multiple rounds of building and refinement were performed to generate the final validated structural model.

### 2.5.9 Validation and analysis

Once all interpretable features in the electron density and difference maps were built, which indicating the model was close to completions with acceptable  $R_{\text{work}}$  and  $R_{\text{free}}$ , the model was validated using MolProbity (Williams *et al.*, 2018), to check for steric clashes, rotamer conformations, Ramachandran angles, bond lengths and angles. This generates a validation report and overall score compared to other structures at similar resolution. Riding hydrogens were added, and sidechains flipped to satisfy hydrogen bonding as needed based on validation. If residues or side chains had poor geometry or rotamers these were checked against the electron density, where this was not clear, low energy conformations were chosen. Further rounds of rebuilding and refinement in Coot and Refmac5 (Emsley and Cowtan, 2004; Murshudov *et al.*, 2011) produced the final model. Figures with protein models were made in PyMOL (Schrödinger, and DeLano, 2020), while figures showing density were made in COOT.

## 2.6 Thermofluor assay

In order to measure the stability and possible binding partners for the different SAG proteins a Thermofluor assay was used (Pinz *et al.*, 2022). The technique utilizes a fluorescent dye, commonly SYPRO Orange, which is highly sensitive to environmental changes. In the presence of the protein, the dye remains quenched in an aqueous solution. However, as the protein unfolds and exposes hydrophobic regions, the dye binds to these regions, resulting in a significant increase in fluorescence.

Thermofluor based assays were used to determine the melting temperature of SAGs protein. The SAG proteins were buffer-exchanged into 100 mM Tris-HCl at pH 7.5, 5.0 and 6.0. 50  $\mu$ l assay mixtures of 25  $\mu$ M SAGs, 2.5 mM ligand, 0.1  $\mu$ l of 5000x SYPRO orange solution (SIGMA). A 96-well plate was set up in triplicate for each condition and fluorescence measured on a Stratagene MX3005P RTPCR machine running MXPro software as over a range of 25-98 °C. Melting temperatures ( $T_m$ ) were calculated from the data using excel.

## Chapter3 Cloning, Overexpression, Purification, and Crystallisation of Surface Antigen Protein (SAG).

### 3.1 Overview

This chapter describes the overexpression, purification, and crystallisation trials of surface antigen (SAG) protein in two species of *Eimeria* parasite (*Eimeria tenella* and *Eimeria brunetti*). Eight SAG proteins from family A of *E. tenella* were selected based on the hypothesis that these members exhibit sequence differences compared to SAG-B19, and some possess RRL motifs in a substantial insertion in the sequence alignment with SAG-B19. Additionally, as the SAG B family are very similar to each other, four SAGs from family B of *E. tenella* were chosen that were most different from SAG-B19. Since *E. tenella* has only one SAG-C, this protein was also included in the study with three SAG-C members from *E. brunetti*, included to increase the diversity of this study [Table 3.1].

Table 3.1: SAG proteins of *Eimeria tenella* and *Eimeria brunetti* of different family.

<b><i>Eimeria</i> species</b>	<b>Family</b>	<b>Gene name</b>	<b>Short name</b>
<i>Eimeria tenella</i>	A	SAG 10	SAG-A10
<i>Eimeria tenella</i>	A	SAG 31	SAG-A31
<i>Eimeria tenella</i>	A	SAG 7	SAG-A7
<i>Eimeria tenella</i>	A	SAG 6	SAG-A6
<i>Eimeria tenella</i>	A	ETH 00023375	SAG-A91
<i>Eimeria tenella</i>	A	ETH 00010870	SAG-A870
<i>Eimeria tenella</i>	A	ETH 00034905	SAG-A905
<i>Eimeria tenella</i>	A	SAG 1	SAG-A1
<i>Eimeria tenella</i>	B	SAG 13	SAG-B13
<i>Eimeria tenella</i>	B	SAG 16	SAG-B16
<i>Eimeria tenella</i>	B	SAG 22	SAG-B22
<i>Eimeria tenella</i>	B	SAG 41	SAG-B41
<i>Eimeria tenella</i>	C	ETH 00001975	SAG-C75
<i>Eimeria brunetti</i>	C	EBH 0036680 (EBC1)	EBC1
<i>Eimeria brunetti</i>	C	EBH 0015570 (EBC2)	EBC2
<i>Eimeria brunetti</i>	C	EBH 0027980 (EBC3)	EBC3

### 3.2 Construct Design

As all *E. tenella* SAG proteins contain an N-terminal signal sequence to deliver the SAG to the Endoplasmic Reticulum, which would be most likely cleaved off from the mature protein together with a C-terminal GPI anchor sequence to present the SAG on the outside of the parasite cell membrane, it was decided to focus the structural work on the core, folded component of the SAGs [Figure 3.1]. The full length of each SAG sequence as a part of this study is found in Appendix A.1.



Figure 3.1: **The full length of SAG sequence.** Schematic to show the full length of SAG protein sequence with predicted core domain and the N- and C-terminus.

Previous work on the structure of *E. tenella* SAG proteins by Zaza Ramly in Sheffield (Ramly, 2012) used a construct design where the length of the core SAG protein was estimated from sequence comparisons and predictions of the N-terminal signal sequence and C-terminal GPI anchor sequence. This core SAG was cloned from genomic DNA and placed in a plasmid (pET32b) with a N-terminal-Trx-His-S tag and SAGs-C-terminal-His tag, to aid folding, an enterokinase (EK) protease recognition sequence to allow cleavage of the Trx-tag and a His-tag (at the linker AM residue and the sequence start at AAAPDF) to aid purification [Figure 3.2]. However, there were a number of problems with this construct design including uncertainty over the length of the core folded SAG region, and multiple different enterokinase cleavage sites leading to heterogeneous purified proteins, which appeared to have adverse effects on solubility, purification and crystallisation, with only one SAG (SAG-B19) from the Ramly study leading to a structure from six different target SAGs (Ramly .2012). In order to counteract these problems different strategies were employed for construct design.



Figure 3.2: **The original Trx-construct.** Schematic to show the previous trx-construct that include the non-specific cleavage site (Enterokinase) with an additional his-tag at the C-terminal.

### 3.2.1 Core SAG Length

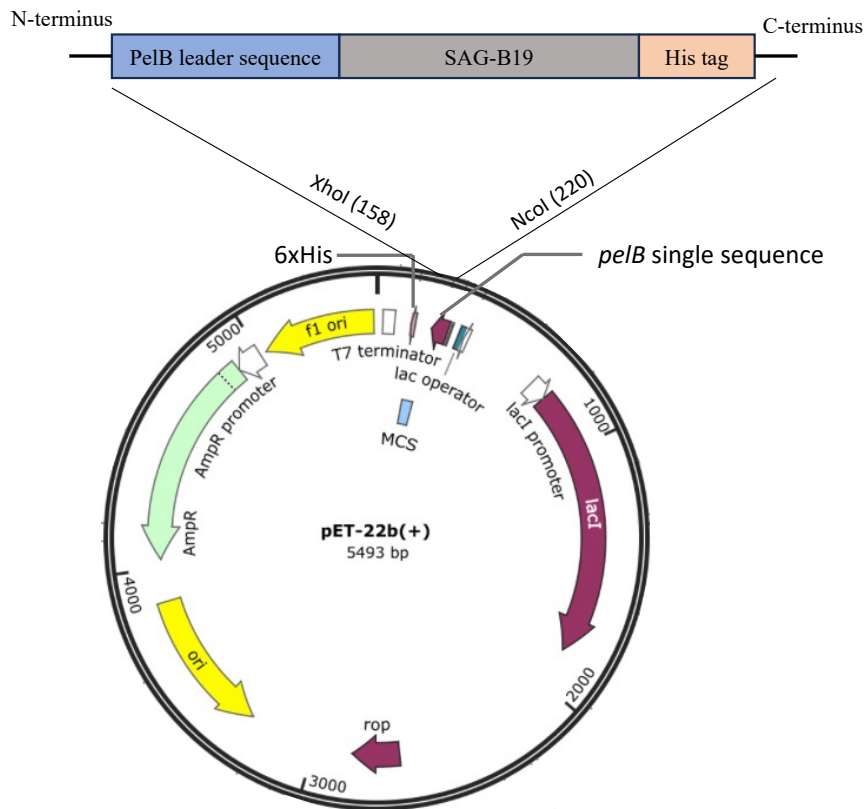
As the structure of SAG19 had been successfully determined using the Ramly protocol, the structure of SAG 19 was inspected to determine how much of the construct had actually folded in the crystal structure. The SAG-B19 construct included the whole predicted core domain (Ala17 – Ala248), but residues Ala17 - Gln53 were not visible in the electron density map at the N-terminus and similarly residues beyond Leu246 were not visible at the C-terminus. Thus, the predicted core domain for each SAG to be expressed was chosen to be equivalent to the folded part of the SAG-B19 structure, with the start and end points of each core SAG protein being equivalent to residues Thr54 – Leu246 identified from a multiple sequence alignment of the SAG proteins.

### 3.2.2 Codon Optimization

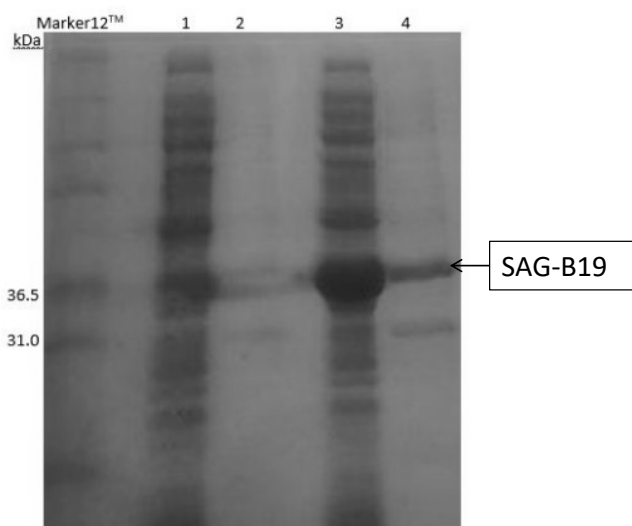
It was also decided to purchase plasmids containing the required SAG insert from Genscript as this would allow optimization of the DNA sequence to use commonly occurring *E. coli* codons, rather than the *E. tenella* genomic codons that would arise from cloning the genes from genomic DNA. Codon optimization has been shown to improve yields of overexpressed proteins (Mauro, 2018).

### 3.2.3 *PelB* expression system

As all *E. tenella* proteins are predicted to contain disulphide bonds, a *PelB* expression system was tested to exploit the natural periplasmic export system of *E. coli*, with the SAG proteins hopefully folding up with correct disulphide formation in the oxidising conditions of the periplasm. SAG-B19 was used as a control for this expression system. A construct of *E. tenella* SAG-B19 was designed, incorporating residues 54–246 of the SAG-B19 protein, starting from the ATG start codon and ending with the TGA stop codon, into the *pET22b* vector via the *NcoI* (CCATGG) and *XhoI* (CTCGAG) restriction sites. This design allowed for the inclusion of an N-terminal *pelB* leader sequence on the mature protein [Figure 3.3].



**Figure 3.3: The *PelB* construct inserted into the *pET-22b* vector.** The *PelB* construct contains *PelB* leader sequence to aid the SAG protein to be correctly folded in the periplasm environment (oxidising conditions), with an his-tag at the C-terminus for purification process.



**Figure 3.4: SDS-PAGE analysis the SAG-B19 overexpression on M9 media at 16°C using the *PelB* expression system.** The first row indicates to the protein marker. Lane-1 SAG-B19 insoluble pre-induction, Lane-2 SAG-B19 soluble pre- induction, Lane-3 SAG-B19 insoluble post- induction, Lane-4 SAG-B19 soluble post- induction with 1mM IPTG.

Using this *PelB* system, SAG19 protein was successfully expressed but accumulated in the insoluble fraction. No clear overexpression of protein in the soluble fraction was observed using LB media, even using different concentrations of IPTG or with a variety of temperatures, in attempts to reduce expression levels to optimise levels in the periplasm. It was assumed that the *PelB* export system was being overloaded, resulting in most protein remaining in the cytoplasm, resulting in incorrect folding as the protein disulphide bonds would not form. To try to improve the solubility of the protein, further experiments using M9 minimal media to slow the growth rate even more did produce a very small amount of soluble SAG19 at 16°C, which corresponds to a tiny band on SDS-PAGE [Figure 3.4]. However, the very small amounts of protein produced could not be purified sufficiently for crystallisation and so this method of protein expression was abandoned.

### 3.2.4 *Trx* expression system in *Origami* cells

As the *PelB* expression system had not proved tractable a different strategy was developed, optimising the original construct design of Ramly. An overexpression construct was designed with a N-terminal thioredoxin (*Trx*-tag), a 6xHis-tag, a TEV cleavage site, and the core SAG sequence at the C-terminal, inserted into *pET32a*, a vector that allows for the production of a fusion protein. Replacing the enterokinase cleavage sequence with a TEV cleavage site was predicted to lead to more homogeneous purified proteins as TEV protease has a higher specificity compared to other cleavage enzymes (Sreejith *et al.*, 2017). This addresses concerns about potential incomplete folding, which could lead to heterogeneities that might interfere with crystallization. Placing the 6xHis-tag at the C-terminus of the *Trx*-tag facilitated the removal of the N-terminal thioredoxin leaving behind the core SAG protein [Figure 3.5].

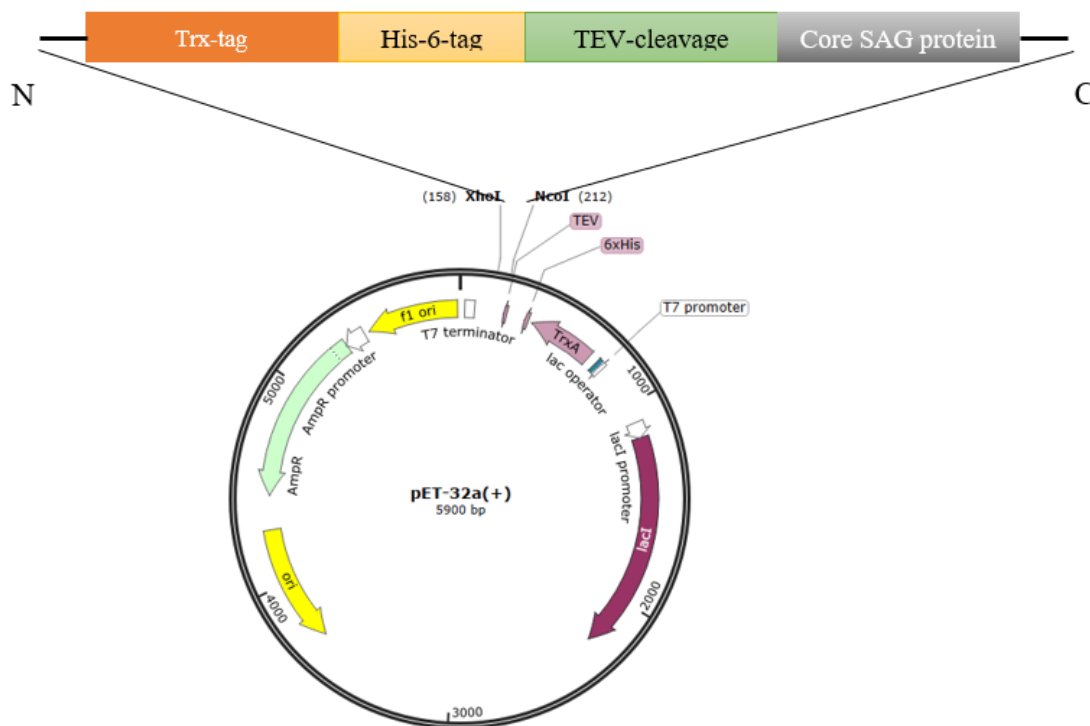


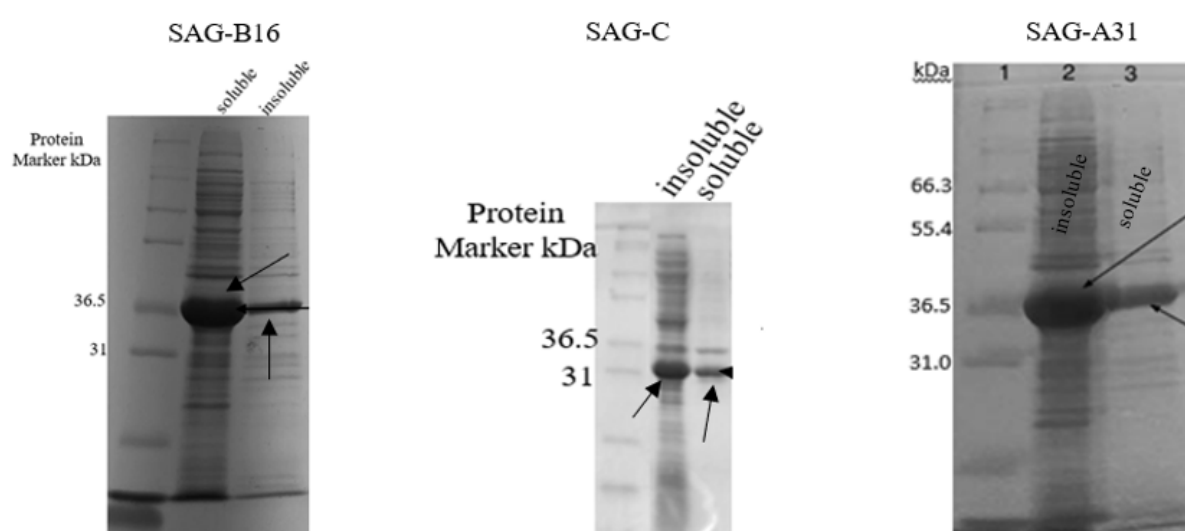
Figure 3.5: *Trx*-construct insertion into *pET32a*. *pET32a* vector containing the *Trx* sequence and *His*-tag sequence. The complete construct, including *Trx*-tag, 6His-tag, TEV cleavage site, and SAGs protein with restriction enzyme sites (*NcoI* and *XhoI*).

### 3.3 SAG Protein Overexpression

Plasmids containing the core sequence of *E. tenella* SAGs were transformed into *E. coli origami* cells (DE3), to facilitate correct disulphide bond formation in the expressed protein. The transformation was performed as described in Chapter 2 (section 2.2.7).

Subsequently, all SAGs were overexpressed in batches of 500 ml LB cultures in 2 L flasks overnight at 17°C, using 1mM IPTG, following the procedures outlined in Chapter 2 (section 2.3.1). The level of overexpression was then examined by cell lysis followed by examining the amount of expressed protein in the soluble and insoluble fractions using SDS/PAGE electrophoresis (section 2.3.4).

Successful overexpression was indicated by the appearance of a band corresponding to the expected molecular weight of each SAG in the soluble fraction on the gel. All the SAG proteins (SAG-B members, SAG-A members and the members of SAG-C family) gave good expression of the soluble form, with an approximate ranged between 50-70 % insoluble to 50-30 % soluble ratio. One example of each family was chosen as representative to show the expression solubility level [Figure 3.6].



**Figure 3.6: SDS-PAGE analysis of one example of each SAG family overexpression were tested at 17°C with 1mM IPTG induction.** SAG-B16 as an example of SAG-B family shows the soluble and insoluble fraction of the protein at 17°C overnight incubated with 1mM IPTG. SAG-C as an example of SAG-C SAG family shows the soluble and insoluble fraction of the protein at 17°C overnight incubated with 1mM IPTG. SAG-A31 as an example of SAG-A family shows the soluble and insoluble fraction of the protein at 17°C overnight incubated with 1mM IPTG.



### 3.4 Purification of the SAG proteins

The purification of each SAG protein followed the same protocol as detailed below:

Step 1: Affinity chromatography using a Ni-NTA column.

Affinity chromatography was used as the first step in the purification process, exploiting the His6 tag between the N-terminal Trx and the TEV cleavage site of the construct. The cell paste was lysed by sonication and following centrifugation, cell free extract (CFE) was checked for protein concentration before being applied to a Ni-NTA column (His-Trap<sup>TM</sup>HP cartridge, 5 ml), which was pre-equilibrated with 50 mM Tris pH 8.0 and 0.5 M NaCl (Buffer A). The protein was eluted using 0.5 M imidazole in 50 mM Tris pH 8.0 and 0.5 M NaCl (Buffer B). The protein concentrations of selected fractions were estimated using the Bradford method (Bradford, 1976). Subsequently, the fractions were further analysed by SDS-PAGE, and those containing SAGs proteins were pooled prior to cleavage of the Trx tag using TEV protease.

Step 2: Cleavage with TEV protease (Tobacco Etch Virus)

SAGs were cleaved overnight using TEV protease at room temperature (22°C) using 50  $\mu$ g of TEV per 1 mg of protein into buffer A twofold. The expected cleavage occurred at the Ser of the [Glu-Asn-Leu-Tyr-Phe-Gln-(Gly/Ser)] recognition sequence. The extent of the cleavage was estimated using SDS-PAGE.

Step 3: Ni-NTA 2

To remove un-cleaved protein, the N-terminal-Trx-His-tag and His-tagged TEV protease from the core SAG protein a second Ni-NTA (1 ml His-Trap column) was used. The protein was loaded onto a 1 ml His-Trap column that was pre-equilibrated with buffer A. After loading, the unbound target core SAG protein was washed out using buffer A and the concentration was estimated using the Bradford method (Bradford, 1976). Then the column was further eluted with buffer B. Both the unbound and bound fractions were checked and analysed by SDS-PAGE.

#### Step 4: Hi Load™ Superdex™ 200 gel filtration

The core SAG proteins were concentrated to 2 ml using a VivaSpin concentrator fitted with a 10 kDa cut-off filter. The proteins were then loaded onto a 16x60 Hi-Load™ Superdex™200 column (GE Healthcare) pre-equilibrated with 50 mM Tris-HCl pH 8.0 and 0.5 M NaCl (buffer A). After loading, the protein was eluted using the same buffer, and the concentration of the selected fraction containing SAGs was estimated (Bradford, 1976). The combined fraction containing SAGs was further concentrated using a VivaSpin concentrator fitted with a 10 kDa cut-off filter, and the buffer was exchanged to a crystallization buffer (buffer A diluted tenfold) using a diafiltration cup to reach the target concentration, which fell between 20-25 mg/ml. The concentration of the final protein was checked using an IMPLEN P300 Nanophotometer and further analysed by SDS-PAGE. The details for each of the Sag proteins are given below.

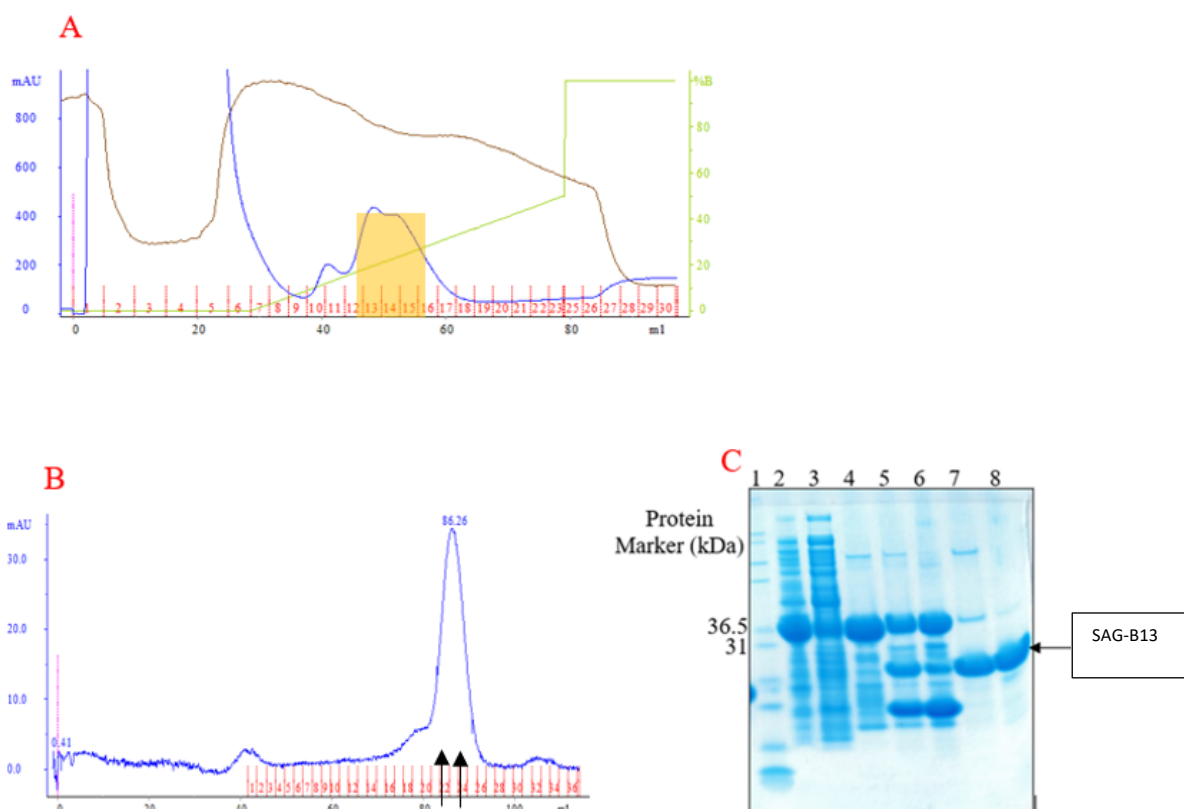
### 3.5 SAG-B Family

*Table 3.2: Four SAG-B family proteins were chosen for overexpression.*

B-family SAG	Construct mw (kDa)	Core SAG mw (kDa)	Core SAG residues	Sequence identity to SAG19 core (%)
SAG-B13	34.5	19.7	Ala52-Gln239	40
SAG-B16	35.3	20.7	Ala50-Arg243	44
SAG-B22	35.7	20.9	Ser51-Gly248	63
SAG-B41	35.8	21.3	Leu51-Gly249	72

#### 3.5.1 Surface Antigen 13 (SAG-B13)

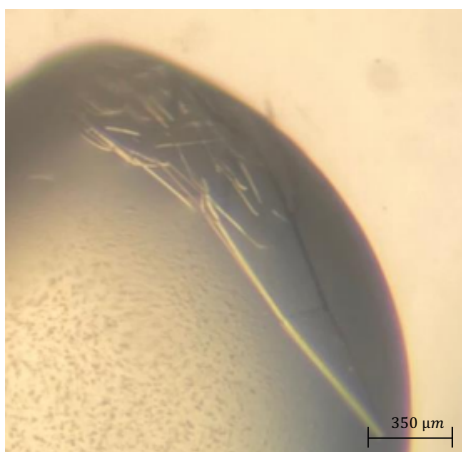
SAG-B13 was selected for study as it only shared 40% sequence identity with the control SAG (SAG-B19), highlighting a notable difference within this otherwise highly similar family. The 34.5 kDa SAG-B13 protein was overexpressed in a 2 L culture (4x 500 ml flasks) as described section 2.3.1. Approximately 30% of the protein was in the soluble fraction, that indicated as CFE, which contained 120 mg of total protein [Figure 3.6]. The final preparation of SAG-B13, produced ~78  $\mu$ l at a concentration of 21.6 mg/ml in buffer C (5 mM Tris PH 8.0, 50 mM NaCl), which is suitable for the targeted protein crystallization [Figure 3.7].



**Figure 3.7: *SAG-B13* purification process.** A) 280 nm absorption trace for the affinity chromatography of *SAG-B13* on 5 ml His-Trap<sup>TM</sup> HP cartridge (GH Healthcare). 120 mg of CFE was applied to the column; Flow rate- 5 ml/min; 50 ml gradient 0 to 0.25 M imidazole in buffer A; Fraction 3 ml. B) 280 nm absorption trace for the gel filtration of *SAG-B13* on a 1.6x60 Hi-load<sup>TM</sup> Superdex<sup>TM</sup> 200 column (GH Healthcare). 5.2 mg of *SAG-B13*; buffer 50 mM Tris-HCl, PH 8.0 and 0.5 M NaCl; Flow rate- 1.5 ml/min; fraction- 2 ml. C) SDS-PAGE analysis of *SAG-B13* after purification steps. Lane-1 protein marker (Mark12<sup>TM</sup>); Lane-2 cells debris; Lane-3 cell-free extract of *SAG-B13*; Lane-4 *SAG-B13* before cleavage (after Ni-NTA 1); Lane-5 *SAG-B13* after cleavage; Lane-6 Ni-NTA 2 elution (imidazole peak fraction); Lane-7 gel filtration loading sample; Lane-8 final prep of *SAG-B13*.

### ❖ Crystallisation of *SAG-B13*

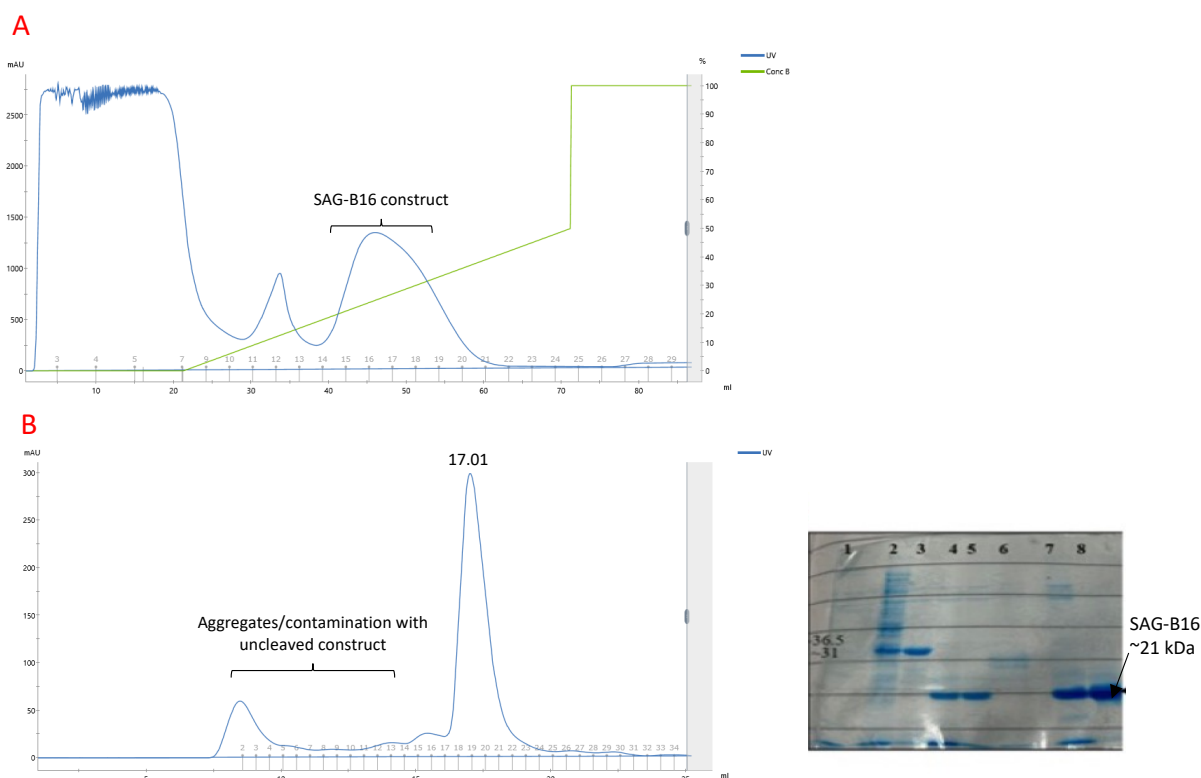
The *SAG-B13* sample was used in four different crystallisation screens (PACT, JCSG, PH Clear, and AmSO<sub>4</sub> screens), as described in Chapter 2, Part II. Crystals were observed in the JCSG A12 condition (0.2 M Potassium nitrate PH 6.9 + 20% w/v PEG 3350) after two weeks [Figure 3.8], reaching a size sufficient for diffraction analysis at the Diamond Synchrotron. Several crystals were carefully mounted using the appropriate loop and immersed in cryoprotectant (20 % ethylene glycol + A12 mother solution) before being cooled to 100K in liquid nitrogen. These crystals were then sent to the Diamond Synchrotron under the code MX31850 trip number 4, using beamline I03.



*Figure 3.8: **Crystal of SAG-B13.** The grown crystal on JCSG A12 condition (0.2 M Potassium nitrate PH 6.9 + 20% w/v PEG 3350).*

### **3.5.2 Surface Antigen 16 (SAG-B16)**

The second B-family SAG was SAG-B16, which again had a low sequence identity to SAG-B19 (44%). Approximately 75% of the expressed SAG-B16 protein (from 4xb500 ml cultures) was soluble, facilitating its purification. For the purification process, half the CFE (182 mg) was loaded onto the first Ni-NTA column at a flow rate of 5 ml/min and a 50 ml gradient of 0 to 0.25 M imidazole in buffer A was used for elution, collecting fractions every 3 ml. Gel filtration followed by buffer exchange (into buffer C) and concentration yielded ~75  $\mu$ l of 23.2 mg/ml protein for crystallisation [Figure 3.9],



**Figure 3.9: SAG-B16 purification process.** A) 280 nm absorption trace for the affinity chromatography of SAG-B16 on 5 ml His-Trap<sup>TM</sup>HP cartridge (GH Healthcare). 182 mg of CFE was applied to the column; Flow rate- 5 ml/min; 50 ml gradient 0 to 0.25 M imidazole in buffer A; Fraction 3 ml. 280 nm absorption trace for the gel filtration of SAG-B16 on a 1.6x60 Hi-load<sup>TM</sup> Superdex<sup>TM</sup> 200 column (GH Healthcare). 3.2 mg of SAG-B16 was loaded into the column and the collected 2 ml fraction eluted with buffer A; Flow rate- 1.5ml/min. SDS-PAGE analysis of SAG-B16 after purification steps. Lane-1 protein marker (Mark12<sup>TM</sup>); Lane-2 CFE of SAG-B16; Lane-3 SAG-B16 before cleavage (after Ni-NTA); Lane-4 and 5 SAG-B16 after cleavage; Lane- 6 Ni-NTA 2 elution (imidazole peak fraction); Lane-7 gel filtration loading sample; Lane-8 final prep of SAG-B16.

### ❖ Crystallization of SAG-B16

The same four crystallisation screens were used to screen for SAG-B16 crystal growth. Crystals grew in a number of different conditions (PACT C11, B11, B10, F6 and F3 conditions, JCSG D6 and D9 conditions and AmSO<sub>4</sub> A9, C10, D7, D12 and H2 conditions) [Figure 3.10]. Crystals were cryoprotected in solutions (20 % ethylene glycol + mother solution of each condition) and sent to the Diamond Synchrotron under code MX24447-trip number 73 and 93 at beamline i03 for data collection.

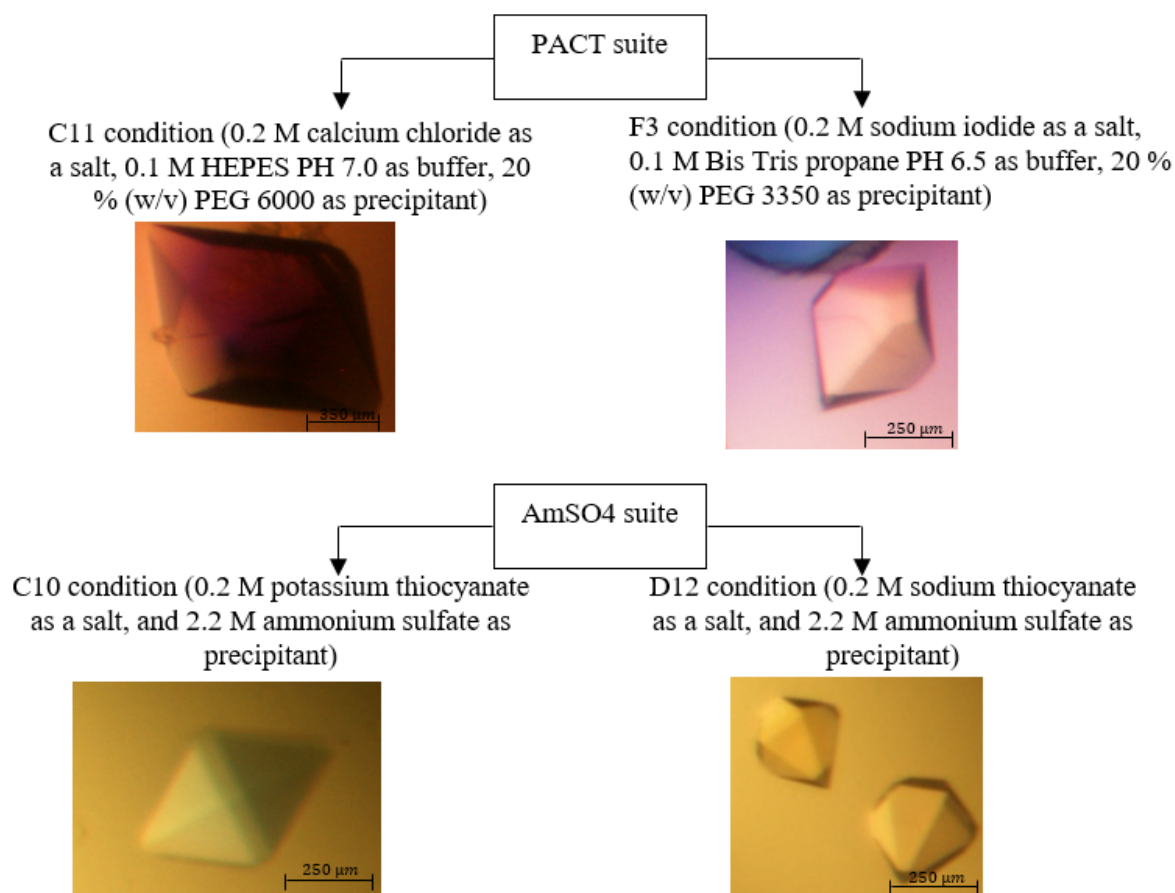
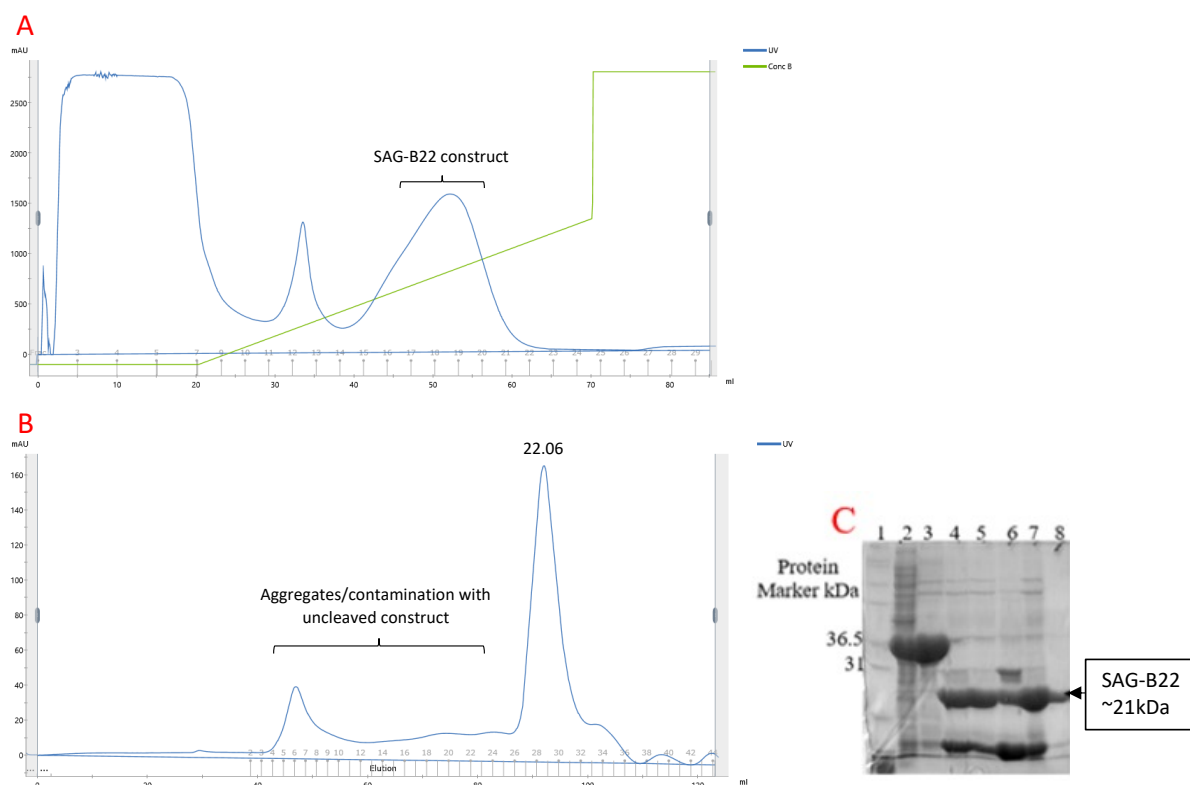


Figure 3.10: **Crystals of SAG-B16.** The crystals grown on different condition of AmSO<sub>4</sub> (C110 and D12) and PACT (C11 and F3) suites.

### 3.5.3 Surface Antigen 22 (SAG-B22)

The third SAG B-family protein was SAG-B22, which had a much closer sequence identity to SAG-B19 (63%). 6x 500 ml cultures were used in the overexpression to produce 210 mg of CFE. Following purification, the SAG-B22 sample was concentrated to 17 mg/ml in buffer C for crystallisation [Figure 3.11].

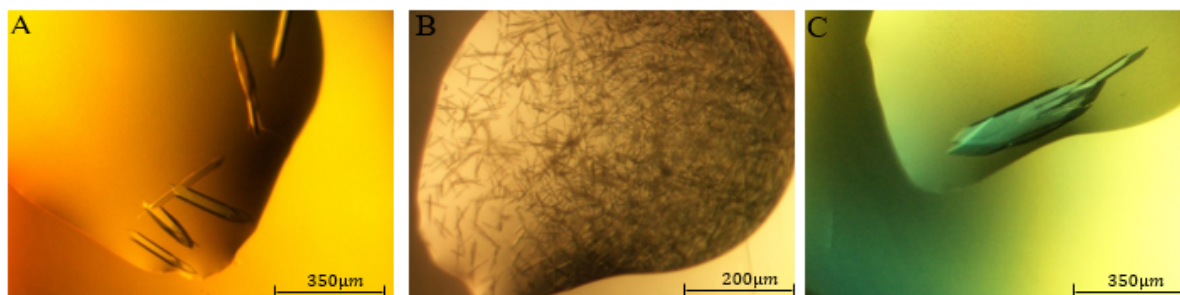


**Figure 3.11: *SAG-B22* purification process.** A) 280 nm absorption trace for the affinity chromatography of *SAG-B22* on 5 ml His-Trap<sup>TM</sup>HP cartridge (GH Healthcare). 210 mg of CFE was applied to the column; Flow rate- 5 ml/min; 50 ml gradient 0 to 0.25 M imidazole in buffer A; Fraction 3 ml. B) 280 nm absorption trace for the gel filtration of *SAG-B22* on a 1.6x60 Hi-load<sup>TM</sup> Superdex<sup>TM</sup> 200 column (GH Healthcare). 9 mg of *SAG-B22* was loaded; buffer A was used to elute the 2 ml collected fraction; Flow rate- 1.5ml/min. C) SDS-PAGE analysis of *SAG-B22* after purification steps. Lane-1 protein marker (Mark12<sup>TM</sup>); Lane-2 CFE of *SAG22*; Lane-3 *SAG-B22* before cleavage (after Ni-NTA 1); Lane-4 *SAG-B22* after cleavage; Lane-5 unbound material (*SAG* after Ni-NTA 2); Lane- 6 Ni-NTA 2 elution (imidazole peak fraction); Lane-7 gel filtration loading sample; Lane-8 final prep of *SAG-B22*.

### ❖ Crystallization of *SAG-B22*

Following crystallisation using the same four screens (JCSG, PH-Clear, AmSO<sub>4</sub> and PACT suites), crystals were observed in PACT A2, B1 and B2 conditions. The A2 condition (0.1 M SPG PH 5.0 + 25% (w/v) PEG 1500) gave the possible crystals that was used for further optimized to improve the quality of the crystal [Figure 3.12]. A second 96 well plate was set up using the Formulatrix Robot (FORMULATRIX<sup>®</sup>), by varying the PH of the SPG buffer (ranged from 5-8) and the PEG 1500 concentration (20, 25, 30, and 35%). The crystals of A2, B1 and B2 were carefully mounted using the appropriate loop and immersed in cryoprotectant

(20 % ethylene glycol + mother solution of each condition) and were sent to the Diamond Synchrotron under code MX24447 trip number 84, on beamline i03.

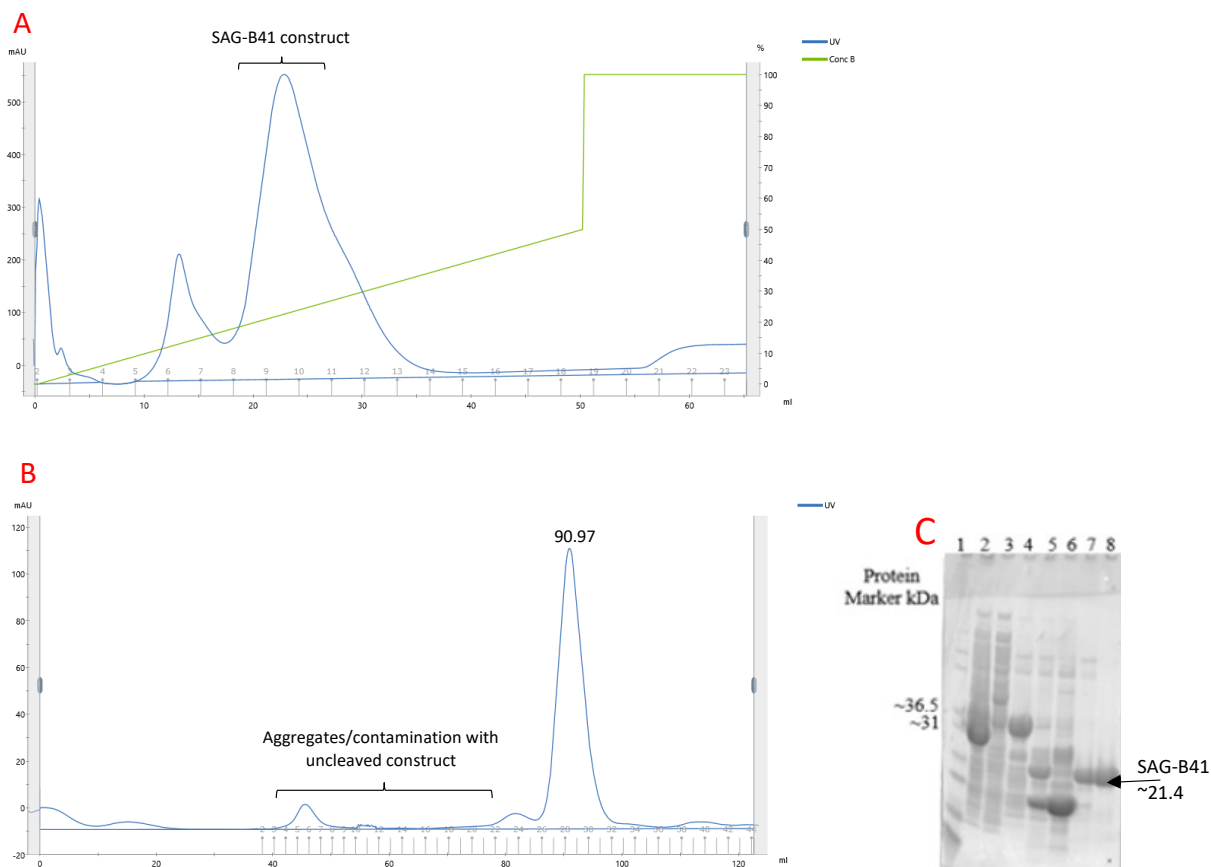


*Figure 3.12: Crystals of SAG-B22 grown in different condition from PACT suite. A) crystal of B2 condition (0.1 M MIB PH 5.0 + 25% (w/v) PEG 1500). B) crystal of A2 condition (0.1 M SPG PH 5.0 + 25% (w/v) PEG 1500). C) crystal of B1 condition (0.1 M MIB PH 4.0 + 25% (w/v) PEG 1500).*

#### 3.5.4 Surface Antigen 41 (SAG-B41)

The fourth B-family Sag to be chosen was SAG-B41 as it has 72% sequence identity to SAG-B19. The idea was to determine a similar structure to SAG-B19 to see if the disordered residues on SAG-B19 can be observed in the SAG-B41 structure. The overexpression of this SAG used 6x 500ml cultures to produce 165 mg of CFE. 50 % of the protein was in the soluble fraction and following purification, gel-filtration and buffer exchange ~71µl of 23 mg/ml SAG-B41 was prepared in buffer C for crystallisation [Figure 3.13].





**Figure 3.13: *SAG-B41* purification process.** A) 280 nm absorption trace for the affinity chromatography of *SAG-B41* on 5 ml His-Trap<sup>TM</sup> HP cartridge (GH Healthcare). 165 mg of CFE was applied to the column; Flow rate- 5 ml/min; 50 ml gradient 0 to 0.25 M imidazole in buffer A; Fraction 3 ml. B) 280 nm absorption trace for the gel filtration of *SAG-B41* on a 1.6x60 Hi-load<sup>TM</sup> Superdex<sup>TM</sup> 200 column (GH Healthcare). 3 mg of *SAG-B41* was loaded; buffer A was used to elute the 2 ml collected fraction; Flow rate- 1.5ml/min. C) SDS-PAGE analysis of *SAG41* after purification steps. Lane-1 protein marker (Mark12<sup>TM</sup>); Lane-2 cell debris Lane-3 CFE of *SAG-B41*; Lane-4 *SAG-B41* before cleavage (after Ni-NTA); Lane-5 *SAG-B41* after cleavage; Lane-6 Ni-NTA 2 elution (imidazole peak fraction); Lane-7 gel filtration loading sample; Lane-8 final prep of *SAG-B41*.

### ❖ Crystallization of *SAG-B41*

Using the same crystallisation suites, crystals were observed after a week in PH-Clear D9 (0.1M MES PH 6.0 + 30% (w/v) PEG600) and D12 conditions (0.1M Bicine PH 9.0 + 30% (w/v) PEG6000), as well as in the AmSO<sub>4</sub> H8 condition (3.0M Ammonium sulphate + 1% (w/v) MPD). The crystals were immersed in the cryoprotectant (20 % ethylene glycol + mother solution of each condition) mounted, cooled to 100k, and sent to the Diamond Synchrotron under the same code and the same beamline as *SAG-B22* [Figure 3.14].

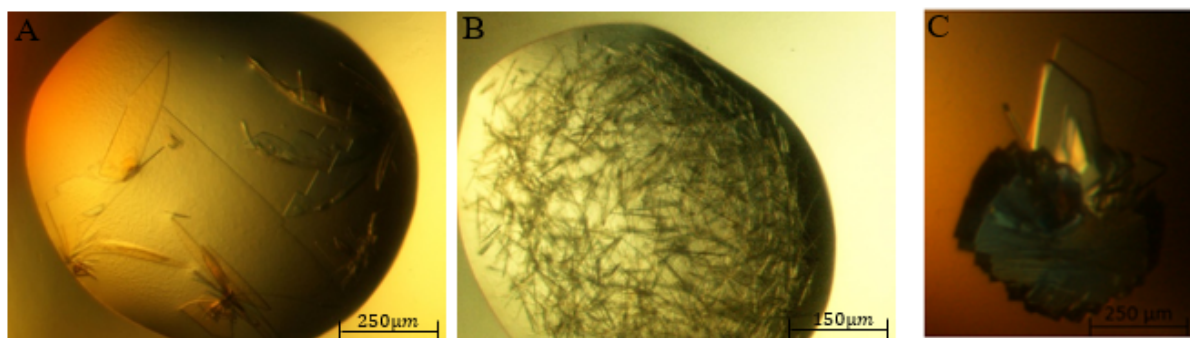


Figure 3.14: **Grown crystals of SAG-B41.** A) crystal from D9 condition (0.1M MES PH 6.0 + 30% (w/v) PEG600). B) crystal from H8 condition (3.0M Ammonium sulphate + 1% (w/v) MPD). C) crystal from D12 condition (0.1M Bicine PH 9.0 + 30% (w/v) PEG6000).

### 3.6 SAG-C family

One of the objectives of this study is to analyse the structural differences among the three families of SAG proteins within the *Eimeria* parasite. This may lead to an understanding of the important roles these proteins play, whether they are involved in invasion or disrupting the immune system of the chicken. *Eimeria tenella* has only one SAG-C protein, which has five exons and two disulphide bonds that distinguish it from the A and B families of SAG proteins (Reid *et al.* 2014). Based on this, this particular SAG protein was chosen for study along with another three Sag-C family proteins from *Eimeria brunetti*, this *Eimeria* species shows the most SAG-C family proteins. As the majority of the SAG-C family proteins in *E. brunetti* have about 26-32% identity to the *E. tenella* SAG-C it was not possible to choose the closest or furthest relative to SAG-C75 and the selection was made randomly based on the most differences [Table 3.3].

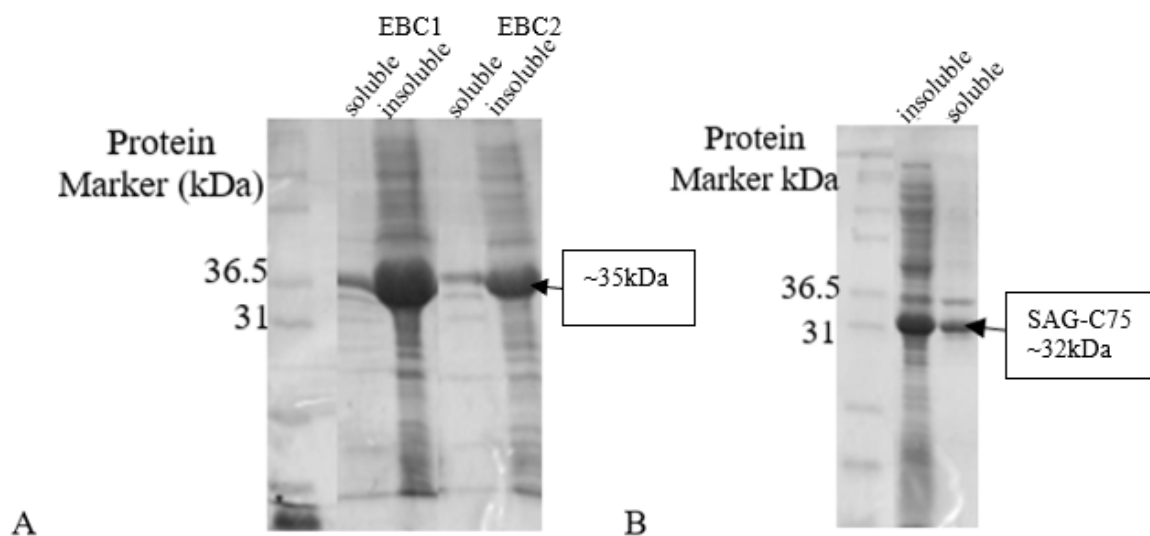
Table 3.3: Details fo333r SAG-C family proteins used in this study.

C-family SAG	Construct mw (kDa)	Core SAG mw (kDa)	Core SAG residues	Sequence identity to SAG-19 core (%)	Sequence identity to SAG-C75 (%)
<i>E. tenella</i> SAG-C75	32.4	18	Ser30-Glu200	25	100
<i>E. brunetti</i> EBC1	35.8	21.4	Ala24-Ala220	22	29
<i>E. brunetti</i> EBC2	35.8	21.4	Ala22-Ala218	19	26

<i>E. brunetti</i> EBC3	36.2	21.8	Ala25-Leu226	21	30
-------------------------	------	------	--------------	----	----

### 3.6.1 SAG-C family Overexpression

Overexpression of the SAG-C proteins used the same construct design as that for the SAG-B proteins (section 3.2.4), but the amount of soluble protein expressed was lower (20-25 %) that that observed for the SAG-B family proteins [Figure 3.15, Table 3.4]. Nevertheless, using larger culture volumes, sufficient soluble protein was present in the cell free extract to enable purification of SAG-C75, SAG-EBC1, SAG-EBC2 and SAG-EBC3 using the procedure described above (section 3.4) [Figure 3.16 – 3.19]. For SAG-EBC2 and SAG-EBC3, where the amount of soluble expressed protein was lowest (20%) an effective strategy was implemented by co-expressing chaperone systems with these proteins, attempting to aid in improving protein folding and solubility. Although the solubility of these two proteins was still lower compared to the previous SAGs some purified proteins was obtained [Table 3.4, Figure 3.15].



**Figure 3.15: SDS-PAGE analysis of SAG-C proteins overexpression which was tested at 18°C O/N with 1mM IPTG.** A) The gel lanes started with protein marker; soluble fraction of EBC1 followed with insoluble fraction; afterword soluble fraction of EBC2 followed with insoluble fraction at the expected molecular wight 36 kDa. B) showed the SAG-C75 expression which was around 70% insoluble and 30% soluble.

Table 3.4: Overexpression and Purification Summary for Sag-C family proteins

SAG	Culture Volume	Expression Temperature	Approximate soluble expression (%)	Final Purified Protein		Crystals	Data
				Volume (ul)	Concentration (mg/ml)		
SAG-C75	6 x 500 ml	17 C°	35	50-60	20.7	✓	✓
EBC1	8 x 500 ml	17 C°	30	45-55	16.5	✓	weak
EBC2	8 x 500 ml	17 C°	20	35-40	14.6	No	
EBC3	8 x 500 ml	17 C°	20	40-45	9.2	No	

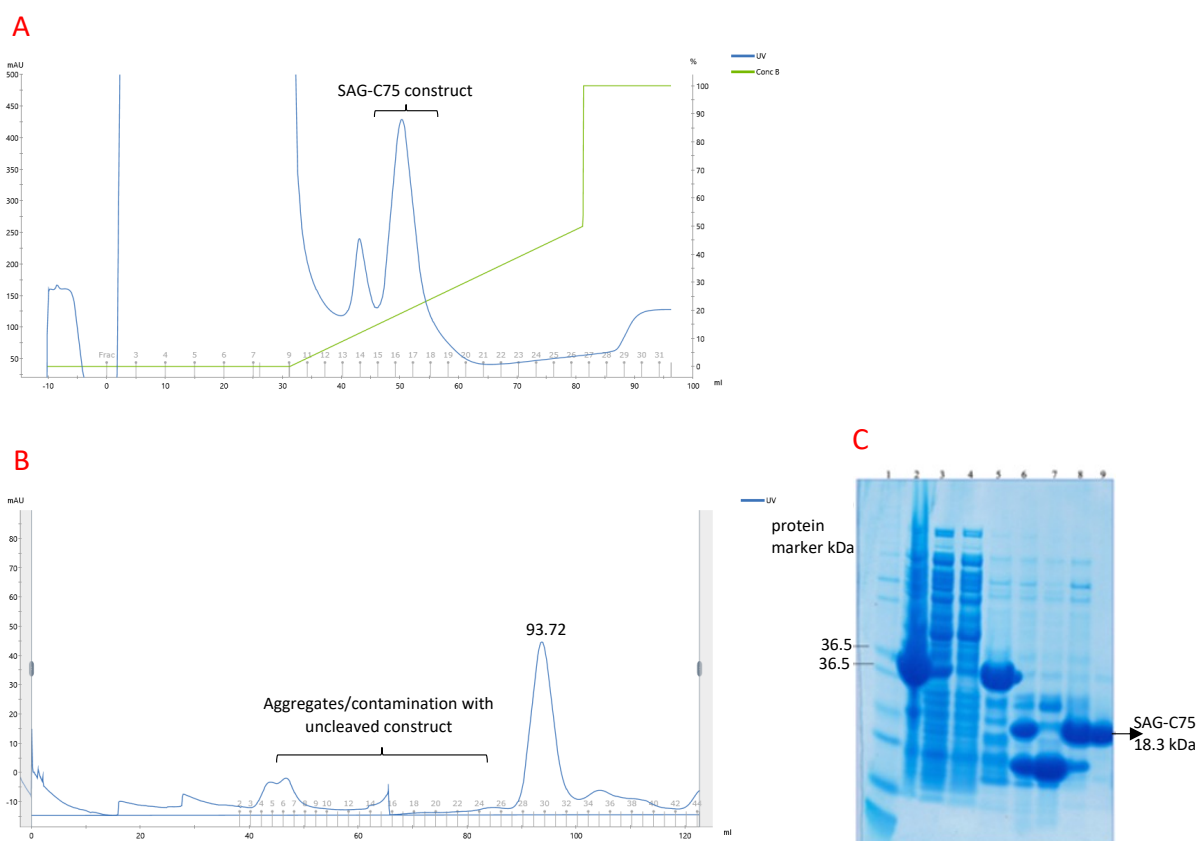
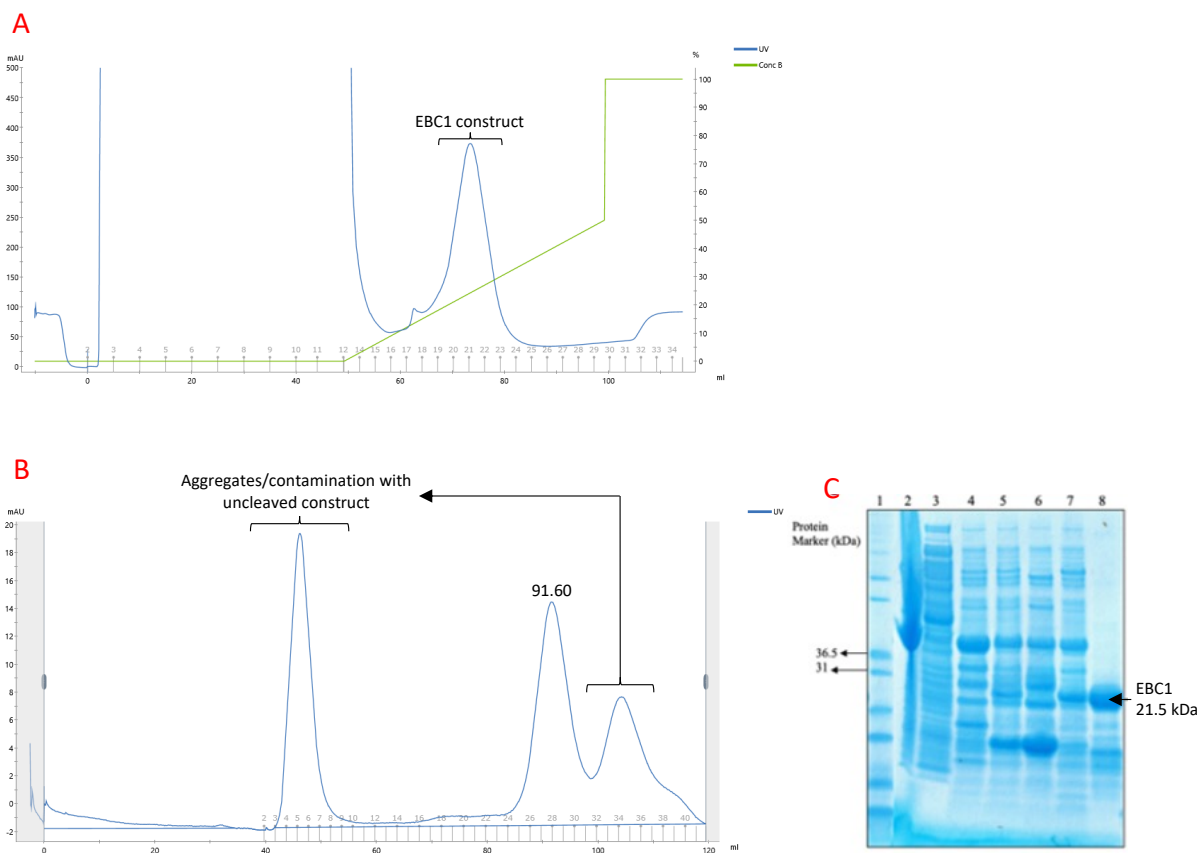
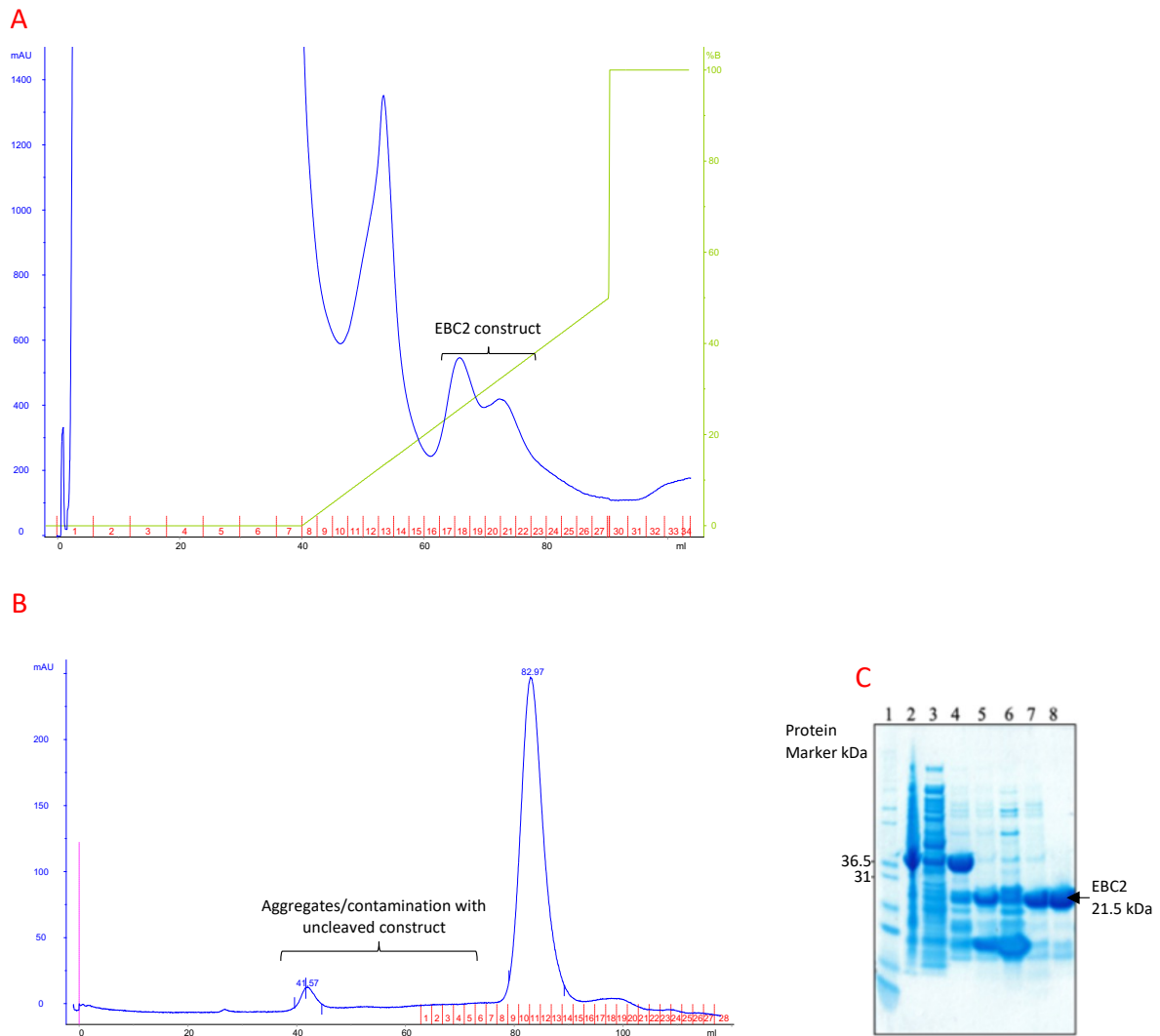


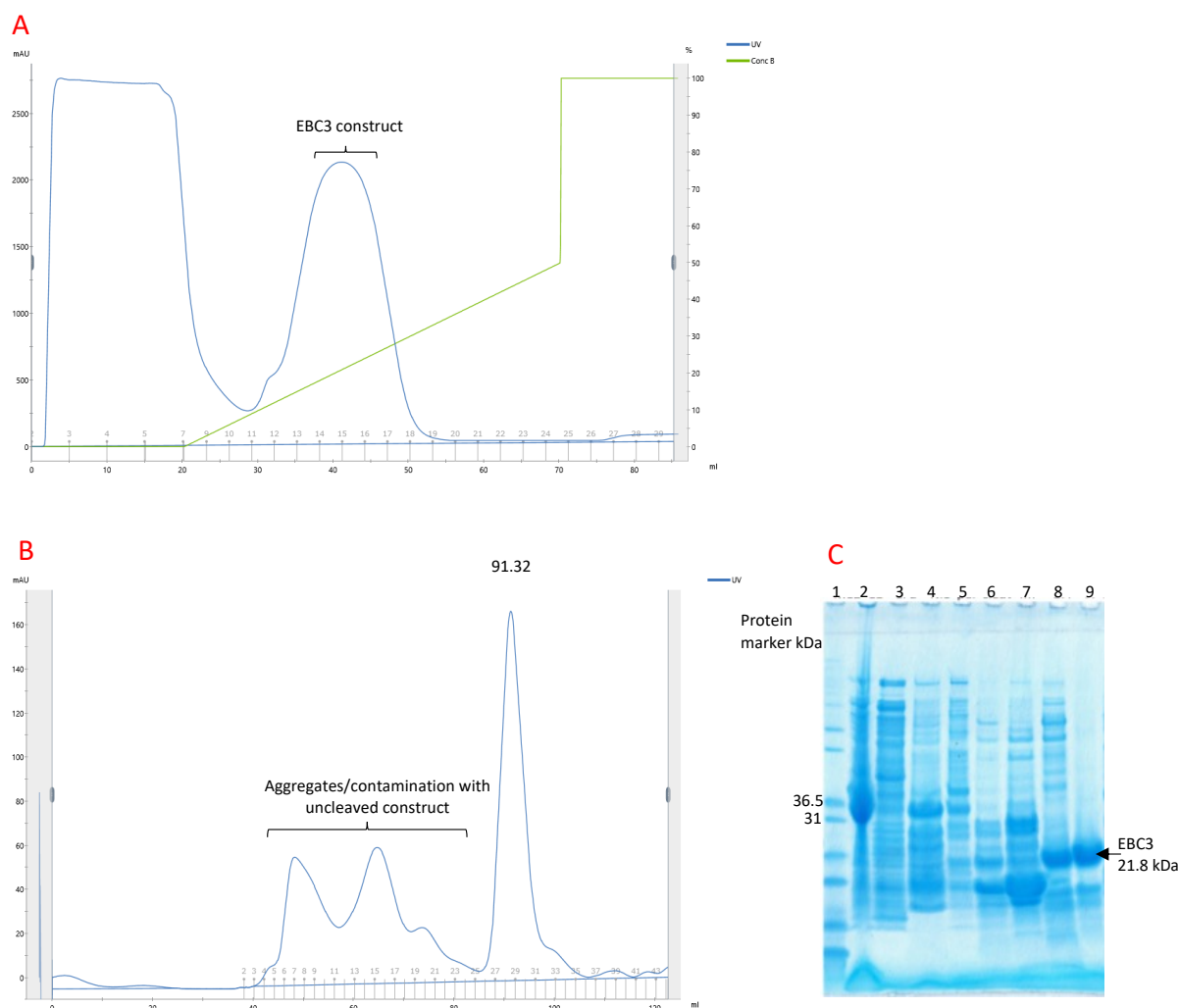
Figure 3.16: **SAG-C75 purification processes.** A) 280 nm absorption trace for the affinity chromatography of SAG-C75 on 5 ml His-Trap<sup>TM</sup> HP cartridge (GH Healthcare). 200 mg of CFE was loaded to the column; Flow rate- 5 ml/min; 50 ml gradient 0 to 0.25 M imidazole in buffer A; Fraction 3 ml. B) 280 nm absorption trace for the gel filtration of SAG-C75 on a 1.6x60 Hi-load<sup>TM</sup> Superdex<sup>TM</sup> 200 column (GH Healthcare). 11 mg of SAG-C75 was loaded to the column; buffer A was used to elute the 2 ml collected fractions; Flow rate- 1.5ml/min. C) SDS-PAGE analysis of cleaved SAG-C75 after purification steps. Lane-1 protein marker (Mark12TM); Lane-2 cells debris; Lane-3 and 4 CFE of SAG-C75; Lane-5 SAG-C75 before cleavage (after Ni-NTA 1); Lane-6 SAG-C75 after cleavage; Lane- 7 Ni-NTA 2 elution (imidazole peak fraction); Lane-8 gel filtration loading sample; Lane-9 final prep of SAG-C75.



**Figure 3.17: EBC1 SAG purification processes.** A) 280 nm absorption trace for the affinity chromatography of EBC1 on 5 ml His-Trap<sup>TM</sup> HP cartridge (GH Healthcare). 450 mg of CFE was loaded to the column; Flow rate- 5 ml/min; 50 ml gradient 0 to 0.25 M imidazole in buffer A; Fraction 2.5 ml. B) 280 nm absorption trace for the gel filtration of EBC1 on a 1.6x60 Hi-load<sup>TM</sup> Superdex<sup>TM</sup> 200 column (GH Healthcare). 5 mg of EBC1 was loaded to the column; buffer A was used to elute the 2 ml collected fraction; Flow rate- 1.5 ml/min. C) SDS-PAGE analysis of cleaved EBC1 after purification steps. Lane-1 protein marker (Mark12TM); Lane-2 cells debris; Lane-3 CFE of EBC1; Lane-4 EBC1 before cleavage (after Ni-NTA 1); Lane-5 EBC1 after cleavage; Lane-6 Ni-NTA 2 elution (imidazole peak fraction); Lane-7 gel filtration loading sample; Lane-8 final prep of EBC1.



**Figure 3.18: EBC2 SAG purification processes.** A) 280 nm absorption trace depicts the affinity chromatography process using a 5 ml His-Trap<sup>TM</sup> HP cartridge (GE Healthcare). 130 mg of EBC2 CFE was loaded onto the column with a flow rate of 5 ml/min. A 50 ml gradient ranging from 0 to 0.25 M imidazole in buffer A was employed, and fractions of 3 ml were collected. B) 280 nm absorption trace shows the gel filtration process using a 1.6x60 Hi-load<sup>TM</sup> Superde<sup>TM</sup> 200 column (GE Healthcare). 4 mg of EBC2 was loaded to the column, and buffer A was used as eluted buffer for the 2 ml collected fraction. C) SDS-PAGE analysis of cleaved EBC2 after purification steps. Lane-1 protein marker (Mark12TM); Lane-2 cells debris; Lane-3 CFE of EBC2; Lane-4 EBC2 before cleavage (after Ni-NTA 1); Lane-5 EBC2 after cleavage (before Ni-NTA 2); Lane-6 Ni-NTA 2 elution (imidazole peak fraction); Lane-7 gel filtration loading sample; Lane-8 final prep of EBC2.

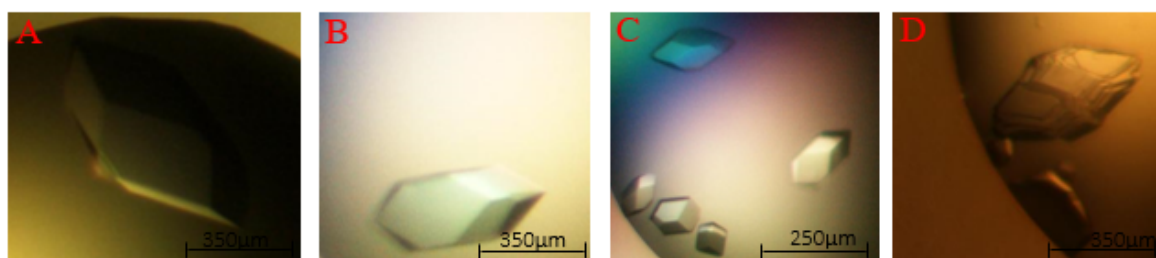


**Figure 3.19: EBC3 SAG purification processes.** A) 280 nm absorption trace depicts the affinity chromatography process using a 5 ml His-Trap<sup>TM</sup> HP cartridge (GE Healthcare). 200 mg of EBC3 CFE was loaded onto the column with a flow rate of 5 ml/min. A 50 ml gradient ranging from 0 to 0.25 M imidazole in buffer A was employed, and fractions of 3 ml were collected. B) 280 nm absorption trace shows the gel filtration process using a 1.6x60 Hi-load<sup>TM</sup> Superde<sup>TM</sup> 200 column (GE Healthcare). 4.3 mg of EBC3 was loaded to the column, and buffer A was used to elute the 2 ml collected fraction. C) SDS-PAGE analysis of cleaved EBC3 after purification steps. Lane-1 protein marker (Mark12<sup>TM</sup>); Lane-2 cells debris; Lane-3 CFE of EBC3; Lane-4 EBC3 before cleavage (after Ni-NTA 1); Lane-5 unbound protein; Lane-6 EBC3 after cleavage (before Ni-NTA 2); Lane- 7 Ni-NTA 2 elution (imidazole peak fraction); Lane-8 gel filtration loading sample; Lane-9 final prep of EBC3.

### 3.6.2 Crystallization of SAG-C Family Proteins

#### 3.6.2.1 SAG-C75

The purified protein at a concentration 20.7 mg/ml in buffer C (5 mM Tris PH 8.0, 50 mM NaCl) was used for crystallisation using the JCSG, AmSO<sub>4</sub>, PACT and PH-Clear crystallization suites. Crystals grew in five days in several conditions of PH-Clear suite {F11 condition (0.1 M Tris PH 8.0 + 3.2 M ammonium sulphate), F4 condition (0.1 M HEPES PH 7.0 + 2.4 M ammonium sulphate), F3 condition (0.1 M MES PH 6.0 + 2.4 ammonium sulphate), F6 condition (0.1 M Bicine PH 9.0 + 2.4 ammonium sulphate)}, which these were the best crystals from the PH-Clear suite conditions [Figure 3.20] were mounted, immersed in a cryoprotectant (20 % ethylene glycol + mother solution of each condition) before cooling in liquid nitrogen. The crystals were sent to the Diamond Synchrotron under code MX24447 trip number 96 on beamline i04 for data collection.



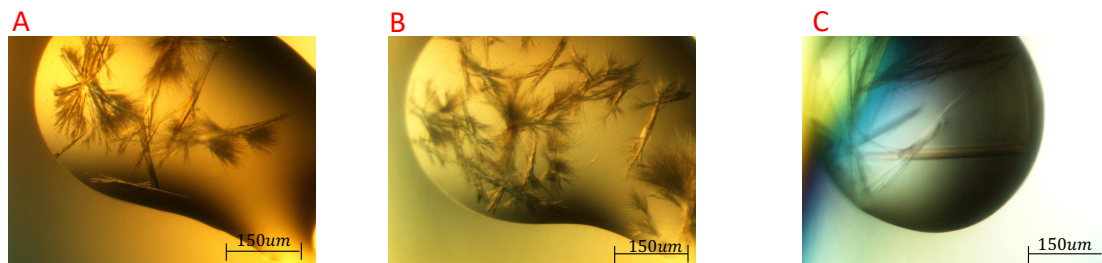
*Figure 3.20: Grown crystals of SAG-C75 on PH-Clear suite. A) crystal on F11 condition (0.1 M Tris PH 8.0 + 3.2 M Ammonium sulphate). B) crystal on F4 condition (0.1 M HEPES PH 7.0 + 2.4 M Ammonium sulphate). C) crystal on F3 condition (0.1 M MES PH 6.0 + 2.4 Ammonium sulphate). D) crystal on F6 condition (0.1 M Bicine PH 9.0 + 2.4 Ammonium sulphate).*

#### 3.6.2.2 SAG-EBC1

Purified SAG-EBC1 protein at 16.5 mg/ml in buffer C was used to set up four 96-well crystallisation plates of the protein, using the same screens as before. After two weeks, crystals were observed in PACT C4, JCSG G10, and PH-Clear H3 conditions. All the crystals were mounted and immersed in cryoprotectant (20 % ethylene glycol + mother solution of each condition except the H3 crystal which contain 40% MPD) before being cooled in liquid



nitrogen and sent to the Diamond Synchrotron under code MX24447 trip number 90 on beamline i03 for data collection [Figure 3.21]. However, it was not possible to determine the structure of SAG-EBC1 using this data.



**Figure 3.21: Crystal of EBC1.** A) crystal on PACT C4 condition (0.1 M PCB buffer PH 7.0 + 25% (w/v) PEG1500; B) crystal on PH-Clear H3 condition (0.1 M MES PH 6.0 + 40% (v/v) MPD); C) crystal on JCSG G10 condition (0.15 M potassium bromide + 30% (w/v) PEG MME 2000).

### 3.6.2.3 SAG-EBC2 and SAG-EBC3

Only a small amount of purified protein was obtained for each of these proteins (40 µl of 14.6 mg/ml for EBC2 and 45 µl of 9.2 mg/ml for EBC3, respectively), which was insufficient to set up four 96-well plates for each. Thus, the crystallization screens were divided between them: PACT and PH-Clear suites were used to set up two 96-well plates for EBC2, while JCSG and AmSO<sub>4</sub> suites were used for two 96-well plates for EBC3. Despite regular monitoring, no crystals were observed for either protein. For these two SAG-C family proteins additional methods are required to improve solubility. Unfortunately, there was not enough time to explore different expression protocols and further study has been paused for these two proteins.

### 3.7 *Eimeria Tenella* SAG A family

The majority of the SAG proteins in *E. tenella* belong to the SAG A family (60 members) (Reid *et al.* 2014) and eight SAG-A family proteins were selected for study as similarities and differences to the SAG-B and SAG-C families may shed light on understanding the function of these proteins, which until now is still unclear.

The SAG-A proteins all have a big insertion in their sequences in loop L10 compared to SAG-B19, with some SAG-A proteins having a conserved RRL motif in this insertion. In choosing the SAG-A members for study, it was attempted to cover the differences in the insertion length, the presence of the RRL motif as well as the diverse sequence identity to SAG-B19, in an attempt to understand the variance in this family [Table 3.5].

Table 3.5: shows the similarity of SAG-A members to SAG-B19 with those having the RRL motifs.

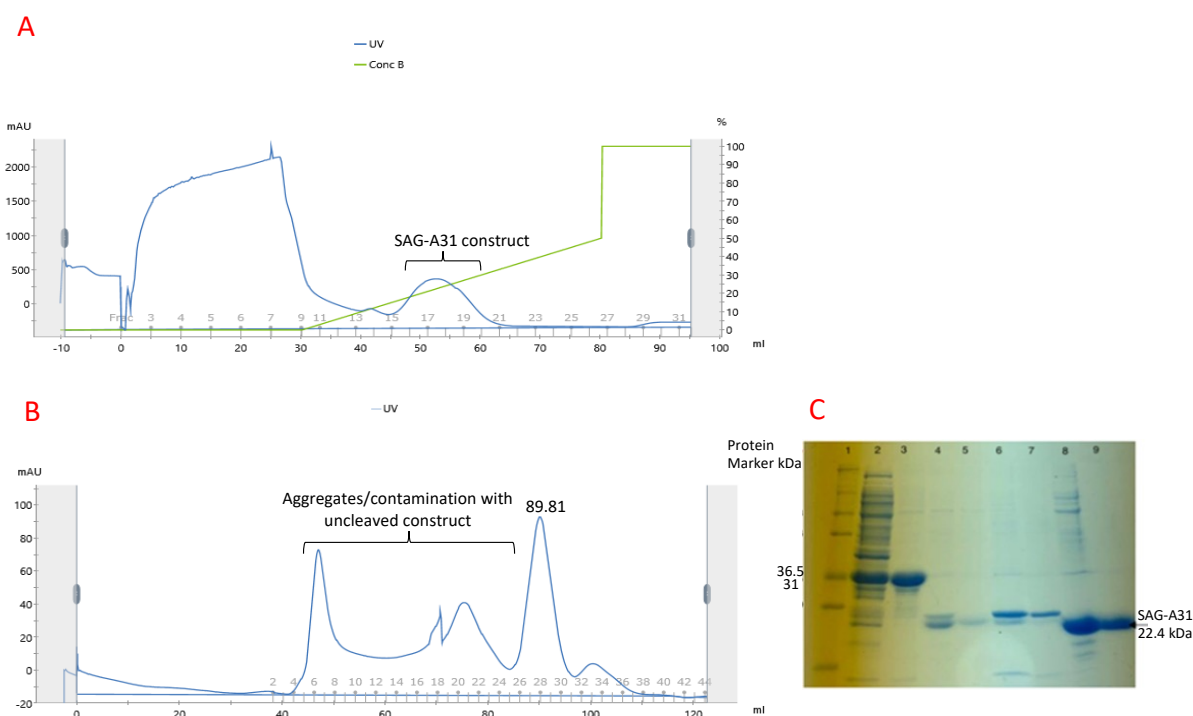
SAG ID	Construct mw (kDa)	Core SAG mw (kDa)	Core SAG residues	Sequence identity to SAG-B19 core (%)	RRL motifs
SAG-A10	36.7	22.3	Glu29-Ser236	18%	X
SAG-A31	36.8	22.4	Lys19-Ala226	22%	√
SAG-A7	38.7	24.3	Gln26-Leu253	20%	X
SAG-A91	34	19.6	Pro28-Gly216	20%	X
SSAG -A1	36	21.6	Gln24-Ser223	25%	√
SAG-A905	37	22.5	Thr32-Gly240	19%	X
SAG-A870	36.3	22	Thr23-Asn229	20%	√
SAG-A6	36.4	22	Pro30-Gly234	16%	X

#### 3.7.1 Overexpression and Purification

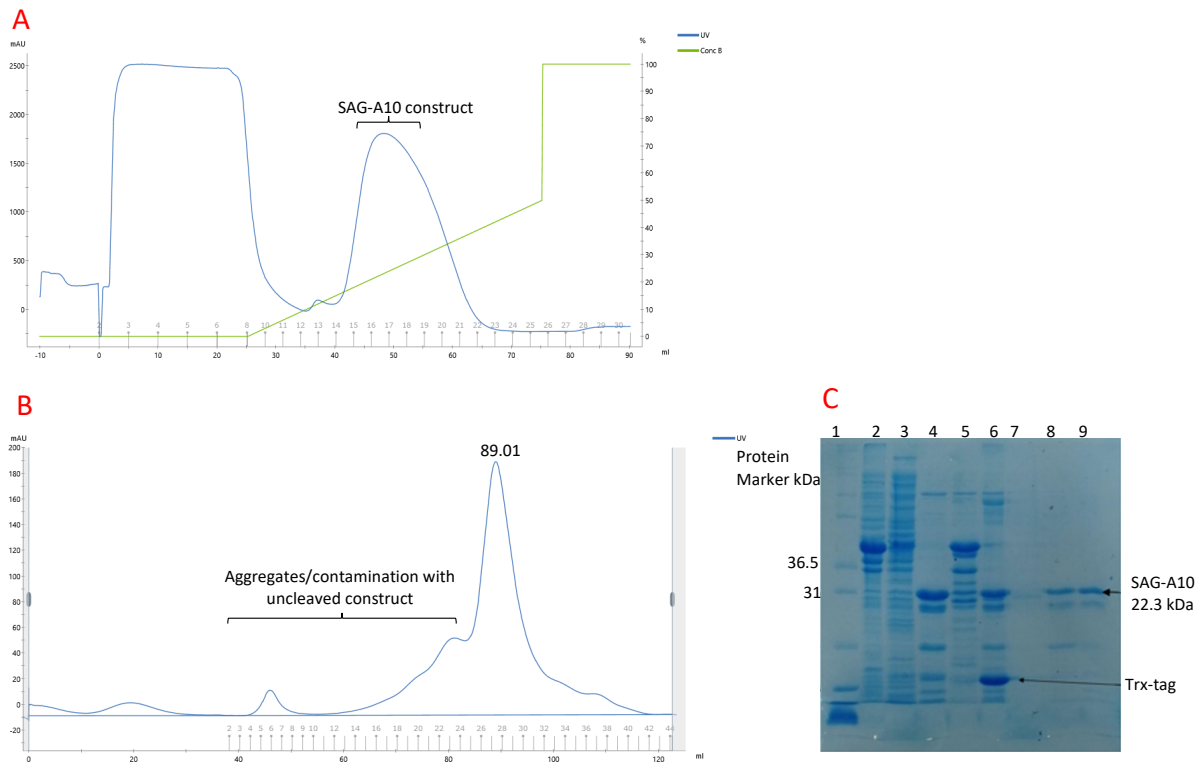
The same strategies were followed as that used for the previous SAGs, with *E. tenella* SAG-A constructs designed and transformed into *E. coli* origami cells (DE3). After overexpression in 500 ml LB cultures at 17°C, the SAG-A family proteins showed successful expression, with varying solubility between 30 and 70 % [Table 3.6]. Purification of each SAG-A protein followed the same protocol as before, yielding sufficient quantity of each protein for crystallization studies [Table 3.6, Figures 3.22 – 3.29]. Although 8 SAG-A family proteins were successfully expressed, the structure of only one (SAG-A91) was determined. The results of all the crystallisation experiments are detailed for each SAG-A protein below, as a record for future study.

Table 3.6: Overexpression and Purification Summary for SAG-A family proteins

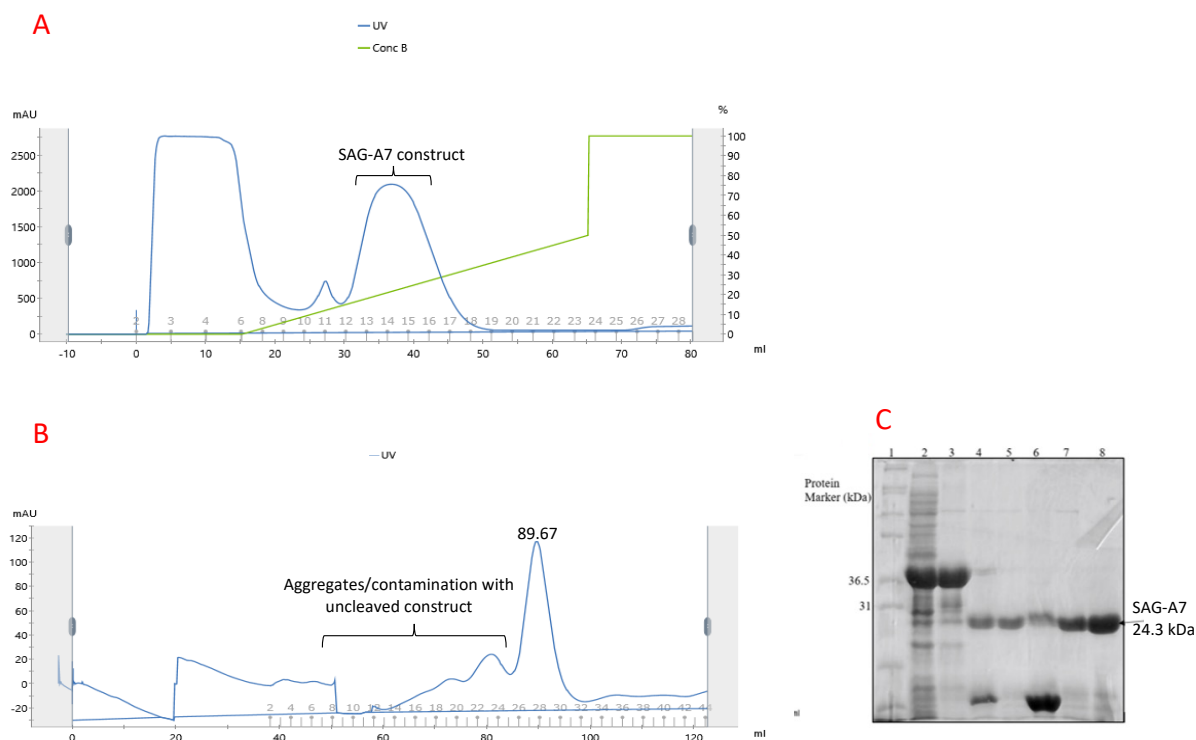
A-family SAG	Culture Volume	Expression temperature	Approximate soluble expression (%)	Final Purified Protein		Crystals	Data
				Volume (ul)	Concentration (mg/ml)		
SAG-A31	6 x 500 ml	18 C°	35	60-75	18	√	No
SAG-A10	6x 500 ml	17 C°	45	60-70	22	√	No
SAG-A7	6x 500 ml	17 C°	45	45-55	23	No	
SAG-A6	8 x 500 ml	17 C°	30	25-30	13.6	No	
SAG-A870	8 x 500 ml	17 C°	30	20-30	13	No	
SAG-A905	4 x 500 ml	17 C°	70	55-65	23	No	
SAG-A1	6 x 500 ml	18 C°	50	35-45	9	No	
SAG-A91	6x 500 ml	18 C°	50	60-70	18	√	√



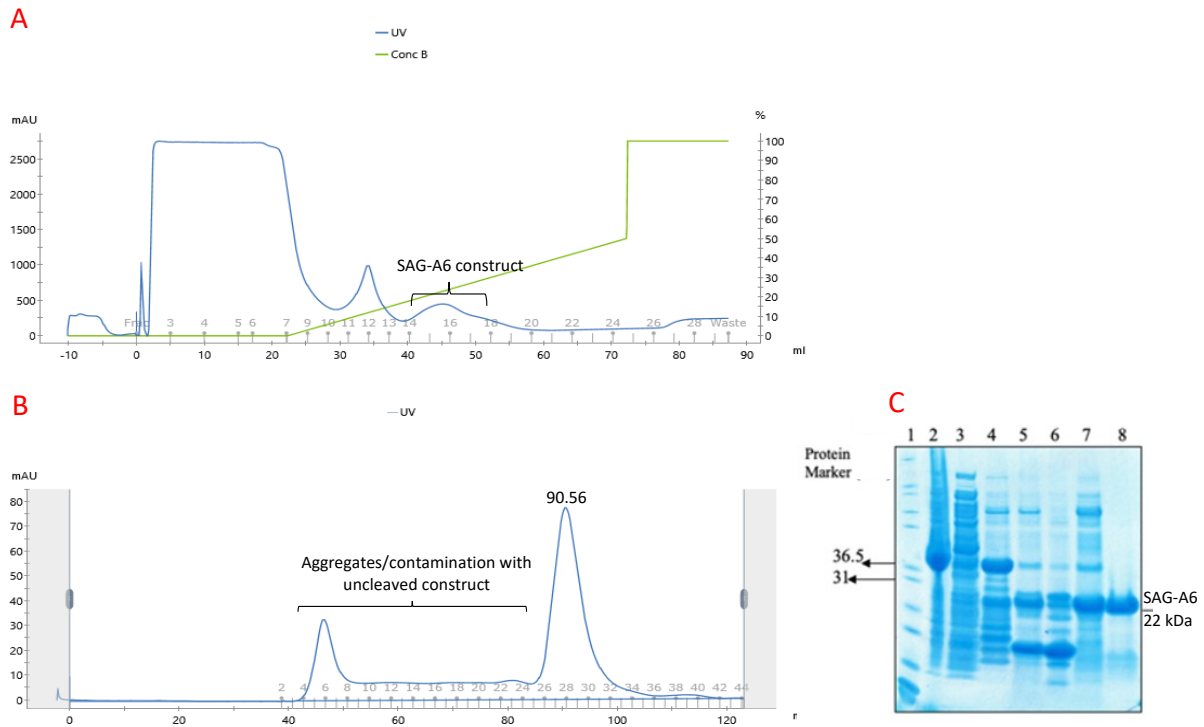
**Figure 3.22: SAG-A31 purification process.** A) The 280 nm absorption trace depicts the affinity chromatography using a 5 ml His-Trap<sup>TM</sup>HP cartridge (GE Healthcare). 117 mg of SAG-A31 CFE was loaded onto the column with a flow rate of 5 ml/min. A 50 ml gradient ranging from 0 to 0.25 M imidazole in buffer A was employed, and fractions of 3 ml were collected. B) 280 nm absorption trace shows the gel filtration process using a 1.6x60 Hi-load<sup>TM</sup> Superdex<sup>TM</sup> 200 column (GE Healthcare). 1.65 mg of SAG-A31 was loaded to the column and buffer A was used to elute the 2 ml collected fraction. C) SDS-PAGE analysis of cleaved SAG-A31 after purification steps. Lane-1 protein marker (Mark12TM); Lane-2 CFE of SAG-A31; Lane-3 SAG-A31 before cleavage (after Ni-NTA 1); Lane-4 unbound material; Lane-5 SAG-A31 after cleavage (before Ni-NTA 2); Lane- 6 and 7 Ni-NTA 2 elution (imidazole peak fraction); Lane-8 gel filtration loading sample; Lane-9 final prep of SAG-A31.



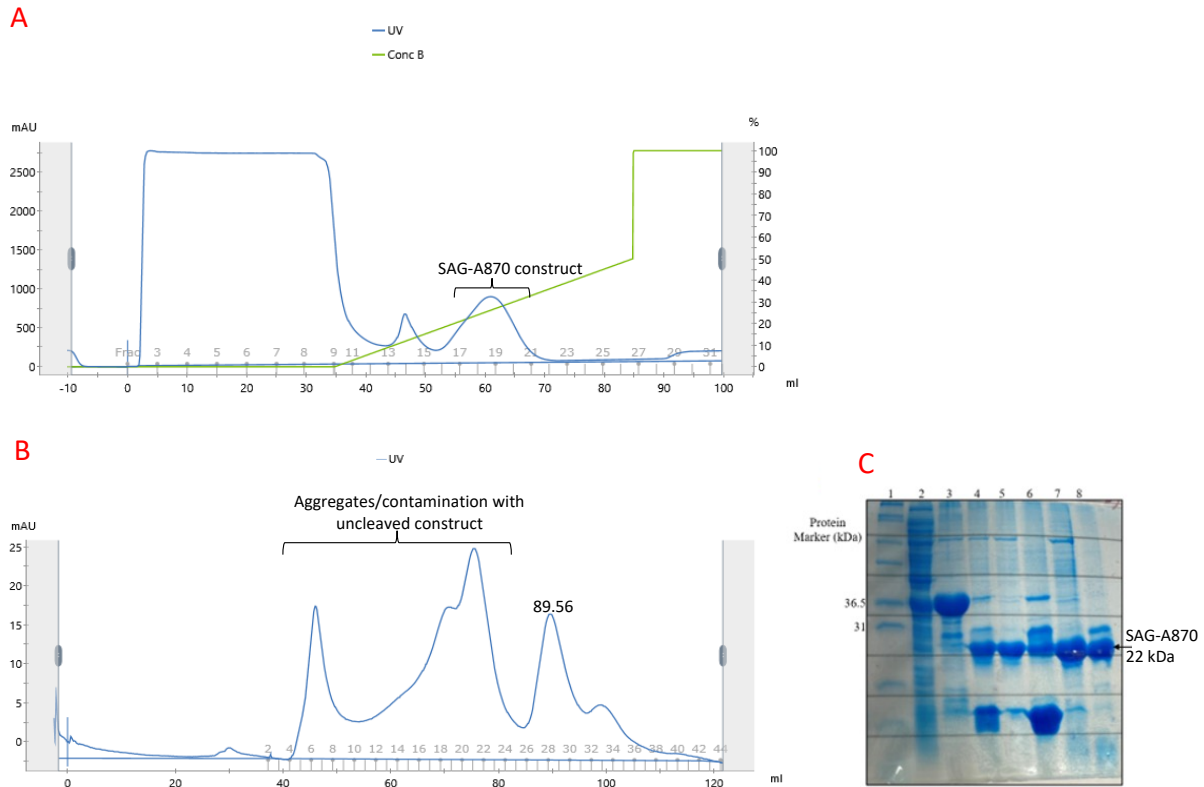
**Figure 3.23: SAG-A10 purification process.** A) The 280 nm absorption trace depicts the affinity chromatography using a 5 ml His-Trap<sup>TM</sup> HP cartridge (GE Healthcare). 182 mg of SAG-A10 CFE was loaded onto the column with a flow rate of 5 ml/min. A 50 ml gradient ranging from 0 to 0.25 M imidazole in buffer A was employed, and fractions of 3 ml were collected. B) 280 nm absorption trace shows the gel filtration process using a 1.6x60 Hi-load<sup>TM</sup> Superdex<sup>TM</sup> 200 column (GE Healthcare). 4 mg of SAG-A10 was loaded to the column and buffer A was used to elute the 2 ml collected fraction. C) SDS-PAGE analysis of cleaved SAG-A10 after purification steps. Lane-1 protein marker (Mark12<sup>TM</sup>); Lane-2 cells debris; Lane-3 CFE of SAG-A10; Lane-4 SAG-A10 unbound material; Lane-5 before cleavage (after Ni-NTA 1); Lane-6 SAG-A10 after cleavage (before Ni-NTA 2); Lane-7 Ni-NTA 2 elution (imidazole peak fraction); Lane-8 gel filtration loading sample; Lane-9 final prep of SAG-A10 at expect molecular wight.



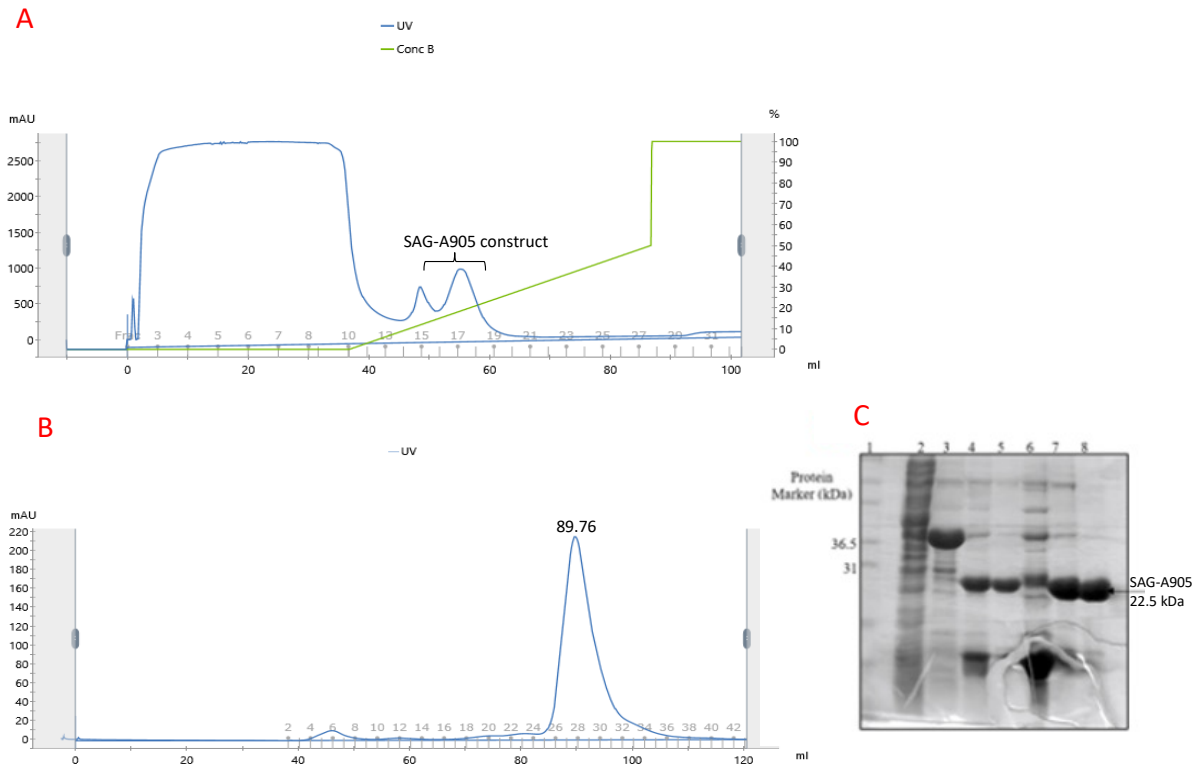
**Figure 3.24: A) SAG-A7 purification process.** A) The 280 nm absorption trace depicts the affinity chromatography using a 5 ml His-Trap<sup>TM</sup> HP cartridge (GE Healthcare). 153 mg of SAG-A7 CFE was loaded onto the column with a flow rate of 5 ml/min. A 50 ml gradient ranging from 0 to 0.25 M imidazole in buffer A was employed, and fractions of 3 ml were collected. B) 280 nm absorption trace shows the gel filtration process using a 1.6x60 Hi-load<sup>TM</sup> Superdex<sup>TM</sup> 200 column (GE Healthcare). 5 mg of SAG-A7 was loaded to the column and buffer A was used to elute the 2 ml collected fraction. C) SDS-PAGE analysis of cleaved SAG-A7 after purification steps. Lane-1 protein marker (Mark12<sup>TM</sup>); Lane-2 CFE of SAG-A7; Lane-3 SAG-A7 before cleavage (after Ni-NTA 1); Lane-4 post-TEV cleavage; Lane-5 SAG-A7 after cleavage (before Ni-NTA 2); Lane- 6 peak fractions eluted with Imidazole on Ni-NTA 2; Lane-7 gel filtration loading sample; Lane-8 final prep of SAG-A7 at expect molecular wight.



**Figure 3.25: SAG-A6 purification process.** A) The 280 nm absorption trace depicts the affinity chromatography using a 5 ml His-Trap<sup>TM</sup> HP cartridge (GE Healthcare). 170 mg of SAG-A6 CFE was loaded onto the column with a flow rate of 5 ml/min. A 50 ml gradient ranging from 0 to 0.25 M imidazole in buffer A was employed, and fractions of 3 ml were collected. B) 280 nm absorption trace shows the gel filtration process using a 1.6x60 Hi-load<sup>TM</sup> Superdex<sup>TM</sup> 200 column (GE Healthcare). 1.8 mg of SAG-A6 was loaded to the column and buffer A was used to elute the 2 ml collected fraction. C) SDS-PAGE analysis of cleaved SAG-A6 after purification steps. Lane-1 protein marker (Mark12TM); Lane-2 cell debris, Lane-3 CFE, Lane-4 pre-cleaved (post-Ni-NTA 1), Lane-5 post-cleaved, Lane-6 Ni-NTA 2 elution (representing the imidazole peak fraction), Lane-7 gel filtration loading sample, and Lane-8 the ultimate purified SAG-A6.

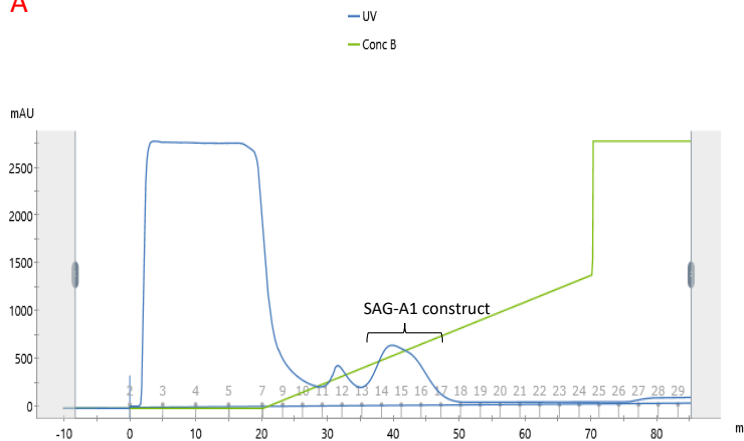
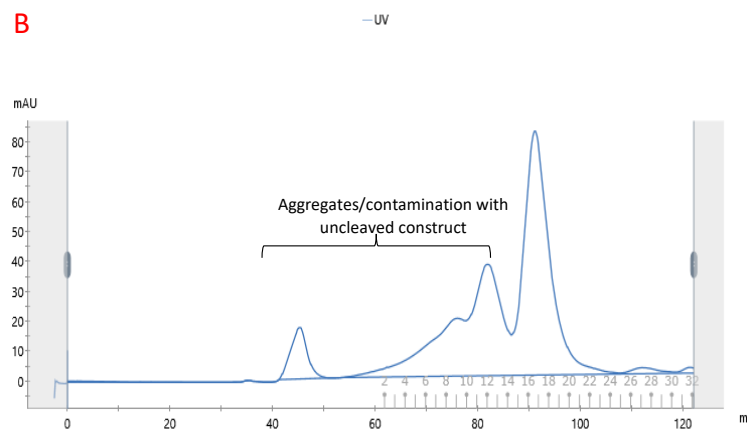
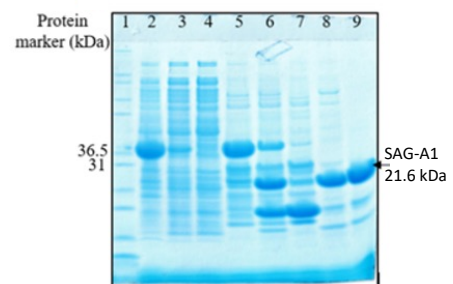


**Figure 3.26: SAG-A870 purification process.** A) The 280 nm absorption trace depicts the affinity chromatography using a 5 ml His-Trap<sup>TM</sup> HP cartridge (GE Healthcare). 250 mg of SAG-A870 CFE was loaded onto the column with a flow rate of 5 ml/min. A 50 ml gradient ranging from 0 to 0.25 M imidazole in buffer A was employed, and fractions of 3 ml were collected. B) 280 nm absorption trace shows the gel filtration process using a 1.6x60 Hi-load<sup>TM</sup> Superdex<sup>TM</sup> 200 column (GE Healthcare). 1 mg of SAG-A870 was loaded to the column and buffer A was used to elute the 2 ml collected fraction. C) SDS-PAGE analysis of cleaved SAG-A870 after purification steps. Lane-1 protein marker (Mark12<sup>TM</sup>); Lane-2 cells debris; Lane-3 CFE of SAG-A870; Lane-4 SAG-A870 before cleavage (after Ni-NTA); Lane-5 SAG-A870 after cleavage; Lane-6 Ni-NTA 2 elution (imidazole peak fraction); Lane-7 gel filtration loading sample; Lane-8 final prep of SAG-A870.

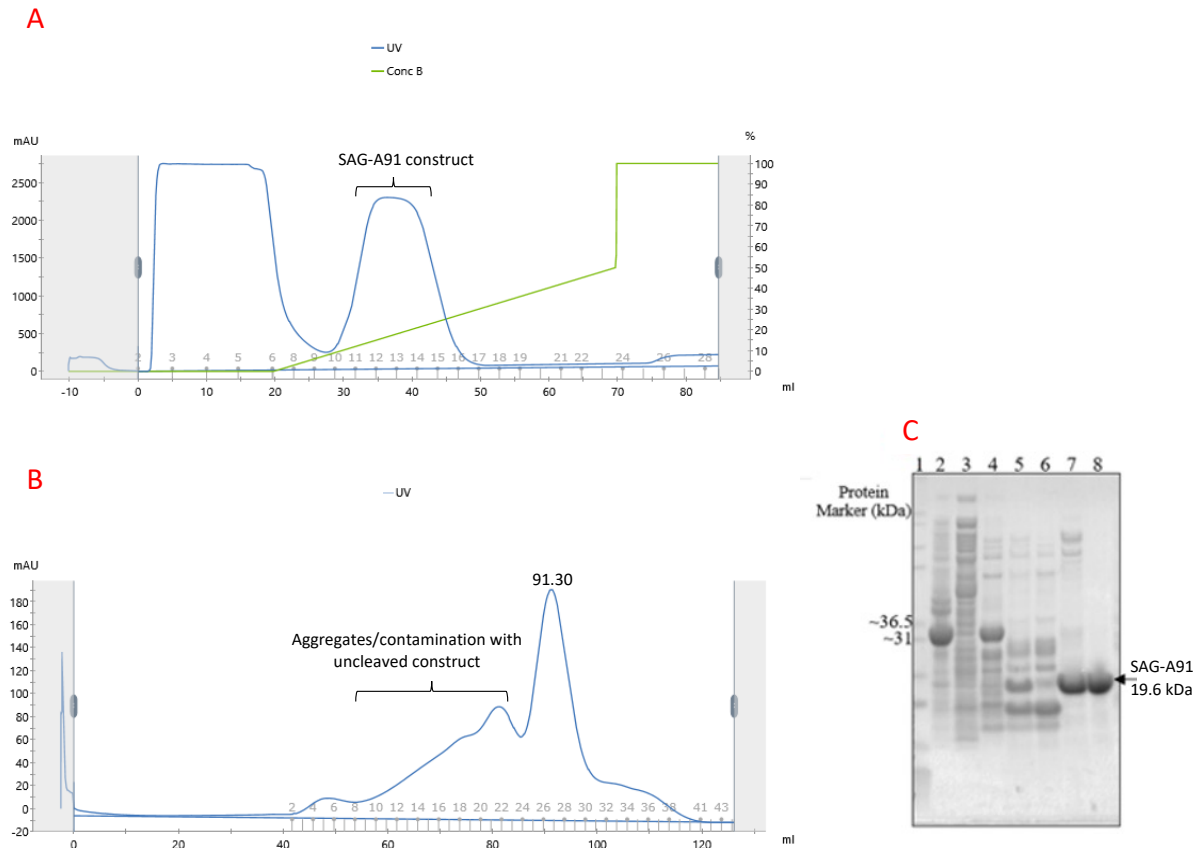


**Figure 3.27: *SAG-A905* purification process.** A) The 280 nm absorption trace depicts the affinity chromatography using a 5 ml His-Trap<sup>TM</sup> HP cartridge (GE Healthcare). 255 mg of *SAG-A905* CFE was loaded onto the column with a flow rate of 5 ml/min. A 50 ml gradient ranging from 0 to 0.25 M imidazole in buffer A was employed, and fractions of 3 ml were collected. B) 280 nm absorption trace shows the gel filtration process using a 1.6x60 Hi-load<sup>TM</sup> Superdex<sup>TM</sup> 200 column (GE Healthcare). 4.5 mg of *SAG-A905* was loaded to the column and buffer A was used to elute the 2 ml collected fraction. C) SDS-PAGE analysis of cleaved *SAG-A905* after purification steps. Lane-1 protein marker (Mark12<sup>TM</sup>); Lane-2 cells debris; Lane-3 CFE of *SAG-A905*; Lane-4 *SAG-A905* before cleavage (after Ni-NTA); Lane-5 *SAG-A905* after cleavage; Lane-6 Ni-NTA 2 elution (imidazole peak fraction); Lane-7 gel filtration loading sample; Lane-8 final prep of *SAG-A905*.



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

**Figure 3.28: SAG-A1 purification process.** A) The 280 nm absorption trace depicts the affinity chromatography using a 5 ml His-Trap<sup>TM</sup> HP cartridge (GE Healthcare). 130 mg of SAG-A1 CFE was loaded onto the column with a flow rate of 5 ml/min. A 50 ml gradient ranging from 0 to 0.25 M imidazole in buffer A was employed, and fractions of 3 ml were collected. B) 280 nm absorption trace shows the gel filtration process using a 1.6x60 Hi-load<sup>TM</sup> Superdex<sup>TM</sup> 200 column (GE Healthcare). 4 mg of SAG-A1 was loaded to the column and buffer A was used to elute the 2 ml collected fraction. C) SDS-PAGE analysis of cleaved SAG-A1 after purification steps. Lane-1 protein marker (Mark12<sup>TM</sup>); Lane-2 cells debris; Lane-3 CFE of SAG-A1; Lane-4 Ni-unbound; Lane-5 SAG-A1 before cleavage (after Ni-NTA); Lane-6 SAG-A1 after cleavage; Lane-7 Ni-NTA 2 elution (imidazole peak fraction); Lane-8 gel filtration loading sample; Lane-9 final prep of SAG-A1.

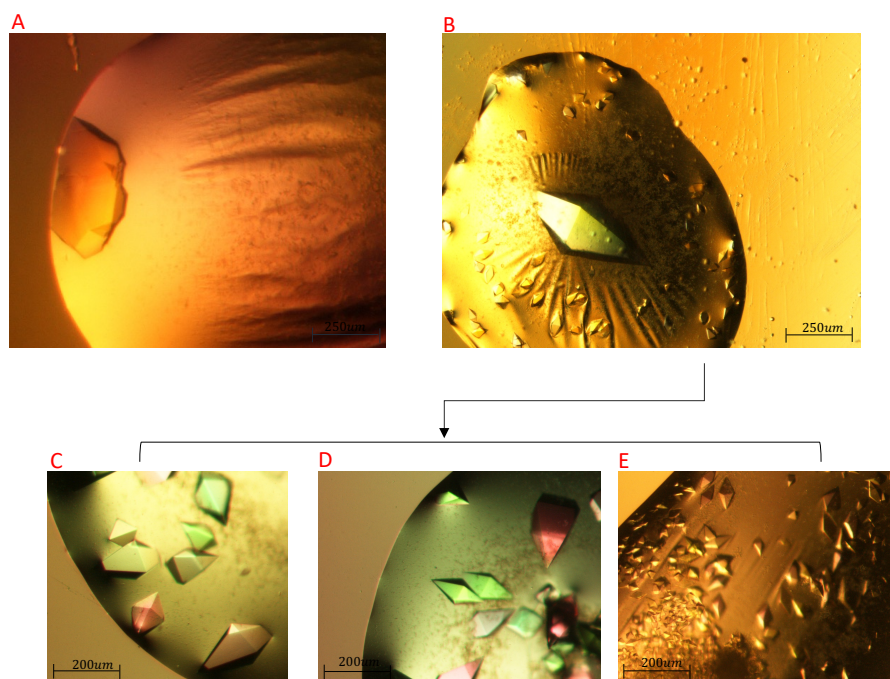


**Figure 3.29: SAG-A91 purification process.** A) The 280 nm absorption trace depicts the affinity chromatography using a 5 ml His-Trap<sup>TM</sup> HP cartridge (GE Healthcare). 390 mg of SAG-A91 CFE was loaded onto the column with a flow rate of 5 ml/min. A 50 ml gradient ranging from 0 to 0.25 M imidazole in buffer A was employed, and fractions of 3 ml were collected. B) 280 nm absorption trace shows the gel filtration process using a 1.6x60 Hi-load<sup>TM</sup> Superdex<sup>TM</sup> 200 column (GE Healthcare). 3.6 mg of SAG-A91 was loaded to the column and buffer A was used to elute the 2 ml collected fraction. C) SDS-PAGE analysis of cleaved SAG-A91 after purification steps. Lane-1 protein marker (Mark12<sup>TM</sup>); Lane-2 CFE of SAG-A91; Lane-3 Ni-unbound; Lane-4 SAG-A91 before cleavage (after Ni-NTA); Lane-5 SAG-A91 after cleavage; Lane-6 Ni-NTA 2 elution (imidazole peak fraction); Lane-7 gel filtration loading sample; Lane-8 final prep of SAG-A91.

### 3.7.2 Crystallization of SAG-A Proteins

#### 3.7.2.1 Crystallization of SAG-A91

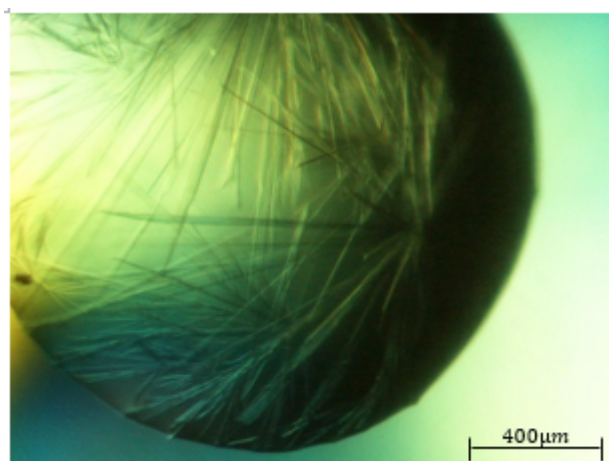
The final SAG-A family protein to be tried was ETH\_00023375 (SAG-A91). The sequence alignment with SAG-B19 showed that SAG-A91 had one of the smallest insertions in loop L10 for the SAG-A family proteins. Expression of SAG-A91 was good, with 50% of this SAG in the soluble fraction [Table 3.6] and the purification resulted in Figure 3.29. Following crystallisation screening using the normal four suites, crystals were observed in conditions F2 and H2 condition of the AmSO<sub>4</sub> suite (0.1 M Citric acid pH 4.0 + 2.4 M Ammonium sulphate, 0.1 M Sodium acetate PH 4.6 + 2.0 M Ammonium sulphate, respectively) [Figure 3.30 A and B]. Optimization of the F2 condition, varying the ammonium sulphate concentration and the PH of the citric acid ranged 4-6, and by using much larger drops (2  $\mu$ l) in a hanging drop experiment, produced much of bi-pyramidal crystals [Figure 3.30 C, D, and E]. A number of these were mounted on loops, cryoprotected in 20 % ethylene glycol + the mother solution of each condition, cooled in liquid nitrogen and sent to the Diamond Synchrotron under code MX24447 trip 73 on the i03 beamline for data collection.



**Figure 3.30: Crystals of SAG-A91.** A) grown crystal on AmSO<sub>4</sub> H2 condition (0.1 M Sodium acetate PH 4.6 + 2.0 M ammonium sulphate); B) grown crystal on AmSO<sub>4</sub> F2 condition (0.1 M Citric acid PH 4.0 + 2.4 M ammonium sulphate. C) the optimized crystal of AmSO<sub>4</sub> F2 condition (0.1 M Citric acid PH 5.0 + 2.2 M ammonium sulphate.); D) (0.1 M Citric acid PH 5.0 + 2.0 M ammonium sulphate); E) (0.1 M Citric acid PH 5.5 + 2.4 M ammonium sulphate.).

### 3.7.2.2 Crystallization of SAG-A31

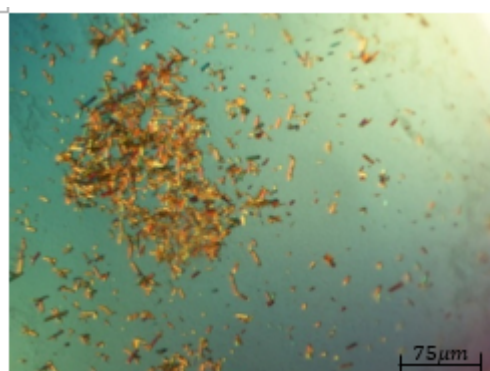
Purified SAG-A31 protein at 18 mg/ml was screened using the PACT, JCSG, PH clear, and AmSO<sub>4</sub> crystallisation screens, as described on chapter 2.7. Needle crystals were observed in condition D12 (0.1 M Bicine PH 9.0 + 30% (w/v) PEG 6000) of the PH-Clear suite [Figure 3.31]. Several crystals were cryo-protected in 20 % ethylene glycol + mother solution of D12 condition and then sent to the Diamond Synchrotron under code MX24447 trip 86 on beamline i03, but the diffraction quality was poor and data could not be collected from these crystals. Attempts were made to improve the crystal quality by varying the concentration of PEG 6000 and the PH of the bicine buffer. However, in these new conditions, no crystal growth was observed and so structural determination of SAG-A31 was paused.



*Figure 3.31: **Crystal of SAG-A31.** The grown crystal of PH-Clear D12 condition (0.1 M Bicine PH 9.0 + 30% (w/v) PEG 6000).*

### 3.7.2.3 Crystallization of SAG-A10

Small clusters of crystals of SAG-A10 were observed in the H7 condition of the PH-Clear suite (0.1 M citric acid PH 4.0 + 65% (v/v) MPD) [Figure 3.32], when purified SAG-A10 protein at 22 mg/ml was screened against the same four screens as those used with SAG-A31. Attempts to improve the crystals by varying the pH range (4.5 – 6.5), and MPD concentration (15 - 65%), failed to yield crystals and thus structure determination of this SAG-A protein was also paused.



*Figure 3.32: Crystal of SAG-A10. The grown crystal of PH-Clear H7 condition (0.1 M citric acid PH 4.0 + 65% (v/v) MPD)*

### 3.7.2.4 Crystallisation of SAG-A7

Despite obtaining good quantities of purified SAG-A7 protein at 23 mg/ml, robot crystallization screening using the standard four screens did not produce crystals in any of the conditions, and thus the structure determination was paused.

### 3.7.2.5 Crystallization of SAG-A6

The expression and purification of soluble SAG-A6 was not as straightforward as for some of the other SAG proteins and only 28µl of 13.6 mg/ml protein was produced, insufficient to try all four crystallisation screens. Consequently, only the PH-Clear and AmSO<sub>4</sub> suites were used. Regular monitoring of the screening plates revealed no crystal formation and the wells remained clear, indicating that the protein had not precipitated. In an attempt to remedy this, efforts were made to increase the concentration of the purified SAG-A6, by repeating the

expression and purification steps. However, it was not possible to increase the concentration of the purified protein and thus the solubility limit was not reached in the crystallisation conditions used.

#### **3.7.2.6 Crystallization of SAG-A870**

The protein expression and purification of SAG-A870 faced multiple problems [Figure 3.26], starting with low solubility and resulted in just ~30 µl of a 13 mg/ml solution in buffer C, which was sufficient to undertake just two crystallization screens. The PACT and JCSG suites were chosen as these were not used for SAG-A6. Unfortunately, no crystal formation was observed and due to time constraints further steps in the process were paused.

#### **3.7.2.7 Crystallization of SAG-A905**

The expression and purification of SAG-A905 was straightforward and aided by the high levels of protein in the soluble cell fraction (~70%). The purified protein was diluted to 23 mg/ml and the standard four screens (JCSG, PACT, AmSO<sub>4</sub> and PH-Clear) were used in crystallisation trials, but disappointingly no protein crystals were observed.

#### **3.7.2.8 Crystallization of SAG-A1**

The *E. tenella* SAG-A1 protein has been extensively studied and is proposed to have ability to attachment to mammalian cells because it highly expressed on the sporozoite stage of the life cycle (Tabarés *et al.*, 2004) and was thus chosen as one of the family A SAG proteins for this study. Expression proved fairly straightforward, with ~ 50 % of the protein in the soluble cell fraction, however, following purification a maximum concentration of 9 mg/ml could be obtained in ~38 µl. Two plates (AmSO<sub>4</sub> and PACT suites) were screened, but no crystals were observed. It is possible that the low molecular weight fragments in the final sample [Figure 3.28] indicating the presence of some clipped molecules could have negatively affected the crystallisation process, and attempts should be made to purify the protein further, however, again there was insufficient time to undertake further experiments.

### **3.8 Summary**

Of the 16 SAG proteins described in this chapter, all proved possible to express and purify, although in varying amounts and data-quality crystals were obtained for six SAG proteins, representatives from each of the three *Eimeria* SAG families. The determination of the structures of these SAG proteins is detailed in the following chapter.

## Chapter4 Structure determination and preliminary crystallographic analysis of *E. tenella* SAG proteins.

### 4.1 Overview

This chapter describes data collection, experimental phasing of and structural description of the *E. tenella* SAGs that had successful structural determination. These include four SAG-B family proteins, one SAG-A family protein and the single SAG-C protein in this *Eimeria tenella* [Table 4.1]. For each protein a number of different crystals were sent to the Diamond synchrotron for data collection and processing, this chapter describes the details of data processing and structure determination for the crystal that gave the best quality data for each protein.

Table 4.1: Crystallization and data collection overview of successful structure determinations of SAG proteins.

SAG ID	Crystallization condition	Resolution	Spacegroup	Beamline
SAG-B13	0.2 M potassium nitrate PH 6.9+ 20% (w/v) PEG 3350	1.71 Å	C2 2 <sub>1</sub>	i03
SAG-B16	0.2 M potassium thiocyanate + 2.2 M Ammonium sulphate	1.52 Å	P2 <sub>1</sub>	i03
SAG-B22	0.1 M SPG buffer PH 5 + 25% (w/v) PEG 1500	1.87 Å	P4 <sub>3</sub> 2 <sub>1</sub> 2	i03
SAG-B41	0.1 M MES PH 6 + 30% (w/v) PEG 6000	1.18 Å	P1	i03
SAG-A91	0.1 M sodium acetate PH 4.6 + 2.0 M Ammonium sulphate	1.40 Å	P4 <sub>1</sub> 2 <sub>1</sub> 2	i03
SAG-C75	0.1 M Tris PH 8.0 + 3.2 M Ammonium sulphate	1.32 Å	P4 <sub>1</sub> 2 <sub>1</sub> 2	i04

### 4.2 Data collection and processing

Crystals of each SAG were transferred to a cryoprotectant solution, mounted on a loop and immersed in liquid nitrogen, prior to transfer to the Diamond synchrotron for data collection as described on chapter 3. Data were collected using the unattended data collection protocol and processed through the various pipelines as described in chapter 2 section 2.5.6.

### 4.3 Structure Determination of SAG-B family Proteins

Diffraction quality crystals were obtained for all four of the B subfamily SAGs that had been successfully purified (SAG-B13, SAG-B16, SAG-B22 and SAG-B41) with varying resolution. Each of these proteins is described in turn below:

#### 4.3.1 SAG-B13 Data Processing

A total of 3600 images were collected with a rotation range of  $0.1^\circ$  per image. The diffraction images showed clear spots extending to high resolution, indicating good crystal quality. The processing revealed a space group  $C2\ 2\ 2_1$  with cell dimensions of  $a = 49.57\ \text{\AA}$ ,  $b = 65.16\ \text{\AA}$ ,  $c = 114.60\ \text{\AA}$ ,  $\alpha = 90^\circ$ ,  $\beta = 90^\circ$ ,  $\gamma = 90^\circ$ . Inspection of the processing statistics for the different pipelines showed that the xia2 3dii was the better pipeline based on the provided metrics, which indicate a resolution range of  $39.45 - 1.71\ \text{\AA}$ , and a completeness of 100%. The  $Mn\langle I/\sigma(I) \rangle$  value stands at 7.5, with  $R_{pim}$  0.997. [Table 4.2].

Table 4.2: SAG-B13 Data processing statistics.

Overall	
Diamond Beamline	I03
Wavelength ( $\text{\AA}$ )	0.97623
Space group	$C2\ 2\ 2_1$
Unit cell lengths( $\text{\AA}$ )	$a = 49.57$ , $b = 65.16$ , $c = 114.60$
Unit cell angles ( $^\circ$ )	$\alpha = 90$ , $\beta = 90$ , $\gamma = 90$
Molecules per asymmetric unit	
Resolution range ( $\text{\AA}$ )	39.45 – 1.71 (1.74 - 1.71)
Total Reflections measured	271325 (14522 - 12896)
Unique Reflections	20477 (1127 - 982)
Completeness (%)	100.00 (100.00)
Multiplicity	13.25 (13.13)
CC-1/2	0.9969 (0.3012)
Mean $I/\sigma$	7.55 (0.82)
$R_{merge}$ (I)	0.2315 (2.9403)
$R_{pim}$	0.997 (0.301)



#### 4.3.1.1 Molecular Replacement for SAG-B13

A solvent content analysis, using the method of Matthews (Potterton *et al.*, 2018), showed that it was likely that a single copy for SAG-B13 was present in the asymmetric unit, with a solvent content of 47.2 %. The scaled and merged data file from the xia2 3dii pipeline was used as input for molecular replacement using *Phaser*, with SAG-B19 as the search model. The *Phaser* analysis yielded a unique solution with the space group C 2 2 2<sub>1</sub>, with a high translational function Z-score (TFZ) of 19.2 and a refined TFZ-equivalent of 19.2, alongside a substantial overall log likelihood gain (LLG) of 474. Despite a single packing clash, the model, had a good R-factor (0.38), and clear differences could be seen between the electron density and the model, showing where the sequence of SAG-B13 differs from SAG-B19, confirming that the molecular replacement was successful.

#### 4.3.1.2 Refinement of SAG-B13

To change the correctly placed molecular replacement model into that of SAG-B13, Buccaneer was used to auto-build a model with the sequence of SAG-B13 into the *Phaser* electron density map. After running Buccaneer, the model was selected from cycle 7 due to it having the lowest free-R factor. The model consisted of 189 residues built into 3 fragments, with all residues assigned to the sequence. This model had an R-factor of 0.21 and a free-R factor of 0.25, with an RMS bond deviation of 0.008 Å. These metrics suggest that the model was approaching completion. However, inspection of the fit of the model to the density showed that a number of residues had been automatically built into very weak density and these were removed from the model. Further rounds of manual model building in COOT and refinement in Refmac5, were used to link the three fragments together to build a single chain of SAG-B13, with water molecules added as appropriate in the |Fo-Fc| electron density map as this map reflection the water molecules and any other substrate if it found. This process was repeated until no new features could be interpreted in either the |2mFo-DFc| electron density map or the |Fo-Fc| difference electron density map. This produced a final model that fitted the density well [Figure 4.2], with an R<sub>work</sub> and R<sub>free</sub> of 0.18 and 0.20, respectively [Table 4.3].

Table 4.3: Final refinement statistics for SAG-B13.

Refinement	
Resolution	1.71 Å
Number of non-H atoms	
Protein	2668
Water	154
R <sub>work</sub> /R <sub>free</sub>	0.188/0.245
Average B factor (Å <sup>2</sup> )	
Main chain/side chain	18.59/22.48
Water	30.4
Rmsd bond length (Å) / angle (°)	0.0113/1.986
Ramachandran favoured/allowed (%)	98.88/1.12

In the final refined map, the density was very weak between residues Pro95 - Ser99 and these residues were not built into the final model [Figure 4.1A]. At the N-terminus, clear density could be seen for the alanine residue, but not the preceding serine from the TEV cleavage site. For another three residues (Asn124, Gln125 and Gln127), the electron density was weak for the side chains, but these residues have been included in the model. [Figure 4.1B]. The final model of SAG-B13 thus contained 182 residues (Ala52 – Gln94 and Gly100 – Gln239).

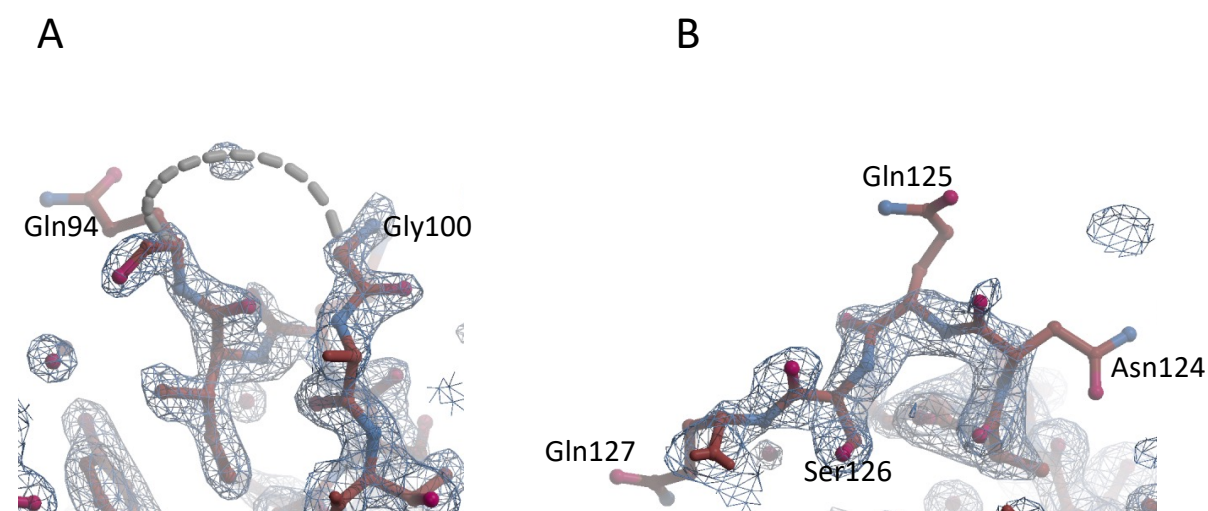


Figure 4.1: **Weak electron density of SAG-B13 map.** A) shows the five disordered residues (Pro95, Leu96, Ala97, Arg98 and Ser99) where no electron density can be seen. B) indicates to the weak density around the main chain of Asn124, Gln125 and Gln127. The map contoured at 1.2 Å.

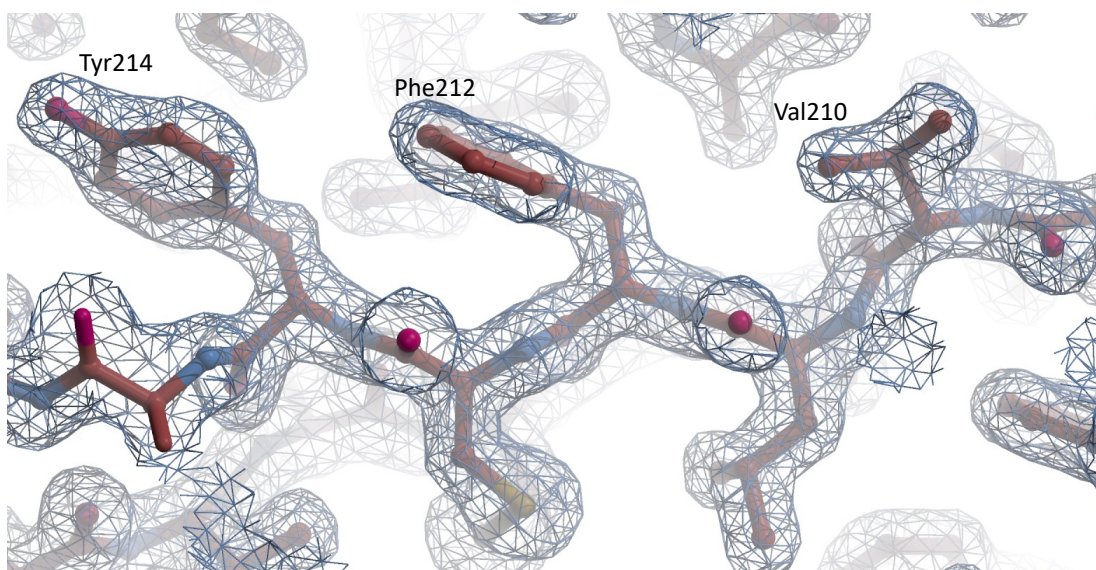


Figure 4.2: *The initial electron density map for SAG-B13. The confirmation of the map around all atoms in strand  $\beta$ D can be seen, such as Tyr214, Phe212 and Val210. The map contoured at 1.2 sigma.*

#### 4.3.1.3 Structure Validation of SAG-B13

The geometry of the SAG-B13 model was analysed using the MolProbity server (Williams *et al.*, 2018). For the 182 residues in the model, 98.9 % (180/182) had favored Ramachandran angles, two residues in the allowed region (Ser117 and Gln125) and no residues were found to be Ramachandran outliers. The protein structure had a clash score of 1.87, ranking in the 99<sup>th</sup> percentile, which indicates a very low number of serious steric overlaps per 1000 atoms. Protein geometry was favorable, with no poor side chain rotamers and a high percentage of favored rotamers at 98.88 %, above the recommended level of >98%. The MolProbity score was notably high at 1.16, achieving the 100<sup>th</sup> percentile for structures of this resolution, reflecting the overall structural quality. In terms of bond and angle accuracy, no bad bonds were observed, and bad angles were present in only 0.16% of cases (2 residues). Overall, the analysis indicates a structurally reliable and well-refined protein model, with metrics largely meeting or exceeding quality goals [Table 4.4].

Table 4.4: MolProbity analysis of the SAG-B13 final model including the geometry.

All-Atom Contacts	Clashscore, all atoms:	1.87	99 <sup>th</sup> percentile* (N=793, 1.71Å ± 0.25Å)	
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Protein Geometry	Poor rotamers	0	0.00%	Goal: <0.3%
	Favoured rotamers	140	98.88	Goal: >98%
	Ramachandran outliers	0	0.00%	Goal: <0.05%
	Ramachandran favoured	180	99.45	Goal: > 98%
	Rama distribution Z-score	0.53 ± 0.58		Goal: abs (Z score) < 2
	MolProbity score <sup>^</sup>	1.16		100 <sup>th</sup> percentile*(N=9166, 1.71Å ± 0.25Å)
	Cβ deviations >0.25Å	0	0.00%	Goal: 0
	Bad bonds:	0 / 1365	0.00%	Goal: 0%
	Bad angles:	2/ 1822	0.16%	Goal: <0.1%
Peptide Omegas	Cis Prolines:	1/ 6	16.67%	Expected: ≤1 per chain, or ≤5%
	Cis non-Prolines	1/175	0.57%	Goal: <0.05%

#### 4.3.1.4 SAG-B13 Overall Fold

The final model of SAG-B13 showed that it had folded into a  $\alpha\beta\alpha$  sandwich structure, in a similar way to that of the previously determined structure of SAG-B19 [Figure 4.3]. The SAG-B13 structure consisted of a four stranded anti-parallel  $\beta$ -sheet surrounded by five  $\alpha$ -helices and connected by ten loops including the N- and C-terminal loops. Using the SAG-B19 labelling convention, the four strands that form the  $\beta$ -sheet involve residues 56-58 (strand A), residues 131-136 (strand B), residues 194-201 (strand C) and 208-216 (strand D). The five  $\alpha$ -helices are constructed from residues 60-72 ( $\alpha$ I), residues 103-111 ( $\alpha$ III), residues 143-161 ( $\alpha$ IV), 180-189 ( $\alpha$ V) and residues 229-237 ( $\alpha$ VI). In SAGB-13 a long loop (L3) joins  $\alpha$ I to  $\alpha$ III, whereas in SAG-B19 an extra helix ( $\alpha$ II) is present in this area. These differences in secondary structure occur despite both SAG-B13 and SAG B-19 being from the same B subfamily of the *E. tenella* SAGs. A full comparison of all the *E. tenella* SAG structures determined is described in Chapter 5.

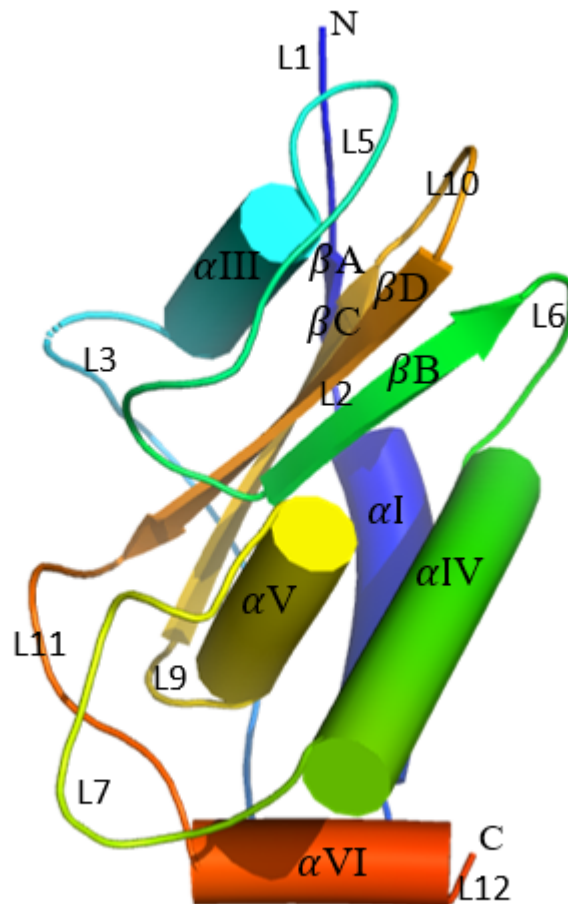


Figure 4.3: **SAG-B13 structure.** Cartoon representation of the SAG-B13 crystal structure coloured from blue (N-terminus) to red (C-terminus) with secondary structural elements labelled (for ease comparison secondary elements numbering followed SAG-B19 structure).

#### 4.3.2 SAG-B16 Data Processing

The second B-family SAG whose structure was determined was SAG-B16. A total of 3600 images were collected from a single crystal grown from ammonium sulphate solutions [Table 4.1] with rotation range of  $0.1^\circ$  per image, with clear spots visible on the diffraction images across the entire resolution range, indicating good crystal quality. For this crystal the xia2 dials pipeline gave the best quality data with a resolution range of 77.82 Å to 1.52 Å and completeness of 100%. The crystal was in space group  $P2_1$  with cell dimensions  $a=44.75$  Å,  $b=77.82$  Å,  $c=52.00$  Å,  $\alpha=90^\circ$ ,  $\beta=98.93^\circ$ ,  $\gamma=90^\circ$  and an overall mean  $\langle I/\sigma(I) \rangle$  value of 12.5 with an  $R_{pim}$  of 0.038 [Table 4.5].

Table 4.5: SAG-B16 detailed statistics for dataset SAD

Overall	
Diamond Beamline	I03
Wavelength (Å)	0.9686
Space group	P2 <sub>1</sub>
Unit cell lengths(Å)	a= 44.75, b=77.82, c=52.00
Unit cell angles (°)	$\alpha= 90, \beta= 98.93, \gamma= 90$
Molecules per asymmetric unit	
Resolution range (Å)	77.82 – 1.52 (1.55 - 1.52)
Total Reflections measured	371811 (18918 - 15417)
Unique Reflections	54120 (2782 - 2671)
Completeness (%)	99.97 (98.56)
Multiplicity	6.87 (5.77)
CC-1/2	0.9990 (0.3349)
Mean I/sigma	12.48 (0.44)
R <sub>merge</sub> (I)	0.092 (2.4721)
R <sub>pim</sub>	0.038 (1.112)

#### 4.3.2.1 Molecular replacement for SAG-B16

Unlike SAG-B13, the Matthews calculation for a single copy of SAG-B16 in the asymmetric unit had a high solvent content (71.3%), with a Matthews coefficient of 4.3 and a very low Matthews probability of 0.03. In contrast, with two copies of the protein in the asymmetric unit, the solvent content decreases significantly to 42.7%, and the Matthews coefficient halves to 2.14, reflecting a more typical protein packing density; this scenario is strongly favoured with a Matthews probability of 0.97. With three copies in the asymmetric unit the solvent content (14.0 %) and the Matthews coefficient of 1.43, were both unrealistic for protein structures, as indicated by a Matthews probability of 0.0. Therefore, the most plausible model for SAG-B16 appears to be with two copies in the asymmetric unit, balancing both solvent content and structural compactness.

The SAG-B19 model was again used for molecular replacement using *Phaser*, but this time searching for two copies, in both space group P2 and P2<sub>1</sub>. The solution in P2<sub>1</sub> gave a rotational function Z-score of 6.5, a translational function Z-score of 12.3 and a substantial LLG of 309, with a significantly higher refined TFZ of 28.8 and overall LLG of 1025, alongside only four packing clashes, which suggested a correct solution. The molecular replacement model had an initial R-factor of 0.32.

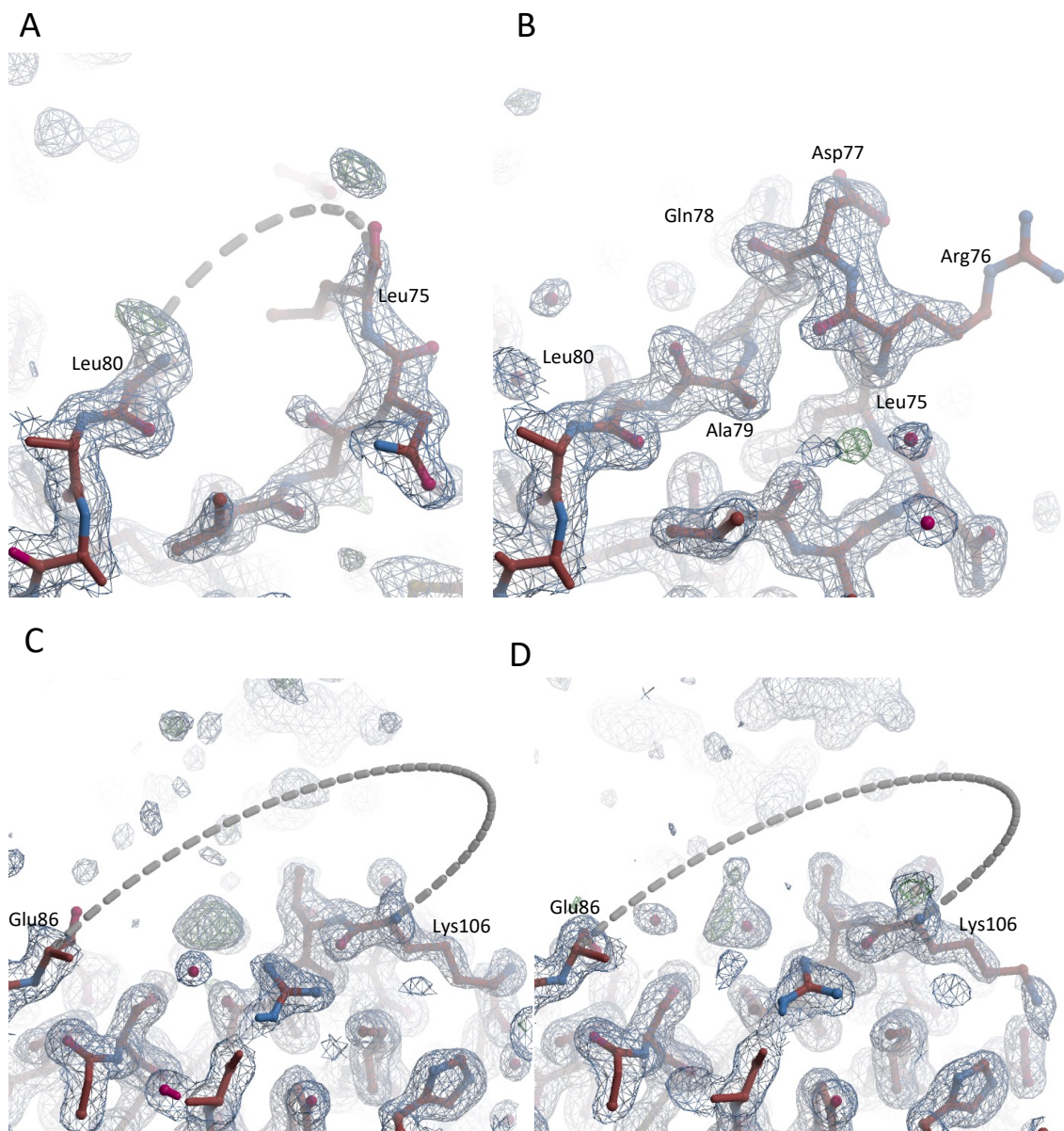
#### 4.3.2.2 Auto-building and Refinement of SAG-B16 *Phaser* Model

In a similar way to SAG-B13, the *Phaser* electron density map was used to auto-build a model of SAG-B16, using Buccaneer. The resultant auto-built model had 364 residues in 4 fragments all successfully assigned to the sequence, with a R-factor of 0.24 and a free-R factor of 0.27. Two SAG-B16 chains could be easily identified, with approximately 93.8% of residues built within these chains. Rounds of model building using Coot and refinement using Refmac5, were used to complete the four Buccaneer fragments into two separate SAG-B16 chains, add solvent molecules and optimise the fit of the model to the density whilst maintaining good model geometry [Figure 4.5]. This resulted in a final model with an  $R_{\text{work}}$  and  $R_{\text{free}}$  of 0.19 and 0.22, respectively [Table 4.6]. 20 residues for both Chain A and Chain B (Glu86-Ser105) could not be modelled into the final electron density map indicating they were disordered, and these residues were omitted from the final model [Figure 4.4 C and D]. In addition, there was also poor map density for residues Arg76, Asp77, Gln78 and Ala79 of chain A [Figure 4.4A] in one loop and these were also omitted from the final model, but can be seen in chain B [Figure 4.4B].

Table 4.6: shows the details of last model for SAG-B16 structure.

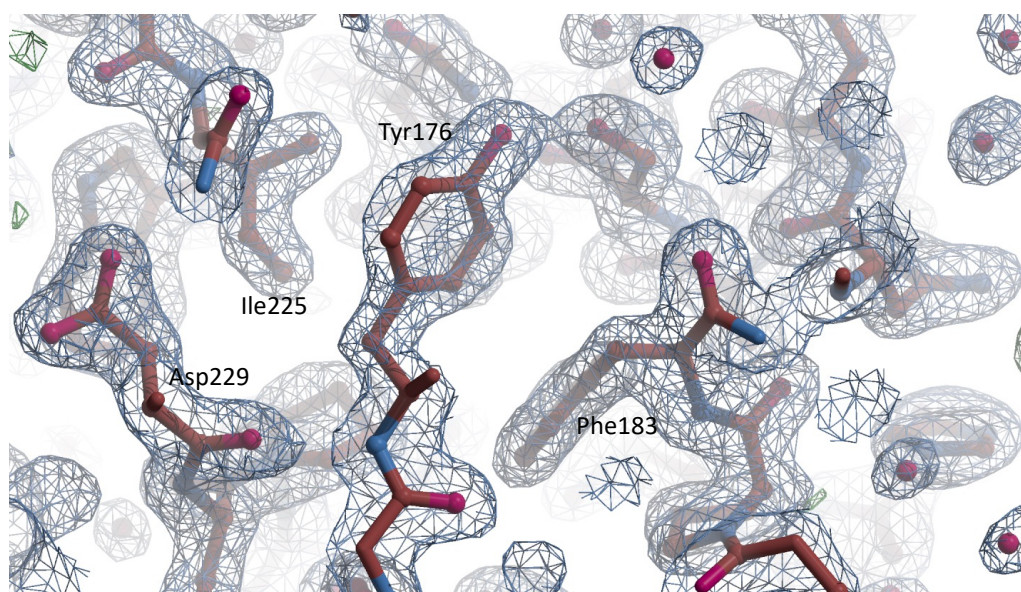
Refinement	
Resolution	1.52
Number of non-H atoms	
Protein chain A/B	2490/2551
Water	226
$R_{\text{work}}/R_{\text{free}}$	0.187/0.224
Average B factor ( $\text{\AA}^2$ )	
Main chain/side chain	17.85/22.48
Water	32.90
Rmsd bond length ( $\text{\AA}$ ) / angle ( $^\circ$ )	0.011/1.987
Ramachandran favoured/allowed (%)	98.80/1.20





**Figure 4.4: The poor electron density of the SAG-B16 map.** A) The final electron density map of SAG-B16 contoured at 1 sigma to show the disordered residues (75Leu – 80Leu) in chain A, whereas (B) these residues can be seen on chain B. (C and D) shows the disordered residues (86 Glu – 106 Lys) of the final model of SAG-B16 on both chain A and B, respectively.





*Figure 4.5: The initial electron density map for SAG-B16 in chain B. The confirmation of the map around all atoms can be seen, such as Tyr176, Phe183, Ile225 and Asp229. The map contoured at 1.2 sigma.*

#### 4.3.2.3 Structure Validation of SAG-B16

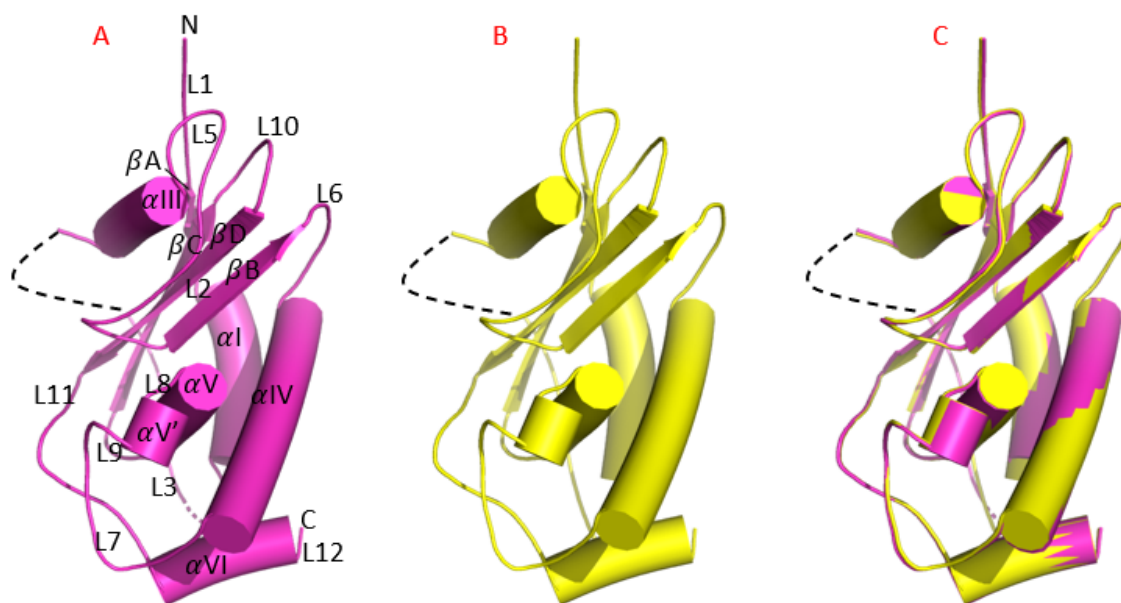
The geometry of the SAG-B16 model was analysed using the MolProbity server. For the 170 residues in chain A and 174 residues of chain B in the model, 98.8 % had favored Ramachandran angles, two residues for chain A in the allowed region (Val72 and Cys123), and another two residues for chain B also in the allowed region (Arg76 and Gly174), and no residues were found to be Ramachandran outliers. The protein structure had a clash score of 2.74, favourable compared to a median of 4.5 within the resolution bin, ranking the structure in the 99<sup>th</sup> percentile. Protein geometry was favorable, with only a single residue (His156 A) with a poor side chain rotamer (corresponding to very weak density for this side chain) and a high percentage of favored rotamers at 98.5 %, above the recommended level of >98%. The overall MolProbity score was notably high at 1.06, achieving the 100<sup>th</sup> percentile for structures of this resolution, reflecting the overall structural quality. In terms of bond and angle accuracy, no bad bonds were observed, and bad angles were present in only 0.2% of cases (2 residues). Overall, the analysis indicates a structurally reliable and well-refined protein model, with metrics largely meeting or exceeding quality goals [Table 4.7].

Table 4.7: MolProbity analysis of the SAG-B16 final model including the geometry.

All-Atom Contacts	Clashscore, all atoms:	2.74	99 <sup>th</sup> percentile* (N=792, 1.52Å ± 0.25Å)	
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Protein Geometry	Poor rotamers	1	0.36%	Goal: <0.3%
	Favored rotamers	274	98.5	Goal: >98%
	Ramachandran outliers	0	0.00%	Goal: <0.05%
	Ramachandran favored	330	98.80	Goal: > 98%
	Rama distribution Z-score	-0.07 ± 0.43		Goal: abs (Z score) < 2
	MolProbity score^	1.06	100 <sup>th</sup> percentile*(N=8806, 1.52Å ± 0.25Å)	
	Cβ deviations >0.25Å	0	0.00%	Goal: 0
	Bad bonds:	0 / 2626	0.00%	Goal: 0%
	Bad angles:	2/ 3169	0.2%	Goal: <0.1%
Peptide Omegas	Cis Prolines:	2/16	12.50%	Expected: ≤1 per chain, or ≤5%
	Cis non-Prolines	2/323	0.62%	Goal: <0.05%

#### 4.3.2.4 Overall Fold of the SAG-B16 Structure

The structure of SAG-B16 showed the same overall fold as SAG-B13 and SAG-B19. The fold is an  $\alpha$ - $\beta$ - $\alpha$  sandwich [Figure 4.6], constructed of a four stranded anti-parallel  $\beta$ -sheet surrounded by five  $\alpha$ -helices with an additional  $3_{10}$  helix, and connected by number of loops. The four strands that form the  $\beta$ -sheet involve residues 56-58 (strand A), residues 137-142 (strand B), residues 200-207 (strand C) and 214-222 (strand D). The five  $\alpha$ -helix include residues 60-72 ( $\alpha$ I), residues 108-116 ( $\alpha$ III), residues 149-167 ( $\alpha$ IV), 182 -184 ( $\alpha$ V'), 186-195 ( $\alpha$ V) and residues 234-242 ( $\alpha$ VI). In the SAG-B16 structure residues Thr87 - Ser105 are disordered and thus it is not known whether helix  $\alpha$ II is present as seen in SAG-B19 or is replaced by a long loop as seen in SAG-B13. The fold of both chain A and Chain B are very similar in the SAG-B16 structure, with the superimposed structures of both chains having an overall rms deviation in C-alpha position of 0.5Å for the 170 residues equivalent in both chains [Figure 4.6 C],



**Figure 4.6: SAG-B16 structure.** A) cartoon representation (purple) of the SAG-B16 chain A with secondary structural elements labelled; B) cartoon representation (yellow) of the SAG-B16 chain B; C) overlaps the SAG-B16 structure chain A (purple) and chain B (yellow). The black dots indicated the disordered residues on both chains.

#### 4.3.3 SAG-B22 Data Processing

The third B-subfamily SAG that had its structure determined was that of SAG-B22. Following successful crystallization from PEG solutions, data were collected from a single crystal in a similar way as before. A total of 3600 images of  $0.1^\circ$  per image were collected to a resolution of  $1.79 \text{ \AA}$  and processed using xia2 dials pipeline at Diamond. The processing revealed a space group  $P4_1 2_1 2$  with cell dimensions of  $a = 57.68 \text{ \AA}$ ,  $b = 57.68 \text{ \AA}$ ,  $c = 209.59 \text{ \AA}$ ,  $\alpha = 90^\circ$ ,  $\beta = 90^\circ$ ,  $\gamma = 90^\circ$ . The data were 100 % complete in the range  $57.68 - 1.79 \text{ \AA}$ , with a mean  $\langle I/\sigma(I) \rangle$  of 6.8, and  $R_{pim}$  of 0.056[Table 4.8].

Table 4.8: SAG-B22 detailed statistics for dataset SAD.

Overall	
Diamond Beamline	I03
Wavelength (Å)	0.9795
Space group	P4 <sub>1</sub> 2 <sub>1</sub> 2
Unit cell lengths(Å)	a= 57.68, b=57.68, c=209.59
Unit cell angles (°)	$\alpha = 90, \beta = 90, \gamma = 90$
Molecules per asymmetric unit	
Resolution range (Å)	57.68 – 1.79 (1.82 – 1.79)
Total Reflections measured	906786 (47358 - 42016)
Unique Reflections	34630 (1982-1678)
Completeness (%)	100.00 (100.00)
Multiplicity	26.18 (25.04)
CC-1/2	0.9970 (0.3327)
Mean I/sigma	6.81 (0.36)
R <sub>merge</sub> (I)	0.283 (6.624)
R <sub>pim</sub>	0.056 (1.341)

#### 4.3.3.1 Molecular Replacement for SAG-B22

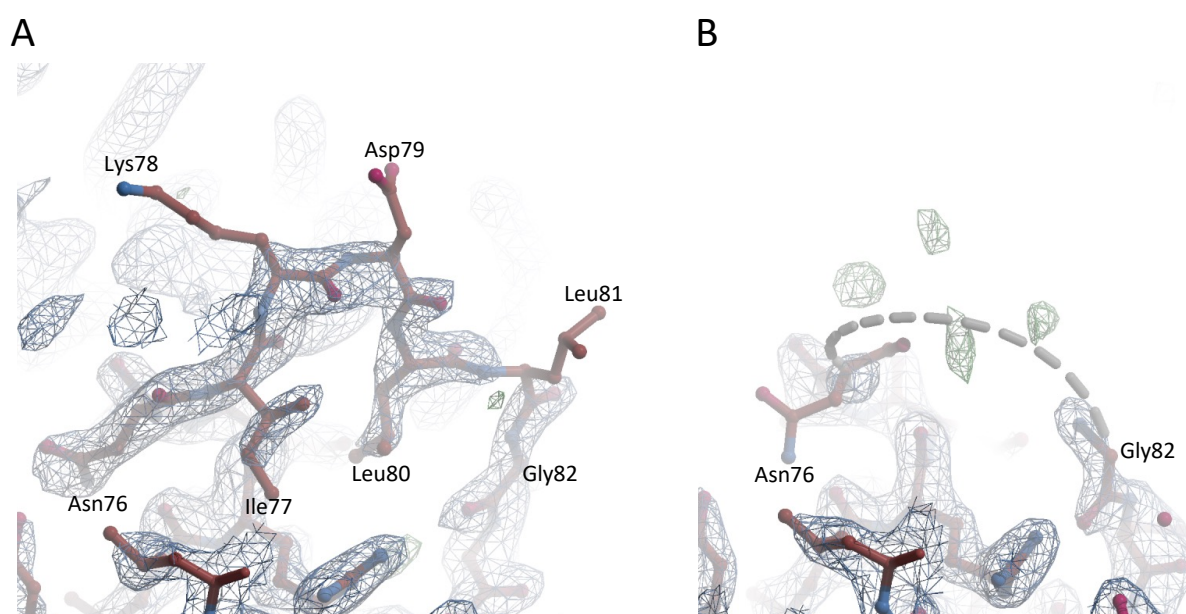
Analysis of the AU contents showed a solvent content of 70.1% for a single copy of SAG-B22, and 40.2% solvent for two copies, with a Matthews probability 0.94. Molecular replacement was thus attempted searching for two copies of the SAG-B19 structure, using Phaser. The molecular replacement was run in both P4<sub>1</sub> 2<sub>1</sub> 2 and P4<sub>3</sub> 2<sub>1</sub> 2 space groups as both these space groups have the same systematic absences in the diffraction pattern and could not be distinguished in the data processing. A clear solution for two chains was found in space group P4<sub>3</sub> 2<sub>1</sub> 2, with a Rotational Function Z-score of 6.8 and a Translational Function Z-score of 11.1, leading to a refined TFZ-equivalent of 56.6. Despite encountering three packing clashes, the model demonstrated a substantial LLG of 199, resulting in an overall LLG of 5032, indicating a strong and favourable fit to the experimental data.

#### 4.3.3.2 Refinement of SAG-B22 Phaser model

The SAG-B22 construct consisted of 198 residues, plus the Ser-Ala linker from the TEV cleavage site at the N-terminus. Buccaneer was used with the construct sequence to auto-build two chains of SAG-B22 into the molecular replacement electron density. This auto-built model consisted of 367 residues built into 2 fragments, with all residues successfully assigned to the sequence, with a refinement R-factor for the model at 0.26, and the free-R factor slightly higher

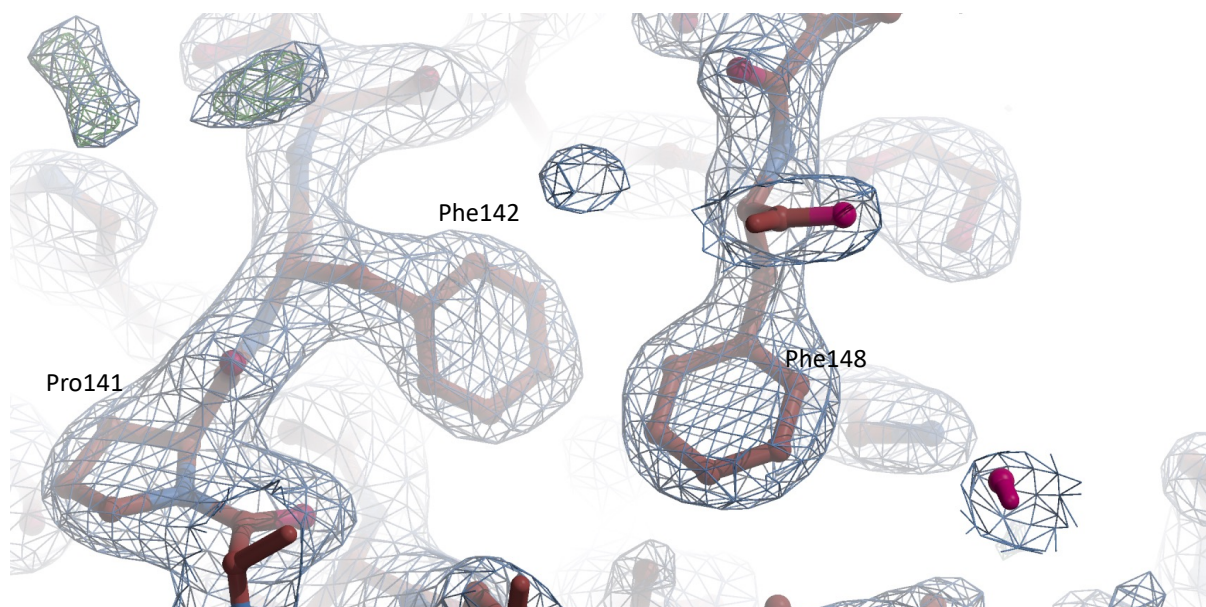
at 0.30. Rounds of manual model building in Coot, followed by refinement using Refmac5 were used to adjust and complete the auto-built model. Some residues had been placed incorrectly by Buccaneer as it can be seen on the map and these were removed from the model. The fit of the side chains and main chain to the density was checked and improved where appropriate, whilst keeping good geometry for the model and water molecules were added to distinct density features where the placed water molecules made sensible hydrogen bonds to the protein. The model underwent multiple rounds of refinement, with each round trying to improve the  $R_{\text{work}}$  and  $R_{\text{free}}$  values. Refinement criteria applied included TLS parametrization, use of non-crystallographic symmetry (NCS) restraints, and a manual weight of 0.12 for restraints versus experimental data. The model was thus improved to a final  $R_{\text{work}}$  and  $R_{\text{free}}$  of 0.20 and 0.25, respectively [Table 4.9 Figure 4.8].

The final model of SAG-B22 consisted of two chains with chain A having 184 residues correctly placed and chain B with 176 residues. Both chains had a considerable portion disordered at the C-terminus (Phe238-Gly248 in chain A and Asn235-Gly248 in chain B) with three residues missing (Ser51, Leu52 and Arg53), at the N-terminus for both chains and a further disordered loop in chain B (Ile77, Lys78, Asp79, Leu80 and Leu81), with the electron density around this region very poor [Figure 4.7].



**Figure 4.7: The poor electron density map of the SAG-B22 final model.** A) shows the weak density of side chain for five residues in chain A. B) shows the poor density for the same five residues (Ile77, Lys78, Asp79, Leu80 and Leu81) but cannot be seen on chain B. the map contoured at 1.0 sigma.





**Figure 4.8:** *The initial electron density map for SAG-B22 in chain A. The confirmation of the map around all atoms can be seen at the start of Loop L6, such as Pro141, Phe142 and Phe148. The map contoured at 1.2 sigma.*

*Table 4.9: shows the details of last good model for SAG-B22 structure.*

Refinement	
Resolution	1.79 Å
Number of non-H atoms	
Protein chain A/B	2626/2466
Ligands (EDO)	45
Water	94
R <sub>work</sub> /R <sub>free</sub>	0.190/0.249
Average B factor (Å <sup>2</sup> )	
Main chain/side chain	22.62/24.63
Ligands	27.39
Water	37.79
Rmsd bond length (Å) / angle (°)	0.0121/1.928
Ramachandran favoured/allowed (%)	97.74/2.86

### 4.3.3.3 SAG-B22 Structural Validation

Unlike the SAG-B13 and SAG-B16, the resolution for SAG-B22 was lower than the others, but at 1.79Å resolution still of sufficient resolution to clearly see the course of the main chain and side chain positions in the electron density map. Using the MolProbity server the geometry of the SAG-B22 model was analysed. For the 184 residues in chain A and 176 residues of chain B in the model, 98.26 % had favored Ramachandran angles, with no residues found to be Ramachandran outliers indicating a well-conformed backbone structure. The protein structure had a clash score of 1.94, favourable compared to a median of 4.5 within the resolution bin, ranking the structure in the 98<sup>th</sup> percentile. Although 0.36% of the side chains had unfavorable rotamers, suggesting minor deviations in some side-chain conformations, the overall protein geometry was favorable. There was a high percentage of favored rotamers at 98.5 %, above the recommended level of >98%. The overall MolProbity score was 0.96, in the 100<sup>th</sup> percentile for structures of this resolution, reflecting the overall structural quality. In terms of bond and angle accuracy, no bad bonds were observed, and bad angles were present in only 0.23% of cases (6 residues) indicating that there might be specific areas that could benefit from further refinement. In summary, the structural validation of SAG-B22 indicates a generally good enough quality protein model with appropriate backbone conformation and reasonable geometric accuracy. [Table 4.10].

Table 4.10: MolProbity analysis of the SAG-B22 final model including the geometry.

All-Atom Contacts	Clashscore, all atoms:	1.94	100 <sup>th</sup> percentile* (N=715, 1.79Å ± 0.25Å)	
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Protein Geometry	Poor rotamers	1	0.36%	Goal: <0.3%
	Favoured rotamers	273	98.5%	Goal: >98%
	Ramachandran outliers	0	0.00%	Goal: <0.05%
	Ramachandran favoured	339	98.26%	Goal: > 98%
	Rama distribution Z-score	-0.62 ± 0.40		Goal: abs (Z score) < 2
	MolProbity score <sup>^</sup>	0.96		100 <sup>th</sup> percentile*(N=12522, 1.79Å ± 0.25Å)
	Cβ deviations >0.25Å	0	0.00%	Goal: 0
	Bad bonds:	0/2654	0.00%	Goal: 0%
	Bad angles:	6	0.23%	Goal: <0.1%
Peptide Omegas	Cis Prolines:	4/13	30.77%	Expected: ≤1 per chain, or ≤5%
	Cis nonProlines	2/339	0.59%	Goal: <0.05%

#### 4.3.3.4 The Overall Fold of the SAG-B22

As expected, the structure of SAG-B22 had the same overall fold as SAG-B19, SAG-B13 and SAG-B16, with the two chains displaying very similar folds (rmsD of 0.7 Å for 176 residues overlapped C- $\alpha$ ) [Figure 4.9]. The four strands that form the  $\beta$ -sheet involved residues 58-60 (strand A), residues 138-142 (strand B), residues 201-207 (strand C) and 216-224 (strand D). The five  $\alpha$ -helices include residues 62-74 ( $\alpha$ I), residues 89-99 ( $\alpha$ II), residues 109-117 ( $\alpha$ III), residues 150-166 ( $\alpha$ IV), 183-185 ( $\alpha$ V') and 187-196 ( $\alpha$ V). As about ten residues in each chain were disordered at the C-terminus of the SAG-B22 structure the final alpha helix ( $\alpha$ VI) in the other B-family SAGs was missing. However, this helix might be present in the full-length SAG-B22 protein and perhaps might be visible in a different crystal form or by using a different length construct.

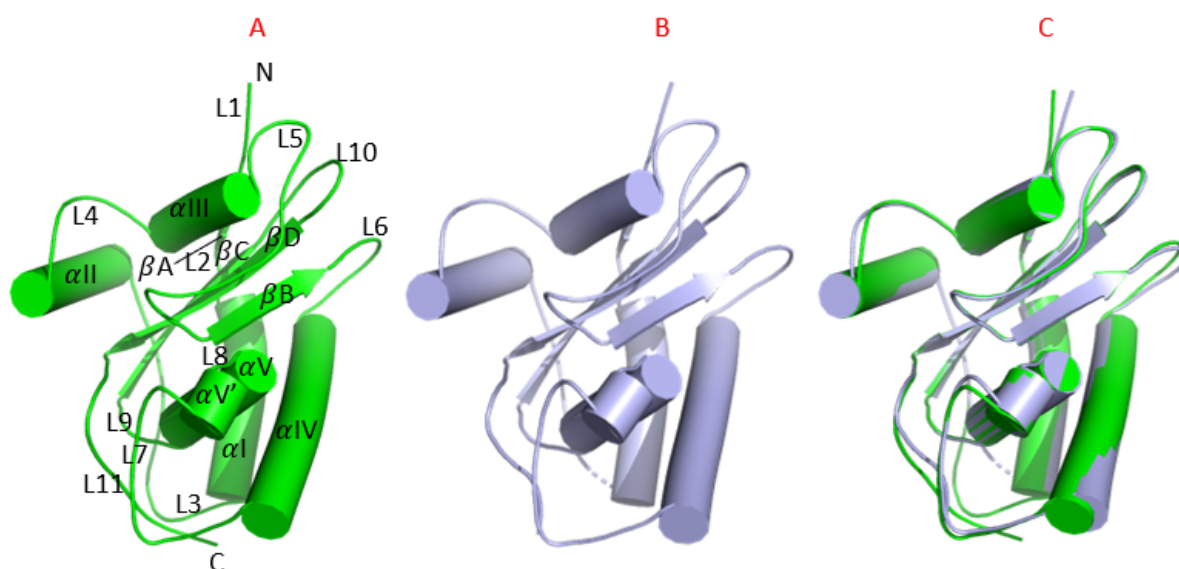


Figure 4.9: **SAG-B22 structure.** A) cartoon representation (green) of the SAG-B22 chain A with secondary structural elements labelled; B) cartoon representation (light-blue) of the SAG-B22 chain B; C) overlaps the SAG-B22 structure chain A (green) and chain B (light-blue). The dots on chain B indicated the disordered residues.



#### 4.3.4 SAG-B41 Data Processing

The final B-family SAG to have its structure determined was SAG-B41. For this SAG, the crystals that grew in PEG 6000 solution diffracted very well, and data collected at Diamond from 3600 images with a rotation range of  $0.1^\circ$  per image were again processed in the xia2/dials pipeline. For these data the space group was P1, with cell dimensions of  $a = 38.44 \text{ \AA}$ ,  $b = 47.80 \text{ \AA}$ ,  $c = 51.37 \text{ \AA}$ ,  $\alpha = 116.90^\circ$ ,  $\beta = 92.77^\circ$ ,  $\gamma = 107.03$ . Although the data extended to  $1.18 \text{ \AA}$  the completeness in the higher resolution shells decreased markedly as the detector had been set for complete data at  $1.5 \text{ \AA}$  [Table 4.12]. Nevertheless, good quality data did extend to  $1.18 \text{ \AA}$  resolution and was included in the refinement, even though the completeness was less than ideal [Table 4.12]. The overall data quality in the range  $44.77 - 1.18 \text{ \AA}$  showed a mean  $\langle I/\sigma(I) \rangle$  value of 23.9 and  $R_{\text{pim}}$  of 0.031. [Table 4.11].

Table 4.11: SAG-B41 detailed statistics for dataset SAD.

Overall	
Diamond Beamline	I03
Wavelength ( $\text{\AA}$ )	0.979539
Space group	P1
Unit cell lengths( $\text{\AA}$ )	$a = 38.44$ , $b = 47.80$ , $c = 51.37$
Unit cell angles ( $^\circ$ )	$\alpha = 116.90$ , $\beta = 92.77$ , $\gamma = 107.03$
Molecules per asymmetric unit	
Resolution range ( $\text{\AA}$ )	44.77 – 1.18 (1.20 – 1.18)
Total Reflections measured	233606 (17851 - 230)
Unique Reflections	66753 (4932-186)
Completeness (%)	66.88 (3.71)
Multiplicity	3.50 (1.24)
CC-1/2	0.9982 (-0.0516)
Mean $I/\sigma$	23.90 (0.32)
$R_{\text{merge}}$ (I)	0.049 (1.130)
$R_{\text{pim}}$	0.031 (1.131)

Table 4.12: the merging-statistics data for SAG-B41 crystal.

d-max	d-min	#obs	#uniq	mult.	%comp	<I>	<I/sI>	r_mrg	r_meas	r_pim	cc1/2
44.80	3.21	17851	4932	3.62	99.16	889.2	97.4	0.027	0.032	0.016	0.998
3.21	2.55	18359	4930	3.72	98.32	363.2	63.5	0.036	0.043	0.022	0.998
2.55	2.22	16847	4889	3.45	97.78	210.2	44.4	0.046	0.055	0.029	0.996
2.22	2.02	17746	4849	3.66	97.23	153.5	35.9	0.057	0.066	0.034	0.996
2.02	1.88	18079	4831	3.74	96.81	90.2	25.1	0.072	0.084	0.043	0.994
1.88	1.77	17039	4789	3.56	96.2	48.6	15.9	0.103	0.121	0.064	0.986
1.77	1.68	16286	4761	3.42	95.7	32.9	11.2	0.128	0.152	0.081	0.981
1.68	1.60	17306	4770	3.63	95.46	25.5	9.4	0.162	0.191	0.099	0.975
1.60	1.54	17446	4713	3.7	94.87	19.7	7.6	0.192	0.225	0.116	0.962
1.54	1.49	17666	4746	3.72	94.64	16.2	6.2	0.222	0.26	0.134	0.957
1.49	1.44	15750	4605	3.42	92.14	12.2	4.4	0.285	0.34	0.183	0.739
1.44	1.40	12270	3869	3.17	77.27	8.7	3.2	0.36	0.434	0.239	0.878
1.40	1.36	8887	2736	3.25	55.37	6.5	2.5	0.469	0.559	0.301	0.826
1.36	1.33	6930	2130	3.25	42.63	5	1.9	0.595	0.709	0.38	0.774
1.33	1.30	5328	1696	3.14	33.87	4	1.6	0.701	0.841	0.457	0.683
1.30	1.27	3965	1283	3.09	25.73	3.4	1.3	0.823	0.984	0.531	0.594
1.27	1.25	2944	953	3.09	19	2.9	1.1	0.974	1.172	0.64	0.482
1.25	1.22	1803	662	2.72	13.44	2.2	0.8	1.267	1.547	0.871	0.382
1.22	1.20	874	423	2.07	8.48	2.1	0.6	1.193	1.56	0.99	0.348
1.20	1.18	230	186	1.24	3.71	1.1	0.3	1.131	1.599	1.131	-0.052
44.77	1.18	233606	66753	3.5	66.88	137.8	23.9	0.05	0.058	0.031	0.998

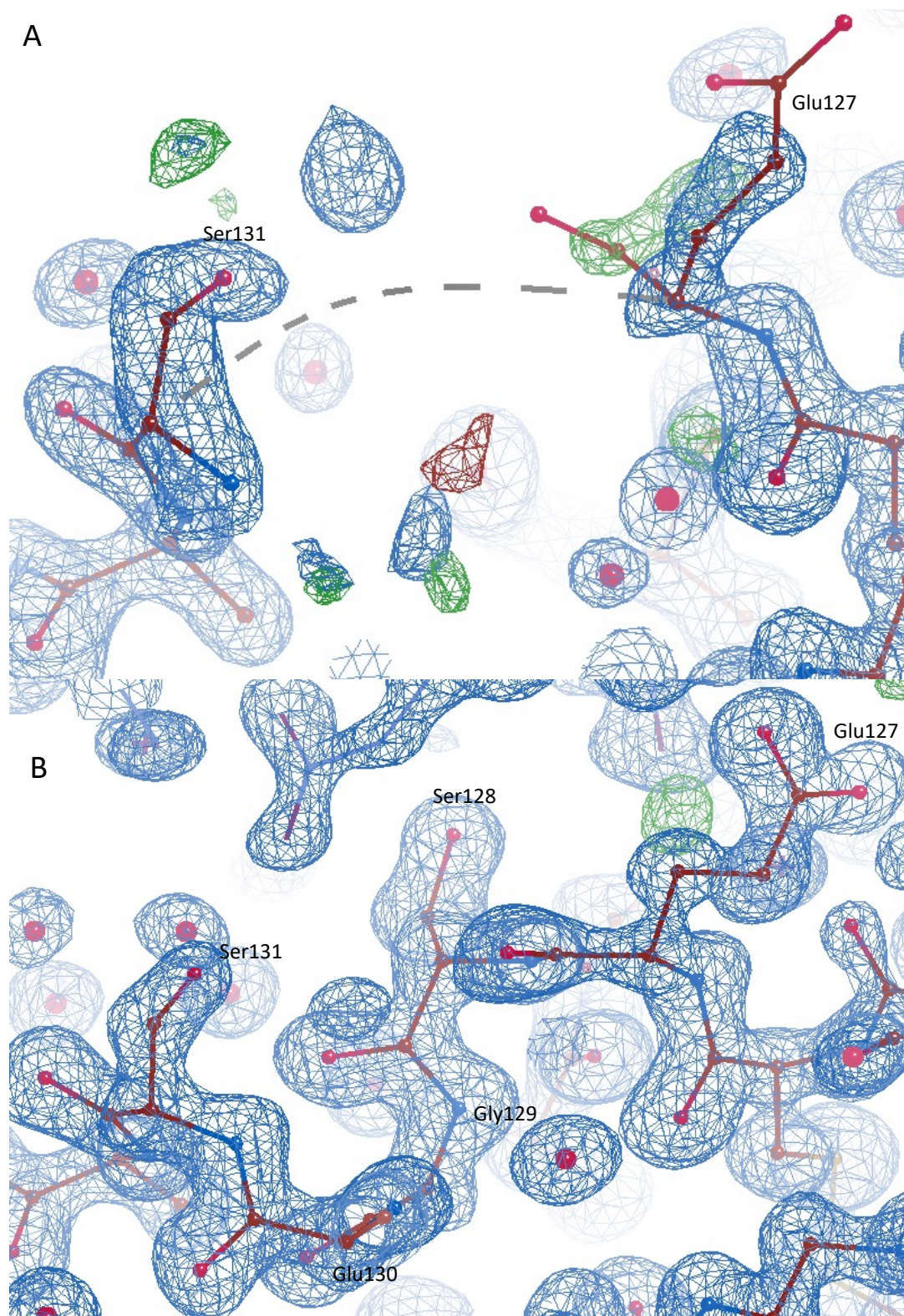
#### 4.3.4.1 Molecular Replacement for SAG-B41

The same logic was followed as that used for the structure determination of the previous SAG protein. The solvent content analysis suggested there could be either one or two (66.4% and 32.86% solvent, respectively) copies of SAG-B41 in the asymmetric unit for this protein, with the Matthews probability for both fairly close (0.51 vs. 0.49). It was thus decided to run Phaser with the SAG-B19 structure as search model in two separate jobs, one searching for a single copy and one searching for two copies. The molecular replacement searching for two copies of SAG-B19 successfully identified a unique solution in space group P1, with a high rotational function Z-score of 14.2 and a translational function Z-score equivalent of 72.8. and refined LLG of 3868, with a refined R-factor of 46.96. Importantly, no clashes were noted in the final solution. These results collectively confirm a correct solution for the molecular replacement step.

#### 4.3.4.2 Refinement of *Phaser* Model of SAG-B41

In a similar way to the other B-family SAG structure determinations, Buccaneer was used to auto-build a model into the molecular replacement electron density map that had been generated from *Phaser*. This auto-built model comprised 395 residues in 2 fragments, each of which was a single chain of SAG-B41. As the construct consisted of 201 residues (including the SAG N-terminal linker) 98.3% of the residues in the construct had been built and assigned to a chain, to give an initial model with an R-factor of 0.20 and a free-R factor of 0.23 suggesting that the model was nearing completion. Subsequent rounds of manual model building and refinement using COOT and Refmac5 were used to improve the model based on the features that appeared on the  $|2mFo-DFc|$  electron density and the  $|Fo-Fc|$  different electron density maps, including the addition of water molecules. This process was repeated until no more electron density features could be interpreted [Figure 4.11].

This process resulted in a final model for chain A of 192 residues (Leu51-Gly245). Three residues (Ser128, Gly129 and Glu130) [Figure 4.10] on loop region and 4 residues at the C-terminal (Leu246, Glu247, Ala248 and Gly249) could not be assigned to density which are disordered and were omitted from the model. For chain B of SAG-B41 191 residues could be placed (Thr55-Gly245), and similarly 4 residues at the N-terminal (Leu51, Leu52, Arg53 and Thr54) and 4 residues at the C-terminal (Leu246, Glu247, Ala248 and Gly249) were omitted from the model due to weak density and presumed disorder. The final model had an R-factor of 0.15 and Free R of 0.19 [Table 4.13].



**Figure 4.10: The poor electron density map of the SAG-B41 final model.** (A) shows the disordered residues on loop region (Ser128, Gly129 and Glu130), whereas (B) shows these residues can be seen, the map contoured at 1.0 sigma.



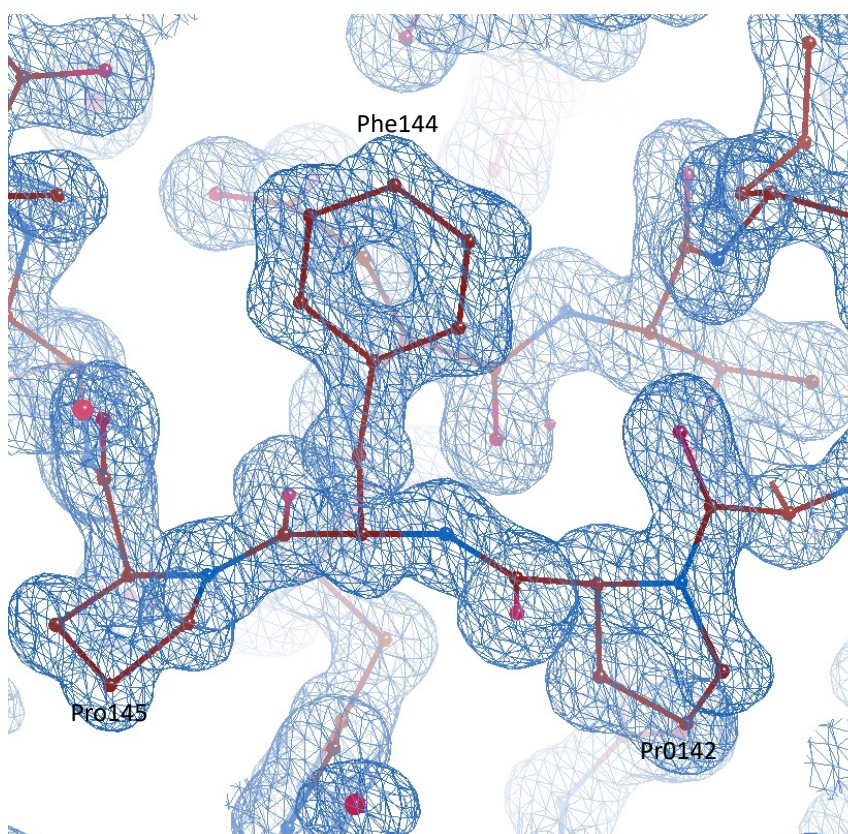


Figure 4.11: *The initial electron density map for SAG-B41 in chain B. The confirmation of the map around all atoms can be seen at the start of Loop L6, such as Pro142, Phe144 and Pro145. The map contoured at 1.2 sigma.*

Table 4.13: *Final refinement statistics for SAG-B41 structure.*

Refinement	
Resolution	1.18 Å
Number of non-H atoms	
Protein chain A/B	2810/2807
Water	339
R <sub>work</sub> /R <sub>free</sub>	0.146/0.188
Average B factor (Å <sup>2</sup> )	
Main chain/side chain	12.70/18.35
Water	23.54
Rmsd bond length (Å) / angle (°)	0.0125/1.957
Ramachandran favoured/allowed (%)	98.67/1.60

#### 4.3.4.3 SAG-B41 Structure Validation

The final model of SAG-B41 structure consisted of two subunits in the symmetric unit, each chain comprised 189 residues. Inspection of the SAG-B41 model quality after using the MolProbity server that the all-atom contacts resulted in a clash score of 2.97, positioning the model in the 94<sup>th</sup> percentile, indicating a relatively low number of serious steric overlaps per 1000 atoms. The protein geometry revealed no poor side chain rotamers and a high 99.1% of favoured rotamers, 98.7% of the residues were in favoured Ramachandran conformations, with 1.6 % (six residues) in allowed regions, and no Ramachandran outliers. The overall MolProbity score was 1.1, placing the structure in the 98<sup>th</sup> percentile, suggesting a high-quality model overall. There were no C $\beta$  deviations greater than 0.25Å, and minimal bad bonds and angles at 0.05% and 0.13%, respectively. The peptide omega analysis showed there were two *cis* proline residues and one non-*cis* proline on both chains [Table 4.14].

Table 4.14: MolProbity analysis of the SAG-B41 final model structure including the geometry.

All-Atom Contacts	Clashscore, all atoms:	2.97	94 <sup>th</sup> percentile* (N=256, 1.18Å ± 0.25Å)	
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Protein Geometry	Poor rotamers	0	0.00%	Goal: <0.3%
	Favoured rotamers	321	98.77%	Goal: >98%
	Ramachandran outliers	0	0.00%	Goal: <0.05%
	Ramachandran favoured	372	98.67%	Goal: > 98%
	Rama distribution Z-score	-0.78 ± 0.38		Goal: abs (Z score) < 2
	MolProbity score^	1.11		98 <sup>th</sup> percentile*(N=9456, 1.18Å ± 0.25Å)
	Cβ deviations >0.25Å	1	0.25%	Goal: 0
	Bad bonds:	01/ 3000	0.05%	Goal: 0%
	Bad angles:	4/ 4090	0.13%	Goal: <0.1%
Peptide Omegas	Cis Prolines:	2/ 14	16.67%	Expected: ≤1 per chain, or ≤5%
	Cis nonProlines	2/370	0.54%	Goal: <0.05%

#### 4.3.4.4 The Overall Fold of SAG-B41

Both chains of the SAG-B41 structure had a very similar structure (rmsD of 0.5 Å for 187 residues overlapped C- $\alpha$ ) [Figure 4.12], which unsurprisingly was the same as that of the other B-family SAGs. The four strands that form the  $\beta$ -sheet involve residues 60-62 (strand A), residues 140-145 (strand B), residues 203-210 (strand C) and 217-225 (strand D), with the six  $\alpha$ -helices formed from residues 64-76 ( $\alpha$ I), residues 91-101 ( $\alpha$ II), residues 111-119 ( $\alpha$ III), residues 152-169 ( $\alpha$ IV), 185 -187 ( $\alpha$ V'), 189-198 ( $\alpha$ V) and residues 237-243 ( $\alpha$ VI). A few residues at both the N and C termini were disordered, but all secondary structure elements could be seen in the model. Both chains of SAG-B41 structure showed the same fold without any dissimilar features. Therefore, the SAG-B41 chain B was chosen for farther analysis as it was more complete [Figure 4.12C].

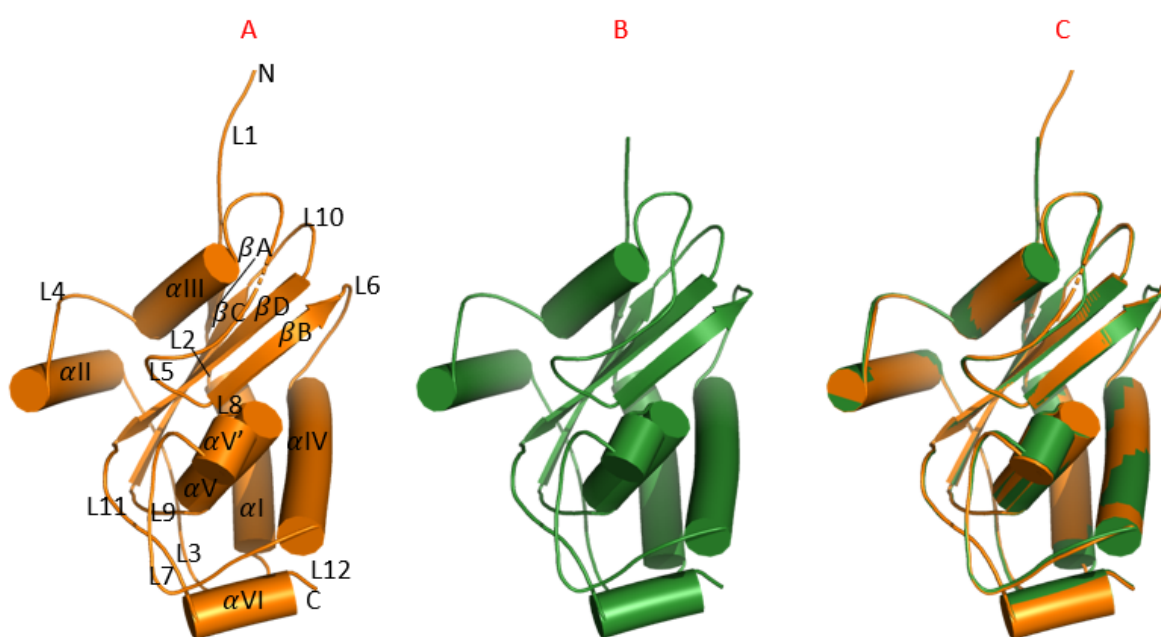


Figure 4.12: **SAG-B41 structure.** A) cartoon representation (orange) of the SAG-B41 chain A with secondary structural elements labelled; B) cartoon representation (green) of the SAG-B41 chain B; C) overlaps the SAG-B41 structure chain A (orange) and chain B (green). The dots on chain A indicated the 3 disordered residues.

### 4.3.5 SAG-A91 Data Processing

Although eight family A SAGs had been successfully expressed and purified, diffraction quality crystals were only obtained for SAG-A91. These crystals grew from solutions of ammonium sulphate [Table 4.1] and data were again collected at the Diamond synchrotron from a single crystal to a resolution of 1.4 Å. A total of 3600 images with a rotation range of 0.1° were processed using the xia2 dials pipeline in space group of P4<sub>1</sub>2<sub>1</sub>2, with unit cell dimensions of a = 58.24 Å, b = 58.24 Å, c = 157.28 Å,  $\alpha = 90^\circ$ ,  $\beta = 90^\circ$ ,  $\gamma = 90^\circ$ . The processed data over a resolution range of 54.61 - 1.40 Å, had a mean  $\langle I/\sigma(I) \rangle$  value of 15.6, an R<sub>pim</sub> of 0.017 and a completeness of 97.4% indicating good data quality [Table 4.15].

Table 4.15: SAG-A91 detailed statistics for dataset SAD.

Overall	
Diamond Beamline	I03
Wavelength (Å)	0.9686
Space group	P4 <sub>1</sub> 2 <sub>1</sub> 2
Unit cell lengths(Å)	a= 58.24, b=58.24, c=157.28
Unit cell angles (°)	$\alpha = 90$ , $\beta = 90$ , $\gamma = 90$
Molecules per asymmetric unit	
Resolution range (Å)	54.23 – 1.40 (1.42 – 1.40)
Total Reflections measured	1118550 (73317 - 13594)
Unique Reflections	52956 (3040-2030)
Completeness (%)	97.41 (76.23)
Multiplicity	21.12 (6.70)
CC-1/2	0.9997 (0.3289)
Mean I/sigma	15.59 (0.17)
R <sub>merge</sub> (I)	0.082 (2.786)
R <sub>pim</sub>	0.017 (1.103)

#### 4.3.5.1 Molecular Replacement

Matthews analysis for this SAG-A91 data showed a solvent content of 63.4 % for one copy of SAG-A91 in the asymmetric unit and a solvent content of 26.9 % for two copies. The predicted solvent content for two copies is well outside the normal range for protein crystals, reflected in a Matthews probability of 0.03. it was thus likely that a single SAG-A91 chain was present in the asymmetric unit and the molecular replacement step searched for just one chain, using the SAG-B19 structure as the search model as before. Similar to the SAG-B22 molecular replacement, two *Phaser* jobs were run to distinguish between space groups P4<sub>1</sub>2<sub>1</sub>2 and P4<sub>3</sub>2<sub>1</sub>2.



A clear solution was found for a single chain in space group  $P4_12_12$ , with a rotation function z-score of 5.3 and translational function z-score of 14.5, culminating in a refined TFZ-equivalent of 37.9. There was only one packing clash and the solution demonstrated a considerable LLG of 244, leading to an overall LLG of 877 and R-factor of 0.32, indicating a favourable fit of the MR model to the experimental data.

#### 4.3.5.2 Refinement of SAG-A91 Model

After obtaining the initial electron density map from *Phaser*, *Buccaneer* was used to auto build a model of SAG-A91 into the molecular replacement map. A total of 185 residues were built in two fragments, with all 185 residues assigned to the sequence. The SAG-A91 construct consisted of 189 residues (Pro28-Gly216), plus the Ser-Ala N-terminal linker and 94.2% of the residues were successfully built, to give an initial model with an R-factor of 0.24 and a free R-factor of 0.28. Manual building in COOT was needed to edit together the two-fragment generated from the *Buccaneer* job, add missing residues where there were features in both the  $|2mFo-DFc|$  electron density and the  $|Fo-Fc|$  different electron density maps. The model then underwent rounds of rebuilding and refinement (using COOT and Refmac5) including addition of water molecules until no new features could be interpreted in the difference map. The final model fitted the density well [Figure 4.13], with an  $R_{work}$  and  $R_{free}$  of 0.17 and 0.21, respectively [Table 4.16].

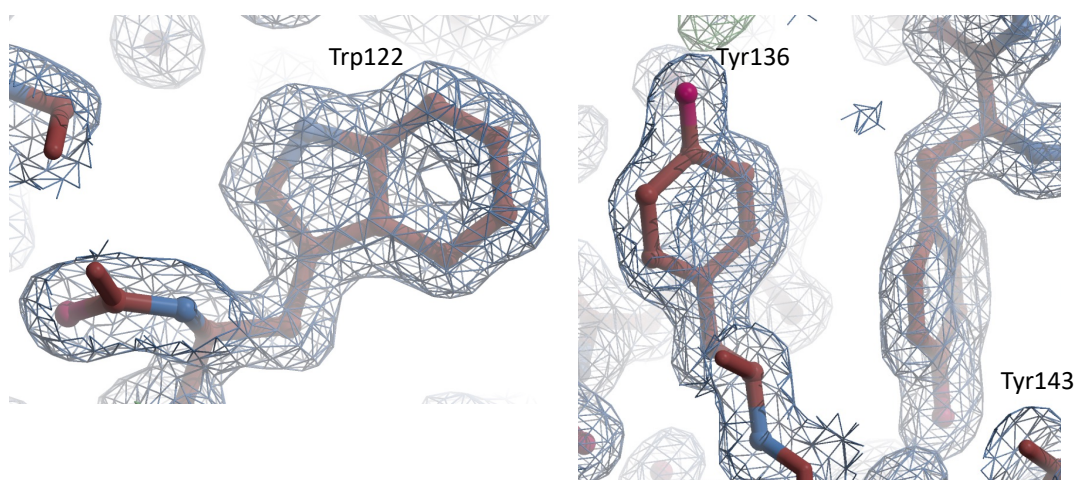


Figure 4.13: *The initial electron density map for SAG-A91. The confirmation of the map around all atoms can be seen, such as Trp122, Tyr136 and Tyr143. The map contoured at 1.2 sigma.*

Table 4.16: shows the details of last good model for SAG-A91 structure.

Refinement	
Resolution	1.40 Å
Number of non-H atoms	
Protein	2682
Water	223
R <sub>work</sub> /R <sub>free</sub>	0.171/0.206
Average B factor (Å <sup>2</sup> )	
Main chain/side chain	17.64/20.73
Water	35.01
Rmsd bond length (Å) / angle (°)	0.0090/1.644
Ramachandran favoured/allowed (%)	98.38/1.62

#### 4.3.5.3 SAG-A91 Structure validation

Analysis of the final model of this SAG-A91 protein showed that 183/187 residues (98.92 %) had favoured Ramachandran angles, with no residues categorized as Ramachandran outliers. Similarly, 145 of the side chains were in favoured rotamers, with no of all residues were found to be in poor side chain rotamer. During the model building and improvement stages side chain rotamers for some residues (e.g. Ser 111, Lys 65, and Ser 176) were chosen to prevent clashes with surrounding atoms or water molecules but still in the favoured rotamers. Nevertheless, the overall MolProbity clash score was 1.84, but this did rank the structure in the 99<sup>th</sup> percentile (N=718, 1.4 Å ± 0.25 Å). None of the 187 residues, were displayed bond length irregularities [Table 4.17]. Additionally, four residues showed bond angle issues (Arg87, Asp127, Glu62, Thr178 and Lys109), because the electron density of the map around these residues was quite weak, and perhaps more refinement could be done to settle these residues angle. Even given these small areas of poor geometry, the overall MolProbity score of 0.94 ranked the SAG-A91 structure in the 100<sup>th</sup> percentile of structures of a similar resolution [Table 4.17].

Table 4.17: MolProbity analysis of the SAG-A91 final model structure including the geometry.

All-Atom Contacts	Clashscore, all atoms:	1.84	99 <sup>th</sup> percentile* (N=718, 1.40Å ± 0.25Å)	
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Protein Geometry	Poor rotamers	0	0.00%	Goal: <0.3%
	Favoured rotamers	145	99.32%	Goal: >98%
	Ramachandran outliers	0	0.00%	Goal: <0.05%
	Ramachandran favoured	183	98.92%	Goal: > 98%
	Rama distribution Z-score	-0.04± 0.59		Goal: abs (Z score) < 2
	MolProbity score <sup>^</sup>	0.94		100 <sup>th</sup> percentile*(N=7266, 1.40Å ± 0.25Å)
	Cβ deviations >0.25Å	0	0.00%	Goal: 0
	Bad bonds:	0/1397	%	Goal: 0%
	Bad angles:	5/1900	0.32%	Goal: <0.1%
Peptide Omegas	Cis Prolines:	1/12	8.33%	Expected: ≤1 per chain, or ≤5%

#### 4.3.5.4 The Overall Fold of the SAG-A91 Structure

Although SAG-A91 shared the same overall  $\alpha\beta\alpha$  sandwich fold of the SAG-B family structures [Figure 4.14], there were some notable differences. First differences was spotted on Loop 10 (SAG-B19 labelling convention) that links strands  $\beta$ C and  $\beta$ D was quite extended with two further short strands (C1 and C2) present in loop L10. These two extra short strands extend the  $\beta$ -sheet away from the C-terminus and the presumed location of the *Eimeria* cell membrane [Figure 4.14]. In SAG-A91 the secondary structure elements are formed by residues 34 -36 (strand-A), residues 102 – 107 (strand-B), residues 162 – 169 (strand-C), residues 173 – 175 (strand-C1), residues 178 – 180 (strand-C2), residues 183 – 191 (strand-D), residues 41 – 51 ( $\alpha$ I), residues 76 – 88 ( $\alpha$ III), residues 115 – 128 ( $\alpha$ IV), residues 146 – 155 ( $\alpha$ V) and residues 204 – 215 ( $\alpha$ VI).

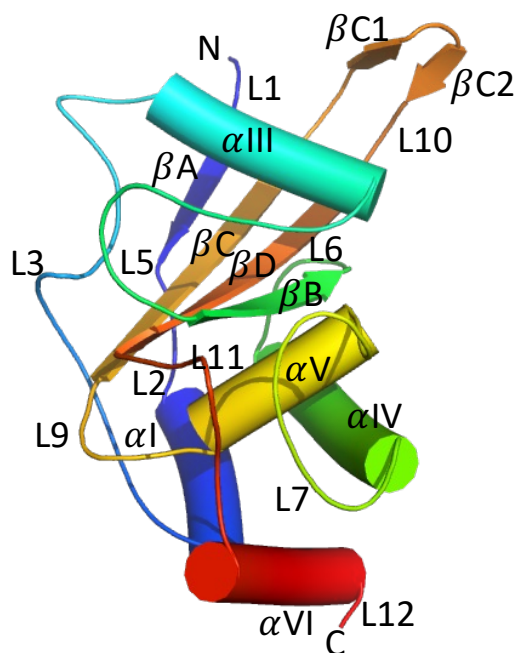


Figure 4.14: **SAG-A91 Structure.** Cartoon representation of the SAG-A91 crystal structure coloured from blue (N-terminus) to red (C-terminus) with secondary structural elements labelled (for ease comparison secondary elements numbering followed SAG-B19 structure).

#### 4.3.6 SAG-C75 Data Processing

The *E. tenella* genome contains only a single instance of a family C SAG, and it was fortunate that diffraction quality crystals were obtained from ammonium sulphate solutions for SAG-C75, this single member [Table 4.1]. Several crystals were sent to the Diamond synchrotron for data collection, each using 3600 images of  $0.1^\circ$  rotation. For the best crystal, data extended to  $1.32 \text{ \AA}$ , processed using the xia2 3dii pipeline in space group  $P4_12_12$  with cell dimensions  $a = 60.44$ ,  $b = 60.44$ ,  $c = 112.10 \text{ \AA}$  and  $\alpha = 90^\circ$ ,  $\beta = 90^\circ$ ,  $\gamma = 90^\circ$ . The data over the range  $42.74 - 1.32 \text{ \AA}$  were complete and had a mean  $\langle I/\sigma(I) \rangle$  value of 15.7, with  $R_{pim}$  of 0.017 indicating good quality data, although quite weak in the highest resolution shell [Table 4.18].

Table 4.18: Table 4.12: SAG-C75 detailed statistics for dataset SAD.

Overall	
Diamond Beamline	I04
Wavelength (Å)	0.8517
Space group	P4 <sub>1</sub> 2 <sub>1</sub> 2
Unit cell lengths(Å)	a= 60.44, b=60.44, c=112.10
Unit cell angles (°)	$\alpha = 90, \beta = 90, \gamma = 90$
Molecules per asymmetric unit	
Resolution range (Å)	42.74 – 1.32 (1.34 – 1.32)
Total Reflections measured	1310322 (67702 - 60687)
Unique Reflections	49651 (2751-2434)
Completeness (%)	100.00 (100.00)
Multiplicity	26.39 (24.93)
CC-1/2	0.9998 (0.2930)
Mean I/sigma	15.69 (0.24)
R <sub>merge</sub> (I)	0.084 (5.254)
R <sub>pim</sub>	0.017 (1.069)

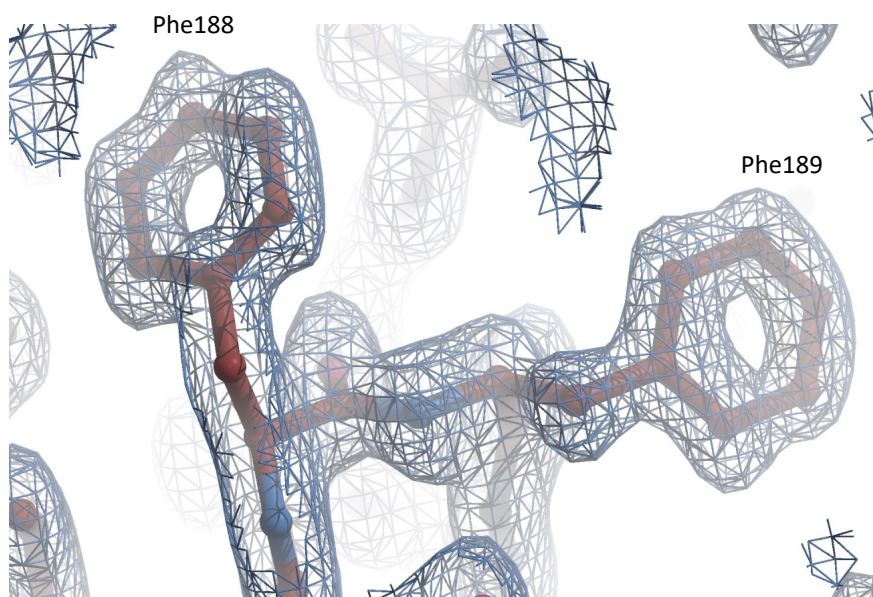
#### 4.3.6.1 Molecular Replacement

The molecular replacement process for SAG-C75 again used the same methodology as used for the other *E. tenella* SAGs. Matthews analysis suggested that a single copy of SAG-C75 was present in the asymmetric unit (56% solvent, 0.99 Matthews probability). *Phaser* was run with SAG-B19 as the search model, looking to place a single copy in either space group P4<sub>1</sub>2<sub>1</sub>2 or P4<sub>3</sub>2<sub>1</sub>2. A clear solution was achieved in space group P4<sub>1</sub>2<sub>1</sub>2, resolving the space group ambiguity, with a rotation function Z score of 4.8, a translation function Z score of 8.1 a refined TFZ-equivalent score of 18.8, an overall LLG gain of 375 and R-factor of 0.43. This all indicated that a correct solution had been found.

#### 4.3.6.2 Refinement of SAG-C75 model

Buccaneer was again used to auto build a model of SAG-C75 into the molecular replacement electron density map. This resulted in a model of 171 residues (98.3% of the construct) built in three fragments, with all 171 residues successfully assigned to the SAG-C75 sequence, with an R-factor of 0.35 and a free-R factor of 0.37 indicating that although correct, the model needed substantial manual rebuilding. Rounds of manual rebuilding and refining using COOT and Refmac5, including addition of water molecules, TLS refinement produced a final model with R-factor and R-free at 0.17 and 0.21, respectively, and RMS deviations for bonds and angles

are measured at 0.0124 and 1.857, respectively, which underscored the quality of the model [Table 4.19 Figure 4.15].



*Figure 4.15: The initial electron density map for SAG-C75. The confirmation of the map around all atoms can be seen, for example, Phe188 and Phe189. The map contoured at 1.2 sigma.*

Table 4.19: shows the details of last good model for SAG-C75 structure.

Refinement	
Resolution	1.32 Å
Number of non-H atoms	
Protein	2530
ligands	32
Water	143
R <sub>work</sub> /R <sub>free</sub>	0.175/0.209
Average B factor (Å <sup>2</sup> )	
Main chain/side chain	27.85/32.21
Water	41.49
ligands	49.52
Rmsd bond length (Å) / angle (°)	0.0120/1.86
Ramachandran favoured/allowed (%)	96.97/1.21

#### 4.3.6.3 SAG-C75 Structure Validation

The final model of this SAG-C75 protein consisted of 171 residues with 164 with favoured Ramachandran angles, no residues categorized as Ramachandran outliers and with Ser166, Thr124, Ser196, Thr104, Arg52 and Asn133 with allowed Ramachandran angles. In terms of all-atom contacts, the clash score stands at 0.77, indicating a minimal occurrence of serious steric overlaps ( $> 0.4$  Å) per 1000 atoms. Protein geometry analysis highlighted the quality of the side chain rotamers, with 138 residues, (97.9%) in allowed conformations with no residues considered to be having poor rotamers [Table 4.20]. These findings collectively demonstrate a reliable quality of the validated SAG-C75 structure, reflecting its suitability for further analysis and eventual deposition in the Protein Data Bank (PDB).

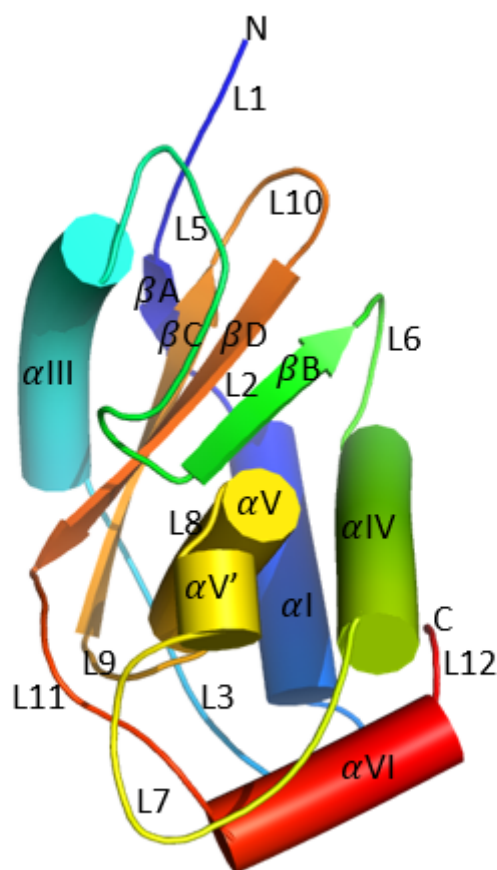
Table 4.20: MolProbity analysis of the SAG-C75 final model structure including the geometry.

All-Atom Contacts	Clashscore, all atoms:	0.77		99 <sup>th</sup> percentile* (N=374, 1.32Å ± 0.25Å)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Protein Geometry	Poor rotamers	0	0.00%	Goal: <0.3%
	Favoured rotamers	138	97.87%	Goal: >98%
	Ramachandran outliers	0	0.00%	Goal: <0.05%
	Ramachandran favoured	164	96.47%	Goal: >98%
	Rama distribution Z-score	-0.39 ± 0.61		Goal: abs (Z score) < 2
	MolProbity score^	0.97		99 <sup>th</sup> percentile* (N=2291, 1.32Å ± 0.25Å)
	Cβ deviations >0.25Å	0	0.00%	Goal: 0
	Bad bonds:	0 / 1298	0.00%	Goal: 0%
	Bad angles:	3 / 1665	0.28%	Goal: <0.1%
Peptide omegas	Cis Prolines:	2 / 8	25.00%	Expected: ≤1 per chain, or ≤5%

#### 4.3.6.4 The Overall Fold of SAG-C75

The overall fold of the SAG-C75 was again similar to that seen in both the SAG-A and SAG-B family proteins, but the loop regions did show differences in length and positions [Figure 4.16]. The secondary structure was folded from residues 36-38 (strand-A), residues 98-102 (strand-B), residues 154-161 (strand-C), residues 168-175 (strand-D), residues 41-52 ( $\alpha$ I), residues 66-79 ( $\alpha$ III), residues 108-122 ( $\alpha$ IV), residues 136-138 ( $\alpha$ V'), residues 140-149 ( $\alpha$ V), residues 186-195 ( $\alpha$ IV).





*Figure 4.16: **SAG-C75 structure.** Cartoon representation of the SAG-C75 crystal structure coloured from blue (N-terminus) to red (C-terminus) with secondary structural elements labelled (for ease comparison secondary elements numbering followed SAG-B19 structure).*

#### 4.4 Summary

The successful determination of high-quality structures for representatives from each of the three *E. tenella* SAG families described in this chapter has shown they all share the same basic fold [Figure 4.17] and has allowed a thorough comparison of the structure function relationships between them, which is detailed in the following chapter. For ease of comparison a table of the residues in the secondary structure elements for each SAG structure described is shown below [Table 4.21].

Table 4.21: Showed the  $\beta$  and  $\alpha$  length of residues for each SAG.

	Residues number of secondary structure									
	$\beta$ A	$\beta$ B	$\beta$ C	$\beta$ D	$\alpha$ I	$\alpha$ II	$\alpha$ III	$\alpha$ IV	$\alpha$ V	$\alpha$ VI
SAG-B13	3	6	8	9	13	na	9	19	10	10
SAG-B16	3	6	8	9	13	na	9	19	10	9
SAG-B22	3	5	8	8	15	11	9	19	10	na
SAG-B41	3	6	8	9	14	11	8	18	10	7
SAG-C75	3	5	8	8	12	na	15	15	10	10
SAG-A91	4	6	9	9	11	na	13	14	10	12
SAG-B19	3	6	7	9	14	12	11	20	10	9

na—indicates to the disordered/missing region.

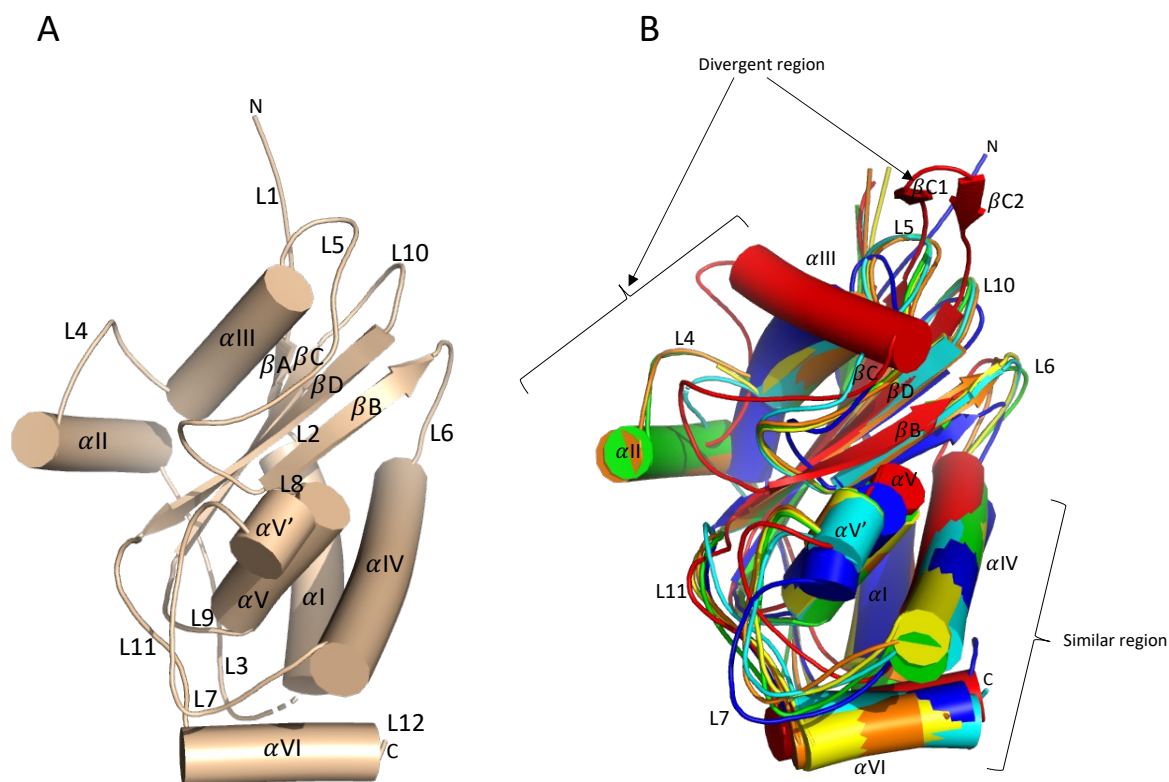


Figure 4.17: **The *E. tenella* SAG structure.** A) SAG-B19 structure, with secondary element labelled (wheat, PDB:6zzb). B) A superposition of the six *E. tenella* structures described in this chapter illustrating the similarities differences between them: SAG-B13 (cyan), SAG-B16 (yellow), SAG-B22 (green), SAG-B41 (orang), SAG-A91 (red), SAG-C75 (blue).

## Chapter5 Structural analysis of *Eimeria Tenella* SAG proteins

### 5.1 Overview

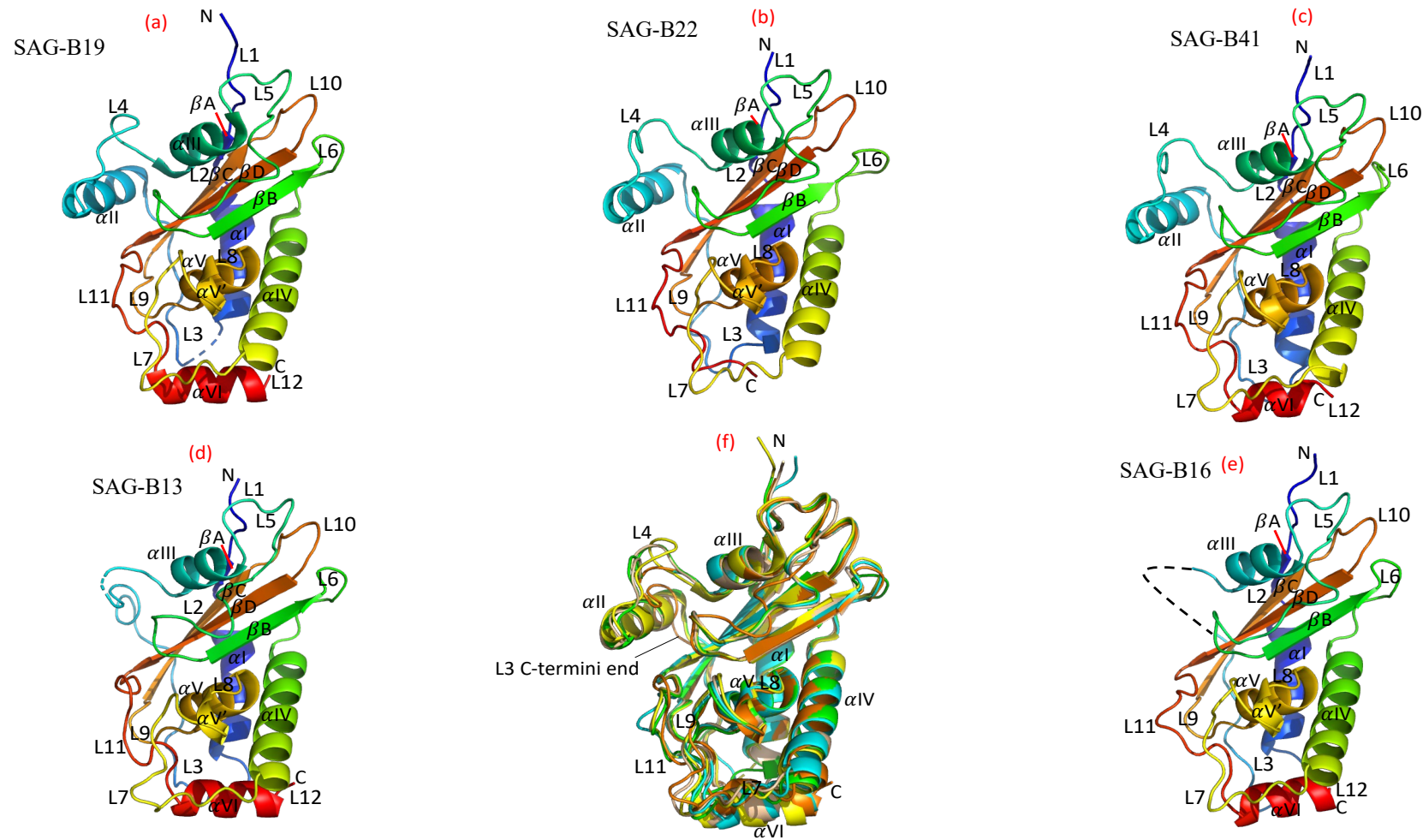
To aid in the analysis of the structure-function relationships of the SAG family of proteins, six additional *Eimeria* SAG proteins have had their structures successfully determined at high resolution as part of this thesis to complement the pre-existing structure of SAG-B19 (see chapter 4). Four of these new SAG structures belong to the SAG-B family (SAG-B13, SAG-B16, SAG-B22, and SAG-B41), with one example of each of a SAG-A and a SAG-C family member (SAG-A91 and SAG-C75). In this chapter the similarities and differences in the sequences and molecular structures of these SAGs are compared to deepen the understanding the potential roles of these proteins in host-parasite interactions. As three of the SAG B proteins (SAG-B16, SAG-B22 and SAG-B41) were crystallised with two molecules in the asymmetric unit, the most complete chain was used for the purpose of the comparison [Table 5.1].

*Table 5.1: Summary of the SAG family members for which a structure has been determined.*

SAG	Family	Chain in A.U	Chain used for the analysis	PDB code
SAG-B19	B	1	Chain A	6zzb
SAG-B13	B	1	Chain A	TBC
SAG-B16	B	2	Chain B	TBC
SAG-B22	B	2	Chain A	TBC
SAG-B41	B	2	Chain B	TBC
SAG-A91	A	1	Chain A	TBC
SAG-C75	C	1	Chain A	TBC

## 5.2 Exploring the Structural Similarity in the SAG-B Family.

Comparison of the four proteins from the SAG-B family whose structures have been determined (SAG-B13, SAG-B16, SAG-B22, and SAG-B41) revealed that they share a very similar 3D fold to SAG-B19 [Figure 5.1 and 5.2], each containing four anti-parallel  $\beta$ -strands in the core, surrounded by five or six  $\alpha$ -helices. In all of these structures the loop regions are substantial, accounting for approximately 50% of the entire structure, with some loops being particularly large, for example, the ~20-residue region that forms loop 5, which links the helices  $\alpha$ III and strand  $\beta$ B in all the SAG-B family members. Inspection of the five proteins showed that on one face of the  $\beta$ -sheet the flanking structures comprising  $\alpha$ I,  $\alpha$ IV,  $\alpha$ V,  $\alpha$ V',  $\alpha$ VI, L7, L8, L9 and L11 are very similar [Figure 5.1 and 5.2]. However, on the other face of the  $\beta$ -sheet there are some differences at the C-terminal end of L3,  $\alpha$ II and L4 [Figure 5.1 and 5.2]. In SAG-B19 four residues of loop L3 are disordered (Glu77-Lys80), but in SAG-B41 and SAG-B22, all the residues in this loop can be clearly seen in the density. However, in SAG-B13 no helix equivalent to  $\alpha$ II exists and instead L3 and L4 merge into one long twenty-seven residue loop in which five residues (Pro95-Ser99) disordered. For SAG-B16, eighteen residues (Thr87-Ser105) which are equivalent to those forming  $\alpha$ II and L4 in SAG-B19 are disordered.



**Figure 5.1: Structure of the five SAG-B members.** (a-e) SAG-B19, SAG-B22, SAG-B41, SAG-B13, and SAG-B16) The 3D structure of these proteins labelled following the convention used with SAG-B19. Each structure consists of four beta-strands (labelled A-D) forming an anti-parallel beta sheet flanked by six alpha-helices (labelled I-VI) together with a single region of  $3^{10}$  helix labelled as V'. Twelve connecting loops (L1-L12) join the regions of regular secondary structure. Panel (f) shows the overlap of the structure of SAG-B19 (yellow) with SAG-B41(magentas), SAG-B22 (green), SAG-B16 (cyan) and SAG-B13 (oranges).

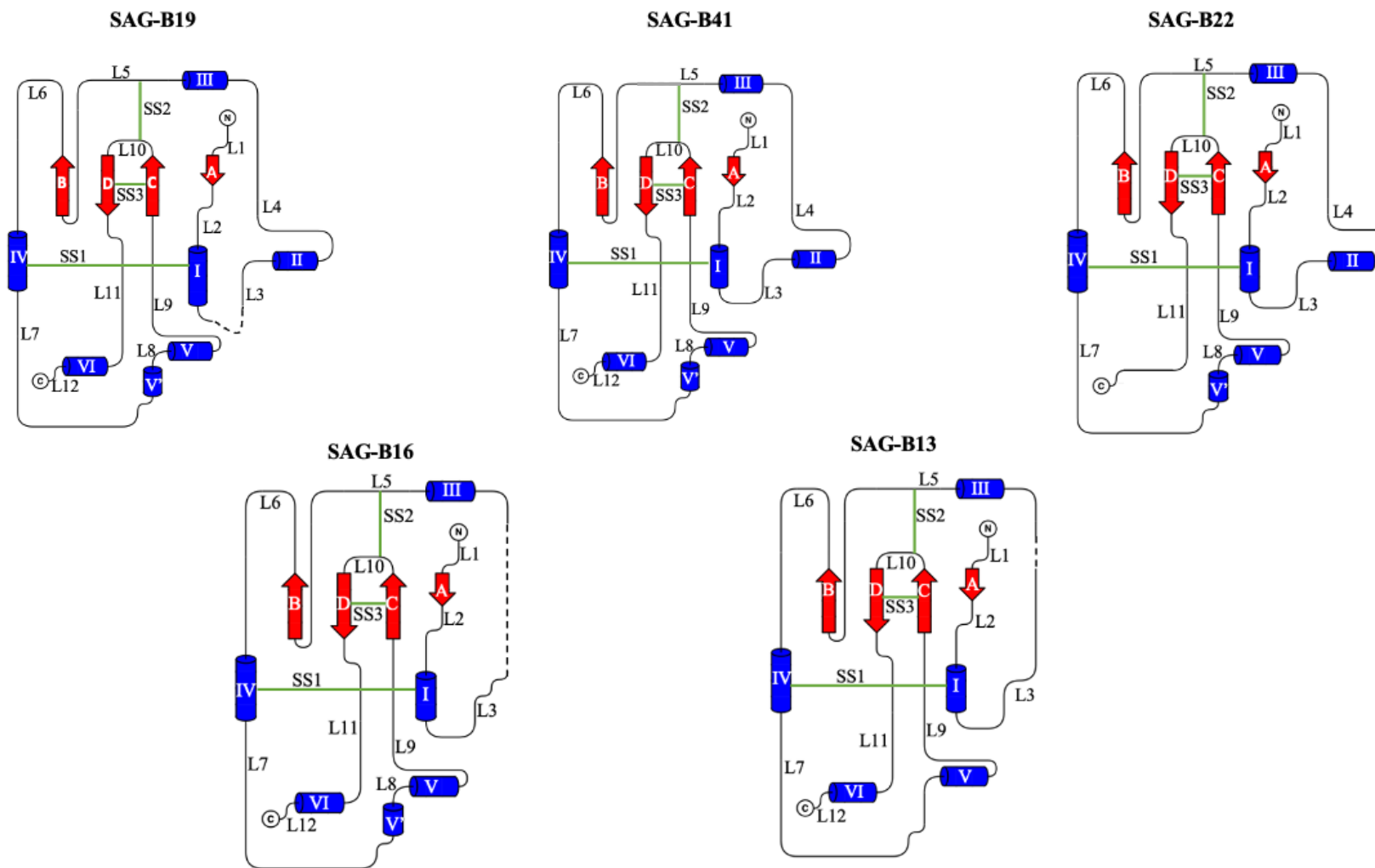


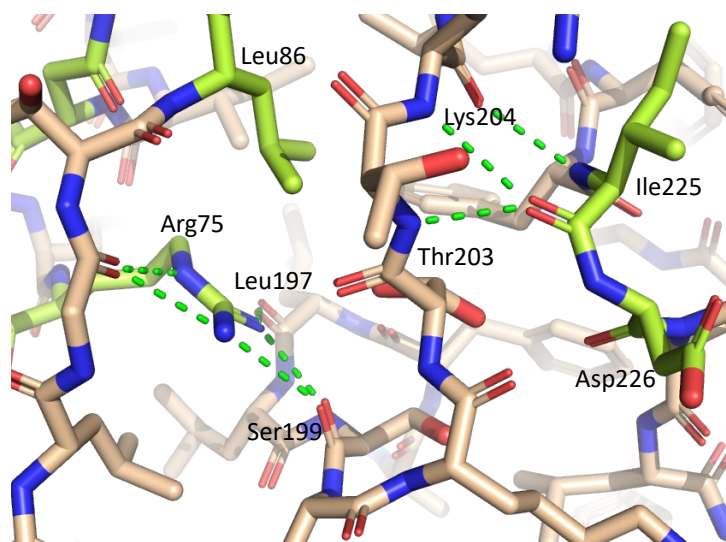
Figure 5.2: *The 2D topology diagram of five SAG-B members. The relative position and connection of secondary structural elements are shown with dashes indicating regions corresponding to disordered residues. The positions of the three disulphide bonds are represented by green bars and are labelled.*

### 5.2.1 Disulphide bonds and cis-peptides in the SAG-B members.

The structures of the five SAG-B proteins all contain six conserved cysteine residues [Table 5.2], which form three disulphide bonds linking the same secondary structure elements and that presumably contribute to the stability of the fold. These disulphide bonds are indicated as SS1, which connects  $\alpha$ I to  $\alpha$ IV; SS2 that links loop 5 to loop 10 and SS3 that connects  $\beta$ C to  $\beta$ D. These SAG-B protein structures also contain either two or three *cis* peptides with two of these are conserved in the B-family members. One of the conserved *cis* peptides involves a *cis*-proline residue (Pro185 in SAG-B19 and its equivalents in the other SAGs). The other is a highly unusual non-proline *cis* peptide (Ile225 - Asp226) that forms a tight turn in the all five SAG-B structures in which the peptide N-H and C-O groups are involved in multiple hydrogen bonds to the main chain of residues in the 197-204 loop (SAG-B41 numbers). In turn the main chain of these residues' forms part of the pocket that accommodates the arginine side chain of the NxxR motif with the neighbouring side chain of Leu86, a residue conserved in the family A-, B- and C- SAGs interacting with the guanidinyll group of the arginine [Figure 5.3]. This pattern of interactions explains the very strong conservation of these structural features. In addition, both SAG-B22 and SAG-B19 contain a second *cis* proline on the L4 loop but this *cis* proline is not conserved in the sequence of the other family B SAGs whose structures have been determined [Table 5.2].

Table 5.2: Disulphide bridges and cis-peptides in the SAG-B family structures.

SAG	SS1	SS2	SS3	Conserved non-Pro <i>cis</i> peptide	Conserved <i>cis</i> -Pro	other <i>cis</i> -Pro
SAG-B19	C67-C152	C126-C213	C207-C222	I225-D226	A184-P185	E105-P106,
SAG-B41	C67-C152	C126-C213	C207-222	I225-D226	T184-P185	na
SAG-B22	C65-C150	C124-C211	C205-C220	I223-E224	A182-P183	I99-P100,
SAG-B16	C63-C149	C123-210	C204-C219	V222-E223	A181-P182	na
SAG-B13	C63-C143	C118-C204	C198-C213	I216-Q217	P175-P176	na



*Figure 5.3: **Key interactions creating the pocket for the conserved buried arginine.** The hydrogen bonds between the main chain carbonyl of Ile225 that forms part of the Ile225-Asp226 non-Pro cis-peptide in SAG-B41 and the main chain NHs of Thr203 and Lys204 that lie in the loop next to the guanidiny group of the buried arginine Arg75. The green dots indicated the hydrogen bonding.*

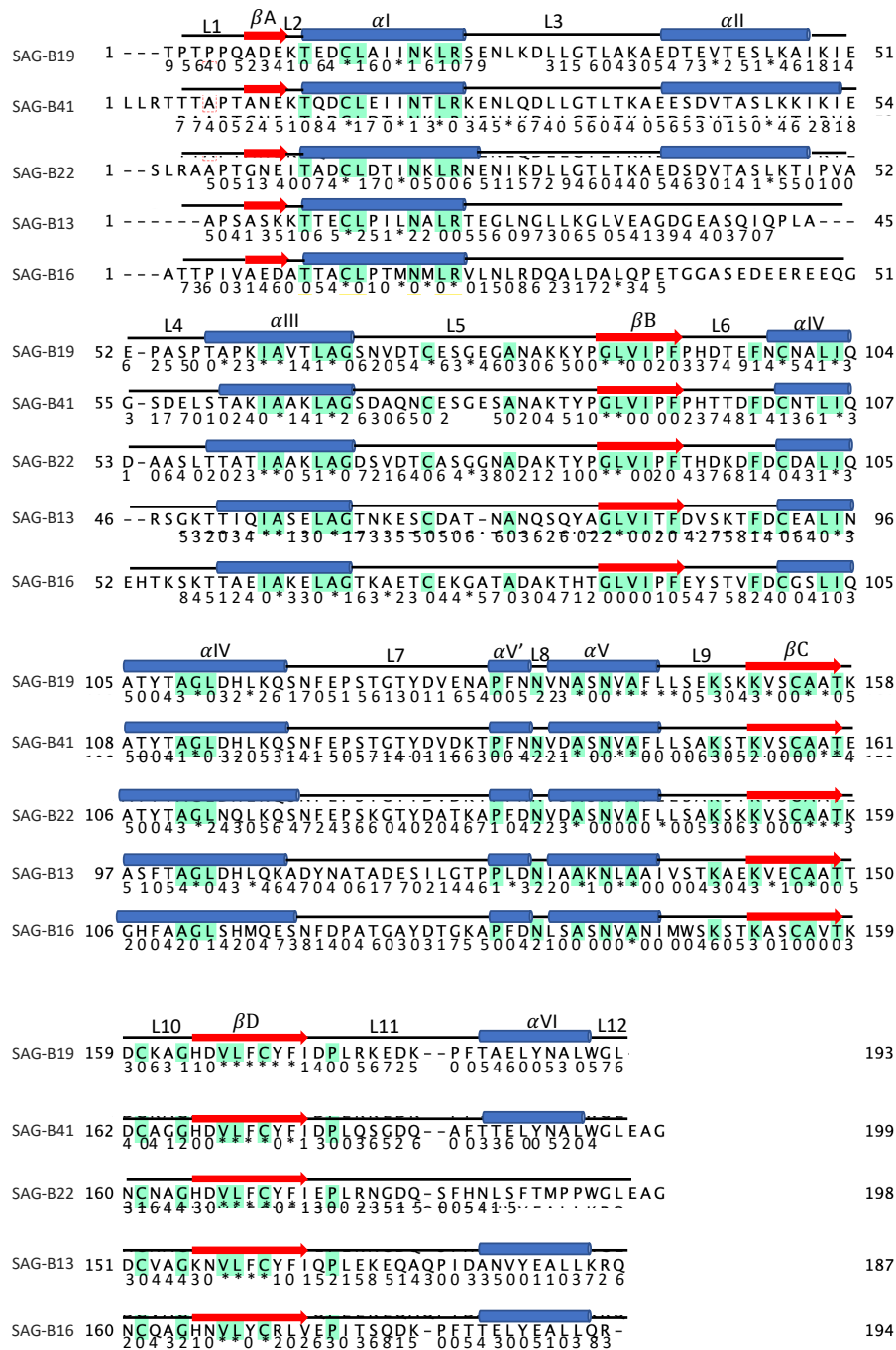
## 5.2.2 Exploring the Divergence within the SAG-B Family

To further analyse the similarities and differences between the five SAG B family proteins their structures were superimposed using GESAMT (Krissinel, 2012), and this superposition was used to produce a structure-based sequence alignment for the five proteins [Figure 5.4]. Comparing each structure to that of SAG-B19, the rmsD for C $\alpha$  positions were closely related and varied between 0.84 Å to 1.2 Å [Table 5.6] with, as expected, the greatest similarity being between SAG-B19 and SAG-B41, the two proteins sharing the highest sequence identity (74%) [Table 5.5]. Comparison of the overlapped structures revealed that the major differences occur in the surface loop regions that decorate the structures, with differences in both the conformation of the loops and of the loop lengths being seen [Table 5.3]. The biggest difference between the structures lies in the region of the L4 loop [Figure 5.5] with minor differences occurring in L1, L3 and L11.

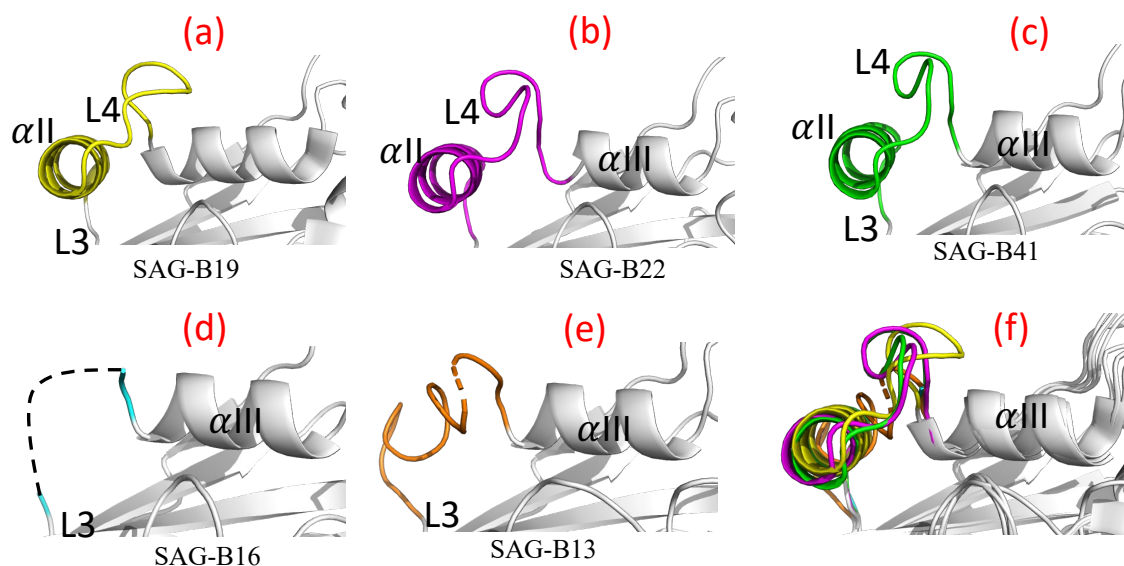


Table 5.3: Shows the length of the loop region of SAG-B group.

	Number of residues in each loop											
SAG	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	L11	L12
SAG-B19	6	1	9	8	19	5	14	1	5	6	10	2
SAG-B41	5	1	13	10	20	6	15	1	4	6	11	1
SAG-B22	4	1	12	9	20	6	14	1	4	6	14	disordered
SAG-B16	5	1	14	disordered	20	6	14	1	4	6	11	1
SAG-B13	4	1	22	disordered	19	6	14	1	4	6	12	1



**Figure 5.4: Multiple sequence alignment of the five SAG-B members that were successfully structures determined.** The sequences alignment crosses the full length of each SAG protein, the 41 conserved residues are highlighted in green. The positions of the loops and secondary structure elements in these proteins are shown above the sequences and labelled. The solvent accessible area of residues in these proteins were calculated by the method of Lee and Richards (Lee and Richards, 1971). The resulting accessible areas were then expressed as a fraction of the total solvent accessible surface area for each amino acid in a GXG tripeptide and are given as a sequence of integers 0-9 under each SAG sequence (where each number represent residues which have 1-10%, 11-20%, 21-30%, etc of the surface solvent accessible. Totally buried residues indicated by \*



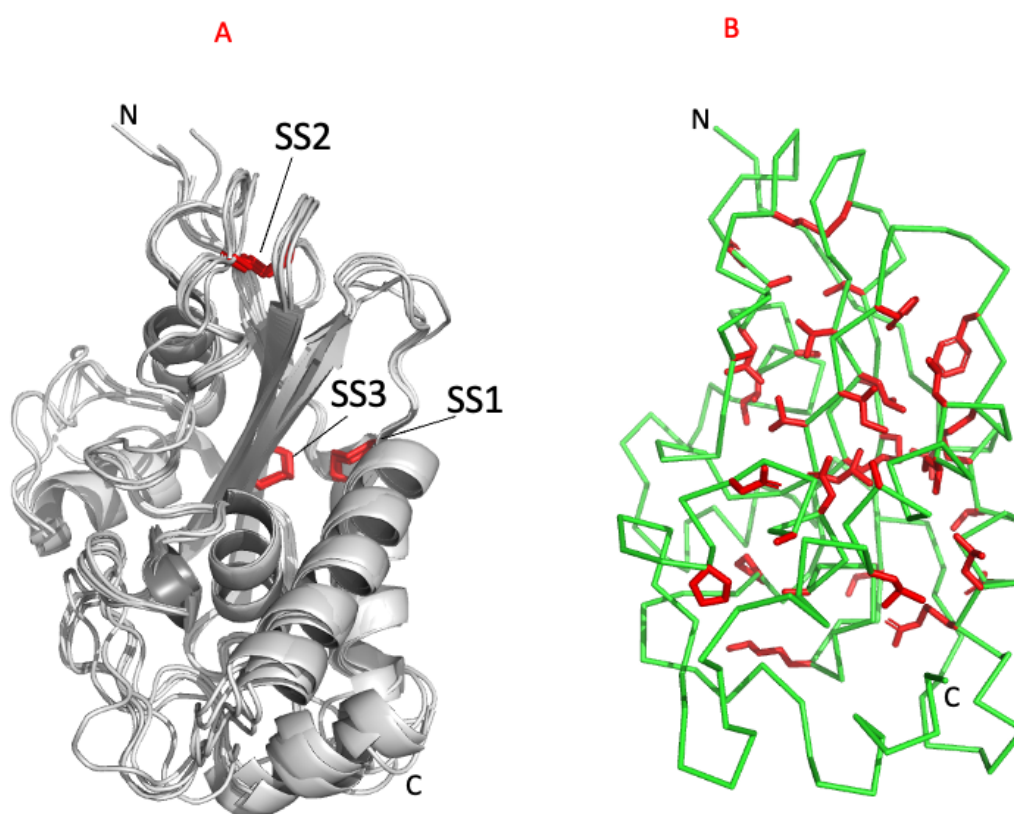
**Figure 5.5: Diagram to show the variation in the structure of the L4 loop in the five SAG-B proteins.** (a-e) SAG-B19, SAG-B22, SAG-B41, SAG-B16 and SAG-B13. (f) a superposition of the five structures. Note that the position of  $\alpha$ III is very similar but the L4 loop shows considerable variation.

### 5.2.3 Structure-Based Sequence Alignment of Family B SAG Proteins

The structure-based alignment of the five SAG-B proteins whose structures have been determined revealed that 41 residues are absolutely conserved [Figure 5.4B and Figure 5.6]. Of these, the majority (33) are hydrophobic, including the 6 cysteines, 7 Alanines, 4 Glycines, and 16 bulky hydrophobic residues with only 8 of the conserved residues being hydrophilic. Moreover, every one of these 41 conserved residues, including the hydrophilic ones, are largely buried (defined as having an accessible surface area of less than 20%). As the conserved residues have no appreciable surface accessibility, they are all presumably involved in maintaining the overall structure of the SAG proteins, rather than forming the sites of possible interactions to other molecules. In addition, the solvent accessible surface for each residue was calculated using the method of Lee and Richards (Lee and Richards, 1971), as implemented in AreaMol ('The CCP4 suite: programs for protein crystallography.', 1994) and plotted on the alignment to identify whether specific residues were exposed or buried.

Of the eight conserved hydrophilic residues of the SAG-B family (Thr64, Asn72, Arg75, Asn 188, Asn193, Lys 201, Lys 204 and Thr210), the  $\text{NH}_3^+$  groups on the two Lysine are solvent accessible and, in addition, in SAG-B41 each makes a salt bridge interaction to an adjacent residues (Asp232 and Asp93, respectively), with the negative charge of these two residues

mostly conserved across the SAG-B family [Figure 5.7]. Of the remaining six buried conserved hydrophilic residues the hydrophilic side chains form hydrogen bonds [Table 5.4] to adjacent main chain NH and carbonyl groups, linking the secondary structures. Finally, the buried conserved arginine residue (Arg75) forms part of the conserved NxxR motif seen in all *Eimeria* SAG proteins and in the wider CAP protein superfamily (Gibbs *et al.* 2008). The guanidyl group of the arginine makes four hydrogen bonds to Gly84, Ser199, and Leu197 [Table 5.4]. This intricate array of hydrogen bonds underscores the structural and functional importance of these hydrophilic residues, anchoring them within the protein's core. Their role is presumably very important in stabilizing the protein's three-dimensional structure and may also influence its biological activity, reflecting an evolutionary adaptation to leverage these polar side chains for critical internal interactions within the SAG-B protein family.



**Figure 5.6: The conserved 41 residues across the five SAG-B members.** A) The overlaps of five SAG-B members with the three conserved disulphide bonds drawn as red sticks on the cartoon representation of the structure. B) The ribbon representation of SAG-B41 structure showing the position of the 41 conserved residues drawn as red sticks.

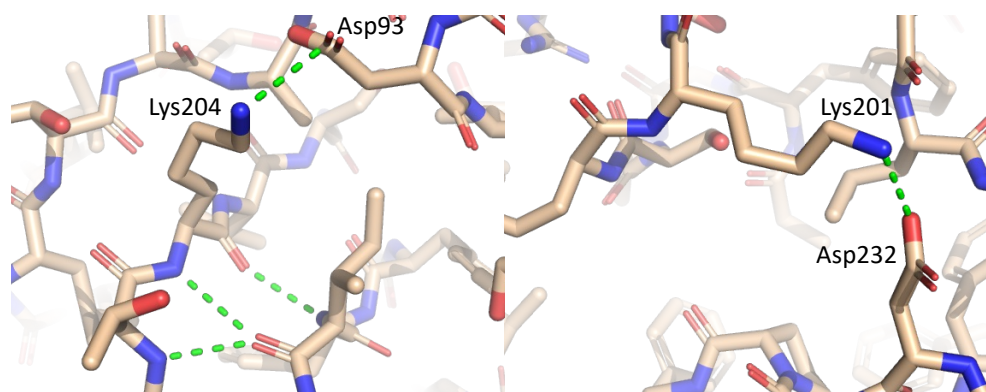


Figure 5.7: **Key interaction creating the salt bridges.** The  $\text{NH}_3^+$  groups on the two lysine residues (Lys201 and Lys204) are making the salt bridges with the negatively charge residues (Asp232 and Asp93).

Table 5.4: Hydrogen bonds of conserved hydrophilic residues across the five SAG-B family showed in SAG-B19 structure.

Hydrogen bond	Distance (Å)	Hydrogen bond	Distance (Å)
(N72) OD---N L86	2.77	(T64) OG---N (C207)	2.90
(N72) ND---O (L86)	3.03	(N188) OD1---N (A191)	2.93
(R75) NE---O (G84)	2.77	(N188) OD2---N (A190)	3.01
(R75) NH1---O (S199)	2.87	(N193) ND---O (V141)	2.96
(R75) NH1---O (L197)	2.81	(N193) OD---N V141	2.91
(R75) NH2---O (S199)	3.12	(T210) OG---N (V219)	3.01
(R75) NH2---O (G84)	3.04		

Overall, the analysis of the pattern of sequence conservation showed that, whilst the buried residues have a tendency to be conserved, the major areas of sequence difference between these proteins including the positions of insertions/deletions occurred in the surface loop regions, where most residues have an accessible surface area of greater than 20% [Figure 5.7]. To discover whether the lack of sequence identity in the surface accessible residues also extended to different patterns of surface charge, the surface electrostatic properties for each of the five SAG-B structures were calculated in PyMOL (Schrödinger & DeLano. 2020), [Figure 5.8]. It can be clearly seen that there is no common pattern of surface charge across these five SAG-B structures, with the chemical characteristics of their surfaces being quite different. Thus, there do not appear to be any conserved surface charged patches that could be involved in interactions with host molecules. For example, the clear negative surface patch around the C-terminal end of SAG-B13 (patch 1) is of mixed character in SAG-B41 and SAG B-19, positive and negative in SAG-B16 and slightly negative in SAG-B22 [Figure 5.8]. The differences in surface charge are more pronounced in the parts of the structure remote from the C-terminus that are presumed distal from the *Eimeria* cell membrane. For example, the largely positive surface adjacent to the N-terminus of SAG-B19 (patch 2) is of mixed or negative character in the other SAG-B structures [Figure 5.8]. Similar differences in the surface characteristics can be seen elsewhere on the structures.

*Table 5.5: Sequence identity and structural similarity between the SAG-B members.*

	SAG-B19	SAG-B41	SAG-B22	SAG-B16	SAG-B13
SAG-B19 (%)	100	74 (0.84Å) *	67 (1.08Å) *	49 (1.04Å) *	44 (1.20Å) *
SAG-B41 (%)		100	71	50	43
SAG-B22 (%)			100	50	42
SAG-B16 (%)				100	43
SAG-B13 (%)					100

\*numbers in parentheses refer to the rmsD superpositions against SAG-B19 (see Table 5.6)

Table 5.6: The structural similarity between the structures of the SAG B proteins determined as part of this thesis against SAG-B19 generated by using GESAMT.

<i>E. tenella</i> SAG ID	Residues in construct	Number of disordered residues	Residues in X-ray structure	Equivalenced residues to SAG-B19	Sequence Identity of equivalenced residues (%)	C-alpha rmsD
SAG-B13	186	5	181	170	44.1	1.20 Å
SAG-B16 Chain B	192	18	174	168	49.4	1.04 Å
SAG-B22 Chain A	194	0	194	175	66.9	1.08 Å
SAG-B41 Chain A	198	0	198	182	74.2	0.84 Å
SAG-A91	188	0	188	144	16.7	1.94 Å
SAG-C75	170	0	170	153	27.0	1.95 Å

While suggestions for the function of the SAG proteins in *Eimeria* parasites have included possible roles in cell adhesion or parasite invasion (Reid *et al.*, 2014; Wang *et al.*, 2023), the precise function of these proteins is currently unknown. Taken together, the buried nature of all the conserved residues and the lack of any patterns of sequence conservation or surface charge on the surface of the protein suggests that these five SAG-B members are not involved in binding to a host receptor as this would require conservation of a patch of exposed residues and/or a patch of surface charge.

Although the functional implications of these structural differences remain uncertain, it is possible that the role of these B-family SAG proteins that have quite similar structures, but quite different surface properties, is linked to the parasite's ability to evade the host's immune system, then give a chance for a particular member of these SAG to play the role in the invasion process. (Ramly *et al.*, 2021).

#### 5.2.4 All *E. tenella* SAG-B Family

The structural and sequence analysis of these five SAG-B proteins was extended to cover all the 26 SAG proteins belong to the SAG-B family in *Eimeria tenella*, (Reid *et al.*, 2014). The extended family displays a blend of similarity and diversity, as evidenced by their percentage identity [Appendix A.2]. For example, some SAG-B family proteins (ETH\_00026045 and ETH\_00008700) exhibit a striking 99% sequence identity, suggesting that they have essentially identical structures and functionalities. In contrast, other B-family members, such as

ETH\_00013160, stand out with notably lower similarity (around 35-40%) compared to other proteins in the family. This allows for the possibility that these proteins may have unique features or specialized roles or perhaps not expressed at any stage of the lifecycle. This combination of high similarity within specific subgroups and broader variability across the entire SAG-B family indicates the presence of both conserved and divergent regions, which could be key to understanding the family's characteristic functions, unique functionalities, and evolutionary relationships. The other possibility of these SAG-B family might be play a role in hormonal stimulation of the parasite.



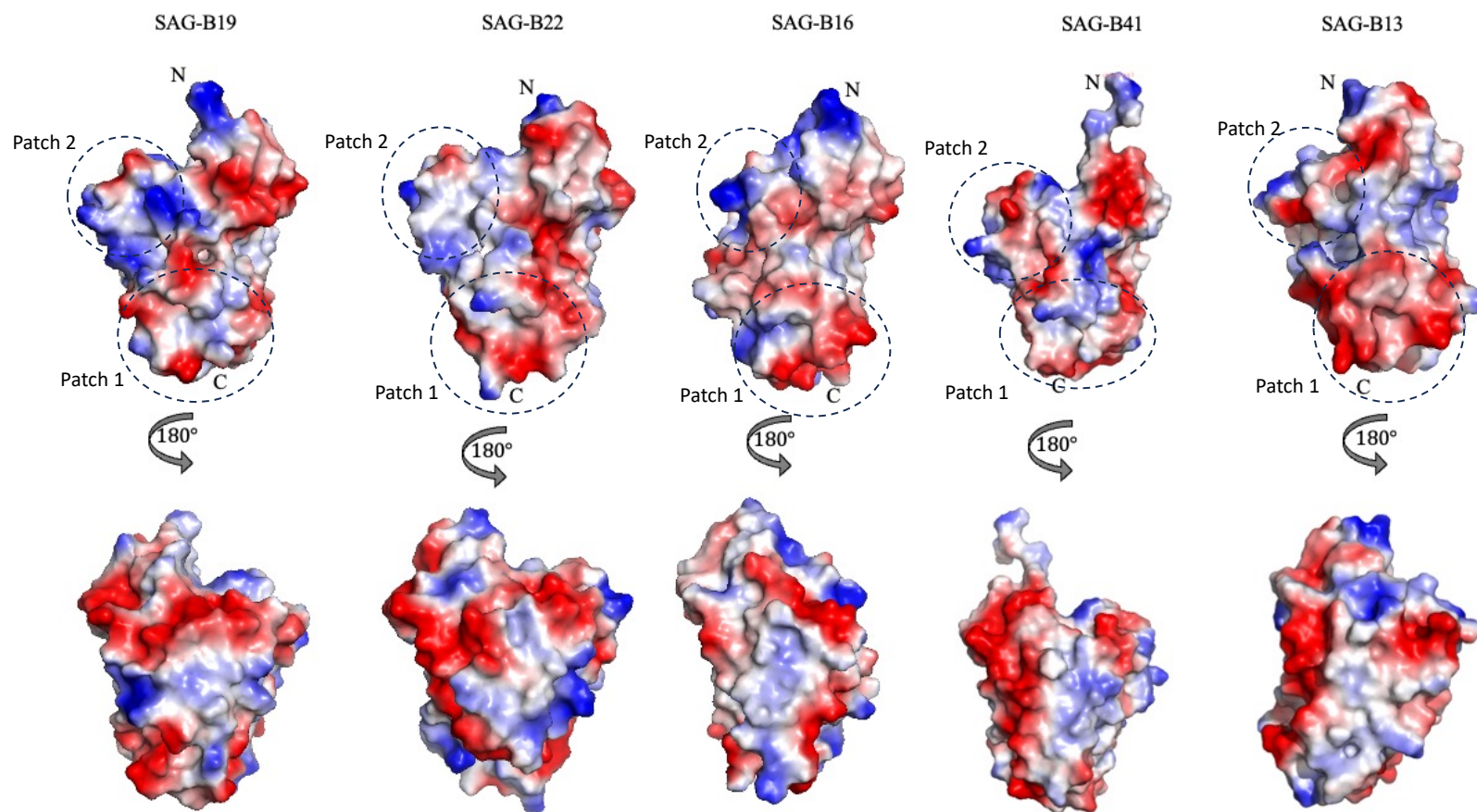
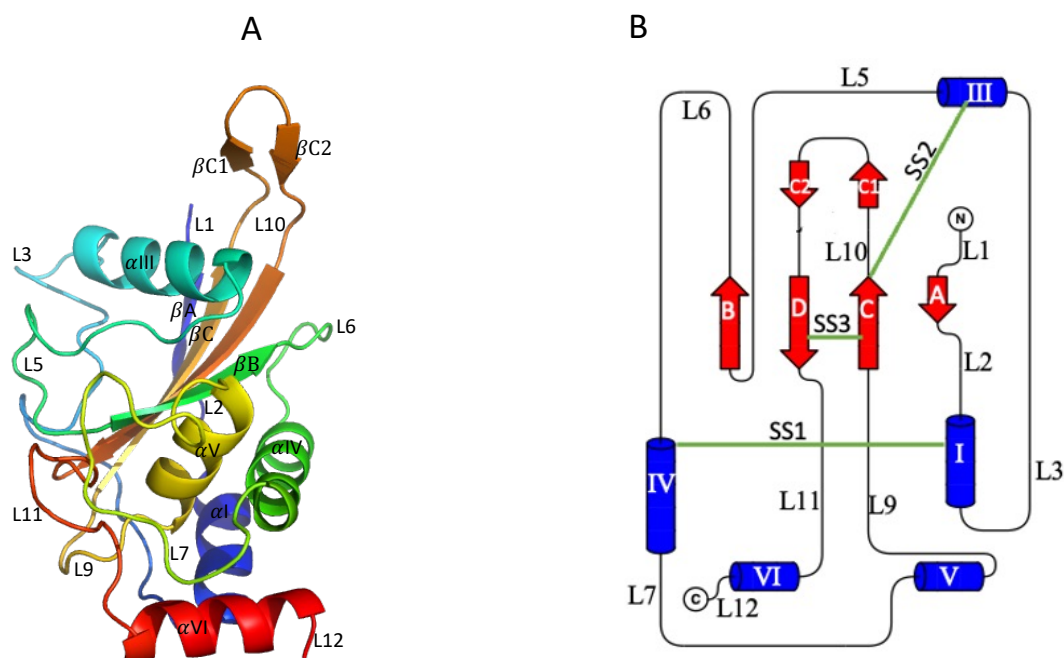


Figure 5.8: **Electrostatic surfaces diagrams of five *E. tenella* SAG-B members.** The electrostatic surfaces of SAG-B19, SAG-B22, SAG-B16, SAG-B41 and SAG-B13 as calculated by PYMOL. The ranges of  $kbT/e-$  vary between dark red (negative) to dark blue (positive) that are set by the programme defaults. Two regions of the surface (labelled as Patch 1 and Patch 2) are identified and described in the text (see section 5.2.2).

### 5.3 The Structure of SAG-A91 an SAG-A family member

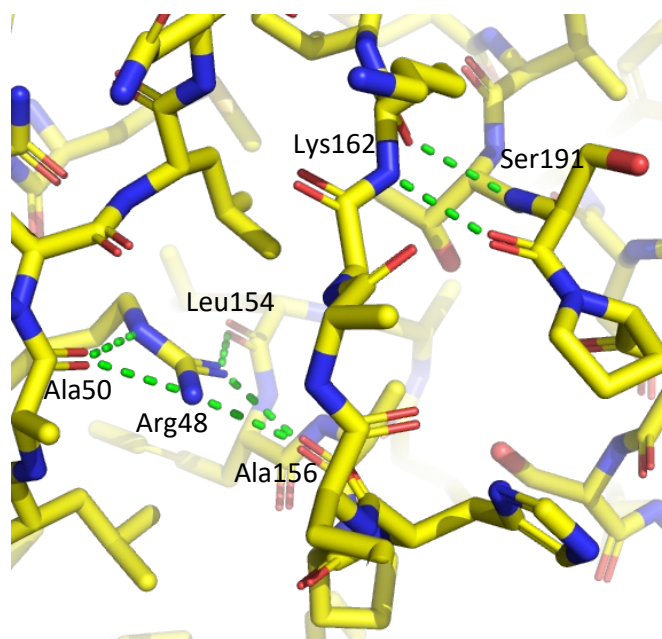
In addition to the group of SAG-B structures described above, one SAG-A member (SAG-A91) was also successfully crystallised, and its structure determined at a resolution of 1.4 Å, as described in chapter 4. The structure of this SAG comprised 184 residues, revealed a similar fold to those of the SAG-B members with a central 4-stranded antiparallel  $\beta$ -sheet surrounded by five  $\alpha$ -helices and 12 loops, but with two additional surface exposed strands ( $\beta$ C1 and  $\beta$ C2) that extend loop L10 [Figure 5.9]. For ease of comparison the numbering of the elements of regular structure in SAG-A91 follows the convention set by that of SAG-B19.



**Figure 5.9: 3D structure of SAG-A91.** A) The structure consists of four  $\beta$ -strands (labelled A-D) forming an anti-parallel  $\beta$ -sheet flanked by five  $\alpha$ -helices. Twelve connecting loops (L1-L12) join the regions of regular secondary structure. The additional residues in L10 form two short antiparallel  $\beta$ -strands ( $\beta$ C1 and  $\beta$ C2). B) The 2D topology diagram displays the relative position and connection of secondary structural elements. The positions of the three disulphide bonds are represented by green bars and are labelled, (SAG-B19 numbering convention).

As seen in the structures of the SAG-B members, the loop regions in SAG-A91 are extensive, comprising ~ 50% of the entire protein structure. Some loops are quite large such as that of L3 which links  $\alpha$ I to  $\alpha$ III, and that consists of 24 residues many of which are hydrophobic. Similarly, L7 is also large comprising 17 residues that connect  $\alpha$ IV to  $\alpha$ V. As seen in the SAG-B members, SAG-A91 contains six cysteines which form three disulphide bonds. Two of these disulphides, SS1 (C40 – C115 that links L2 to the N-terminal end of  $\alpha$ IV), and SS3 (C160 – C188 that joins  $\beta$ C to  $\beta$ D) are present in the same place as in the SAG-B family structures. However, the third disulphide, SS2 (C83 – C170) occurs in a position similar to that of SS2 in the SAG-B family but instead connects  $\alpha$ III to the C-terminal end of  $\beta$ C, rather than connecting loop L5 to loop L10.

The structure of SAG-A91 contains a single *cis*-peptide between Ser191 and Pro 192 at the C-terminal end of  $\beta$ D. Interestingly, this *cis*-Pro occurs in the same place as that of the conserved non-proline *cis*-peptide in the SAG-B family which is, as described earlier (see section 5.2.1), involved in positioning the side chain of the conserved arginine belonging to the NxxR motif [Figure 5.10].

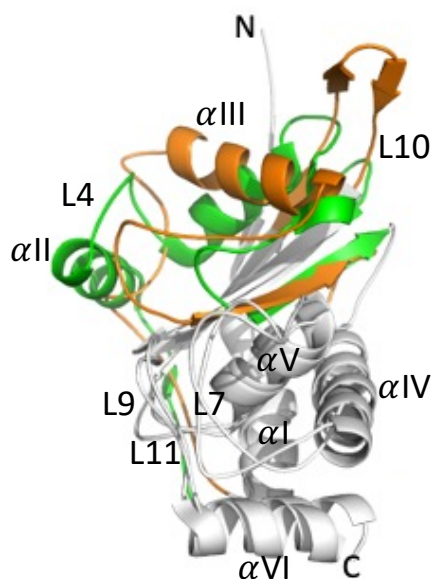


**Figure 5.10: Key interactions creating the pocket for the conserved buried arginine in SAG-A91.** The hydrogen bonds between the main chain carbonyl and NHs of Ser191 that forms part of the Ser191-Pro192 *cis*-Pro peptide and the main chain NHs and carbonyl of Lys162 that lie in the loop next to the guanidiny group of the buried arginine Arg48. The green dots indicated the hydrogen bonding.

### 5.3.1 The Comparison between SAG-A91 and SAG-B family structures

The structure of SAG-A91 was superimposed with the structure of SAG-B19 ( $C\alpha$  rmsD of 1.94 Å), showing that they share a similar folding pattern with approximately 50 % of both structures being accounted for by loops. Although the structures are quite similar overall there are a number of significant differences. In the SAGA-91 structure helices I, IV, V and VI together with loops L9 and L11, that pack on one side of the central 4-stranded antiparallel  $\beta$ -sheet, all occur in similar positions to SAG-B19 [Figure 5.11]. However, on the other side of the sheet the fold of SAG-A91 deviates from that of SAG-B19. Helix  $\alpha$ II is not present in the SAG-A91 structure, with loops L3 and L4 merging into one long loop, L3 and  $\alpha$ III is rotated by approximately 70 degrees [Figure 5.11]. Whilst  $\alpha$ II is also not present in SAG-B13, in this protein  $\alpha$ III is in the same as that in the other members of the SAG-B family. A further difference between SAG-A91 and the SAG-B-family is the extension of loop L10, part of which folds into two short antiparallel  $\beta$ -strands. The fold of SAG-A91 in the C-terminal region that links with the GPI-anchor is similar to that adopted in the SAG-B family, indicating a similar presentation of the structures from the surface of the membrane [Figure 5.11].

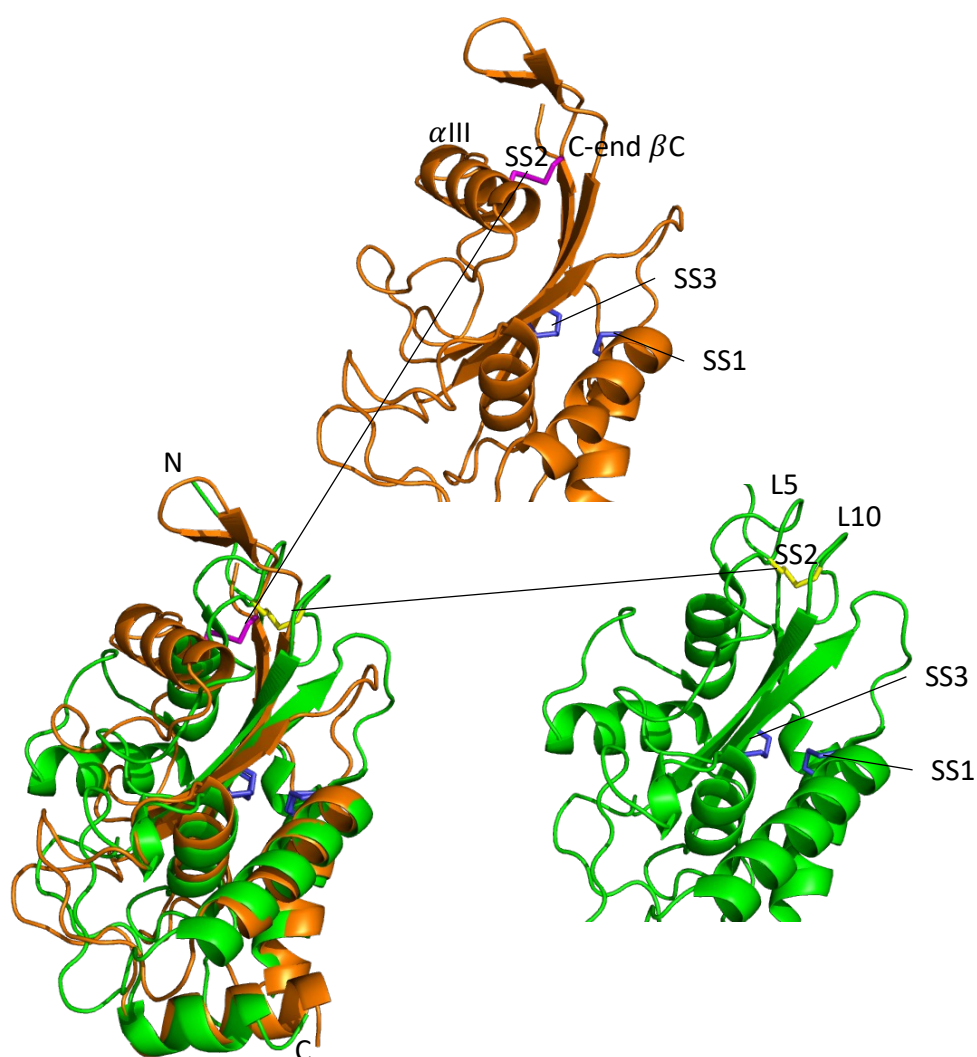
Comparison of the pattern of disulphide bonds between the SAG-A91 and the members of the SAG-B family reveals that SS2 occurs in a similar region of the molecule but, in detail the disulphide is different [Figure 5.12]. This arises because of the rotation of  $\alpha$ III which provides one of the residues forming the interaction (Cys 83 in SAG-A91) to the cysteine at the C-terminal end of  $\beta$ C (Cys 170), whereas in SAG-B19, the interaction is between Cys 126 in loop L5 and Cys 213 which lies at the extension of  $\beta$ C into loop L10 in SAG-B19. The similar sequence pattern between SAG-A91 and the other SAG-A proteins in the region of  $\alpha$ III suggests that this helix is rotated in all the SAG-A family members and hence the structure of SAG-A91 is a good model for all the SAG-A family.



*Figure 5.11: **Superposition of the SAG-A91 and SAG-B19 Structures:** A cartoon diagram to show the overlap between SAG-A91 and SAG-B19 and illustrate their similarities and differences. A grey colour for the is used to indicate regions of similarity in both structures, while green and orange is used to highlight regions of difference between SAG-B19 and SAG-A91, respectively. The loss of  $\alpha$ II, the rotation of  $\alpha$ III and the extension to the L10 loop in SAG-A91 are clearly visible.*

### 5.3.2 Sequence Diversity and Similarity within the SAG-A family.

As the structure of only a single SAG-A protein was determined, the sequences and of other *E. tenella* SAG-A proteins were analysed to see if the differences between the SAG-B family and SAG-A91 were also present. Analysing 51 sequences (9 members were excluded from the analysis as the sequences were either incomplete or possibly represented pseudogenes) of the 60 SAG-A members showed a wide range of sequence identities from 18% to 95%. As expected, the analysis revealed that of the 14 conserved residues that can be identified [Figure 5.13], most of them are largely buried in the core of the protein (defined as having an accessible surface area of less than 20%). The only exceptions are Pro133 and Pro134 which have solvent accessible surfaces of 30% and 50% respectively. The analysis further suggests that the sequences can be subdivided into two subfamilies which differ in the length of loop L10. In both subfamilies loop L10 is longer than in the B-family forming two extra short antiparallel strands as discussed above [Figure 5.11]. The two subfamilies can be distinguished on the basis of the presence of an RRL motif (for example residues 187-189 in SAG-A31 (Appendix A.3) and the fact that the L10 loop is larger. The role of this motif is as yet not understood.



*Figure 5.12: Superposition of SAG-B19 and SAG-A91 conserved disulphides bonds. The diagram shows the differences in positions of the SS2 disulphide bridge indicated as magenta stick in SAG-A91 (orange) and as yellow stick in SAG-B19 (green). In both structures SS3 and SS1 indicated as blue sticks.*



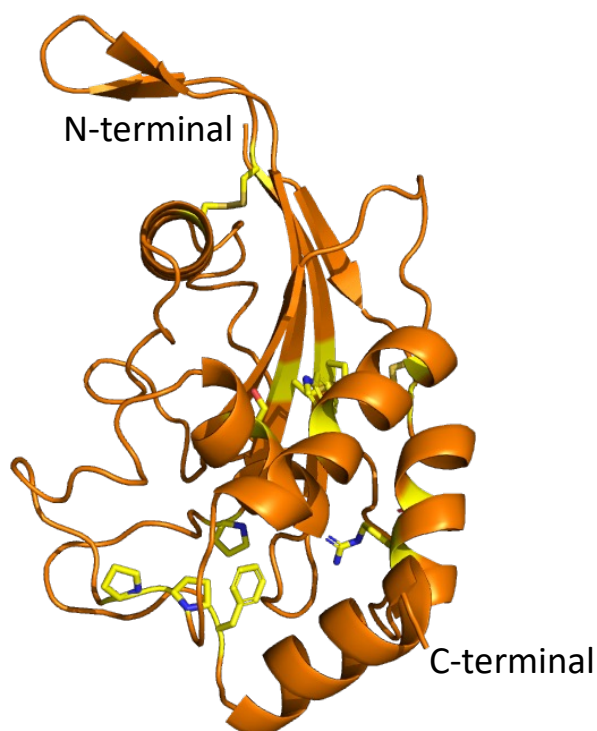
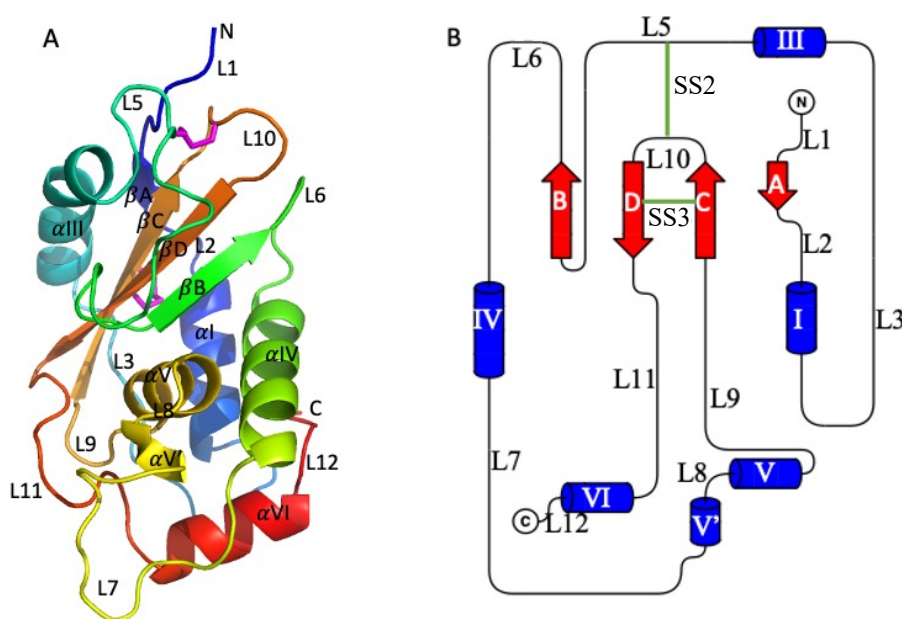


Figure 5.13: **Conserved residues across the 51 members of SAG-A family.** The cartoon representative of SAG-A91 structure, the yellow sticks indicating the conserved residues across 51 SAG-A family of *E. tenella*, including the conserved three disulphide bonds.

#### 5.4 The structure of SAG-C75

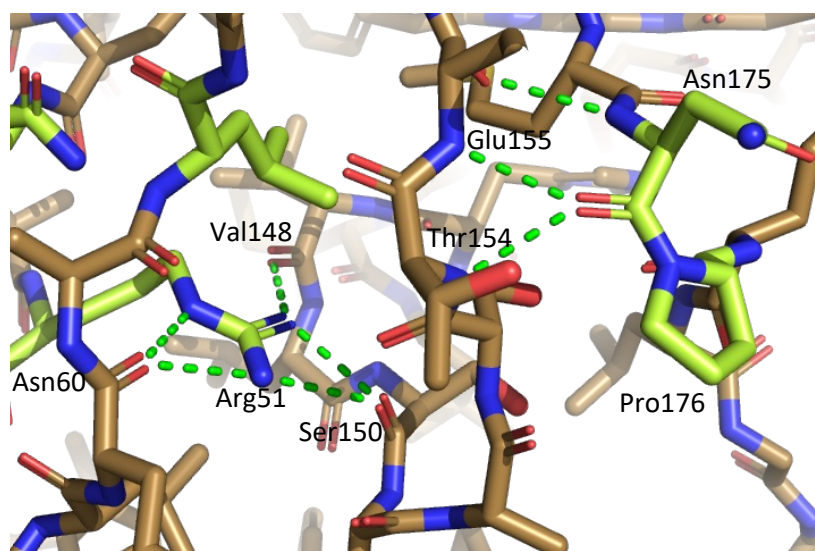
Analysis of the genome of the single member of *Eimeria tenella* suggests that it contains a further SAG protein that has been classified as belonging to family C (SAG-C75). This protein was over expressed, purified, crystallised and its structure determined at a resolution of 1.32 Å, as described in chapter 4. The structure of this SAG protein consists of 170 residues arranged in four anti-parallel  $\beta$ -strands, five  $\alpha$ -helices, and 11 loops with a similar fold that been seen in the SAG-B and SAG-A families [Figure 5.14]. However, the loop regions are less extensive than those seen in SAG-A and SAG-B members comprising only 40% of the polypeptide chain. Interestingly, the structure of SAG-C75 contains only two disulphide bonds corresponding to SS2 and SS3 in the SAG-B family. For ease of comparison the numbering of the elements of regular structure in SAG-C75 follows the convention set by that of SAG-B19 and hence these two disulphides are referred to as SS2 and SS3 (C84 - C164 and C158 - C172 in SAG-C75, respectively). The structure of SAG-C75 contains two *cis*-peptides involving proline (the first between Ala135-Pro136, and the second between Asn175-Pro176). The first *cis*-Pro (Ala135-

Pro136) occurs in the same place as that of a conserved *cis*-proline in the SAG-B family. In contrast, the second *cis*-Pro (Asn175-Pro176) of SAG-C75 occurs in the same place as that of the conserved non-proline *cis*-peptide in the SAG-B family that has previously been shown to form key interactions involving the pocket for the buried arginine side chain of the conserved arginine residue in the NxxR motif [Figure 5.15].



**Figure 5.14: Schematic diagrams of the structure of SAG-C75 to show the overall fold and the positions of elements of regular secondary structure.** (A) Chainbow representation of the 3D structure of SAG-C75 indicating the positions of the major  $\beta$ -strands,  $\alpha$ -helices and connecting loop regions all of which are labelled, and the magenta sticks indicating the two disulphide bonds. (B) 2D topology diagram with the N- and C-termini and the elements of secondary structure labelled. The locations of the disulphide bonds are shown by green bars, (SAG-B19 numbering convention).





**Figure 5.15: Key interactions creating the pocket for the conserved buried arginine in SAG-C75.** The hydrogen bonds between the main chain carbonyl and the NHs of Asn175 that forms part of the Asn175-Pro176 *cis*-Pro peptide and the main chain NHs and carbonyl of Glu155 and the main chain NHs of Thr154 that lie in the loop next to the guanidiny group of the buried arginine Arg51. The green dots indicated the hydrogen bonding.

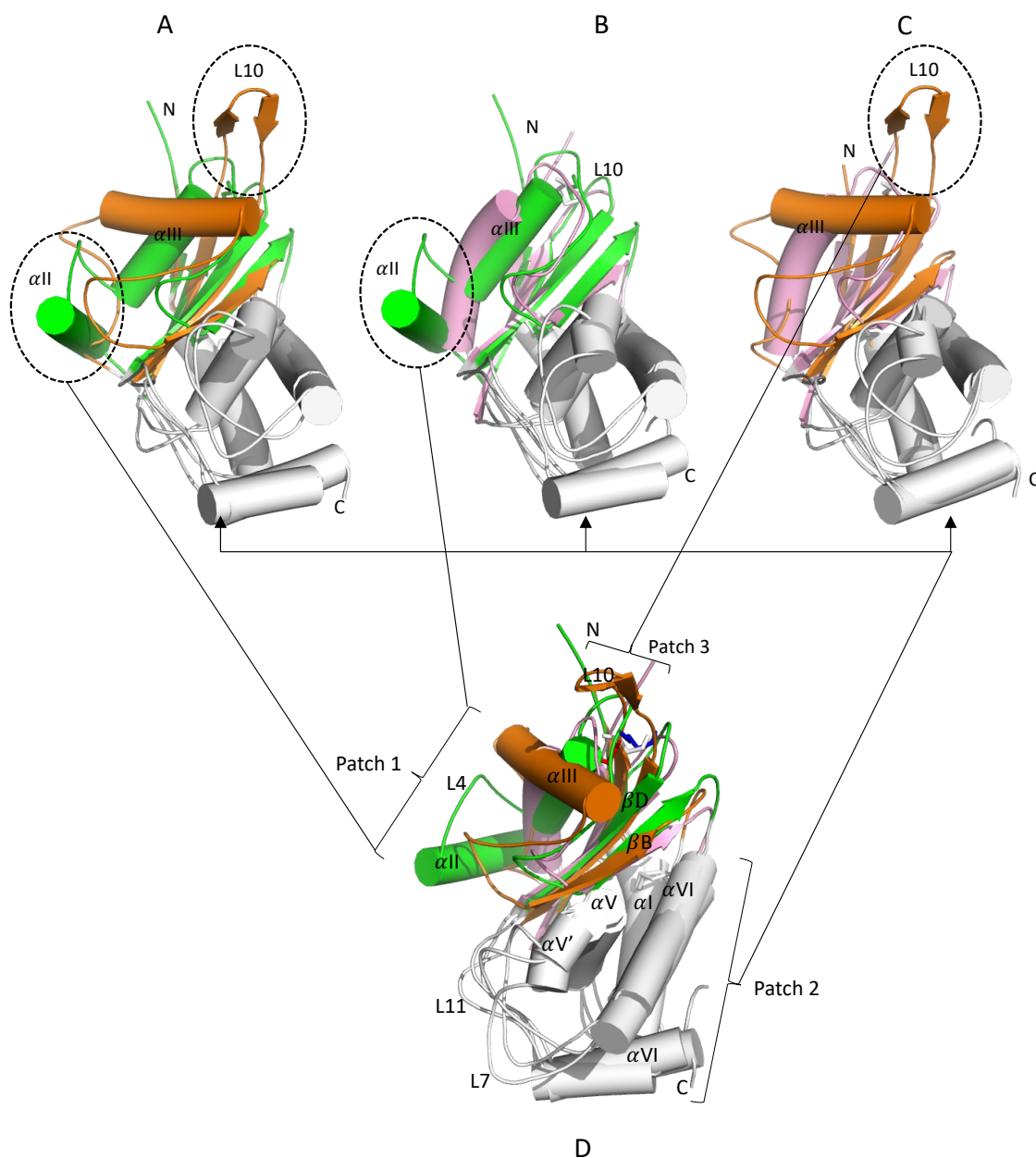
## 5.5 Comparison of the structures of the SAG-A, SAG-B and SAG-C families

Within the three families of *E. tenella* SAGs, the fold of these proteins is very closely related. To facilitate a detailed comparison the structures of SAG-B19, SAG-A91 and SAG-C75 were superimposed using GESAMT (C $\alpha$  rmsD of 1.94Å between the SAG-B19 and SAG-A91; 1.95Å between SAG-B19 and SAG-C75 and 2.03Å between SAG-A91 and SAG-C75 [Table 5.7]). These rmsD values are higher than those observed when comparing the members of the SAG-B family (typically rmsD values of 1 Å [Table 5.6] consistent with the lower level of sequence identity between SAG-B19 and the family -A and -C proteins [Table 5.7]). In the three structures, helices I, IV, V and VI together with loops L7 and L11, all of which pack on one face of the central 4-stranded antiparallel  $\beta$ -sheet, lie in similar positions [Figure 5.16 A (patch 2)]. However, on the other face of the sheet, the three SAGs structure show more significant differences with helix  $\alpha$ II in SAG-B19 being absent in both SAG-A91 and SAG-C75 L3 [Figure 5.16 (patch 1)] and the loops L3 and L4 merging into one long loop. In addition, in SAG-A91, helix  $\alpha$ III is rotated by approximately 70° compared to its position in both SAG-B19 and SAG-C75 which are more closely related [Figure 5.16 B]. Another clear difference can be seen between SAG-C75 and SAG-B19, both of which have a very similar loop L10 as opposed to SAG-A91, where this loop extends to form another two short antiparallel  $\beta$ -strands

as mentioned above [Figure 5.16 (patch 3)]. The three structures share a similar fold in the C-terminal region that lies near the site for the attachment of the GPI-anchor [Figure 5.16].

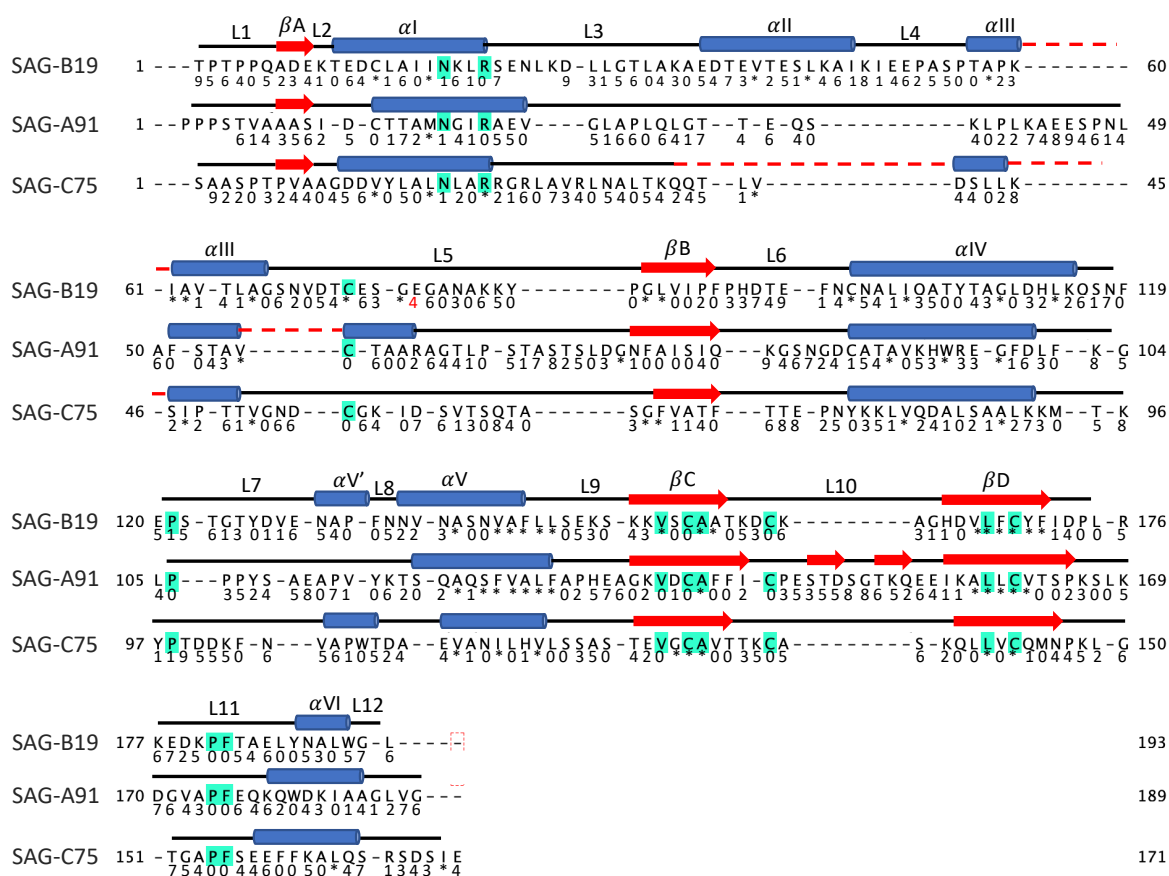
*Table 5.7: Sequence identity between the E. tenella three SAG family.*

	SAG-B19 (%) / C $\alpha$ rmsD	SAG-A91 (%) / C $\alpha$ rmsD	SAG-C75 (%) / C $\alpha$ rmsD
SAG-B19 (%) / C $\alpha$ rmsD	100	17/ 1.94 Å	26/ 1.95 Å
SAG-A91 (%) / C $\alpha$ rmsD		100	15/ 2.03 Å
SAG-C75 (%) / C $\alpha$ rmsD			100

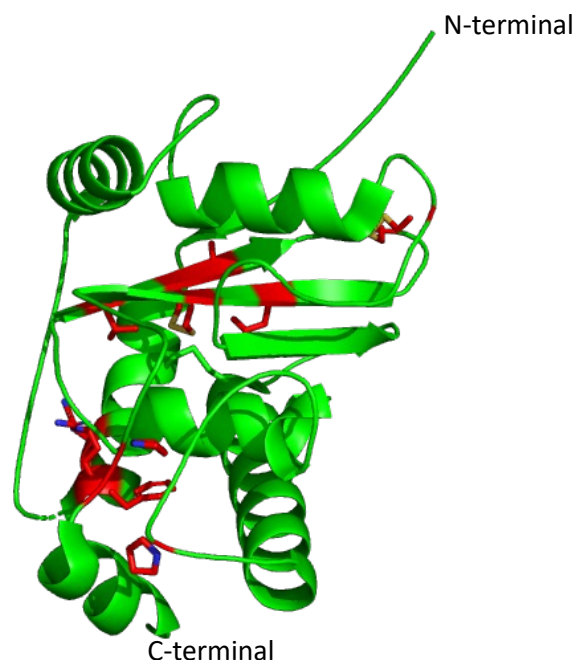


**Figure 5.16: Superposition of the SAG-A91, SAG-B19 and SAG-C75 structures:** A) overlap between SAG-B19 and SAG-A91; B) overlap between SAG-B19 and SAG-C75; C) overlap between SAG-C75 and SAG-A91. D) A cartoon diagram to show the overlap between three SAG proteins and illustrate their similarities and differences. A grey colour is used to indicate regions of similarity in the three structures (patch 2), while green (SAG-B19), orange (SAG-A91) and pink (SAG-C75) is used to highlight regions of difference between them (patch 1 and 3).

The structure-based alignment of the SAG-B19, SAG-A91 and SAG-C75 sequences [Figure 5.17] revealed that 12 residues are absolutely conserved across the three families (N72, R75, C126, P174, V205, C207, A208, C213, L220, C222, P234 and F235 (SAG-B19 numbers). [Figure 5.18]. All of these conserved residues are largely buried in the core of the protein (defined as having an accessible surface area of less than 20%) consistent with them playing a role in maintaining the overall structure of the SAG proteins, rather than forming any interactions to other host molecules.



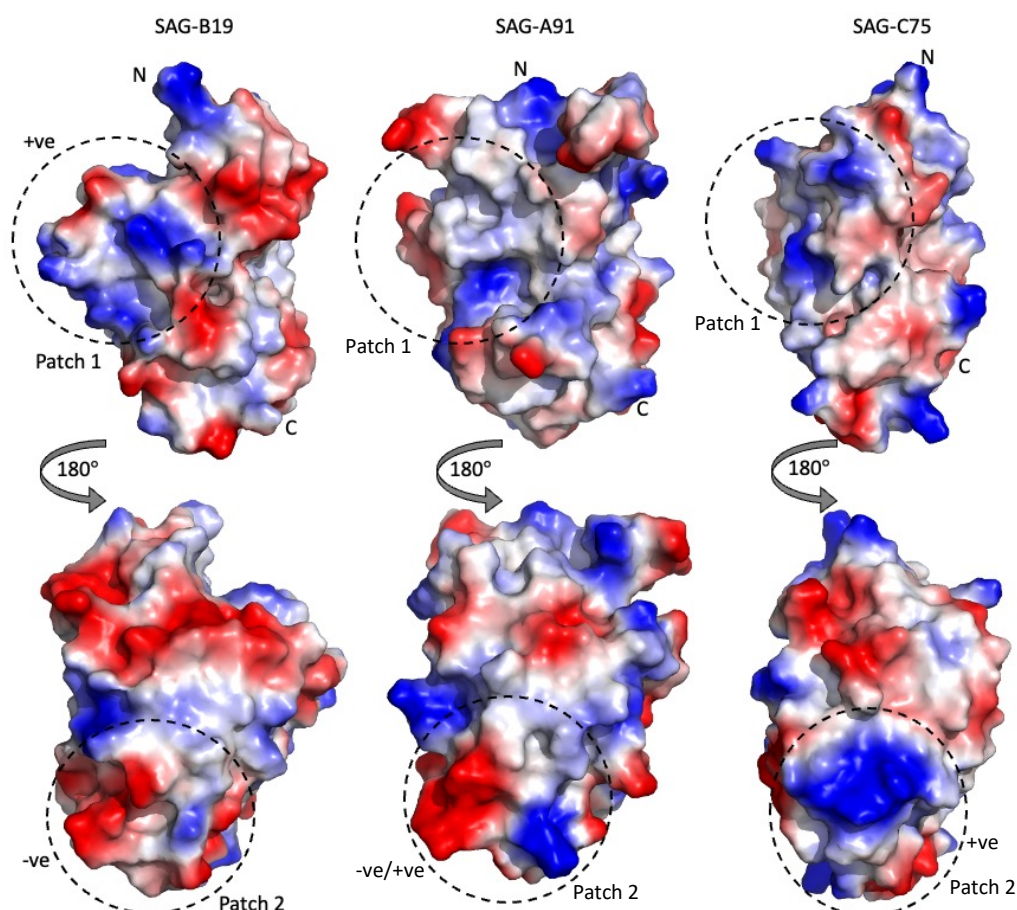
**Figure 5.17: Sequence alignment of SAG-B19, SAG-A91 and SAG-C75.** The sequences alignment crosses the core domain of each SAG protein. 9 conserved residues between them and shows as sandy boxes. The positions of the loops and secondary structure elements in these proteins are shown above the sequences and labelled. The solvent accessible area of residues in these proteins were calculated by the method of Lee and Richards. The resulting accessible areas were then expressed as a fraction of the total solvent accessible surface area for each amino acid in a GXG tripeptide and are given as a sequence of integers 0-9 under each SAG sequence (where each number represent residues which have 1-10%, 11-20%, 21-30%, etc of the surface solvent accessible. Totally buried residues indicated by \*. The red dots indicated that helix continues.



*Figure 5.18: The position of the conserved residues across the three SAG families. The 12 conserved residues are highlighted as red sticks in SAG-B19 structure including the conserved disulphide bonds.*

To examine whether the lack of sequence identity between these proteins also extends to different patterns of surface charge, the surface electrostatic properties of each of these SAG protein structures were calculated using PyMOL [Figure 5.19]. Inspection of these surfaces shows that there is no common pattern of surface charge, and their chemical characteristics are quite distinct. For instance, a positively charged surface patch around the in SAG-B19 (patch 1) is quite different from that in SAG-C75, and in SAG-A91, where it exhibits a polar pattern but with less positive charge [Figure 5.19]. Further differences can be seen for a negatively charged surface patch in SAG-B19 (patch 2) which shows a mixed grouping of positive and negative residues in SAG-A91, and a the positively charged surface in SAG-C75 [Figure 5.19]. Consequently, there are no conserved surface charged patches that could be implicated in host interactions. Although various functions have been proposed for the SAG proteins in *Eimeria tenella* parasites, including potential roles in cell adhesion or parasite invasion, their precise functions remain unknown. Overall, the buried nature of all the conserved residues and the absence of any patterns of sequence conservation or surface charge on the protein surface strongly suggest that these three SAG families do not share a common binding partner in the host by binding to either host receptors or their exposed sugar residues. However, they may

play a similar functional role in stimulating the immune system albeit with a different group a residues.



**Figure 5.19: Electrostatic surfaces diagrams within the *E. tenella* family.** The electrostatic surfaces of SAG-B19, SAG-A91 and SAG-C75 as calculated by PYMOL. The ranges of  $kbT/e^-$  vary between dark red (negative) to dark blue (positive) that are set by the programme defaults. Two regions of the surface (labelled as Patch 1 and Patch 2) are identified and described in the text (see section 5.5).



## 5.6 *E. tenella* Binding Partner for SAGs protein

*In vitro* studies have indicated that SAG proteins, particularly from family A, play a pivotal role in attaching to epithelial cells found in the chicken intestine and in inducing protective immunity against similar challenges (Song *et al.*, 2015). This attachment is a crucial step in the parasite's subsequent invasion of the host cell surface. The epithelial mucins secreted by intestinal cells have distinct O- and N-linked glycosylation that affects the protective mucus layer in chickens. O-linked glycans are more abundant than N-linked glycans on chicken intestinal mucins (Stanley *et al.*, 2017). The O-glycans have variable sequences and branching composed of fucose, mannose, galactose, N-acetyl-D-galactosamine, N-acetylglucosamine, and sialic acid. In contrast, N-glycans have a conserved chitobiose core of two N-acetyl-D-glucosamines with high mannose and hybrid type branching (Brockhausen and Stanley, 2015). The complex O-glycans comprise over 80% of mucin molecular mass and their composition and density influences the structure and function of the mucosal barrier. Clearly, in principle, the different sugars on the cell surface could act as targets for interaction with the different SAGs. However, this would imply the existence of some pattern of sequence conservation on the surface of the SAG proteins, a suggestion not consistent with the results of the structure comparison. Therefore, to further exclude the possibility that the different SAG proteins might bind to similar sugars on the host cell surface the purified proteins were tested for their ability to interact with monosaccharides (Fuc, Man, Gal, GalNAc, GlcNAc and sialic acid) found in chicken mucins. All the SAGs were concentrated to 25  $\mu$ M and tested for changes in the melting temperature of the protein following equilibration with different sugars at 2.5 mM see section 2.6 (chapter 2). However, whilst the melting temperatures ( $T_m$ ) of the different SAGs varied considerably (30-71 C°) these experiments did not provide any evidence for the interaction of any of the sugars with the SAG proteins as their individual melting temperatures were unchanged [Table 5.8].

Interestingly, the wide range of melting temperatures across the *Eimeria* SAGs does not correlate with the number of disulphides present in the structure, despite all these SAG proteins sharing effectively the same core fold, this reflecting the possible reasons for biological function. For example, the  $T_m$  varies between 36 and 71 C° in the SAG-A family and between 30 and 62 C° in the SAG-B family, despite all these proteins containing three disulphide bridges. In addition,

although the SAG-C family has only two disulphide bridges the  $T_m$  of SAG-C75 is 60 C°, considerably higher than some of the proteins in the SAG-A and SAG-B families. The SAG-C family proteins SAG-EBC1 and SAG-EBC3 have an even higher  $T_m$  (65 C° and 70 C°, respectively) and clearly the temperature stability of the *Eimeria* family of SAG proteins depends on more than just the number of disulphide bridges.

Table 5.8: The melting point of *E. tenella* SAGs protein including the testing temperature (C°) to the monosaccharides.

SAG	Family	Melting Temperature ( $T_m$ ) $\pm 2$ C°	Fuc	Man	Gal	GalNAc	GlcNAc	sialic acid
SAG 10	A	40	39.0	41.0	40.1	39.5	40.1	40.4
SAG 31	A	36	36.4	36.5	36.7	35.9	36.3	36.5
SAG 7	A	59	58.8	58.4	59.7	59.6	59.8	59.4
SAG 6	A	71	71.1	70.9	70.7	71.4	71.0	71.2
SAG-A91	A	48	48.4	48.6	48.3	49.1	48.6	49.0
SAG-A870	A	52	52.7	52.9	52.4	52.8	52.4	52.5
SAG-A905	A	61	61.8	61.7	60.9	61.8	61.7	61.3
SAG-A1	A	65	65.8	65.7	65.1	65.7	65.0	66.2
SAG-B13	B	62	62.9	62.1	62.4	62.8	62.1	62.1
SAG-B16	B	61	61.8	62.0	62.7	61.7	60.9	61.4
SAG-B22	B	51	51.0	51.2	51.9	51.3	51.7	50.9
SAG -B41	B	30	30.0	30.4	29.8	29.7	30.5	30.6
SAG-C75	C	60	60.1	60.8	60.3	60.7	60.5	60.4
EBC1	C	65	65.9	65.8	65.3	65.7	64.8	64.9
EBC2	C	47	47.5	47.6	47.8	47.9	47.6	47.5
EBC3	C	70	70.8	70.5	70.8	70.6	70.9	70.2



## Chapter6 Overexpression, Purification, Crystallisation and Structure Determination of CAP-like SAG in *Plasmodium Vivax* (*P. vivax*)

### 6.1 Overview

On the basis of the analysis of the genome sequences of various apicomplexans including *Eimeria*, *Toxoplasma* and *Plasmodium* (Kissinger & DeBarry, 2011) it had been established that each of these species expressed a large superfamily of SAGs on their parasite surfaces. Moreover, detailed comparison of the sequences of these SAGs combined with structural studies of representative SAGs from each of these species (Abrahamsen *et al.*, 2004; Owusu and Bennett, 2015; Swapna and Parkinson, 2017) showed that whilst the structures of the SAGs from *Toxoplasma* and *Plasmodium* were similar, belonging to the SRS-related protein superfamily, the SAGs from *Eimeria* had a completely different fold based on the CAP-superfamily fold. This had led to the suggestion that each parasite expressed one type of SAG belonging to one or other of these families but not both (Ramly *et al.*, 2021). However, more detailed analysis of the genome sequences coupled with studies using alphafold has suggested that a CAP-like SAG might be present and expressed in *P. vivax* alongside the expression of SAGs belonging to the SRS superfamily. Further comparisons have led to the identification of possible homologues of the *P. vivax* CAP-like SAG in a number of apicomplexan parasites including different strains of *Plasmodium*, *Babesia*, *Toxoplasma* and *Besnoitia* species. Having identified these possible CAP-like SAGs it was therefore decided to attempt to determine the structure of one or more of these proteins to confirm that they did indeed have a CAP-like fold and to investigate their structural and functional relationship to the *Eimeria* SAGs.

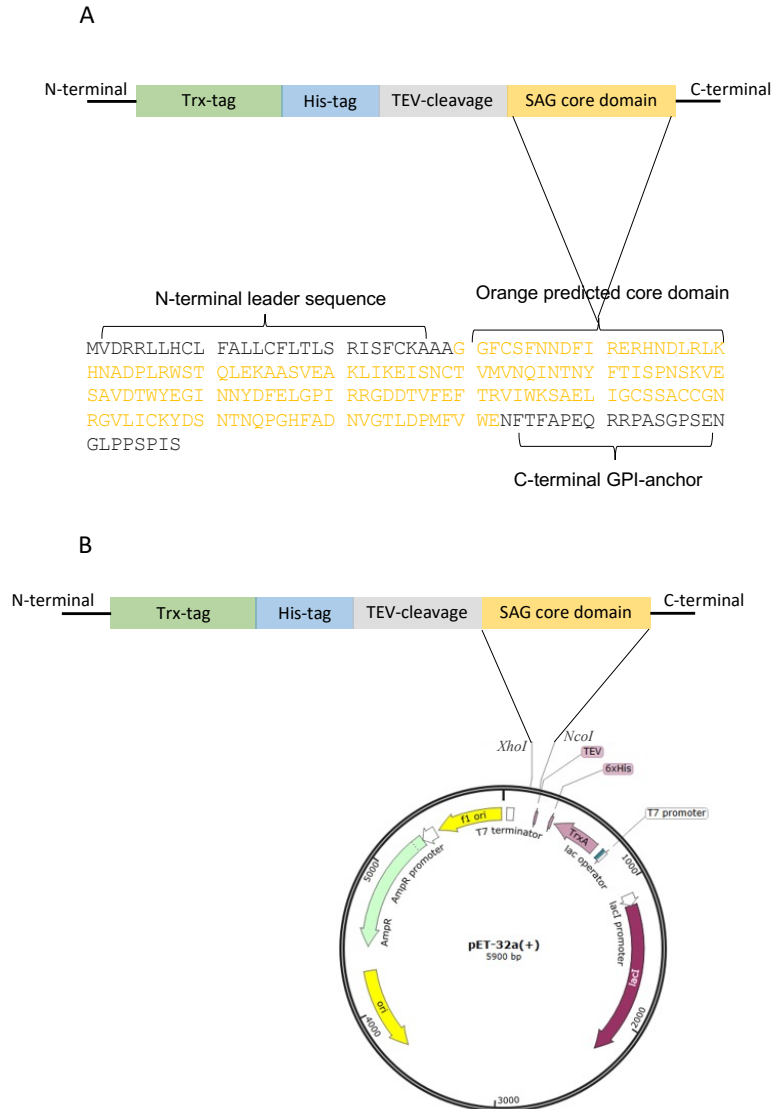
This chapter describes the overexpression, purification, and crystallization trials of these CAP-like SAGs. However, expression tests on the *Toxoplasma*, *Babesia*, and *Besnoitia* proteins showed low solubility at the 17 C° that the optimized temperature for previous SAGs including the SAGs from *Plasmodium* species. Consequently, further work is required to optimize the solubility and expression levels of these proteins. This may involve varying the IPTG concentration, changing the growth media, or experimenting with different temperatures. Despite the low solubility challenges, *P. vivax* CAP-like SAG displayed favourable characteristics during purification steps.

Therefore, it was focused how to get enough protein at the target concentration that *Eimeria tenella* SAG were crystallized. In addition, the chapter also include the data collection from the *P. vivax* protein crystals that facilitated successful structure determination. Additionally, it covers the analysis of the structure of this CAP-like SAG in *P. vivax* and its comparison with similar proteins from other apicomplexan parasites and members of the CAP superfamily, which is essential for understanding the unique features and potential functional roles of these proteins across different species.

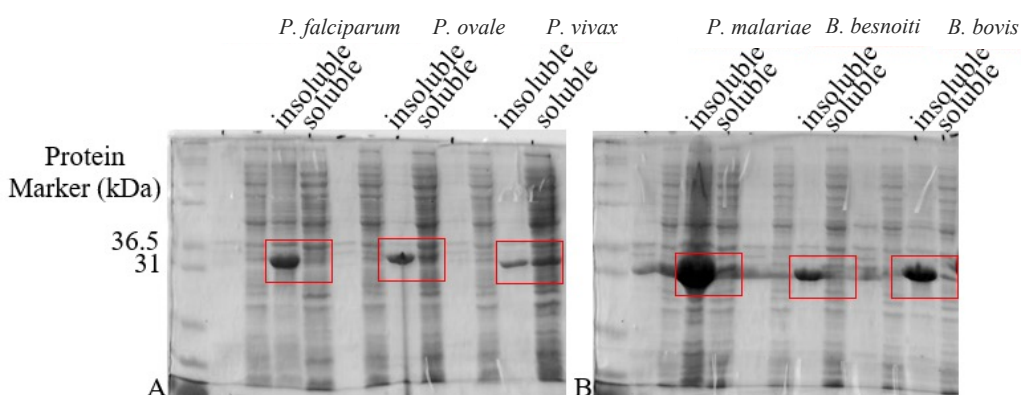
## 6.2 Overexpression

Constructs for the predicted core domain, excluding the N-terminal leader sequence and the C-terminal GPI anchor sequence, of each protein were designed in the same way as that for the *Eimeria* proteins, as described in Chapter 3 (section 3.2.4). The sequence for each construct was inserted into a *pET32a* vector for expression [Figure 6.1].

Plasmids containing *Plasmodium falciparum* (*P. falciparum*), *Plasmodium voale* (*P. ovale*), *Plasmodium vivax* (*P. vivax*), *Plasmodium malariae* (*P. malariae*), *Babesia bovis* (*B. bovis*), and *Besnoitia besnoiti* (*B. besnoiti*) SAGs were transformed into *origami* cells (DE3) as outlined in the materials and methods chapter 2 (section 2.2.7). Subsequently, all SAGs were overexpressed in a small scale (50 ml flasks) culture at 17 C°, following the same procedures used for the *Eimeria* SAG proteins as detailed in chapter 2 (section 2.3.1). All of the constructs for these predicted SAG proteins were successfully expressed, but analysis of SDS-PAGE gels showed that the proteins from *P. falciparum*, *B. bovis*, and *B. besnoiti* were mostly expressed in inclusion bodies rendering them insoluble [Figure 6.2]. However, soluble protein could be obtained from the expression trials using plasmids containing the *P. vivax*, protein (~30% soluble protein) and *P. ovale* and *P. malariae* (~10% soluble) [Figure 6.2]. As the latter two proteins would require very large scale growths to get enough protein to be purified and crystallised, it was decided to concentrate on the *P. vivax*, protein.



**Figure 6.1: The CAP-like SAG in Trx-construct.** A) *P. vivax* CAP-like SAG construct including the *trx*-tag, His-tag, TEV cleavage and *P. vivax* CAP-like SAG predicated core domain. B) The designed construct insertion into the pET32a vector by using the two restriction enzymes *NcoI* and *XhoI*.



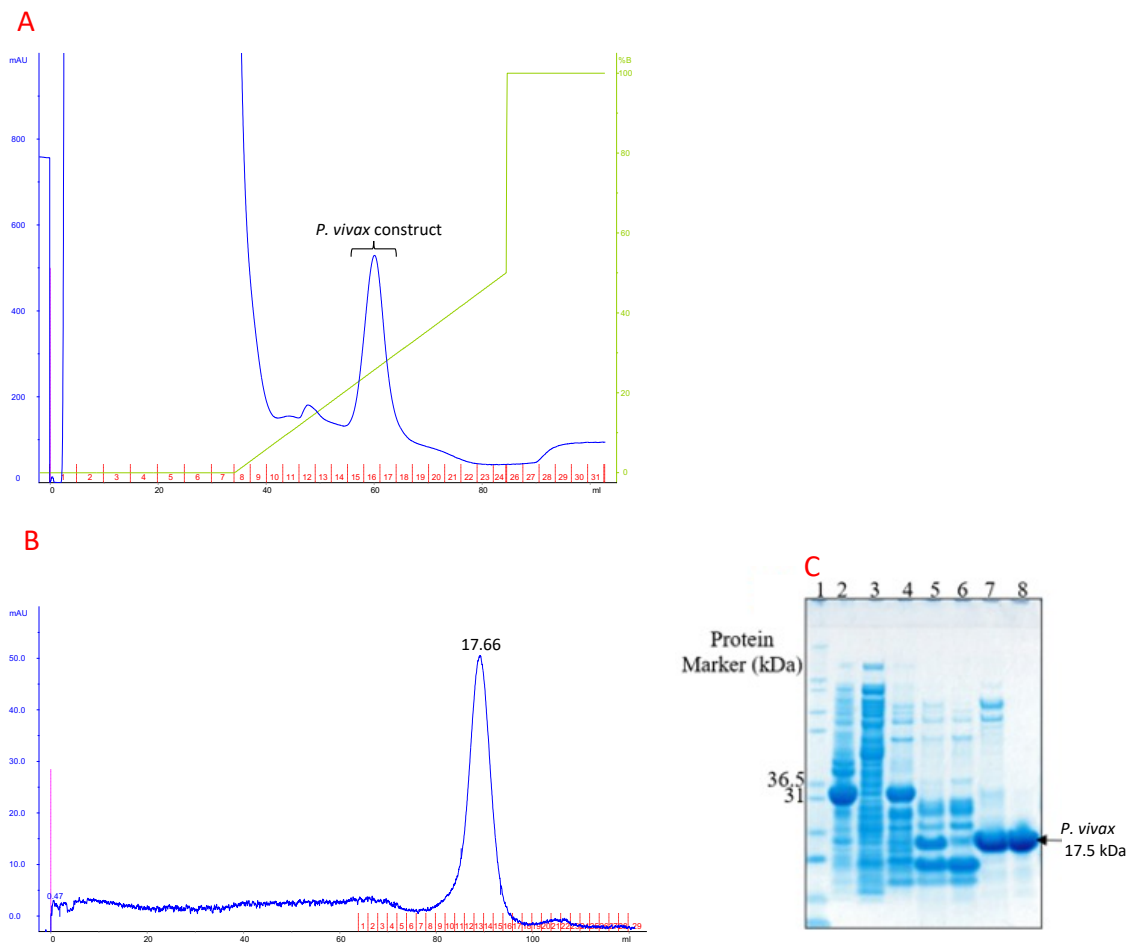
**Figure 6.2: Expression of CAP like SAG proteins.** A) SDS-PAGE gel shows the overexpressed of *P. falciparum*, *P. ovale* and *P. vivax*. B) SDS-PAGE gel shows the overexpressed of *P. malariae*, *B. besnoiti* and *B. bovis*. All the proteins were tested at the optimized temperature (17°C) with 1mM IPTG induction. The red boxes indicate each SAG (insoluble and soluble) at the expected molecular weight.

### 6.3 Purification of *P. vivax* CAP-like SAG

As the amount of soluble protein was lower than the *Eimeria* SAGs a much larger scale overexpression (8-10 500ml flasks) were required to obtain enough pure protein (~ 6-9 g of cell pellet). The purification of the *P. vivax* CAP-like SAG followed that developed for the *Eimeria* SAGs as detailed in Chapter 3. Following sonication of the cell paste and centrifugation, the CFE (290 mg/ml) was applied to the Ni-NTA column (His-Trap<sup>TM</sup>HP cartridge, 5 ml), pre-equilibrated with Buffer A (0.5M NaCl + 50 mM Tris PH 8.0) and the protein eluted using 0.5 M imidazole in buffer A [Figure 6.3A]. The fractions corresponding to the second peak of affinity chromatography trace were checked by SDS/PAGE for the expected construct of *P. vivax* CAP-like SAG concentration and subsequently were stirred overnight in a dialysis buffer (buffer A diluted two-fold) containing TEV protease at 22°C, using a ratio of 50 mg of TEV per 1 mg of protein. The reaction mix was then checked for protein concentration and uploaded on the Ni-NTA second column (His-Trap<sup>TM</sup>HP cartridge, 1 ml) to separate the trx-tag from the core domain of the protein.

The flow-through from the second His-Trap column, containing the *P. vivax* CAP-like SAG, was concentrated to 2ml using VivaSpin concentrator fitted with 5 kDa cut-off filter and loaded onto a 16x60 Hi-Load<sup>TM</sup> Superdex<sup>TM</sup>200 column (GE Healthcare), pre-equilibrated with buffer A. The

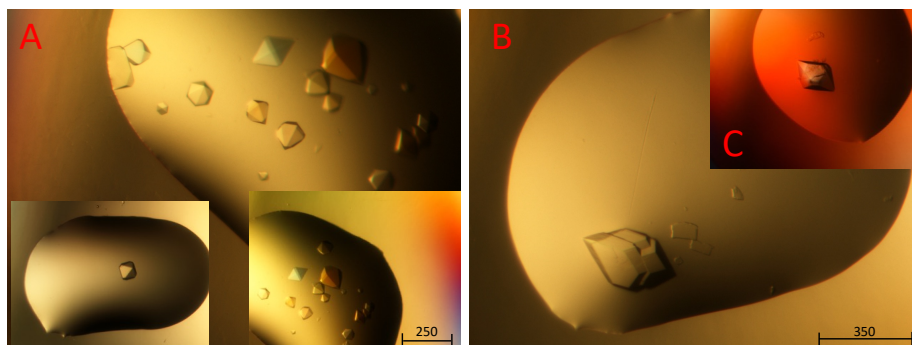
protein was eluted using the same buffer [Figure 6.3B]. The gel filtration showed two peaks the first peak corresponding to the protein aggregating, whereas the second peak corresponding to the combined fractions at the expected molecular weight of the *P. vivax* CAP-like SAG, which were concentrated to 14.4 mg/ml and buffer exchanged in buffer A diluted tenfold for crystallization.



**Figure 6.3: *P. vivax* CAP-like SAG Purification process.** A) 280 nm absorption trace for the affinity chromatography of *P. vivax* CAP-like SAG on 5 ml His-Trap<sup>TM</sup> HP cartridge (GH Healthcare). 290 mg of CFE was applied to the column; Flow rate- 5 ml/min; 50 ml gradient 0 to 0.25 M imidazole in buffer A; Fraction 3 ml. B) 280 nm absorption trace for the gel filtration of *P. vivax* CAP-like SAG on a 1.6x60 Hi-load<sup>TM</sup> Superdex<sup>TM</sup> 200 column (GH Healthcare). 1.2 mg of *P. vivax* CAP-like SAG was loaded to the column; buffer A was used to elute the 2 ml collected fraction; Flow rate- 1.5ml/min. C) SDS-PAGE analysis of *P. vivax* CAP-like SAG after purification steps. Lane-1 protein marker (Mark12<sup>TM</sup>); Lane-2 cells debris; Lane-3 CFE; Lane-4 *P. vivax* before cleavage (after Ni-NTA); Lane-5 *P. vivax* after cleavage; Lane- 6 Ni-NTA 2 elution (imidazole peak fraction); Lane-7 gel filtration loading sample; Lane-8 final prep of *P. vivax* CAP-like SAG.

## 6.4 Crystallization of *P. vivax* CAP-like SAG

The protein was screened against the same set of crystallisation buffers that been used with the *Eimeria* SAG proteins (JCSG, PACT, AmSO4, and PH-Clear). Plates were then incubated at 17C° and checked every week. Crystals only formed after three months of incubation, unlike the *Eimeria* SAG crystals which usually grew within a week. Crystals of *P. vivax* CAP-like SAG grew in several conditions Figure 6.4). The best crystals of *P. vivax* CAP-like SAG from conditions PACT H8, JCSG A9 and JCSG A7 were transferred to a cryoprotectant solution composed of the well solution supplemented with 20% Ethelene glycol, before mounting on a cryo-loop, immersed in liquid nitrogen, and transferred to the Diamond synchrotron for data collection.



**Figure 6.4: Crystals of CAP-like SAG in *P. vivax* SAG.** (A) crystal grew on JCSG A7 condition [0.1M CHES PH 9.5 + 20% (w/v) PEG 8000]. (B) crystal grew on H 8 condition [0.2M sodium sulphate + 0.1M Bis Tris propane PH 8.5 + 20% (w/v) PEG3350] from PACT suite. (C) crystal grew on JCSG A9 condition [0.2M Ammonium chloride PH 6.3 + 20% (w/v) PEG 3350].

## 6.5 *P. vivax* CAP-like SAG Data Collection and Processing

The best data was obtained from a crystal in the PACT H8 condition with diffraction observed beyond 1.3 Å resolution. A total of 3600 images were collected with a rotation range of 0.1° per image at a wavelength of 0.9762 Å on beamline I03. The diffraction images showed clear spots extending to high resolution, indicating good crystal quality. The data were processed by the automatic pipeline at the Diamond synchrotron using the Xia2/Dials protocol in space group P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub>, with unit cell dimensions of a = 30.48 Å, b = 61.64 Å, c = 77.00 Å to give a data set to

1.25 Å [Table 6.1]. However, analysis of the data processing statistics showed that the diffraction data was quite anisotropic, with effective resolution limits along the h, k and l axes of 1.54 Å, 1.58 Å, 1.65 Å respectively. The data were thus cut to a resolution of 1.5 Å [Table 6.1], to give a data set with an overall  $R_{\text{pim}}$  of 0.34 and I/Sig of 16.9 and with a CC1/2 of 1.00 for the highest resolution data.

*Table 6.1: Data collection of Merging statistics for CAP-like SAG in P. vivax crystal*

Overall		
Diamond Beamline	I03	
Wavelength (Å)	0.9762 Å	
Space group	P 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	
Unit cell lengths(Å)	a=30.48, b=61.64, c=77.00	
Unit cell angles( <sup>o</sup> )	α=90,β=90,γ=90	
Molecules per asymmetric unit		
Resolution range (Å)	48.12 – 1.25 (1.27 – 1.25)	48.09 – 1.50 (1.53-150)
Total Reflections measured	535157 (26471)	315021 (15698)
Unique Reflections	40691 (1966)	23919 (1150)
Completeness (%)	99.14 (95.39)	99.70 (97.40)
Multiplicity	13.15 (13.46)	13.20 (13.71)
CC-1/2half	0.9995 (0.4297)	1.000 (0.914)
Mean I/sigma	10.28 (0.22)	16.9 (1.7)
R <sub>merge</sub> (I)	0.1041 (5.4116)	0.071 (1.202)
R <sub>pim</sub>	0.030 (1.515)	0.021 (0.343)

### 6.5.1 Molecular Replacement

Analysis of the unit cell volume, space group symmetry and the molecular weight of the *P. vivax* CAP-like SAG construct (17,509.4 kDa), showed that there was most likely to be a single chain in the asymmetric unit with a solvent content of 40.5% and a Matthews coefficient, which is a measure of the protein packing density in the crystal, of 2.07 Å<sup>3</sup>/Da. The data were used as input for molecular replacement using *Phaser* with the SAG-B19 structure as the search model. The results from the molecular replacement revealed a clear and unique solution in space group P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub>, with a Z scores of 7.5 and 17.5 for the rotational function translation functions, respectively. The log-likelihood gain (LLG), a measure of the probability improvement with the model, was significant at 250 for both the initial and refined stages. The packing of the model within the unit cell was efficient, with only 3 clashes observed, which is relatively minimal and suggests a good fit of the model in the crystal lattice. The molecular replacement solution gave an initial R-factor of 0.46 reflecting that the diffraction was high quality and enough to resolve and initial the first model of *P. vivax* CAP-like SAG but need a slight improvement upon refinement and underscoring the model's suitability.

### 6.5.2 *P. vivax* CAP-like SAG Refinement

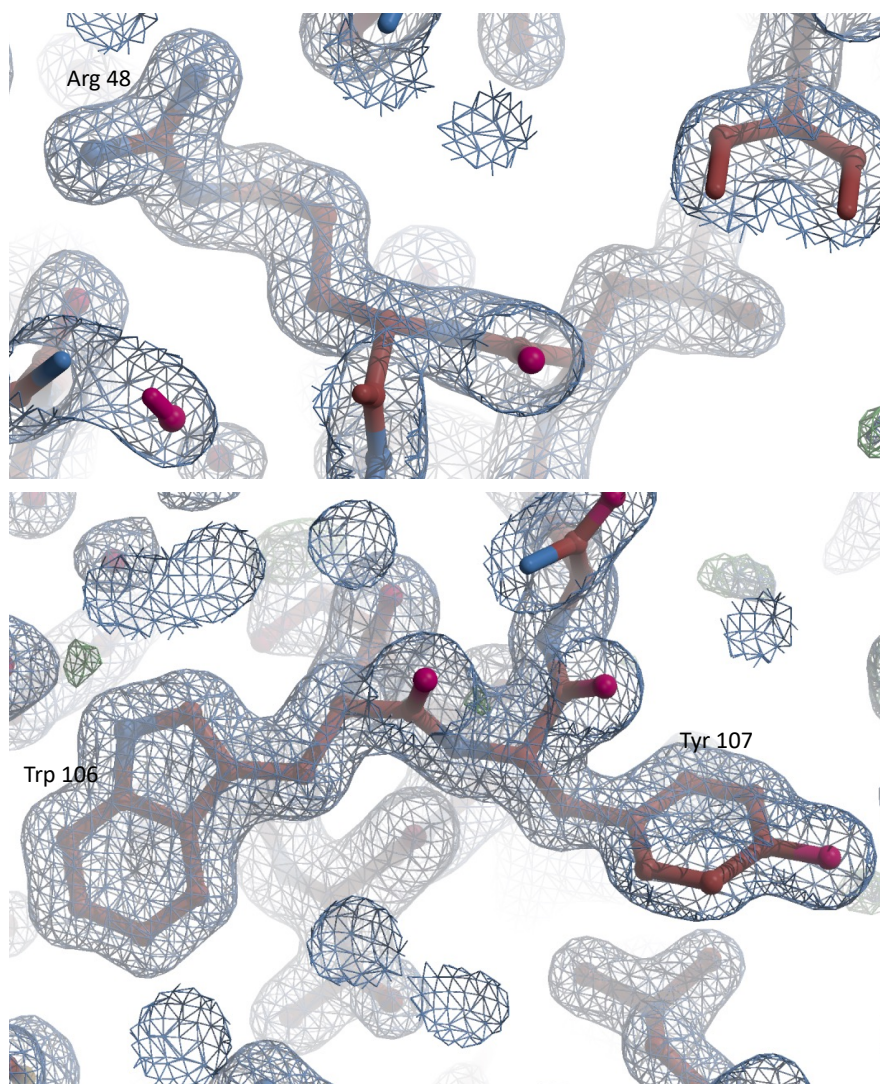
After obtaining the initial electron density map from *Phaser*, Buccaneer (Kovalevskiy *et al.*, 2018; Murshudov *et al.*, 2011) was used to automatically improve the model, by fitting the backbone and side chains of the *P. vivax* CAP-like SAG protein sequence to the electron density map. This process produced a model for all 151 residues of the construct after the TEV cleavage site, in one continuous fragment, with an R-factor of 0.24 and a free-R factor of 0.27, indicating a good level of agreement between the model and the observed data, [Figure 6.5]. However, in this model the final five residues (MFVWE) had been placed in very weak density, and it was not clear from the density whether this was correct, so these five residues were removed, to leave the C-terminal end of the polypeptide as Pro177, where the density was clear for both side chain and main chain [Figure 6.6]. Similarly, at the N-terminus the automatic model building had added the Ser and Ala residues that occur after the TEV cleavage site and before the protein sequence, in very weak density [Figure 6.7] and these two residues were also removed from the model. Subsequent rounds of manual model building, addition of solvent and refinement using COOT and Refmac improved



the model, to give a final model with an R factor of 0.15, Free R of 0.22 and rms bonds/angle deviations of 0.011 and 1.78, respectively [Table 6.2].

*Table 6.2: shows the final model refinement of P. vivax CAP-like SAG.*

<b>Refinement</b>	
Resolution	1.5 Å
Number of non-H atoms	
Protein	2272
Water	89
R <sub>work</sub> /R <sub>free</sub>	0.155/0.219
Average B factor (Å <sup>2</sup> )	
Main chain/side chain	25.32/30.80
Water	34.20
Rmsd bond length (Å) / angle (°)	0.011/1.78
Ramachandran favoured/allowed (%)	97.96/2.04



*Figure 6.5: The initial electron density map for the *P. vivax* CAP-like SAG. The confirmation of map around all atoms can be seen, for examples, Arg48, Trp106 and Tyr107. The map contoured at 1.2 sigma.*

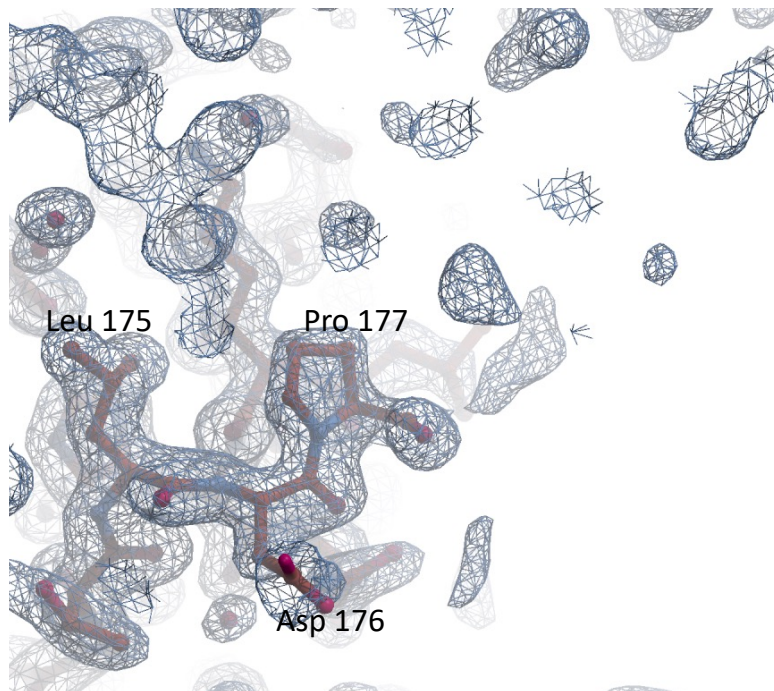


Figure 6.6: **The C-terminus region.** The final model of *P. vivax* CAP-like SAG where no electron density can be seen after Pro177, resulting the disordered of the last five residues (MFVWE) from the construct. The map contoured at 1.2 sigma.

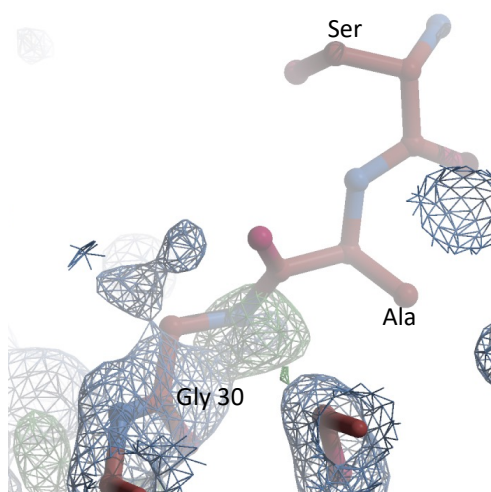


Figure 6.7: **The N-terminus region.** The final model of *P. vivax* CAP-like SAG where the electron density map started at Gly30, and no density can be seen for Ser and Ala that occurs after TEV cleavage site. The map contoured at 1.2 sigma.

### 6.5.3 Structure Validation of *P. vivax* CAP-like SAG.

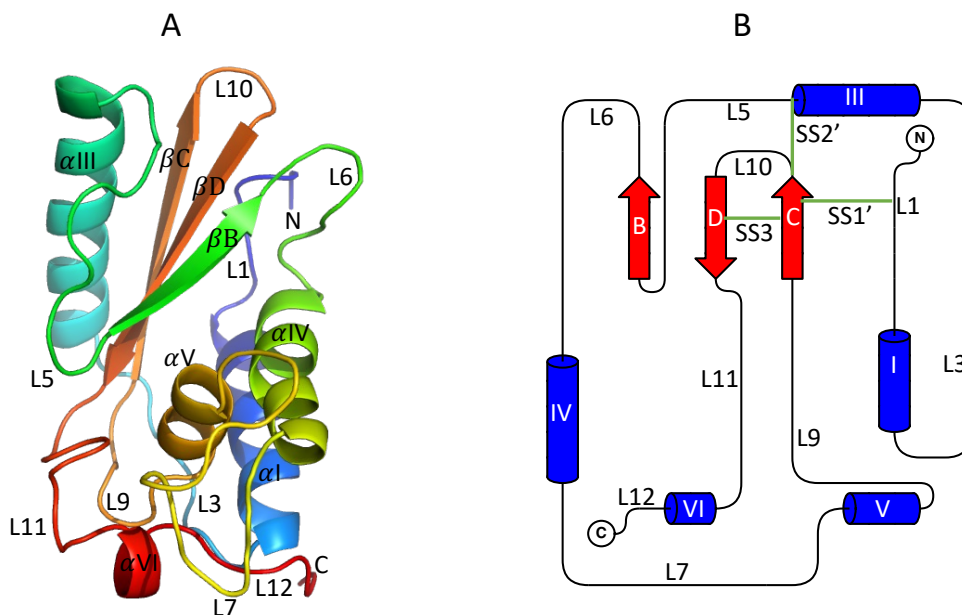
The final model of *P. vivax* CAP-like SAG consisted of 148 residues, 144 of which were in the favoured Ramachandran angles and 4 in the allowed region [Table 6.3]. Most of the residues (98.7%) had side chains with favoured rotamers. However, the residues Arg122 and Gly123 together with the side chains of residues Glu75 and Ile76, occurred in weak areas of density, reflected in higher B-factors for these residues. In the MolProbity analysis, the overall clash score was 3.0, and the structure ranked in the 98th percentile (N=598,  $1.50\text{\AA} \pm 0.25\text{\AA}$ ). Asp124 and Asn52 had a single bond angle in their respective side chains that was slightly outside the normal range.

Table 6.3: MolProbity analysis of the final refined model structure of *P. vivax* CAP-like SAG.

All-Atom Contacts	Clashscore, all atoms:	3.04		98 <sup>th</sup> percentile* (N=598, 1.50Å ± 0.25Å)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Protein Geometry	Poor rotamers	0	0.00%	Goal: <0.3%
	Favoured rotamers	140	98.67	Goal: >98%
	Ramachandran outliers	0	0.00%	Goal: <0.05%
	Ramachandran favoured	144	97.96%	Goal: >98%
	Rama distribution Z-score	-0.34 ± 0.67		Goal: abs (Z score) < 2
	MolProbity score <sup>^</sup>	1.06		99 <sup>th</sup> percentile* (N=4836, 1.50Å ± 0.25Å)
	Cβ deviations >0.25Å	0	0.00%	Goal: 0
	Bad bonds:	0 / 1196	0.00%	Goal: 0%
	Bad angles:	2 / 1634	0.12%	Goal: <0.1%
Peptide omegas	Cis Prolines:	0 / 5	0.00%	Expected: ≤1 per chain, or ≤5%

#### 6.5.4 Structure Description of *P. vivax* CAP-like SAG

For ease of comparison the labelling convention of the *E. tenella* SAG family is used in the following comparison. As expected, the structure of the *P. vivax* CAP-like SAG reveals that its overall fold is that of an  $\alpha\beta\alpha$  sandwich which is very similar to that of the *Eimeria* SAG family with a central  $\beta$ -sheet surrounded by four  $\alpha$ -helices (with a single  $3^{10}$  helix) and with nine surface loops [Figure 6.8]. When compared to the *Eimeria* SAG structures, the *P. vivax* CAP-like SAG structure is closest to that of SAG-C75 [Table 6.4]. Helix  $\alpha$ II of the *Eimeria* SAG-B family is absent in *P. vivax* CAP-like SAG being replaced by a longer loop L3, as seen in the SAG-C75 and SAG-A91 structures. Another difference between the *P. vivax* CAP-like SAG and the *Eimeria* SAGs is that the central  $\beta$ -sheet is constructed from three rather than four strands, with strand  $\beta$ A of the *Eimeria* family being absent and replaced by a longer loop, L1. In addition, although the *P. vivax* CAP-like SAG structure contains the two disulphides present in SAG-C75 (SS2' (C79-C148) connecting L3 to the C-terminal end of  $\beta$ B and SS3 (C143-C156) connecting  $\beta$ B to  $\beta$ C). Thus, SS1 is not conserved rather an extra disulphide (SS1'), not seen in the *Eimeria* SAG structures links loop L1 with  $\beta$ C [Figure 6.8B].



**Figure 6.8: *P. vivax* CAP-like SAG structure.** (A) Chainbow representation of the 3D structure of *P. vivax* CAP-like SAG showing the central three  $\beta$ -strands (labelled  $\beta B$ ,  $\beta C$  and  $\beta D$ ) and the four surrounding  $\alpha$ -helices. (B) 2d topology diagram that illustrates the relative positions of the elements of secondary structure of the *P. vivax* CAP-like SAG indicating the connecting loops (SAG-B19 numbering convention).

In the *Eimeria* SAG structures, the presence of a *cis* peptide (Asn175-P176 in SAG-C75) is crucial for setting up the binding pocket for the buried arginine side chain (Arg51) of the conserved NxxR motif, as described in chapter 5 (section 5.2.1). Interestingly, this crucial *cis*-peptide is not conserved in the *P. vivax* CAP-like SAG structure, where a *trans*-peptide is present instead between the equivalent residues Asp159 and Ser160. However, despite this difference, the hydrogen bonding pattern in the two equivalent regions shares some remarkable similarities. The net effect of the change from a *cis* peptide in the *Eimeria* SAGs to the *trans* peptide in the *P. vivax* CAP-like SAG places OD1 of Asp159 in the latter in the same position as the *cis* peptide carbonyl oxygen of the main chain of the Asn175 in SAG-C75. This results in the same hydrogen bonding pattern in both structures, with OD1 of Asp159 in *P. vivax* CAP-like SAG interacting with the main chain N-Hs of Leu140 and Glu139, compared to the Asn175 carbonyl oxygen hydrogen bonding to the main chain N-H groups of Glu155 and Thr154 in SAG-C75 [Table 6.5, Figure 6.9]. Thus, the important interactions from the 134-140 loop which surround the guanidinyll group of the conserved arginine of the NxxR motif is preserved in the *P. vivax* SAG and the OD1 of the

side chain of Asp159 in the trans-peptide of the *P. vivax* CAP-like SAG structure mimicks the main chain carbonyl of Asn175 in the *cis*-peptide *E. tenella* SAG-C75 structure [Figure 6.9]. Interestingly, in the structure of Tablysin-15 (Xu *et al.*, 2012), another member of the CAP superfamily, a trans-peptide is also found at this position (Ser189-Ser190) with the side chain hydroxyl of Ser189 performing an equivalent role to that of Asp159 in *P. vivax* SAG and forming hydrogen bonds to the NH groups of Thr165 and Ala166 [Figure 6.10]. In other members of the CAP superfamily a *cis* configuration is found.

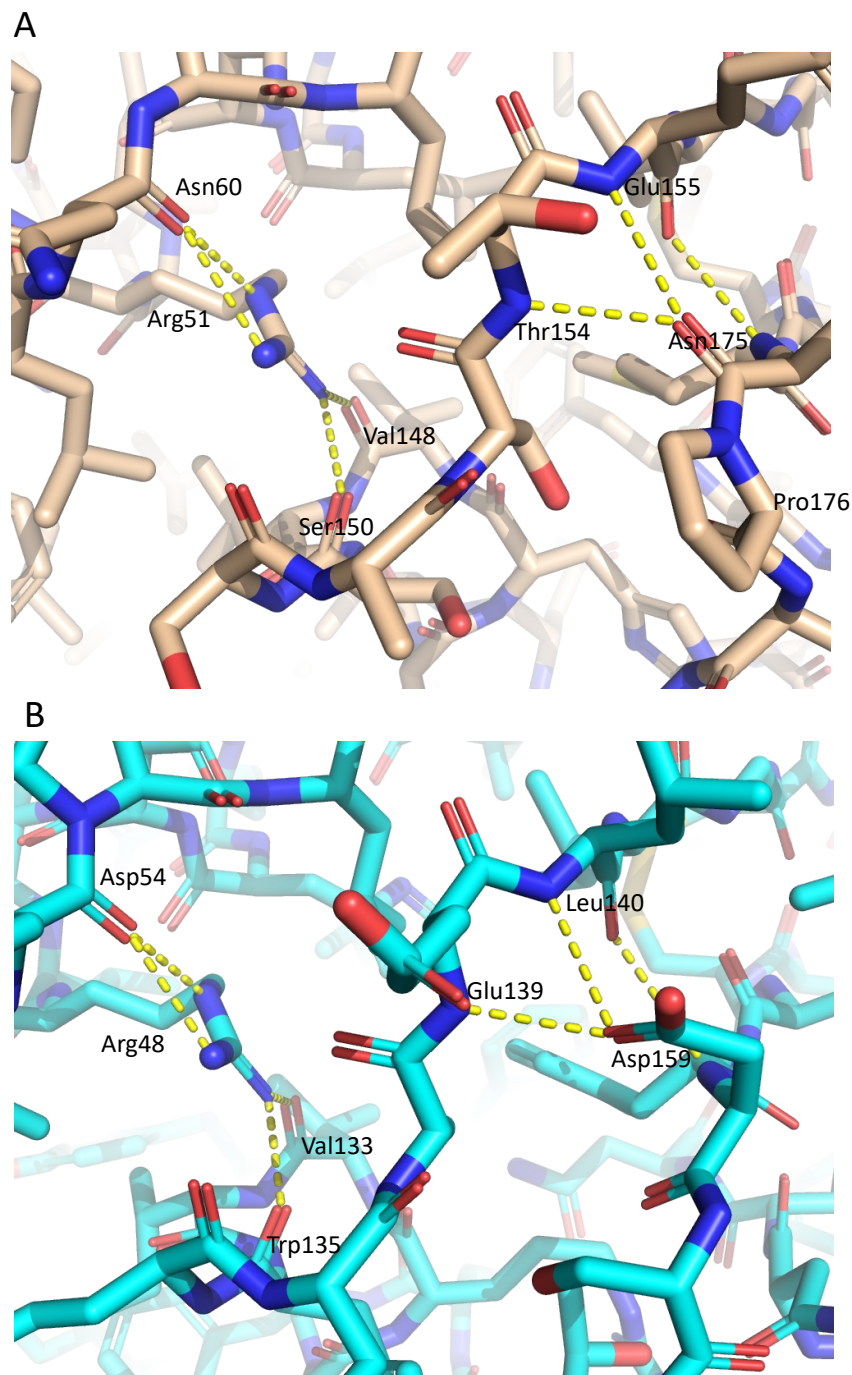
Table 6.4: Similarity between the structure of the *P. vivax* CAP-like SAG and *E. tenella* SAG families A, B and C.

Protein Comparison	rmsD C $\alpha$ (Å)	Sequence Identity (%)	Number of Equivalenced residues
<i>P. vivax</i> CAP-like SAG vs <i>E. tenella</i> SAG-A91	2.6	15	126
<i>P. vivax</i> CAP-like SAG vs <i>E. tenella</i> SAG-B41	2.5	11	126
<i>P. vivax</i> CAP-like SAG vs <i>E. tenella</i> SAG-C75	2.0	12	124

Table 6.5: Hydrogen bonding pattern of the *cis*-peptide N175-P176 in SAG-C75 compared to the trans-peptide D159-S160 in *P. vivax* CAP-like SAG.

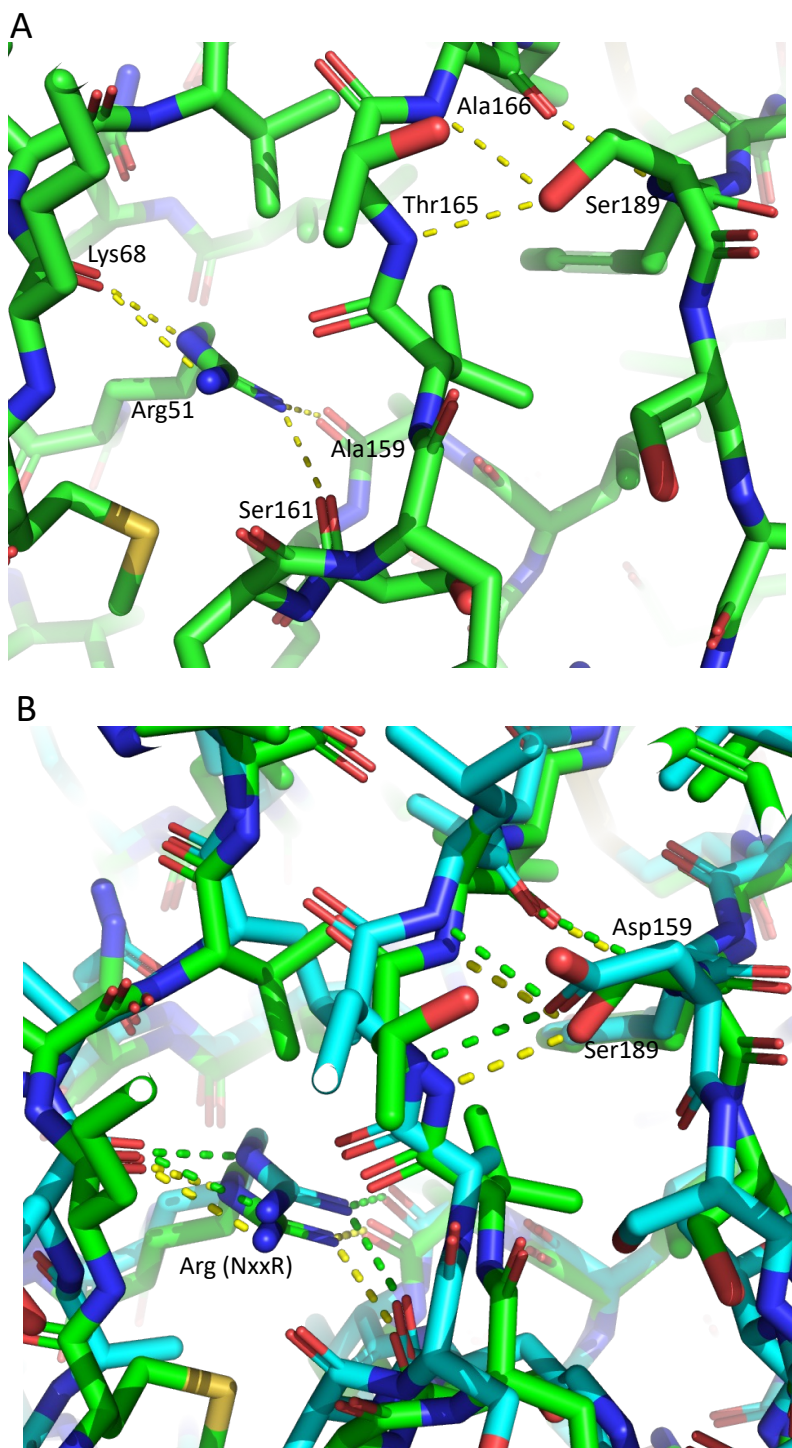
	SAG-C75 ( <i>cis</i> peptide)	<i>P. vivax</i> CAP-like SAG (trans-peptide)
H-bond 1	N175 N-H-----O=C E155	D159 N-H-----O=C L140
H-bond 2	N175 C=O-----H-N E155	D159 OD1-----H-N L140
H-bond 3	N175 C=O-----H-N T154	D159 OD1-----H-N E139
H-bond 4	P176 C=O-----H-O S153	n/a





**Figure 6.9: Key interactions creating the pocket for the conserved buried arginine.** A) The hydrogen bonds between the main chain carbonyl of Asn175 that forms part of the Asn 175-Pro 176 cis-peptide in SAG-C75 and the main chain NHs of Glu155 and Thr154 that lie in the loop next to the guanidinyll group of the buried arginine, Arg 51. B) In *P. vivax* CAP-like SAG an equivalent hydrogen bonding pattern is made by the side chain oxygen of Asp159 and the main chain NHs of Leu140 and Glu139 with the Asp159-Ser160 peptide having that a trans -rather than the cis- configuration more commonly in the other SAGs.





**Figure 6.10: Key interactions creating the pocket for the conserved buried arginine in Tablysin-15 structure (PDB 3U3U).** The hydrogen bonds between the side chain hydroxyl of Ser189 and the main chain NHs of Thr165 and Ala 166 with the Ser189-Ser190 peptide having that a trans configuration as seen in *P. vivax* CAP-like SAG. (B) superposition of the *P. vivax* CAP-like SAG and Tablysin-15 in the region of the conserved arginine showing the equivalent hydrogen bonding patterns of Asp159 and Ser189 that permit the adoption of the trans peptide.

## 6.6 Conservation of the *P. vivax*-like SAG across different species of *Plasmodia*

Analysis of the genomes sequences of range of *Plasmodium* species, including the five species that are most important in causing disease in man (*P. vivax*, *P. knowlesi*, *P. malariae*, *P. ovale*, and *P. falciparum*) shows that they each contain a single CAP-like SAG equivalent to that found in *Eimeria*. Close inspection of the aligned sequences of the important human parasites shows that can be broadly grouped into two groups. In one group, which includes the pathogens *P. vivax* and *P. knowlesi*, the sequences of this SAG are very similar (88% identity over the region covered by the ordered part of the structure). In the other group, which includes the other human pathogens *P. malariae*, *P. ovale*, and *P. falciparum*, the sequences, are more divergent (60%, 64% and 60% identical respectively). These similarities and differences reflect in part the close evolutionary distance between *P. vivax* and *P. knowlesi* [Figure 6.11].

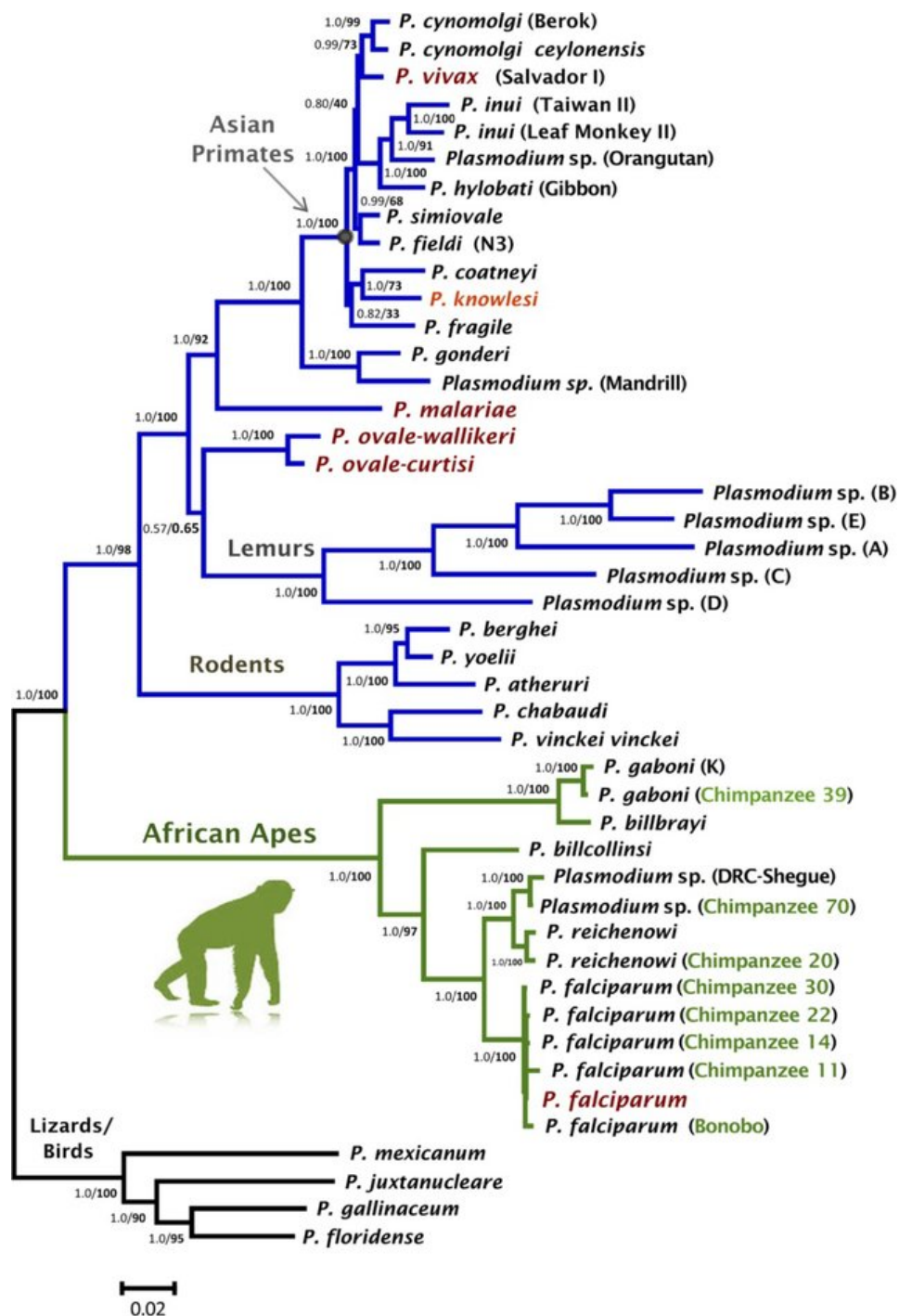


Figure 6.11: *Phylogenetic tree of Plasmodium spp based on the analysis of sequences of complete mitochondrial genomes.* Branches indicated in green correspond to plasmodium parasites from African Apes, whilst those in blue and black are from other mammals and birds/reptiles, respectively. The most important species that cause disease in man are highlighted in light or dark red. The figure is adapted from (Pacheco et al., 2013).

Analysis of the positions of the conserved residues between the *P. vivax* and *P. knowlesi* SAGs shows that 59 of the 61 largely buried residues (defined as having an accessible surface area of less than 20%) are identical (97%). However, given the high sequence identity between CAP-like SAGs of these species a significant number of the surface exposed residues (defined as having an accessible surface area of greater than 20%) are also conserved (74/87, 85%) [Figure 6.12, 6.14]. Therefore, it would be expected that antibodies raised against *P. vivax* SAG could show considerable cross-reactivity against the SAG from *P. knowlesi*. Therefore, if the function of this SAG is important in the life cycle of the parasite, a vaccine using this protein might be useful against these strains.

Analysis of the position of the 69 residues that are conserved across all five *Plasmodium* revokes the above reveals that 44 of them are largely buried and only 25 lies on the surface [Figure 6.13, 6.14]. The most prominent patch of these surface residues lies close to  $\alpha$ IV,  $\beta$ C,  $\beta$ D and the L7 loop. This region includes the conservation of residues such as Trp106, Tyr107, Glu108, Gly109, Ile110, Asn111 from helix  $\alpha$ IV, Tyr113, Asp114, Phe115, Glu116, Leu117 and Gly118 from the N-terminal part of L7. Whether this region represents an important functional part of the molecule to account for its conservation is unclear. However, the fact that sequences on the surface of these other three strains are so different would suggest that a *P. vivax* SAG vaccine would probably not be useful against every strain of *Plasmodium* and an approach using a multivalent vaccine would be necessary. There might also be problems associated with the emergence of mutant strains that could escape any vaccine. Clearly the establishment of the role of this protein and its expression level in different life cycle stages of the parasite would be important in predicting the usefulness of using this protein as a target for vaccine development.

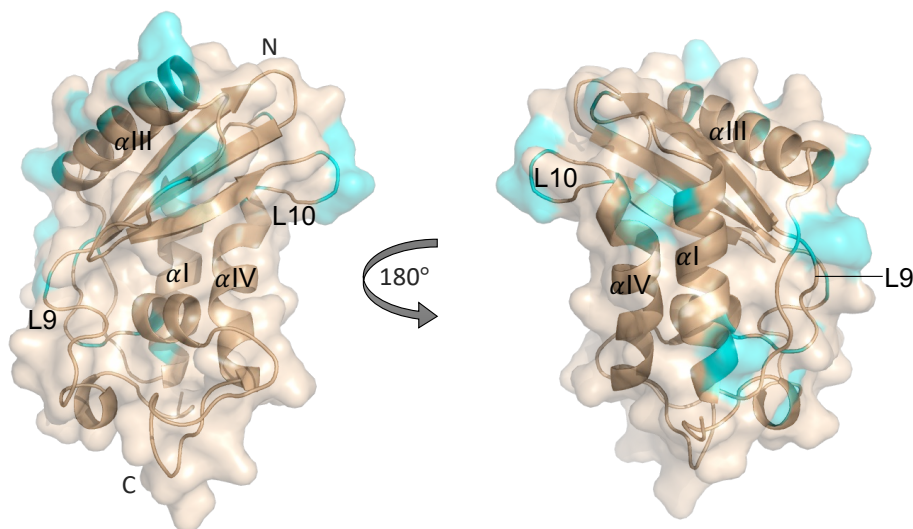


Figure 6.12: **Pattern of conservation on the surface of the *P. vivax* and *P. knowlesi* CAP-like SAGs.** Two different views of the molecular surface (solid) of the *P. vivax* CAP-like SAG is drawn on a backbone cartoon. A cyan colour indicates regions of sequence difference between this protein and that from the *P. knowlesi* homologue whereas a brown colour represents regions of similarity.

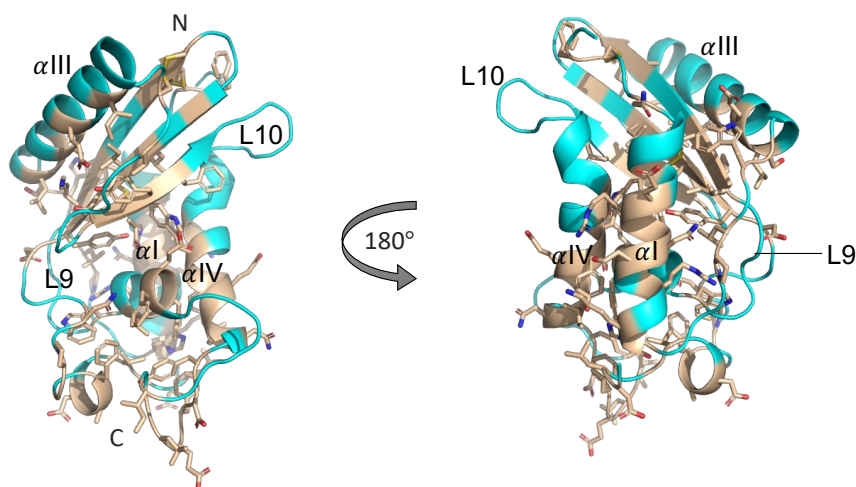


Figure 6.13: **The positions of sequence conservation and sequence difference across the CAP-like SAGs from the important human plasmodium species.** Two different views of a cartoon of the structure of the *P. vivax* CAP-like SAG with residues that are conserved coloured brown and drawn with their side chains shown as sticks. The cyan colour indicates regions of sequence difference. Some of the loops and regions of regular secondary structure are labelled. The most prominent patch of conserved surface residues is indicated.

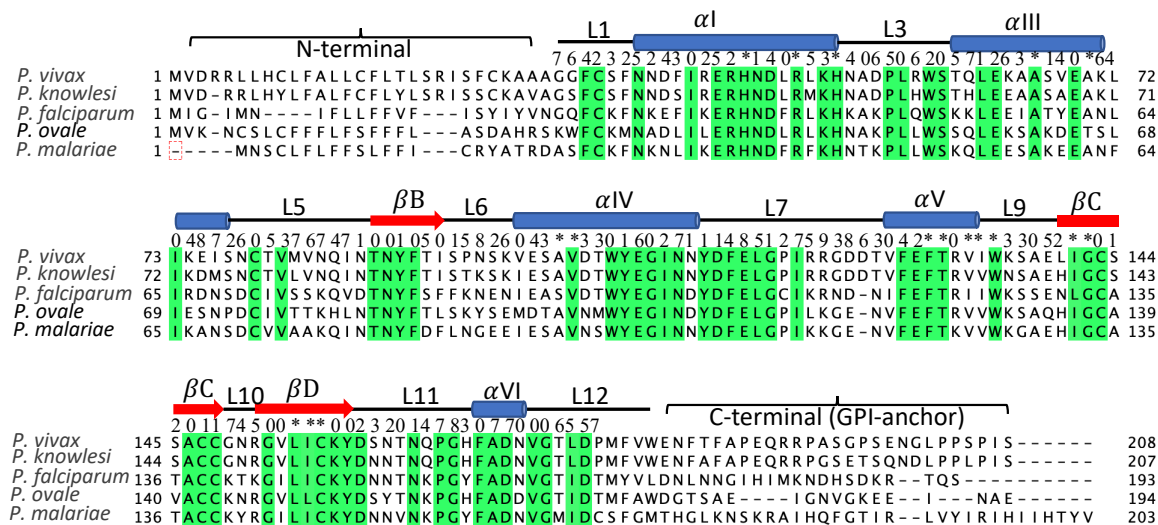


Figure 6.14: Multiple sequence alignment of the CAP-like SAGs in *Plasmodium* strains that are important in human disease. The sequences compared are those of the core region of the CAP-like SAGs from *P. vivax* (PVC01\_010012600), *P. knowlesi* (PKNH\_0104300), *P. falciparum* (PF3D7\_0705800), *P. ovale* (PocGH01\_01013600) and *P. malariae* (PmUG01\_01017300). The positions of non-conserved residues are indicated by an orange background. Residues conserved in all five sequences are highlighted in green. The positions of the loops and secondary structure elements in the CAP-like SAG from *P. vivax* are shown above the sequences and labelled. The last five residues at the C-termini highlighted with orange background cannot be seen on the electron density of the structure. The solvent accessible area of residues in the CAP-like SAG from *P. vivax* were calculated by the method of Lee and Richards (Lee and Richards, 1971). The resulting accessible areas were then expressed as a fraction of the total solvent accessible surface area for each amino acid in a GXG tripeptide and are given as a sequence of integers 0-9 above the sequences (where each number represent residues which have 1-10%, 11-20%, 21-30%, etc of the surface solvent accessible. Totally buried residues indicated by \*).

## Chapter 7 Discussion

### 7.1 Design of a new Trx construct

The initial objective of the work described in this thesis was to design a new, more efficient construct for the expression of a wider range of SAGs and, in particular, to avoid problems caused by the non-specific cleavage of the expressed SAG proteins arising from the use of the enterokinase cleavage site in the original construct (Ramly, 2012). Initial work focussed on the replacing the entire construct with a *PelB* expression system which directs the expression of the polypeptide to the periplasm. However, whilst attempts to use this system on a control protein (SAG-B19) showed expression, unfortunately the protein was insoluble and so this approach was abandoned.

A new construct was therefore created based on the Ramly system, but replacing the enterokinase site by a TEV site together with a further modification by removing the His-tag from the C-terminus as this was unnecessary. In addition, by using the structure of SAG-B19, it was possible to make improved predictions for the N- and C-terminal ends of the construct so as to produce a fragment that represented the “core” domain of the SAG protein omitting the secretion signal and the GPI-anchor regions for structural studies. This strategy was successful as judged by the soluble expression of twenty two proteins studied as part of this thesis and the structure determination at high resolution of the seven SAGs proteins described in the earlier chapters. Nevertheless, analysis of the aligned sequences of the wider family of SAGs does indicate some variation in their length in the region of both the N- and the C- areas of the “core” domain. This gives rise to some uncertainty as to the optimal expression construct for some of the proteins. For example, a construct was designed to express the core domain of the CAP-like SAG from *T. gondii* (Uniprot S7UPE9) (Ala141–Arg302), but in expression tests the protein was always in the insoluble fraction, irrespective of temperature and IPTG concentration. The structure of this protein predicted by alpha-fold was inspected and this suggests that the C-terminal helix is longer than those observed in the other SAG structures determined to date, with the helix ending at Gly310. Thus, the construct used which ended at Arg 302 might be too short. Therefore, would be useful to redesign the construct for this particular SAG, perhaps ending at Arg315, and experiments to test expression on constructs of different length might well be necessary for this and possibly other SAGs. It might also be useful to investigate whether it is necessary to include the trx-tag as part of



the construct to confer solubility of the product. Furthermore, it would also be interesting to see whether the use of codon optimised genes, as trialled in the constructs described in this thesis, is better for solubility and expression compared to the use of genomic DNA sequences as used in the earlier work of Ramly and co-workers (Ramly, 2012). Other factors that might be considered for investigation would be whether removing the N-terminal his-tag is essential. However, it is clear that the impact of the latter might vary from one protein to another due to effects on crystal packing of the need for the additional residues to be accommodated in the crystal lattice (Yamada *et al.*, 2017).

## **7.2 CAP-like SAGs in the wider apicomplexan parasite family.**

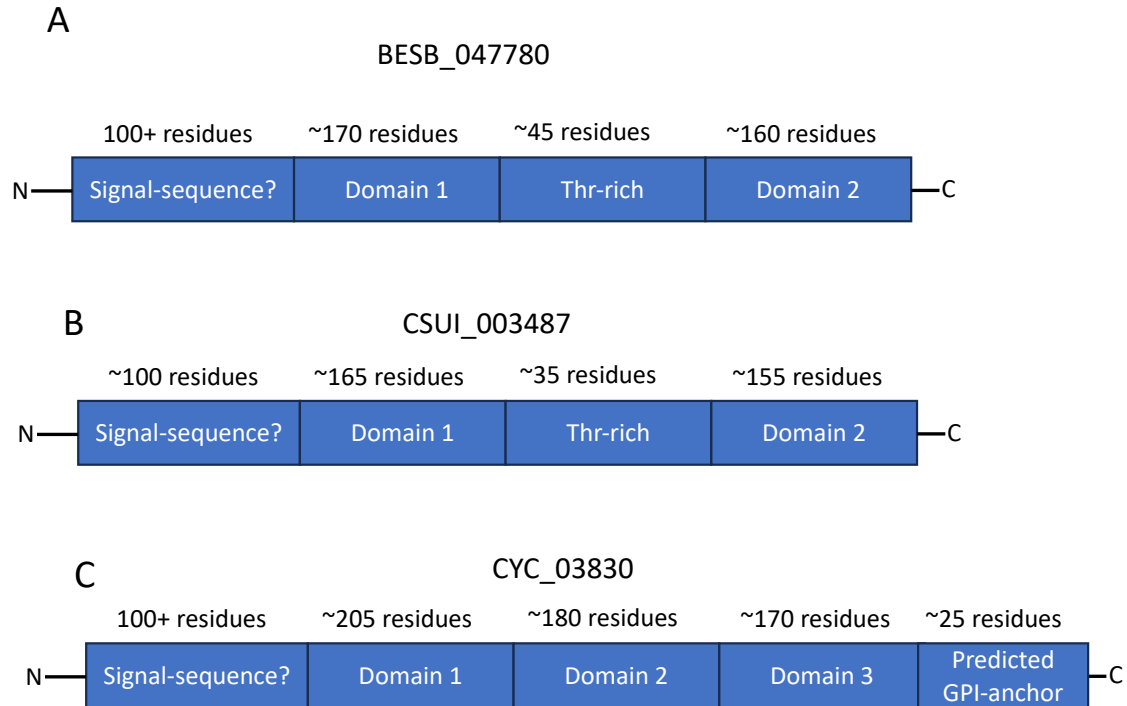
The programme of structure determination on the superfamily of *Eimeria tenella* SAGs succeeded in producing atomic models for 6 new members including representatives of each of the A, B and C families all of which were shown to possess a closely related CAP-like fold [Table 7.3]. In contrast, in other apicomplexans such as *Plasmodium* and *Toxoplasma* they express a superfamily of cysteine-rich proteins belonging to the SRS-class rather than the CAP-class. This had led to the suggestion that the different parasites might exclusively have cysteine-rich superfamilies of proteins belonging to one or the other class but not to both (Ramly *et al.*, 2021). However, the discovery of a CAP-like SAG in *P. vivax* and other *plasmodium* species with a similar fold to that adopted by the *Eimeria* SAGs clearly showed that this is not the case.

Therefore, a wider sequence analysis was undertaken to search for CAP-like SAGs in the genomes of other apicomplexan parasites that cause disease in either humans and/or animals. This search included an analysis of the genomes of the parasites *Babesia bovis*, *Theileria orientalis*, *Cyclospora cayetanensis*, *Toxoplasma gondii*, *Besnoitia besnoiti*, *Neospora caninum* and *Cystoisospora suis*. Using the sequence of *P. vivax* CAP-like SAG as a probe, it quickly emerged that potential CAP-like homologues could be identified in the genomes of each of these parasites. This analysis showed that the genomes of *B. bovis*, *T. orientalis* contained a single CAP-like SAG, whereas, in *C. cayetanensis*, *T. gondii*, *B. besnoiti*, *N. caninum* and *C. suis*, between five and nine copies of a CAP-like SAG sequence could be found [Table 7.5,]. In a number of cases the CAP-like SAG genes could be seen to be located in tandem on the parasite genome. For example, in *Besnoitia besnoiti* strain Ger-1, the four genes BESB\_017010 – BESB\_017040 are next to each



other on chromosome X. Of the remaining 5 CAP-like SAG genes one is on chromosome II (BESB\_039070) and the chromosome on which the others are located is unknown.

Interestingly, some of these CAP-like SAGs proteins were found to have two or three SAG domains within one long polypeptide. For example, the gene (CYC\_03830) of *C. cayetanensis* encodes a protein of 726 residues containing three CAP-like SAG domains (domain 1 Lys135 – Lys340, domain 2 Ala345 – Pro525, domain 3 Arg530 – Pro698) connected by short linkers. Moreover, analysis of the sequence of the full-length protein using the NetGPI server indicates that it contains a GPI-anchor attachment sequence at the C-terminal region (Ser700 – Phe726). While this suggests that the product of the *C. cayetanensis* CYC\_03830 gene may be expressed on the cell surface, the absence of a predicted signal sequence for the region covered by the additional 100+ residues N-terminal to the first CAP-like domain [Figure 7.1] leaves room for uncertainty as to whether there is a GPI anchor attached to the mature protein. Other examples of proteins containing multidomain CAP modules includes the gene products of BESB\_047780 of *B. besnoiti* and CSUI\_003487 of *C. suis* which appear to code for proteins that contain two CAP-like SAG domains in one polypeptide of 518 and 454 residues, respectively. In both of these proteins the two domains are separated by a block ~ 40 residues in which more than 75% of the residues are threonine and which is predicted to be unstructured by alphafold. However, neither sequence appears to encode a GPI-anchor attachment site, although they do contain an additional 100+ residues of unknown function at the N-terminus which again are not predicted to be an N-terminal signal sequence. Taken together the sequence data suggest that these proteins are not expressed on the cell surface.



**Figure 7.1: Multiple CAP-like SAG domains in proteins from different apicomplexan parasites.** A) A schematic to show the domain structure of *B. besnoiti* BESB\_04778. The two CAP-like SAG domains are separated by a region of 40+ residues that is very rich in threonine. An additional domain of 100 residues of unknown fold lies at the N-terminus; B) A schematic to show the domain structure of *C. suis* CSUI\_003487. As with *B. besnoiti* BESB\_04778, the two CAP-like SAG domains of *C. suis* CSUI\_003487 are separated by a threonine rich region with an additional domain of approximately 100 residues at the N-terminus. C) The domain structure of *C. cayetanensis* CYC\_03830. Three CAP-like SAG domains are separated by short linkers with an additional domain of unknown fold and function at the N-terminus. The position of the predicted GPI-anchor attachment site at the C-terminus is indicated.

A multiple sequence alignment was constructed to include a single representative CAP-like sequence from each of these parasites against the sequence of the *P. vivax* CAP-like SAG [Table 7.1, Figure 7.2]. This revealed that whilst the sequences were clearly related and could be confirmed to belong to a CAP-like SAG, only 16 of the 148 residues of *P. vivax* CAP-like SAG were identical across the homologues from all these species. Further analysis reveals that all of these sixteen conserved residues are largely buried (an exposed surface area of less than 20%) in the *P. vivax* CAP-like SAG and are thus mainly involved in maintenance of the three-dimensional structure [Figure 7.4]. The conserved residues include the NxxR motif and associated residues that form the buried arginine pocket (N45, R 48 and L56); five aromatic and two aliphatic hydrophobic residues (Ala 102, Trp106, Tyr107, Phe130, Trp135, Leu 154 and Phe168) that pack in the hydrophobic core of the protein, an asparagine residue, Asn89, whose carboxamide side chain forms a hydrogen bond to the with the side chain of a conserved arginine, Arg132, that is packed against the side chain of another conserved asparagine, Asn163. The final four conserved residues are the four cysteines (Cys79, Cys148, Cys143, Cys156; *P. vivax* numbering) that form the SS2 and SS3 disulphide bonds in *P. vivax* CAP-like SAG, [Figure 7.2]. However, the third disulphide bond in the *P. vivax* CAP-like SAG (SS1', Cys33 - Cys147, which connects loop L1 to  $\beta$ C is not completely conserved across the homologues from each of these other parasites. Instead, the sequences are subdivided into two distinct conservation groups. Group one (haemoparasites) includes *P. vivax*, *B. bovis*, and *T. orientalis* which conserve SS1' [Figure 7.2]. However, in group two (coccidia), comprising *T. gondii*, *N. caninum*, *C. suis*, *B. besnoiti* and *C. cayetanensis*, the position of the third disulphide is very similar to that of SS1 in the *E. tenella* SAG-A and SAG-B families linking  $\alpha$ I and  $\alpha$ IV [Figure 7.3]. This division of the pattern of sequence conservation into two groups is interesting in that it matches the evolutionary relationship and division of the apicomplexan parasites into the two distinct groups that require (*Plasmodium*, *Theileria*, and *Babesia*) or do not require an intermediate host (*Toxoplasma*, *Neospora*, *Cystoisospora*, *Cyclospora*, and *Besnoitia*).

Whilst this study clearly identified the presence of one or more a CAP-like SAGs in each of the apicomplexan parasites, this leaves open the question as to whether or not they all contain an example of an SRS-like SAG as seen for *Toxoplasma* and *Plasmodium* (He *et al.*, 2002; Dietrich *et al.*, 2022). In previous work SRS-like SAG proteins have been identified in *Besnoitia besnoiti*, *Neospora caninum* (O'Toole and Jeffrey, 1987) and *Cystoisospora suis* (Palmieri N, 2017). To

confirm the presence of this class of proteins in other apicomplexans including *Babesia* and *Theileria* a BLAST was conducted using the sequence of the *Plasmodium falciparum* (Pf12 PDB 2YMO) SRS-like SAG. This led to the identification of a number of homologues in *Babesia* (for example, Uniport D5FW38) and *Theileria* (for example, Uniport A0A976M8D1) species. However, no SRS-like SAG homologues were found in *Cyclospora cayetanensis* or any of the *Eimeria* species using either the sequence of *Toxoplasma* SAG1 SRS-like SAG or *Plasmodium falciparum* (Pf12) SRS-like SAG. This suggests that the genomes of these parasites only encode CAP-like SAGs though it remains possible that these organisms contain SRS-like SAGs with very low sequence similarity such that they have been overlooked.

*Table 7.1: The level of identity between an example of the sequences of the CAP-like SAGs in selected apicomplexan parasites and the P. vivax homologue.*

Parasite	(%) to <i>P. vivax</i>	Gene ID
<i>Babesia bovis</i>	23	BBOV_III003800
<i>Theileria orientalis</i>	25	TOT_020000205
<i>Besnoitia besnoiti</i>	18	BESB_017040
<i>Toxoplasma gondii</i>	12	TGGT1_288220
<i>Neospora caninum</i>	16	NCLIV_012410
<i>Cystoisospora suis</i>	12	CSUI_002672
<i>Cyclospora cayetanensis</i>	21	CYC_03341



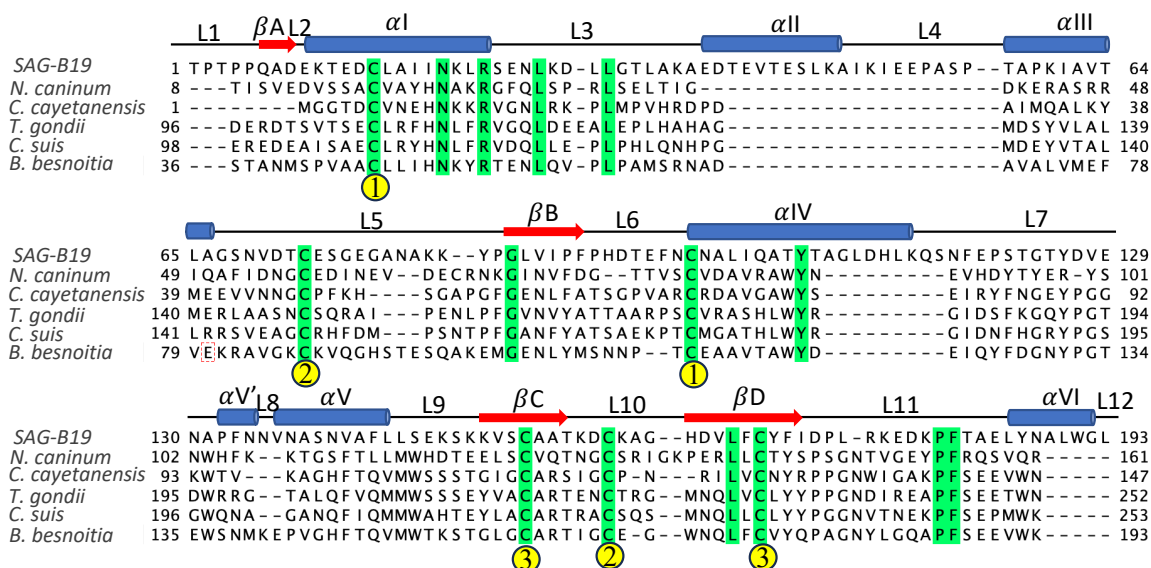


Figure 7.3: Alignment of CAP-like SAGs from a range of Apicomplexans including SAG-B19. The sequences are aligned with the aid of the core domain of SAG-B19, the position of whose secondary structure are shown above the aligned sequences. The yellow circled number indicated to the disulphide bonds (numbering in SAG-B19).

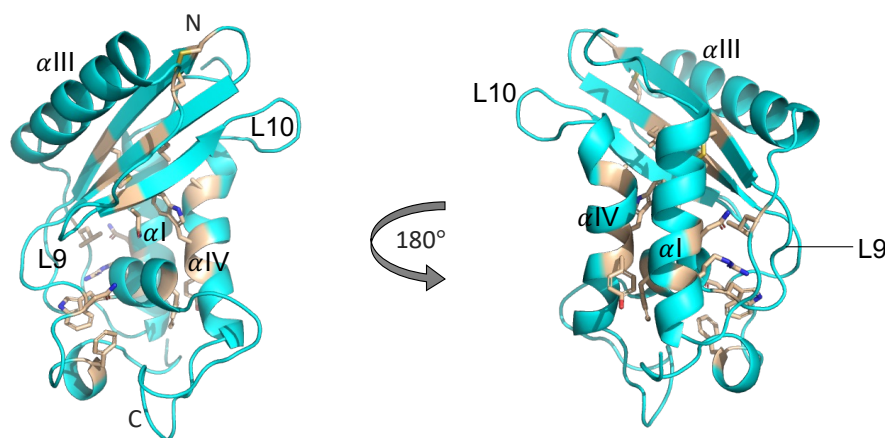


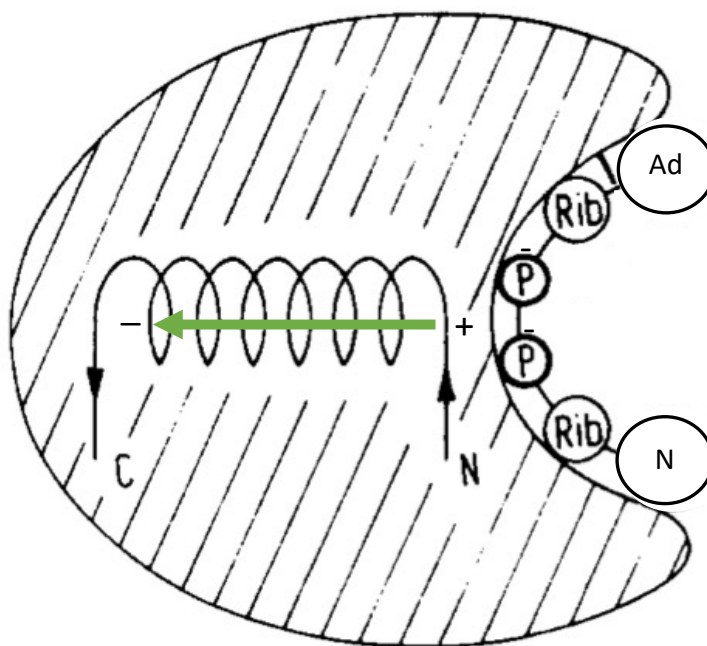
Figure 7.4: The conserved residues between the *P. vivax* CAP-like SAG and the SAG proteins from a range of apicomplexan parasites. The cyan colour showed the divergence between the proteins in *P. vivax* CAP-like SAG structure, while the brown sticks showed the identical residues between them.

The analysis of these newly identified CAP-like SAG proteins revealed that the majority of the conserved residues are largely buried [Figure 7.4]. As a result, their molecular surfaces do not contain a conserved region as might have been required if the role of these SAGs is involved binding to either a conserved host receptor or to other molecules on the cell surface such as surface glycans. Indeed, analysis of their molecular surfaces showed that they were quite divergent with not only no conserved patches of exposed residues but also no similarities in the distribution of patches of positively or negatively charged residues. Taken together these results suggest that the primary function of some, or perhaps all, of these proteins is more likely to be associated with the stimulation of the immune response as part of the process of parasite invasion. If this is indeed the case, then this mirrors the antigenic variation role played by the hundreds of variant surface glycoproteins (VSGs) displayed by trypanosomes during infection (Onyilagha and Uzonna, 2019). Might be the same situation linked to the SAGs in *Eimeria*, even with 1% of the trypanosome genome is made up of the genes for VSGs.

### **7.3 Critical conserved features of the CAP- superfamily structure.**

The successful structure determination of the range of SAGs presented in thesis has permitted a wider comparison of the key structural features found in members of CAP superfamily. This includes an analysis of the positions and numbers of disulphide bonds, the conservations of key residues and motifs and the positions of more unusual *cis* peptides. This analysis is presented in [Table 7.4] which shows that while many of proteins contain disulphide bonds, some, for example with the human CAP-like protein, GAPR-1, and the bacterial protein, BB0689, do not. In those proteins with disulphides the only strongly conserved disulphide is SS3 and across the entire CAP-family the position of all of the other disulphides can vary widely. One of the most prominent features of the CAP structures is the conservation of the side chain of the buried arginine which forms part of the NxxR motif. To satisfy the positive charge of this residue its guanidiny group is positioned at the C-terminal end of helix  $\alpha V$ , where the negatively charged helix dipole will help stabilise the guanidyl group [Figure 7.6]. This interaction is similar to the way in which the positive end of the helix dipole can stabilise the binding of negatively charged molecules as seen for example in the binding site of the pyrophosphate group the NAD(P) moiety in the dehydrogenases [Figure 7.5] (Hol, 1985). The finding that the position of these arginine residues are so similar

suggests that this interaction may be a critical part of the protein folding pathway of proteins with the CAP fold perhaps by the interaction of the positively charged guanidinyll group with the negatively charged end of the growing dipole of  $\alpha V$  as it folds. Indeed, an early analysis of the type of residues found in the N- and C-terminal regions of alpha helices (Chou and Pasman, 1974) points to the higher probability of finding negatively charged residues at the N-terminal end of a helix and positively charged residues at the C-terminal end [Table 7.2]. Whilst this is not the same situation as seen with the conserved arginine in the CAP structures as this residue interacts with the C-terminal end of one helix ( $\alpha V$ ) but is positioned at in the C-terminal region of a different helix ( $\alpha I$ ), nevertheless the same principle might be at work.



*Figure 7.5: **Schematic diagram of helix dipole.** The pyrophosphate moiety is close to the N-terminus of the  $\alpha$ -helix while the negative charges of the C-end of  $\alpha$ -helix compensated by one, or more, positively charged side chains (The NAD(H) components are shown in the schematic as Ad = adenine, N = nicotinamide, Rib = ribose and P = phosphate). Figure adapted from (Hol, 1985).*

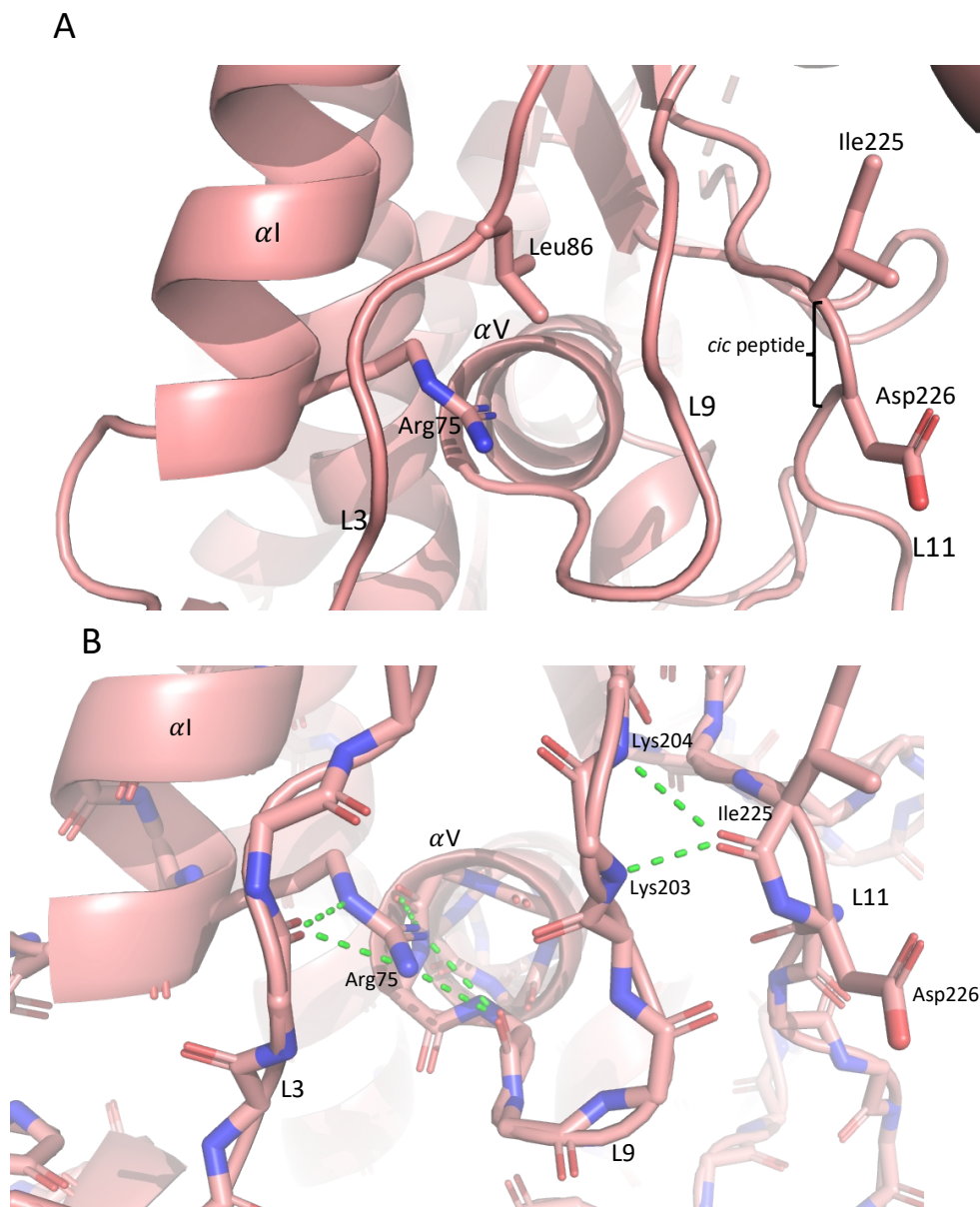


Table 7.2: Frequency of helical boundary residues in 15 proteins\*

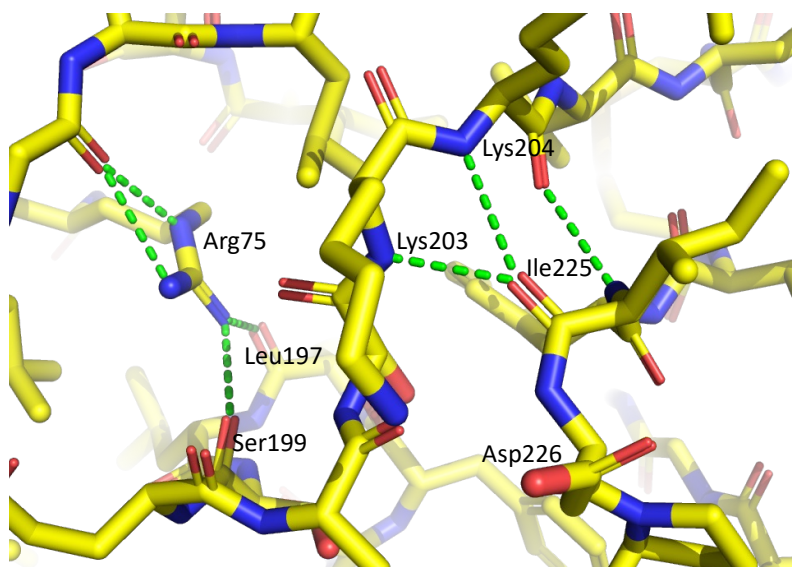
	$f_{hN^+}$		$f_{hC^{++}}$
Pro	0.212	His (+)	0.216
Asp (-)	0.207	Lys (+)	0.160
Glu (-)	0.195	Gln	0.158
Ala	0.140	Arg (+)	0.154
Trp	0.136	Cys	0.148
Thr	0.122	Met	0.143
Gln	0.116	Glu (-)	0.124
Phe	0.098	Ala	0.118
Asn	0.090	Val	0.116
Ser	0.079	Phe	0.110
Cys	0.074	Leu	0.102
Met	0.071	Asn	0.090
Tyr	0.070	Ser	0.084
Ile	0.066	Ile	0.075
Val	0.061	Asp (-)	0.054
Gly	0.060	Tyr	0.050
Lys (+)	0.057	Thr	0.045
Leu	0.056	Trp	0.045
His (+)	0.054	Gly	0.039
Arg (+)	0.038	Pro	0.000

\*Table taken from (Chou and Fasman. 1974)

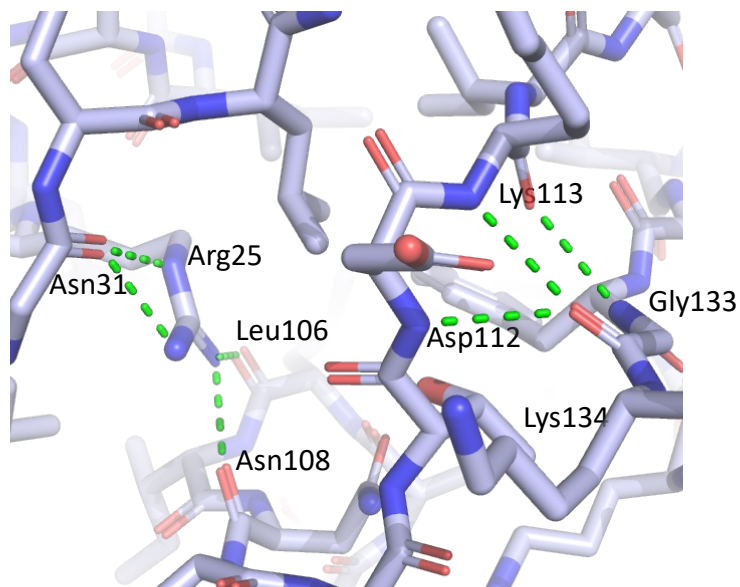
In addition, the guanidinyll group forms an intricate pattern of hydrogen bonding to nearby carbonyl groups of the protein main chain whose local structure is commonly maintained by the conservation of a nearby *cis* peptide [Figure 7.7]. Indeed, in some of the structures examples of a rare non-Pro *cis* peptide can be found at this position [Figure 7.7]. This is highly unusual as only one in ten thousand peptides are thought to adopt this conformation (Williams *et al.*, 2018). Interestingly however, in some of the CAP structures including the *P. vivax* CAP-like SAG and in the structure of a different CAP-like protein, tablysin-15 (Xu *et al.*, 2012), sequence changes allow this unusual interaction to be replaced by a more conventional *trans* peptide at the equivalent position and an interaction with the side chain of an aspartate or serine, respectively, as described in chapter 6 [Figure 6.9 and 6.10]. In the structure of BB0689 (PDB 4D53), a member of the bacterial CAP superfamily (Brangulis *et al.*, 2015), a *trans* peptide is also found but this time the main chain carbonyl of the peptide acts as the hydrogen bond acceptor in yet another variant of this interaction maintaining the binding site of the guanidyl group [Figure 7.8].



**Figure 7.6: the environment of the critical arginine in the CAP fold of SAG-B41.** A) A cartoon of the structure of SAG-B41. The arginine is located at the C-terminal end of  $\alpha I$  with its guanidiny group positioned at the C-terminal end of  $\alpha V$  where it can be stabilised by the negative end of helix dipole. The nearby cic peptide Ile225-Asp226 is important in stabilising the conformation of the main chain of residues in loop L9, which in turn form hydrogen bonds (dotted green lines) to the guanidiny group. B) A diagram to show the key hydrogen bonds to the guanidiny group from residues in L3,  $\alpha V$  and L9 including the cis peptide with their interaction residues. The hydrogen bonds from the cis peptide carbonyl that interact with residues from L9 is also shown. Key residues and elements of secondary structure are labelled. Other side chain are omitted for clarity.



*Figure 7.7: **Key interactions creating the pocket for the conserved buried arginine in SAG-B41.** The hydrogen bonds (dotted green lines) between the main chain carbonyl of Ile225 that forms part of the Ile225-Asp226 cis-peptide (non-Pro cis) in SAG-B41 and the main chain NHs of Lys203 and Lys204 that lie in the loop L9 next to the guanidiny group of the buried arginine, Arg75.*



*Figure 7.8: Interactions creating the pocket for the conserved buried arginine in the structure of the bacterial protein, BB0689. The hydrogen bonds (dotted green lines) between the main chain carbonyl of the trans peptide between gly133 and Lys 134 and the main chain NHs of Asp112 and Lys113 that lie at the boundary of a loop and the following strand (equivalent to loop L9 and strand C in SAGB-41). These interactions lead to the correct positioning of the two carbonyl groups of Asp 112 and Lys 113 that form the wall of the pocket for the guanidiny group of the buried arginine, Arg25.*

Table 7.3: The summary of SAG proteins that were part of this study.

SAG ID	Soluble Expression	Purification	Crystallisation	Data collection	Structure determined (Å)
SAG-A1	Yes	Yes	No		
SAG-A6	Yes	Yes	No		
SAG-A7	Yes	Yes	No		
SAG-A10	Yes	Yes	Yes*		
SAG-A31	Yes	Yes	Yes*		
SAG-A91	Yes	Yes	Yes	Yes	Yes (1.40 Å)
SAG-A870	Yes	Yes	No		
SAG-A905	Yes	Yes	No		
SAG-B13	Yes	Yes	Yes	Yes	Yes (1.75 Å)
SAG-B16	Yes	Yes	Yes	Yes	Yes (1.67 Å)
SAG-B22	Yes	Yes	Yes	Yes	Yes (2.0 Å)
SAG-B41	Yes	Yes	Yes	Yes	Yes (1.18 Å)
SAG-C75	Yes	Yes	Yes	Yes	Yes (1.32 Å)
EBC1	Yes	Yes	Yes*		
EBC2	Yes	Yes	No		
EBC3	Yes	Yes	No		
CAP-like SAGs from other Apicomplexans					
<i>P. vivax</i>	Yes	Yes	Yes	Yes	Yes (1.50 Å)
<i>P. falciparum</i>	Yes				
<i>P. ovale</i>	Yes	Yes			
<i>P. malariae</i>	Yes	Yes			
<i>B. bovis</i>	Yes				
<i>B. besnoiti</i>	Yes				

\* Preliminary crystal obtained but further optimization needed.

Table 7.4: Key structural features found in the *Eimeria tenella* SAGs compared to other proteins of the CAP superfamily.

CAP-family member	Conservation of SS-bridges compared to <i>E. tenella</i> SAGs*			Conservation of the NxxR motif	Conserved <i>cis</i> peptide
	SS1	SS2	SS3		
SAG-A91	Yes	#	Yes	Yes	Ile225-Asp226
SAG-B13	Yes	Yes	Yes	Yes	Ile216-Gln217
SAG-B16	Yes	Yes	Yes	Yes	Val222-Leu223
SAG-B22	Yes	Yes	Yes	Yes	Ile223-Glu224
Sag-B19	Yes	Yes	Yes	Yes	Ile225-Asp226
SAG-B41	Yes	Yes	Yes	Yes	Ile225-Asp226
SAG-C75	No	Yes	Yes	Yes	Asn175-Pro176
<i>P. vivax</i> SAG <sup>§</sup>	No	#	Yes	Yes	No**
Other CAP-like SAGs					
3U3U <sup>§</sup>	No	No	Yes	Yes	No***
1U53 <sup>§</sup>	No	#	Yes	Yes	Gly159-Pro160
1QNX <sup>§</sup>	No	No	Yes	Yes	Gly190-Pro191
1SMB <sup>§</sup>	No	No	No	Yes	Phe134-Pro135

# SS-bridge in a similar but not identical position.

\*SS1, SS2 and SS3 numbers refer to SAG-B19

<sup>§</sup> *P. vivax* SAG, 3U3U, 1U53 and 1QNX also contain an additional 1, 4, 3 and 3 non-conserved disulphides, respectively. 1SMB has no disulphide.

\*\* the structurally equivalent peptide, Asp159-Ser160, is in a *trans* configuration

\*\*\* the structurally equivalent peptide, Ser189-Ser190, is in a *trans* configuration

Table 7.5: Disruption of SAGs in selected apicomplexan parasites

Parasite	Hematozoa Group	Coccidia Group	<i>Eimeria</i> -like Coccidia Group	SRS-related SAGs	Number of CAP-like SAGs	Disease	Human/Animals
<i>Plasmodium vivax</i>	√			√	1	Malariae	Human
<i>Plasmodium malariae</i>	√			√	1	Malariae	Human
<i>Plasmodium ovale</i>	√			√	1	Malariae	Human
<i>Plasmodium falciparum</i>	√			√	1	Malariae	Human
<i>Plasmodium knowlesi</i>	√			√	1	Malariae	Human
<i>Babesia bovis</i>	√			√	1	Babesiosis	Animals
<i>Theileria orientalis</i>	√			√	1	Oriental theileriosis	Animals
<i>Besnoitia besnoiti</i>		√		√	9 <sup>a</sup>	Bovine besnoitiosis	Animals
<i>Neospora caninum</i>		√		√	5 <sup>b</sup>	Neosporosis	Animals
<i>Cystoisospora suis</i>		√		√	5 <sup>c</sup>	Porcine neonatal coccidiosis	Animals
<i>Toxoplasma gondii</i>		√		√	5 <sup>d</sup>	Toxoplasmosis	Human/Animals
<i>Cyclospora cayetanensis</i>		√	√	X	7 <sup>e</sup>	Cyclosporiasis	Human
<i>Eimeria tenella</i>		√	√	X	89	Coccidiosis	Animals

<sup>a</sup> the nine identified genes are (BESB\_017040, BESB\_017030, BESB\_017010, BESB\_017020, BESB\_039070, BESB\_021660, BESB\_047780, BESB\_036050, BESB\_009920)

<sup>b</sup> the five identified genes are (NCLIV\_012410, NCLIV\_000710, NCLIV\_051180, NCLIV\_051190, NCLIV\_040740)

<sup>c</sup> the five identified genes are (CSUI\_002672, CSUI\_001518, CSUI\_003904, CSUI\_003487, CSUI\_001885)

<sup>d</sup> the five identified genes are (TGGT1\_288220, TGGT1\_237450, TGGT1\_220280, TGGT1\_239890, TGGT1\_237425)

<sup>e</sup> the seven identified genes are (CYC\_03341, CYC\_08792, CYC\_03830, CYC\_04187, CYC\_01094, CYC\_00713, CYC\_06337)



#### **7.4 Discovery of ETH-00002020 is the closest homologue to CAP-like SAGs (*Babesia bovis*, *Theileria*, *Besnoitia*, *Neospora*, *Cystoisospora*, *Toxoplasma* and *Cyclospora*).**

Having discovered CAP-like SAGs in the genomes of a number of apicomplexan parasites that also possess SRS-like SAGs [Table 7.5] a BLAST search was undertaken to discover their closest homologues in *E. tenella*. The sequence of a representative CAP-like SAG from each organism identified in this study was searched against the *E. tenella* proteome. Interestingly, for each BLAST search undertaken the *E. tenella* ETH-00002020 gene product was found to be the closest SAG homologue for each representative CAP-like SAG of the different search species. Although the ETH-00002020 gene was not included in the 89 members of the *E. tenella* SAG family (Reid *et al.* 2014), all the sequence identifiers of the CAP-like SAG are present in its sequence such as NxxR and the three disulphide bonds. Furthermore, the ETH-00002020 gene has five exons which place it in the SAG B family. Further analysis was carried out to investigate which chromosome this SAG comes from. An inspection of this analysis revealed that the ETH\_00002020 and the single type C SAG, SAG-C75, appear on the same DNA genome scaffold fragment Eth\_scaff106 indicating that they lie on the same chromosome although which chromosome this is has not been determined.

Comparing the sequence of *E. tenella* ETH-00002020 against the sequences of the *E. tenella* SAGs in this study, showed pairwise sequence identities between 13-22% [Table 7.6] a similarity very much in line with that for any of the *E. tenella* SAGs. For example, the identity between the SAG-A91 and SAG-B19 is 17 % and between the SAG-A91 and SAG-C75 is 15 % [chapter 5, Table 5.7]. Further analysis of the genomes of different *Eimeria* species showed that ETH-00002020 has 92% identity to ENH\_00005250 SAG from *Eimeria necatrix*. Recent studies on this protein have shown that *E. necatrix* ENH\_00005250 shows potential to be used as a vaccine against *Eimeria* infections in chickens (Wang *et al.*, 2023). Thus, immunization with ENH\_00005250 SAG conferred some protection, as evidenced by reduced intestinal pathology, decreased oocyst shedding, and minimized weight loss following coccidial exposure. It would be interesting to see if a similar protective effect could be observed with *E. tenella* ETH-00002020. Indeed, given that these two proteins are the most similar to the CAP-like SAGs found in *Plasmodium* and other SRS-SAG containing apicomplexan parasites, it would be interesting to see if a protective immune response could also be obtained by inoculating with these specific proteins.

Table 7.6: The identity (%) between ETH\_00002020 and the SAGs studied as part of this thesis.

SAGs	(%) ETH 00002020	SAGs	(%) ETH 00002020	SAGs	(%) ETH 00002020	SAGs	(%) ETH 00002020
SAG-A1	19	SAG-B13	18	SAG-C75	20	<i>P. vivax</i>	19
SAG-A6	18	SAG-B16	20	EBC1	15	<i>B. bovis</i>	14
SAG-A7	18	SAG-B19	22	EBC2	21	<i>B. besnoiti</i>	35
SAG-A10	19	SAG-B22	21	EBC3	19	<i>T. orientalis</i>	12
SAG-A31	13	SAG-B41	22			<i>T. gondii</i>	23
SAG-A91	16					<i>N. caninum</i>	17
SAG-A870	15					<i>C. suis</i>	20
SAG-A905	19					<i>C. cayetanensis</i>	42

## 7.5 The Function of Apicomplexan CAP-like SAGs

The structural work described in this thesis has clearly shown that the family of CAP-like SAGs share the same basic fold, yet display entirely different surface properties to each other. This suggests that the family of proteins do not have a common binding partner on the host cells. Nevertheless, studies elsewhere have identified that some members of the SAG A family have the ability to bind to cultured cells (Jahn *et al.*, 2009), suggesting a potential role for the SAG A family specifically in mediating attachment to host cells. Other studies have indicated that some of the SAGs can bind small ligands, for example *Eimeria tenella* SAG1. However, it must be remembered that each of the SAGs contain crevices on their surface that might, by chance, provide a region that would have chemical properties suitable for binding a complementary ligand. If it is assumed that each CAP-like SAG domain contains 10 such crevices (estimated as being due to non-overlapping sites between different elements of secondary structure), each with different chemical character, then combined with the diversity in the surface properties, a family of 100+ SAGs would give rise to 1000+ regions of different chemical properties. It would therefore not be surprising if at least some of these provided suitable binding sites for some of the ligands found in the cell. Therefore, any binding activity detected *in vitro* might not represent a biologically relevant interaction further confusing such studies.

One hypothesis arising from this work is that the majority of the CAP-like SAG proteins in *Eimeria* are involved in providing the parasite with means to evade the immune response of the host, by presenting a wide variety of temporally different surfaces to the host organism during infection. One question that then arises is whether any of the *Eimeria* SAGs play a different role. One way to gain information on possible functional role(s) of these SAG

proteins is by analysing transcriptome data assembled from RNA-seq experiments on these different apicomplexan parasites [Table 7.5]. A preliminary investigation of such transcriptome data in various stages of parasite development has shown that, in some of the studies, some of the CAP-like SAG genes were found to be differentially expressed under various conditions (Buchholz *et al.*, 2011; Horcajo *et al.*, 2017; Ueti *et al.*, 2020; Cruz-Bustos *et al.*, 2022; Garfoot *et al.*, 2019; Ramaprasad *et al.*, 2015; Ramakrishnan *et al.*, 2022). However, no clear consistent pattern of differential expression has yet emerged. For example in a study comparing gene expression in *Besnoitia* strain Lisbon14, one of the nine CAP-like SAGs in this organism (BESB\_039070) was upregulated in tissue cysts by approximately 45 fold compared to the level expressed in tachyzoites (Ramakrishnan *et al.*, 2022). Similarly, in a study on the transcriptome of *Cystoisospora suis* tracking gene expression from asexual to sexual development *in vitro* at three different time points (T1, merozoites; T2, merozoites and immature sexual stages and T3, mature sexual stage and oocysts), 891 and 1860 differentially expressed genes (> 2 fold) at T2 and T3 compared to T1 and 823 genes at T3 compared to T2 were identified (Cruz-Bustos *et al.*, 2022). Amongst these identified genes, one of the CAP-like SAGs (CSUI\_003904) was down regulated approximately seven-fold in T3 compared to T1, indicating that this particular SAG might not be involved in the sexual development of *C. suis*. However, the other four CAP-like SAGs of *C. suis* do not feature in this list of differentially expressed genes, and there is therefore no consistent picture for the function of this family of CAP-like SAGs.

For *Babesia bovis*, a study comparing levels of gene expression in asexual and sexual *B. bovis* life stages including bovine erythrocytes (asexual) and tic haemolymph (sexual), the single CAP-like SAG of *B. bovis* was found to be upregulated by approximately three-fold in the tick stage of the life cycle compared to the blood stage (Ueti *et al.*, 2020). Although the single CAP-like SAG of *Theileria orientalis* was identified as being expressed in both tic salivary glands (sporozoite stage) and infected cattle blood (piroplasm stage) the expression levels were not significantly different in the two stages (Hayashida *et al.*, 2018).

The single CAP-like SAG gene in *P. vivax* and *P. falciparum* has been highlighted in a number of transcriptome studies in the PlasmoDB (<https://plasmodb.org/plasmo/app>). For example, in a study tracking the intraerythrocyte developmental stage of the parasite (asexual

multiplication) over 48 hours from samples collected from two patients, the expression level of the *P. vivax* CAP-like SAG gene (PVP01\_0106000) increased between 2- and 4-fold at 48h compared to that at the start (Zhu *et al.*, 2016). Similarly, in a study analysing the transcriptomes of *P. vivax* infected mosquito salivary gland sporozoites (Muller *et al.*, 2019), the level of gene expression of this protein varied by up to 8-fold in the different clinical isolates. Comparison of the gene expression levels during cultivation of clinical *P. vivax* isolates in differing culture media showed that the *P. vivax* CAP-like SAG was upregulated by up to 8-fold at 36h compared to initial levels (Rangel *et al.*, 2020). In similar studies by (Gural *et al.*, 2018; Bozdech *et al.*, 2008; Roth *et al.*, 2018), no significant variations in gene expression for *P. vivax* (PVP01\_0106000) could be identified.

An analysis of the PlasmoDB data base for the *P. falciparum* homologue of the *P. vivax* CAP-like SAG (PF3D7\_0705800) again identified a number of conditions during infectious stages where the gene was upregulated. For example, a 2-fold upregulation was observed for oocyst sporozoites compared to salivary gland sporozoites (Lasonder *et al.*, 2016). Other studies on gametocyte maturation showed a 4-fold upregulated gene expression level for the *P. falciparum* CAP-like SAG gene at 10 days maturation, compared to the initial level (Young *et al.*, 2005). In addition, an up to 4-fold increase in *P. falciparum* CAP-like SAG occurred during a ten day *in vitro* maturation of gametocytes (Van Biljon *et al.*, 2019). Interestingly a study comparing the transcriptomes of male and female gametocytes in *P. falciparum* showed that the CAP-like SAG gene was expressed >16-fold more in female gametocytes compared to male gametocytes (Lasonder *et al.*, 2016). Further studies on the life cycle of *P. falciparum* showed that there was a >100-fold upregulation of this gene in gametocytes compared to trophozoites or rings (López-Barragán *et al.*, 2011), with a similar 16-fold upregulation in oocysts compared to sporozoites or rings also observed (Zhang *et al.*, 2021), with the *P. falciparum* data suggesting some potential role for the CAP-like SAG in the sexual development stages of the parasite.

For *Toxoplasma gondii*, a study was carried out comparing the expression levels between acute-stage tachyzoites and chronic-stage bradyzoites that form intracellular cysts resistant to immune clearance and existing therapies. This study showed that three of the five the CAP-like SAGs in this organism, TGME49\_288220, TGME49\_237450 and TGME49\_220280 were upregulated in bradyzoites compared to tachyzoites but to different extents (22-fold, 2-fold and 10-fold, respectively) (Waldman *et al.*, 2020). However, the expression levels of the remaining CAP-like SAGs, TGGT1\_239890 and TGGT1\_237425, were broadly similar. These data

suggest that some of these CAP-like SAG in *Toxoplasma gondii* are not involved in the development of this part of the life cycle of this parasite.

## 7.6 Concluding remarks

The determination of the structures of the *Eimeria* SAGs presented in this thesis has clearly shown that they are all based on a CAP-like fold irrespective of the subfamily to which they belong. Moreover, the analysis of the patterns of sequence and structural conservation show that the majority of the similarities between them are associated with largely buried residues leaving their exposed surfaces to be chemically very diverse. These data support the view that the function of the majority of this superfamily of proteins is not to bind a specific ligand but rather point to a role related to evading the immune system or could there be value in goading the immune system? Either misdirection or benefiting from the consequences of inflammation. Such a role could then be analogous to the function of the SRS-like superfamily of surface proteins found in *Toxoplasma* or *Plasmodium* parasites. However, even if the role of the majority of the *Eimeria* SAGs is to act in this way, this does not mean that some of them might have other ligand binding properties even if such properties have arisen initially by chance. Indeed, were this to be the case the ability of one or more specific SAGs to function in the binding of a ligand might well have become hard-wired into the organism in such a way that this binding might have a biological role, for example in some form of regulation. Clearly, currently it is still not possible to define a function for these SAGs and further experiments are essential.

Whilst the summary presented above is restricted to the *Eimeria* SAGs the discovery of a single CAP-like SAG in *Plasmodium* species and of a small number of CAP-like SAGs in other apicomplexans raises yet more interesting questions. For example, why have these other apicomplexans expanded the proteins with an SRS-like fold and left only a small number of CAP-like SAGs? What is the significance of the fact that in *Plasmodium* and other parasites belonging to the hematazoan group including *Babesia* and *Theileria* there is only a single CAP-like SAG? Is the function of this protein similar to that in *Eimeria* or not? Is there a possibility that some of these apicomplexan SAGs might form suitable targets for the development of a therapeutically useful vaccine? Clearly many questions remain as to the role and possible therapeutic exploitation of this intriguing family of proteins. However, their widespread distribution in apicomplexans suggest that they play an important role in their biology and as such merit further investigation.

## 8. References

- Abrahamsen, M.S. *et al.* (2004) 'Complete Genome Sequence of the Apicomplexan, *Cryptosporidium parvum*', *Science*, 304(5669), pp. 441–445. Available at: <https://doi.org/10.1126/science.1094786>.
- Ahouidi, A.D. *et al.* (2010) 'Population genetic analysis of large sequence polymorphisms in *Plasmodium falciparum* blood-stage antigens', *Infection, Genetics and Evolution*, 10(2), pp. 200–206. Available at: <https://doi.org/10.1016/j.meegid.2009.11.008>.
- Aikawa, M. *et al.* (1978) 'Erythrocyte entry by malarial parasites. A moving junction between erythrocyte and parasite', *Journal of Cell Biology*, 77(1), pp. 72–82. Available at: <https://doi.org/10.1083/jcb.77.1.72>.
- Ajioka, J.W., Fitzpatrick, J.M. and Reitter, C.P. (2001) 'Toxoplasma gondii genomics: Shedding light on pathogenesis and chemotherapy', *Expert Reviews in Molecular Medicine*, 3(1), pp. 1–19. Available at: <https://doi.org/10.1017/S1462399401002204>.
- Arredondo, S.A. and Kappe, S.H.I. (2017) 'The s48/45 six-cysteine proteins: mediators of interaction throughout the *Plasmodium* life cycle', *International Journal for Parasitology*. Elsevier Ltd, pp. 409–423. Available at: <https://doi.org/10.1016/j.ijpara.2016.10.002>.
- Asojo, O.A. *et al.* (2005) 'X-ray structure of Na-ASP-2, a pathogenesis-related-1 protein from the nematode parasite, *Necator americanus*, and a vaccine antigen for human hookworm infection', *Journal of Molecular Biology*, 346(3), pp. 801–814. Available at: <https://doi.org/10.1016/j.jmb.2004.12.023>.
- Aurrecoechea, C. *et al.* (2009) 'PlasmoDB: A functional genomic database for malaria parasites', *Nucleic Acids Research*, 37(SUPPL. 1), pp. 539–543. Available at: <https://doi.org/10.1093/nar/gkn814>.
- Van Biljon, R. *et al.* (2019) 'Hierarchical transcriptional control regulates *Plasmodium falciparum* sexual differentiation', *BMC Genomics*, 20(1), pp. 1–16. Available at: <https://doi.org/10.1186/s12864-019-6322-9>.
- Blake, D.P. *et al.* (2011) 'Genetic mapping identifies novel highly protective antigens for an apicomplexan parasite', *PLoS Pathogens*, 7(2). Available at: <https://doi.org/10.1371/journal.ppat.1001279>.
- Blake, D.P., Worthing, K. and Jenkins, M.C. (2020) 'Exploring *Eimeria* genomes to understand population biology: Recent progress and future opportunities', *Genes*. MDPI AG, pp. 1–14. Available at: <https://doi.org/10.3390/genes11091103>.
- Britez, J.D. *et al.* (2023) 'What Do We Know about Surface Proteins of Chicken Parasites *Eimeria*?', *Life*, 13(6), p. 1295. Available at: <https://doi.org/10.3390/life13061295>.
- Burrell, A. *et al.* (2020) 'Life cycle stages, specific organelles and invasion mechanisms of *Eimeria* species', *Parasitology*, 147(3), pp. 263–278. Available at: <https://doi.org/10.1017/S0031182019001562>.
- Burrell, A. *et al.* (2022) 'Cellular electron tomography of the apical complex in the apicomplexan parasite *Eimeria tenella* shows a highly organised gateway for regulated secretion', *PLoS Pathogens*, 18(7 July), pp. 1–16. Available at: <https://doi.org/10.1371/journal.ppat.1010666>.

- Bozdech, Z. *et al.* (2008) 'The transcriptome of *Plasmodium vivax* reveals divergence and diversity of transcriptional regulation in malaria parasites', *Proceedings of the National Academy of Sciences of the United States of America*, 105(42), pp. 16290–16295. Available at: <https://doi.org/10.1073/pnas.0807404105>.
- Bradford, M.M. (1976) *A Rapid and Sensitive Method for the Quantitation of Microgram Quantities of Protein Utilizing the Principle of Protein-Dye Binding*, *ANALYTICAL BIOCHEMISTRY*.
- Bradley, P.J. and Sibley, L.D. (2007) 'Rhoptries: an arsenal of secreted virulence factors', *Current Opinion in Microbiology*, 10(6), pp. 582–587. Available at: <https://doi.org/10.1016/j.mib.2007.09.013>.
- Brangulis, K. *et al.* (2015) 'Structural and functional analysis of BB0689 from *Borrelia burgdorferi*, a member of the bacterial CAP superfamily', *Journal of Structural Biology*, 192(3), pp. 320–330. Available at: <https://doi.org/10.1016/j.jsb.2015.09.007>.
- Brinkmann, V., Remington, J.S. and Sharma, S.D. (1993) 'Vaccination of mice with the protective F3G3 antigen of *Toxoplasma gondii* activates Cd4<sup>+</sup> but not Cd8<sup>+</sup> T cells and induces toxoplasma specific IgG antibody', *Molecular Immunology*, 30(4), pp. 353–358. Available at: [https://doi.org/10.1016/0161-5890\(93\)90064-I](https://doi.org/10.1016/0161-5890(93)90064-I).
- Britez, J.D. *et al.* (2023) 'What Do We Know about Surface Proteins of Chicken Parasites *Eimeria*?', *Life*, 13(6), p. 1295. Available at: <https://doi.org/10.3390/life13061295>.
- Buchholz, K.R. *et al.* (2011) 'Identification of tissue cyst wall components by transcriptome analysis of in vivo and in vitro *Toxoplasma gondii* bradyzoites', *Eukaryotic Cell*, 10(12), pp. 1637–1647. Available at: <https://doi.org/10.1128/EC.05182-11>.
- Burrell, A. *et al.* (2020) 'Life cycle stages, specific organelles and invasion mechanisms of *Eimeria* species', *Parasitology*, 147(3), pp. 263–278. Available at: <https://doi.org/10.1017/S0031182019001562>.
- Burrell, A. *et al.* (2022) 'Cellular electron tomography of the apical complex in the apicomplexan parasite *Eimeria tenella* shows a highly organised gateway for regulated secretion', *PLoS Pathogens*, 18(7 July), pp. 1–16. Available at: <https://doi.org/10.1371/journal.ppat.1010666>.
- Cai, X. *et al.* (2003) 'Apicoplast genome of the coccidian *Eimeria tenella*', *Gene*, 321(1–2), pp. 39–46. Available at: <https://doi.org/10.1016/j.gene.2003.08.008>.
- Carruthers, V. and Boothroyd, J.C. (2007) 'Pulling together: an integrated model of *Toxoplasma* cell invasion', *Current Opinion in Microbiology*, 10(1), pp. 83–89. Available at: <https://doi.org/10.1016/j.mib.2006.06.017>.
- Carruthers, V.B. and Tomley, F.M. (2008) 'Microneme proteins in apicomplexans', *Subcellular Biochemistry*, 47, pp. 33–45. Available at: [https://doi.org/10.1007/978-0-387-78267-6\\_2](https://doi.org/10.1007/978-0-387-78267-6_2).
- Cavalier-Smith, T. (1993) *Kingdom protozoa and its 18 phyla*, *Microbiological Reviews*. Available at: <https://doi.org/10.1128/mmbr.57.4.953-994.1993>.
- Cerutti, A., Blanchard, N. and Besteiro, S. (2020) 'The bradyzoite: A key developmental stage for the persistence and pathogenesis of toxoplasmosis', *Pathogens*, 9(3), pp. 1–21. Available at: <https://doi.org/10.3390/pathogens9030234>.

Ben Chaabene, R., Lentini, G. and Soldati-Favre, D. (2021) 'Biogenesis and discharge of the rhoptries: Key organelles for entry and hijack of host cells by the Apicomplexa', *Molecular Microbiology*, 115(3), pp. 453–465. Available at: <https://doi.org/10.1111/mmi.14674>.

Chapman, H.D. (1998) 'Evaluation of the efficacy of anticoccidial drugs against *Eimeria* species in the fowl', *International Journal for Parasitology*, 28(7), pp. 1141–1144. Available at: [https://doi.org/10.1016/S0020-7519\(98\)00024-1](https://doi.org/10.1016/S0020-7519(98)00024-1).

Chou, P.Y. and Pasman, G.D. (1974) 'for Amino', 13(2), pp. 211–222.

Chow, Y.-P. *et al.* (2011) 'Immunogenic *Eimeria tenella* glycosylphosphatidylinositol-anchored surface antigens (SAGs) induce inflammatory responses in avian macrophages', *PLoS ONE*. Edited by E.M. Braga, 6(9), p. e25233. Available at: <https://doi.org/10.1371/journal.pone.0025233>.

Crawford, J. *et al.* (2009) 'Structural characterization of the bradyzoite surface antigen (BSR4) from *Toxoplasma gondii*, a unique addition to the surface antigen glycoprotein 1-related superfamily', *Journal of Biological Chemistry*, 284(14), pp. 9192–9198. Available at: <https://doi.org/10.1074/jbc.M808714200>.

Crawford, J. *et al.* (2010) 'Structural and functional characterization of SporoSAG: A SAG2-related surface antigen from *Toxoplasma gondii*', *Journal of Biological Chemistry*, 285(16), pp. 12063–12070. Available at: <https://doi.org/10.1074/jbc.M109.054866>.

Cruz-Bustos, T. *et al.* (2022) 'The transcriptome from asexual to sexual in vitro development of *Cystoisospora suis* (Apicomplexa: Coccidia)', *Scientific Reports*, 12(1), pp. 1–17. Available at: <https://doi.org/10.1038/s41598-022-09714-8>.

Dalloul, R.A. *et al.* (2007) 'Unique responses of the avian macrophage to different species of *Eimeria*', *Molecular Immunology*, 44(4), pp. 558–566. Available at: <https://doi.org/10.1016/j.molimm.2006.02.004>.

Debierre-Grockiego, F. *et al.* (2007) 'Activation of TLR2 and TLR4 by Glycosylphosphatidylinositols Derived from *Toxoplasma gondii*', *The Journal of Immunology*, 179(2), pp. 1129–1137. Available at: <https://doi.org/10.4049/jimmunol.179.2.1129>.

Delbecq, S. (2022) 'Major Surface Antigens in Zoonotic *Babesia*', *Pathogens*, 11(1). Available at: <https://doi.org/10.3390/pathogens11010099>.

Dietrich, M.H. *et al.* (2022) 'Nanobodies against Pfs230 block *Plasmodium falciparum* transmission', *The Biochemical journal*, 479(24), pp. 2529–2546. Available at: <https://doi.org/10.1042/BCJ20220554>.

Doolan, D.L. (2011) 'Plasmodium immunomics', *International Journal for Parasitology*, 41(1), pp. 3–20. Available at: <https://doi.org/10.1016/j.ijpara.2010.08.002>.

Döşkaya, M. *et al.* (2014) 'Diagnostic value of a Rec-ELISA Using *Toxoplasma gondii* recombinant SporoSAG, BAG1, and GRA1 proteins in murine models infected orally with tissue cysts and oocysts', *PLoS ONE*, 9(9), pp. 1–7. Available at: <https://doi.org/10.1371/journal.pone.0108329>.

Dowse, T.J. *et al.* (2005) 'Apicomplexan rhomboids have a potential role in microneme protein cleavage during host cell invasion', *International Journal for Parasitology*, 35(7), pp. 747–756. Available at: <https://doi.org/10.1016/j.ijpara.2005.04.001>.



- Dubey, J.P. (2019) *Coccidiosis in Livestock, Poultry, Companion Animals, and Humans*, *Coccidiosis in Livestock, Poultry, Companion Animals, and Humans*. Available at: <https://doi.org/10.1201/9780429294105>.
- Dubremetz, J.F. *et al.* (1998) 'Apical organelles and host-cell invasion by Apicomplexa', *International Journal for Parasitology*, 28(7), pp. 1007–1013. Available at: [https://doi.org/10.1016/S0020-7519\(98\)00076-9](https://doi.org/10.1016/S0020-7519(98)00076-9).
- Emsley, P. and Cowtan, K. (2004) 'Coot: Model-building tools for molecular graphics', *Acta Crystallographica Section D: Biological Crystallography*, 60(12 I), pp. 2126–2132. Available at: <https://doi.org/10.1107/S0907444904019158>.
- Evans, P. (2006) 'Scaling and assessment of data quality', *Acta Crystallographica Section D: Biological Crystallography*, 62(1), pp. 72–82. Available at: <https://doi.org/10.1107/S0907444905036693>.
- Evans, P. and McCoy, A. (2007) 'An introduction to molecular replacement', *Acta Crystallographica Section D: Biological Crystallography*, 64(1), pp. 1–10. Available at: <https://doi.org/10.1107/S0907444907051554>.
- Frénal, K. *et al.* (2017) 'Gliding motility powers invasion and egress in Apicomplexa', *Nature Reviews Microbiology*, 15(11), pp. 645–660. Available at: <https://doi.org/10.1038/nrmicro.2017.86>.
- Frénal, K. and Soldati-Favre, D. (2009) 'Role of the Parasite and Host Cytoskeleton in Apicomplexa Parasitism', *Cell Host and Microbe*, pp. 602–611. Available at: <https://doi.org/10.1016/j.chom.2009.05.013>.
- Garfoot, A.L. *et al.* (2019) 'Proteomic and transcriptomic analyses of early and late-chronic *Toxoplasma gondii* infection shows novel and stage specific transcripts', *BMC Genomics*, 20(1), pp. 1–11. Available at: <https://doi.org/10.1186/s12864-019-6213-0>.
- Ghosh, A., Edwards, M.J. and Jacobs-Lorena, M. (2000) 'The journey of the malaria parasite in the mosquito: Hopes for the new century', *Parasitology Today*, 16(5), pp. 196–201. Available at: [https://doi.org/10.1016/S0169-4758\(99\)01626-9](https://doi.org/10.1016/S0169-4758(99)01626-9).
- Gibbs, G.M., Roelants, K. and O'Bryan, M.K. (2008) 'The CAP superfamily: Cysteine-rich secretory proteins, antigen 5, and pathogenesis-related 1 proteins - Roles in reproduction, cancer, and immune defense', *Endocrine Reviews*, 29(7), pp. 865–897. Available at: <https://doi.org/10.1210/er.2008-0032>.
- Gilson, P.R. and Crabb, B.S. (2009) 'Morphology and kinetics of the three distinct phases of red blood cell invasion by *Plasmodium falciparum* merozoites', *International Journal for Parasitology*, 39(1), pp. 91–96. Available at: <https://doi.org/10.1016/j.ijpara.2008.09.007>.
- Giovannini, D. *et al.* (2011) 'Independent roles of apical membrane antigen 1 and rhoptry neck proteins during host cell invasion by apicomplexa', *Cell Host and Microbe*, 10(6), pp. 591–602. Available at: <https://doi.org/10.1016/j.chom.2011.10.012>.
- Gowda, D.C. (2007) 'TLR-mediated cell signaling by malaria GPIs', *Trends in Parasitology*, 23(12), pp. 596–604. Available at: <https://doi.org/10.1016/j.pt.2007.09.003>.
- Grigg, M.E. and Sundar, N. (2009) 'Sexual recombination punctuated by outbreaks and clonal expansions predicts *Toxoplasma gondii* population genetics', *International Journal for Parasitology*, 39(8), pp. 925–933. Available at: <https://doi.org/10.1016/j.ijpara.2009.02.005>.

- Grimwood, J. and Smith, J.E. (1996) 'Toxoplasma gondii: The role of parasite surface and secreted proteins in host cell invasion', *International Journal for Parasitology*, 26(2), pp. 169–173. Available at: [https://doi.org/10.1016/0020-7519\(95\)00103-4](https://doi.org/10.1016/0020-7519(95)00103-4).
- Gubbels, M.J. *et al.* (2020) 'Fussing About Fission: Defining Variety Among Mainstream and Exotic Apicomplexan Cell Division Modes', *Frontiers in Cellular and Infection Microbiology*, 10(June). Available at: <https://doi.org/10.3389/fcimb.2020.00269>.
- Gural, N. *et al.* (2018) 'In vitro culture, drug sensitivity, and transcriptome of Plasmodium vivax hypnozoites', *HHS Public Access*, 23(3), pp. 395–406. Available at: <https://doi.org/10.1016/j.chom.2018.01.002>.
- Guttery, D.S. *et al.* (2015) 'Commit and Transmit: Molecular Players in Plasmodium Sexual Development and Zygote Differentiation', *Trends in Parasitology*, 31(12), pp. 676–685. Available at: <https://doi.org/10.1016/j.pt.2015.08.002>.
- Harding, C.R. and Frischknecht, F. (2020) 'The Riveting Cellular Structures of Apicomplexan Parasites', *Trends in Parasitology*, 36(12), pp. 979–991. Available at: <https://doi.org/10.1016/j.pt.2020.09.001>.
- Hayashida, K. *et al.* (2018) 'Establishment of a mouse-tick infection model for Theileria orientalis and analysis of its transcriptome', *International Journal for Parasitology*, 48(12), pp. 915–924. Available at: <https://doi.org/10.1016/j.ijpara.2018.05.012>.
- He, X.L. *et al.* (2002) 'Structure of the immunodominant surface antigen from the toxoplasma gondii SRS superfamily', *Nature Structural Biology*, 9(8), pp. 606–611. Available at: <https://doi.org/10.1038/nsb819>.
- Heitlinger, E. *et al.* (2014) 'The genome of Eimeria falciformis - reduction and specialization in a single host apicomplexan parasite', *BMC Genomics*, 15(1), pp. 1–17. Available at: <https://doi.org/10.1186/1471-2164-15-696>.
- Henriksen, A. *et al.* (2001) 'Major venom allergen of yellow jackets, ves v 5: Structural characterization of a pathogenesis-related protein superfamily', *Proteins: Structure, Function and Genetics*, 45(4), pp. 438–448. Available at: <https://doi.org/10.1002/prot.1160>.
- Hol, W.G.J. (1985) 'The role of the  $\alpha$ -helix dipole in protein function and structure', *Progress in Biophysics and Molecular Biology*, 45(3), pp. 149–195. Available at: [https://doi.org/10.1016/0079-6107\(85\)90001-X](https://doi.org/10.1016/0079-6107(85)90001-X).
- Hol, W.I.M.G.J. (1985) 'Effects of the  $\alpha$ -helix dipole upon the functioning and structure of proteins and peptides Laboratory of Chemical Physics, Department of Chemistry, University of Groningen, Nijenborgh, Groningen, The Netherlands', 19, pp. 133–165.
- Horcajo, P. *et al.* (2017) 'Transcriptome modulation of bovine trophoblast cells in vitro by Neospora caninum', *International Journal for Parasitology*, 47(12), pp. 791–799. Available at: <https://doi.org/10.1016/j.ijpara.2017.08.007>.
- Jackson, A.P. *et al.* (2014) 'The evolutionary dynamics of variant antigen genes in Babesia reveal a history of genomic innovation underlying host-parasite interaction', *Nucleic Acids Research*, 42(11), pp. 7113–7131. Available at: <https://doi.org/10.1093/nar/gku322>.
- Jacquet, A. *et al.* (2001) 'The surface antigen SAG3 mediates the attachment of Toxoplasma gondii to cell-surface proteoglycans', *Molecular and Biochemical Parasitology*, 116(1), pp. 35–44. Available at: [https://doi.org/10.1016/S0166-6851\(01\)00297-3](https://doi.org/10.1016/S0166-6851(01)00297-3).

- Jahn, D. *et al.* (2009) 'Model structure of the immunodominant surface antigen of *Eimeria tenella* identified as a target for sporozoite-neutralizing monoclonal antibody', *Parasitology Research*, 105(3), pp. 655–668. Available at: <https://doi.org/10.1007/s00436-009-1437-6>.
- Jeanes, C. *et al.* (2013) 'Two new eimeria species parasitic in corncrakes (*Crex crex*) (Gruiformes: Rallidae) in the United Kingdom', *Journal of Parasitology*, 99(4), pp. 634–638. Available at: <https://doi.org/10.1645/12-52.1>.
- Jenkins, M.C. *et al.* (2019) 'Viable *Eimeria* oocysts in poultry house litter at the time of chick placement', *Poultry Science*, 98(8), pp. 3176–3180. Available at: <https://doi.org/10.3382/ps/pez147>.
- Jung, C., Lee, C.Y.F. and Grigg, M.E. (2004) 'The SRS superfamily of *Toxoplasma* surface proteins', *International Journal for Parasitology*, 34(3), pp. 285–296. Available at: <https://doi.org/10.1016/j.ijpara.2003.12.004>.
- Kabsch, W. (2010) 'Biological crystallography', *Acta Crystallographica Section D Biological Crystallography*, 49(1), pp. 1–1. Available at: <https://doi.org/10.1107/s090744499201028x>.
- Kafsack, B.F.C. *et al.* (2009) 'Rapid membrane disruption by a perforin-like protein facilitates parasite exit from host cells', *Science*, 323(5913), pp. 530–533. Available at: <https://doi.org/10.1126/science.1165740>.
- Karsten, V. *et al.* (1998) 'The protozoan parasite *Toxoplasma gondii* targets proteins to dense granules and the vacuolar space using both conserved and unusual mechanisms', *Journal of Cell Biology*, 141(6), pp. 1323–1333. Available at: <https://doi.org/10.1083/jcb.141.6.1323>.
- Kerlin, D.H. *et al.* (2012) 'An Analytical Method for Assessing Stage-Specific Drug Activity in *Plasmodium vivax* Malaria: Implications for Ex Vivo Drug Susceptibility Testing', *PLoS Neglected Tropical Diseases*, 6(8). Available at: <https://doi.org/10.1371/journal.pntd.0001772>.
- Khanaliha, K. *et al.* (2014) 'Evaluation of recombinant SAG1, SAG2, and SAG3 antigens for serodiagnosis of toxoplasmosis', *Korean Journal of Parasitology*, 52(2), pp. 137–142. Available at: <https://doi.org/10.3347/kjp.2014.52.2.137>.
- Kissinger, J.C.J.D. (2011) 'Genome cartography: charting the apicomplexan genome', *NIH Public Access*, 27(8), pp. 345–354. Available at: <https://doi.org/10.1016/j.pt.2011.03.006>.
- Koreny, L. *et al.* (2023) 'Stable endocytic structures navigate the complex pellicle of apicomplexan parasites', *Nature Communications*, 14(1). Available at: <https://doi.org/10.1038/s41467-023-37431-x>.
- Kovalevskiy, O. *et al.* (2018) 'Overview of refinement procedures within REFMAC 5: Utilizing data from different sources', *Acta Crystallographica Section D: Structural Biology*, 74, pp. 215–227. Available at: <https://doi.org/10.1107/S2059798318000979>.
- Krissinel, E. (2012) 'Enhanced fold recognition using efficient short fragment clustering', 1(2), pp. 76–85.
- Kuo, C.H. and Kissinger, J.C. (2008) 'Consistent and contrasting properties of lineage-specific genes in the apicomplexan parasites *Plasmodium* and *Theileria*', *BMC Evolutionary Biology*, 8(1), pp. 1–16. Available at: <https://doi.org/10.1186/1471-2148-8-108>.

- Lang, M. *et al.* (2009) 'Veterinary Parasitology Inhibition of host cell apoptosis by *Eimeria bovis* sporozoites', 160, pp. 25–33. Available at: <https://doi.org/10.1016/j.vetpar.2008.10.100>.
- Lasonder, E. *et al.* (2016) 'Integrated transcriptomic and proteomic analyses of *P. Falciparum* gametocytes: Molecular insight into sex-specific processes and translational repression', *Nucleic Acids Research*, 44(13), pp. 6087–6101. Available at: <https://doi.org/10.1093/nar/gkw536>.
- Lee, B. and Richards, F.M. (1971) 'The interpretation of protein structures: Estimation of static accessibility', *Journal of Molecular Biology*, 55(3). Available at: [https://doi.org/10.1016/0022-2836\(71\)90324-X](https://doi.org/10.1016/0022-2836(71)90324-X).
- Lekutis, C. *et al.* (2001) 'Surface antigens of *Toxoplasma gondii*: Variations on a theme', *International Journal for Parasitology*, 31(12), pp. 1285–1292. Available at: [https://doi.org/10.1016/S0020-7519\(01\)00261-2](https://doi.org/10.1016/S0020-7519(01)00261-2).
- Levine, N.D. (1986) 'The Taxonomy of Sarcocystis (Protozoa, Apicomplexa) Species Author (s): Norman D. Levine Published by: Allen Press on behalf of The American Society of Parasitologists Stable URL : <https://www.jstor.org/stable/3281676> THE TAXONOMY OF SARCOCYSTIS (', *The Journal of Parasitology*, 72(3), pp. 372–382. Available at: <https://www.jstor.org/stable/3281676>.
- Lillehoj, H.S. (1998) 'Role of T lymphocytes and cytokines in coccidiosis', *International Journal for Parasitology*, 28(7), pp. 1071–1081. Available at: [https://doi.org/10.1016/S0020-7519\(98\)00075-7](https://doi.org/10.1016/S0020-7519(98)00075-7).
- Lillehoj, H.S. (2016) 'Immune response to Coccidia', *IX International Coccidiosis Conference*, pp. 1–21.
- Lim, S.S. *et al.* (2014) 'Structure and dynamics of apical membrane antigen 1 from plasmodium falciparum FVO', *Biochemistry*, 53(46), pp. 7310–7320. Available at: <https://doi.org/10.1021/bi5012089>.
- Ling, K.H. *et al.* (2007) 'Sequencing and analysis of chromosome 1 of *Eimeria tenella* reveals a unique segmental organization', *Genome Research*, 17(3), pp. 311–319. Available at: <https://doi.org/10.1101/gr.5823007>.
- Liu, S. *et al.* (2016) 'Comparative genomics reveals *Cyclospora cayetanensis* possesses coccidia-like metabolism and invasion components but unique surface antigens', *BMC Genomics*, 17(1). Available at: <https://doi.org/10.1186/s12864-016-2632-3>.
- López-Barragán, M.J. *et al.* (2011) 'Directional gene expression and antisense transcripts in sexual and asexual stages of *Plasmodium falciparum*', *BMC Genomics*, 12(1), p. 587. Available at: <https://doi.org/10.1186/1471-2164-12-587>.
- Matsubayashi, M. *et al.* (2005) 'Molecular characterization of crane Coccidia, *Eimeria gruis* and *E. reichenowi*, found in feces of migratory cranes', *Parasitology Research*, 97(1), pp. 80–83. Available at: <https://doi.org/10.1007/s00436-005-1404-9>.
- Mauro, V.P. (2018) 'Codon Optimization in the Production of Recombinant Biotherapeutics: Potential Risks and Considerations', *BioDrugs*, 32(1), pp. 69–81. Available at: <https://doi.org/10.1007/s40259-018-0261-x>.
- Mayor, S. and Riezman, H. (2004) 'Sorting GPI-anchored proteins', *Nature Reviews Molecular Cell Biology*, 5(2), pp. 110–120. Available at: <https://doi.org/10.1038/nrm1309>.

- McCoy, A.J. *et al.* (2007) 'Phaser crystallographic software', *Journal of Applied Crystallography*, 40(4), pp. 658–674. Available at: <https://doi.org/10.1107/S0021889807021206>.
- Mcdonald, V., Shirley, M.W. and Millard, B.J. (1986) 'A comparative study of two lines of eimeria tenella attenuated either by selection for precocious development in the chicken or by growth in chicken embryos', *Avian Pathology*, 15(3), pp. 323–335. Available at: <https://doi.org/10.1080/03079458608436296>.
- Mongui, A. *et al.* (2010) 'Identification and characterization of the Plasmodium vivax thrombospondin-related apical merozoite protein', *Malaria Journal*, 9(1). Available at: <https://doi.org/10.1186/1475-2875-9-283>.
- Moreira, C.K., Marrelli, M.T. and Jacobs-Lorena, M. (2004) 'Gene expression in Plasmodium: From gametocytes to sporozoites', *International Journal for Parasitology*, 34(13–14), pp. 1431–1440. Available at: <https://doi.org/10.1016/j.ijpara.2004.10.007>.
- Morgan, J.A.T. and Godwin, R.M. (2017) 'Mitochondrial genomes of Australian chicken Eimeria support the presence of ten species with low genetic diversity among strains', *Veterinary Parasitology*, 243(October 2016), pp. 58–66. Available at: <https://doi.org/10.1016/j.vetpar.2017.05.025>.
- Morris, G.M. and Gasser, R.B. (2006) 'Biotechnological advances in the diagnosis of avian coccidiosis and the analysis of genetic variation in Eimeria', *Biotechnology Advances*, 24(6), pp. 590–603. Available at: <https://doi.org/10.1016/j.biotechadv.2006.06.001>.
- Morrisette, N.S. and Sibley, L.D. (2002) 'Cytoskeleton of Apicomplexan Parasites', *Microbiology and Molecular Biology Reviews*, 66(1), pp. 21–38. Available at: <https://doi.org/10.1128/mmbr.66.1.21-38.2002>.
- Muller, I. *et al.* (2019) 'Transcriptome and histone epigenome of Plasmodium vivax salivary-gland sporozoites point to tight regulatory control and mechanisms for liver-stage differentiation in relapsing malaria', *International Journal for Parasitology*, 49(7), pp. 501–513. Available at: <https://doi.org/10.1016/j.ijpara.2019.02.007>.
- Murshudov, G.N. *et al.* (2011) 'REFMAC5 for the refinement of macromolecular crystal structures', *Acta Crystallographica Section D: Biological Crystallography*, 67(4), pp. 355–367. Available at: <https://doi.org/10.1107/S0907444911001314>.
- Nagel, S.D. and Boothroyd, J.C. (1989) 'The major surface antigen, P30, of Toxoplasma gondii is anchored by a glycolipid', *Journal of Biological Chemistry*, 264(10), pp. 5569–5574. Available at: [https://doi.org/10.1016/s0021-9258\(18\)83584-0](https://doi.org/10.1016/s0021-9258(18)83584-0).
- O'Donnell, R.A. *et al.* (2006) 'Intramembrane proteolysis mediates shedding of a key adhesin during erythrocyte invasion by the malaria parasite', *Journal of Cell Biology*, 174(7), pp. 1023–1033. Available at: <https://doi.org/10.1083/jcb.200604136>.
- O'Toole, D. and Jeffrey, M. (1987) 'Congenital sporozoan encephalomyelitis in a calf.', *The Veterinary record*, 121(24), pp. 563–566.
- Oddoux, O. *et al.* (2011) 'Identification of the five human Plasmodium species including P. knowlesi by real-time polymerase chain reaction', *European Journal of Clinical Microbiology and Infectious Diseases*, 30(4), pp. 597–601. Available at: <https://doi.org/10.1007/s10096-010-1126-5>.
- Onyilagha, C. and Uzonna, J.E. (2019) 'Host Immune Responses and Immune Evasion

Strategies in African Trypanosomiasis', *Frontiers in Immunology*, 10(November). Available at: <https://doi.org/10.3389/fimmu.2019.02738>.

Opitz, C. *et al.* (2002) 'Intramembrane cleavage of microneme proteins at the surface of the apicomplexan parasite *Toxoplasma gondii*', *EMBO Journal*, 21(7), pp. 1577–1585. Available at: <https://doi.org/10.1093/emboj/21.7.1577>.

Owusu, C.K. and Bennett, H.M. (2015) 'How apicomplexans became free-riders', *Nature Reviews Microbiology*, 13(10), p. 603. Available at: <https://doi.org/10.1038/nrmicro3551>.

Pacheco, M.A. *et al.* (2013) 'Malarial parasite diversity in chimpanzees: The value of comparative approaches to ascertain the evolution of *Plasmodium falciparum* antigens', *Malaria Journal*, 12(1), pp. 1–20. Available at: <https://doi.org/10.1186/1475-2875-12-328>.

Palmieri, N. *et al.* (2017) 'The genome of the protozoan parasite *Cystoisospora suis* and a reverse vaccinology approach to identify vaccine candidates', *International Journal for Parasitology*, 47(4), pp. 189–202. Available at: <https://doi.org/10.1016/j.ijpara.2016.11.007>.

Pinz, S., Daskocil, E. and Seufert, W. (2022) 'Thermofluor-Based Analysis of Protein Integrity and Ligand Interactions', in K.-D. Entian (ed.) *Ribosome Biogenesis: Methods and Protocols*. New York, NY: Springer US, pp. 247–257. Available at: [https://doi.org/10.1007/978-1-0716-2501-9\\_15](https://doi.org/10.1007/978-1-0716-2501-9_15).

Portman, N. and Šlapeta, J. (2014) 'The flagellar contribution to the apical complex: A new tool for the eukaryotic swiss army knife?', *Trends in Parasitology*, 30(2), pp. 58–64. Available at: <https://doi.org/10.1016/j.pt.2013.12.006>.

Potterton, L. *et al.* (2018) 'CCP 4 i 2: The new graphical user interface to the CCP 4 program suite', *Acta Crystallographica Section D: Structural Biology*, 74(August 2017), pp. 68–84. Available at: <https://doi.org/10.1107/S2059798317016035>.

Ramakrishnan, C. *et al.* (2022) 'Dissection of *Besnoitia besnoiti* intermediate host life cycle stages: From morphology to gene expression', *PLoS Pathogens*, 18(11), pp. 1–29. Available at: <https://doi.org/10.1371/journal.ppat.1010955>.

Ramaprasad, A. *et al.* (2015) 'Comprehensive evaluation of *Toxoplasma gondii* VEG and *Neospora caninum* LIV genomes with tachyzoite stage transcriptome and proteome defines novel transcript features', *PLoS ONE*, 10(4), pp. 1–16. Available at: <https://doi.org/10.1371/journal.pone.0124473>.

Ramly, N.Z. (2012) 'Molecular analysis on the supefamily of surface antigens from the Apicomplexan parasite *Eimeria tenella*. A thesis submitted in part fulfillment of the requirements for the degree of Doctor of Philosophy By Nur Zazarina Ramly B . Sc . ( Hons ) National Uni', *Department of Molecular Biology and Biotechnology University of Sheffield* [Preprint], (June).

Ramly, N.Z. *et al.* (2021) 'The structure of a major surface antigen SAG19 from *Eimeria tenella* unifies the *Eimeria* SAG family', *Communications Biology*, 4(1), pp. 1–9. Available at: <https://doi.org/10.1038/s42003-021-01904-w>.

Rangel, G.W. *et al.* (2020) '*Plasmodium vivax* transcriptional profiling of low input cryopreserved isolates through the intraerythrocytic development cycle', *PLoS Neglected Tropical Diseases*, 14(3), pp. 1–18. Available at: <https://doi.org/10.1371/journal.pntd.0008104>.

- Ravindran, S. and Boothroyd, J.C. (2008) 'Secretion of proteins into host cells by Apicomplexan parasites', *Traffic*, 9(5), pp. 647–656. Available at: <https://doi.org/10.1111/j.1600-0854.2008.00723.x>.
- Reid, A. *et al.* (2021) 'The complete genome sequence of *Eimeria tenella* (Tyzzer 1929), a common gut parasite of chickens', *Wellcome Open Research*, 6(Tyzzer 1929), pp. 1–12. Available at: <https://doi.org/10.12688/wellcomeopenres.17100.1>.
- Reid, A.J. *et al.* (2012) 'Comparative genomics of the apicomplexan parasites *Toxoplasma gondii* and *Neospora caninum*: Coccidia differing in host range and transmission strategy', *PLoS Pathogens*, 8(3). Available at: <https://doi.org/10.1371/journal.ppat.1002567>.
- Reid, A.J. *et al.* (2014) 'Genomic analysis of the causative agents of coccidiosis in domestic chickens', *Genome Research*, 24(10), pp. 1676–1685. Available at: <https://doi.org/10.1101/gr.168955.113>.
- Ritzi, M.M. *et al.* (2016) 'Combination of probiotics and coccidiosis vaccine enhances protection against an *Eimeria* challenge', *Veterinary Research*, 47(1), pp. 1–8. Available at: <https://doi.org/10.1186/s13567-016-0397-y>.
- Rochell, S.J., Parsons, C.M. and Dilger, R.N. (2016) 'Effects of *Eimeria acervulina* infection severity on growth performance, apparent ileal amino acid digestibility, and plasma concentrations of amino acids, carotenoids, and  $\alpha$ 1-acid glycoprotein in broilers', *Poultry Science*, 95(7), pp. 1573–1581. Available at: <https://doi.org/10.3382/ps/pew035>.
- Rommereim, L.M. *et al.* (2016) 'Phenotypes associated with knockouts of eight dense granule gene loci (GRA2-9) in virulent *Toxoplasma gondii*', *PLoS ONE*, 11(7). Available at: <https://doi.org/10.1371/journal.pone.0159306>.
- Roth, A. *et al.* (2018) 'Unraveling the *Plasmodium vivax* sporozoite transcriptional journey from mosquito vector to human host', *Scientific Reports*, 8(1), pp. 1–20. Available at: <https://doi.org/10.1038/s41598-018-30713-1>.
- Rowe, J.A. *et al.* (2009) 'Adhesion of *Plasmodium falciparum*-infected erythrocytes to human cells: Molecular mechanisms and therapeutic implications', *Expert Reviews in Molecular Medicine*, 11(May), pp. 1–29. Available at: <https://doi.org/10.1017/S1462399409001082>.
- Sam-Yellowe, Y. (1996) Reviews Rhoptry Organelles of the Apicomplexa : Their Role in Host Cell Invasion and Intracellular Survival.
- Sarah E. Macdonald , Matthew J. Nolan, Kimberley Harman, Kay Boulton, David A. Hume, Fiona M. Tomley, Richard A. Stabler, D.P.B. (2017) 'Effects of *Eimeria tenella* infection on chicken caecal microbiome diversity, exploring variation associated with severity of pathology', *Acta Chimica Slovenica*, 12. Available at: <https://doi.org/10.1371/journal.pone.0184890>.
- Serrano, R.L. *et al.* (2004) 'Structural analysis of the human Golgi-associated plant pathogenesis related protein GAPR-1 implicates dimerization as a regulatory mechanism', *Journal of Molecular Biology*, 339(1), pp. 173–183. Available at: <https://doi.org/10.1016/j.jmb.2004.03.015>.
- Shen, B. and Sibley, L.D. (2012) 'The moving junction, a key portal to host cell invasion by apicomplexan parasites', *Current Opinion in Microbiology*, 15(4), pp. 449–455. Available at: <https://doi.org/10.1016/j.mib.2012.02.007>.

- Shirley, M.W. and Harvey, D.A. (1996) 'Eimeria tenella: Infection with a single sporocyst gives a clonal population', *Parasitology*, 112(6), pp. 523–528. Available at: <https://doi.org/10.1017/s0031182000066099>.
- Sibley, L.D. (2010) 'How apicomplexan parasites move in and out of cells', *Current Opinion in Biotechnology*, 21(5), pp. 592–598. Available at: <https://doi.org/10.1016/j.copbio.2010.05.009>.
- Smith, T.G., Walliker, D. and Ranford-Cartwright, L.C. (2002) 'Sexual differentiation and sex determination in the Apicomplexa', *Trends in Parasitology*, 18(7), pp. 315–323. Available at: [https://doi.org/10.1016/S1471-4922\(02\)00292-4](https://doi.org/10.1016/S1471-4922(02)00292-4).
- Song, X. *et al.* (2015) 'Partial protection against four species of chicken coccidia induced by multivalent subunit vaccine', *Veterinary Parasitology*, 212(3–4), pp. 80–85. Available at: <https://doi.org/10.1016/j.vetpar.2015.08.026>.
- Soutter, F. *et al.* (2020) 'Poultry Coccidiosis: Design and Interpretation of Vaccine Studies', *Frontiers in Veterinary Science*, 7(February), pp. 1–12. Available at: <https://doi.org/10.3389/fvets.2020.00101>.
- Spence, P.J. *et al.* (2013) 'Vector transmission regulates immune control of Plasmodium virulence', *Nature*, 498(7453), pp. 228–231. Available at: <https://doi.org/10.1038/nature12231>.
- Sreejith Raran-Kurussi, Scott Cherry, Di Zhang, D.S.W.A. (2017) 'Removal of Affinity Tags with TEV Protease Sreejith', 1586, pp. 221–230. Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7974378/pdf/nihms-1671727.pdf>.
- Stanley, P., Taniguchi, N. and Aebi, M. (2017) 'N-glycans', *Essentials of Glycobiology [Internet]*. 3rd edition [Preprint].
- Striepen, B. *et al.* (2007) 'Building the perfect parasite: Cell division in apicomplexa', *PLoS Pathogens*, 3(6), pp. 0691–0698. Available at: <https://doi.org/10.1371/journal.ppat.0030078>.
- Swapna, L.S. and Parkinson, J. (2017) 'Genomics of apicomplexan parasites', *Critical Reviews in Biochemistry and Molecular Biology*, 52(3), pp. 254–273. Available at: <https://doi.org/10.1080/10409238.2017.1290043>.
- Tabarés, E. *et al.* (2004) 'Eimeria tenella sporozoites and merozoites differentially express glycosylphosphatidylinositol-anchored variant surface proteins', *Molecular and Biochemical Parasitology*, 135(1), pp. 123–132. Available at: <https://doi.org/10.1016/j.molbiopara.2004.01.013>.
- Tachado, S.D., Mazhari-Tabrizi, R. and Schofield, L. (1999) 'Specificity in signal transduction among glycosylphosphatidylinositols of Plasmodium falciparum, Trypanosoma brucei, Trypanosoma cruzi and Leishmania spp.', *Parasite Immunology*, 21(12), pp. 609–617. Available at: <https://doi.org/10.1046/j.1365-3024.1999.00268.x>.
- 'The CCP4 suite: programs for protein crystallography.' (1994) *Acta crystallographica. Section D, Biological crystallography*, 50(Pt 5), pp. 760–763. Available at: <https://doi.org/10.1107/S0907444994003112>.
- Tilney, L.G. *et al.* (2004) 'Bacterial Invasion: In Vivo Veritas', *Science*, 304(April), pp. 248–253.
- Tobin, C.M. and Knoll, L.J. (2012) 'A patatin-like protein protects Toxoplasma gondii from



- degradation in a nitric oxide-dependent manner', *Infection and Immunity*, 80(1), pp. 55–61. Available at: <https://doi.org/10.1128/IAI.05543-11>.
- Ueti, M.W. *et al.* (2020) 'Transcriptome dataset of Babesia bovis life stages within vertebrate and invertebrate hosts', *Data in Brief*, 33, p. 106533. Available at: <https://doi.org/10.1016/j.dib.2020.106533>.
- Waldman, B.S. *et al.* (2020) 'Identification of a Master Regulator of Differentiation in Toxoplasma.', *Cell*, 180(2), pp. 359–372.e16. Available at: <https://doi.org/10.1016/j.cell.2019.12.013>.
- Wallach, M. (2010) 'Role of antibody in immunity and control of chicken coccidiosis', *Trends in Parasitology*, 26(8), pp. 382–387. Available at: <https://doi.org/10.1016/j.pt.2010.04.004>.
- Wallach, M., Frölich, S. and Entzeroth, R. (2012) 'Comparison of protective immune responses to apicomplexan parasites', *Journal of Parasitology Research*, 2012. Available at: <https://doi.org/10.1155/2012/852591>.
- Wang, F. *et al.* (2023) 'Expression of a SAG protein with a CAP domain from Eimeria necatrix and its role in invasion and immunoprotection', *Veterinary Parasitology*, 324(September), p. 110060. Available at: <https://doi.org/10.1016/j.vetpar.2023.110060>.
- Wang, Y. and Yin, H. (2014) 'Research progress on surface antigen 1 (SAG1) of Toxoplasma gondii', *Parasites and Vectors*, 7(1), pp. 1–9. Available at: <https://doi.org/10.1186/1756-3305-7-180>.
- West, S.A., Smith, T.G. and Read, A.F. (2000) *Sex allocation and population structure in apicomplexan (protozoa) parasites*.
- Williams, C.J. *et al.* (2018) 'Cis-nonPro Peptides: Genuine Occurrences and their Functional Roles', *bioRxiv* [Preprint], (Videau 2004). Available at: <https://doi.org/10.1101/324517>.
- Williams, Christopher J. *et al.* (2018) 'MolProbity: More and better reference data for improved all-atom structure validation', *Protein Science*, 27(1), pp. 293–315. Available at: <https://doi.org/10.1002/pro.3330>.
- Winter, G. *et al.* (2018) 'DIALS: Implementation and evaluation of a new integration package', *Acta Crystallographica Section D: Structural Biology*, 74, pp. 85–97. Available at: <https://doi.org/10.1107/S2059798317017235>.
- Witcombe, D.M. *et al.* (2004) 'Eimeria maxima TRAP family protein EmTFP250: Subcellular localisation and induction of immune responses by immunisation with a recombinant C-terminal derivative', *International Journal for Parasitology*, 34(7), pp. 861–872. Available at: <https://doi.org/10.1016/j.ijpara.2004.03.006>.
- Wu, K. *et al.* (2009) 'Diagnosis of human toxoplasmosis by using the recombinant truncated surface antigen 1 of Toxoplasma gondii', *Diagnostic Microbiology and Infectious Disease*, 64(3), pp. 261–266. Available at: <https://doi.org/10.1016/j.diagmicrobio.2009.02.009>.
- Xu, E. *et al.* (2012) 'Structure of protein having inhibitory disintegrin and leukotriene scavenging functions contained in single domain', *Journal of Biological Chemistry*, 287(14), pp. 10967–10976. Available at: <https://doi.org/10.1074/jbc.M112.340471>.
- Yamada, K.D. *et al.* (2017) 'Designing better diffracting crystals of biotin carboxyl carrier protein from Pyrococcus horikoshii by a mutation based on the crystal-packing propensity of

amino acids', *Acta Crystallographica Section D: Structural Biology*, 73(9), pp. 757–766. Available at: <https://doi.org/10.1107/S2059798317010932>.

Young, J.A. *et al.* (2005) 'The *Plasmodium falciparum* sexual development transcriptome: A microarray analysis using ontology-based pattern identification', *Molecular and Biochemical Parasitology*, 143(1), pp. 67–79. Available at: <https://doi.org/10.1016/j.molbiopara.2005.05.007>.

Zhang, M. *et al.* (2021) 'The apicoplast link to fever-survival and artemisinin-resistance in the malaria parasite', *Nature Communications*, 12(1), pp. 1–15. Available at: <https://doi.org/10.1038/s41467-021-24814-1>.

Zhou, B. hua *et al.* (2020) 'Effects of *Eimeria tenella* infection on the barrier damage and microbiota diversity of chicken cecum', *Poultry Science*, 99(3), pp. 1297–1305. Available at: <https://doi.org/10.1016/j.psj.2019.10.073>.

Zhu, L. *et al.* (2016) 'New insights into the *Plasmodium vivax* transcriptome using RNA-Seq', *Scientific Reports*, 6(February), pp. 1–13. Available at: <https://doi.org/10.1038/srep20498>.

## 9. Appendix A

### A.1 SAG proteins full length sequence

#### *Eimeria tenella* SAG-A family

##### **SAG10. ETH\_00034975. Uniport ID Q70CC2**

###### **TEST sequence:**

```
SAEETVTTYTASLGNNVECLSDINAAREAAAGLHDFGQASGDAELSIPAPEEGELSEKWKKL  
CEYLIPQTETAAKTSSANPFEKGTYAFKSLTAEQPNCKETIDYWKAAYENFTGLPPSKKE  
GGTLYDDQDNVSFVAVYNPSSSATADCRVVTCTQTNTTTTTPGPTRVQADGGSETTKKGY  
ALLCKTMPTAFASDTSAPFTQAQWDKIMSS
```

###### **Full length**

```
>tr|Q70CC2|Q70CC2_EIMTE SAG family member (Sag10) OS=Eimeria tenella  
OX=5802 GN=sag10 PE=2 SV=1  
MLQRKLPPILRISFLMLSATTLGNSQETEETVTTYTASLGNNVECLSDINAAREAAAGLHDF  
GQASGDAELSIPAPEEGELSEKWKKLCEYLIPQTETAAKTSSANPFEKGTYAFKSLTAEQ  
PNCKETIDYWKAAYENFTGLPPSKKEGGTLYDDQDNVSFVAVYNPSSSATADCRVVTCTQ  
TNTTTTTPGPTRVQADGGSETTKKGYALLCKTMPTAFASDTSAPFTQAQWDKIMSSLTGS  
GSIAAPSLIALAIVTFGIMTL
```

##### **SAG31. ETH\_00010850. Uniport ID U6KYYK1**

###### **Test Sequence:**

```
SAKADGATTTGTAQTLDCCTEMNEVRKAAGLSEFEKASRETEVLPEYPGTTKKISAENLWK  
QTCQKLMGENVEITEAGSLVGTVAHYAGAKDCKEAVQYWKDGFSLFKNELPPKYTALGDPD  
VYTDRAVSFVALYNPKASPVASCAFTVCTKGTAVAAQEMSRRHDSSPLRRLQDGAQTKTAV  
ICLTNPDALAPGTAPFEDDVWQKIVHA
```

###### **Full length**

```
>tr|U6KYYK1|U6KYYK1_EIMTE SAG family member OS=Eimeria tenella OX=5802  
GN=ETH_00010850 PE=4 SV=1  
MLRLWFIASVMSVFCGHKADGATTTGTAQTLDCCTEMNEVRKAAGLSEFEKASRETEVL  
PEYPGTTKKISAENLWKQTCQKLMGENVEITEAGSLVGTVAHYAGAKDCKEAVQYWKDGF  
SLFKNELPPKYTALGDPDVYTDRAVSFVALYNPKASPVASCAFTVCTKGTAVAAQEMSRR  
HDSSPLRRLQDGAQTKTAVICLTNPDALAPGTAPFEDDVWQKIVHAIVGLEESNRASPIR  
PSLAVGFIVTILAHGLL
```

##### **SAG7. ETH\_00034950. Uniprot ID Q70CE1**

###### **Test sequence:**

```
SAQQAPTYTASLGKSNKCLSELNAAAREAAAGLPNFTEATDGKKLSDPEQQQLQEGSEWMKVCK  
HLVPTEQKDPVAAAAGATNPFQDGTAFKSLTAAEPNCKETVDHWKAAFENFTGLPPSKTEG  
ANLYKNQDNVSFVALYNPSSDATADCKVVTCTKATASEAALQSDSDQSSSENGYALICKTMP  
SAFPDDKSPFFTQDQWNKIVSSLTGSASTVIPGFCALFIAVSSWIAL
```

###### **Full length**

```
>tr|Q70CE1|Q70CE1_EIMTE SAG family member (Sag7) OS=Eimeria tenella OX=5802  
GN=sag7 PE=2 SV=1  
MLSLYICASLLVLSASLLKISQAGQQAPTYTASLGKSNKCLSELNAAAREAAAGLPNFTEA  
TDGKKLSDPEQQQLQEGSEWMKVCKHLVPTEQKDPVAAAAGATNPFQDGTAFKSLTAAEPN  
CKETVDHWKAAFENFTGLPPSKTEGANLYKNQDNVSFVALYNPSSDATADCKVVTCTKAT  
ASEAALQSDSDQSSSENGYALICKTMPSAFPDDKSPFFTQDQWNKIVSSLTGSASTVIPGF  
CALFIAVSSWIAL
```

## SAG 91. ETH-00023375. Uniprot ID [U6L7N4](#)

### Crystallised Sequence:

```
SAPPPSTVAAASIDCTTAMNGIRAEVGLAPLQLGTTEQSKLPLKAEESPNLAFSTAVCTA
ARAGTLPSTASTSLDGNFAISIQKGSNGDCATAVKHWREGFDLFDKGLPPPYSAEAPVYKT
SQAQSFVALFAPHEAGKVDCAFFICPESTDSTGKQEEIKALLCVTSPKSLKDGVAPEQK
QWDKIAAGLVG
```

### Full length

```
>tr|U6L7N4|U6L7N4_EIMTE SAG family member OS=Eimeria tenella OX=5802
GN=ETH_00023375 PE=4 SV=1
MRSLKLLFLAGSAVFFRDTQGAQGSSGPPSTVAAASIDCTTAMNGIRAEVGLAPLQLGT
TEQSKLPLKAEESPNLAFSTAVCTAARAGTLPSTASTSLDGNFAISIQKGSNGDCATAVK
HWREGFDLFDKGLPPPYSAEAPVYKTSQAQSFVALFAPHEAGKVDCAFFICPESTDSTGKQ
EEIKALLCVTSPKSLKDGVAPEQKQWDKIAAGLVGGASAAVPTFLVFAVTAVGVALF
```

## SAG1 uniprot ID [Q546B5](#)

### Test sequence

```
SAQDYPTAVTLDCKEAMNKLKRAAGLPAFEDAVGDTFVLPAYSHEESRAAPVAETLWKTEICPKVLGGGRSRNVT
EAVKLTGNFAYYPVTDGKKECSDAVEYWKGGLSQFNDTIPPTFQALNDPVVYNDRAVSFVALYNPKTSPVVSCLV
LQCPNAGVGGRRLAAGTTDAVICLTNPAPLEARSQPFDDQWKKIVDSLSSLS
```

### Full length

```
>tr|Q546B5|Q546B5_EIMTE Major sporozoite surface antigen OS=Eimeria tenella
OX=5802 GN=sag1 PE=2 SV=1
MARLSFVSLLSLSLLFGQQAVRAQDYPTAVTLDCKEAMNKLKRAAGLPAFEDAVGDTFVL
PAYSHEESRAAPVAETLWKTEICPKVLGGGRSRNVT EAVKLTGNFAYYPVTDGKKECSDA
VEYWKGGLSQFNDTIPPTFQALNDPVVYNDRAVSFVALYNPKTSPVVSCLVLLQCPNAGVG
GRRLAAGTTDAVICLTNPAPLEARSQPFDDQWKKIVDSLSSLSSEEEEEKGGVSPVVPVSA
LISA AVISAFALF
```

## SAG905. ETH\_00034905 uniprot ID [H9B926](#)

### Test sequence

```
SATTTVTTYTASVGDSAQCLDEVNAVREVAGLSNFTQASDPNKLQTPGTEGLQENTEWKRLCEHLIATQKAKAPSS
SEAENPFKDGTYAFKSLTDAQPKCKDTVDFWKA AFKNFTGLPPSKKQANGLYDKQDNVSFVALYNPSSSATADCR
VVTCTKKTSTKENELAAAEGNATSEYGYALICKTMPDALADENSAPFTQKQWDGIVSSLTG
```

### Full length

```
>tr|H9B926|H9B926_EIMTE SAG family member OS=Eimeria tenella OX=5802
GN=ETH\_00034905 PE=2 SV=1
MPMMLPRIGALLSTSLVL SAPIAGGSQQSETTTVTTYTASVGDSAQCLDEVNAVREVAGL
SNFTQASDPNKLQTPGTEGLQENTEWKRLCEHLIATQKAKAPSSSEAENPFKDGTYAFKS
LTDAQPKCKDTVDFWKA AFKNFTGLPPSKKQANGLYDKQDNVSFVALYNPSSSATADCRV
VTCTKKTSTKENELAAAEGNATSEYGYALICKTMPDALADENSAPFTQKQWDGIVSSLTG
SASVAVPKLVGIFILAVGMVAL
```

## SAG870. ETH\_00010870 uniport ID U6KV63

### Test sequence

SATTTGKAETLNCLGEMNEARMAAGLPKFEEATQVGQILPEHSSSGKDVAASTLWDQICKKIMGVAGEITEIQKL  
KGMPAYYPGEKDCKAAVQSWKDGFSLFNNQLPPTYVELGNPEVYNDRGISFVALYNPKASPVASCAFATCTTTGA  
GPAAAAAQPKSQGGHSPRRLQAEAQTLTAVICLTNPEALTAGAAPPFKEDEWQKIAHAMN

### Full length

>tr|U6KV63|U6KV63\_EIMTE SAG family member OS=Eimeria tenella OX=5802  
GN=ETH\_00010870 PE=4 SV=1  
MLRVLHMSIVFASVFCGQTESATTTGKAETLNCLGEMNEARMAAGLPKFEEATQVGQILP  
EHSSSGKDVAASTLWDQICKKIMGVAGEITEIQKLKGMPAYYPGEKDCKAAVQSWKDGFS  
LFNNQLPPTYVELGNPEVYNDRGISFVALYNPKASPVASCAFATCTTTGAGPAAAAAQPK  
SQGGHSPRRLQAEAQTLTAVICLTNPEALTAGAAPPFKEDEWQKIAHAMNGTESVKGSWAL  
QARPSLALGLVITLFAAYGLF

## SAG6. ETH\_00034945 Uniport ID Q70CD8

### Test sequence

SAPTIKYTASLGGGAKCLSEVNAAARGAAGLKNFAEATNDKKLSAPSDDLENDTEWKKVCEHLIPTQKEPVEATSG  
TNPFEKGTYAFKSLTTAEPNCKEIVNYWKAAPKNFTGLPPSESQAGDLYKSYNNVSFVALYNTSSNATADCQVVT  
CTKTTTTPGDSSIRDSPSGSQEYGYAMICKTMPAAFADKNSAPFTQDQWDRIISSLTG

### Full length

>tr|Q70CD8|Q70CD8\_EIMTE SAG family member (Sag6) OS=Eimeria tenella OX=5802  
GN=sag6 PE=2 SV=1  
MLPLRIPSVFSASIFLLSVSYLGTSQQSAPTIKYTASLGGGAKCLSEVNAAARGAAGLKNF  
AEATNDKKLSAPSDDLENDTEWKKVCEHLIPTQKEPVEATSGTNPFEKGTYAFKSLTTAE  
PNCKEIVNYWKAAPKNFTGLPPSESQAGDLYKSYNNVSFVALYNTSSNATADCQVVTCTK  
TTTTPGDSSIRDSPSGSQEYGYAMICKTMPAAFADKNSAPFTQDQWDRIISSLTGSASAAI  
PGFGAFFIVVLSMAVL

## *Eimeria Tenella* SAG-B family

## SAG19 ETH\_00041510 (used as a control for this study) uniport ID Q70CD0

### Test sequence

MTHVGLLACYAGLLAGAAAPDFSSALSLRSSTATSQQNSLSTNIFASGDVSPQTPTPPQA  
DEKTEDCLAIINKLRSENLKDLLGTLAKAEDTEVTESLKAIEEPASPTAPKIAVTLAG  
SNVDTCESGEGANAKKYPGLVIPFPDTEFNCNALIQATYTAGLDHLKQSNFEPSTGTID  
VENAPFNNVNASNVAFLLEKSKKVSAAATKDCKAGHDVLCYFIDPLRKEDKPFTAELY  
NALWGL

### Full length

>tr|Q70CD0|Q70CD0\_EIMTE SAG family member (Sag19) OS=Eimeria tenella  
OX=5802 GN=sag19 PE=1 SV=1  
MTHVGLLACYAGLLAGAAAPDFSSALSLRSSTATSQQNSLSTNIFASGDVSPQTPTPPQA  
DEKTEDCLAIINKLRSENLKDLLGTLAKAEDTEVTESLKAIEEPASPTAPKIAVTLAG  
SNVDTCESGEGANAKKYPGLVIPFPDTEFNCNALIQATYTAGLDHLKQSNFEPSTGTID  
VENAPFNNVNASNVAFLLEKSKKVSAAATKDCKAGHDVLCYFIDPLRKEDKPFTAELY  
NALWGLEAGAASISVPSVVTVLLALALIIRA

### **SAG13. ETH\_00013178 uniport ID Q70CD1**

#### **Test sequence**

SAAPSASKKTTECLPILNALRTEGLNGLLKGLVEAGDGEASQIQPLARSGKTTIQIASELAGTNKESCDATNANQSQYAGLVITFDVSKTFDCEALINASFTAGLDHLQKADYNATADESILGTPPLDNI  
AAKNLAAIVSTKAKEKVECAATDDCVAGKNVLCYFIQPLEKEQAQPIDANVYEALLKRQ

#### **Full length**

>tr|Q70CD1|Q70CD1\_EIMTE SAG family member (Sag13) OS=Eimeria tenella  
OX=5802 GN=sag13 PE=1 SV=1  
MSRLGLLACYAGLFAGAAAPDFSTAVPIRSAAVNPHHKILDTEGSSLAQVATAPSASKKT  
TECLPILNALRTEGLNGLLKGLVEAGDGEASQIQPLARSGKTTIQIASELAGTNKESCDATNANQSQYAGLVITFDVSKTFDCEALINASFTAGLDHLQKADYNATADESILGTPPLDNI  
AAKNLAAIVSTKAKEKVECAATDDCVAGKNVLCYFIQPLEKEQAQPIDANVYEALLKRQR  
GSASIAVPGITAMLFSLALILLS

### **SAG 16. ETH\_00013140. Uniprot ID Q70CD2**

#### **Crystallised Sequence:**

SAATTPIVAEDATTACLPTMMLRVNLNRDQALDALQPETGGASEDEEREEQGEHTKSKTTAEIAKELAGTKAETCEKGATADAKTHTGLVIPFEYSTVFDGSLIQGHFAAGLSHMQESNFDPATGAYDTGKAPFDNLSASNVANIMWSKSTKASCAVTKNCQAGHNVLYCRLVEPITSQDKPFTTELYEALLQR

#### **Full length**

>tr|Q70CD2|Q70CD2\_EIMTE SAG family member (Sag16) OS=Eimeria tenella  
OX=5802 GN=sag16 PE=2 SV=1  
MLRPGLLACYIGLLAGAATASFSGAIIITRSANTNPHYDIVDAETAQVQATTPIVAEDATTACLPTMMLRVNLNRDQALDALQPETGGASEDEEREEQGEHTKSKTTAEIAKELAGTKAETCEKGATADAKTHTGLVIPFEYSTVFDGSLIQGHFAAGLSHMQESNFDPATGAYDTGKAPFDNLSASNVANIMWSKSTKASCAVTKNCQAGHNVLYCRLVEPITSQDKPFTTELYEALLQRQAGSSSIALTSIATTTFFCAAWLLST

### **SAG22. ETH\_00008675. Uniprot ID Q70CC3**

#### **Crystallised Sequence:**

SASLRAAPTGNEITADCLDTINKLRNENIKDLLGTLTKAEDSDVTASLKTIPVADAASLT  
TATIAAKLAGDSVDTCASGGNADAKTYPGLVIPFTHDKDFDCDALIQATYTAGLNQLKQSNFEP  
SKGTYDATKAPFDNVDASNVAFLLSAKSKKVSCAATKNCNAGHDVLCYFIEPLRNGDQSFH  
NLSFTMPPWGLEAG

#### **Full length**

>tr|Q70CC3|Q70CC3\_EIMTE Surface antigen 22 OS=Eimeria tenella OX=5802  
GN=sag22 PE=2 SV=1  
MTHLGLLACYAGLLASAAAPHFSLALSLSRSGTAASQQSSLSANLFASGQVSLRAAPTGNE  
ITADCLDTINKLRNENIKDLLGTLTKAEDSDVTASLKTIPVADAASLTATIAAKLAGDS  
VDTCASGGNADAKTYPGLVIPFTHDKDFDCDALIQATYTAGLNQLKQSNFEP  
SKGTYDATKAPFDNVDASNVAFLLSAKSKKVSCAATKNCNAGHDVLCYFIEPLRNGDQSFH  
NLSFTMPPWGLEAGAASTAVPSVATVLLTLALIIQP

### **SAG41. ETH\_00008720. Uniprot ID U6L630**

#### **Crystallised sequence:**

SALLRTTTTAPTANAKTQDCLEIINTLRKENLQDLLGTLTKAEESDVTASLKKIKIEGSDE  
LSTAKIAAKLAGSDAQNCESGESANAKTYPGLVIPFPHTTDFDCNTLIQATYTAGLDHLK  
QSNFEPSTGTYDVKTPFNNVDASNVAFLLSAKSTKVSCAATEDCAGGHDVLCYFIDPL  
QSGDQAFTTELYNALWGLEAG

#### **Full length**

>tr|U6L630|U6L630\_EIMTE SAG family member OS=Eimeria tenella OX=5802  
GN=ETH\_00008720 PE=4 SV=1  
MTYVGLLACYAGLLASAAAPHFSSAISLRAGTATSQKSSLRTNLFASGQDLLRTTTTAPTA  
NEKTQDCLEIINTLRKENLQDLLGTLTKAEESDVTASLKKIKIEGSDELSTAKIAAKLAG  
SDAQNCESGESANAKTYPGLVIPFPHTTDFDCNTLIQATYTAGLDHLKQSNFEPSTGTYD  
VDKTPFNNVDASNVAFLLSAKSTKVSCAATEDCAGGHDVLCYFIDPLQSGDQAFTTELY  
NALWGLEAGAASISVPSVATILLVLALGIWN

### ***Eimeria Tenella* SAG-C**

### **SAGC. ETH\_00001975. Uniport ID U6KHG0**

#### **Test sequence:**

SASAASPTPVAAGDDVYLALNLARRGRLAVRLNALTKQOTLVDSLLKSIPTTVGNDCGKI  
DSVTSQTASGFVATFTTEPNYKKLVQDALSAALKKMTKYPTDDKFNVPWTDAEVANILH  
VLSSASTEVGCAVTTKASKQLLVCQMNPKLGTGAPFSEEFFKALQSRSDSIE

#### **Full length**

>tr|U6KHG0|U6KHG0\_EIMTE SAG family member OS=Eimeria tenella OX=5802  
GN=ETH\_00001975 PE=4 SV=1  
MAPIFQSAALCFMALCGLKSAHAAGGGGSAASPTPVAAGDDVYLALNLARRGRLAVRLN  
ALTKQOTLVDSLLKSIPTTVGNDCGKIDSVTSQTASGFVATFTTEPNYKKLVQDALSAAL  
KKMTKYPTDDKFNVPWTDAEVANILHVLSSASTEVGCAVTTKASKQLLVCQMNPKLGT  
GAPFSEEFFKALQSRSDSIEDMTEADLKTGSNSGIVAVPSVLFAGLVAMLATAAA

### ***Eimeria brunetti* SAG-C family**

### **EBC1. EBH\_0036680. Uniport ID U6LD01**

#### **Test sequence:**

SAAATITYKFTPVDVDDAGYLAANLVRNGKLPVHISTVEKAESIVTALTNKVKSRTATKN  
ESDVSDGACDELVKQDEVKIDIFHYTFDYEEGLDYSKLLQKALDAGLEVFKETQONQNKWET  
IWQDGDGANLAYLLGANSTTIGCVIGQCTTKTNQAPSRETPDGSTTGKAVLFCELKPAAD  
KGKAPFDDEYFNGLIARTA

#### **Full length**

>tr|U6LD01|U6LD01\_9EIME SAG family member OS=Eimeria brunetti OX=51314  
GN=EBH\_0036680 PE=4 SV=1  
MASLYKTALAVCLLGHYGLQTEAAATITYKFTPVDVDDAGYLAANLVRNGKLPVHISTVE  
KAESIVTALTNKVKSRTATKNESDVSDGACDELVKQDEVKIDIFHYTFDYEEGLDYSKLLQ  
KALDAGLEVFKETQONQNKWETIWQDGDGANLAYLLGANSTTIGCVIGQCTTKTNQAPSRE  
TPDGSTTGKAVLFCELKPAADKGKAPFDDEYFNGLIARTAKLADMTEEDLKAPSNDGTAA  
AAAPTILAAGFVAVLSALSV

## EBC2. EBH\_0015570. Uniport ID U6L643

### Test sequence:

SAAASPQTYYKFKAVEVTDDAYIAANLVRNGKLEVHISEVSKDTNLPGLQKEVVASASV  
EQPEGTTESCKKLMEQSGLGIFHHAFSYKEKNDYRELFQAALDAGIAVFKEKGYQNKWD  
EIWASDAGASLAYLLGANSTKIGCVIGECIQVQTQDEGEPSSEESTGNAFLFCDLSPAVD  
KSKAPFDEEYFNVLVARTA

### Full length

>tr|U6L643|U6L643\_9EIME SAG family member OS=Eimeria brunetti OX=51314  
GN=EBH\_0015570 PE=4 SV=1  
MAFCKTAAAVCLVALYGLQSEAASPQTYYKFKAVEVTDDAYIAANLVRNGKLEVHISEVS  
KDTNLPGLQKEVVASASVEQPEGTTESCKKLMEQSGLGIFHHAFSYKEKNDYRELFQA  
ALDAGIAVFKEKGYQNKWDEIWASDAGASLAYLLGANSTKIGCVIGECIQVQTQDEGEPS  
SEESTGNAFLFCDLSPAVDKSKAPFDEEYFNVLVARTAKLAEMTEEDLKAPSNDGTAAGA  
FPTILVAGLVAMLTAAFA

## EBC3. EBH\_0027980 uniport ID U6LM73

### Test sequence

SAAQGTVKLLVDDAGPGSVLAANLARGGKISVPTKLLKEDESIKEGLKKALGETQNVLTGNACDSVTVKPDFVKK  
KFFVQFKSEEDSKDFRKDVQGALDKGVKLLKDMKTYPTDAAQWQTFWGNPDGANVANLLWSNSTKVGCAVGVCVE  
VQASSDVP LLGTNAYLFCQLNPEAEENKAPFDKKYYDALIARTTPLTAMTKDDL

### Full length

>tr|U6LM73|U6LM73\_9EIME SAG family member OS=Eimeria brunetti OX=51314  
GN=EBH\_0027980 PE=4 SV=1  
MAPIFKSAAAFCLVALCGLQSTAAQGTVKLLVDDAGPGSVLAANLARGGKISVPTKLLK  
EDESKEGLKKALGETQNVLTGNACDSVTVKPDFVKKKFFVQFKSEEDSKDFRKDVQGAL  
DKGVKLLKDMKTYPTDAAQWQTFWGNPDGANVANLLWSNSTKVGCAVGVCVEVQASSDVP  
LLGTNAYLFCQLNPEAEENKAPFDKKYYDALIARTTPLTAMTKDDLPSKNGATAVAVPSL  
LLTGLAAILATAAA

## Other CAP-like SAG Sequences

### *Plasmodium falciparum* SAG. Uniport ID AF-A0A024WSS3-F1

#### Test sequence:

SAGQFCKFNKEFIKERHNDFRLKHKAKPLQWSKKLEEIATYEANLIRDNSDCIVSSKQVDTNYF  
SFFKNENIEASVDTWYEGINDYDFELGCIKRNDNIFEFTRIIWKSSENLCATACCKTKGILIC  
KYDNNTNKPgyFADNVGTIDTMYVLD

#### Full length

>tr|A0A2I0BNY8|A0A2I0BNY8\_PLAFO Cysteine-rich secretory protein  
OS=Plasmodium falciparum (isolate NF54) OX=5843 GN=CK202\_5340 PE=4 SV=1  
MIGIMNIFLLFFVFISYIYVNGQFCKFNKEFIKERHNDFRLKHKAKPLQWSKKLEEIATY  
EANLIRDNSDCIVSSKQVDTNYFSFFKNENIEASVDTWYEGINDYDFELGCIKRNDNIFE  
FTRIIWKSSENLCATACCKTKGILICKYDNNTNKPgyFADNVGTIDTMYVLDNLNNGIH  
IMKNDHSDKRTQS



### ***Plasmodium ovale* SAG. Uniport ID [A0A1D3KYL5](#)**

#### **Test sequence:**

SAKWFCNMNADLILERHNDLRLKHNAKPLLWSSQLEKSAKDETSLIESNPDCIVTTKHLNTN  
YFTLSKYSEMDTAVNMWYEGINDYDFELGPILKGENVFEFTRVVWKSQAHIGCAVACCKN  
RGVLLCKYDSYTNKPGHFADDVGTIDTMFAWD

#### **Full length**

>tr|A0A1D3KYL5|A0A1D3KYL5\_9APIC Cysteine-rich secretory protein, putative  
OS=Plasmodium ovale OX=36330 GN=PocGH01\_01013600 PE=4 SV=1  
MVKNCSLCFFFLFSFFFLASDAHRSKWFCNMNADLILERHNDLRLKHNAKPLLWSSQLEK  
SAKDETSLIESNPDCIVTTKHLNTNYFTLSKYSEMDTAVNMWYEGINDYDFELGPILKGE  
NVFEFTRVVWKSQAHIGCAVACCKNRGVLLCKYDSYTNKPGHFADDVGTIDTMFAWDGTS  
AEIGNVGKEEINAE

### ***P. vivax* SAG. Uniport ID [A0A1G4GQR1](#)**

#### **Test sequence:**

SAGGFCSFNND FIRERHNDLRLKHNA DPLRWSTQLEKAASVEAKLIKEISNCTVMVNQINTN  
YFTISPNSKVESAVDTWYEGINNYDFELGPIRRGDDTVFEFTRVIWKS AELIGCSSACCG  
NRGVLLICKYDSNTNQPGHFADNVGTLDPMFVWE

#### **Full length**

>tr|A0A1G4GQR1|A0A1G4GQR1\_PLAVI Cysteine-rich secretory protein, putative  
OS=Plasmodium vivax OX=5855 GN=PVC01\_010012600 PE=4 SV=1  
MVDRLRLHCLFALLCFLTL SRISFCKAAAGGFCSFNND FIRERHNDLRLKHNA DPLRWST  
QLEKAASVEAKLIKEISNCTVMVNQINTNYFTISPNSKVESAVDTWYEGINNYDFELGPI  
RRGDDTVFEFTRVIWKS AELIGCSSACCGNRGVLLICKYDSNTNQPGHFADNVGTLDPMFV  
WENFTFAPEQRRPASGPS ENGLPPSPIS

### ***Plasmodium malariae* SAG. Uniport ID [A0A1D3JK29](#)**

#### **Test sequence:**

SAASFCKFNKNLIK ERHNDFRFKHNTKPLLWSKQLEESAKEE ANFIKANSDCVVA AKQIN  
TNYFD FLNGEEIESAVNSWYEGINNYDFELGPIKKGENVFEFTKV VWKGA EHIGCATACC  
KYRGILICKYDNNV NKPGYFADNVGMIDCSFGMT

#### **Full length**

>tr|A0A1D3JK29|A0A1D3JK29\_PLAMA Cysteine-rich secretory protein, putative  
OS=Plasmodium malariae OX=5858 GN=PmUG01\_01017300 PE=4 SV=1  
MNSCLFLFFSLFFICRYATRDASFCKFNKNLIK ERHNDFRFKHNTKPLLWSKQLEESA KE  
EANFIKANSDCVVA AKQINTNYFD FLNGEEIESAVNSWYEGINNYDFELGPIKKGENVFE  
FTKV VWKGA EHIGCATACC KYRGILICKYDNNV NKPGYFADNVGMIDCSFGMTHGLKNSK  
RAIHQFGTIRLVYIRIHIIHTYV

## **Babesia bovis SAG. Uniport ID A7AN09**

### **Test sequence:**

SAQIKSRCHFEEGRMLAALNDRREFHSSPSLKWNADLATSARNMASELSRRINCQLPLY  
REELGTNYLSADLEHFSESLAAEFWYEGHLDYDFEKGGLNRNPNVLSFTQMVRSTREV  
GCAVACCDGRQVVLVCRFHPPGNIQGQFIGNVLEKFNRLKNR

### **Full length**

>tr|A7AN09|A7AN09\_BABBO SCP-like extracellular protein family protein  
OS=Babesia bovis OX=5865 GN=BBOV\_III003800 PE=4 SV=1  
MRIVNDDTEVYSRQIKSRCHFEEGRMLAALNDRREFHSSPSLKWNADLATSARNMASELS  
RRINCQLPLYREELGTNYLSADLEHFSESLAAEFWYEGHLDYDFEKGGLNRNPNVLSF  
TQMVRSTREVGCAVACCDGRQVVLVCRFHPPGNIQGQFIGNVLEKFNRLKNRESGISTEL

## **Besnoitia besnoiti SAG. Uniport ID A0A2A9M912**

### **Test sequence:**

SAIMSPVAACLLIHNKYRTENLQVPLPAMSRNADAVALVMEFVEKRAVGKCKVQGHSTES  
QAKEMGENLYMSNNPTCEAAVTAWYDEIQYFDGNYPGTEWSNMKEPVGHFTQVMWTKSTG  
LGCARTIGCEGWNQLFCVYQPAGNYLGQAPFSEEVWKAIKKRD

### **Full length**

>tr|A0A2A9M912|A0A2A9M912\_9APIC SCP family extracellular subfamily protein  
OS=Besnoitia besnoiti OX=94643 GN=BESB\_017040 PE=4 SV=1  
MAPVLVLVTKFGTAFVGALCCSSPLFGTATHAGQLSTANSAARASLRAAESESHELGA  
KDVTDGIYDHSSSIMSPVAACLLIHNKYRTENLQVPLPAMSRNADAVALVMEFVEKRAVGKCK  
VQGHSTESQAKEMGENLYMSNNPTCEAAVTAWYDEIQYFDGNYPGTEWSNMKEPVGHFTQ  
VMWTKSTGLGCARTIGCEGWNQLFCVYQPAGNYLGQAPFSEEVWKAIKKRDGISGAMALA  
PVSTVAVAVATSALVYVLVA

## **Trx-tag sequence including the His-tag and TEV at the end of the sequence.**

MSDKIIHLTDDSFDTDLKADGAILVDFWAEWCGPCKMIAPILDEIADEYQGKLTVAKLNIDQNP  
GTAPKYGIRG  
IPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSGSGHMHSSGENLYFQ/SA

## A.2 *Eimeria tenella* SAG-B family identity

2: B9_SAG14/1-265	50.00	100.00	57.48	57.03	42.19	41.92	42.31	40.77	42.59	40.30	39.30	42.05	44.32	42.80	44.11	42.91	42.59	45.45	43.35	43.51	46.21	44.32	44.32	45.45	45.45	45.08	45.08
3: ETH_13170/1-254	59.19	57.48	100.00	60.71	37.60	39.92	38.74	38.74	41.90	40.32	38.21	39.92	42.69	42.29	41.67	41.04	42.46	39.13	40.87	42.46	41.90	41.90	41.90	43.87	43.87	43.87	43.87
4: B8_SAG13/1-263	55.17	57.03	60.71	100.00	41.34	41.86	42.64	41.86	42.91	40.23	39.22	40.84	40.84	38.17	41.76	42.08	42.53	41.60	40.61	41.54	41.60	43.89	43.89	44.66	44.27	43.89	44.27
5: B5_SAG15/1-263	34.35	42.19	37.60	41.34	100.00	75.29	75.67	73.76	43.80	39.08	38.67	40.30	42.21	39.16	41.98	42.08	41.92	41.44	43.30	40.77	41.76	42.59	42.59	43.73	43.73	44.87	44.87
6: B2_SAG17/1-270	32.91	41.92	39.92	41.86	75.29	100.00	80.97	79.85	40.46	40.00	38.85	41.57	43.45	40.82	42.48	41.83	40.91	43.07	42.26	41.67	44.15	43.07	43.07	44.19	44.19	44.94	44.94
7: B3_SAG25/1-268	31.62	42.31	38.74	42.64	75.67	80.97	100.00	82.84	41.60	38.49	39.23	39.70	40.82	40.07	42.11	40.30	42.05	41.20	42.64	40.91	43.40	41.95	41.95	43.07	43.07	43.82	43.82
8: B4_SAG16/1-268	30.34	40.77	38.74	41.86	73.76	79.85	82.84	100.00	41.98	40.38	39.23	40.45	42.70	40.82	43.61	40.30	42.05	41.20	43.77	40.53	43.40	41.20	41.20	42.70	42.70	43.82	43.82
9: B7_SAG18/1-268	35.62	42.59	41.90	42.91	43.80	40.46	41.60	41.98	100.00	56.77	56.76	60.15	61.28	60.15	61.51	57.20	60.38	59.02	59.62	58.87	58.65	60.53	60.53	60.90	60.53	60.15	60.53
10: A2_SAG22/1-270	32.20	40.30	40.32	40.23	39.08	40.00	38.49	40.38	56.77	100.00	64.89	69.52	71.38	69.89	71.64	70.41	67.67	63.94	68.54	69.40	67.42	68.40	68.40	69.89	69.52	68.77	69.14
11: ETH_08680/1-264	33.62	39.30	38.21	39.22	38.67	38.85	39.23	39.23	56.76	64.89	100.00	65.15	64.02	65.91	69.96	67.31	66.67	65.15	63.74	72.80	61.45	71.97	71.97	71.97	72.35	71.97	71.59
12: ETH_08715/1-271	35.71	42.05	39.92	40.84	40.30	41.57	39.70	40.45	60.15	69.52	65.15	100.00	74.54	75.28	69.26	70.04	68.66	68.27	72.49	69.78	69.52	70.85	70.85	69.37	69.37	69.00	69.00
13: A8_SAG41/1-271	34.87	44.32	42.69	40.84	42.21	43.45	40.82	42.70	61.28	71.38	64.02	74.54	100.00	80.44	73.33	71.91	70.15	70.48	73.23	71.27	73.98	73.80	73.80	73.80	73.80	73.06	73.06
14: A10_SAG38/1-271	35.71	42.80	42.29	38.17	39.16	40.82	40.07	40.82	60.15	69.89	65.91	75.28	80.44	100.00	72.59	71.91	68.66	70.11	73.98	71.27	72.49	74.91	74.91	73.43	73.80	73.43	73.06
15: A1_SAG23/1-270	35.02	44.11	41.67	41.76	41.98	42.48	42.11	43.61	61.51	71.64	69.96	69.26	73.33	72.59	100.00	72.56	73.41	70.37	72.39	72.66	73.51	73.70	73.70	72.59	72.22	72.22	72.59
16: ETH_08690/1-267	35.90	42.91	41.04	42.08	42.08	41.83	40.30	40.30	57.20	70.41	67.31	70.04	71.91	71.91	72.56	100.00	68.18	71.16	73.96	74.44	69.06	71.91	71.91	74.16	74.16	73.78	73.78
17: ETH_08725/1-268	35.59	42.59	42.46	42.53	41.92	40.91	42.05	42.05	60.38	67.67	66.67	68.66	70.15	68.66	73.41	68.18	100.00	70.52	69.66	69.81	71.54	71.27	71.27	73.88	73.51	73.13	73.51
18: A12_SAG20/1-271	34.87	45.45	39.13	41.60	41.44	43.07	41.20	41.20	59.02	63.94	65.15	68.27	70.48	70.11	70.37	71.16	70.52	100.00	70.26	72.01	74.72	73.80	73.80	72.69	72.69	72.32	72.32
19: A9_SAG39/1-269	33.47	43.35	40.87	40.61	43.30	42.26	42.64	43.77	59.62	68.54	63.74	72.49	73.23	73.98	72.39	73.96	69.66	70.26	100.00	71.16	71.91	71.00	71.00	74.35	74.35	74.35	74.35
20: A3_SAG64/1-268	36.17	43.51	42.46	41.54	40.77	41.67	40.91	40.53	58.87	69.40	72.80	69.78	71.27	71.27	72.66	74.44	69.81	72.01	71.16	100.00	71.80	75.00	75.00	75.37	75.37	74.25	74.25
21: A11_SAG21/1-269	36.02	46.21	41.90	41.60	41.76	44.15	43.40	43.40	58.65	67.42	61.45	69.52	73.98	72.49	73.51	69.06	71.54	74.72	71.91	71.80	100.00	75.09	75.09	76.58	76.21	75.46	75.84
22: A7_SAG42/1-271	36.13	44.32	41.90	43.89	42.59	43.07	41.95	41.20	60.53	68.40	71.97	70.85	73.80	74.91	73.70	71.91	71.27	73.80	71.00	75.00	75.09	100.00	100.00	85.98	85.61	83.39	83.76
23: A6_SAG43/1-271	36.13	44.32	41.90	43.89	42.59	43.07	41.95	41.20	60.53	68.40	71.97	70.85	73.80	74.91	73.70	71.91	71.27	73.80	71.00	75.00	75.09	100.00	100.00	85.98	85.61	83.39	83.76
24: A5_SAG19/1-271	36.13	45.45	43.87	44.66	43.73	44.19	43.07	42.70	60.90	69.89	71.97	69.37	73.80	73.43	72.59	74.16	73.88	72.69	74.35	75.37	76.58	85.98	85.98	100.00	99.63	97.42	97.79
25: ETH_26040/1-271	35.71	45.45	43.87	44.27	43.73	44.19	43.07	42.70	60.53	69.52	72.35	69.37	73.80	73.80	72.22	74.16	73.51	72.69	74.35	75.37	76.21	85.61	85.61	99.63	100.00	97.79	97.42
26: A4_SAG62/1-271	35.71	45.08	43.87	43.89	44.87	44.94	43.82	43.82	60.15	68.77	71.97	69.00	73.06	73.43	72.22	73.78	73.13	72.32	74.35	74.25	75.46	83.39	83.39	97.42	97.79	100.00	99.63
27: ETH_26045/1-271	36.13	45.08	43.87	44.27	44.87	44.94	43.82	43.82	60.53	69.14	71.59	69.00	73.06	73.06	72.59	73.78	73.51	72.32	74.35	74.25	75.84	83.76	83.76	97.79	97.42	99.63	100.00

### A.3 Eimeria tenella SAG-A family alignment (51 members) continues pages 203-206

tr A0A089WXE4 A0A089WXE4_EIMTE/1-253	1 MARLS ----- FVS ----- PLSLSLLFGQQA ----- VR ----- A 23
tr A0A089YJ3 A0A089YJ3_EIMTE/1-253	1 MARLS ----- FVS ----- LLSLSLLFGQQA ----- VR ----- A 23
tr H989Z8 H989Z8_EIMTE/1-264	1 MAGFW ----- FLA ----- LASTLLLPKV ----- SS ----- A 21
tr Q70CC1 Q70CC1_EIMTE/1-255	1 MFRLL ----- SIP ----- LVLPVPFYGHKA ----- AS ----- A 23
tr Q70CD4 Q70CD4_EIMTE/1-270	1 MSRLA ----- SVP ----- IFVSLFCGHEA ----- VS ----- A 23
tr Q70CD9 Q70CD9_EIMTE/1-253	1 MARVA ----- FFS ----- LVFVPLVFNQA ----- I ----- A 21
tr U6KUL8 U6KUL8_EIMTE/1-264	1 MAGFW ----- FLA ----- LASTLLLPKV ----- SS ----- A 21
tr U6KV39 U6KV39_EIMTE/1-256	1 MFRLL ----- FVP ----- IFVSVFCGHEA ----- VS ----- A 23
tr U6KV43 U6KV43_EIMTE/1-254	1 ---- LI ----- CIA ----- LSVSVFYGHQT ----- AS ----- A 20
tr U6KV49 U6KV49_EIMTE/1-257	1 MLCLA ----- FIP ----- LFVVLGFCGGR ----- ES ----- A 23
tr U6KV63 U6KV63_EIMTE/1-260	1 MLRLV ----- HMS ----- IVFASVFCGQ ----- T ----- ES ----- A 22
tr U6KV71 U6KV71_EIMTE/1-260	1 MFRLL ----- SVP ----- IFVSLFCGHET ----- VS ----- A 23
tr U6KVJ2 U6KVJ2_EIMTE/1-251	1 ---- ---- IP ----- LVLPVPICGHHK ----- AS ----- A 17
tr U6KVJ6 U6KVJ6_EIMTE/1-246	1 MIYLA ----- FIA ----- LAAYTVFCGHKA ----- EG ----- A 23
tr U6KVK9 U6KVK9_EIMTE/1-262	1 MTRLG ----- FVS ----- LFTVSLLLRQDA ----- VR ----- A 23
tr U6KYI6 U6KYI6_EIMTE/1-255	1 MFRFA ----- LVP ----- IFVSVFCGHEA ----- VG ----- A 23
tr U6KYJ0 U6KYJ0_EIMTE/1-257	1 MLHFV ----- RVA ----- LAVSVFYGHQA ----- AS ----- A 23
tr U6KYJ3 U6KYJ3_EIMTE/1-257	1 MLCLV ----- FP ----- CLVSVFCGHRT ----- AG ----- A 22
tr U6KYK1 U6KYK1_EIMTE/1-257	1 MLRLW ----- FIA ----- SVVMSVFCGHKA ----- DG ----- A 23
tr U6L2K4 U6L2K4_EIMTE/1-253	1 MARLS ----- FVS ----- LLSLSLLFGQQA ----- VR ----- A 23
tr H98A19 H98A19_EIMTE/1-317	63 MI RLV ----- IVS ----- LAAYLAFCEHGT ----- AS ----- A 85
tr U6L2K7 U6L2K7_EIMTE/1-255	1 MI RLV ----- IVS ----- LAAYLAFCEHGT ----- AS ----- A 23
tr Q70CE0 Q70CE0_EIMTE/1-260	1 MWLPS ----- PLA ----- LVTTSLLLVAF ----- SSGSDGTPEA ----- S 30
tr Q70CE1 Q70CE1_EIMTE/1-253	1 ML ----- SLV ----- ICASLLVLVSAS ----- LLKISQA ----- GQ ----- Q 27
tr Q70CC2 Q70CC2_EIMTE/1-264	1 MLQRR ----- LPP ----- LLRISFLMLSAT ----- TLGNSQE ----- TE ----- E 30
tr H98918 H98918_EIMTE/1-264	7 MFRMN ----- LAT ----- FLSVSLW ----- L ----- SEKSSQAAAA ----- T 34
tr Q70CD6 Q70CD6_EIMTE/1-263	1 ML SVA ----- FAS ----- FLSASLL ----- L ----- SENSSQAIAS ----- A 28
tr U6KR63 U6KR63_EIMTE/1-260	1 MLQRR ----- LPP ----- LLRISFLMLSAT ----- TLGNSQE ----- TE ----- E 30
tr U6KR66 U6KR66_EIMTE/1-264	7 MFRMN ----- LAT ----- FLSVSLW ----- L ----- SEKSSQAAAA ----- T 34
tr U6KR45 U6KR45_EIMTE/1-253	1 MASFL ----- KCI ----- CLASALLVLRAS ----- AEVEQPPSG ----- S 30
tr H98901 H98901_EIMTE/1-258	1 MLP - R ----- ITA ----- LLSTLLVLVSAS ----- TPGGSSQSGG ----- T 30
tr H98926 H98926_EIMTE/1-262	4 MLP - R ----- IGA ----- LLSTLLVLVSAP ----- IAGGSSQSET ----- T 33
tr H989J1 H989J1_EIMTE/1-254	4 LTLT ----- LFS ----- AFASVFVLSRA ----- QEEEGSIP ----- E 32
tr Q70CD7 Q70CD7_EIMTE/1-265	1 MLHWS ----- LPL ----- VLICWLLVLVSAS ----- ALGASPG ----- GG ----- P 30
tr Q70CD8 Q70CD8_EIMTE/1-256	1 MLPLR ----- IPS ----- VFSASIFLLSVS ----- YLGTSQQ ----- SA ----- P 30
tr Q70CE2 Q70CE2_EIMTE/1-257	1 MLHRN ----- IRK ----- LLSASLLVLVSAS ----- ILGSSQG ----- NQ ----- P 30
tr U6KR57 U6KR57_EIMTE/1-257	1 MPHWW ----- FQR ----- LLSASLLVLVSAS ----- RWGSSQE ----- TP ----- P 30
tr U6KRJ2 U6KRJ2_EIMTE/1-258	1 ---- ---- RLP ----- VFASLLLA ----- A ----- TENAGQTIDV ----- E 24
tr U6KRK3 U6KRK3_EIMTE/1-256	1 MLRRS ----- LPT ----- LSTSLFMLSAA ----- TLGSSQN ----- GQ ----- A 30
tr U6KRMO U6KRMO_EIMTE/1-262	1 MFRVN ----- LAS ----- FLGVSLLV ----- L ----- SEKSSQAAAA ----- T 28
tr U6KUL4 U6KUL4_EIMTE/1-262	1 MFHWG ----- LPL ----- LLSLFFVLSTS ----- TLGSSPA ----- TG ----- P 30
tr U6KUM5 U6KUM5_EIMTE/1-254	1 MVRLS ----- FV ----- FLSASLL ----- L ----- CDNNASAVAGQ ----- N 27
tr U6KW71 U6KW71_EIMTE/1-258	1 MTRFN ----- LFS ----- GVIGPLLLLSGA ----- SVGTGEPADQ ----- Q 31
tr U6KW76 U6KW76_EIMTE/1-260	1 MLP - R ----- INA ----- LVSTLLVLVSNS ----- ISGSLQQGGT ----- A 30
tr U6KW95 U6KW95_EIMTE/1-258	1 MLHWN ----- LPP ----- LLSLFFVLSTS ----- KLGSSQT ----- SA ----- P 30
tr U6KYM1 U6KYM1_EIMTE/1-258	1 MLP - R ----- ITA ----- LLSTLLVLVSAS ----- TPGGSSQSGG ----- T 30
tr U6KYM8 U6KYM8_EIMTE/1-262	1 MFHWG ----- LPL ----- LLSLFFVLSTS ----- TLGSSPA ----- TG ----- P 30
tr U6L7N4 U6L7N4_EIMTE/1-238	1 MRSLK ----- LLF ----- LAGSAVFVRDT ----- QGAQGSSG ----- P 28
tr U6L7Q0 U6L7Q0_EIMTE/1-257	1 MAPSL ----- KFF ----- YLANACILLLAACPRAQSEPADS ----- P 33
tr U6LAF7 U6LAF7_EIMTE/1-254	1 MASFL ----- KCI ----- CLASALLVLRAS ----- AEVEQPPSG ----- S 30
tr U6KRL3 U6KRL3_EIMTE/1-248	1 MLQWN ----- LSP ----- LLRISLLMLGAT ----- TLGSSQE ----- TE ----- E 30
tr A0A089WXE4 A0A089WXE4_EIMTE/1-253	24 QD --- YPTAVTLDGKEAMNKLKKA ----- AGLPAFEDAVGDTFVLPAYSHEESR ----- 69
tr A0A089YJ3 A0A089YJ3_EIMTE/1-253	24 QN --- YPTAVTLDGKEAMNKLKKA ----- AGLPAFEDAVGDTFVLPAYSHEESR ----- 69
tr H989Z8 H989Z8_EIMTE/1-264	22 SE --- DDSEI EVVDGMSLELNKVR ----- VGLPKFSTFKDRT ----- PNYGDEEAN ----- 64
tr Q70CC1 Q70CC1_EIMTE/1-255	24 GT --- TGTAEITISGLEMNVRAA ----- AGLAAFKGATDATQVLPKPTPGT ----- 66
tr Q70CD4 Q70CD4_EIMTE/1-270	24 AA --- TGTAAALDGLAEMNEVRAA ----- AGLAAFGNATADVQLPESSSS ----- 66
tr Q70CD9 Q70CD9_EIMTE/1-253	22 QQAATPDAAKLLNGLAEMNALRTA ----- AGLAEFKGATSTATQILPEK ----- AVEK ----- 68
tr U6KUL8 U6KUL8_EIMTE/1-264	22 SE --- DDSEI EVVDGMSLELNKVR ----- VGLPKFSTFKDRT ----- PNYGDEEAN ----- 64
tr U6KV39 U6KV39_EIMTE/1-256	24 AV --- TATAVTLDGLAEMNEVRET ----- AGFAAFGNATANEQVLPQHPVQ ----- 66
tr U6KV43 U6KV43_EIMTE/1-254	21 TT --- TGKAEITLDGLADMNEVRTV ----- AGLAKFGEPSSDKQLPNNPTL ----- 63
tr U6KV49 U6KV49_EIMTE/1-257	24 TI --- TGKAEITLDGLTHMNEARRA ----- AGLPAFGETAEEGKTLPEHTES ----- 66
tr U6KV63 U6KV63_EIMTE/1-260	23 TT --- TGKAEITLDGLGEMNEARRA ----- AGLPKFEEATQVGQILPEHSSS ----- 65
tr U6KVJ2 U6KVJ2_EIMTE/1-251	24 TV --- TGTAVRLDGLAEMNEVRAA ----- AGLAAFGNATADGQVLPPEPSLP ----- 66
tr U6KVJ6 U6KVJ6_EIMTE/1-246	18 GT --- TGTAEITIDFREMNEERKA ----- TGLPEFRQASEETQVLLPEHPS ----- 60
tr U6KVK9 U6KVK9_EIMTE/1-262	24 SI --- TGTAKTLDGLAAMNEART ----- AGLAEFKPAEEQGMPLPENPAA ----- 66
tr U6KYI6 U6KYI6_EIMTE/1-255	24 QDGTSPRATTLGLAEMNGLRTA ----- AGLAEFKAASNPQVLPKYEKNSQR ----- 72
tr U6KYJ0 U6KYJ0_EIMTE/1-257	24 AV --- TGTAVRLDGLAEMNEVRAA ----- AGLAAFGNATTTQVMPPEHTVQ ----- 66
tr U6KYJ3 U6KYJ3_EIMTE/1-257	24 ST --- TGTAEITLDGLADMNEART ----- AGLAKFGEPSSDKQLPESSPTL ----- 66
tr U6KYK1 U6KYK1_EIMTE/1-257	23 TI --- TGKAEITLDGLTEMNEARVA ----- AGLPVFGNATEAGQTLPEHSP ----- 65
tr U6L2K4 U6L2K4_EIMTE/1-253	24 TT --- TGTATLDGLTEMNEARVA ----- AGLSEFEKASRETEVLPPEYPT ----- 66
tr H98A19 H98A19_EIMTE/1-317	24 QN --- YPTAVTLDGKEAMNKLKKA ----- AGLPAFEDAVGDTFVLPAYSHEESR ----- 69
tr U6L2K7 U6L2K7_EIMTE/1-255	86 AT --- TGKAEITLDGLTEMNEARAA ----- AGLAGFEKAKEGGQVLPPEYSTE ----- 128
tr Q70CE0 Q70CE0_EIMTE/1-260	24 AT --- TGKAEITLDGLTEMNEARAA ----- AGLAGFEKAKEGGQVLPPEYSTE ----- 66
tr Q70CE1 Q70CE1_EIMTE/1-253	31 TVQYKATLGTGPQCFDEINAAEA ----- AGLAEFIAATSDQKLPEPGQE ----- 76
tr H98918 H98918_EIMTE/1-264	28 APTYTASLGKSNKGLSELNAAEA ----- AGLPNFTFATDGGKLSDPGE ----- 71
tr Q70CC2 Q70CC2_EIMTE/1-261	31 TVTYTASLGNNVCLSDINAAEA ----- AGLHDFGQASGDAELSI PAPE ----- 75
tr U6KR63 U6KR63_EIMTE/1-260	35 TVKYTAKLGSSVCLGEVNNAQA ----- AGLANFIKATNDGDKISDPGSA ----- 80
tr U6KR66 U6KR66_EIMTE/1-264	29 TVKYTAKLGTTVQCLSEINGAEA ----- AGLANFADAATKEEKISEPSS ----- 74
tr U6KR45 U6KR45_EIMTE/1-253	31 TVTYTASLGNNVCLSDINAAEA ----- AGLDDDFGRFS ----- DAELSDPAPE ----- 74
tr H98901 H98901_EIMTE/1-258	35 TVKYTAKLGSSVCLGEVNNAQA ----- AGLANFIKATNDGDKISDPGSA ----- 80
tr H98926 H98926_EIMTE/1-262	31 TISYSVKLGSDGCLAEVNNAEA ----- AKLSKLQPPGSGEGAKRLP ----- 73
tr H989J1 H989J1_EIMTE/1-254	31 TVKYTASLGNSTQCLDEINAAEA ----- AGLSKFAEAI ----- TNNKLDPGSA ----- 75
tr Q70CD7 Q70CD7_EIMTE/1-265	34 TVTYTASVGDQAQCLDEVNAVREV ----- AGLSNFTQAS ----- DPNKLQTPGTE ----- 78
tr Q70CD8 Q70CD8_EIMTE/1-256	33 SIQLTVKFESGMLLEEINAAARK ----- AGFSNFTQPA ----- SGSDAALPSE ----- 76
tr Q70CE2 Q70CE2_EIMTE/1-257	31 TIKYTASLGQSVCLSEVNNAEA ----- AGLPNFTEAS ----- DNNKLTPDNGT ----- 75
tr U6KR57 U6KR57_EIMTE/1-257	31 TIKYTASRGENVICLGEVNNAEA ----- AGLKNFAEAT ----- NDKKLSAPS ----- D ----- 74
tr U6KRJ2 U6KRJ2_EIMTE/1-258	31 TIKYTASRGGSVCLSEINAVREA ----- AGLPNFGEAS ----- GDGQLSDPGTS ----- 75
tr U6KRK3 U6KRK3_EIMTE/1-256	25 PVSYTVSLGRDGLCLGEVNNAALA ----- AGLPNFGEAS ----- GDGQLSDPGTS ----- 75
tr U6KRMO U6KRMO_EIMTE/1-262	21 TIQYTASLSENAECLSEVNNAEA ----- AGLPKFKDAEA ----- GENMLVAPSA ----- 75
tr U6KUL4 U6KUL4_EIMTE/1-262	29 TNYNTAKIGGSVCLGEVNNAEA ----- AGLPNFKDAEA ----- KETEKETLDPGST ----- 74
tr U6KUM5 U6KUM5_EIMTE/1-254	31 TIKYTASLGEDVCLSEVNNAEA ----- AGLTNFTTAS ----- DGTKLTAPDAS ----- 75
tr U6KW71 U6KW71_EIMTE/1-258	28 TVTYTAKLGGTVQCLREVNNAEL ----- GHFTFATTEDEKKISDPGEG ----- 70
tr U6KW76 U6KW76_EIMTE/1-260	32 TIKYAVKLGENAQCLDEVNAEA ----- AALAKFTGAADDEKRLPTPSA ----- 76
tr U6KW95 U6KW95_EIMTE/1-258	31 TIKYATLGGATVQCLNEINTARE ----- AGLPNFTKAT ----- DSNALSPGEQ ----- 75
tr U6KYM1 U6KYM1_EIMTE/1-258	31 TVKYTATLGENVCLSEFNAAEA ----- AGLANFIKAP ----- KENELTDPQT ----- 74
tr U6KYM8 U6KYM8_EIMTE/1-262	31 TVKYTASLGNSTQCLDEINAAEA ----- AGLSKFAEAI ----- TNNKLDPGSA ----- 75
tr U6L7N4 U6L7N4_EIMTE/1-238	31 TIKYTASLGEDVCLSEVNNAEA ----- AGLTNFTTAS ----- DGTKLTAPDAS ----- 75
tr U6L7Q0 U6L7Q0_EIMTE/1-257	29 PP --- STVAASIDCTAMNGIRAE ----- VGLAPLQLGTTQESKLPAKEE ----- 72
tr U6LAF7 U6LAF7_EIMTE/1-254	34 TISYSVNLGDDGCLAEVNNAEA ----- AKLSKLQPPGSGEGAKRLP ----- 76
tr U6KRL3 U6KRL3_EIMTE/1-248	31 TISYSVNLGSDGCLAEVNNAEA ----- AKLSKLQPPGSGEGAKRLP ----- 73
	31 TVTYTVTLGNFCLSEVNNAEA ----- AGLANFKEMG ----- EEDKLSSQPPE ----- 74



```

tr|A0A089WXE4|A0A089WXE4_EIMTE/1-253 70 -----AAPVAETLWKTEI|CPKVLGGGRS----- 92
tr|A0A089YJ3|A0A089YJ3_EIMTE/1-253 70 -----AAPVAETLWKTEI|CPKVLGGGRS----- 92
tr|H989Z8|H989Z8_EIMTE/1-264 65 -----DDVALCSALEKGGD----- 78
tr|Q70CC1|Q70CC1_EIMTE/1-255 67 -----ERTITPATLWN-E|CKVIAKEAE----- 88
tr|Q70CD4|Q70CD4_EIMTE/1-270 67 -----TKTITPATLWA-E|CKMKMGDQT----- 88
tr|Q70CD9|Q70CD9_EIMTE/1-253 69 -----DATVKPGTLWE-E|CPKVRGTET----- 90
tr|U6KUL8|U6KUL8_EIMTE/1-264 65 -----DDVALCSALEKGGD----- 78
tr|U6KV39|U6KV39_EIMTE/1-256 67 -----GKAIITAANLWT-E|CEKMTGGQT----- 88
tr|U6KV43|U6KV43_EIMTE/1-254 64 -----KETVNATLWN-E|CTIITTKEG----- 85
tr|U6KV49|U6KV49_EIMTE/1-257 67 -----GRAISPKTLWK-QACQKIMGEEG----- 88
tr|U6KV63|U6KV63_EIMTE/1-260 66 -----GKDVAASTLWD-Q|CKKIMGVAG----- 87
tr|U6KV7|U6KV7_EIMTE/1-260 67 -----TKAITAEDLWN-E|CQKMTGGQT----- 88
tr|U6KVJ2|U6KVJ2_EIMTE/1-251 61 -----GRKITAATLWS-E|CKKLAGEAG----- 82
tr|U6KVJ6|U6KVJ6_EIMTE/1-246 67 -----KSDIRAADLWE-E|CQEIAGSDS----- 88
tr|U6KVK9|U6KVK9_EIMTE/1-262 73 -----DVQITAETLWG-E|CPKIATADS----- 94
tr|U6KY6|U6KY6_EIMTE/1-255 67 -----GKAKTAEDLWN-E|CEEI--GDS----- 86
tr|U6KYJ0|U6KYJ0_EIMTE/1-257 67 -----KETVNATLWN-E|CKIIAEEEE----- 88
tr|U6KYJ3|U6KYJ3_EIMTE/1-252 66 -----VLQITAETLWG-Q|CKKIFRS----- 85
tr|U6KYK1|U6KYK1_EIMTE/1-257 67 -----TKKISAENLWK-QTCQLMGENV----- 88
tr|U6L2K4|U6L2K4_EIMTE/1-253 70 -----AAPVAETLWKTEI|CPKVLGGGRS----- 92
tr|H98A19|H98A19_EIMTE/1-317 129 -----ERQIVASDLWN-E|CQNLVGEES----- 150
tr|U6L2K7|U6L2K7_EIMTE/1-255 67 -----ERQIVASDLWN-E|CQNLVGEES----- 88
tr|Q70CE0|Q70CE0_EIMTE/1-260 77 -----ALA-DGKWK-DLCEYLIPMQV-ETL----- 98
tr|Q70CE1|Q70CE1_EIMTE/1-253 72 -----QLQEGSEWM-KVC|KHLVPTKEQDPV----- 95
tr|Q70CC2|Q70CC2_EIMTE/1-261 76 -----EGELSEKWK-KLCEYLIPQET--A----- 97
tr|H98918|H98918_EIMTE/1-264 81 -----DLT-DGDWK-E|CEYLVPQTQ-FAH----- 102
tr|Q70CD6|Q70CD6_EIMTE/1-263 75 -----EL--ESEWK-KLCEYLIPPPS-QEE----- 95
tr|U6KR63|U6KR63_EIMTE/1-260 75 -----EGELNEKWR-KFCEHVIPTQET--A----- 96
tr|U6KR66|U6KR66_EIMTE/1-264 81 -----DLT-DGDWK-E|CEYLVPQTQ-FAH----- 102
tr|U6KR45|U6KR45_EIMTE/1-253 74 -----GVEPVTWE-PLCKELIPPAQ-GTP----- 96
tr|H98901|H98901_EIMTE/1-258 76 -----QLTEDSEWL-KLCKHLVPTKEASDQ----- 99
tr|H98926|H98926_EIMTE/1-262 79 -----GLQENTEWK-KLCEHLIATQKAKAP----- 102
tr|H989J1|H989J1_EIMTE/1-254 77 -----DNVELWY-PVGLAATPGGTI----- 95
tr|Q70CD7|Q70CD7_EIMTE/1-265 76 -----ELDDANDWK-KVCTHLIPTQAPKTM----- 99
tr|Q70CD8|Q70CD8_EIMTE/1-256 75 -----DLENDTEWK-KVCEHLIPTQ-KEPV----- 97
tr|Q70CE2|Q70CE2_EIMTE/1-257 76 -----ELQPDTEWR-KLCEYLIPQ-TEPA----- 98
tr|U6KR57|U6KR57_EIMTE/1-257 76 -----ELGKDTWR-KLCEYLIPQ-TEPV----- 98
tr|U6KRJ2|U6KRJ2_EIMTE/1-258 70 -----DPVDPWK-PLCEGLIRMPEDNTS----- 91
tr|U6KRK3|U6KRK3_EIMTE/1-256 76 -----LQEGPWE-DLCTYLIPTQQTQTV----- 97
tr|U6KRM0|U6KRM0_EIMTE/1-262 75 -----ELP-EGWK-DLCEHLIPAPP-EAY----- 96
tr|U6KUL4|U6KUL4_EIMTE/1-262 76 -----ELQDSEWK-KVCTHLIPTQETAAM----- 99
tr|U6KUM5|U6KUM5_EIMTE/1-254 71 -----ELA-ESWK-DLCEYLIPPS-ETY----- 92
tr|U6KW71|U6KW71_EIMTE/1-258 77 -----ELEEGWK-KLCEYLIPTE-DNG----- 97
tr|U6KW76|U6KW76_EIMTE/1-260 76 -----ELQDSEWR-KLCEHLIPTQEA1YQ----- 99
tr|U6KW95|U6KW95_EIMTE/1-258 75 -----ELEEEWK-KVCTHLIPSEETEGA----- 96
tr|U6KYM1|U6KYM1_EIMTE/1-258 76 -----QLTEDSEWL-KLCKHLVPTKEASDQ----- 99
tr|U6KYM8|U6KYM8_EIMTE/1-262 76 -----ELQDSEWK-KVCTHLIPTQETAAM----- 99
tr|U6L7N4|U6L7N4_EIMTE/1-238 73 -----S-PNLAFST-AVCTAARAGTL----- 91
tr|U6L7Q0|U6L7Q0_EIMTE/1-257 77 -----ALVPVTWE-PLCKELIPPAQ-GKS----- 99
tr|U6LAF7|U6LAF7_EIMTE/1-254 74 -----GVEPVTWE-PLCKELIPPAQ-GTP----- 96
tr|U6KRL3|U6KRL3_EIMTE/1-248 75 -----DLEDGSQWK-KLCEHLIPKETET--A----- 96

tr|A0A089WXE4|A0A089WXE4_EIMTE/1-253 93 -----RNVTEAVKLTGNFAYYPVTDGKKE|SDAVE 122
tr|A0A089YJ3|A0A089YJ3_EIMTE/1-253 93 -----RNVTEAVKLTGNFAYYPVTDGKKE|SDAVE 122
tr|H989Z8|H989Z8_EIMTE/1-264 79 -----PHVSMP|TLGVSLAAMVQEGETRD|CAAAVR 107
tr|Q70CC1|Q70CC1_EIMTE/1-255 89 -----ESEEAKKLQGTFAYYR--GEND|CAAAVQ 114
tr|Q70CD4|Q70CD4_EIMTE/1-270 89 -----EGTEAKKLQGTFAYYP--DGKD|CAAAVQ 114
tr|Q70CD9|Q70CD9_EIMTE/1-253 91 -----DNITEAKKL|TGTFAYYPVADGKKD|NAAVE 120
tr|U6KUL8|U6KUL8_EIMTE/1-264 79 -----PHVSMP|TLGVSLAAMVQEGETRD|CAAAVR 107
tr|U6KV39|U6KV39_EIMTE/1-256 89 -----EGTEAKKLQGTFAYYP--GKGD|CAAAVQ 114
tr|U6KV43|U6KV43_EIMTE/1-254 86 -----DSAEAKKLEGTFAYYP--GKGD|CAAAVQ 111
tr|U6KV49|U6KV49_EIMTE/1-257 89 -----EVTEIEQLSGIPAYYP--GNKD|CAAAVQ 114
tr|U6KV63|U6KV63_EIMTE/1-260 88 -----EITEIQLKLGMPAYYP--GKGD|CAAAVQ 113
tr|U6KV7|U6KV7_EIMTE/1-260 89 -----EGTQAKKLQGTFAYYP--DGKD|CAAAVQ 114
tr|U6KVJ2|U6KVJ2_EIMTE/1-251 83 -----DTTQAKKLEGTFAYY--GKGD|CAAAVQ 108
tr|U6KVJ6|U6KVJ6_EIMTE/1-246 89 -----DGVEAKKLKGTFLYYP--GEDD|CAAAVQ 114
tr|U6KVK9|U6KVK9_EIMTE/1-262 95 -----TRTTEASKL|TGTFAYYPVADGKND|NAAVE 124
tr|U6KY6|U6KY6_EIMTE/1-255 87 -----EGTEAHLRGTFAYYP--DSKD|CAAAVQ 112
tr|U6KYJ0|U6KYJ0_EIMTE/1-257 89 -----DSAEAKKLEGTFAYYA--GKGD|CAAAVQ 114
tr|U6KYJ3|U6KYJ3_EIMTE/1-252 86 -----SEIDQLVGTPAYYA--GNDG|CAAAVQ 109
tr|U6KYK1|U6KYK1_EIMTE/1-257 89 -----EITEAGSLVGTVAHYA--GAKD|KEAVQ 114
tr|U6L2K4|U6L2K4_EIMTE/1-253 93 -----RNVTEAVKLTGNFAYYPVTDGKKE|SDAVE 122
tr|U6L2K7|U6L2K7_EIMTE/1-255 151 -----KPKETLKL|VGTPAHYL--GEKN|CAAAVE 176
tr|Q70CE0|Q70CE0_EIMTE/1-260 89 -----KPKETLKL|VGTPAHYL--GEKN|CAAAVE 114
tr|Q70CE1|Q70CE1_EIMTE/1-253 99 -----ADQA-PTNPFSGGTYAFKSLTAEQPN|KETVD 129
tr|Q70CC2|Q70CC2_EIMTE/1-261 96 -----AAAG-ATNPFDGTYAFKSLTAAEPN|KETVD 126
tr|H98918|H98918_EIMTE/1-264 98 -----AKTS-SANPFEGTYAFKSLTAEQPN|KETID 133
tr|Q70CD6|Q70CD6_EIMTE/1-263 103 -----SSQA-ATETFKDGYAFKALTAEPN|KETID 128
tr|U6KR63|U6KR63_EIMTE/1-260 96 -----SVQA-TAEPFKDGYAFKSLTAEQPD|KETVD 126
tr|U6KR66|U6KR66_EIMTE/1-264 97 -----AKTS-SANPFEGTYAFKSLTAEQPN|KETID 127
tr|U6KR45|U6KR45_EIMTE/1-253 103 -----SSQA-ATETFKDGYAFKALTAEPN|KETID 133
tr|H98901|H98901_EIMTE/1-258 97 -----ANAK-SGNEFNSGTYAIALNSATPN|GSAAVD 127
tr|H98926|H98926_EIMTE/1-262 100 -----VDSH-ANSPFKDGYAFKSLTDEKPD|CKSI|VG 130
tr|H989J1|H989J1_EIMTE/1-254 103 -----SSSE-AENPFKDGTYAFKSLTDAQPD|CKD|TVD 133
tr|Q70CD7|Q70CD7_EIMTE/1-265 96 -----DTHRFPPGTF|AVAPQSS|TKVDC|DETVD 122
tr|Q70CD8|Q70CD8_EIMTE/1-256 100 -----AASS-SLKPFKDGTYAFKSLTAAKPD|KETVD 130
tr|Q70CE2|Q70CE2_EIMTE/1-257 98 -----EATS-GTNPFEGTYAFKSLTAAEPN|KEI|VN 128
tr|U6KR57|U6KR57_EIMTE/1-257 99 -----KAAS-AAKPFEGQIYAFKSLTAAEPN|KEAVD 129
tr|U6KRJ2|U6KRJ2_EIMTE/1-258 99 -----EAAS-AAKPFEGQTYAFKSLTAAEPN|KEAVD 129
tr|U6KRK3|U6KRK3_EIMTE/1-256 92 -----AKSTEYQ|SLGGTTYARMPVVS|TDPD|CATVN 123
tr|U6KRM0|U6KRM0_EIMTE/1-262 98 -----AART-TEDTFEKGTYAFRPLTS|TDPK|TESVE 128
tr|U6KUL4|U6KUL4_EIMTE/1-262 97 -----NAQT-SAEAFQDGYAFKSLADAPDN|KETVD 127
tr|U6KUM5|U6KUM5_EIMTE/1-254 100 -----AEPS-AASPFQDGYAFKSLTAAKPN|KETVD 130
tr|U6KW71|U6KW71_EIMTE/1-258 93 -----GAQK-AVDPFKDGTYAFKSLTAEQPN|KETID 123
tr|U6KW76|U6KW76_EIMTE/1-260 98 -----EAKV-SGSPFKSGTYAVNVLTS|ETPQ|CRETVD 128
tr|U6KW95|U6KW95_EIMTE/1-258 100 -----AATK-AENPFKDGTYAFKSLTDAQPS|CND|TVD 130
tr|U6KYM1|U6KYM1_EIMTE/1-258 97 -----AKSS-SANPFEDGTYCFKRLADAEPN|KEAVD 127
tr|U6KYM8|U6KYM8_EIMTE/1-262 100 -----VDSH-ANSPFKDGYAFKSLTDEKPD|CKSI|VG 130
tr|U6L7N4|U6L7N4_EIMTE/1-238 100 -----AEPS-AASPFQDGYAFKSLTAAKPN|KETVD 130
tr|U6L7Q0|U6L7Q0_EIMTE/1-257 92 -----PS-TASTSLDGNF|AISI|QKGSNGD|CATAVK 120
tr|U6LAF7|U6LAF7_EIMTE/1-254 100 -----ASER-TGNEFSSGTYAIALNSATPN|GSAAVD 130
tr|U6KRL3|U6KRL3_EIMTE/1-248 97 -----ANAK-SGNEFNSGTYAIALNSATPN|GSAAVD 127
97 -----AQSK-STNPFEGTYAFKSLTAEQPN|KETVA 127

```



```

tr|A0A089WXE4|A0A089WXE4_EIMTE/1-253 123 -----YWKGGLSQFN-DTIPPTFQALNDPVPVYNDRAV FVALYNPKTSPVVS CVLLQ-CPNA 177
tr|A0A089YJV3|A0A089YJV3_EIMTE/1-253 123 -----YWKGGLSQFN-DTIPPTFQALNDPVPVYNDRAA FVALYNPKTSPVVS CVLLQ-CPNA 177
tr|H989Z8|H989Z8_EIMTE/1-264 108 -----YWKDVYSQFG-SSLPPEYEVGVKPY-DDWRTV FMAFLT TQDDPAAT CVNIK-PPR 161
tr|Q70CC1|Q70CC1_EIMTE/1-255 115 -----YWKDGFSLFK-NELPPTYALNDPKIYTDQAT FVALYNPQASPVAS CAFVT-CTTA 169
tr|Q70CD4|Q70CD4_EIMTE/1-270 115 -----YWKEGFSLFK-NELPPTYTASNTPAVYTDRAV FVALYNPQSPPLAS CALVT-CTTA 169
tr|Q70CD9|Q70CD9_EIMTE/1-253 121 -----YWKGGFSLFK-NEIPPEFT EANKTTVYNDRAV FVALYNPKPDPVVS CVLLQ-CPNA 175
tr|U6KUL8|U6KUL8_EIMTE/1-264 108 -----YWKDVYSQFG-SSLPPEYEVGVKPY-DDWRTV FMAFLT TQDDPAAT CVNIK-PPR 161
tr|U6KV39|U6KV39_EIMTE/1-256 115 -----HWKEGFSLFK-NELPKPYTASSTPAVYTDKAV FVALYNPQSPPLAS CSLVT-CTKG 169
tr|U6KV43|U6KV43_EIMTE/1-254 112 -----YWKDGFSLFK-NELPPTYAAPNEPDVYTDRAV FVALYNPKTSPVAS CVFVT-CTTA 166
tr|U6KV49|U6KV49_EIMTE/1-257 115 -----YWKEGFSLFK-NEIPPTYQEIGNPDVYTDKGI FVALYNPQVNPVAS CAFVT-CTKA 169
tr|U6KV63|U6KV63_EIMTE/1-260 114 -----YWKDGFSLFK-NQLPPTYVELGNPEVYNDRGI FVALYNPKASPVAS CAFAT-CTTT 168
tr|U6KV7|U6KV7_EIMTE/1-260 115 -----YWKEGFSLFK-NELPPTYTASNTPEVYTDRAV FVALYNPQSSPLAS CSLVT-CTKA 169
tr|U6KVJ2|U6KVJ2_EIMTE/1-251 109 -----YWKDGFSLFK-NELPPTYTALNDPQIYTDQAV FVALYNPKASPVAS CSFVT-CTGG 163
tr|U6KVJ6|U6KVJ6_EIMTE/1-246 115 -----YWKDGFSLFK-NELPKMYKVLGKPEVYTNKAV FVALYNPKADPVAS CAFVT-CTKG 169
tr|U6KVK9|U6KVK9_EIMTE/1-262 125 -----YWKGGFSLFK-DEIPPKYEGTNLPTVYSDRAV FVALYNPKPDPVVS CVLLQ-CPA 179
tr|U6KY6|U6KY6_EIMTE/1-255 113 -----YWKDGFSLFK-DELPPTYTEPDNPVYVTEKAV FVALYNPQASPVAR CALIT-CTQG 167
tr|U6KYJ0|U6KYJ0_EIMTE/1-257 115 -----YWKDGFSLFK-NELPPAYTTSNQSEVYTDRAV FVSLYNPQASPVAS CVFVT-CTTA 169
tr|U6KYJ3|U6KYJ3_EIMTE/1-252 110 -----YWKDGSLSLK-NEIPPTYQALNDPDPVYTDKAV FVALYNPQANPVAS CASVT-CTKA 164
tr|U6KYK1|U6KYK1_EIMTE/1-257 115 -----YWKDGFSLFK-NELPKPYTALGDPDVYTDRAV FVALYNPKASPVAS CAFVT-CTKG 169
tr|U6L2K4|U6L2K4_EIMTE/1-253 123 -----YWKGGLSQFN-DTIPPTFQALNDPVPVYNDRAV FVALYNPKTSPVVS CVLLQ-CPNA 177
tr|H98A19|H98A19_EIMTE/1-317 177 -----YWNEGFSLFK-NEIPPAYTASSRDPVYNDKAV FVALYSPKESPVAS CVFVT-CTTP 231
tr|U6L2K7|U6L2K7_EIMTE/1-255 115 -----YWNEGFSLFK-NEIPPAYTASSRDPVYNDKAV FVALYSPKESPVAS CAFVT-CTTP 169
tr|Q70CE0|Q70CE0_EIMTE/1-260 130 -----YWKAAFKNFT-GLPPSK--NDAGELYN SQDNV FVALYNPSSANASAD CRVVT-CTK 182
tr|Q70CE1|Q70CE1_EIMTE/1-253 127 -----HWKAAEFNFT-GLPPSK--TEGANLYKKNQDNV FVALYNPSSDATAD CRVVT-CTKA 179
tr|Q70CC2|Q70CC2_EIMTE/1-261 129 -----YWKAAEFNFT-GLPPSK--KEGGTLYDDQDNV FVAVYNPSSSATAD CRVVT-CTQT 181
tr|H98918|H98918_EIMTE/1-264 134 -----YWKAAEFNFT-GLPPSK--KGTGKLYENQDNV FVALYNPSSSNATAD CRVVT-CTQK 186
tr|Q70CD6|Q70CD6_EIMTE/1-263 127 -----YWKAAEFNFT-GLPPSK--SEDTSIYSKQDNV FVALYNPISKATAD CRVVT-CTQT 179
tr|U6KR63|U6KR63_EIMTE/1-260 128 -----YWKAAEFNFT-GLPPSK--KEGGTLYDDQDNV FVAVYNPSSSATAD CRVVT-CTQT 180
tr|U6KR66|U6KR66_EIMTE/1-264 134 -----YWKAAEFNFT-GLPPSK--KGAGKLYENQDNV FVALYNPSSSNATAD CRVVT-CTQK 186
tr|U6KR45|U6KR45_EIMTE/1-253 128 -----HWKAAEFNFT-GLPPSK--GVPAY--KDNNTLYSKQDNV FVALYNPSSKDATAD CRVVT-CTET 180
tr|H98901|H98901_EIMTE/1-258 131 -----SWKAAEFNFT-GLPPSK--NQAGLYNNQDNV FVALYNPQSSATAD CRVVT-CTKT 183
tr|H98926|H98926_EIMTE/1-262 134 -----HWKAAEFNFT-GLPPSK--KQANGLYDKQDNV FVALYNPSSSATAD CRVVT-CTKK 186
tr|H989J1|H989J1_EIMTE/1-254 123 -----HWKAAEFNFT-GPPPK--VASS E IYNNLNDL FVAMYNPSSADATAD CRVMT-CTQT 175
tr|Q70CD7|Q70CD7_EIMTE/1-265 131 -----YWKAAEFNFT-GLPPSK--TDNETLYKKNQDNV FVSLYNPSSSATAD CRVVT-CTQT 183
tr|Q70CD8|Q70CD8_EIMTE/1-256 129 -----YWKAAEFNFT-GLPPSK--SQAGDLYKSYNNV FVALYNTSSSNATAD CRVVT-CTKT 181
tr|Q70CE2|Q70CE2_EIMTE/1-257 130 -----YWKSAFKNFT-GLPPSK--SQAGLYNNQDNV FVALYNPSSSATAD CRVIT-CTKT 182
tr|U6KR57|U6KR57_EIMTE/1-257 130 -----YWKSAFKNFT-GLPPSK--SQAGLYNNQDNV FVALYNPSSSATAD CRVIT-CTKT 182
tr|U6KRJ2|U6KRJ2_EIMTE/1-258 124 -----HWKAAEFNFT-GLPPSK--EKADLYGDPDNV FVALYNPSSPNAYAD CRVVT-CNEL 176
tr|U6KRK3|U6KRK3_EIMTE/1-256 129 -----FWESA FKNFT-GLPPSK--SKGALYGI RENV FIALYNAQTDATVD CRVVT-CTQT 181
tr|U6KRMO|U6KRMO 128 -----YWKAAEFNFT-GLPPSK--SKDEKLYNNQDNV FVALYNPSSSNATAD CRVVT-CTQT 180
tr|U6KUL4|U6KUL4_EIMTE/1-262 131 -----SWKAAEFNFT-GLPPSK--TEDENLYKKNQDNV FVALFNPSSSRATAD CRVVT-CTQT 183
tr|U6KUM5|U6KUM5_EIMTE/1-254 124 -----YWKAAEFNFT-GLPPAK--GEAGTLYDNQDNV FVALYNPSSSATAD CRVVT-CTQK 176
tr|U6KW71|U6KW71_EIMTE/1-258 129 -----YWKAAEFNFT-GLPPSK--SQDPEIYRKQDNV FVALYNPSSSTEGAD CRVAT-CTKT 181
tr|U6KW76|U6KW76_EIMTE/1-260 131 -----YWKAAEFNFT-GLPPSK--SQANGLYDKQDNV FVALYNPSSSATAD CRVVT-CTKT 183
tr|U6KW95|U6KW95_EIMTE/1-258 128 -----YWKAAEFNFT-GLPPSK--TEGTDLYKKNQDNV FVALYNPSSSNATAD CRVVT-CTQT 180
tr|U6KYM1|U6KYM1_EIMTE/1-258 131 -----YWKAAEFNFT-GLPPSK--NQAGLYNNQDNV FVALYNPQSSATAD CRVAT-CTKT 183
tr|U6KYM8|U6KYM8_EIMTE/1-262 131 -----YWKAAEFNFT-GLPPSK--TEDENLYKKNQDNV FVALFNPSSSRATAD CRVVT-CTQT 183
tr|U6L7N4|U6L7N4_EIMTE/1-238 121 -----HWREGDFLKG-GLPPPY--SAEAPVYKTSQAQ FVALFAPHEAGKVDCAFFI-CPES 173
tr|U6L7Q0|U6L7Q0_EIMTE/1-257 131 -----HWKAAEFNFT-GLPPSK--KDNNTLYSKQDNV FVALYNPSSKDATAD CRVVT-CTET 183
tr|U6LAF7|U6LAF7_EIMTE/1-254 128 -----HWKAAEFNFT-GLPPSK--KDNNTLYSKQDNV FVALYNPSSKDATAD CRVVT-CTET 180
tr|U6KRL3|U6KRL3_EIMTE/1-248 128 -----YWKAAEFNFT-GLPPSK--KEAGTLYNKNQDNV FVAVYNPSSSNATAD CRVVT-CTQT 180

```

```

tr|A0A089WXE4|A0A089WXE4_EIMTE/1-253 178 GV----- 179
tr|A0A089YJV3|A0A089YJV3_EIMTE/1-253 178 GV----- 179
tr|H989Z8|H989Z8_EIMTE/1-264 162 PSNEEPTGPGVSGPPTTMRNPSQDGE----- 188
tr|Q70CC1|Q70CC1_EIMTE/1-255 170 TES--TAPDKGLST----- 181
tr|Q70CD4|Q70CD4_EIMTE/1-270 170 AAL--AAQDPLKSSGGP----- 184
tr|Q70CD9|Q70CD9_EIMTE/1-253 176 TSPGVPG----- 182
tr|U6KUL8|U6KUL8_EIMTE/1-264 162 PSNEEPTGPGVSGPPTTMRNPSQDGE----- 188
tr|U6KV39|U6KV39_EIMTE/1-256 170 GTSASPEIAESQKGR----- 184
tr|U6KV43|U6KV43_EIMTE/1-254 167 SGFASPLSKGEEGP----- 181
tr|U6KV49|U6KV49_EIMTE/1-257 170 TASADQDTS TSQDGH----- 184
tr|U6KV63|U6KV63_EIMTE/1-260 169 GAGPAAAAA QPKSQGGH----- 185
tr|U6KV7|U6KV7_EIMTE/1-260 170 AAL--ASPRMLKSHDVR----- 184
tr|U6KVJ2|U6KVJ2_EIMTE/1-251 164 AVF--AAPDL SKSHERR----- 178
tr|U6KVJ6|U6KVJ6_EIMTE/1-246 170 DVF--TAAALSKSHGGP----- 184
tr|U6KVK9|U6KVK9_EIMTE/1-262 180 SAAAPSVPGVPG----- 191
tr|U6KY6|U6KY6_EIMTE/1-255 168 GAF--AARRML----- 176
tr|U6KYJ0|U6KYJ0_EIMTE/1-257 170 SGF--AASALP----- 178
tr|U6KYJ3|U6KYJ3_EIMTE/1-252 165 TVF--AAQDMS----- 173
tr|U6KYK1|U6KYK1_EIMTE/1-257 170 TAV--AAQEMS----- 178
tr|U6L2K4|U6L2K4_EIMTE/1-253 178 GV----- 179
tr|H98A19|H98A19_EIMTE/1-317 232 GANATAPPAQPEDPD SG----- 248
tr|U6L2K7|U6L2K7_EIMTE/1-255 170 GANATAPPAQPEDPD SG----- 186
tr|Q70CE0|Q70CE0_EIMTE/1-260 183 TGVA SKSASTSADTGSA----- 199
tr|Q70CE1|Q70CE1_EIMTE/1-253 180 TASEAALQSDSDQ----- 192
tr|Q70CC2|Q70CC2_EIMTE/1-261 182 NTTTTTTP-GPTRVQADGGSGE----- 200
tr|H98918|H98918_EIMTE/1-264 187 TSAAAVSASVSSGDSGD----- 203
tr|Q70CD6|Q70CD6_EIMTE/1-263 180 ITAASVPAAAAAQGDGD--SEADTR----- 202
tr|U6KR63|U6KR63_EIMTE/1-260 181 NTTTTTTP-GPTRVQADGGSGE----- 199
tr|U6KR66|U6KR66_EIMTE/1-264 187 TSAAAVSASVSSGDSGD----- 203
tr|U6KR45|U6KR45_EIMTE/1-253 181 TTPQSP----- 186
tr|H98901|H98901_EIMTE/1-258 184 TSFAPS SFSDDTSA----- 197
tr|H98926|H98926_EIMTE/1-262 176 TSTKENELAAEAGNA----- 201
tr|H989J1|H989J1_EIMTE/1-254 176 PKATGDPDVS GRQSGDQA----- 193
tr|Q70CD7|Q70CD7_EIMTE/1-265 184 TTSTTGAGGGQLSTSAEDG----- 203
tr|Q70CD8|Q70CD8_EIMTE/1-256 182 TTPGDS SIRDSPSG----- 195
tr|Q70CE2|Q70CE2_EIMTE/1-257 183 TTTGGSEQQSDPQE----- 196
tr|U6KR57|U6KR57_EIMTE/1-257 183 TTPGESVVS SDPQG----- 196
tr|U6KRJ2|U6KRJ2_EIMTE/1-258 177 ENKPYIDLSSHTMKS NLF----- 195
tr|U6KRK3|U6KRK3_EIMTE/1-256 182 TTTGDAAEHSDPPA----- 195
tr|U6KRMO|U6KRMO 181 NTTTPQPDVTSRDNGTE----- 197
tr|U6KUL4|U6KUL4_EIMTE/1-262 184 TTSATDDAVGSTYSGSG----- 200
tr|U6KUM5|U6KUM5_EIMTE/1-254 177 ASPAAL SAGVSSGGND----- 192
tr|U6KW71|U6KW71_EIMTE/1-258 182 VSGGAGVLN----- 190
tr|U6KW76|U6KW76_EIMTE/1-260 184 TTDAGEEADE----- 193
tr|U6KW95|U6KW95_EIMTE/1-258 181 TTSGLSPATAT----- 191
tr|U6KYM1|U6KYM1_EIMTE/1-258 184 TSAPSSLS DTSAA----- 197
tr|U6KYM8|U6KYM8_EIMTE/1-262 184 TTSATDDAVGST----- 195
tr|U6L7N4|U6L7N4_EIMTE/1-238 174 TDSG TKQ----- 180
tr|U6L7Q0|U6L7Q0_EIMTE/1-257 184 TTNQSRQGGEGSG----- 196
tr|U6LAF7|U6LAF7_EIMTE/1-254 181 TTPQSPNGQRATG----- 193
tr|U6KRL3|U6KRL3_EIMTE/1-248 181 NTSTAGPVEVTTDGGSA----- 197

```



```

tr|A0A089WXE4|A0A089WXE4_EIMTE/1-253 180 -----GRRRLA--AGTTY-----AVICLTN 197
tr|A0A089YJ3|A0A089YJ3_EIMTE/1-253 180 -----GRRRLA-AG-TTY-----AVICLTN 197
tr|H989Z8|H989Z8_EIMTE/1-264 189 -----GKSRVQASSPAAQTRRL--AAQFY-----ALLCIID 217
tr|Q70CC1|Q70CC1_EIMTE/1-255 182 -----RRLQGEEDTASTSY-----AVICLTN 202
tr|Q70CD4|Q70CD4_EIMTE/1-270 185 -----SARRLQGEEDSTTTPTPTTTTAVICLTN 216
tr|Q70CD9|Q70CD9_EIMTE/1-253 183 -----AGRRLLSSSSTTVY-----AVICLTN 202
tr|U6KUL8|U6KUL8_EIMTE/1-264 189 -----GKSRVQASSPAAQTRRLAAQFY-----ALLCIID 217
tr|U6KV39|U6KV39_EIMTE/1-256 185 -----STRRLQAAQETTY-----AVICLTN 204
tr|U6KV43|U6KV43_EIMTE/1-254 182 -----TLRRLGGGTQTTY-----AVICLTN 201
tr|U6KV49|U6KV49_EIMTE/1-257 185 -----SLRRLQDEPQITY-----AVICLTN 204
tr|U6KV63|U6KV63_EIMTE/1-260 186 -----SPRRLQAEAQTLTY-----AVICLTN 205
tr|U6KV7|U6KV7_EIMTE/1-260 185 -----STRRLQGEEDPTTVTY-----AVICLTN 207
tr|U6KVJ2|U6KVJ2_EIMTE/1-251 179 -----AVRRLLEELEADAY-----AVICLTN 198
tr|U6KVJ6|U6KVJ6_EIMTE/1-246 185 -----ALRRLSGEKPTY-----AVICLTN 203
tr|U6KVK9|U6KVK9_EIMTE/1-262 192 -----SSRRLSDSSTTVY-----AVICLTN 211
tr|U6KY6|U6KY6_EIMTE/1-255 177 -----KSHEGRSTRRLQGEEDTTY-----AVICLTN 202
tr|U6KYJ0|U6KYJ0_EIMTE/1-252 179 -----KTDERRTLRLQGEEDPAY-----AVICLTN 204
tr|U6KYJ3|U6KYJ3_EIMTE/1-252 174 -----RSPNGQSPRRLQDGNQATY-----AVICLTN 199
tr|U6KYK1|U6KYK1_EIMTE/1-257 179 -----RRHDSPLRRLQDGAQTKY-----AVICLTN 204
tr|U6L2K4|U6L2K4_EIMTE/1-253 180 -----GRRRLAAGTTY-----AVICLTN 197
tr|H98A19|H98A19_EIMTE/1-317 249 -----PSRSLQEESKITY-----AVMCLN 268
tr|U6L2K7|U6L2K7_EIMTE/1-255 187 -----PSRSLQEESKIT-----AVMCLN 205
tr|Q70CE0|Q70CE0_EIMTE/1-260 200 -----TTKSGY-----ALLCKTV 212
tr|Q70CE1|Q70CE1_EIMTE/1-253 193 -----SSENGY-----ALLCKTM 205
tr|Q70CC2|Q70CC2_EIMTE/1-261 201 -----TTKGY-----ALLCKTM 213
tr|H98918|H98918_EIMTE/1-264 204 -----TTKLG-----ALLCKTM 216
tr|Q70CD6|Q70CD6_EIMTE/1-263 203 -----TKTNGY-----ALLCKTM 215
tr|U6KR63|U6KR63_EIMTE/1-260 200 -----TTKKG-----ALLCKTM 211
tr|U6KR66|U6KR66_EIMTE/1-264 204 -----TTKLG-----ALLCKTM 215
tr|U6KR45|U6KR45_EIMTE/1-253 187 -----QGEGSGTQRHGY-----ALLCMTT 205
tr|H98901|H98901_EIMTE/1-258 198 -----TPEKGY-----ALLCKTM 210
tr|H98926|H98926_EIMTE/1-262 202 -----TSEYGY-----ALLCKTM 214
tr|H989J1|H989J1_EIMTE/1-254 194 -----EEKQGY-----ALVCRMT 206
tr|Q70CD7|Q70CD7_EIMTE/1-265 204 -----TTKNGY-----ALLCKTM 216
tr|Q70CD8|Q70CD8_EIMTE/1-256 196 -----SQEYGY-----AMICKTM 208
tr|Q70CE2|Q70CE2_EIMTE/1-257 197 -----SKEYGY-----AMICKT 209
tr|U6KR57|U6KR57_EIMTE/1-257 197 -----SKENGY-----AMICKT 209
tr|U6KRJ2|U6KRJ2_EIMTE/1-258 196 -----TGNSA-----ALLCLIT 207
tr|U6KRK3|U6KRK3_EIMTE/1-256 196 -----NEKYG-----ALLCKTM 207
tr|U6KRM0|U6KRM0_EIMTE/1-262 198 -----ATKKG-----ALLCKTT 209
tr|U6KUL4|U6KUL4_EIMTE/1-262 201 -----TTKNG-----ALLCKTM 212
tr|U6KUM5|U6KUM5_EIMTE/1-254 193 -----ANTKLG-----AMLCCKT 205
tr|U6KW71|U6KW71_EIMTE/1-258 191 -----GGETSQPTGKG-----ALLCMTT 208
tr|U6KW76|U6KW76_EIMTE/1-260 194 -----DTRAGQSSQSG-----ALLCKTM 211
tr|U6KW95|U6KW95_EIMTE/1-258 192 -----SGGESGTTKNG-----ALLCKTM 209
tr|U6KYM1|U6KYM1_EIMTE/1-258 198 -----TPEKG-----ALLCKTM 209
tr|U6KYM8|U6KYM8_EIMTE/1-262 196 -----YSGSGTTKNG-----ALLCKTM 212
tr|U6L7N4|U6L7N4_EIMTE/1-238 181 -----EEI-----ALLCVTS 190
tr|U6L7Q0|U6L7Q0_EIMTE/1-257 197 -----TQRHG-----ALLCMTT 208
tr|U6LAF7|U6LAF7_EIMTE/1-254 194 -----TERLG-----ALLCMTT 205
tr|U6KRL3|U6KRL3_EIMTE/1-248 198 -----TTKKG-----ALLCKTM 209

tr|A0A089WXE4|A0A089WXE4_EIMTE/1-253 198 P A P L E A -----R S Q P F D D E Q W K -----K I V D S L S L S E E E E E K G G 231
tr|A0A089YJ3|A0A089YJ3_EIMTE/1-253 198 P A P L E A -----R S Q P F D D E Q W K -----K I V D S L S L S E E E E E K G G 231
tr|H989Z8|H989Z8_EIMTE/1-264 218 P V A L T Q -----G K A P F T E S Q W A -----K I R Q V L Q S S S P -----V 246
tr|Q70CC1|Q70CC1_EIMTE/1-255 203 P A A L S K -----G Q Q P F K E E V W R -----K I V Q A L V G T E E S N -----G 233
tr|Q70CD4|Q70CD4_EIMTE/1-270 217 P K A L T D -----G Q A P F K E E V W Q -----K I V Q A I V S T Q K G N -----G 247
tr|Q70CD9|Q70CD9_EIMTE/1-253 203 P A A L T E -----N A A P F K E D E W K -----K I V E A I S G N K S -----E 231
tr|U6KUL8|U6KUL8_EIMTE/1-264 218 P V A L T Q -----G K A P F T E S Q W A -----K I R Q V L Q S S S P -----V 246
tr|U6KV39|U6KV39_EIMTE/1-256 205 P E A L T E -----G Q A P F K E E V W Q -----K I V Q A V V G T Q K S N G V -----236
tr|U6KV43|U6KV43_EIMTE/1-254 202 P E A L T T -----E E A P F K E D E W Q -----K I V H A I V G T E E S N R V -----233
tr|U6KV49|U6KV49_EIMTE/1-257 205 P E A L K T -----G A A P F K E E W Q -----K I V R A I V G T E E V N R V -----236
tr|U6KV63|U6KV63_EIMTE/1-260 206 P E A L T A -----G A A P F K E D E W Q -----K I A H A M N G T E S V K G S W A 239
tr|U6KV7|U6KV7_EIMTE/1-260 208 P K A L T D -----G Q A P F K E E V W Q -----K I V Q A I V S T Q K G N G V -----239
tr|U6KVJ2|U6KVJ2_EIMTE/1-251 199 P K A L T T -----G A A P F K E D E W Q -----K I V H A I V G T E G S N G V -----230
tr|U6KVJ6|U6KVJ6_EIMTE/1-246 204 P E A F T A -----D A A P F K E E W Q -----K I V H A I V G P E A S T G A -----235
tr|U6KVK9|U6KVK9_EIMTE/1-262 212 P P A L T K -----S A A P F T E D E W N -----K I G E A L S G S K G G V -----241
tr|U6KY6|U6KY6_EIMTE/1-255 203 P V A L S K -----G Q Q P F K E E V W Q -----K I V Q A L V G A E K S N G A -----234
tr|U6KYJ0|U6KYJ0_EIMTE/1-252 205 P Q A L I T -----D T A P F K E D E W Q -----K I V H A V V G T E E R T A -----236
tr|U6KYJ3|U6KYJ3_EIMTE/1-257 200 P K A L V A -----G Q S P F K E E E W K -----K I V D A I V G T E E V N R V -----231
tr|U6KYK1|U6KYK1_EIMTE/1-257 205 P D A L A P -----G T A P F E D D V W Q -----K I V H A I V G L E E S N R A -----236
tr|U6L2K4|U6L2K4_EIMTE/1-253 198 P A P L E A -----R S Q P F D D E Q W K -----K I V D S L S L S E E E E E K G G 231
tr|H98A19|H98A19_EIMTE/1-317 269 P E A L D P -----N E A P F T E D E W K -----K I V R A L G S S G I -----296
tr|U6L2K7|U6L2K7_EIMTE/1-255 206 P E A L D P -----N E A P F T E D E W K -----K I V R A L G S S G I -----233
tr|Q70CE0|Q70CE0_EIMTE/1-260 213 P T V F A N -----D T T A P F T Q E Q W D -----K I V S S L T G S T S -----241
tr|Q70CE1|Q70CE1_EIMTE/1-253 206 P S A F P D -----D K S P F T Q D Q W N -----K I V S S L T G S A S -----234
tr|Q70CC2|Q70CC2_EIMTE/1-261 214 P T A F A S -----D T S A P F T Q A Q W D -----K I M S S L T G S G S -----242
tr|H98918|H98918_EIMTE/1-264 217 P T A F G N -----G S I A P F T Q E Q W D -----K I K Y S L T G S A S -----245
tr|Q70CD6|Q70CD6_EIMTE/1-263 216 P A A F E S -----D N T P F T E E Q W D -----K I K Y S L T G S A S -----244
tr|U6KR63|U6KR63_EIMTE/1-260 212 P T A F A S -----D T F A P F T Q A Q W D -----K I M S S L T G S G S -----240
tr|U6KR66|U6KR66_EIMTE/1-264 216 P T A F G N -----G S I A P F T Q E Q W D -----K I K Y S L T G S A S -----244
tr|U6KR45|U6KR45_EIMTE/1-253 206 P D I L K E -----N E S A P F T Q D Q W N -----A I S R S I T G S A S -----234
tr|H98901|H98901_EIMTE/1-258 211 P T A F Q N -----A E T A P F T Q D Q W N -----M I T S S L T G S T T -----239
tr|H98926|H98926_EIMTE/1-262 215 P D A L A D -----E N S A P F T Q K Q W D -----G I V S S L T G S A S -----243
tr|H989J1|H989J1_EIMTE/1-254 207 P N A L P I -----N G S G A P F T E E Q W D -----L I A F S L I G S A T -----235
tr|Q70CD7|Q70CD7_EIMTE/1-265 217 P A V F G S -----D G S A L P F T E Q W D -----K I V S S L T G S A S -----245
tr|Q70CD8|Q70CD8_EIMTE/1-256 209 P A A F A D -----K N S A P F T Q D Q W D -----R I S S L T G S A S -----237
tr|Q70CE2|Q70CE2_EIMTE/1-257 210 P T A F T N -----N D S V P F T Q D Q W D -----R I S S L T G S A S -----238
tr|U6KR57|U6KR57_EIMTE/1-257 210 P T P F T S -----N D S A P F T Q D Q W D -----G I A S S L T G S A A -----238
tr|U6KRJ2|U6KRJ2_EIMTE/1-258 208 P D V L Q A -----G E G D P F T Q K Q W D -----Q I K A S L G N S S R -----236
tr|U6KRK3|U6KRK3_EIMTE/1-256 208 P A A L P D -----D D S A P F T Q D Q W D -----R I S S L T G S S S -----236
tr|U6KRM0|U6KRM0_EIMTE/1-262 210 P A A F E N -----D A S A P F T Q E Q W D -----K I R S S L T G S A S -----238
tr|U6KUL4|U6KUL4_EIMTE/1-262 213 P A A F G T -----G S S A P F T Q E Q W E -----K I R S S V T G S A S -----241
tr|U6KUM5|U6KUM5_EIMTE/1-254 206 P T A F E N -----D S T A P F T Q E Q W D -----K I K Y S L T G S A S -----234
tr|U6KW71|U6KW71_EIMTE/1-258 209 P D V F Q D -----A S S T P F T P E Q W D -----K I I T S L T G S A S -----237
tr|U6KW76|U6KW76_EIMTE/1-260 212 P A A F A D -----E N A A P F T Q Q W D -----A I V S S L T G S A A -----240
tr|U6KW95|U6KW95_EIMTE/1-258 210 P T A F A N -----D T S A P F T Q E Q W D -----R I K S S L T G S A S -----238
tr|U6KYM1|U6KYM1_EIMTE/1-258 210 P T A F Q N -----A E T A P F T Q D Q W N -----M I T S S L T G S T T -----238
tr|U6KYM8|U6KYM8_EIMTE/1-262 213 P P A F G T -----G S S A P F T Q E Q W E -----K I V L S L T G T A S -----241
tr|U6L7N4|U6L7N4_EIMTE/1-238 191 P K S L K D -----G V A P F E Q K Q W D -----K I A A G L V G G A S -----218
tr|U6L7Q0|U6L7Q0_EIMTE/1-257 209 P D I L K E -----N E S A P F T Q D Q W N -----A I S R S I T G S A S -----237
tr|U6LAF7|U6LAF7_EIMTE/1-254 206 P D I L K E -----N E S A P F T Q D Q W N -----A I S R S I T G S A S -----234
tr|U6KRL3|U6KRL3_EIMTE/1-248 210 P A A G A V -----G Q D R V F A H R F W I -----S S S S T P D C S C H -----238

```

## A.4 Molecular Dimensions for crystallisation trails

**AmSO<sub>4</sub> Suite Refill-Hit Solutions (4 x 12.5 ml tubes)**

Number	Salt	Buffer	Precipitant	Cat. no.
1			2.2 M Ammonium sulfate	134401
2	0.2 M Ammonium acetate		2.2 M Ammonium sulfate	134402
3	0.2 M Ammonium chloride		2.2 M Ammonium sulfate	134403
4	0.2 M Ammonium phosphate		2.2 M Ammonium sulfate	134404
5	0.2 M Ammonium fluoride		2.2 M Ammonium sulfate	134405
6	0.2 M Ammonium formate		2.2 M Ammonium sulfate	134406
7	0.2 M di-Ammonium citrate		2.2 M Ammonium sulfate	134407
8	0.2 M di-Ammonium phosphate		2.2 M Ammonium sulfate	134408
9	0.2 M Ammonium iodide		2.2 M Ammonium sulfate	134409
10	0.2 M Ammonium nitrate		2.2 M Ammonium sulfate	134410
11	0.2 M di-Ammonium tartrate		2.2 M Ammonium sulfate	134411
12	0.2 M Cadmium chloride		2.2 M Ammonium sulfate	134412
13	0.2 M Cadmium sulfate		2.2 M Ammonium sulfate	134413
14	0.2 M Cesium chloride		2.2 M Ammonium sulfate	134414
15	0.2 M Cesium sulfate		2.2 M Ammonium sulfate	134415
16	0.2 M Ammonium bromide		2.2 M Ammonium sulfate	134416
17	0.2 M Lithium acetate		2.2 M Ammonium sulfate	134417
18	0.2 M Lithium chloride		2.2 M Ammonium sulfate	134418
19	0.2 M tri-Lithium citrate		2.2 M Ammonium sulfate	134419
20	0.2 M Lithium nitrate		2.2 M Ammonium sulfate	134420
21	0.2 M Lithium sulfate		2.2 M Ammonium sulfate	134421
22	0.2 M Potassium acetate		2.2 M Ammonium sulfate	134422
23	0.2 M Potassium bromide		2.2 M Ammonium sulfate	134423
24	0.2 M Potassium chloride		2.2 M Ammonium sulfate	134424
25	0.2 M tri-Potassium citrate		2.2 M Ammonium sulfate	134425
26	0.2 M Potassium phosphate		2.2 M Ammonium sulfate	134426
27	0.2 M Potassium fluoride		2.2 M Ammonium sulfate	134427
28	0.2 M Potassium formate		2.2 M Ammonium sulfate	134428
29	0.2 M di-Potassium phosphate		2.2 M Ammonium sulfate	134429
30	0.2 M Potassium iodide		2.2 M Ammonium sulfate	134430
31	0.2 M Potassium nitrate		2.2 M Ammonium sulfate	134431
32	0.2 M K <sub>2</sub> /Na tartrate		2.2 M Ammonium sulfate	134432
33	0.2 M Potassium sulfate		2.2 M Ammonium sulfate	134433
34	0.2 M Potassium thiocyanate		2.2 M Ammonium sulfate	134434
35	0.2 M Sodium acetate		2.2 M Ammonium sulfate	134435
36	0.2 M Sodium bromide		2.2 M Ammonium sulfate	134436
37	0.2 M Sodium chloride		2.2 M Ammonium sulfate	134437
38	0.2 M tri-Sodium citrate		2.2 M Ammonium sulfate	134438
39	0.2 M Sodium phosphate		2.2 M Ammonium sulfate	134439
40	0.2 M Sodium fluoride		2.2 M Ammonium sulfate	134440
41	0.2 M Sodium formate		2.2 M Ammonium sulfate	134441
42	0.2 M di-Sodium phosphate		2.2 M Ammonium sulfate	134442
43	0.2 M Sodium iodide		2.2 M Ammonium sulfate	134443
44	0.2 M Sodium malonate		2.2 M Ammonium sulfate	134444
45	0.2 M Sodium nitrate		2.2 M Ammonium sulfate	134445
46	0.2 M Sodium sulfate		2.2 M Ammonium sulfate	134446
47	0.2 M di-Sodium tartrate		2.2 M Ammonium sulfate	134447
48	0.2 M Sodium thiocyanate		2.2 M Ammonium sulfate	134448



Number	Salt	Buffer	Precipitant 1	Precipitant 2	Cat. no.
49		0.1 M Citric acid pH 4.0	0.8 M Ammonium sulfate		134449
50		0.1 M Citric acid pH 5.0	0.8 M Ammonium sulfate		134450
51		0.1 M MES pH 6.0	0.8 M Ammonium sulfate		134451
52		0.1 M HEPES pH 7.0	0.8 M Ammonium sulfate		134452
53		0.1 M Tris pH 8.0	0.8 M Ammonium sulfate		134453
54		0.1 M BICINE pH 9.0	0.8 M Ammonium sulfate		134454
55		0.1 M Citric acid pH 4.0	1.6 M Ammonium sulfate		134455
56		0.1 M Citric acid pH 5.0	1.6 M Ammonium sulfate		134456
57		0.1 M MES pH 6.0	1.6 M Ammonium sulfate		134457
58		0.1 M HEPES pH 7.0	1.6 M Ammonium sulfate		134458
59		0.1 M Tris pH 8.0	1.6 M Ammonium sulfate		134459
60		0.1 M BICINE pH 9.0	1.6 M Ammonium sulfate		134460
61		0.1 M Citric acid pH 4.0	2.4 M Ammonium sulfate		134461
62		0.1 M Citric acid pH 5.0	2.4 M Ammonium sulfate		134462
63		0.1 M MES pH 6.0	2.4 M Ammonium sulfate		134463
64		0.1 M HEPES pH 7.0	2.4 M Ammonium sulfate		134464
65		0.1 M Tris pH 8.0	2.4 M Ammonium sulfate		134465
66		0.1 M BICINE pH 9.0	2.4 M Ammonium sulfate		134466
67		0.1 M Citric acid pH 4.0	3.2 M Ammonium sulfate		134467
68		0.1 M Citric acid pH 5.0	3.2 M Ammonium sulfate		134468
69		0.1 M MES pH 6.0	3.2 M Ammonium sulfate		134469
70		0.1 M HEPES pH 7.0	3.2 M Ammonium sulfate		134470
71		0.1 M Tris pH 8.0	3.2 M Ammonium sulfate		134471
72		0.1 M BICINE pH 9.0	3.2 M Ammonium sulfate		134472
73	0.1 M tri-Sodium citrate		0.5 M Ammonium sulfate	1.0 M Lithium Sulfate	134473
74			1.0 M Ammonium sulfate		134474
75		0.1 M Sodium acetate pH 4.6	1.0 M Ammonium sulfate		134475
76		0.1 M HEPES sodium salt pH 7.5	1.0 M Ammonium sulfate	2 %(w/v) PEG 400	134476
77		0.1 M Tris.HCl pH 8.5	1.0 M Ammonium sulfate		134477
78	0.05 M tri-Sodium citrate		1.2 M Ammonium sulfate	3 %(w/v) Isopropanol	134478
79		0.1 M Tris.HCl pH 8.5	1.5 M Ammonium sulfate	15 %(w/v) Glycerol	134479
80	0.5 M Lithium chloride		1.6 M Ammonium sulfate		134480
81	1.0 M Lithium sulfate		1.6 M Ammonium sulfate		134481
82	0.2 M Sodium chloride	0.1 M HEPES sodium salt pH 7.5	1.6 M Ammonium sulfate		134482
83		0.1 M HEPES sodium salt pH 7.5	1.6 M Ammonium sulfate	2 %(w/v) PEG 1000	134483
84		0.1 M MES sodium salt pH 6.5	1.8 M Ammonium sulfate		134484
85	2.0 M Sodium chloride		2.0 M Ammonium sulfate		134485
86		0.1 M Sodium acetate pH 4.6	2.0 M Ammonium sulfate		134486
87		0.1 M MES sodium salt pH 6.5	2.0 M Ammonium sulfate	5 %(w/v) PEG 400	134487
88		0.1 M Tris.HCl pH 8.5	2.0 M Ammonium sulfate		134488
89			2.2 M Ammonium sulfate		134489
90			2.2 M Ammonium sulfate	20 %(w/v) Glycerol	134490
91	0.1 M tri-Sodium citrate		2.4 M Ammonium sulfate		134491
92			3.0 M Ammonium sulfate	1 %(w/v) MPD	134492
93			3.0 M Ammonium sulfate	10 %(w/v) Glycerol	134493
94		0.1 M HEPES sodium salt pH 7.5	3.5 M Ammonium sulfate		134494
95		0.1 M MES sodium salt pH 6.5	3.5 M Ammonium sulfate	1 %(w/v) MPD	134495
96			3.5 M Ammonium sulfate		134496

**The PACT Suite Composition Table**

Number	Salt	Buffer	Precipitant	Cat. no. (Refill-Hit Solution, 4 x 12.5 ml tubes)
1		0.1 M SPG buffer pH 4	25% (w/v) PEG 1500	135701
2		0.1 M SPG buffer pH 5	25% (w/v) PEG 1500	135702
3		0.1 M SPG buffer pH 6	25% (w/v) PEG 1500	135703
4		0.1 M SPG buffer pH 7	25% (w/v) PEG 1500	135704
5		0.1 M SPG buffer pH 8	25% (w/v) PEG 1500	135705
6		0.1 M SPG buffer pH 9	25% (w/v) PEG 1500	135706
7	0.2 M Sodium chloride	0.1 M Sodium acetate pH 5	20% (w/v) PEG 6000	135707
8	0.2 M Ammonium chloride	0.1 M Sodium acetate pH 5	20% (w/v) PEG 6000	135708
9	0.2 M Lithium chloride	0.1 M Sodium acetate pH 5	20% (w/v) PEG 6000	135709
10	0.2 M Magnesium chloride	0.1 M Sodium acetate pH 5	20% (w/v) PEG 6000	135710
11	0.2 M Calcium chloride	0.1 M Sodium acetate pH 5	20% (w/v) PEG 6000	135711
12	0.01 M Zinc chloride	0.1 M Sodium acetate pH 5	20% (w/v) PEG 6000	135712
13		0.1 M MIB buffer pH 4	25% (w/v) PEG 1500	135713
14		0.1 M MIB buffer pH 5	25% (w/v) PEG 1500	135714
15		0.1 M MIB buffer pH 6	25% (w/v) PEG 1500	135715
16		0.1 M MIB buffer pH 7	25% (w/v) PEG 1500	135716
17		0.1 M MIB buffer pH 8	25% (w/v) PEG 1500	135717
18		0.1 M MIB buffer pH 9	25% (w/v) PEG 1500	135718
19	0.2 M Sodium chloride	0.1 M MES pH 6	20% (w/v) PEG 6000	135719
20	0.2 M Ammonium chloride	0.1 M MES pH 6	20% (w/v) PEG 6000	135720
21	0.2 M Lithium chloride	0.1 M MES pH 6	20% (w/v) PEG 6000	135721
22	0.2 M Magnesium chloride	0.1 M MES pH 6	20% (w/v) PEG 6000	135722
23	0.2 M Calcium chloride	0.1 M MES pH 6	20% (w/v) PEG 6000	135723
24	0.01 M Zinc chloride	0.1 M MES pH 6	20% (w/v) PEG 6000	135724
25		0.1 M PCB buffer pH 4	25% (w/v) PEG 1500	135725
26		0.1 M PCB buffer pH 5	25% (w/v) PEG 1500	135726
27		0.1 M PCB buffer pH 6	25% (w/v) PEG 1500	135727
28		0.1 M PCB buffer pH 7	25% (w/v) PEG 1500	135728
29		0.1 M PCB buffer pH 8	25% (w/v) PEG 1500	135729
30		0.1 M PCB buffer pH 9	25% (w/v) PEG 1500	135730
31	0.2 M Sodium chloride	0.1 M HEPES pH 7	20% (w/v) PEG 6000	135731
32	0.2 M Ammonium chloride	0.1 M HEPES pH 7	20% (w/v) PEG 6000	135732
33	0.2 M Lithium chloride	0.1 M HEPES pH 7	20% (w/v) PEG 6000	135733
34	0.2 M Magnesium chloride	0.1 M HEPES pH 7	20% (w/v) PEG 6000	135734
35	0.2 M Calcium chloride	0.1 M HEPES pH 7	20% (w/v) PEG 6000	135735
36	0.01 M Zinc chloride	0.1 M HEPES pH 7	20% (w/v) PEG 6000	135736
37		0.1 M MMT buffer pH 4	25% (w/v) PEG 1500	135737
38		0.1 M MMT buffer pH 5	25% (w/v) PEG 1500	135738
39		0.1 M MMT buffer pH 6	25% (w/v) PEG 1500	135739
40		0.1 M MMT buffer pH 7	25% (w/v) PEG 1500	135740
41		0.1 M MMT buffer pH 8	25% (w/v) PEG 1500	135741
42		0.1 M MMT buffer pH 9	25% (w/v) PEG 1500	135742
43	0.2 M Sodium chloride	0.1 M Tris pH 8	20% (w/v) PEG 6000	135743
44	0.2 M Ammonium chloride	0.1 M Tris pH 8	20% (w/v) PEG 6000	135744
45	0.2 M Lithium chloride	0.1 M Tris pH 8	20% (w/v) PEG 6000	135745
46	0.2 M Magnesium chloride	0.1 M Tris pH 8	20% (w/v) PEG 6000	135746
47	0.2 M Calcium chloride	0.1 M Tris pH 8	20% (w/v) PEG 6000	135747
48		0.1 M Tris pH 8	20% (w/v) PEG 6000	135748

**The PACT Suite Composition Table**

Number	Salt	Buffer	Precipitant	Cat. no. (Refill-Hit Solution, 4 x 12.5 ml tubes)
49	0.2 M Sodium fluoride		20% (w/v) PEG 3350	135749
50	0.2 M Sodium bromide		20% (w/v) PEG 3350	135750
51	0.2 M Sodium iodide		20% (w/v) PEG 3350	135751
52	0.2 M Potassium thiocyanate		20% (w/v) PEG 3350	135752
53	0.2 M Sodium nitrate		20% (w/v) PEG 3350	135753
54	0.2 M Sodium formate		20% (w/v) PEG 3350	135754
55	0.2 M Sodium acetate		20% (w/v) PEG 3350	135755
56	0.2 M Sodium sulfate		20% (w/v) PEG 3350	135756
57	0.2 M Potassium/sodium tartrate		20% (w/v) PEG 3350	135757
58	0.2 M Sodium/potassium phosphate		20% (w/v) PEG 3350	135758
59	0.2 M Sodium citrate		20% (w/v) PEG 3350	135759
60	0.2 M Sodium malonate		20% (w/v) PEG 3350	135760
61	0.2 M Sodium fluoride	0.1 M Bis Tris propane pH 6.5	20% (w/v) PEG 3350	135761
62	0.2 M Sodium bromide	0.1 M Bis Tris propane pH 6.5	20% (w/v) PEG 3350	135762
63	0.2 M Sodium iodide	0.1 M Bis Tris propane pH 6.5	20% (w/v) PEG 3350	135763
64	0.2 M Potassium thiocyanate	0.1 M Bis Tris propane pH 6.5	20% (w/v) PEG 3350	135764
65	0.2 M Sodium nitrate	0.1 M Bis Tris propane pH 6.5	20% (w/v) PEG 3350	135765
66	0.2 M Sodium formate	0.1 M Bis Tris propane pH 6.5	20% (w/v) PEG 3350	135766
67	0.2 M Sodium acetate	0.1 M Bis Tris propane pH 6.5	20% (w/v) PEG 3350	135767
68	0.2 M Sodium sulfate	0.1 M Bis Tris propane pH 6.5	20% (w/v) PEG 3350	135768
69	0.2 M Potassium/sodium tartrate	0.1 M Bis Tris propane pH 6.5	20% (w/v) PEG 3350	135769
70	0.2 M Sodium/potassium phosphate	0.1 M Bis Tris propane pH 6.5	20% (w/v) PEG 3350	135770
71	0.2 M Sodium citrate	0.1 M Bis Tris propane pH 6.5	20% (w/v) PEG 3350	135771
72	0.2 M Sodium malonate	0.1 M Bis Tris propane pH 6.5	20% (w/v) PEG 3350	135772
73	0.2 M Sodium fluoride	0.1 M Bis Tris propane pH 7.5	20% (w/v) PEG 3350	135773
74	0.2 M Sodium bromide	0.1 M Bis Tris propane pH 7.5	20% (w/v) PEG 3350	135774
75	0.2 M Sodium iodide	0.1 M Bis Tris propane pH 7.5	20% (w/v) PEG 3350	135775
76	0.2 M Potassium thiocyanate	0.1 M Bis Tris propane pH 7.5	20% (w/v) PEG 3350	135776
77	0.2 M Sodium nitrate	0.1 M Bis Tris propane pH 7.5	20% (w/v) PEG 3350	135777
78	0.2 M Sodium formate	0.1 M Bis Tris propane pH 7.5	20% (w/v) PEG 3350	135778
79	0.2 M Sodium acetate	0.1 M Bis Tris propane pH 7.5	20% (w/v) PEG 3350	135779
80	0.2 M Sodium sulfate	0.1 M Bis Tris propane pH 7.5	20% (w/v) PEG 3350	135780
81	0.2 M Potassium/sodium tartrate	0.1 M Bis Tris propane pH 7.5	20% (w/v) PEG 3350	135781
82	0.2 M Sodium/potassium phosphate	0.1 M Bis Tris propane pH 7.5	20% (w/v) PEG 3350	135782
83	0.2 M Sodium citrate	0.1 M Bis Tris propane pH 7.5	20% (w/v) PEG 3350	135783
84	0.2 M Sodium malonate	0.1 M Bis Tris propane pH 7.5	20% (w/v) PEG 3350	135784
85	0.2 M Sodium fluoride	0.1 M Bis Tris propane pH 8.5	20% (w/v) PEG 3350	135785
86	0.2 M Sodium bromide	0.1 M Bis Tris propane pH 8.5	20% (w/v) PEG 3350	135786
87	0.2 M Sodium iodide	0.1 M Bis Tris propane pH 8.5	20% (w/v) PEG 3350	135787
88	0.2 M Potassium thiocyanate	0.1 M Bis Tris propane pH 8.5	20% (w/v) PEG 3350	135788
89	0.2 M Sodium nitrate	0.1 M Bis Tris propane pH 8.5	20% (w/v) PEG 3350	135789
90	0.2 M Sodium formate	0.1 M Bis Tris propane pH 8.5	20% (w/v) PEG 3350	135790
91	0.2 M Sodium acetate	0.1 M Bis Tris propane pH 8.5	20% (w/v) PEG 3350	135791
92	0.2 M Sodium sulfate	0.1 M Bis Tris propane pH 8.5	20% (w/v) PEG 3350	135792
93	0.2 M Potassium/sodium tartrate	0.1 M Bis Tris propane pH 8.5	20% (w/v) PEG 3350	135793
94	0.2 M Sodium/potassium phosphate	0.1 M Bis Tris propane pH 8.5	20% (w/v) PEG 3350	135794
95	0.2 M Sodium citrate	0.1 M Bis Tris propane pH 8.5	20% (w/v) PEG 3350	135795
96	0.2 M Sodium malonate	0.1 M Bis Tris propane pH 8.5	20% (w/v) PEG 3350	135796

## JCSG+ Suite Refill-Hit Solutions (4 x 12.5 ml tubes)

Number	Salt(s)	Buffer	Precipitant	Cat. no.
1	0.2 M Lithium sulfate	0.1 M Sodium acetate pH 4.5	50% w/v PEG 400	135901
2		0.1 M tri-Sodium citrate pH 5.5	20% w/v PEG 3000	135902
3	0.2 M di-Ammonium citrate pH 5.0		20% w/v PEG 3350	135903
4	0.02 M Calcium chloride	0.1 M Sodium acetate pH 4.6	30% v/v MPD	135904
5	0.2 M Magnesium formate pH 5.9		20% w/v PEG 3350	135905
6	0.2 M Lithium sulfate	0.1 M Phosphate-citrate pH 4.2	20% w/v PEG 1000	135906
7		0.1 M CHES pH 9.5	20% w/v PEG 8000	135907
8	0.2 M Ammonium formate pH 6.6		20% w/v PEG 3350	135908
9	0.2 M Ammonium chloride pH 6.3		20% w/v PEG 3350	135909
10	0.2 M Potassium formate pH 7.3		20% w/v PEG 3350	135910
11	0.2 M Ammonium phosphate	0.1 M Tris pH 8.5	50% v/v MPD	135911
12	0.2 M Potassium nitrate pH 6.9		20% w/v PEG 3350	135912
13	0.8 M Ammonium sulfate	0.1 M Citric acid pH 4		135913
14	0.2 M Sodium thiocyanate pH 6.9		20% w/v PEG 3350	135914
15		0.1 M BICINE pH 9	20% w/v PEG 6000	135915
16		0.1 M HEPES pH 7.5	10% w/v PEG 8000, 8% v/v Ethylene glycol	135916
17		0.1 M Sodium cacodylate pH 6.5	40% v/v MPD, 5% w/v PEG 8000	135917
18		0.1 M Phosphate-citrate pH 4.2	40% v/v Ethanol, 5% w/v PEG 1000	135918
19		0.1 M Sodium acetate pH 4.6	8% w/v PEG 4000	135919
20	0.2 M Magnesium chloride	0.1 M Tris pH 7	10% w/v PEG 8000	135920
21		0.1 M Citric acid pH 5	20% w/v PEG 6000	135921
22	0.2 M Magnesium chloride	0.1 M Sodium cacodylate pH 6.5	50% v/v PEG 200	135922
23	1.6 M tri-Sodium citrate			135923
24	0.2 M tri-Potassium citrate pH 8.3		20% w/v PEG 3350	135924
25	0.2 M Sodium chloride	0.1 M Phosphate-citrate pH 4.2	20% w/v PEG 8000	135925
26	1 M Lithium chloride	0.1 M Citric acid pH 4	20% w/v PEG 6000	135926
27	0.2 M Ammonium nitrate pH 6.3		20% w/v PEG 3350	135927
28		0.1 M HEPES pH 7	10% w/v PEG 6000	135928
29		0.1 M HEPES pH 7.5	0.8 M Sodium phosphate, 0.8 M Potassium phosphate	135929
30		0.1 M Phosphate-citrate pH 4.2	40% v/v PEG 300	135930
31	0.2 M Zinc acetate	0.1 M Sodium acetate pH 4.5	10% w/v PEG 3000	135931
32		0.1 M Tris pH 8.5	20% v/v Ethanol	135932
33		0.1 M Na/K phosphate pH 6.2	25% v/v 1,2 propandiol, 10% v/v Glycerol	135933
34		0.1 M BICINE pH 9	10% w/v PEG 20000, 2 %w/v Dioxane	135934
35	2 M Ammonium sulfate	0.1 M Sodium acetate pH 4.6		135935
36			10% w/v PEG 1000, 10% w/v PEG 8000	135936
37			24 %w/v PEG 1500, 20% w/v glycerol	135937
38	0.2 M Magnesium chloride	0.1 M HEPES pH 7.5	30% v/v PEG 400	135938
39	0.2 M Sodium chloride	0.1 M Na/K phosphate pH 6.2	50% v/v PEG 200	135939
40	0.2 M Lithium sulfate	0.1 M Sodium acetate pH 4.5	30% w/v PEG 8000	135940
41		0.1 M HEPES pH 7.5	70% v/v MPD	135941
42	0.2 M Magnesium chloride	0.1 M Tris pH 8.5	20% w/v PEG 8000	135942
43	0.2 M Lithium sulfate	0.1 M Tris pH 8.5	40% v/v PEG 400	135943
44		0.1 M Tris pH 8	40% v/v MPD	135944
45	0.17 M Ammonium sulfate		25.5% w/v PEG 4000, 15% v/v Glycerol	135945
46	0.2 M Calcium acetate	0.1 M Sodium cacodylate pH 6.5	40% v/v PEG 300	135946
47	0.14 M Calcium chloride	0.07 M Sodium acetate pH 4.6	14% v/v Isopropanol, 30% v/v Glycerol	135947
48	0.04 M Potassium phosphate		16% w/v PEG 8000, 20% v/v Glycerol	135948

## JCSG+ Suite Refill-Hit Solutions (4 x 12.5 ml tubes)

Number	Salt(s)	Buffer	Precipitant	Cat. no.
49	1 M tri-Sodium citrate	0.1 M Sodium cacodylate pH 6.5		135949
50	0.2 M Sodium chloride	0.1 M Sodium cacodylate pH 6.5	2 M Ammonium sulfate	135950
51	0.2 M Sodium chloride	0.1 M HEPES pH 7.5	10% v/v Isopropanol	135951
52	0.2 M Lithium sulfate	0.1 M Tris pH 8.5	1.26 M Ammonium sulfate	135952
53		0.1 M CAPS pH 10.5	40% v/v MPD	135953
54	0.2 M Zinc acetate	0.1 M Imidazole pH 8	20% w/v PEG 3000	135954
55	0.2 M Zinc acetate	0.1 M Sodium cacodylate pH 6.5	10% v/v Isopropanol	135955
56	1 M di-Ammonium phosphate	0.1 M Sodium acetate pH 4.5		135956
57	1,6 M Magnesium sulfate	0.1 M MES pH 6.5		135957
58		0.1 M BICINE pH 9	10% w/v PEG 6000	135958
59	0.16 M Calcium acetate	0.08 M Sodium cacodylate pH 6.5	14.4 %w/v PEG 8000, 20% v/v Glycerol	135959
60		0.1 M Imidazole pH 8	10% w/v PEG 8000	135960
61	0.05 M Cesium chloride	0.1 M MES pH 6.5	30% w/v Jeffamine M-600	135961
62	3,15 M Ammonium sulfate	0.1 M Citric acid pH 5		135962
63		0.1 M Tris pH 8	20% v/v MPD	135963
64		0.1 M HEPES pH 6.5	20% w/v Jeffamine M-600	135964
65	0.2 M Magnesium chloride	0.1 M Tris pH 8.5	50% v/v Ethylene glycol	135965
66		0.1 M BICINE pH 9	10% v/v MPD	135966
67	0.8 M Succinic acid pH 7.0			135967
68	2,1 M DL-Malic acid pH 7.0			135968
69	2,4 M Sodium malonate pH 7.0			135969
70	1,1 M Sodium malonate pH 7.0	0.1 M HEPES pH 7	0.5 %v/v Jeffamine ED-2001 pH 7.0	135970
71	1 M Succinic acid pH 7.0	0.1 M HEPES pH 7	1% w/v PEG MME 2000	135971
72		0.1 M HEPES pH 7	30% v/v Jeffamine M-600 pH 7.0	135972
73		0.1 M HEPES pH 7	30% v/v Jeffamine ED-2001 pH 7.0	135973
74	0.02 M Magnesium chloride	0.1 M HEPES pH 7.5	22% w/v Polyacrylic acid 5100 sodium salt	135974
75	0.1 M Cobalt chloride	0.1 M Tris pH 8.5	20% w/v Polyvinylpyrrolidone K15	135975
76	0.2 M Trimethylamine N-oxide	0.1 M Tris pH 8.5	20% w/v PEG MME 2000	135976
77	0.005 M Cobalt chloride, 0.005 M Cadmium chloride, 0.005 M Magnesium chloride, 0.005 M Nickel chloride	0.1 M HEPES pH 7.5	12% w/v PEG 3350	135977
78	0.2 M Sodium malonate pH 7.0		20% w/v PEG 3350	135978
79	0.1 M Succinic acid pH 7.0		15% w/v PEG 3350	135979
80	0.15 M DL-Malic acid pH 7.0		20% w/v PEG 3350	135980
81	0.1 M Potassium thiocyanate		30% w/v PEG MME 2000	135981
82	0.15 M Potassium bromide		30% w/v PEG MME 2000	135982
83	2 M Ammonium sulfate	0.1 M bis-Tris pH 5.5		135983
84	3 M Sodium chloride	0.1 M bis-Tris pH 5.5		135984
85	0.3 M Magnesium formate	0.1 M bis-Tris pH 5.5		135985
86	1 M Ammonium sulfate	0.1 M bis-Tris pH 5.5	1% w/v PEG 3350	135986
87		0.1 M bis-Tris pH 5.5	25% w/v PEG 3350	135987
88	0.2 M Calcium chloride	0.1 M bis-Tris pH 5.5	45% v/v MPD	135988
89	0.2 M Ammonium acetate	0.1 M bis-Tris pH 5.5	45% v/v MPD	135989
90	0.1 M Ammonium acetate	0.1 M bis-Tris pH 5.5	17% w/v PEG 10,000	135990
91	0.2 M Ammonium sulfate	0.1 M bis-Tris pH 5.5	25% w/v PEG 3350	135991
92	0.2 M Sodium chloride	0.1 M bis-Tris pH 5.5	25% w/v PEG 3350	135992
93	0.2 M Lithium sulfate	0.1 M bis-Tris pH 5.5	25% w/v PEG 3350	135993
94	0.2 M Ammonium acetate	0.1 M bis-Tris pH 5.5	25% w/v PEG 3350	135994
95	0.2 M Magnesium chloride	0.1 M bis-Tris pH 5.5	25% w/v PEG 3350	135995
96	0.2 M Ammonium acetate	0.1 M HEPES pH 7.5	45% v/v MPD	135996

## The pHClear Suite Composition Table

Number	Buffer	Salt	pH	Cat. no. (Refill-Hit Solution, 4 x 12.5 ml tubes)
1	0.1 M Citric acid	1.0 M Sodium chloride	4.0	134801
2	0.1 M Citric acid	1.0 M Sodium chloride	5.0	134802
3	0.1 M MES	1.0 M Sodium chloride	6.0	134803
4	0.1 M HEPES	1.0 M Sodium chloride	7.0	134804
5	0.1 M Tris	1.0 M Sodium chloride	8.0	134805
6	0.1 M Bicine	1.0 M Sodium chloride	9.0	134806
7	0.1 M Citric acid	2.0 M Sodium chloride	4.0	134807
8	0.1 M Citric acid	2.0 M Sodium chloride	5.0	134808
9	0.1 M MES	2.0 M Sodium chloride	6.0	134809
10	0.1 M HEPES	2.0 M Sodium chloride	7.0	134810
11	0.1 M Tris	2.0 M Sodium chloride	8.0	134811
12	0.1 M Bicine	2.0 M Sodium chloride	9.0	134812
13	0.1 M Citric acid	3.0 M Sodium chloride	4.0	134813
14	0.1 M Citric acid	3.0 M Sodium chloride	5.0	134814
15	0.1 M MES	3.0 M Sodium chloride	6.0	134815
16	0.1 M HEPES	3.0 M Sodium chloride	7.0	134816
17	0.1 M Tris	3.0 M Sodium chloride	8.0	134817
18	0.1 M Bicine	3.0 M Sodium chloride	9.0	134818
19	0.1 M Citric acid	4.0 M Sodium chloride	4.0	134819
20	0.1 M Citric acid	4.0 M Sodium chloride	5.0	134820
21	0.1 M MES	4.0 M Sodium chloride	6.0	134821
22	0.1 M HEPES	4.0 M Sodium chloride	7.0	134822
23	0.1 M Tris	4.0 M Sodium chloride	8.0	134823
24	0.1 M Bicine	4.0 M Sodium chloride	9.0	134824
25	0.1 M Citric acid	5% (w/v) PEG 6000	4.0	134825
26	0.1 M Citric acid	5% (w/v) PEG 6000	5.0	134826
27	0.1 M MES	5% (w/v) PEG 6000	6.0	134827
28	0.1 M HEPES	5% (w/v) PEG 6000	7.0	134828
29	0.1 M Tris	5% (w/v) PEG 6000	8.0	134829
30	0.1 M Bicine	5% (w/v) PEG 6000	9.0	134830
31	0.1 M Citric acid	10% (w/v) PEG 6000	4.0	134831
32	0.1 M Citric acid	10% (w/v) PEG 6000	5.0	134832
33	0.1 M MES	10% (w/v) PEG 6000	6.0	134833
34	0.1 M HEPES	10% (w/v) PEG 6000	7.0	134834
35	0.1 M Tris	10% (w/v) PEG 6000	8.0	134835
36	0.1 M Bicine	10% (w/v) PEG 6000	9.0	134836
37	0.1 M Citric acid	20% (w/v) PEG 6000	4.0	134837
38	0.1 M Citric acid	20% (w/v) PEG 6000	5.0	134838
39	0.1 M MES	20% (w/v) PEG 6000	6.0	134839
40	0.1 M HEPES	20% (w/v) PEG 6000	7.0	134840
41	0.1 M Tris	20% (w/v) PEG 6000	8.0	134841
42	0.1 M Bicine	20% (w/v) PEG 6000	9.0	134842
43	0.1 M Citric acid	30% (w/v) PEG 6000	4.0	134843
44	0.1 M Citric acid	30% (w/v) PEG 6000	5.0	134844
45	0.1 M MES	30% (w/v) PEG 6000	6.0	134845
46	0.1 M HEPES	30% (w/v) PEG 6000	7.0	134846
47	0.1 M Tris	30% (w/v) PEG 6000	8.0	134847
48	0.1 M Bicine	30% (w/v) PEG 6000	9.0	134848

## The pHClear Suite Composition Table

Number	Buffer	Salt	pH	Cat. no. (Refill-Hit Solution, 4 x 12.5 ml tubes)
49	0.1 M Citric acid	0.8 M Ammonium sulfate	4.0	134849
50	0.1 M Citric acid	0.8 M Ammonium sulfate	5.0	134850
51	0.1 M MES	0.8 M Ammonium sulfate	6.0	134851
52	0.1 M HEPES	0.8 M Ammonium sulfate	7.0	134852
53	0.1 M Tris	0.8 M Ammonium sulfate	8.0	134853
54	0.1 M Bicine	0.8 M Ammonium sulfate	9.0	134854
55	0.1 M Citric acid	1.6 M Ammonium sulfate	4.0	134855
56	0.1 M Citric acid	1.6 M Ammonium sulfate	5.0	134856
57	0.1 M MES	1.6 M Ammonium sulfate	6.0	134857
58	0.1 M HEPES	1.6 M Ammonium sulfate	7.0	134858
59	0.1 M Tris	1.6 M Ammonium sulfate	8.0	134859
60	0.1 M Bicine	1.6 M Ammonium sulfate	9.0	134860
61	0.1 M Citric acid	2.4 M Ammonium sulfate	4.0	134861
62	0.1 M Citric acid	2.4 M Ammonium sulfate	5.0	134862
63	0.1 M MES	2.4 M Ammonium sulfate	6.0	134863
64	0.1 M HEPES	2.4 M Ammonium sulfate	7.0	134864
65	0.1 M Tris	2.4 M Ammonium sulfate	8.0	134865
66	0.1 M Bicine	2.4 M Ammonium sulfate	9.0	134866
67	0.1 M Citric acid	3.2 M Ammonium sulfate	4.0	134867
68	0.1 M Citric acid	3.2 M Ammonium sulfate	5.0	134868
69	0.1 M MES	3.2 M Ammonium sulfate	6.0	134869
70	0.1 M HEPES	3.2 M Ammonium sulfate	7.0	134870
71	0.1 M Tris	3.2 M Ammonium sulfate	8.0	134871
72	0.1 M Bicine	3.2 M Ammonium sulfate	9.0	134872
73	0.1 M Citric acid	10% (v/v) MPD	4.0	134873
74	0.1 M Sodium acetate	10% (v/v) MPD	5.0	134874
75	0.1 M MES	10% (v/v) MPD	6.0	134875
76	0.1 M HEPES	10% (v/v) MPD	7.0	134876
77	0.1 M Tris	10% (v/v) MPD	8.0	134877
78	0.1 M Bicine	10% (v/v) MPD	9.0	134878
79	0.1 M Citric acid	20% (v/v) MPD	4.0	134879
80	0.1 M Sodium acetate	20% (v/v) MPD	5.0	134880
81	0.1 M MES	20% (v/v) MPD	6.0	134881
82	0.1 M HEPES	20% (v/v) MPD	7.0	134882
83	0.1 M Tris	20% (v/v) MPD	8.0	134883
84	0.1 M Bicine	20% (v/v) MPD	9.0	134884
85	0.1 M Citric acid	40% (v/v) MPD	4.0	134885
86	0.1 M Sodium acetate	40% (v/v) MPD	5.0	134886
87	0.1 M MES	40% (v/v) MPD	6.0	134887
88	0.1 M HEPES	40% (v/v) MPD	7.0	134888
89	0.1 M Tris	40% (v/v) MPD	8.0	134889
90	0.1 M Bicine	40% (v/v) MPD	9.0	134890
91	0.1 M Citric acid	65% (v/v) MPD	4.0	134891
92	0.1 M Sodium acetate	65% (v/v) MPD	5.0	134892
93	0.1 M MES	65% (v/v) MPD	6.0	134893
94	0.1 M HEPES	65% (v/v) MPD	7.0	134894
95	0.1 M Tris	65% (v/v) MPD	8.0	134895
96	0.1 M Bicine	65% (v/v) MPD	9.0	134896