

The structure-function relationship of Multimerin-1, a platelet protein

S A Rama

MSc by Res 2023

The structure-function relationship of Multimerin-1, a platelet protein

SILVANA APRYL RAMA

A thesis submitted in fulfilment of the
requirements of Manchester Metropolitan
University for the degree of Master of Science
(By Research)

Department of Natural Sciences
Manchester Metropolitan University

2023

Contents

List of Figures	6
List of Tables	8
List of Abbreviations.....	9
Abstract	13
1 Introduction	14
1.1 Multimerin-1 (MMRN1) structure and functions.....	14
1.1.1 MMRN1's physiological role in platelets	14
1.1.2 MMRN1 deficiency in platelets causes Quebec platelet disorder (QPD)	16
1.1.3 MMRN1 as an acute myeloid leukaemia (AML) biomarker.....	16
1.1.4 MMRN1 domains	18
1.2 MMRN1 interactions with pathogenic proteins	24
1.3 Study outline	26
1.3.1 Study aims and objectives.....	27
1.3.2 Study hypothesis	28
2 Material and methods	29
2.1 Materials.....	29
2.1.1 Preparation of growth media.....	29
2.1.2 Preparation of Buffers.....	29
2.2.1 Primer design	31
2.2.2 Growth of bacterial cultures with antibiotics	34
2.2.3 Plasmid miniprep	34
2.2.4 Polymerase Chain Reaction (PCR).....	35
2.2.5 DNA agarose gel electrophoresis.....	36
2.2.6 Restriction enzyme digests	37
2.2.7 Restriction digest Clean-up	37
2.2.8 Size Selected Clean-up of DNA digests	38
2.2.9 DNA Ligation	39
2.2.10 Additional subcloning methods	40
2.2.11 Transformation of vectors into DH5 α cloning cell line and BL21(DE3) chemically competent <i>E. coli</i>	41
2.2.12 Screening for successful expression clones	42
2.2.13 Sequencing.....	43

2.2.14 Small-scale protein expression trials	43
2.2.15 Large-scale protein expression and purification.....	45
2.2.16 Soluble and insoluble sample preparation for SDS-PAGE.....	45
2.2.17 SDS-Polyacrylamide gel electrophoresis (SDS-PAGE)	45
2.2.18 Coomassie Staining	46
2.2.19 Western blot	46
2.2.20 Pull-down assays using Ni ²⁺ -affinity beads.....	47
2.2.21 Immobilised metal affinity chromatography (IMAC) protein purification.....	47
3 Results	49
3.1 Molecular cloning	49
3.1.1 Plasmid miniprep	49
3.1.2 Restriction enzyme digest to prepare vectors for subcloning	51
3.1.3 Screening of optimal PCR annealing temperatures for domain insert primer pairs	53
3.1.4 Subcloning of genes coding for MMRN1 regions into expression vectors	57
3.1.5 Transformation into DH5 α cloning cell line	60
3.1.6 Screening of genes coding for MMRN1 regions following the transformation of ligated products	61
3.1.7 Sequencing construct 1C (gC1q pET28a)	63
3.1.8 TOPO [®] TA cloning [®]	66
3.2 Recombinant expression of MMRN1 constructs in <i>E. coli</i> culture.....	66
3.2.1 Small-scale expression trials	66
3.2.2 Scaling up the expression of construct 1A (EMI) and 1C (gC1q).....	70
3.2.3 Western blot of construct 1A (EMI) and 1C (gC1q)	71
3.2.4 Mini His-bead pull-down purification	72
3.2.5 Affinity chromatography using the Akta for the purification of construct 1C (gC1q)	74
4 Discussion	76
4.1 Subcloning of the MMRN1 domains	77
4.1.1 Non-specific amplification in PCR agarose gels	77
4.1.2 Challenges encountered in sequencing outcomes	78
4.1.3 Cloning the EGF-like domain.....	80
4.2 Expression conditions and expression systems	81
4.2.1 Optimising IPTG concentration as an expression condition	81
4.2.2 Post-translational modifications of the MMRN1 glycoprotein.....	81
4.3 Protein sample analysis	86
4.3.1 SDS-PAGE optimisation	86
4.3.2 Optimising protein processing methods.....	86

4.3.3 Contamination in SDS-PAGE following protein purification of the gC1q domain construct	88
4.4 Future directions	89
4.4.1 Purification methods for gC1q domain homogeneity	90
4.4.2 Current methods of detecting Protein-protein interactions (PPI)	91
5 Conclusion	93
6 Appendix	94
6.1 Appendix 1: Unedited ligation DNA agarose gel	94
6.2 Appendix 2: Unedited purification SDS-PAGE gels of construct 1C (gC1q)	94
6.3 Appendix 3: Chromatogram of construct 1C (gC1q) following Äkta purification	95
7 References	96
Acknowledgements	110

List of Figures

Figure 1. MMRN1 acts as a binding site for FV to regulate thrombin production	15
Figure 2. Granular content and anatomical structure of a platelet (adapted from Selvadurai and Hamilton (2018))	15
Figure 3. Overview of EMILIN family members' protein domains (adapted from Colombatti et al. (2012))	18
Figure 4. A summary of known and proposed functions of the MMRN1 domains (adapted from Posner (2022))	19
Figure 5. A multiple sequence alignment of both EMI (A) and gC1q (B) domains in the EMILIN family using the Jalview programme (Waterhouse et al., 2009)	21
Figure 6. An AI predicted structure of MMRN1 and protein domains using ColabFold-AlphaFold2 (Mirdita et al., 2022)	23
Figure 7. A workflow summary of the proposed experimental workflow of MMRN1 cloning, expression, purification, interaction (co-immunoprecipitation) and structural studies (X-ray crystallography)	27
Figure 8. Plasmid maps of pET vectors with an overview of the MMRN1 domains	33
Figure 9. Overview of plasmid Miniprep preparation	35
Figure 10. Overview of Gel and PCR clean-up	38
Figure 11. Right side size selection overview	39
Figure 12. Overview of transformation	42
Figure 13. Overview of the PCR colony screen protocol	43
Figure 14. Cell growth curve showing phases of cell growth (adapted from Peleg and Corradini (2011))	44
Figure 15. Overview of Western Blotting	47
Figure 16. DNA gels (1%) of plasmid minipreps for MMRN1 and pET vectors	50
Figure 17. DNA gels (1%) of digestion and purification of pET vectors	52
Figure 18. Screening of annealing temperatures and the effect of the DMSO additive	56
Figure 19. Gel electrophoresis of domain insert subcloning	58
Figure 20. DNA gel (1%) of minipreps of 3-4 DH5 α colonies per plasmid	60
Figure 21. DNA gels (1%) of double-digest and PCR colony screening of genes coding for the domain constructs	62
Figure 22. Amino acid sequence of the gC1q pET28a construct	64
Figure 23. ColabFold-AlphaFold2 3D image of the impact on the last amino acid change for construct 1C on protein folding versus the expected protein sequence of the gC1q (Mirdita et al., 2022)	65
Figure 24. Expression of 8 clones within BL21(DE3) with varied induction times and temperatures	69
Figure 25. Scaled up expression of constructs 1A and 1C in 100 mL culture	71
Figure 26. Fluorescence western blot of constructs 1C (a, c) and 1A (b, c) incubated with anti-His antibodies	72
Figure 27. A His-bead pull-down purification of constructs 1A and 1C incubated with HisPur™ Ni-NTA Resin (ThermoFisher)	73

Figure 28. The purification of Construct 1C using affinity chromatography eluted with a Peristaltic Pump across a gradient of 100 mM to 500 mM Imidazole74

Figure 29. Diagram of MMRN1 protein sequence, and domains with confirmed and inferred primary glycan locations (Uniprot, 2023)82

List of Tables

Table 1. Known mechanisms of <i>S. aureus</i> to target platelet proteins.....	25
Table 2. Growth media composition.	29
Table 3. General buffer composition.	30
Table 4. Outline of forward and reverse primer sequences and features for each domain construct referred to in text with a construct ID.....	32
Table 5. Antibiotic stock and working concentrations.	34
Table 6. Composition of PCR reactions (20 μ L).....	36
Table 7. Cycling instructions using a ProFlex™ PCR system thermocycler (ThermoFisher). ...	36
Table 8. FastDigest reaction mixture.	37
Table 9. Ligation reaction mixture set up.	40
Table 10. TOPO cloning reaction mixture.	41
Table 11. Induction times and temperatures of cultures.....	44
Table 12. Purified pET vector concentrations analysed via Nanodrop.....	51
Table 13. Purified pET vector concentration analysed by Nanodrop.....	53
Table 14. Optimal PCR reaction conditions per construct.....	57
Table 15. Concentrations of domain inserts following purification via Nanodrop.	59
Table 16. Chosen expression conditions for each construct.	70
Table 17. Summary of various expression systems.	84

List of Abbreviations

AML	Acute Myeloid Leukaemia
ArnA	Bifunctional Polymyxin Resistance protein
BCA	Bicinchoninic Acid assay
BiFC	Biomolecular Fluorescence Complementation
BLI	Bio-layer Interferometry
CHO	Chinese Hamster Ovary cells
Clf	Clumping factor
Co-IP	Co-immunoprecipitation
Construct 1A	EMI pET19b
Construct 1C	gC1q pET28a
Construct 2A	EMI pET19b
Construct 2B	EMI pET24a
Construct 2C	gC1q pET19b
Construct 2D	gC1q pET24a
Construct 2E	Coiled coil + EMI pET19b
Construct 2F	Coiled coil + EMI pET24a
Construct 2G	EGF-like pET19b
Construct 2H	EGF-like pET24a
D	Insoluble
DMSO	Dimethyl Sulfoxide
Eap	Extracellular Adherence Protein
EC	Endothelial Cells

ECM	Extracellular Matrix
EDTA	Ethylenediaminetetraacetic Acid
Efb	Extracellular fibrinogen-binding protein
EGF-like	Epidermal Growth Factor
EMILIN	Elastin Microfibril Interface-located Protein
FLIPr	Fibrinogen-like Protein A-interacting Protein
Fnbp	Fibronectin-binding Protein
FRET	Fluorescence Resonance Energy Transfer
FV	Factor 5
FX	Factor 10
GlmS	Glucosamine Fructose-6-phosphate Aminotransferase
HEK	Human Embryonic Kidney cells
HUS	Hemolytic Uremic Syndrome
I ₄₅	45-minute induction
IgG	Immunoglobulin G
I ₀	Uninduced
I _{0/n}	Overnight induction
I ₁₂₀	120-minute induction
IPTG	Isopropyl β -d-1-thiogalactopyranoside
IsdB	Iron-regulated determinant B
ITC	Isothermal Titration Calorimetry
LexSy	Leishmania Expression System
MMRN1	Multimerin-1

MMRN 2	Multimerin-2
MW	Molecular weight
PBS	Phosphate-Buffered Saline
PCR	Polymerase Chain Reaction
pI	Isoelectric Point
PLA	Proximity Ligation Assay
PPI	Protein-Protein Interaction
qPCR	Quantitative real-time PCR
QPD	Quebec Platelet Disorder
S	Soluble
Sbi	Staphylococcal Binder of Immunoglobulin
SDS-PAGE	SDS-Polyacrylamide Gel Electrophoresis
SEC	Size Exclusion Chromatography
SdrE	Serine-aspartate Dipeptide Repeat E
SNP	Single Nucleotide Polymorphism
SpA	Protein A
SPR	Surface Plasmon Resonance
SraP	Serine-rich Adhesin for Platelets
TBE	Tris-Borate EDTA
TBST	Tris-Buffered Saline with Tween
TOPO	Topoisomerase Based Cloning
uPA	Urokinase Plasminogen Activator
VacA	Vacuating cytotoxin A

VEGF

Vascular Endothelial Growth Factor

Abstract

Multimerin-1 (MMRN1), a large mammalian glycoprotein, has emerged as a pivotal player in various physiological processes, including blood coagulation, angiogenesis, and vascular haemostasis. As a novel target for pathogenic proteins, such as Extracellular fibrinogen binding protein (Efb) from *Staphylococcus aureus* and Vacuolating cytotoxin A (VacA) from *Helicobacter pylori* and potential as a cancer biomarker with its differential expression monitored across various cancer types, positions MMRN1 as a compelling candidate for further exploration. However, little is known about MMRN1 structure-function relationship and how it links to its physiological function. This thesis aimed to explore the MMRN1 domain's structural characteristics, functional attributes, and molecular interactions by employing biochemical assays to explore the expression and cloning conditions that enhance its soluble production. The analysis of one construct had been successfully confirmed while discussing major obstacles and optimisation for future exploration. By investigating domain-specific growth conditions, this will inform novel studies surrounding their structure, function, and interactions with other proteins to provide insight into MMRN1's role in health and disease and potential implications in diagnostics and therapeutics.

1 Introduction

1.1 Multimerin-1 (MMRN1) structure and functions

1.1.1 MMRN1's physiological role in platelets

MMRN1 is a large multidomain glycoprotein belonging to the Elastin Microfibril Interface-located Protein (EMILIN) family. Encoded by the MMRN1 gene, this glycoprotein is 1228 amino acids long and is expressed in platelets, megakaryocytes, and endothelial cells with various soluble, homopolymeric, and disulfide-linked properties (Leatherdale et al., 2021). MMRN1 is implicated in several roles across the body, with crucial roles in vascular biology and haemostasis (Saini et al., 2020). MMRN1 is a protein of interest due to its potential involvement in platelet function and vascular homeostasis with roles pertaining to haemostasis and coagulation influencing the release of factor V/Va and the generation of thrombin (Parker et al., 2016) (Fig. 1).

Platelets, known as thrombocytes, are derived from megakaryocytes and are a component of blood. While traditionally, their primary physiological role as a haemostatic agent is to work with coagulation factors to clot haemorrhages to preserve haemostasis, platelets have now been acknowledged as key players in immune responses and inflammatory processes (Holinstat, 2017). These multifunctional cells have been linked to the regulation of inflammatory cascades, immunological surveillance, and infection response and support pro-inflammatory mechanisms, including phagocytosis and leukocyte migratory control (Storey and Thomas, 2015). Platelets express surface receptors and release various bioactive molecules from their alpha granules upon activation, such as inflammatory mediators, cytokines, and P-selectin, thereby influencing the recruitment and activation of immune cells (Scherlinger et al., 2023) (Fig. 2). MMRN1 is one of the proteins stored in these granules, and its release upon platelet activation contributes to its role in processes such as angiogenesis and vascular haemostasis.

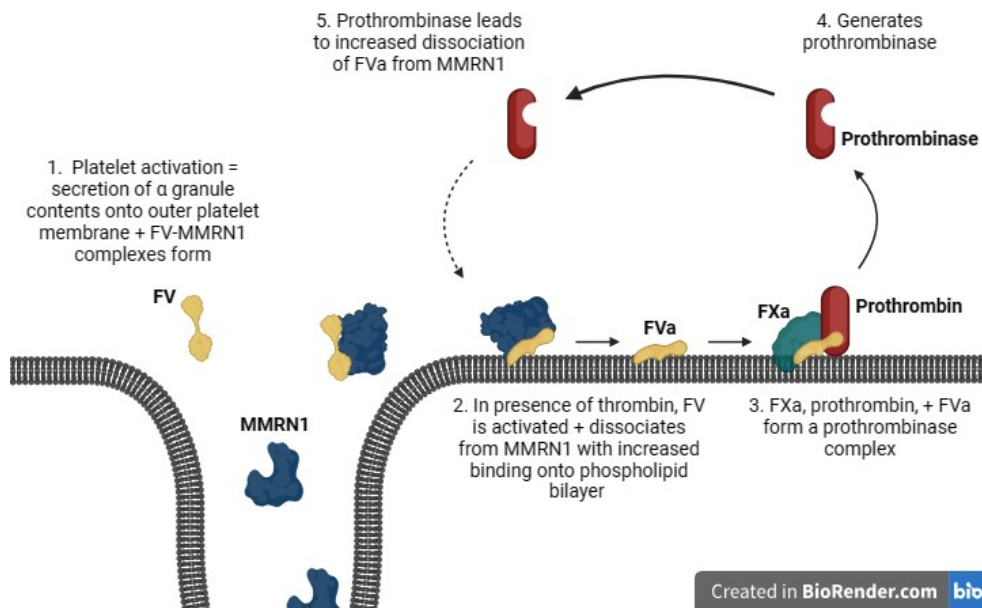


Figure 1. MMRN1 acts as a binding site for FV to regulate thrombin production (adapted from Jeimy et al. (2008)). Alpha granules assemble extra MMRN1 in proportion to FV while at rest. The alpha granules are released, and the FV-MMRN1 complexes are secreted and localised on the outside of the platelet membrane when the platelet is activated. When thrombin is present, FV is activated and separates from MMRN1. By preventing the mechanism of activation, MMRN1 modulates this reaction. A prothrombinase complex is created when Factor Va binds to Factor Xa and prothrombin to encourage the formation of thrombin (figure prepared by author, created with BioRender.com).

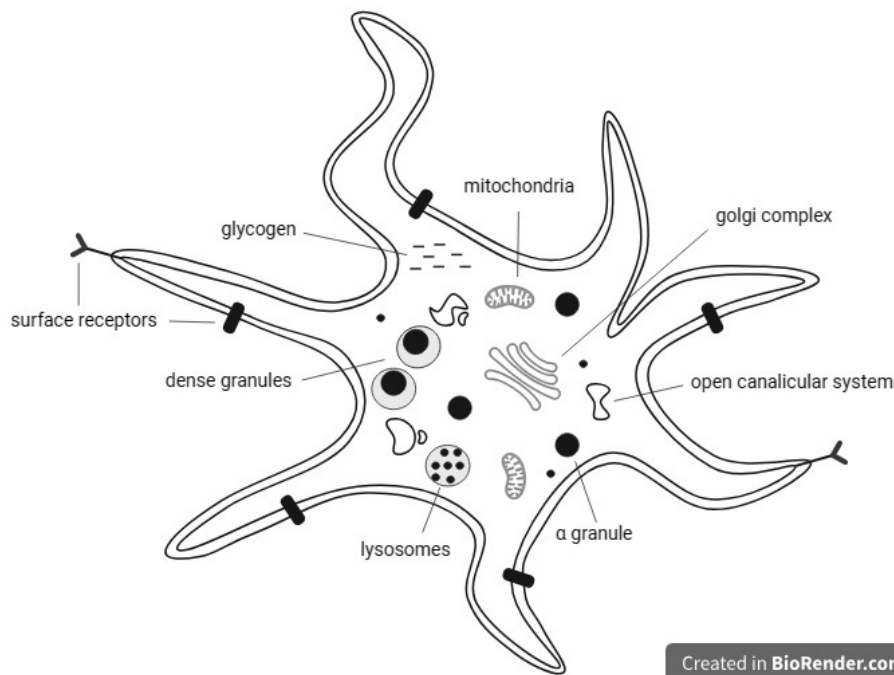


Figure 2. Granular content and anatomical structure of a platelet (adapted from Selvadurai and Hamilton (2018)). Alpha granules are membrane-bound organelles within platelets that contain various proteins involved in blood clotting, wound healing, and other physiological processes. MMRN1 is stored in the alpha granules of platelets (figure prepared by author, created with BioRender.com).

1.1.2 MMRN1 deficiency in platelets causes Quebec platelet disorder (QPD)

QPD is a rare bleeding disorder characterised by abnormally high levels of FV associated with a MMRN1 deficiency (Blavignac et al., 2011). The autosomal dominant disorder was first identified in a French-Canadian family in the province of Quebec. QPD presents with a mild to moderate tendency to bleed, with symptoms ranging from easy bruising and nosebleeds to more severe bleeding episodes. In QPD, tandem duplication of the *PLAU* gene, which codes for urokinase plasminogen activator (uPA), causes upregulation of uPA expression in megakaryocytes through unidentified mechanisms. The degradation of numerous proteins found in platelet alpha granules, including MMRN1 and FV, is linked to the build-up of uPA, causing plasminogen to convert to plasmin in platelets, leading to dysregulated anticoagulant activity and uncontrolled bleeding (Frontini, 2020). While the role of MMRN1 in platelet function is still being investigated, QPD provides insight into the mechanisms of MMRN1 in haemostasis.

1.1.3 MMRN1 as an acute myeloid leukaemia (AML) biomarker

Molecular indicators found in tissues and bodily fluids play a role in detecting, monitoring, and predicting cancer progression (Henry and Hayes, 2012). EMILIN family members have been at the centre of cancer-related studies when compared to MMRN1 which has yet to be thoroughly investigated, such as MMRN2's contribution to vascular stability and permeability in relation to tumour cells, EMILIN-1 inhibiting migration and invasion of gastric cancer cells, and EMILIN-2's proapoptotic effects on tumour cells (Rabajdova et al., 2016; Qi et al., 2019; Pellicani et al., 2020; Mongiat et al., 2010). MMRN1's cellular location (platelets, extracellular matrix, and endothelial cells) suggests it may have active involvement in the tumour microenvironment (Posner, 2022). Beyond its involvement in haemostasis, MMRN1 has been recently proposed as a potential cancer biomarker in the context of AML, with observed upregulated expression (Laszlo et al., 2015). In Leukaemia, the abnormal production of white blood cells interferes with the normal production of platelets, resulting in thrombocytopenia (Qian and Wen-jun, 2013). Additionally, the haematological malignancy may infiltrate the bone marrow, affecting the microenvironment where platelets are produced (Kokkaliaris and Scadden, 2020). This serves as a clear link between AML affecting platelets and MMRN1 being

a platelet protein with an increased risk of bleeding and difficulty clotting. While MMRN1 has not conclusively been established as a driver for cancer, it provides valuable insight into its implications in cancer-related processes.

MMRN1's differential expression has also been monitored across multiple cancer types (Huang et al., 2012; Chokchaichamnankit et al., 2019). MMRN1 expression is seen to be downregulated in bladder, breast, colon, oesophagus, liver, lung, ovary, prostate, rectum, renal, skin, stomach, testis, thyroid, and uterus cancer and has been seen to be upregulated in pancreatic cancer (Posner, 2022). However, whether this correlates with cancer progression is unknown due to gaps in understanding MMRN1's molecular mechanisms. MMRN1 dysregulation raises questions about the significance of MMRN1's involvement and whether it is a result of the tumour's impact on the protein or whether MMRN1 expression aids in tumour development or progression. This makes it a compelling candidate to be further explored in the context of cancer research and to explore its potential as a diagnostic marker for specific malignancies.

The nature of MMRN1's involvement in cancer biology is an ongoing investigation, with findings pointing to its implication in angiogenesis, tumour growth, and metastasis (Zhao et al., 2020; Laszlo et al., 2015; Keeratichamroen et al., 2020). Platelets are associated with metastasis; for example, cancer-cell interaction may result in the evasion of natural killer cells and vascular wall adherence to allow extravasation. Activated platelets may recruit the extracellular matrix (ECM) and granulocytes to provide an environment for the beginnings of cancer growth, and activated platelets contribute to tumour angiogenesis through the release of pro-angiogenic factors, such as vascular endothelial growth factor (VEGF) (Haemmerle et al., 2020; Gkolfinopoulos et al., 2020; Lucotti and Muschel, 2020). Fibroblasts associated with tumours may also remodel the ECM, further supporting the survival and proliferation of tumour cells (Henke et al., 2020; Cromar et al., 2012). Given its expression in endothelial cells (EC), platelets, and the ECM, MMRN1 may influence angiogenic and metastatic processes crucial for tumour sustenance and progression.

1.1.4 MMRN1 domains

Based on sequence-alignment and homology prediction, MMRN1 contains three domains, the gC1q and EMI domain, both shared within the EMILIN family members, and the EGF-like domain being unique to MMRN1 (Colombatti et al., 2011) (Fig. 3).

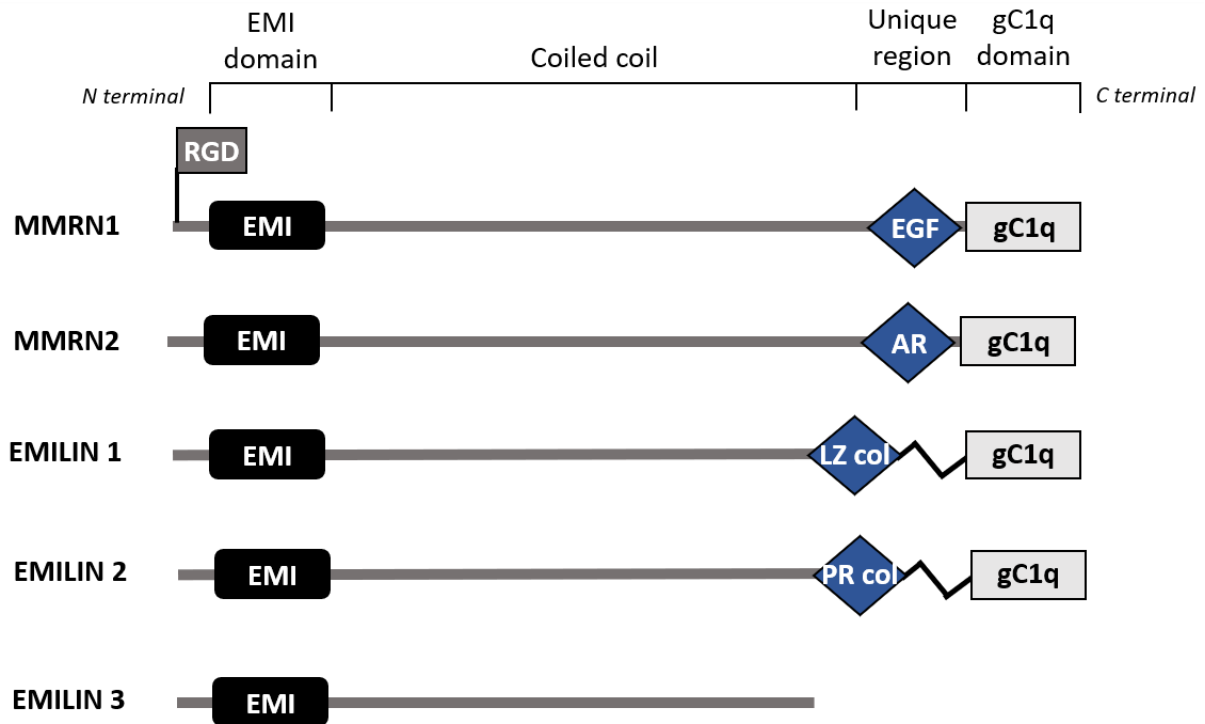


Figure 3. Overview of EMILIN family members' protein domains (adapted from Colombatti et al. (2012)). A domain is the functional section of a protein. The EMI and gC1q domains and a coiled coil are shared by the EMILIN and Multimerin families, although each protein has a unique domain (shown in blue) or region upstream of the gC1q domain. The unique region of EMILIN-1 is comprised of two leucine zippers and a collagenic sequence (LZ col); EMILIN-2 and MMRN2 each contain proline-rich sequences (PR col) and an arginine-rich sequence (AR); MMRN1 contains an EGF-like domain (EGF) and an RGD motif (RGD), and EMILIN 3 has no unique region (figure prepared by author, created with BioRender.com).

Each domain has dedicated functions, however, while investigated *in vitro*, the functions of each domain have yet to be explored *in vivo* (Fig. 4) (Posner, 2022).

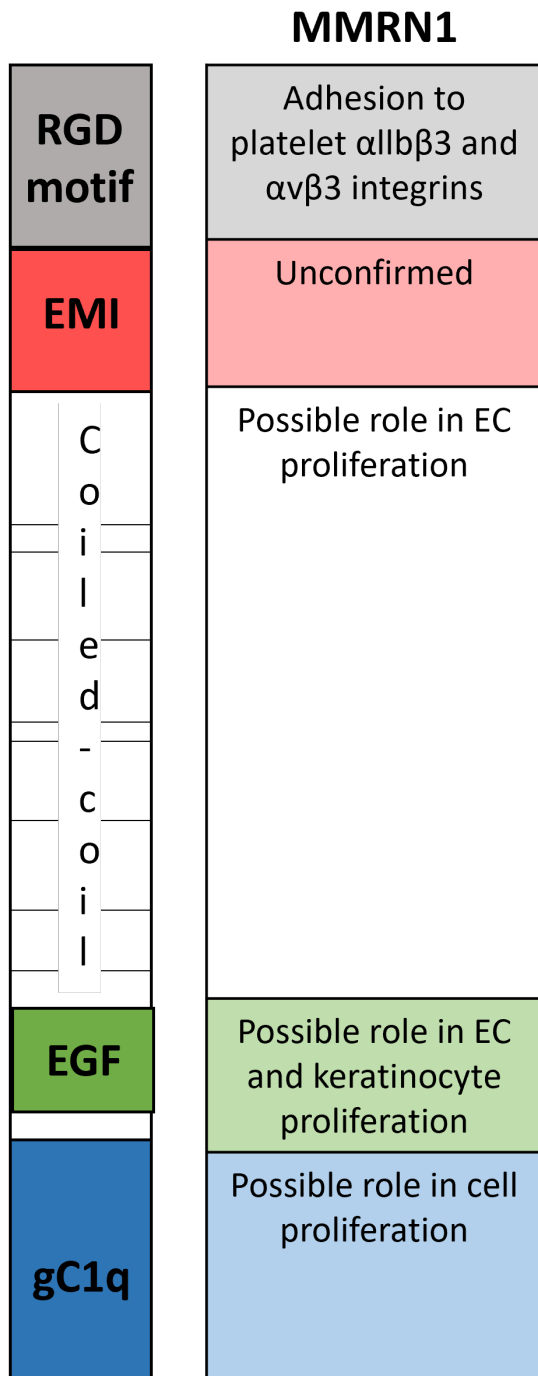


Figure 4. A summary of known and proposed functions of the MMRN1 domains (adapted from Posner (2022)). The MMRN1 domains and RGD motif are outlined on the left, and MMRN1 functions are associated with each segment on the right (figure prepared by author).

The EMI domain is an N-terminal cysteine-rich region found in various other proteins in literature, including EMILIN and fibrillin (Doliana et al., 2000). Across the EMILIN family, the EMI domain has seven cysteine residues (C1-C7) where the distance between C1 and C2, C5 and C6, and C6 and C7 are highly conserved (Fig. 5). MMRN1 has the only EMI domain in the EMILIN family that does not contain the C2 cysteine residue (Doliana et al., 2000). The

conservation of these cysteine residues forming disulphide bonds contributes to the stability and structure of the domain. Comparatively, the gC1q domain has no conserved cysteine residues but has more overall residue conservation (Fig. 5). It is unclear how the EMI domain performs in MMRN1 and whether it participates in protein-protein interactions (PPI) and signalling processes similar to other family members (Colombatti et al., 2011). The EMI domain is found to mediate PPIs across many proteins, for example, EMILIN-3 and herapin and EMILIN-1 and proTGF-B1 (Zacchigna et al., 2006). While the EMI domain of MMRN1 proves unique, it is likely to have similar interactions and signalling events of the EMILIN family due to its strong consensus of forty-five of the seventy-five amino acid positions (Schiavinato et al., 2012; Doliana et al., 2000).

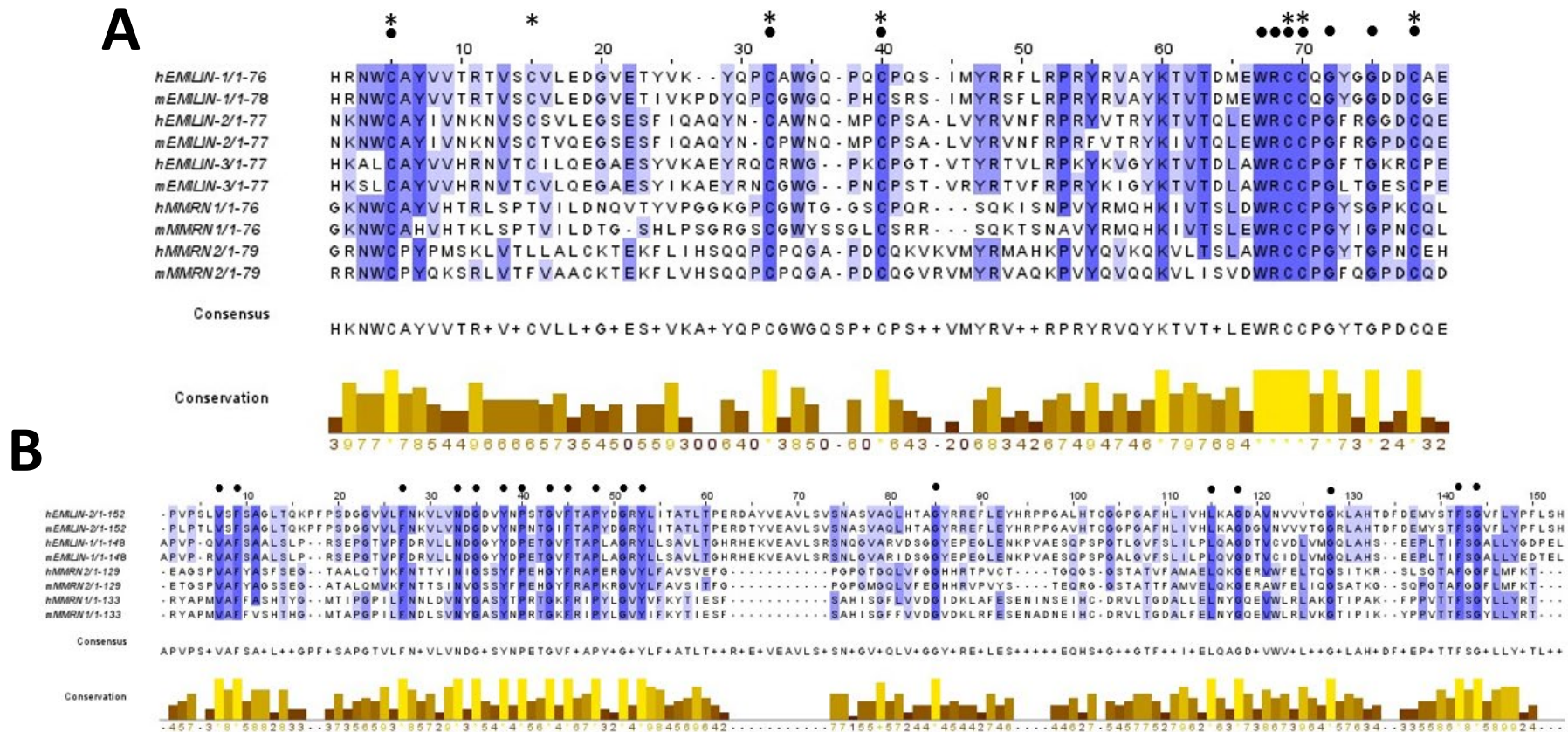


Figure 5. A multiple sequence alignment of both EMI (A) and gC1q (B) domains in the EMILIN family using the Jalview programme (Waterhouse et al., 2009). Protein names are indicated on the left with *m* signifying that it is the mouse protein sequence and *h* being the human protein sequence. The notation following the sequence name indicates the range of residues included in the displayed sequence; for example, ‘hEMILIN-1/1-76’ represents residues 1 to 76 of the original sequence. A gradient of blue is used to mark conservation with highly conserved residues indicated in dark blue and lightly conserved residues highlighted in light blue. Positions marked with dots are the positions with 100% conservation. The highly conserved cysteine residues for EMI are highlighted with an asterisk. The consensus includes residues conserved in the vast majority (>70%) of the aligned sequences. EGF-like domain conservation was not considered as it is unique to MMRN1.

It has been demonstrated that EMILIN-1 and MMRN1 multimerise into larger complexes up to several million Daltons in size, as well as disulfide-linked trimers (Hayward et al., 1991; Mongiat et al., 2000). Long sections forming coiled coil structures and the gC1q domain are among the regions in the EMILIN family capable of oligomerisation. It has also been found that the gC1q domain is necessary for EMILIN-1's cell adhesion characteristics and its involvement in the supramolecular organisation and multimer assembly of EMILIN-1 (Mongiat et al., 2000; Spessotto et al., 2003). However, a study investigating the multimeric assembly of the EMILIN-3 protein revealed that the protein primarily resides as a higher-order oligomer. It was shown through transfection experiments using deletion constructs that EMILIN-3 assembly may occur without the gC1q or EMI domains, implying that the high cysteine residue coiled coil region is essential for this process (Schiavinato et al., 2012).

With the EGF-like domain being unique to MMRN1 in the EMILIN family, its role in MMRN1 cannot be confidently predicted without further investigation. However, it has been found in proteins associated with cell proliferation, differentiation, and cancer, including various ECM proteins (Song et al., 2015). It has also been linked to calcium binding and in mediating PPIs and likely contributes to the structural and functional properties of MMRN1 within the ECM and cellular environment (Sinha et al., 1998).

The domains may be further visualised through theoretical modelling. Figure 6 presents an image of MMRN1 generated by ColabFold-AlphaFold2 (Mirdita et al., 2022).

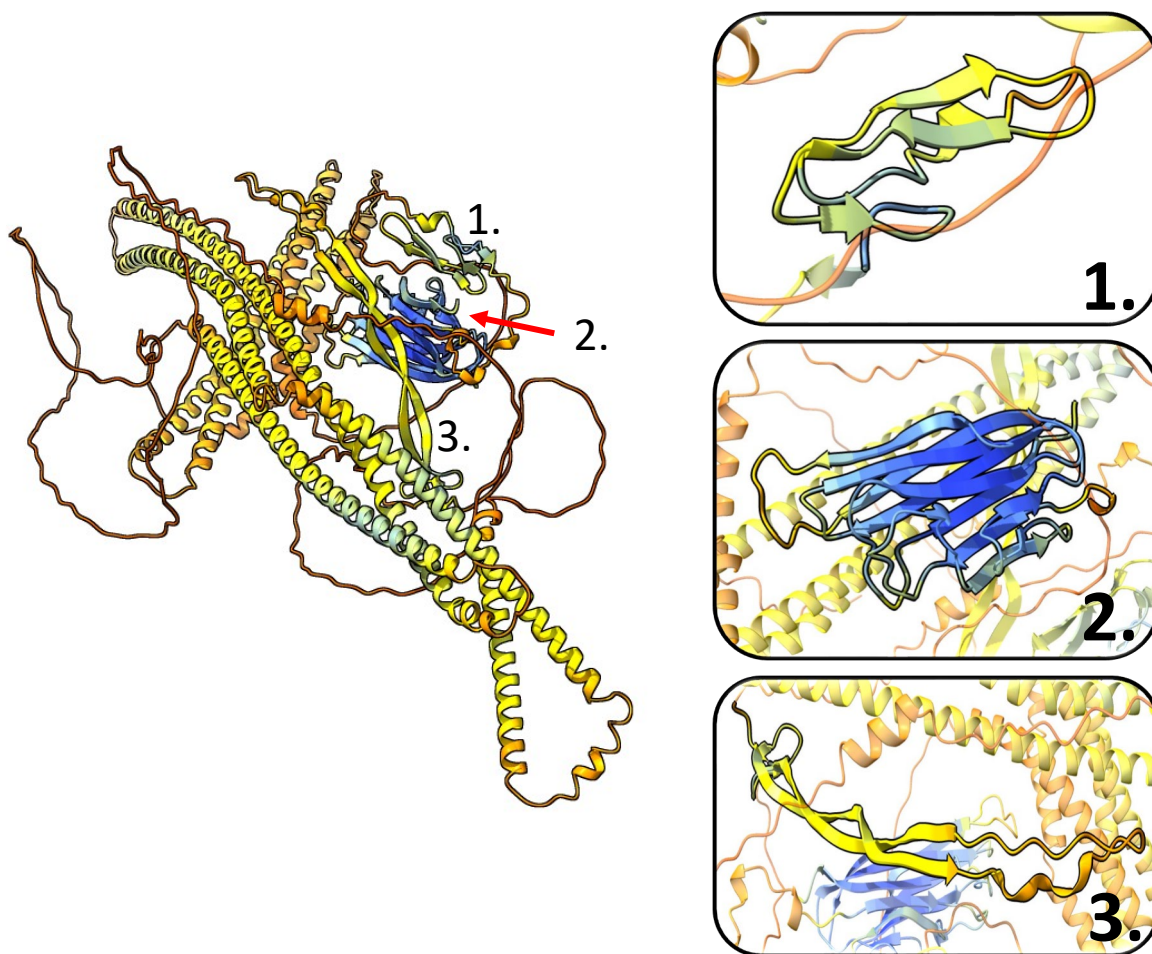


Figure 6. An AI predicted structure of MMRN1 and protein domains using ColabFold-AlphaFold2 (Mirdita et al., 2022). (1) EGF-like (2) gC1q (3) EMI. AlphaFold2 produces a per-residue model confidence score between 0 and 100. Each colour signifies a score bracket: dark blue >90 (very high); light blue 90-70 (high); yellow 70-50 (low); orange <50 (very low). Molecular graphics and analyses performed with UCSF ChimeraX, developed by the Resource for Biocomputing, Visualisation, and Informatics at the University of California, San Francisco, with support from National Institutes of Health R01-GM129325 and the Office of Cyber Infrastructure and Computational Biology, National Institute of Allergy and Infectious Diseases (Meng et al., 2023; Pettersen et al., 2021; Goddard et al., 2018).

AlphaFold2 structure prediction provides information on individual domains, such as the gC1q domain, with high confidence (blue colour) but fails to predict the majority of the MMRN1 structure, as highlighted mostly in yellow/orange (Fig. 6). Therefore, this model is not a representative structure for the glycoprotein as a whole, but in addition to existing literature, it helps to provide a holistic picture when investigating its structural characteristics.

1.2 MMRN1 interactions with pathogenic proteins

Staphylococcus aureus is a gram-positive bacterium that primarily lays dormant on the skin surface; when encountering wounded skin, it may infect deeper in the body (Adalbert et al., 2021). It is a human pathogen responsible for many community and hospital-borne infections, ranging from mild wound infections to severe invasive illnesses such as bacteraemia and endocarditis (Tong et al., 2015). As an invading pathogen, *S. aureus* seeks to evade immune system responses through various mechanisms of propagation, including the production of proteins to modulate Immunoglobulin G (IgG) binding and inhibit its recognition by Fc receptors (Protein A – Spa, Staphylococcal binder of immunoglobulin – Sbi, and Fibrinogen-like Protein A-interacting Protein – FLIPr) and forming a capsule around its cell wall to prevent surface recognition (Kuipers et al., 2016).

To propagate infection, *S. aureus* typically releases virulence factors, such as the Extracellular fibrinogen-binding protein (Efb), known for inhibiting convertase activity, platelet aggregation, and haemostasis inhibition to prevent wound healing, increasing *S. aureus* survivability (Ko et al., 2016). Efb's immunosuppressive role involves interfering with the complement system and reducing bacterial phagocytosis, while its haemostatic abilities are linked to inhibiting fibrinogen binding by platelets or inhibiting platelet activation in response to fibrinogen binding, although the exact mechanism remains unknown (Lee et al., 2004; Jongerius et al., 2012). Table 1 outlines the dynamic interactions between host platelet proteins targeted by bacterial *S. aureus* proteins, namely Efb. Posner et al. (2016) investigated the haemostatic nature of Efb-platelet interactions and found that the N-terminus of Efb interacts with MMRN1 and P-selectin, respectively. While Efb's interaction with P-selectin prevents platelet-leukocyte complex formation, it remains unclear whether Efb directly binds to MMRN1, impacting platelet aggregation, or if the PPI influences clotting processes.

Table 1. Known mechanisms of *S. aureus* to target platelet proteins. *To date, three interactional mechanisms have been identified: (1) indirect binding of bacteria to plasma protein, such as fibrinogen, that acts as a ligand for platelet receptors; (2) direct binding of bacteria to platelet receptors; and (3) the binding of secreted bacterial products, particularly toxins, to platelets (Hamzeh-Cognasse et al., 2015).*

<i>S. aureus</i> proteins	Platelet proteins	Mechanism of interaction	References
Extracellular fibrinogen-binding protein (Efb)	Multimerin-1	(3) The direct binding of Efb is not clearly established.	Posner et al. (2016)
	P-selectin	(3) Efb interacts with P-selectin preventing platelet-leukocyte complex formation.	Wallis et al. (2022)
Serine-rich adhesin for platelets (SraP)	Unknown	High affinity binding for platelets. No identified platelet protein binding partners.	Siboo et al. (2005)
Serine-aspartate dipeptide repeat E (SdrE)	Unknown (fibrinogen mediated)	(1) When binding with platelets, SdrE requires fibrinogen to be present.	Hamzeh-Cognasse et al. (2015)
Protein A (SpA)	FcγRIIIa (Immunoglobulin mediated)	(1) Immune complexes formed with SpA may attach to FcγRIIIa, the immunoglobulin receptor on platelets.	Hamzeh-Cognasse et al. (2015)
	GPIbα (vWF mediated)	(1) SpA may attach to vWF, which binds to GPIbα.	Hamzeh-Cognasse et al. (2015)
	Globular heads of C1q receptor (gC1qR)	(2) Direct adhesion of SpA to gC1qR, a platelet surface receptor.	Nguyen et al. (2000)
Extracellular adherence protein (Eap)	Glycosaminoglycans	(3) Eap, as an oligomer, directly binds to glycosaminoglycans present on platelets.	Heptinstall (2012)
Clumping factor A (ClfA)	GPIIb-IIIa (fibrinogen mediated)	(1) ClfA binds to the C-terminal of the fibrinogen γ-chain to interact with GPIIb-IIIa.	Hamzeh-Cognasse et al. (2015)
Clumping factor B (ClfB)		(1) ClfB binds to the C-terminal of the fibrinogen α-chain to interact with GPIIb-IIIa.	
Fibronectin-binding proteins (FnbpA, FnbpB)	GPIIb-IIIa (fibrinogen mediated)	(1) FnbpA and FnbpB bind to the C-terminal of the fibrinogen γ-chain to interact with GPIIb-IIIa.	Hamzeh-Cognasse et al. (2015)
Iron-regulated surface determinant B (IsdB)	GPIIb-IIIa (fibronectin mediated)	(1) IsdB binds to GPIIb-IIIa in the presence of fibronectin.	Hamzeh-Cognasse et al. (2015)

MMRN1 has also been identified as a target for Vacuolating cytotoxin A (VacA), a virulence factor produced by *Helicobacter pylori* (Satoh et al., 2013). It is unclear how this interaction results in VacA-induced platelet activation. It is hypothesised that VacA binding to MMRN1 may result in decreased regulation of factor V, which in turn increases thrombin generation and platelet activation (Satoh et al., 2013). Alternatively, VacA may bind to MMRN1 to promote interaction with $\alpha v\beta 3$ and $\alpha IIb\beta 3$ platelet activation receptors, increasing platelet activation. However, although the MMRN1-VacA interaction is not the focus of this project, it poses an interesting research question on the structure-function relationship of MMRN1 interactions with both Efb and VacA, as these interactions produce opposite outcomes.

1.3 Study outline

The MMRN1 domains served as the foundation for this research project. This study set out to investigate how the Efb protein interacts with MMRN1 through pull-down assays (Co-Immunoprecipitation) and structural analyses (X-ray crystallography) to further explain MMRN1's physiological roles and implications in other signalling pathways. The problem remains that there is currently a lack of structural information and the structure-function relationship of MMRN1 in the existing literature, with only 93 publications. Studying Efb-MMRN1 interactions in *S. aureus* pathogenesis is necessary to describe MMRN1's molecular mechanism and may provide information for therapeutic targets treating bacterial infections and Efb's inhibitory effect on platelet activation, informing novel anti-thrombotic therapy. Providing insight into MMRN1's structural significance and domain-related functions may also implicate the protein's involvement in various molecular mechanisms and regulatory pathways. However, the protein must first be cloned, expressed, and purified before being taken further for analysis (Fig. 7).

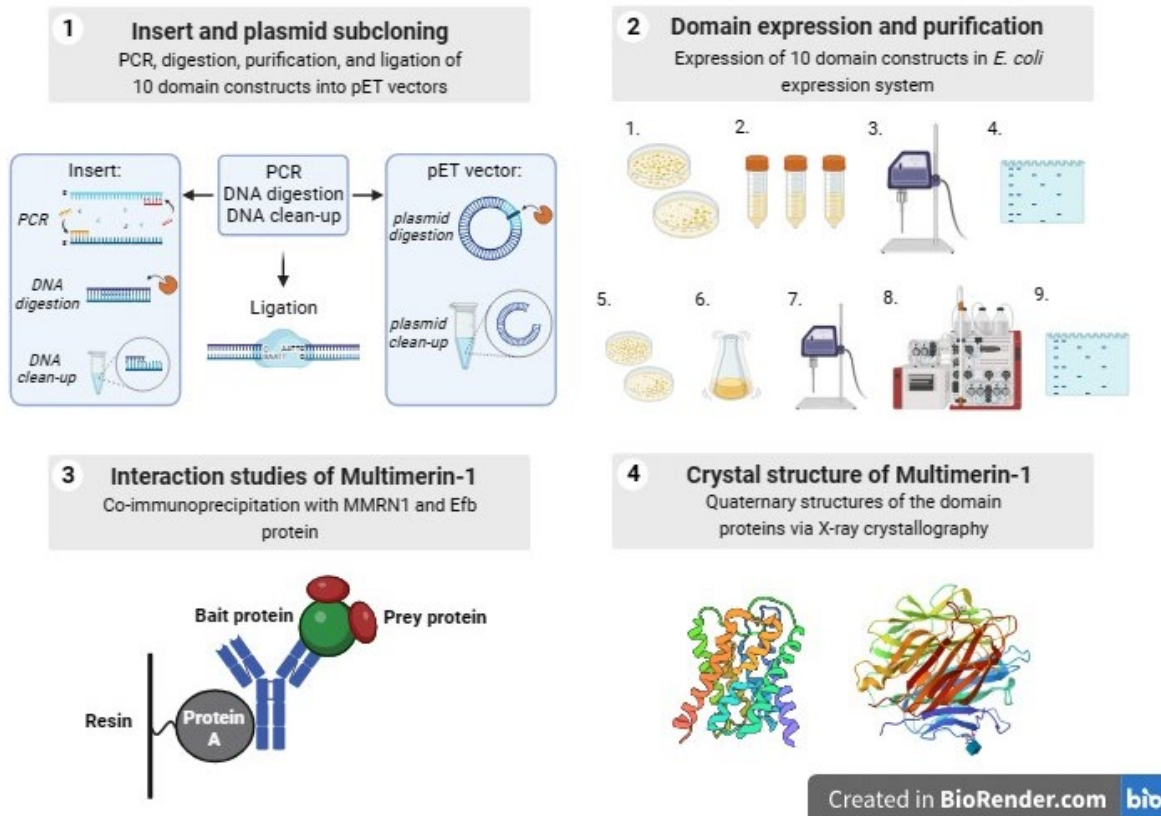


Figure 7. A workflow summary of the proposed experimental workflow of MMRN1 cloning, expression, purification, interaction (co-immunoprecipitation) and structural studies (X-ray crystallography). 1: Outlines the subcloning of the plasmid and insert DNA. 2: Outlines the protein domain expression and purification excluding screening and sequencing with substeps: (1) transformation and growth of bacterial colonies; (2) small-scale expression trials under different conditions; (3) sonication for soluble and insoluble samples; (4) run samples on SDS-PAGE gel to assess optimal expression conditions; (5, 6) pick bacterial colonies for 2L large scale expression; (7) sonication for soluble and insoluble samples; (8) affinity chromatography purification; (9) run purified samples on an SDS-PAGE gel for analysis. 3: Outlines co-immunoprecipitation of purified MMRN1 domain with purified Efb protein. 4: Outlines example X-ray crystallography results for structural analysis (Source: PDB; Berman et al., 2000) (figure prepared by author, created with BioRender.com).

1.3.1 Study aims and objectives

This project aims to identify expression conditions that will lead to the purification of recombinant MMRN1 protein constructs.

To achieve this, the following objectives during the project were as outlined (Fig. 7):

1. Design and optimise primers for specific MMRN1 domains and clone the domains into an expression vector through polymerase chain reaction (PCR), digestion, and ligation.

2. Select an appropriate expression system and optimise expression conditions to enhance the yield of soluble protein for successful expression of MMRN1 domains.
3. Isolate and purify MMRN1 domains to a high level of purity, ready for further analysis.

1.3.2 Study hypothesis

The hypothesis of this study is that the optimised experimental protocols will yield high-quality recombinant protein constructs suitable for structural and functional analysis and characterise the Efb-MMRN1 interaction.

2 Material and methods

2.1 Materials

2.1.1 Preparation of growth media

In accordance with the manufacturer's instructions, LB growth media was prepared and made up per litre of deionised water (Table 2).

Table 2. Growth media composition.

Difco™ LB broth, Miller	Difco™ LB agar, Miller
10 g Tryptone	10 g Tryptone
5 g Yeast Extract	5 g Yeast Extract
10 g NaCl	10 g NaCl
	15 g Agar

2.1.2 Preparation of Buffers

Buffers were prepared and made up using deionised water (Table 3). All buffers and growth media were autoclaved for 20 minutes at 121 °C.

Table 3. General buffer composition.

Application	Name	Composition
DNA agarose gels	1xTBE (pH 8)	0.13 M Tris 45 mM Boric acid 2.5 mM EDTA
Pull down assays (Ni ²⁺ -affinity)	PBS buffer (pH 7.3) (Oxoid™)	8 gm/l NaCl 0.2 gm/l KCl 1.15 gm/l Disodium Hydrogen Phosphate (Na ₂ HPO ₄) 0.2 gm/l Potassium Dihydrogen Phosphate (K ₂ HPO ₄)
Akta Purification (Immobilised metal affinity chromatography)	Binding buffer (pH 8.0)	50 mM Tris 300 mM NaCl 50 mM Imidazole
	Elution buffer (pH 8.0)	50 mM Tris 300 mM NaCl 500 mM Imidazole
SDS-PAGE	2x Laemmli buffer (10 mL)	0.5 M Tris 10% (w/v) SDS 50% (v/v) Glycerol (≥99.5%) 0.5 ml Beta-mercaptoethanol (≥99.0%) 0.1% (w/v) Bromophenol blue
	20x Bolt™ MES SDS Running Buffer (pH 7.3)	50 mM MES 50 mM Tris 0.1% SDS 1 mM EDTA
Western blot	1xTris-Buffered Saline Tween (TBST) (pH 8.0)	150 mM Tris 1.5 M NaCl 0.1% (w/v) Tween 20 (≥10%)
	Western transfer buffer (pH 8.3)	25 mM Tris 200 mM Glycine 10% (v/v) methanol

This project has ethical approval at Manchester Metropolitan University (# 48482). No human material, participants or animals were used in this study.

2.2 Methods

2.2.1 Primer design

Primer sequences to amplify and subclone the EMI, gC1q, EGF-like and coiled coil domains into pET19b, pET24a, and pET28a vectors (Novagen, Merck) were designed using SnapGene. Expsy (Gasteiger et al., 2003) was used to check if the sequences were in frame with the histidine residues to provide the correct amino acids (Table 4) (Fig. 8). The cloning strategy introduced either N-terminal and C-terminal His-tags as the position of the His-tag may affect function or folding of the recombinant protein constructs. Primers were designed following recommended primer conditions such as (Bustin et al., 2020; Ruiz-Villalba et al., 2017):

- 15-30 bases long for specific DNA amplification
- 40-60% GC content to promote binding stability
- 55-65 optimal melting temperature to prevent secondary annealing

Primer pairs belonging to constructs 1A and 1B were synthesised by Sigma-Aldrich, and primer pairs belonging to constructs 2A, 2B, 1C and 1D were synthesised by Integrated DNA technologies. Primers were dissolved in nuclease-free H₂O to a concentration of 100 µM, and aliquots of 0.5 µM were prepared to be used in PCR. Stocks and aliquots were stored at -20 °C.

Table 4. Outline of forward and reverse primer sequences and features for each domain construct referred to in text with a construct ID. Plasmid and vector lengths in base pairs (bp) with N or C terminal His-tags.

Construct ID	Domain /region (MMRN1)	Primer sequence	Destination vector	Plasmid + insert (bp)	Restriction site	GC (%)	Tm (°C)	His-tag (N, C)
1A	EMI (228bp)	Fw: 5' GGCAAGAACTGGTGC GCCTACG 3'	pET19b (5717bp)	5945	<i>NdeI</i>	64	66	N
		Rv: 5' CAGCTGGCACTTGGGGCCGCT 3'			<i>BamHI</i>	71	70	
2A	EMI	Fw: 5' GGCAAGAACTGGTGC GCCTA 3'	pET19b	5945	<i>NdeI</i>	60	63	N
		Rv: 5' CAGCTGGCACTTGGGGC 3'			<i>BamHI</i>	70	61	
2B		Fw: 5' GGCAAGAACTGGTGC GCCTA 3'	pET24a (5310bp)	5540	<i>NdeI</i>	54	61	C
		Rv: 5' CTCAGCTGGCACTTGGGGC 3'			<i>NotI</i>	71	61	
1C	gC1q (396bp)	Fw: 5' AGGTACGCCCCATGGTGGCC 3'	pET28a (5369bp)	5765	<i>NdeI</i>	71	69	C
		Rv: 5' CCTGTACAGCAGGTAGCCGCTGAAGGT 3'			<i>BamHI</i>	59	70	
2C	gC1q	Fw: 5' AGGTACGCCCCATGGTGG 3'	pET19b	6116	<i>NdeI</i>	68	64	N
		Rv: 5' GGTCCTGTACAGCAGGTAGCCG 3'			<i>BamHI</i>	64	65	
2D		Fw: 5' AGGTACGCCCCATGGTGG 3'	pET24a	5709	<i>NdeI</i>	68	64	C
		Rv: 5' GGTCCTGTACAGCAGGTAGCCG 3'			<i>NotI</i>	64	65	
2E	coiled coil + EMI (477bp)	Fw: 5' GGCAAGAACTGGTGC GCCTA 3'	pET19b	6194	<i>NdeI</i>	60	63	N
		Rv: 5' CTCGCTCACCTTGCCCTCCA 3'			<i>BamHI</i>	65	64	
2F		Fw: 5' GGCAAGAACTGGTGC GCCTA 3'	pET24a	5787	<i>NdeI</i>	60	63	C
		Rv: 5' CTCGCTCACCTTGCCCTCCA 3'			<i>NotI</i>	65	64	
2G	EGF-like (111bp)	Fw: 5' GAGTACAGCAGCTGCAGCAGGCA 3'	pET19b	5828	<i>NdeI</i>	61	68	N
		Rv: 5' GGTGCAGTTGTCGCCGGTGAA 3'			<i>BamHI</i>	62	66	
2H		Fw: 5' GAGTACAGCAGCTGCAGCAGGCA 3'	pET24a	5421	<i>NdeI</i>	61	68	C
		Rv: 5' GGTGCAGTTGTCGCCGGTGAA 3'			<i>NotI</i>	62	66	

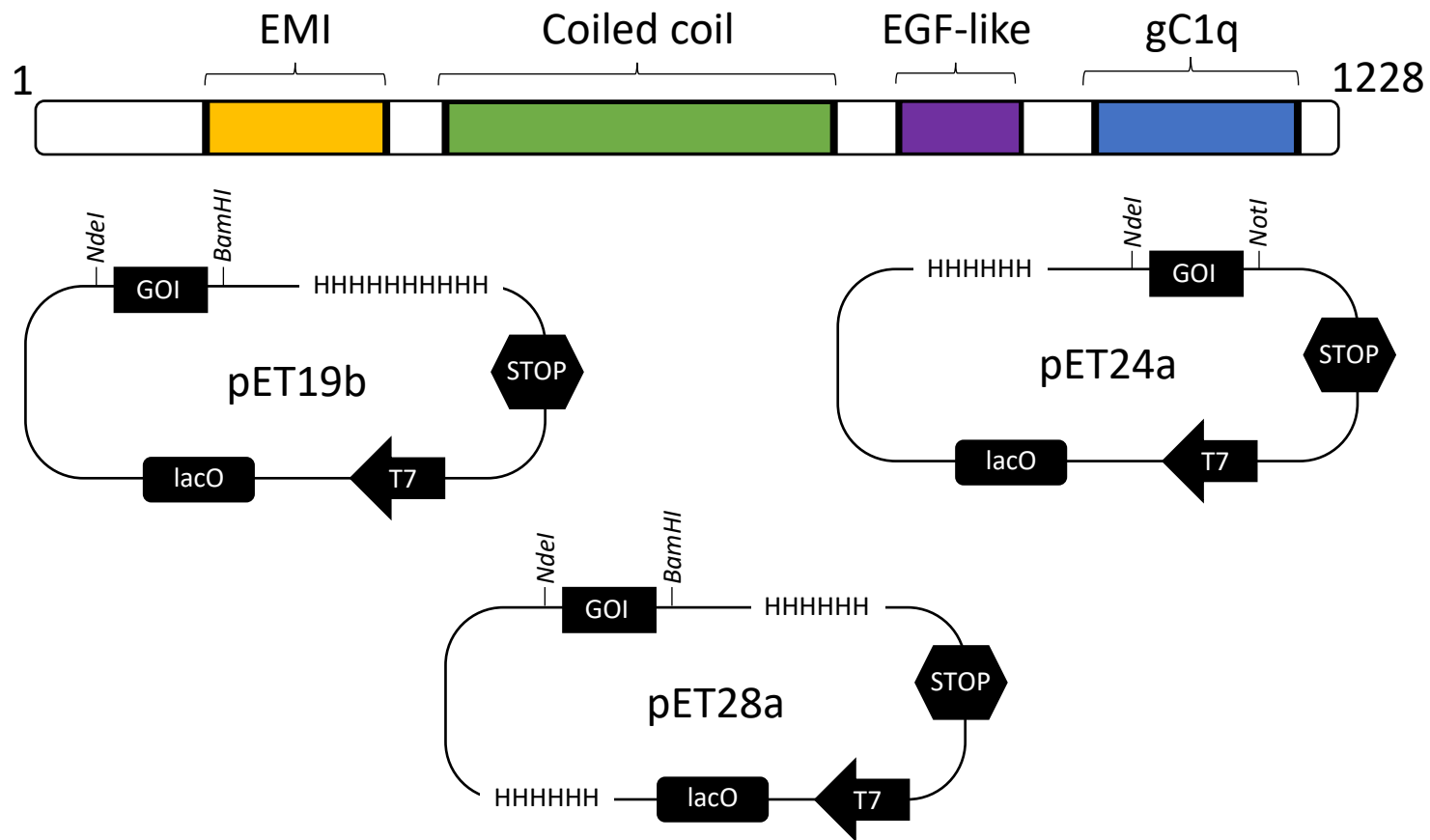


Figure 8. Plasmid maps of pET vectors with an overview of the MMRN1 domains. MMRN1 is 1228 amino acids long with EMI, gC1q, EGF-like and coiled coil domains. Where each domain sequence is inserted into the pET vectors is indicated with gene of interest (GOI). Restriction enzymes used are as indicated in italics, STOP indicates the T7 terminator, T7 indicates the T7 promoter, lacO indicates the lac operator and the His-tag sequence is indicated with either 6xH for pET24a and pET28a or 10xH for pET19b (figure prepared by author).

2.2.2 Growth of bacterial cultures with antibiotics

Ampicillin (Sigma-Aldrich, Merck) for pET19b and Kanamycin (Sigma-Aldrich, Merck) for pET24a and pET28a vectors were the antibiotics added for propagation. Table 5 displays the antibiotic stock concentrations and final working concentrations. Plates were poured and stored in a cold room at 4 °C.

Table 5. Antibiotic stock and working concentrations.

Antibiotic	Stock concentration (1000xstock)	Working concentration
Ampicillin	60 mg/ml	60 µg/ml
Kanamycin	50 mg/ml	50 µg/ml

2.2.3 Plasmid miniprep

Plasmid minipreps were prepared following the manufacturer's instructions (GeneJet, ThermoFisher) (Fig. 9). Two batches of each pET vector were cultured, serving as a backup in the event of issues during the growth/miniprep process. Additional elution steps were carried out to increase the overall yield of plasmid DNA, such as increasing incubation from two to five minutes, heating elution buffer to 70 °C and eluting DNA in two 15 µL elution steps from the silica membrane. The DNA concentration of plasmids was measured using a NanoDrop (ThermoFisher), and plasmid preps were stored at -20 °C following elution.

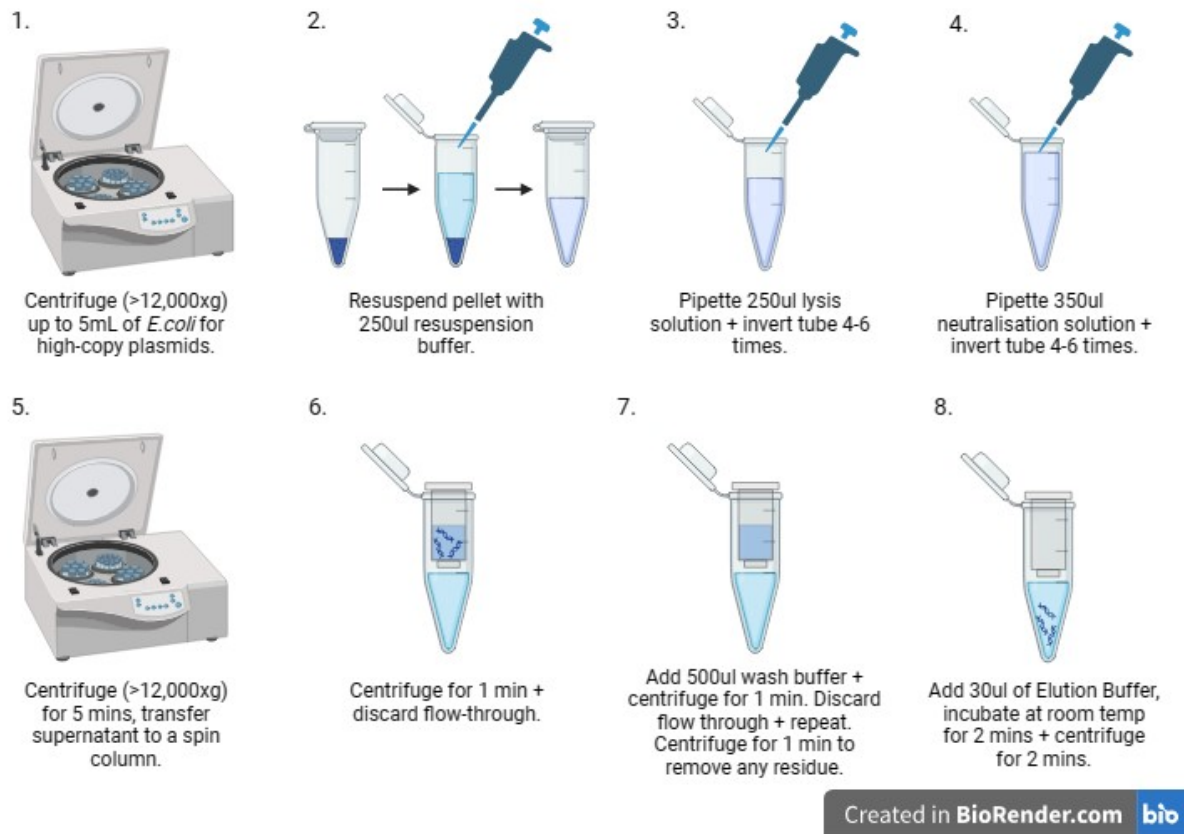


Figure 9. Overview of plasmid Miniprep preparation. *E. coli* containing the *pET* and *MMRN1* plasmids were purified using the GeneJet Plasmid Miniprep Kit (figure prepared by author, created with BioRender.com).

2.2.4 Polymerase Chain Reaction (PCR)

Following the Phusion Flash High-Fidelity PCR master mix protocol (ThermoFisher), the EMI, gC1q, EGF-like domain and coiled coil insert fragments of the MMRN1 coding sequencing were amplified by PCR ready to be digested with corresponding restriction enzymes and ligated into pET19b, pET24a and pET28a plasmids (Table 5). The starting concentration for the MMRN1 template DNA assessed via NanoDrop was 525.6 ng/ μ L and was diluted 1:100. The thermocycler program for insert amplification is summarised in Table 6. Annealing temperatures were screened across a gradient of 60-70 °C for each primer pair.

Table 6. Composition of PCR reactions (20 µL).

Component	20 µL rxn	20 µL rxn (with DMSO)
H ₂ O	add to 20 µL	add to 20 µL
2x Phusion Flash PCR Master Mix	10 µL	10 µL
Forward Primer	0.5 µM	0.5 µM
Reverse primer	0.5 µM	0.5 µM
Template DNA (MMRN1 vector)	12.5 ng	12.5 ng
DMSO	0%	5%

Table 7. Cycling instructions using a ProFlex™ PCR system thermocycler (ThermoFisher).

Cycle step	3-step protocol		Cycles
	Temperature (°C)	Time	
Initial denaturation	98	10s	1
Denaturation	98	0 or 1s	30
Annealing	X	5s	
Extension	72	15s	
Final extension	72 4	1 min hold	1

2.2.5 DNA agarose gel electrophoresis

PCR and plasmid samples were analysed on agarose gels (1%). 1 g of agarose (Meridian Bioscience) was dissolved in 100 mL of 1xTBE buffer mixed with GelRed Nucleic Acid Stain (10,000x DMSO, Millipore) for amplicon visualisation. After the agarose gels were cast, the DNA was loaded into the wells mixed with 5x purple loading dye alongside a 100bp/1kb DNA hyperladder (Bioline) unless stated otherwise. The wide Mini-Sub Cell GT Electrophoresis Cell tank (BioRad) was filled with 1xTBE as the running buffer, and a PowerPac™ HC (BioRad) at 70 V applied an electrical current to move negatively charged DNA from the anode towards

the cathode. The gels were visualised using an Odyssey XF imager (Li-cor), set to the 600 nm wavelength channel.

2.2.6 Restriction enzyme digests

PCR amplicons were digested with restriction enzymes, and the target pET vectors were linearised using matching enzymes for subsequent ligation of inserts into their destination vectors (Table 4). The FastDigest protocol (ThermoFisher) was followed in preparation for digesting the insert DNA (Table 8). All available DNA products were digested to maximise the quantity of DNA at each stage of preparation.

Table 8. FastDigest reaction mixture.

Component	Volume	
	Plasmid DNA (μL)	Unpurified PCR product (μL)
H ₂ O	Add to 20	Add to 30
10x FastDigest Green Buffer	2	2
DNA	17	18
FastDigest enzyme (x2)	1	1
Total volume	20	30

2.2.7 Restriction digest Clean-up

The digested DNA vectors and PCR products were purified using the NucleoSpin® Gel and PCR clean-up kit (Macherey-Nagel) following the manufacturer's instructions (Fig. 10).

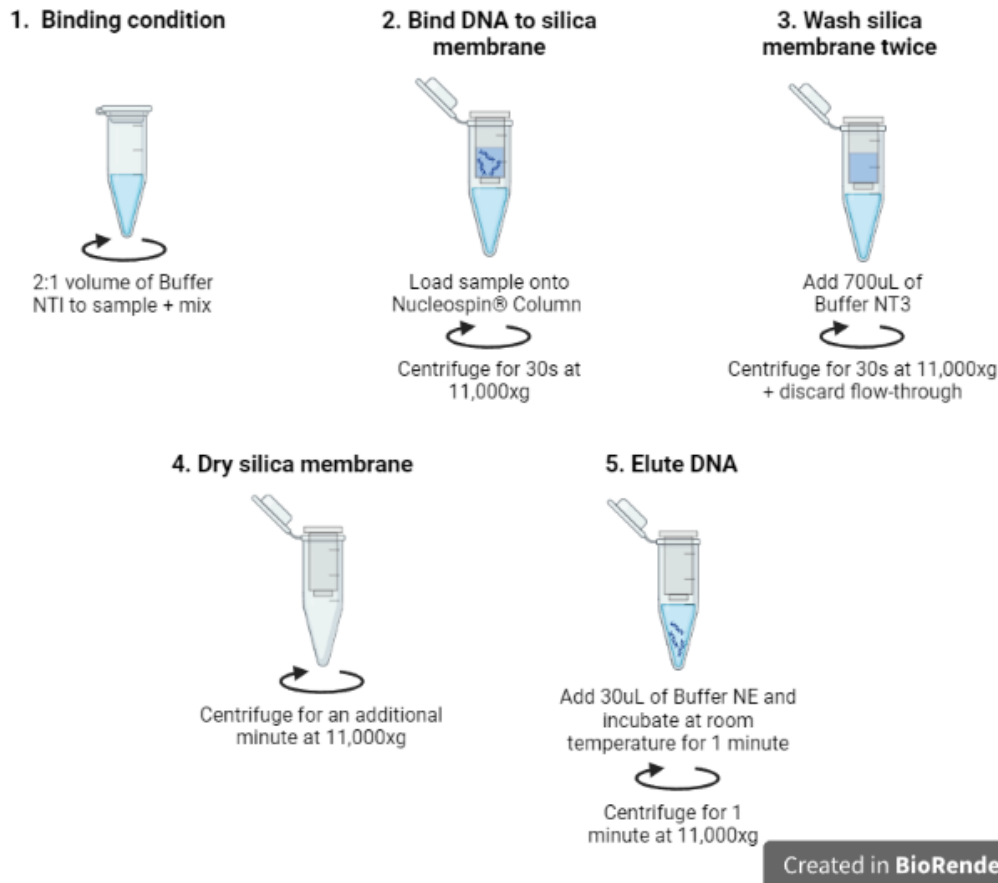


Figure 10. Overview of Gel and PCR clean-up. DNA clean-up was carried out to remove contaminants that may compromise research by purifying the DNA. Washing removes fragments with fewer than 50bp. DNA recovery was increased in step 5 by heating the elution buffer to 70 °C, incubating at room temperature for five minutes and having two elution steps of 15 μ L (figure prepared by author, created with BioRender.com).

2.2.8 Size Selected Clean-up of DNA digests

The coiled coil and EGF-like domain constructs were cleaned up using SPRIselect beads (Beckman Coulter) for size-based selection of DNA (Fig. 11). Right side size selection was conducted for the removal of undesirable larger and smaller fragments of DNA by using DNA charge to separate DNA fragments by length through changing the bead to sample ratio to promote selective separation. The beads have a high positive charge and, therefore, will preferentially bind to larger DNA with more negativity. They bind less efficiently when increasing the bead:DNA ratio, therefore, fewer beads will bind to DNA with higher molecular weight. Right-sided selection is used when wanting to exclude larger fragments selectively. A bead:DNA ratio of 0.7x for EGF-like was used to remove fragments above 400bp, and a bead:DNA ratio of 0.5x for coiled coil was used to remove fragments above 600bp. Following

this, a 1.8x bead:DNA ratio was used to exclude fragments or primer dimers below 50bp.

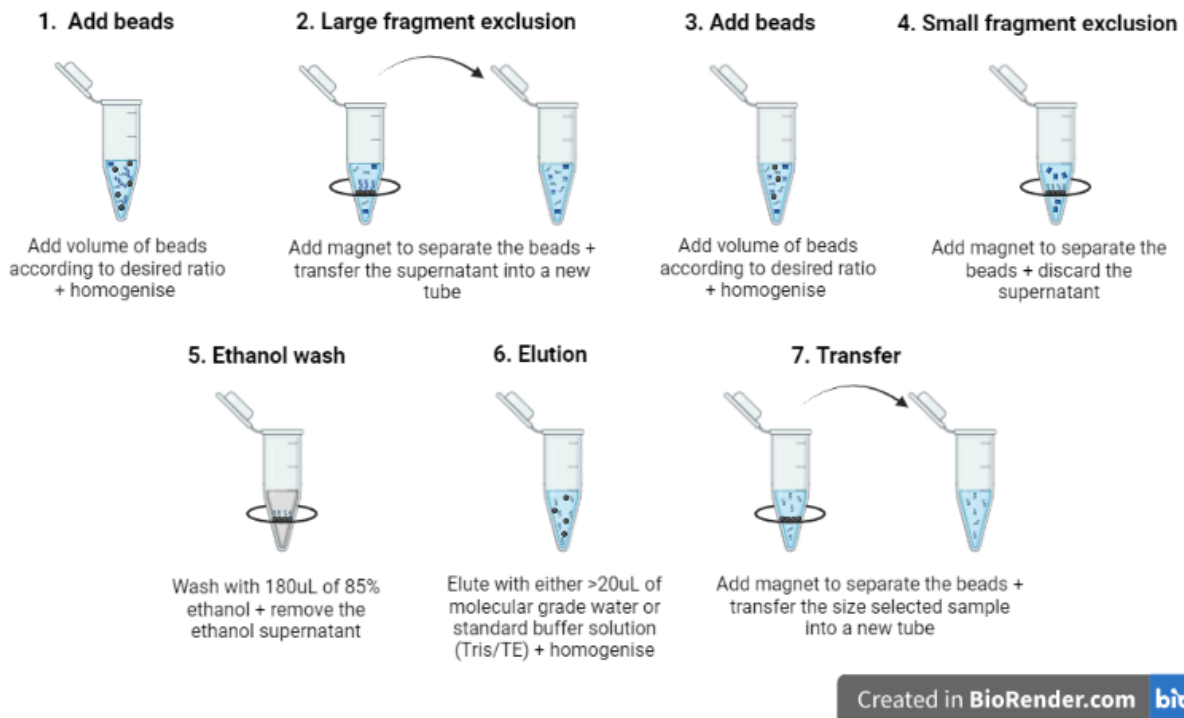


Figure 11. Right side size selection overview. Beads were homogenised via vortex for a minute for consistent size selection. Elution buffer was incubated with the beads for five minutes at 37 °C and vortexed at two-minute intervals to increase elution efficiency (figure prepared by author, created with BioRender.com).

2.2.9 DNA Ligation

Following purification, insert and vector concentrations were determined via NanoDrop to calculate a 5:1 molar ratio for the insert dilution using the NEBio calculator for subsequent ligation (New England Biolabs, 2023) (Table 9). These mixtures were incubated for an hour at 22 °C before storage at -20 °C ready for transformation.

Table 9. Ligation reaction mixture set up. (Catalogue no. EL0011, ThermoFisher)

Component	Volume
Linear vector DNA	20-100 ng (8 μ L)
Insert DNA	5:1 molar ratio over vector
10x T4 DNA ligase buffer	2 μ L
T4 DNA ligase	1 U (0.25 μ L)
H ₂ O	Measure up to 20 μ L
Total volume	20 μL

2.2.10 Additional subcloning methods

In addition to the restriction-ligation method, the TOPO[®] TA cloning[®] kit (Invitrogen[™], ThermoFisher) was used as a method of direct insertion of PCR products into PCR[™]2.1-TOPO[®] cloning vectors for subsequent transformation into One Shot[®] chemically competent *E. coli*. For the PCR products to ligate efficiently with the linearised TOPO vector containing 3' deoxythymidine (T) overhangs, Taq polymerase inserts a deoxyadenosine (A) residue to the 3' ends of PCR products. The TOPO cloning reagents were mixed and incubated for five minutes at room temperature before being placed on ice (Table 10). Following the chemical transformation protocol, 2 μ L of the TOPO cloning reaction was added to 50 μ L of One Shot[™] TOP10 DH5 α cells and incubated on ice for thirty minutes. Cells were heat shocked for thirty seconds at 42 °C and transferred to ice for five minutes. The cells were added to 250 μ L of room temperature S.O.C medium and incubated shaking (200rpm) at 37 °C. 50 μ L of reaction was spread onto plates coated with 80 μ L of X-gal (Invitrogen[™], ThermoFisher) and the appropriate antibiotic to allow for blue-white screening of growth colonies overnight at 37 °C. Several white colonies were picked and grown in culture, ready to be screened for the insert and sent for sequencing.

Table 10. TOPO cloning reaction mixture.

Reagent	Volume
PCR product	0.5-4 μ L
Salt solution	200 mM
H ₂ O	Add to a total volume of 5 μ L
TOPO [®] vector	1 μ L
Final volume	6 μL

2.2.11 Transformation of vectors into DH5 α cloning cell line and BL21(DE3) chemically competent E. coli.

Ligated constructs were transformed into DH5 α (Invitrogen™, ThermoFisher) (Fig. 12). 0.5 μ L of the ligation reaction was added to the cell line and incubated on wet ice for ten minutes before being heat shocked at 42 °C for forty-five seconds, followed by an additional five minutes of incubation on wet ice. The cell line was mixed with 500 μ L of LB broth and shaken for two hours at 37 °C. 150 μ L of the culture was spread onto an agar plate for overnight incubation. Following screening, the same procedure was carried out with BL21(DE3) (ThermoFisher) to transform positively screened clones into an expression cell line. Glycerol stocks were stored using 50% (v/v) glycerol.

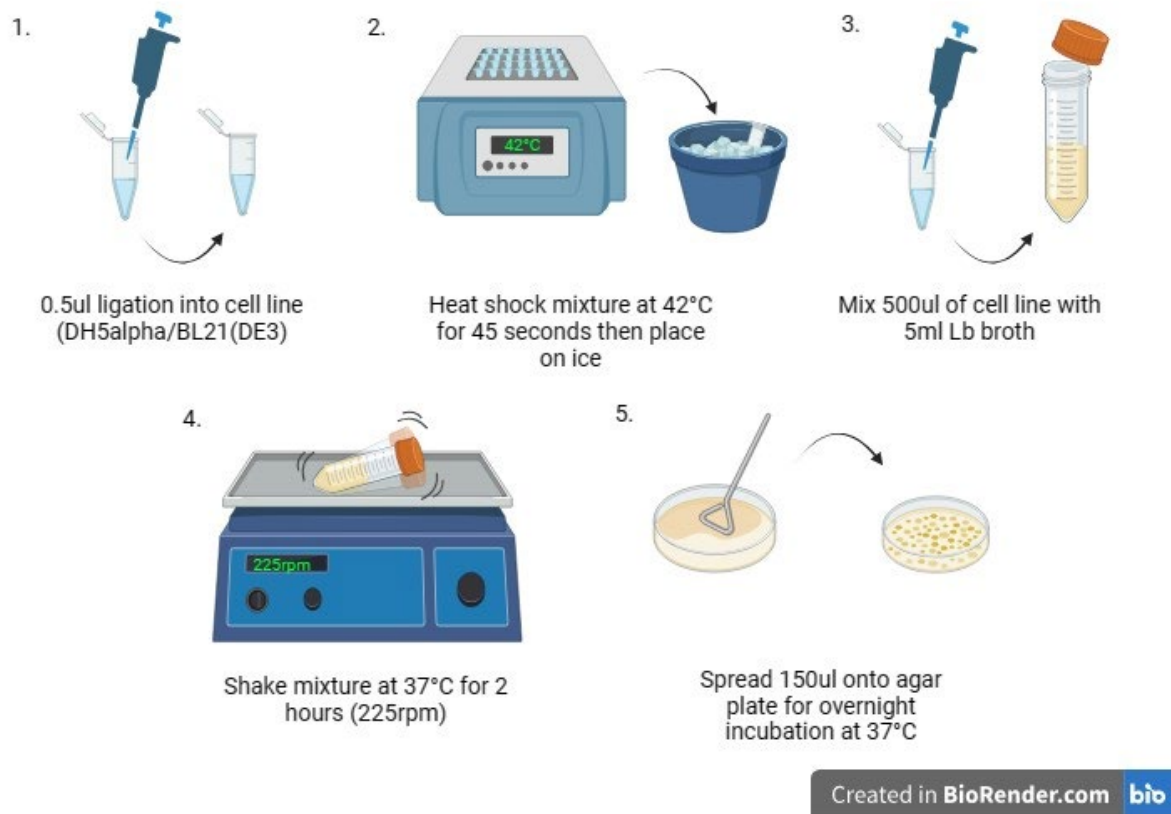


Figure 12. Overview of transformation. (Figure prepared by author, created with BioRender.com).

2.2.12 Screening for successful expression clones

Single colonies of DH5 α , which were transformed with expression vectors, were propagated in liquid culture, and the plasmid was purified using the GeneJet plasmid miniprep kit (Section 2.2.3). Single digests of plasmid DNA were resolved on an agarose gel and compared in size to linearised empty vectors. A drop-down in results comparing the longer domain containing plasmid DNA and the shorter empty plasmids confirmed whether the plasmids contained the target domain inserts. Successful clones were subsequently transformed into chemically competent *E. coli*.

PCR colony screen was also conducted to confirm whether the clones contained the desired domain inserts following the Phusion Flash High-Fidelity PCR master mix protocol (ThermoFisher), however, the first heat step was increased to 98 °C for five minutes to lyse the *E. coli* and release the DNA (Fig. 13). Positive clones from the PCR colony screen were prepared for overnight primary cultures. 5 μ L of LB cultures from all successful screens were

incubated overnight, shaking at 37 °C with 5 mL of LB and the appropriate antibiotic. Following incubation, the primary cultures were purified, and 0.5 µL of plasmid was transformed into chemically competent *E. coli*.

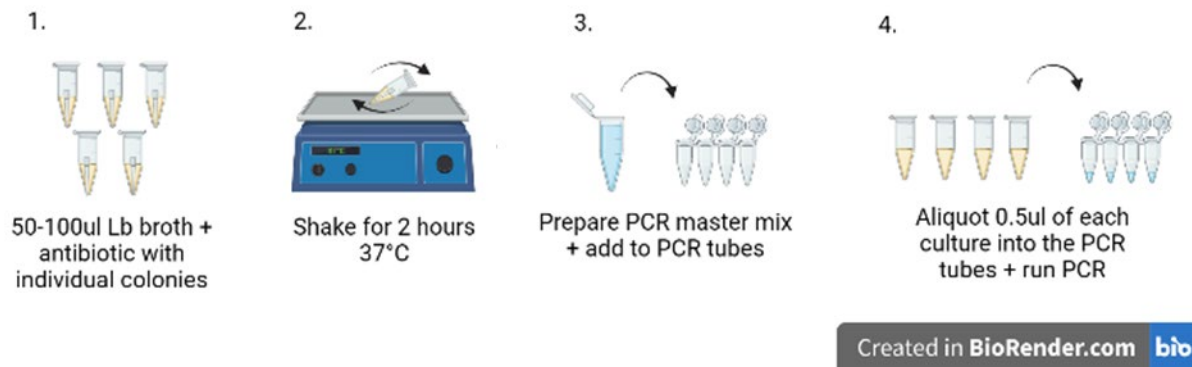


Figure 13. Overview of the PCR colony screen protocol. *PCR colony screen uses the forward and reverse primers specific to each construct to amplify desired inserts if present within the plasmid DNA released from the *E. coli* cells. If positive insert bands were present on the DNA gel, 10 µL of the broth prepared from step 2 was used to inoculate 5-10 mL of LB for subsequent primary cultures (figure prepared by author, created with BioRender.com).*

2.2.13 Sequencing

The most promising clones from initial screening using restriction digest or PCR colony screening were sent for sequencing by Eurofins along with corresponding primers. The sequencing data was compared to the target insert sequence using the multiple sequence alignment software MultAlin (Corpet, 1988).

2.2.14 Small-scale protein expression trials

To identify optimal protein expression conditions, primary cultures were produced using the successful BL21(DE3) colonies and prepared for induction, assessing different culture incubation conditions before scaling up for purification (Table 11).

Table 11. Induction times and temperatures of cultures.

Temperature (°C)	Time (Minutes)		
37	45	90	Overnight
25	120	Overnight	
16	Overnight		

10 mL of LB with the appropriate antibiotic were inoculated with cells from overnight primary cultures (cells harvested from 300 μ L) and incubated for two hours shaking at 37 °C. Once the cultures had reached log phase of growth, OD₆₀₀ of 0.6-0.8, cultures were induced with 0.5 mM Isopropyl β -d-1-thiogalactopyranoside (IPTG) (Invitrogen™, ThermoFisher) (Fig. 14). The BioPhotometer Plus (Eppendorf) was blanked using molecular grade water and the optimal optical density was measured at a wavelength of 600 nm.

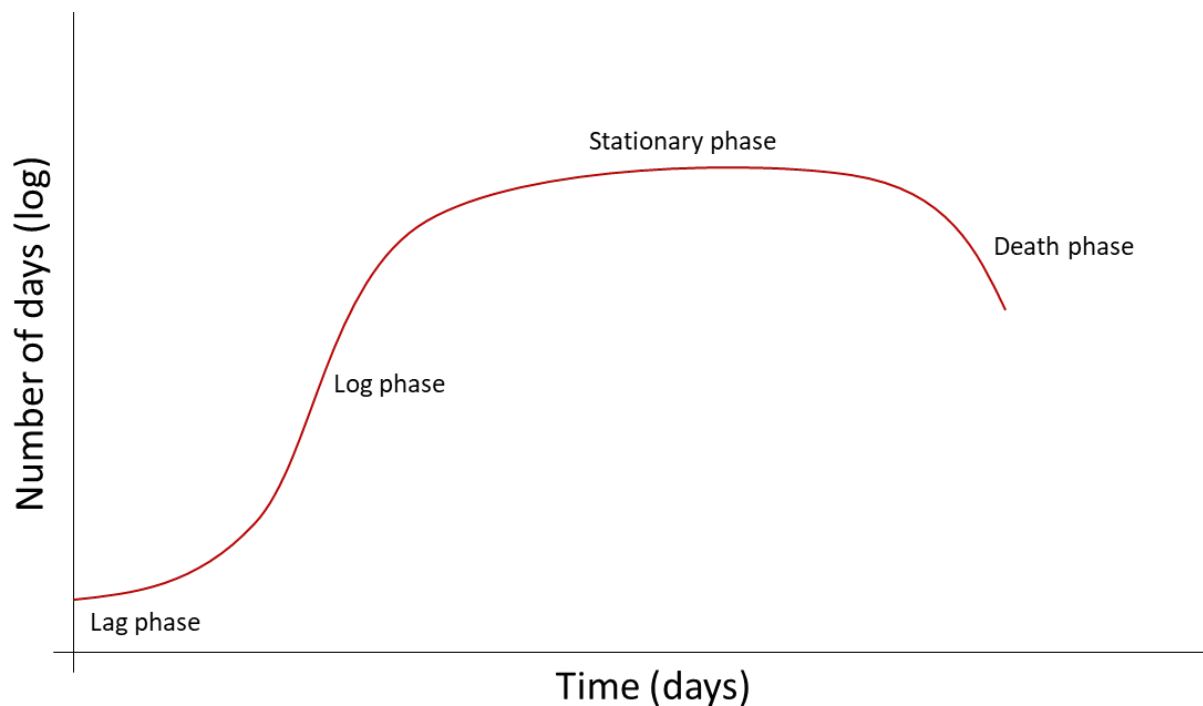


Figure 14. Cell growth curve showing phases of cell growth (adapted from Peleg and Corradini (2011)). Cells were induced during the log phase due to their exponential growth, as they are at the healthiest and most abundant (figure prepared by author).

2.2.15 Large-scale protein expression and purification

Once expression conditions had been established, volumes were scaled up for large-scale expression and purification (100 mL – 2 L). For 100 mL cultures, cells harvested from 1 mL of primary culture was resuspended in 100 mL of LB and incubated for two hours, shaking at 37 °C. For a total of 2 L culture growth, two 1 L cultures were prepared. 20 mL of pelleted primary culture was resuspended in 1 L of LB before a two-hour incubation with shaking at 37 °C. Once the cultures reached an optical density of 0.6-0.8 at a wavelength of 600 nm, the samples were induced with 0.5 mM IPTG and incubated with optimal growth conditions identified in small-scale expression.

2.2.16 Soluble and insoluble sample preparation for SDS-PAGE

500 µL of the uninduced control sample (I_0) was taken from the culture, centrifuged at 11,000xg for two minutes, and mixed with 50 µL of water and 50 µL of 2x Laemmli Buffer. Samples were incubated following each variable condition, as shown in Table 11. Following incubation, 500 µL of whole cell samples were centrifuged at 11,000xg for two minutes before the remaining cultures were prepared for sonication. Both uninduced and induced pelleted samples were mixed with 50 µL dH₂O and 50 µL of 2x Laemmli buffer.

Harvested cells were lysed by sonication in PBS buffer, and the resulting cell lysate was centrifuged at 12,000xg for ten minutes. The supernatant (soluble fraction) was collected, and the pellet (insoluble fraction) was resuspended in the equivalent volume of the supernatant. Samples of each soluble and insoluble fraction were collected for SDS-PAGE analysis, mixing 20 µL of sample with 2x Laemmli buffer.

2.2.17 SDS-Polyacrylamide gel electrophoresis (SDS-PAGE)

Protein samples were loaded into the Bolt™ Bis-Tris Plus Mini Protein gels (4-12%, precast) (ThermoFisher) and placed inside a Mini Gel Tank (Invitrogen™, ThermoFisher) to separate them by molecular weight. The 20x Bolt™ MES SDS Running Buffer (ThermoFisher) was used while the gels were run at 120 A for approximately forty minutes until the dye front reached the bottom of the gel. Samples were heated to 95 °C to denature the proteins before 5 µL of

sample was loaded alongside 5 μ L of the Chameleon[®] Duo Pre-Stained Protein Ladder (Li-cor) unless stated otherwise.

2.2.18 Coomassie Staining

PageBlue[™] Protein Staining Solution (ThermoFisher) was used to stain the proteins following SDS-PAGE. Each gel was washed thrice with 100 mL of dH₂O and had a twenty-second heat step in the microwave between each wash. 10 mL of Coomassie solution was added to the gels and was left to shake overnight at 25 rpm. Following the incubation, the PageBlue[™] solution was discarded, and the gels were rinsed and incubated with H₂O for ten minutes before imaging. The bands were visualised by the ChemiDoc MP Imaging System (BioRad) at a wavelength of 590/110 nm.

2.2.19 Western blot

Western blot was carried out to confirm the expression of the recombinant His-tagged proteins. Following SDS-PAGE, the gel was blotted onto a nitrocellulose membrane via wet transfer (Fig. 15). The proteins were transferred onto the membrane at 100 V for thirty minutes in western transfer buffer using the XCell II[™] Blot Module (ThermoFisher). The membrane was washed with 100 mL of TBST before being blocked with 20 mL of 5% non-fat milk for twenty minutes shaking. Following incubation, the membrane was washed twice with 50 mL of TBST before being incubated with an anti-His-Tag mouse monoclonal IgG antibody (Santa Cruz Biotechnology) (H-3) diluted 1:200 in 15 mL TBST for two hours with gentle agitation. The membrane was washed with four 5-minute washes each in 50 mL TBST for the removal of unbound primary antibody before the Goat anti-Mouse IgG (H+L), Superclonal[™] Recombinant Secondary Antibody (ThermoFisher) was added and diluted 1:1000 in 15 mL TBST for one hour shaking. The membrane was washed for a further four five-minute washes in 50 mL TBST to remove unbound secondary antibody before being visualised by the ChemiDoc MP Imaging System (BioRad) at a wavelength of 680 nm.

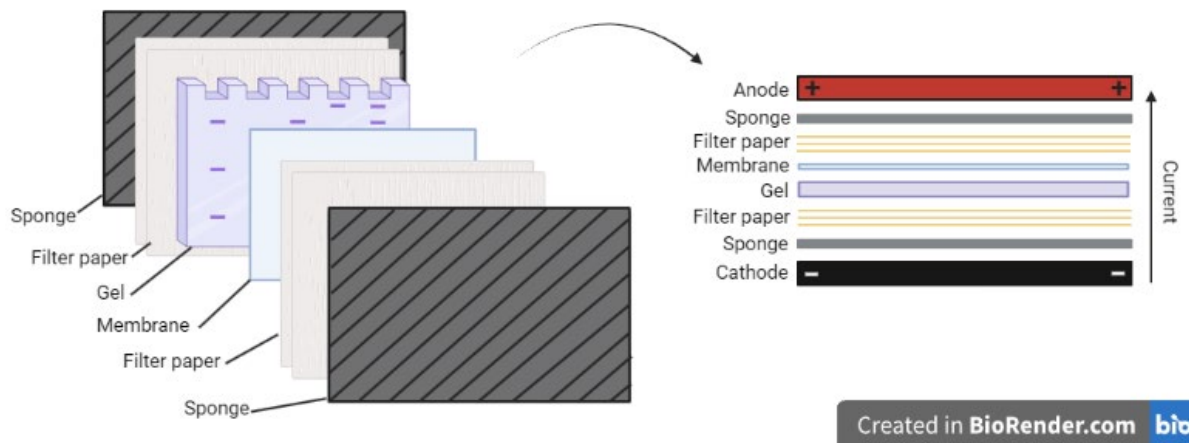


Figure 15. Overview of Western Blotting. *Wet transfer of the proteins is assembled underwater. Filter paper soaked in transfer buffer was placed on top of a fibrous pad, and the gel was placed beneath the membrane before a final stack of filter paper and sponge was placed on top. This transfer sandwich was clamped together with a gel cassette before being placed into the western transfer tank with transfer buffer (figure prepared by author, created with BioRender.com).*

2.2.20 Pull-down assays using Ni²⁺-affinity beads

Another way to detect whether the desired proteins were expressed was via a pull-down assay using the HisPur™ Ni-NTA Resin (ThermoFisher). 10 µL of His resin was equilibrated three times with 1 mL of binding buffer and incubated with 500 µL of sonicated cell lysate for ten minutes with gentle shaking at room temperature. Following incubation, the beads were washed thrice with 500 µL of binding buffer to remove any unbound protein. A sample of the first wash was taken and added to 20 µL of 2x Laemmli buffer to see if the desired protein was being washed out unbound to the beads. 20 µL of 2x Laemmli buffer was added directly to the beads before 5 µL was heated to 95 °C and loaded onto an SDS-PAGE gel ready for analysis.

2.2.21 Immobilised metal affinity chromatography (IMAC) protein purification

The recombinant protein was purified via affinity chromatography on the ÄKTA purifier 10 system (Amersham pharmacia biotech) with the UNICORN software, using the 5 mL HisTrap™ FF Crude (GE Healthcare). At a flow rate of 2.5 mL/min, a blank run was conducted through five-column volumes of water, cleaned with 0.5 M NaOH, followed by five-column volumes of water and ten-column volumes of binding buffer. To allow the protein to bind, the sample was applied at a flow rate of 2 mL/min, and the column was washed with binding

buffer at 2 mL/min for five column volumes to remove any unbound sample. Five column volumes of 20% ethanol run through the column at a rate of 2.5 mL/min. The column was re-equilibrated with two column volumes of elution buffer, washed with two column volumes of 0.5 M NaOH and two column volumes of water, and stored in 20% ethanol to minimise bacterial growth. The sample was eluted in aliquots of 1 mL across a gradient of 100 mM to 500 mM Imidazole (20%-100%).

The Chromatogram demonstrated that the ÄKTA system had not efficiently eluted the fractions containing the target protein and, therefore, the sample was eluted from the column using a Peristaltic Pump P-1 (Cytiva). Once the ÄKTA had loaded the column with the sample, the column was washed with one column volume of binding buffer to ensure that loosely bound protein was removed. 1 mL aliquots were taken across each concentration of Imidazole as follows: 5 mL of 100 mM, followed by 5 mL of 200 mM and finishing with 10 mL of 500 mM at a flow rate of 10 mL/min. 15 mL of 500 mM was collected post-elution to ensure all bound sample had been eluted. The column was washed with two column volumes of water before being stored in 20% ethanol at 4 °C. The protein aliquots were run on an SDS-PAGE ready for analysis.

3 Results

3.1 Molecular cloning

To analyse the EMI, gC1q, EGF-like and coiled coil domains using biochemical and structural approaches, sufficient quantities of purified protein are needed. To achieve this, the coding regions for each domain protein were amplified using PCR, digested, cleaned up and ligated into pET vectors 19b, 24a and 28a, which were then ready to be transformed into DH5 α . The purified colonies were screened for the insert before the recombinant proteins were expressed in *E. coli* BL21(DE3) for protein purification and subsequent biochemical and structural analysis. Each clone was labelled with a construct ID during analysis with the table of constructs summarised in section 2.2.1, table 4.

3.1.1 Plasmid miniprep

The MMRN1-bio-His template DNA (Addgene, #53409) and empty pET19b, pET24a and pET28a vectors (Novagen, Merck) were purified using a miniprep kit from DH5 α *E. coli* cell line as described in section 2.2.3, and purified DNA samples were resolved on 1% DNA agarose gel (Fig. 16).

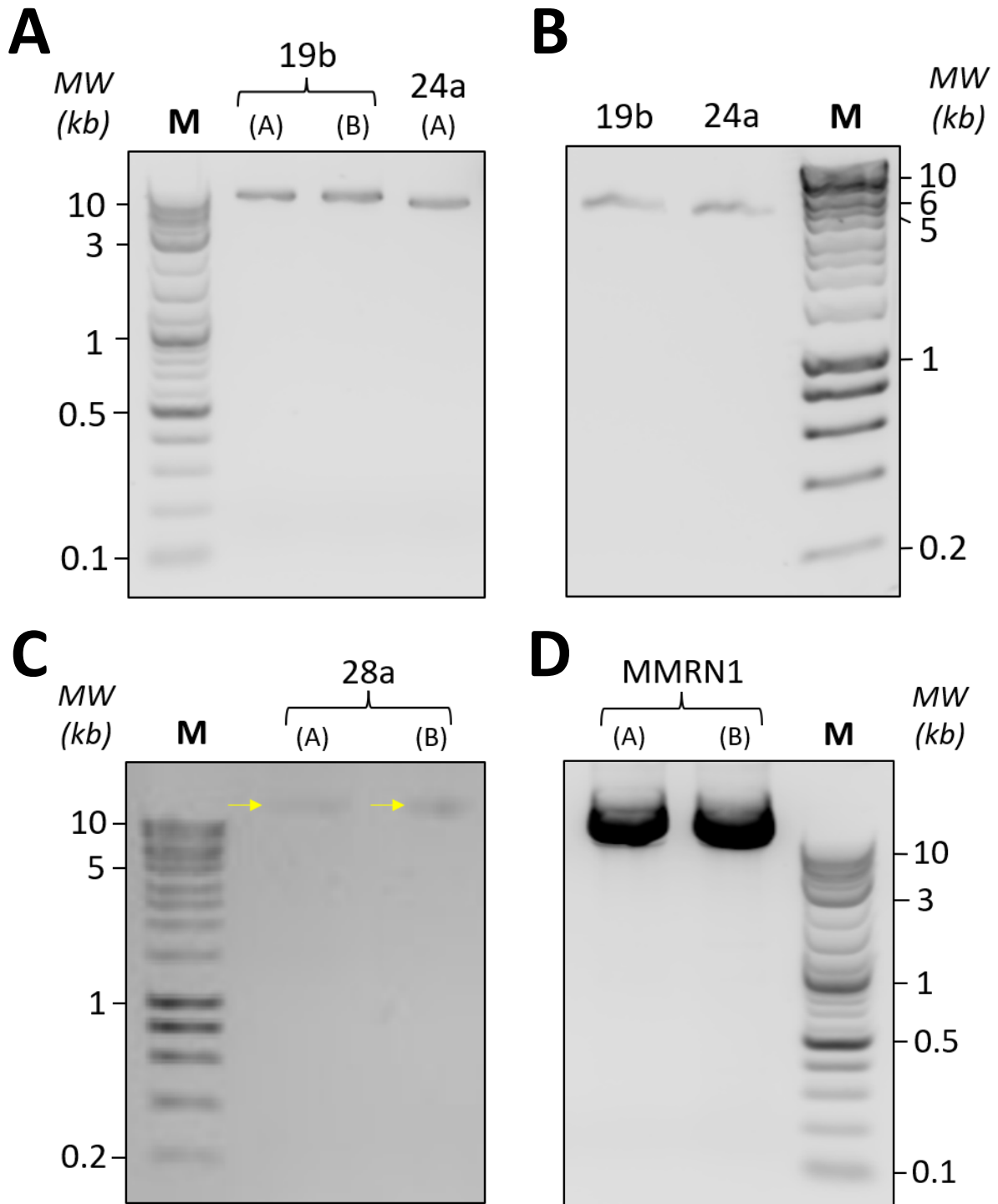


Figure 16. DNA gels (1%) of plasmid minipreps for MMRN1 and pET vectors. *Yellow arrows highlight faint bands. Lane M is represented by a 1kb hyperladder for figures 16b-c and a Quick-load purple 1kb plus DNA ladder (NEB) for figures 16a and 16d. MW., molecular weight.*

- a. Purified pET19b and pET24a vectors.
- b. Purified pET19b and pET24a vectors with additional DNA recovery steps (Section 2.2.3).
- c. Purified pET28a vector.
- d. Purified MMRN1 template DNA.

The purified plasmids from *E. coli* displayed a higher molecular weight that exceeded 10,000bp, which is likely due to their circular nature, causing slower migration through the agarose gel (Fig. 16a, 16c). However, the plasmids in Figure 16b were running at their expected band sizes of 5,310bp for pET24a and 5717bp for pET19b. In contrast to the other plasmids, pET28a appeared to have a lower yield (Fig. 16c). The MMRN1 plasmid appeared to run higher than its expected length of 3,684bp, above 10,000bp, with a concentration of 497.6 ng/ μ L for MMRN1 A and 525.6 ng/ μ L for MMRN1 B as measured by the Nanodrop (Fig. 16d). This could indicate the presence of different topological forms, such as relaxed or open circular structures, that affect its migration through the gel. Additional elution steps (section 2.2.3) increased the overall concentration of the pET vectors (Table 12).

Table 12. Purified pET vector concentrations analysed via Nanodrop.

pET Vector	Concentration (ng/μL)
pET19b B (Fig. 16a)	48.1
pET24a A (Fig. 16a)	55.6
pET19b (Fig. 16b)	63.0
pET24a (Fig. 16b)	71.6
pET 28a A (Fig. 16c)	16.1
pET 28a B (Fig. 16c)	19.2

3.1.2 Restriction enzyme digest to prepare vectors for subcloning

The insert DNA was subcloned into the destination pET vectors, which were digested with the appropriate restriction enzymes and purified, ready for subsequent ligation with their corresponding inserts as outlined in Table 4. The concentration of the purified digested DNA was analysed by Nanodrop to inform insert:vector ligation ratios recommended by manufacturer (ThermoFisher) and was subsequently resolved on DNA agarose gels to confirm the success of the digest (Fig. 17).

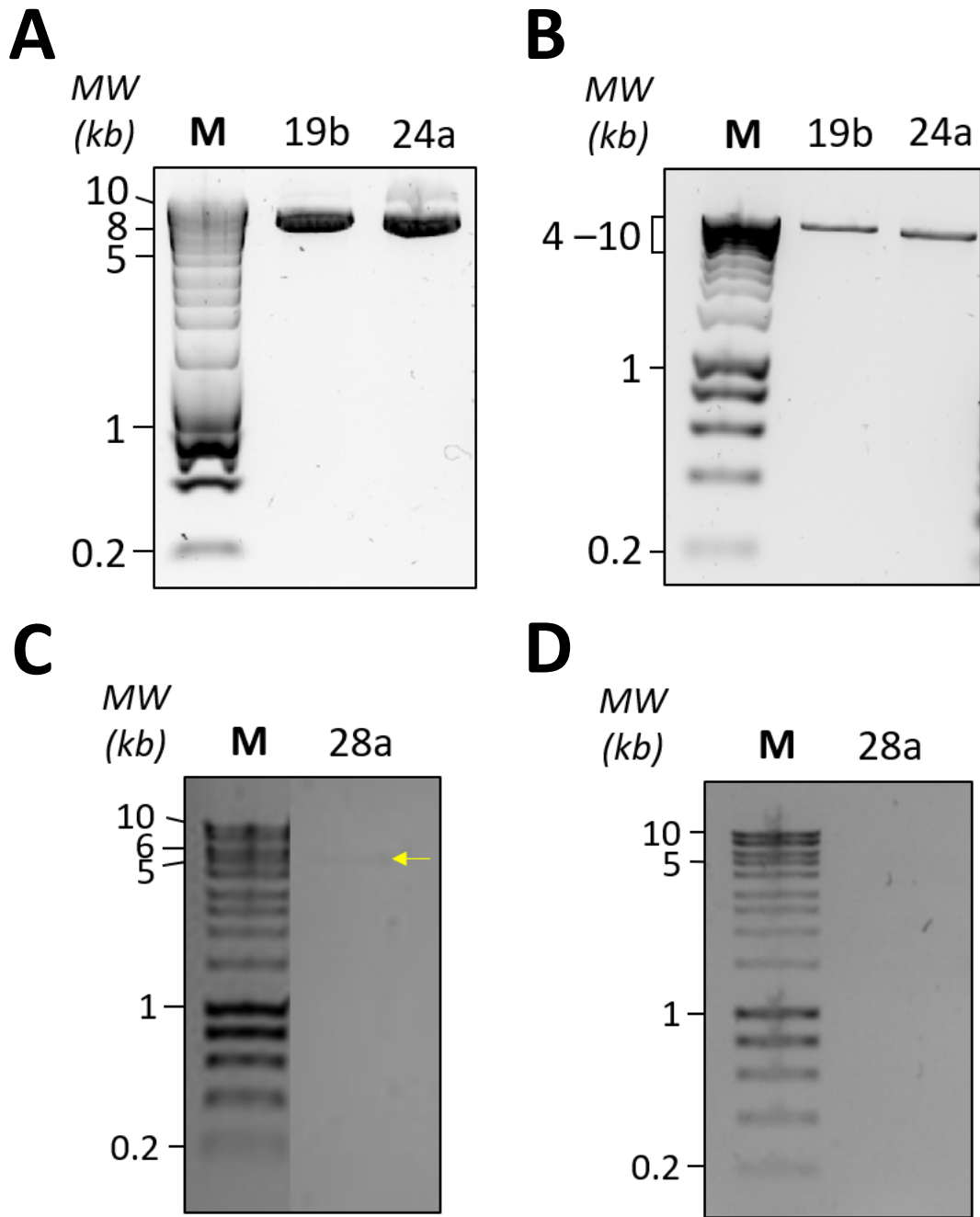


Figure 17. DNA gels (1%) of digestion and purification of pET vectors. Yellow arrows highlight faint bands. Lane M is represented by a 1kb hyperladder. MW., molecular weight.

- a. pET19b digested with NdeI and BamHI and pET24a digested by NdeI and NotI restriction enzymes.*
- b. DNA clean-up of digested pET19b and pET24a vectors.*
- c. Digested pET28a vector with NdeI and BamHI restriction enzymes.*
- d. DNA clean-up of the pET28a vector.*

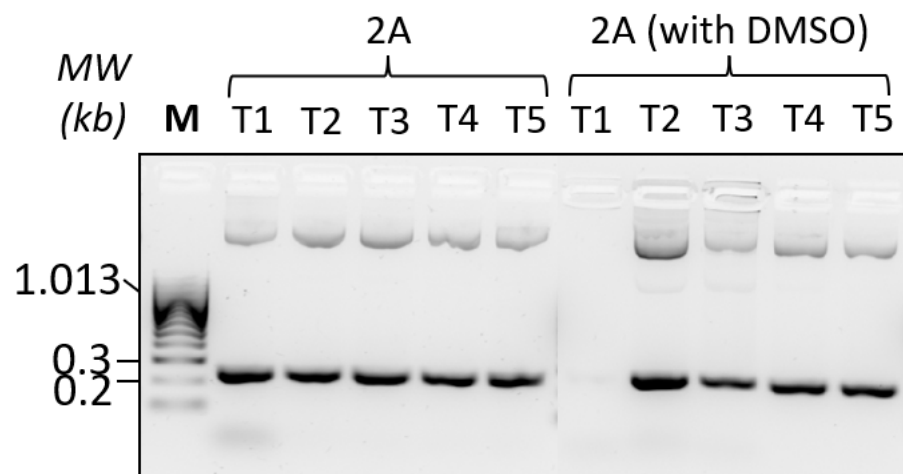
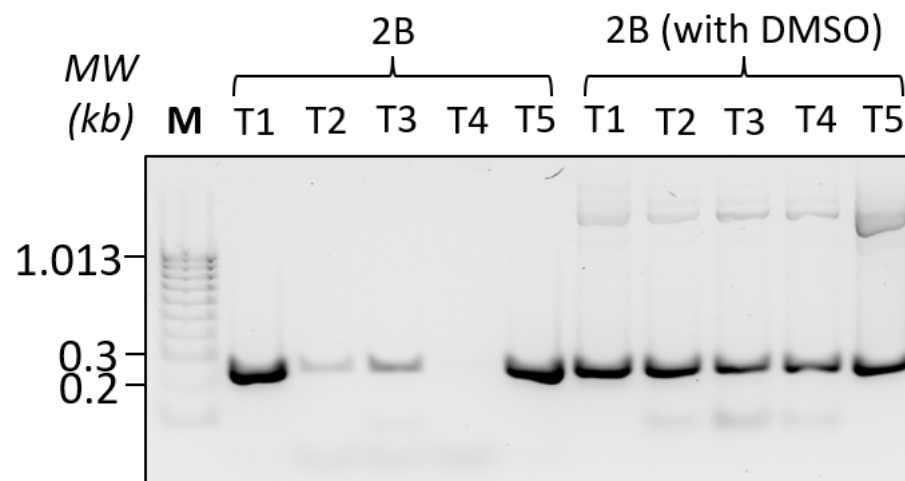
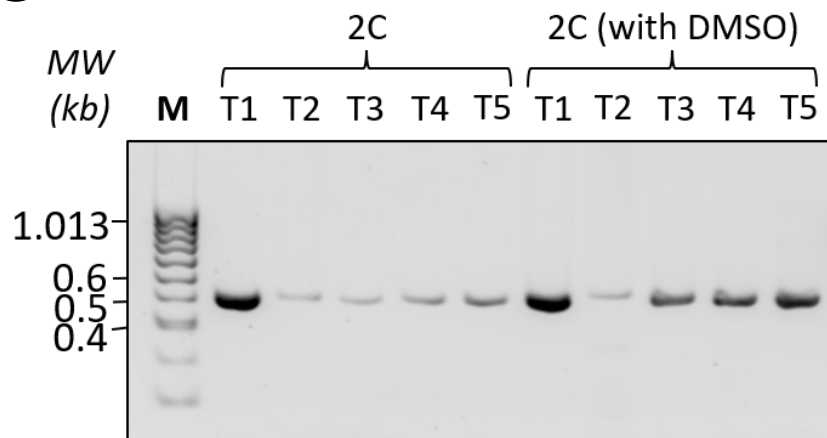
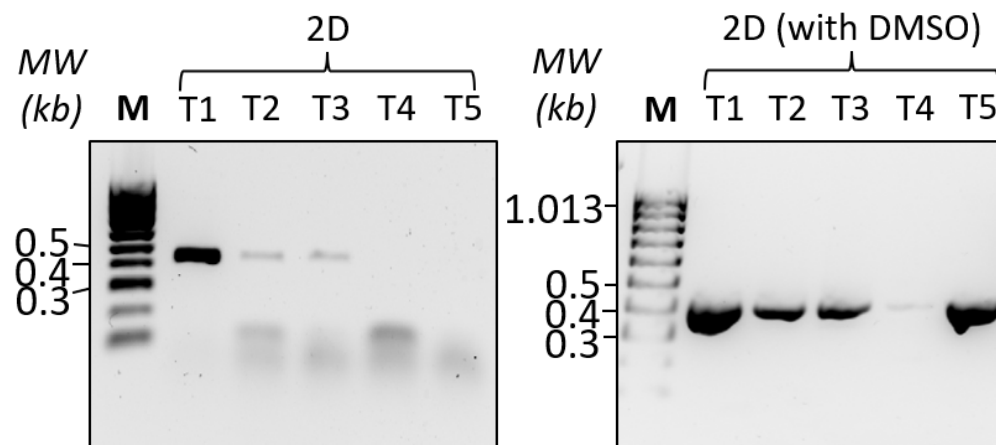
pET vectors 19b and 24a appeared to be running in the appropriate region, though they were running slightly higher compared to their respective sizes of 5717bp and 5310bp (Fig. 17a, 17b). It was seen that they were both running around 6,000bp, with pET24a running smaller than pET19b (Fig. 17a, 17b). With both vector lengths being relatively close to 6,000bp and their clear linearised bands following purification, the results confirmed the successful digestion and purification of the pET19b and pET24a vectors. The pET28a vector was confirmed to be running at the correct length of 5,369bp (Fig. 17c, 17d), while its low yield was further investigated via Nanodrop (Table 13).

Table 13. Purified pET vector concentration analysed by Nanodrop.

pET vector	Concentration (ng/μL)
pET19b	15.2
pET24a	14.4
pET28a	2.6

3.1.3 Screening of optimal PCR annealing temperatures for domain insert primer pairs

Secondary amplification of amplicons above 1000bp was observed through PCR prior to the screening of optimal annealing temperatures. As a result, the reaction conditions were optimised to provide maximal specificity and yield. A range of annealing temperatures were assessed and DMSO was used as a PCR additive to reduce non-specific amplification and enhance PCR amplification (Varadharajan and Parani, 2021). The PCR products were then visualised on an agarose gel (Fig. 18).

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

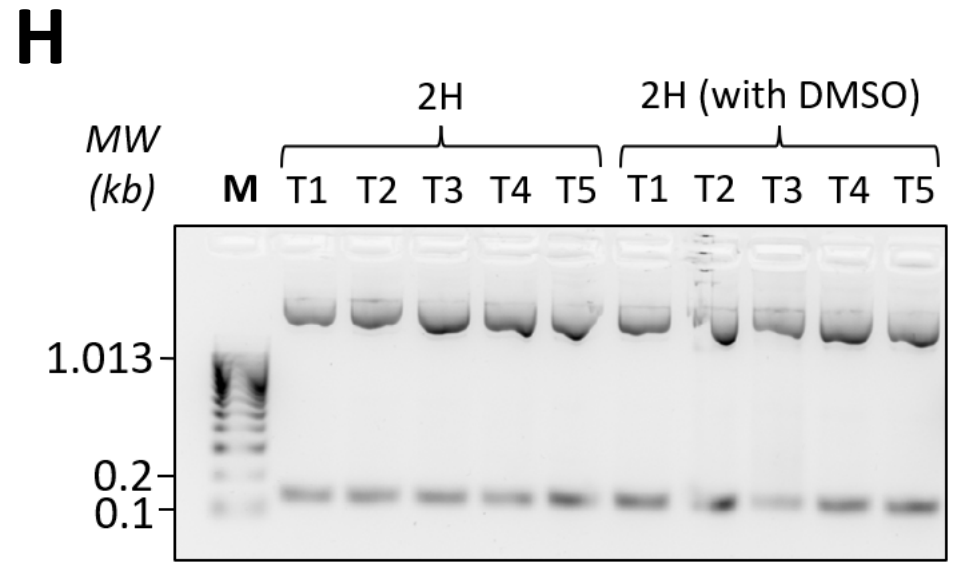
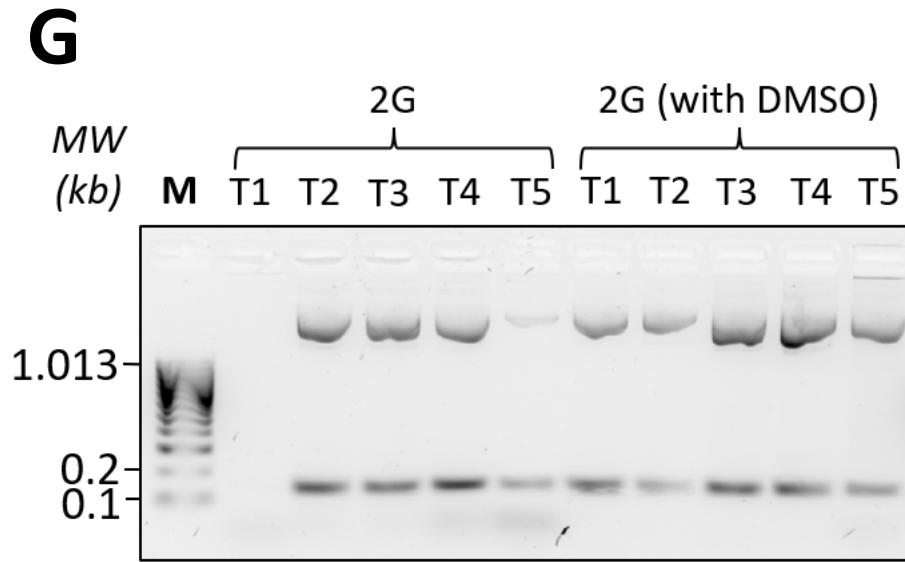
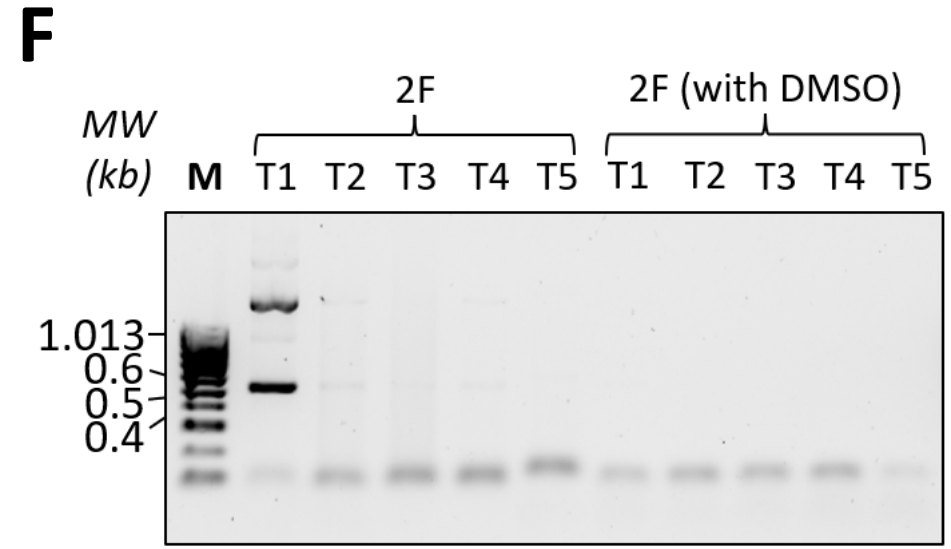
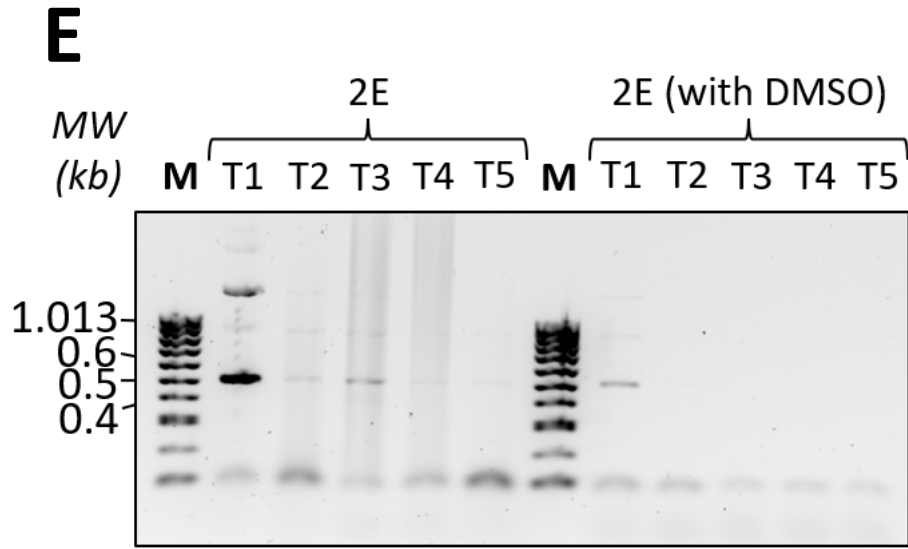


Figure 18. Screening of annealing temperatures and the effect of the DMSO additive. *The temperatures screened were: T1 - 60 °C, T2 - 62.1 °C, T3 - 64.8 °C, T4 - 67.8 °C, T5 – 70 °C. Construct IDs are as described in Table 4. Lane M is represented by a 100bp hyperladder on each gel (1%). MW., molecular weight.*

- a. Construct 2A (EMI pET19b) annealing temperature screen with and without the effect of DMSO.*
- b. Construct 2B (EMI pET24a) annealing temperature screen with and without the effect of DMSO.*
- c. Construct 2C (gC1q pET19b) annealing temperature screen with and without the effect of DMSO.*
- d. Construct 2D (gC1q pET24a) annealing temperature screen with and without the effect of DMSO.*
- e. Construct 2E (Coiled coil pET19b) annealing temperature screen with and without the effect of DMSO.*
- f. Construct 2F (Coiled coil pET24a) annealing temperature screen with and without the effect of DMSO.*
- g. Construct 2G (EGF-like pET19b) annealing temperature screen with and without the effect of DMSO.*
- h. Construct 2H (EGF-like pET24a) annealing temperature screen with and without the effect of DMSO.*

The inserts appeared to be running at the expected sizes, however, all clones other than constructs 2C and 2D (gC1q) appeared to contain non-specific amplification of DNA larger than 1000bp (Fig. 18c-d). Since most of these results are similar across each sample, there may be non-specific binding to the MMRN1 plasmid template (Fig. 18). The target DNA sequence of constructs 2G and 2H (EGF-like, 111bp) is marginally shorter than the other insert sequence lengths, which could increase the likelihood of non-specific binding. Higher annealing temperatures are also seen to decrease the specificity of primer binding, accounting for any non-specific amplification observed across higher temperatures of each construct (Fig. 18) (Wu et al., 1991). It remains unclear as to why constructs 2E and 2F (coiled coil) presented non-specific bands of varied lengths (Fig. 18e, 18f).

Expected sizes for construct 2A, 2B, 2G and 2H had almost consistent amplification across both higher and lower temperatures, demonstrating their stringency (Fig. 18a-b, 18g-h). Almost all constructs, apart from constructs 2G and 2H, seemed to have higher yield during PCR amplification at a lower annealing temperature of around 60 °C (Fig. 18). Constructs 2G and 2H appeared to have better amplification across higher temperatures between 62.1 °C and 70 °C (Fig. 18g, 18h).

Overall, the addition of DMSO did not improve the reaction and likely had a negative effect on primer binding compared to the non-DMSO variable. This may be compared to constructs 2G and 2H, where it seemed DMSO had no overall effect on primer binding as amplification appeared consistent across each well (Fig. 18g-h). The opposite may be seen for constructs

2C and 2D, where DMSO had increased general amplification of the insert compared to without DMSO (Fig. 18c, 18d). The addition of DMSO with constructs 2A and 2B (EMI) appeared as though there was an increase in non-specific binding (Fig. 18a, 18b).

Based on the above data (Fig. 18), the PCR reaction conditions were chosen based on the yield of desired DNA fragments and the presence of non-specific DNA amplicons (Table 14).

Table 14. Optimal PCR reaction conditions per construct. *DMSO containing variable indicated with a tick. The construct IDs are as described in Table 4.*

Construct	Annealing temperature (°C)	DMSO
2G insert (EGF-like pET19b)	62.1	
2H insert (EGF-like pET24a)	67.8	✓
1A, 2A insert (EMI pET19b)	60	
2B insert (EMI pET24a)	60	
1C insert (gC1q pET28a)	65	
2C insert (gC1q pET19b)	60	
2D insert (gC1q pET24a)	60	✓
2E insert (Coiled coil pET19b)	60	✓
2F insert (Coiled coil pET24a)	60	✓

3.1.4 Subcloning of genes coding for MMRN1 regions into expression vectors

Prior to bacterial expression, the EMI, EGF-like, gC1q and coiled coil insert constructs were amplified, digested, and purified, allowing for ligation with the appropriate expression vectors – pET19b for constructs 1A, 2A, 2C, 2E, and 2G; pET24a for constructs 2B, 2D, 2F, and 2H; pET28a for construct 1C (Fig. 19).

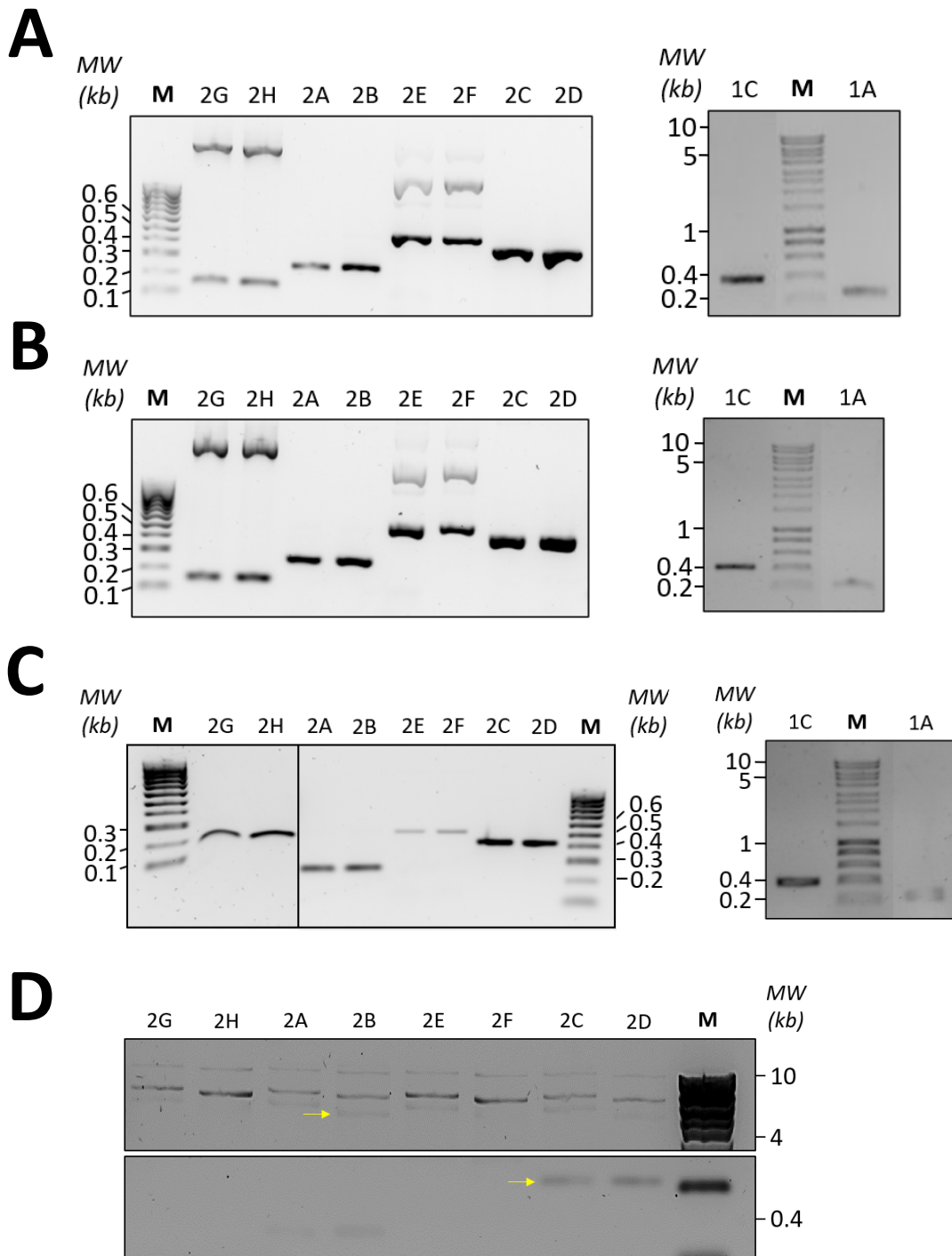


Figure 19. Gel electrophoresis of domain insert subcloning. Construct IDs are as described in Table 4. Lane M is represented by a 100bp hyperladder. MW., molecular weight.

- PCR products of the EGF-like, EMI, Coiled coil and gC1q domain inserts.
- Restriction digest of the EGF-like, EMI, Coiled coil and gC1q domain inserts with the appropriate restriction enzymes (Table 4).
- DNA purification of the EGF-like, EMI, Coiled coil and gC1q domain inserts.
- Ligation of the EGF-like, EMI, Coiled coil and gC1q domains to their respective vectors zoomed and enhanced with 40% contrast and 25% sharpness. Examples of potential unligated product are outlined using yellow arrows (original ligation gel in Appendix 1).

Fragment sizes for each domain construct were seen to be running at the expected size with constructs 2G and 2H (EGF-like) at 111bp, constructs 2A and 2B (EMI) at 228bp, constructs 2E and 2F (coiled coil) at 477bp and constructs 2C and 2D (gC1q) at 399bp (Fig. 19a). Following amplification, large non-specific fragments were still present above 1,000bp in constructs 2E-H which were removed following SPRIselect bead (Beckman Coulter) DNA purification (Section 2.2.8) (Fig. 19c, 19e). During subcloning, consistent amplicon lengths and concentrations were observed, with the exception of constructs 2E and 2F, which had reduced concentrations (Fig 19). This decrease may be due to potential washout during bead purification. This step-intensive process may contribute to the loss of sample concentration compared to constructs purified using a DNA clean-up kit.

Concentrations of each construct were measured before subsequent ligation and transfection of the plasmids (Table 15). Based on the DNA gels, it is unclear whether the ligated plasmids were running at the correct sizes of around 5000-6000bp due to the resolution of the hyperladder (Fig. 19a-c). An enhanced figure (Fig. 19d) of the ligation shows that not all digested PCR and plasmid products had been fully ligated as some unligated plasmid was present at the top of the gel, and insert bands were present at the bottom with an unknown band present above 10,000bp.

Table 15. Concentrations of domain inserts following purification via Nanodrop.

Construct	Concentration (ng/μL)
2G insert (EGF-like pET19b)	31.4
2H insert (EGF-like pET24a)	42.0
1A insert (EMI pET19b)	12.0
2A insert (EMI pET19b)	23.4
2B insert (EMI pET24a)	28.1
2C insert (gC1q pET19b)	47.3
2D insert (gC1q pET24a)	70.0
1C insert (gC1q pET28a)	12.0
2E insert (Coiled coil pET19b)	12.8
2F insert (Coiled coil pET24a)	18.5

3.1.5 Transformation into DH5 α cloning cell line

Corresponding vectors and inserts were ligated together and transformed into a chemically competent DH5 α cloning cell line (Fig. 20). This host strain offers high cloning efficiency and plasmid propagation as a reliable system for producing multiple successful transformants (Kostylev et al., 2015).

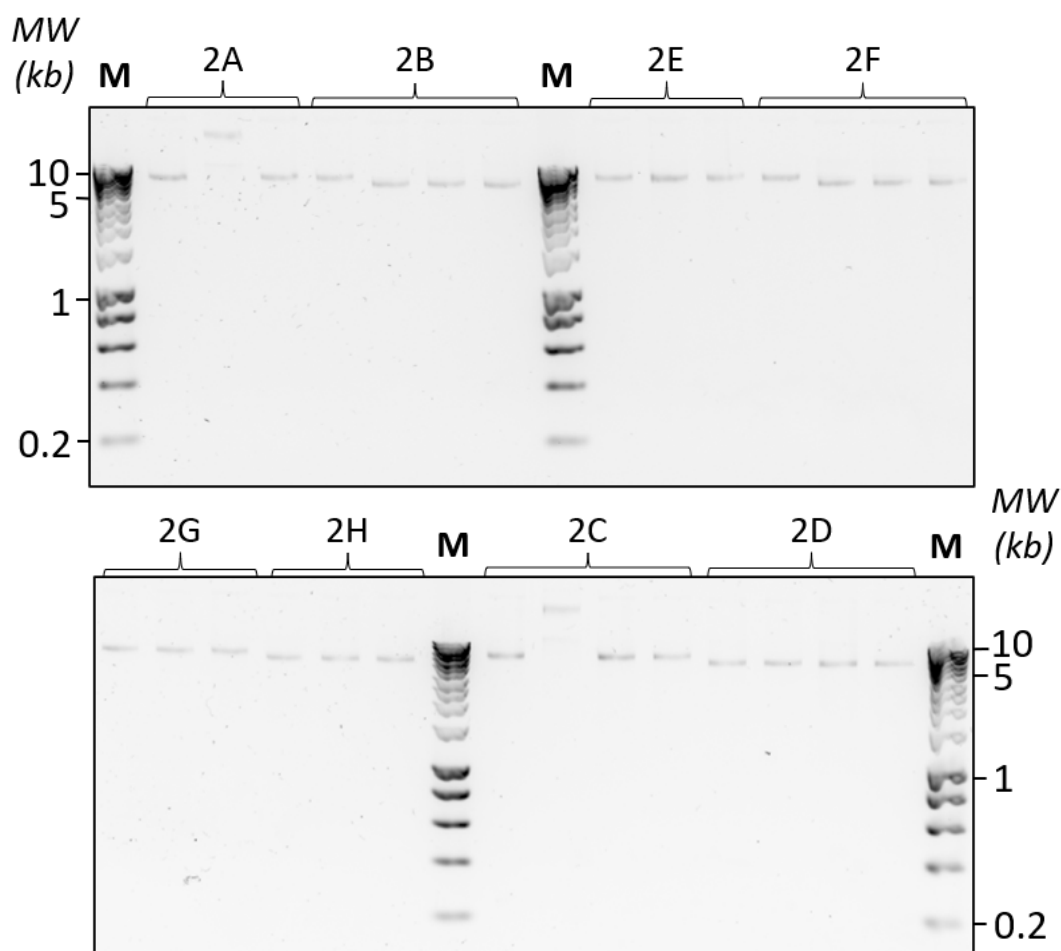


Figure 20. DNA gel (1%) of minipreps of 3-4 DH5 α colonies per plasmid. Following transformation, purified DH5 α colonies of the EGF-like, gC1q, EMI, and coiled coil domain constructs. Construct IDs are as described in Table 4. Lane M is represented by a 1kb hyperladder. MW., molecular weight.

The majority of constructs appeared to have been successfully transformed as the purified DH5 α plasmids were running at similar lengths (Fig. 20). Those that appeared to be running higher than expected, such as colony 2 for both constructs 2C (gC1q) and 2A (EMI) were not chosen to be transformed into chemically competent BL21(DE3) cells.

3.1.6 Screening of genes coding for MMRN1 regions following the transformation of ligated products

Prior to sequencing analysis, both a double-digest and a PCR colony screen were conducted to identify colonies with plasmids containing the insert DNA coding for the EMI, EGF-like, gC1q and coiled coil domains (Fig. 21). Four colonies were chosen for each construct and were grown up in LB broth to confirm the presence and length of the insert sequences to be sent for sequencing and to prepare for further transformation into an expression cell line to express the desired domain proteins.

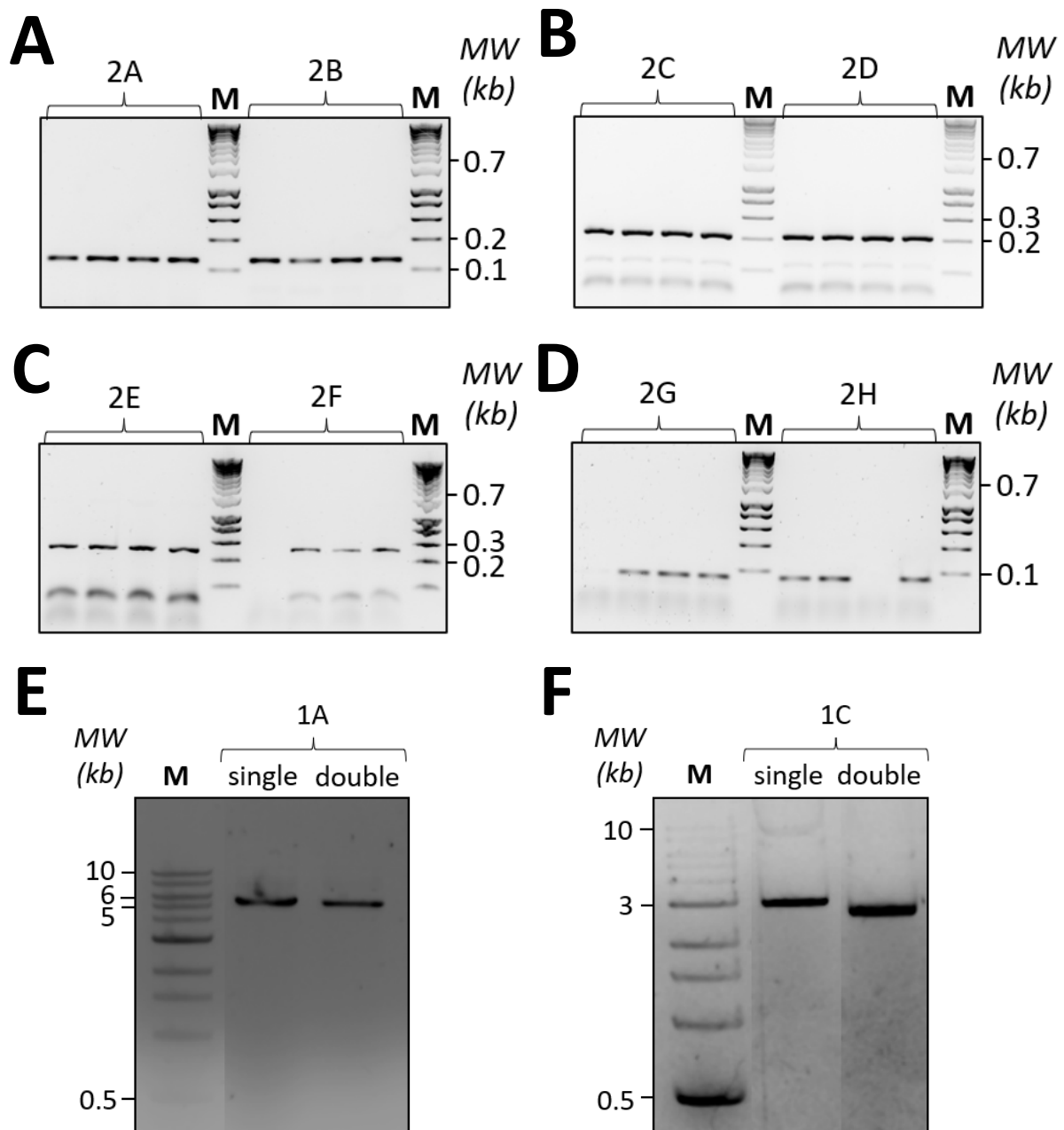


Figure 21. DNA gels (1%) of double-digest and PCR colony screening of genes coding for the domain constructs. Construct IDs are as described in Table 4. Lane M is represented by a 100bp hyperladder for figures 21a-d and a 1kb Quick-Load hyperladder (NEB) for figures 21e-f. MW., molecular weight.

- a. PCR colony screen of constructs 2A and 2B (EMI pET19b and pET24a).
- b. PCR colony screen of constructs 2C and 2D (gC1q pET19b and pET24a).
- c. PCR colony screen of constructs 2E and 2F (Coiled coil pET19b and pET24a).
- d. PCR colony screen of constructs 2G and 2H (EGF-like pET19b and pET24a).
- e. Double digest of construct 1A (EMI pET19b).
- f. Double digest of construct 1C (gC1q pET28a).

The PCR colony screen and the double digest demonstrate that the plasmids contained the desired gene inserts coding for the MMRN1 domains for each construct (Fig. 21). The

constructs analysed by PCR colony screen had been successfully amplified, however, they were all approximately 100bp lower than their expected lengths (Fig. 21a-d). Colony 1 in construct 2F and 2G and colony 3 in construct 2H were not further transformed into an expression cell line as they did not contain a successfully amplified insert on the gels (Fig. 21c, 21d).

Plasmids cut with two restriction enzymes had a noticeable difference in length when compared to plasmids cut with one restriction enzyme, as indicated through the double digest DNA gels (Fig. 21e, 21f). This demonstrates that the constructs containing the MMRN1 domain inserts with corresponding restriction sites had been successfully cut out, ready to be transformed into an expression cell line.

3.1.7 Sequencing construct 1C (gC1q pET28a)

Following screening, all recombinant clones had been submitted to Eurofins Genomics for sequencing along with their respective primers to ensure single nucleotide polymorphisms (SNPs) were not present in MMRN1 insert DNA. The results indicated that, except for construct 1C, none of the sequenced plasmids contained the intended domain inserts, despite positive screening results (Fig. 21). Subsequently, the DNA sequences coding for the gC1q domain and construct 1C were translated into protein sequences using ExPasy for comparative alignment (Fig. 22). Focussing on protein comparison allows for direct comparison of variability between amino acids that dictate protein structure and function. The MultAlin sequencing alignment programme was used to assess the accuracy of the recombinant protein sequence (Corpet, 1988).

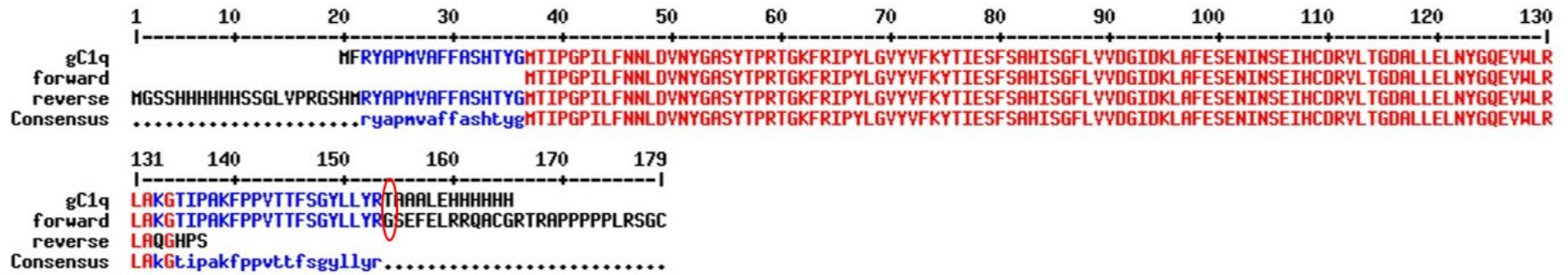


Figure 22. Amino acid sequence of the gC1q pET28a construct. The row labelled gC1q is the expected amino acid sequence, the row labelled forward is the forward DNA sequence with the reverse sequence being labelled as reverse. The red text highlights the high consensus part of the sequence while the blue indicates the low consensus part of the sequence, black is labelled as a neutral consensus. Any incorrect amino acids are outlined using a red oval.

The protein sequence of construct 1C (gC1q), which had been cloned, aligned correctly with its expected sequence (Fig. 22). However, the full length of the sequence should have ended in a threonine, whereas the cloned construct ended in a glycine. These amino acids do not share similar DNA codes and, therefore, was an unexpected base pair change. As this substitution occurred on the C terminus of the gC1q protein instead of the His-tag sequence, the structure and function of the domain protein may have been impacted. For further analysis, the amino acid change was analysed using ColabFold-AlphaFold2 predicted folding (Fig. 23) (Mirdita et al., 2022). The yellow and blue protein backbones align well, showing no significant change in the overall protein conformation. However, the changed amino acid residue is different on the C terminal of the gC1q domain, so it is not expected to align. The degree of its impact would need further experimental and functional validation to understand the implication of the amino acid change on the protein's characteristics. This clone was still taken forward for purification following sequencing.

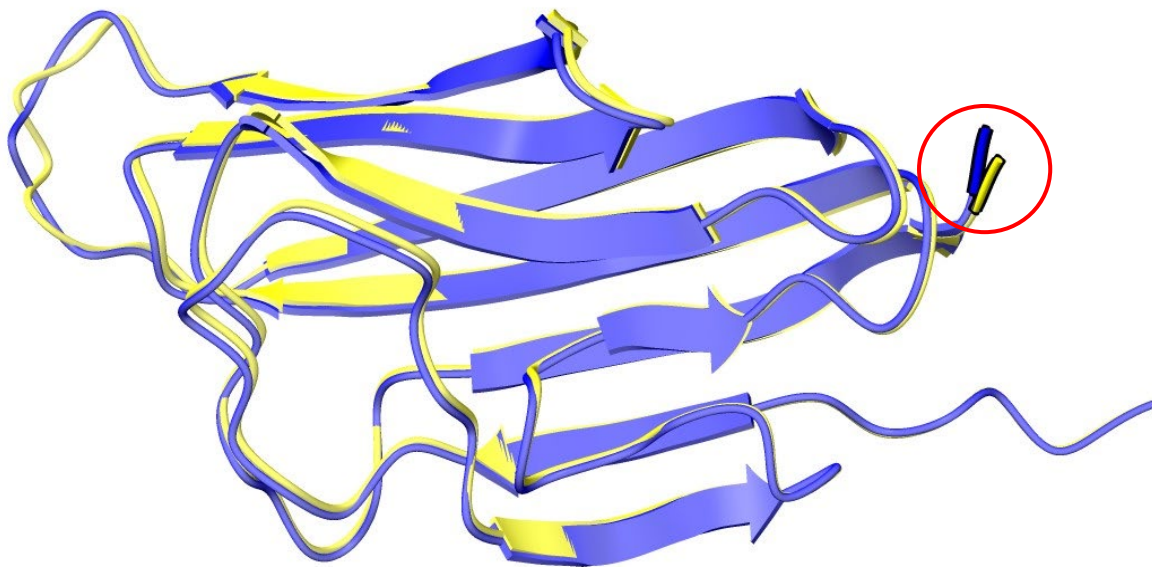


Figure 23. ColabFold-AlphaFold2 3D image of the impact on the last amino acid change for construct 1C on protein folding versus the expected protein sequence of the gC1q (Mirdita et al., 2022). *The sequenced domain (construct 1C) is in yellow, and the actual domain sequence is in blue. The red circle indicates the last residue change. Molecular graphics and analyses performed with UCSF ChimeraX, developed by the Resource for Biocomputing, Visualisation, and Informatics at the University of California, San Francisco, with support from National Institutes of Health R01-GM129325 and the Office of Cyber Infrastructure and Computational Biology, National Institute of Allergy and Infectious Diseases (Meng et al., 2023; Pettersen et al., 2021; Goddard et al., 2018).*

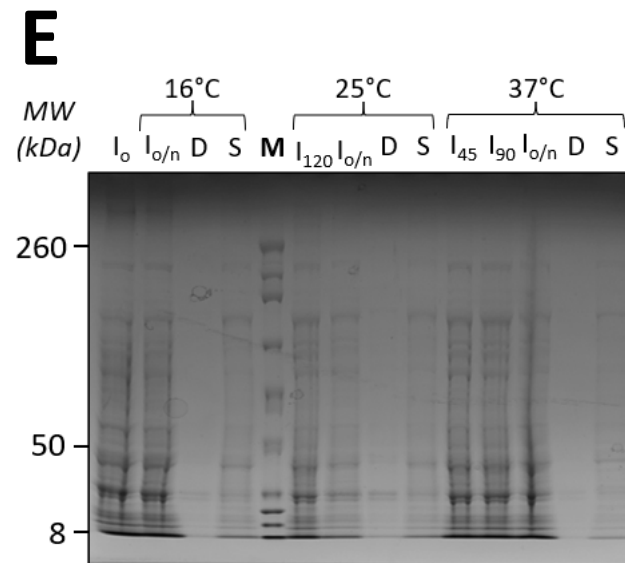
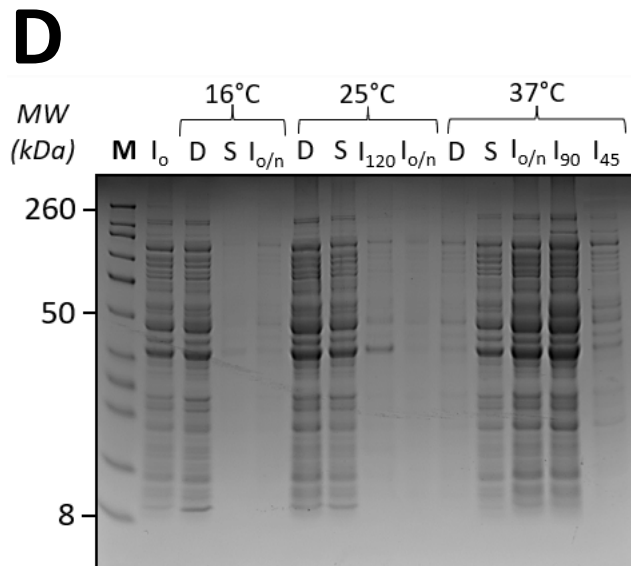
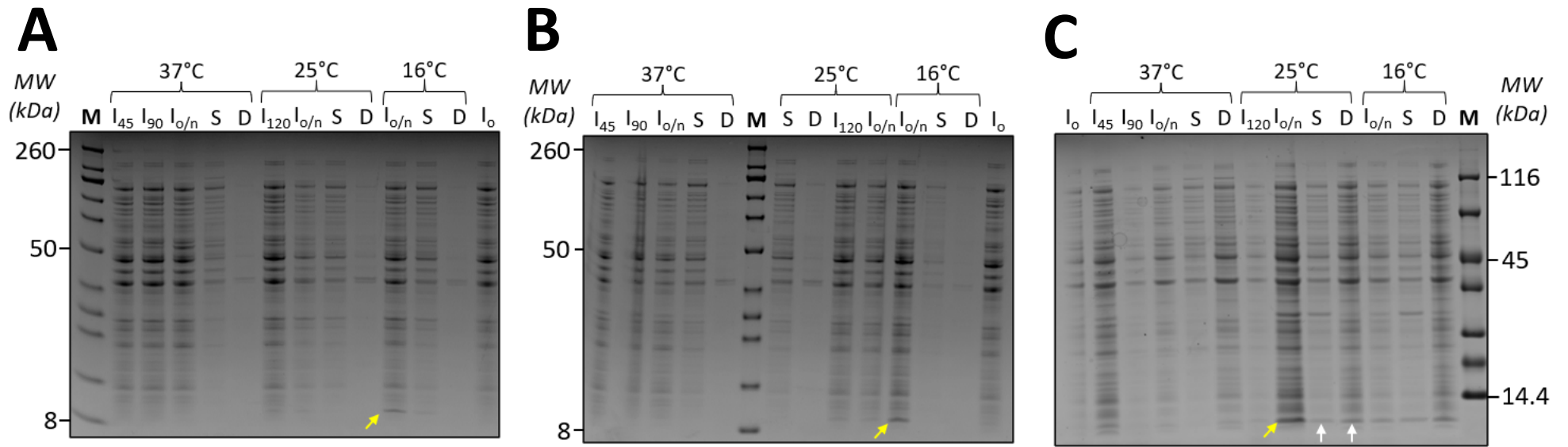
3.1.8 TOPO® TA cloning®

For TA cloning, eight DNA constructs, 2A-2H, were amplified using *Taq* polymerase to create adenine (A) overhangs and ligated into TOPO plasmid vectors containing complementary thymine (T) residues. This method was conducted due to its efficiency in cloning PCR-amplified fragments for optimisation following unsuccessful sequencing results of 7 out of 8 constructs to troubleshoot any annealing or restriction digest issues such as faulty restriction enzymes, restriction site recognition or restriction cutting. However, sequencing results acquired from TA cloning also returned unsuccessful, presenting as an unexpected challenge. Despite the simplicity of this cloning method and the use of blue-white screening, a few potential issues may arise, including the occurrence of false positives, which may indicate the presence of an insert (though not necessarily the desired domain inserts), low transformation efficiency, or the presence of contaminants (Yao et al., 2016).

3.2 Recombinant expression of MMRN1 constructs in *E. coli* culture

3.2.1 Small-scale expression trials

To confirm recombinant protein expression, SDS-PAGE was used to analyse the levels of expression of MMRN1 expression constructs coding for the gC1q (in pET19b, pET24a, pET28a), EMI (in pET19b, pET24a), EGF-like (in pET19b, pET24a) and coiled coil (in pET19b, pET24a) regions in 10 mL of *E. coli* culture (Fig. 24). The expression conditions for each construct were screened for optimum induction duration and temperature (Section 2.2.14).



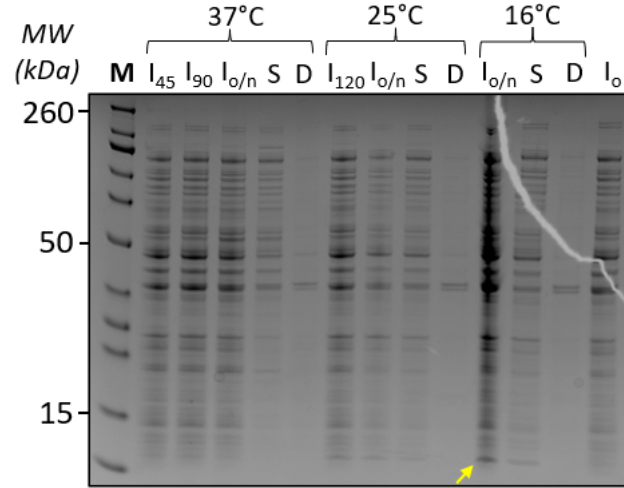
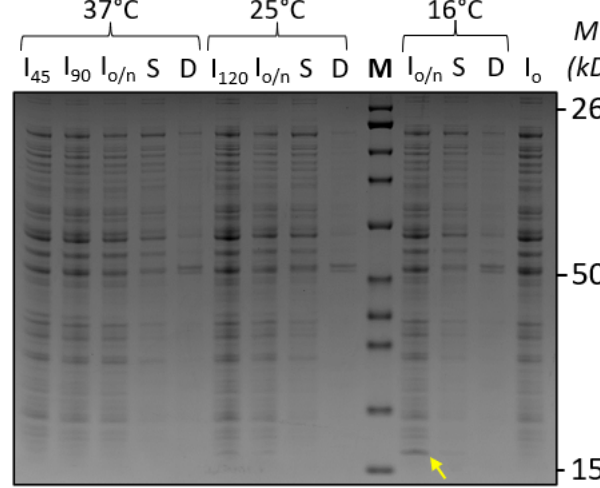
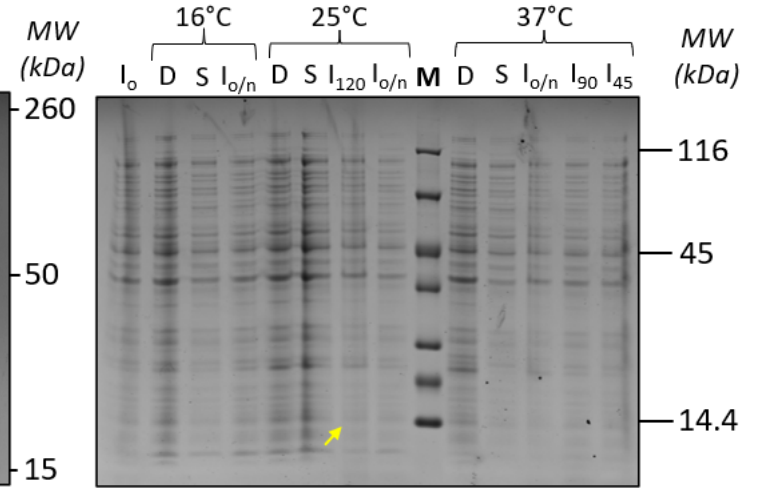
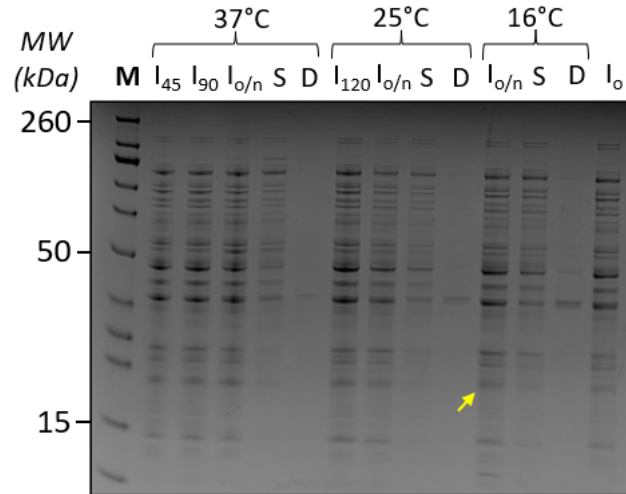
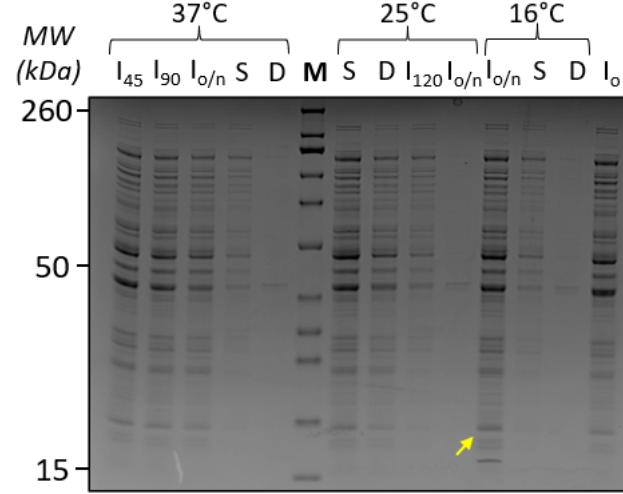
F**G****H****I****J**

Figure 24. Expression of 8 clones within BL21(DE3) with varied induction times and temperatures. 4-12% SDS-PAGE gels were used to compare different expression conditions of each protein. Coomassie staining was used following gel electrophoresis. The positive control (I_0) was a sample taken prior to induction. I_{45} , I_{90} and I_{120} indicate that the samples were taken after a certain number of minutes, while $I_{0/n}$ indicates that the sample was taken after overnight incubation. Soluble (S) and insoluble (D) samples were taken once all whole-cell samples had been collected. The expected sizes of each construct are indicated with a yellow arrow. White arrows highlight any important soluble and debris bands. Construct IDs are as described in Table 4. Lane M is represented by the Chameleon protein ladder. MW., molecular weight.

- a. SDS-PAGE gel of expression conditions for construct 2A (EMI pET19b) with an expected size of 8.3kDa.
- b. SDS-PAGE gel of expression conditions for construct 2B (EMI pET24a) with an expected size of 8.3kDa.
- c. SDS-PAGE gel of expression conditions for construct 1A (EMI pET19b) with an expected size of 8.3kDa.
- d. SDS-PAGE gel of expression conditions for construct 2G (EGF-like pET19b) with an expected size of 4.2kDa.
- e. SDS-PAGE gel of expression conditions for construct 2H (EGF-like pET24a) with an expected size of 4.2kDa.
- f. SDS-PAGE gel of expression conditions for construct 2C (gC1q pET19b) with an expected size of 14.6kDa.
- g. SDS-PAGE gel of expression conditions for construct 2D (gC1q pET24a) with an expected size of 14.6kDa.
- h. SDS-PAGE gel of expression conditions for construct 1C (gC1q pET28a) with an expected size of 14.6kDa.
- i. SDS-PAGE gel of expression conditions for construct 2E (Coiled coil pET19b) with an expected size of 17.6kDa.
- j. SDS-PAGE gel of expression conditions for construct 2F (Coiled coil pET24a) with an expected size of 17.6kDa.

Across all constructs, poor expression at 37 °C had been observed (Fig. 24), whereas 16 °C appeared to be the temperature that promoted the most expression across most of the constructs (Table 16). A lower expression temperature may aid slower and correct protein folding (Mason et al., 2014). Overnight incubation at lower temperatures appeared to increase protein expression compared to short incubation times and long incubation times at higher temperatures, highlighting the influence of expression duration and temperature on the rate of protein production. Figures 24a, 24e and 24f showed little to almost no expression across all expression conditions. Constructs 2G and 2H (EGF-like) were unable to be conclusively analysed using SDS-PAGE. To avoid the possibility of the protein running off the gel, SDS-PAGE was stopped early, but the protein remained undetectable. It is likely that both constructs may not have been expressed, or they were not detectable under the chosen

conditions. Each gel showed no expression in the uninduced fraction compared to those that had been expressed (Fig. 24).

It was seen that there was more expression present in the soluble fraction of each construct compared to the debris (Fig 24c-h). This was beneficial when taking constructs forward for purification. Although, the debris column seemed dilute across all figures, which may have been due to sample preparation. In figure 24c, there appeared to be strong bands across both soluble and debris fractions at the expected length of construct 1A (EMI) (indicated by white arrows in Fig. 24c).

Table 16. Chosen expression conditions for each construct. *I_{o/n}* implies overnight induction.

Construct	Expression condition
1C (gC1q pET28a)	25 °C I _{o/n}
1A (EMI pET19b)	25 °C I _{o/n}
2A (EMI pET19b)	16 °C I _{o/n}
2B (EMI pET24a)	16 °C I _{o/n}
2G (EGF-like pET19b)	N/A
2H (EGF-like pET24a)	N/A
2E (Coiled coil pET19b)	16 °C I _{o/n}
2F (Coiled coil pET24a)	16 °C I _{o/n}
2C (gC1q pET19b)	16 °C I _{o/n}
2D (gC1q pET24a)	16 °C I _{o/n}

3.2.2 Scaling up the expression of construct 1A (EMI) and 1C (gC1q).

The selected constructs showing promising expression were scaled up from 10 mL of media to 100 mL to assess whether there was enough expression of protein to be further purified and analysed via SDS-PAGE (Fig. 25). Due to a lack of glycosylation sites, constructs 1A and 1C trialled for large-scale expression in *E. coli*.

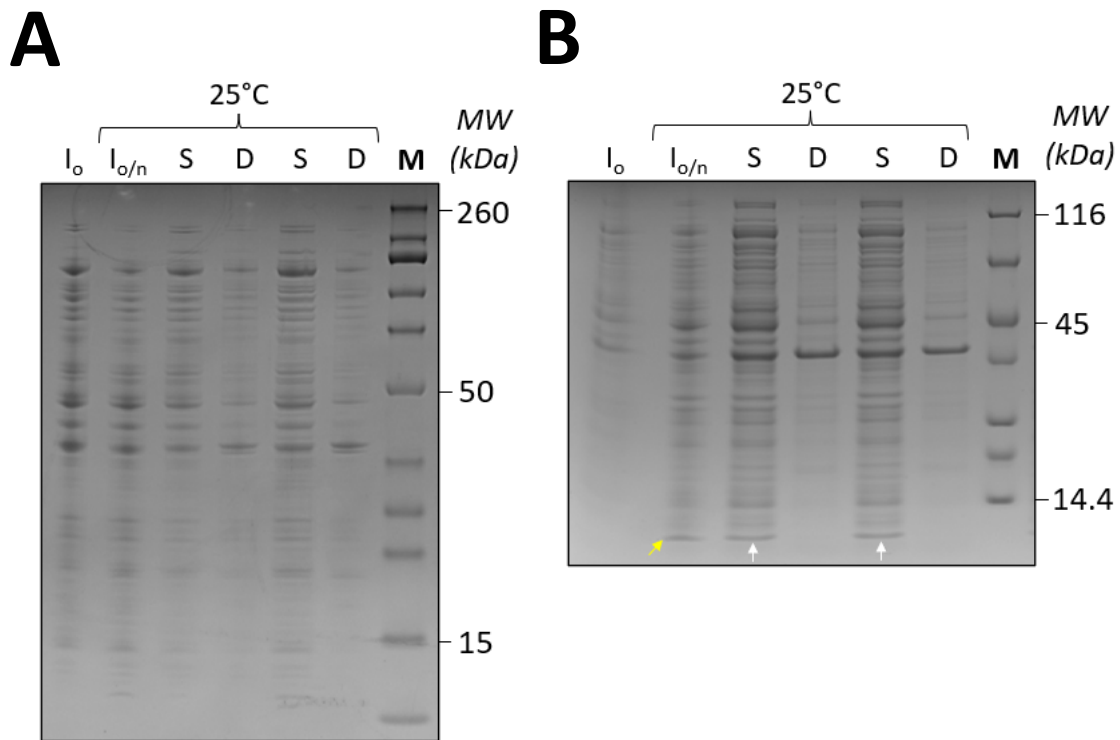


Figure 25. Scaled up expression of constructs 1A and 1C in 100 mL culture. The expected sizes of each construct are indicated with a yellow arrow. White arrows highlight any important soluble and debris bands. Construct IDs are as described in Table 4. Lane M is represented by the Chameleon protein ladder. MW., molecular weight.

- a. SDS-PAGE gel of construct 1C (gC1q) when induced at 25 °C for overnight incubation in 100 mL of media.
- b. SDS-PAGE gel of construct 1A (EMI) when induced at 25 °C for overnight incubation in 100 mL of media.

Low expression levels were observed for construct 1C (gC1q) at an expected size of 14.6kDa, whereas construct 1A (EMI) appeared to have darker bands at what appeared to be its expected length (8.3kDa) in the soluble fractions with no expression in the uninduced fraction (Fig. 25a-b). The expression result of both constructs was followed up with sequencing, and it was found that the EMI DNA was of low quality, whereas the gC1q protein DNA matched its expected DNA sequence (Fig. 22).

3.2.3 Western blot of construct 1A (EMI) and 1C (gC1q)

As sequencing results suggested that the vector contained the correct insert, Western blot analysis was conducted using an anti-His antibody to assess if the recombinant proteins were expressed but below the detection level of Coomassie staining in SDS-PAGE. The western blot

confirmed the presence of the desired recombinant proteins in the experimental samples ready for purification (Fig. 26).

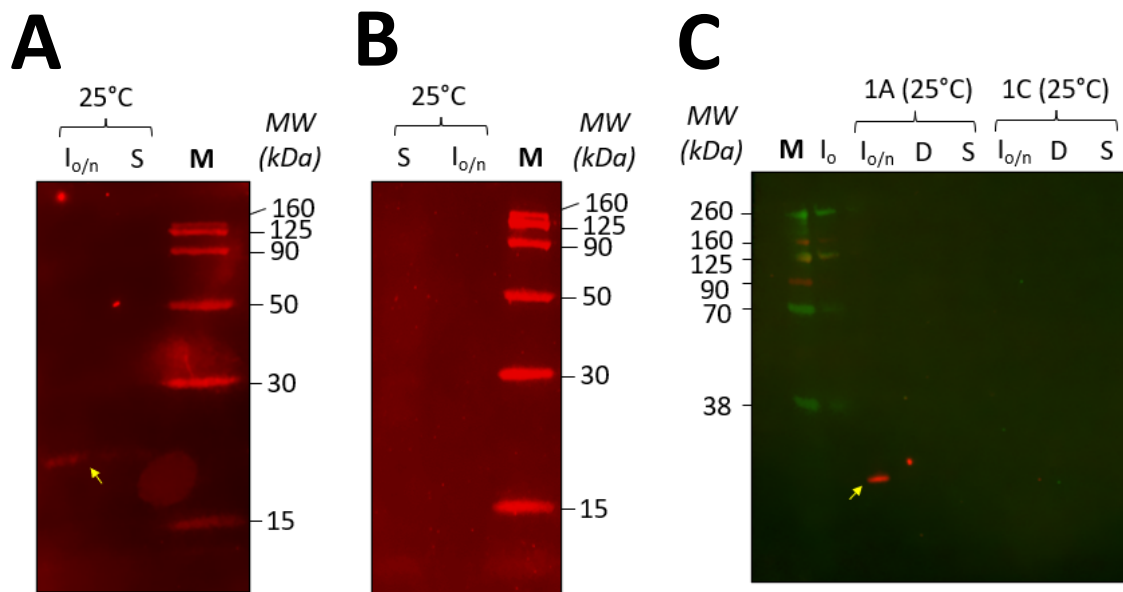


Figure 26. Fluorescence western blot of constructs 1C (a, c) and 1A (b, c) incubated with anti-His antibodies. The expected sizes of each construct are indicated with a yellow arrow. Construct IDs are as described in Table 4. Lane M is represented by the Chameleon protein ladder. MW., molecular weight.

A clear fluorescent band was observed for construct 1C (gC1q), estimated to be higher than its expected size of 14.6kDa, confirming the protein expression of the His-Tag (Fig. 26a, 26c). The inclusion of the His-tag sequence may have contributed to a higher observed molecular weight of 17.6kDa due to the introduction of additional amino acids. The protein seemed to only present a fluorescent signal in the whole cell rather than the soluble fraction, which may be linked to potential protein degradation during sonication. To minimise this effect, protease inhibitors were added to the sample prior to sonication to inhibit proteolytic activity in the sample ready for purification. No fluorescence was detected for construct 1A, consistent with the sequencing results received (Fig. 26b, 26c).

3.2.4 Mini His-bead pull-down purification

To corroborate the sequencing results, an additional mini His-bead pull-down was conducted alongside western blot analysis to confirm and assess recombinant protein expression via the His-tag. The His-beads were pipetted directly onto the SDS-PAGE gel for its analysis (Fig. 27).

A sample was taken from the initial wash of the beads to determine if the protein was being washed out or bound to the His-beads.

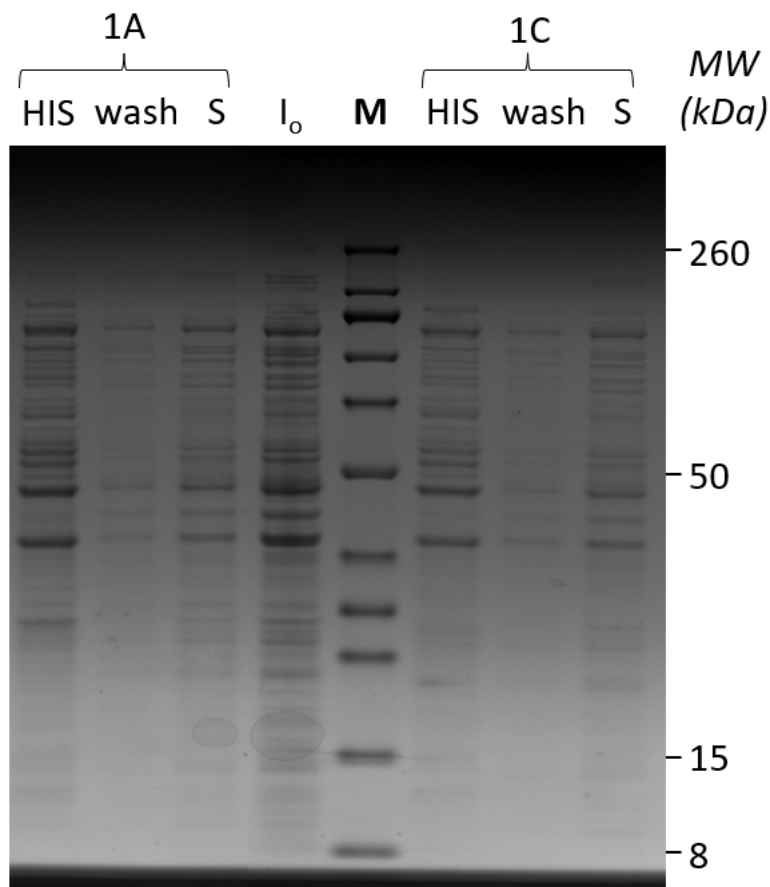


Figure 27. A His-bead pull-down purification of constructs 1A and 1C incubated with HisPur™ Ni-NTA Resin (ThermoFisher). Lanes ending with 'HIS' represent samples incubated with the His-bead resin directly applied to the gel, while lanes ending with 'wash' were samples taken from the initial wash. S represents the soluble fractions. Construct IDs are as described in Table 4. Lane M is represented by the Chameleon protein ladder. MW., molecular weight.

The His-bead pull-down purification indicated an absence of recombinant protein in both the wash and the His-bead samples for both construct 1A and 1C (EMI, gC1q) (Fig. 27). This does not align with the sequencing results and western blot analysis for construct 1C but does align with the results shown for construct 1A. This may be due to multiple factors. Firstly, it is possible that the interaction between the His-tag and the His-bead resin was weaker than expected, which resulted in inefficient binding during the pull-down experiment. Secondly, the His-tag may not be accessible. As observed in literature, bead pull-downs are conducted with folded protein whereas western blot uses denatured protein where the samples are mixed with Laemmli buffer that contains denaturing agents, such as SDS, prior to membrane blotting and therefore, the antibody is able to access the tag (Luo et al., 2014; Brady et al.,

2004; Xu et al., 2019). The SDS-PAGE protein bands were visibly faint, which could suggest a possible issue with sample preparation, such as protein loss during purification steps or low concentrations of the protein samples. It may also be possible that the culture was not large enough for the protein to be detected, and therefore, for IMAC purification, the culture was scaled up to 2 L. The results of this pull-down were later clarified through purification via affinity chromatography.

3.2.5 Affinity chromatography using the Akta for the purification of construct 1C (gC1q)

In response to issues with low protein yield, Construct 1C was scaled up from 100 mL and grown in 2 L of media to increase the protein amount needed for detection for both purification and SDS-PAGE. The recombinant protein was purified using the Äkta purifier and eluted off the HisTrap™ FF Crude column using a Peristaltic Pump in 1 mL aliquots across a gradient of 100 mM to 500 mM Imidazole. The protein aliquots were run on an SDS-PAGE gel ready for analysis (Fig. 28). The chromatogram following Äkta elution is shown in Appendix 3.

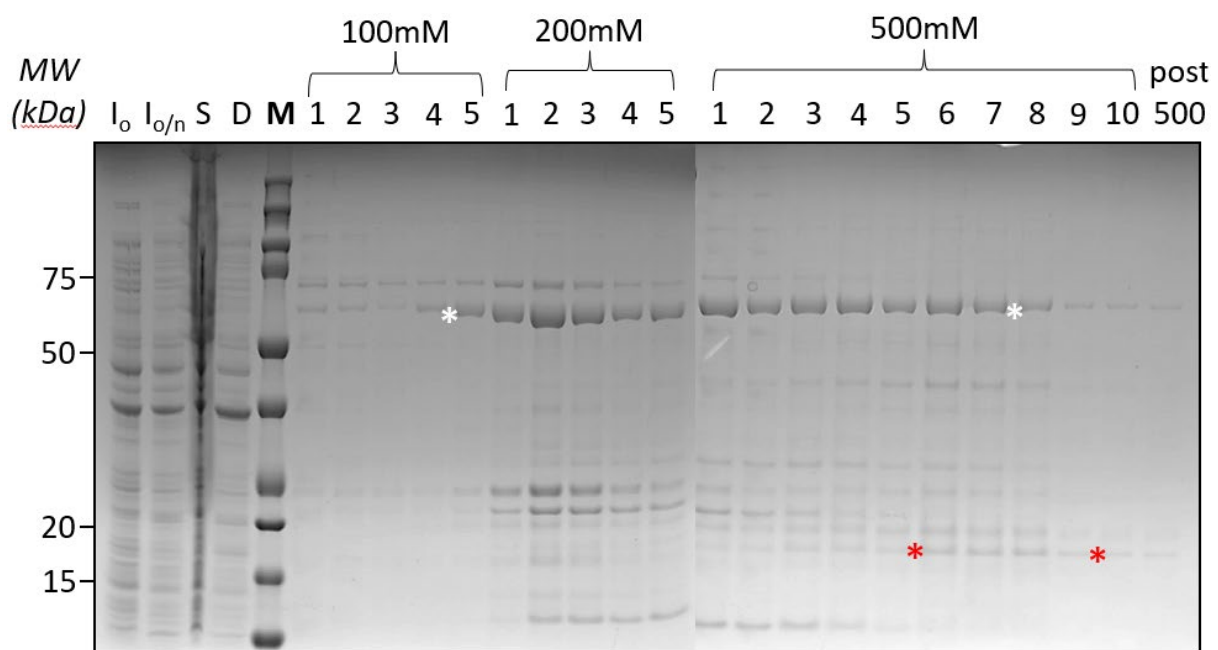


Figure 28. The purification of Construct 1C using affinity chromatography eluted with a Peristaltic Pump across a gradient of 100 mM to 500 mM Imidazole. The number above each lane corresponds to the millilitre aliquot loaded for each Imidazole concentration. ‘Post 500’ indicates the collection of 15 mL of 500 mM following the elution of the initial ten 500 mM aliquots. Gels were edited next to each other, and markers were aligned, unaltered gels are shown in Appendix 2. The expression of the desired recombinant protein is highlighted between two red asterisks, while the 60kDa co-purified protein is highlighted between two white asterisks. *M* is represented by a BioRad Precision Plus Protein™ Standards ladder. *MW.*, molecular weight.

The SDS-PAGE analysis of the eluted fractions shows the expression of construct 1C within the 500 mM Imidazole fractions at a molecular weight of 17.6kDa (Fig. 28). However, the expression level appears to be relatively low when compared to the uninduced fraction. Prominent bands at 60kDa suggest that there is a contaminant protein that has been highly expressed and purified. The presence of this contaminant may be a contributing factor to the low expression of the target recombinant protein, as it may be outcompeting.

4 Discussion

The aim of this project is the recombinant expression and purification of MMRN1 constructs to study the structure-function relationship of MMRN1 with the *S. aureus* protein Efb. To this end, ten MMRN1 domain constructs, carrying either an N-terminal or C-terminal His-tag, were screened for optimum expression conditions in *E. coli* culture (Fig. 24). Sequencing confirmed one correctly cloned construct, the gC1q domain, which was taken forward for expression trials in *E. coli* BL21(DE3) and purification. The gC1q domain was expressed at low amounts alongside a 60kDa contaminant under the tested conditions (Fig. 28). Codon optimisation is an approach aimed to enhance translation efficiency and increase protein yield by adapting the nucleotide sequence of the mammalian gC1q domain gene in humans to better match codon preferences of the host organism (*E. coli*). Alternatively, BL21(DE3) Rosetta cell lines offer an easier strategy by providing additional tRNA molecules that recognise rare codons, eliminating the need for DNA sequence alterations and, therefore, simplifying the experimental process (Burgess-Brown et al., 2008; Heyde and Norholm, 2021). It may be that the absence of glycosylation may have affected the expression, folding and stability of the glycoprotein, which may further degrade and affect the protein's solubility and biological activity. However, trials were initially conducted to express both the EMI and gC1q domains (Constructs 1A and 1C) in *E. coli*, and glycosylation was not expected to be an issue for the gC1q domain construct. Notably, the gC1q domain is also not anticipated to have primary glycosylation sites (Fig. 29) and, therefore, with its observed low expression in *E. coli*, using a Rosetta cell line may offer a better strategy for enhancing protein production for subsequent analysis.

To take this protein forward for structural analysis by X-ray crystallography, the protein should be purified to homogeneity and concentrated between 2-50 mg/mL to obtain crystals that diffract to high resolution during crystallographic analysis (Dessau and Modis, 2011). To do this, additional purification steps such as Ion-exchange chromatography (separation by charge) or size exclusion chromatography (SEC, separation by size) should be tested to separate the 17.6kDa gC1q construct and the 60kDa protein. The interaction between the purified gC1q domain and Efb may be examined using a pull-down assay, such as co-immunoprecipitation (Co-IP). Any purified protein complexes should be trialled for protein crystallography. It is also possible to set up co-crystallisation trials of gC1q and Efb, but in

order to obtain larger, well-diffracting crystals, both proteins need to be expressed and purified to high purity and stability, allowing both to form a complex and set up initial crystallisation trials to identify parameters pH, precipitant, temperature and other additives for the proteins to form crystals, ready for further optimisation and X-ray diffraction analysis (Bolatti and Gourlay, 2022; Hipolito et al., 2014; Schilling et al., 2014).

4.1 Subcloning of the MMRN1 domains

4.1.1 Non-specific amplification in PCR agarose gels

PCR amplification is contingent on meticulous primer design and optimisation of PCR conditions. Hence, in this project, primer length was kept between 15-30bp long to ensure optimal binding to the target DNA sequence (Wu et al., 1991) (Section 2.2.1). Cycling parameters such as the annealing temperatures of each primer pair using gradient PCR and the effect of DMSO as a component in the PCR reactions were fine-tuned to overcome specificity challenges encountered in the experimental process (Section 2.2.4).

Non-specific amplification was observed in agarose gels above 10,000bp during PCR amplification of DNA insert domain regions (Fig. 18, 19). Non-specific DNA amplification was observed for coiled coil and EGF-like domains (constructs 2E, 2F, 2G, 2H), and the potential presence of primer binding sites in unintended genomic regions were examined using the NCBI primer BLAST tool (Ye et al., 2012). During primer analysis, no off-target hits were found in the *E. coli*/Homo Sapiens databases showing that the primer pairs were specific to the MMRN1 plasmid template. Primer pairs had self-complementary scores between 4% and 8%, indicating a low proportion of primer bases that may form secondary structures or primer dimers (Bustin et al., 2020).

Along with no off-target hits and primer design considerations, it seems the appropriate explanation links to template contamination, leading to unexpected amplification. It is unlikely that there was genomic DNA contamination from *E. coli*, such as mitochondrial DNA, as BLAST analysis of the primers showed no complementarity. Therefore, the source may be plasmid, environmental or primer contamination, introduced through shared equipment, residual carryover of DNA from previous reactions on equipment or reagents, and potential

surface or human contamination (Wally et al., 2019; Witt et al., 2009). Repeated PCR attempts could not be carried out within the given project timeframe.

Minimising contamination is crucial to ensure the reliability of the biological experiments (Bastien et al., 2003). In future, template preparation would be conducted in a laminar flow hood, along with consistent refreshment of sterile reagents, taking aliquots of stock reagents and keeping working reagents separate from stocks, incorporating additional steps for template purity checks to minimise the risk of contamination affecting PCR outcomes (Aeschbach and Dion, 2017). Control experiments would be conducted at every step to monitor contamination during DNA extraction/PCR processes. The primary negative control where no amplification should be present is the no template control that uses no DNA template during PCR, replaced with nuclease-free water, to monitor reagent and environmental contamination of foreign DNA (Ruiz-Villalba et al., 2017; Kalle et al., 2014). These steps would significantly reduce the risk of contamination, ensuring the accuracy and reliability of downstream molecular experiments. To purify desired PCR products from non-specific PCR products in the study, PCR products of the desired size were purified using size-exclusion bead purification to mitigate potential downstream experimental effects (Fig. 19). A high-fidelity polymerase was used to reduce non-specific amplification (Brelsford et al., 2012). Specific binding may be crucial for optimising overall protein yield, making it imperative to troubleshoot.

In future, the larger bands present on the gel could also be purified using gel extraction (excision of DNA band containing the PCR product) prior to sequencing to investigate what region is being amplified above 10,000bp. This emphasises the importance of refining experimental approaches to ensure specific amplification in future experiments.

4.1.2 Challenges encountered in sequencing outcomes

To confirm the desired domain inserts had the correct DNA sequence, positive clones in DH5 α *E. coli* that passed the initial screening tests, using PCR colony screen and double digest, were sent for sequencing (Fig. 21). The gC1q domain construct was the only successful clone (Fig. 22). The discrepancy observed between the positive results in the PCR colony screen, and the sequencing results could be due to various factors that have been observed across literature

for PCR colony screen such as false positives, PCR artifacts, and false negatives (Yang and Rothman, 2004).

PCR colony screens, especially for small target sequences like the 111bp EGF-like sequence, are susceptible to false positives due to non-specific binding or primer artifacts (Dallas-Yang et al., 1998). To address this, additional controls, such as no-template PCR negative controls, should be used to detect potential contamination in future. Primers used for colony screening may also produce false-positive results when annealing temperatures are low (Fredriksson-Ahomaa and Korkeala, 2003). When optimising annealing temperatures for enhanced primer binding, 60 °C was observed to be best suited to each primer pair, excluding EGF-like inserts (Table 14). This suggests that even though lower temperatures were optimal for enhancing primer binding, this may contribute to false-positive results in colony screening. Future optimisation could involve designing primers with higher annealing temperatures to increase stringency, reducing the likelihood of non-specific amplification. Quantitative real-time PCR (qPCR) colony screening may be a more reliable alternative method of observing the presence of desired DNA sequences when compared to PCR colony screening because it may differentiate between false-positive and true-positive transformants by comparing Ct values obtained from qPCR amplification curves (Skarratt and Fuller, 2014). However, when time and equipment costs are restricted, qPCR may not be as practical compared to all other optimisation techniques as listed above.

Although a double digest may be less sensitive than a PCR colony screen, it may be considered a more reliable screening as there is a lower risk of false positive results (Evans et al., 2018; Anand et al., 2004). As the only construct to be successfully sequenced in this study was one that was analysed by double digest, it reaffirms the reliability of this method (Fig. 21f). Both techniques may be used in tandem to increase the validity of the result while implementing additional controls as well as sequencing to enhance the accuracy of insert verification in cloning experiments.

The possibility of vector religation leading to background colonies was considered as sequencing results from unsuccessful clones only returned the sequence of the expression vector (Bessa et al., 2012). Dephosphorylating the vector after restriction digestion could help prevent self-ligation (Ukai et al., 2002). Including a self-ligated control and running PCR on self-ligated colonies may reveal non-specific amplification originating from the vector or

reagents. Paired with a no template control, this would determine if the amplification was due to reagent contamination or non-specific amplification originating from the vector. However, this problem was addressed with TA cloning, where the clones still faced unsuccessful sequencing results (results not shown). Since TA cloning uses directional cloning, the complementary adenine and thymine overhangs ensure that only the PCR product may ligate in one orientation, preventing the linearised vector from religating without the presence of the insert (Yao et al., 2016).

Other controls, such as scraping part of the plate with no visible growth, may be implemented to verify the legitimacy of positive colonies. This helps differentiate true clones from potential contamination on the agar plate. Though it is unlikely, diluted ligation may contaminate the plate when bacteria are plated onto the agar plate, making colonies that are chosen to appear positive by PCR but are contaminated by what is present on the plate (Dallas-Yang et al., 1998). Since the plates contained antibiotics for selection, this was not expected. However, agar plates become susceptible to environmental contamination over time, and antibiotic selection decreases the longer the plates are stored (Jenkins and Schuetz, 2012). Freshly prepared agar plates ensure optimal selection pressure (Navarro et al., 2010).

4.1.3 Cloning the EGF-like domain

Non-specific primer binding was observed when amplifying the EGF-like domain, and the addition of DMSO to enhance PCR and primer specificity did not resolve this issue (Fig. 18g, 18h). This may be explained by the domain's short DNA template of 111bp. Short DNA templates may lead to increased instances of mispriming as they have fewer unique sequences and nucleotides to bind to, making it easier for primers to bind non-specifically to regions with partial complementarity. The cloning of this domain may have benefitted from including neighbouring domains or regions, such as the gC1q domain, to create a more stable construct (Gray et al., 2022). This modification may further prevent the EGF-like domain protein from running off the SDS-PAGE gel (Fig. 24d, 24e) to be able to provide insight into the protein's expression conditions.

4.2 Expression conditions and expression systems

4.2.1 Optimising IPTG concentration as an expression condition

Optimisation of expression conditions may involve various parameters, including temperature, induction time and IPTG concentration for plasmid-based expression constructs under a T7 promoter to help maximise the expression of the gC1q domain (Gomes et al., 2020). Optimisation of IPTG concentration is essential to balance induction and potential toxicity. Too much IPTG may lead to metabolic stress and reduced cell viability, while too little may result in insufficient expression of the target protein, therefore, identification of the optimal induction conditions ensures robust protein expression without compromising cell health (Gomes et al., 2020). This approach helps achieve higher yields of soluble protein, reducing the risk of inclusion body formation and improving the overall efficiency of protein production in *E. coli* expression systems (Xu et al., 2023).

4.2.2 Post-translational modifications of the MMRN1 glycoprotein

MMRN1 domains were expressed in *E. coli* culture to generate recombinant protein for structural and functional analysis. Despite its inability to glycosylate proteins, *E. coli* as an expression system offers distinct advantages such as cost-effectiveness, ease of use and rapid growth. With its ability to produce a high yield of protein, this expression system was considered suitable for further X-ray crystallography analyses. Additionally, with two of the domain proteins, the EMI and gC1q domains, lacking in primary glycosylation sites, it seemed advantageous to use *E. coli* as an expression system (Fig. 29). However, while using *E. coli* may be efficient and cost-effective, expressing specific proteins may pose as a problem (Baeshen et al., 2015).

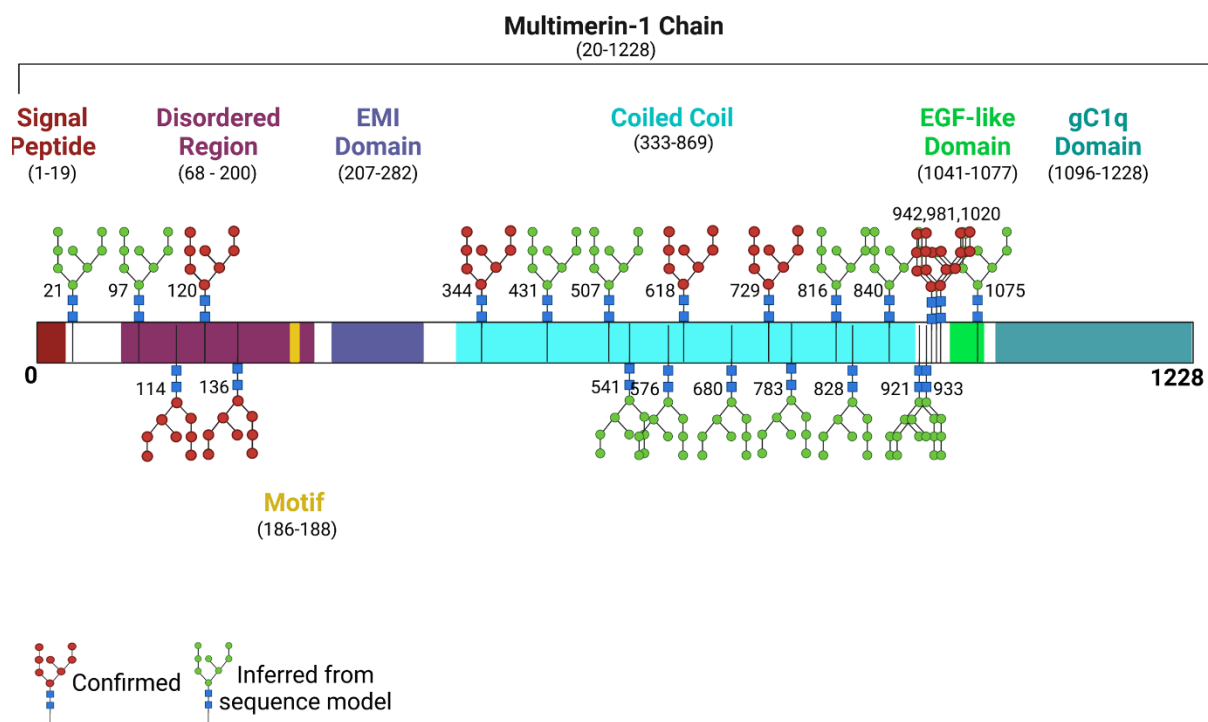


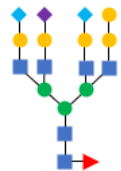
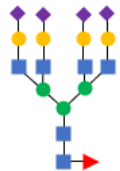
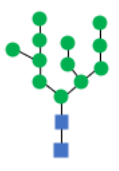



Figure 29. Diagram of MMRN1 protein sequence, and domains with confirmed and inferred primary glycan locations (Uniprot, 2023). The MMRN1 is 1228 amino acids long, comprised of a disordered region, a motif, an EMI domain, a coiled coil, a unique EGF-like domain and the gC1q domain (Colombatti et al., 2011). Its signal peptide is cleaved off after the protein has been newly synthesised. It contains twenty-three glycans along the chain, eight of which are confirmed in red and fifteen have been inferred from a sequence model in green (figure prepared by author, created with Biorender.com).

Recombinant expression of the gC1q domain was low (Fig. 28). To increase gC1q yield, protease inhibitors were used to reduce its potential degradation prior to purification as well as the optimisation of induction and growth parameters (Fig. 24). Various domains were expressed to overcome the structural complexity of the full-length protein, of these domains, EMI and gC1q are predicted to be non-glycosylated (Fig. 29). With the absence of glycosylation in *E. coli*, particularly when expressing the EGF-like and coiled coil domains, it may impact the protein's native conformation and potentially influence its functional characteristics and therefore any structural and functional findings may not be as easily applied with confidence as it does not truly reflect the native state of the protein. Therefore, other expression systems may need to be considered (Fisher et al., 2016) (Table 17).

A potential expression system for MMRN1 domains in the future is the Leishmania expression system (LexSy, Jena Bioscience). Known for its compatibility with eukaryotic post-translational modifications while combining the advantages of prokaryotic cell lines (cost-efficient and

robust), this system may retain the native conformation of the protein through glycosylation while producing a high enough yield for both structural and functional analysis (Breitling et al., 2002). Based on publications, the LexSy system, with 882 published records on PubMed, appears less widely used than other established systems, such as *E. coli* (35,166 PubMed publications), yeast (25,689 PubMed publications), or mammalian cells (464,583 PubMed publications), which may be due to it being relatively new, being available since 2002 (de Oliveira et al., 2019). Also, laboratories that have already established mammalian/bacterial/viral protein purification facilities may be more inclined to use existing systems. One downside to the LexSy expression system is that it has low expression levels compared to common bacterial plasmids, with researchers looking to optimise codons to increase yields (de Oliveira et al., 2019). However, LexSy has been successful for some difficult constructs failing to be expressed in other expression systems with multiple high-impact papers in Nature that have used this system for structural and functional analysis, evidencing that it is capable of effective protein production (Hemayatkar et al., 2010; Dadashipour et al., 2011; Murphy et al., 2020; Zabelskii et al., 2021; Swartz, 2009). For example, recombinant proprotein convertase 4 in mouse and insect cells exhibited poor expression but demonstrated high expression in the LexSy system and recombinant acetyl serotonin methyl transferase in *E. coli* produced insoluble, heavily degraded material when compared to the LexSy system able to express the soluble form (Basak et al., 2008; Ben-Abdallah et al., 2011).

Table 17. Summary of various expression systems. Comparing features of the *E. coli* expression system with Mammalian, LexSy, Yeast, Baculovirus, Rabbit Reticulocyte and HeLa systems when expressing recombinant proteins. The primary cell lines for Mammalian expression are Chinese Hamster Ovary (CHO) cells and Human Embryonic Kidney (HEK) cells. The monosaccharide symbols are depicted according to SNFG-nomenclature (Varki et al., 2015; Neelamegham et al., 2019). The N-glycan building blocks are mannose (green circle), N-acetylglucosamine (blue square), fucose (red triangle), galactose (yellow circle), N-acetylneuraminic acid (purple diamond), and N-glycosylneuraminic acid (light blue diamond) (Frigliou et al., 2021; Khan et al., 2017).

Expression systems	Prokaryotic	Eukaryotic				Cell-free	
Cell lines	<i>E. coli</i>	Mammalian (e.g., HEK293 or CHO)	LexSy	Yeast	Baculovirus	Rabbit Reticulocyte	HeLa
Yield	High	Low	High	High	High	Low	Medium-High
Glycosylation	No	Yes	Yes	Yes	Yes	Limited	Yes
Advantages	High yield Robust system	Represents full biological activity	Fast growth High yield	Rapid reproduction Promotes glycosylation	High yield Promotes glycosylation	Flexible system	High yield Promotes glycosylation
Disadvantages	No glycosylation	Sensitive method – no cell wall Low yield	Lower expression to bacterial counterparts	Expensive High maintenance	Expensive High maintenance	Limited glycosylation Low yield	Expensive Low quantity for structural analysis
N-glycans							
References	Brondyk (2009)	Brondyk (2009) Schmidt (2004)	de Oliveira et al. (2019)	Sodoyer (2004) Schmidt (2004)	Brondyk (2009)	Carlson et al. (2005) Anastasina et al. (2014)	Gagoski et al. (2016)

Scientific literature reveals a recurring theme in selecting expression systems based on the trade-off between their advantages and disadvantages (Ducker et al., 2023; Sodoyer, 2004; Schmidt, 2004; Viktorinova and Wimmer, 2007). While many use the *E. coli* expression system due to its convenience and high yield, many studies employ this system when glycosylation is not a critical factor in their research (Briand et al., 2016; Lozano et al., 2021; Jia and Jeon, 2016). However, when glycosylation is pivotal, alternative expression systems, such as yeast or mammalian systems, are often used to mimic mammalian glycosylation patterns to accurately represent the protein's structure and functionality. Correct protein glycosylation is particularly imperative when designing or assessing proteins for therapeutic use due to the impact of glycan structures on immunogenicity (Kim et al., 2020).

With mammalian and *E. coli* cell lines being widespread and cost-effective methods of expression, both gC1q and EMI domains may be expressed using the BL21(DE3) Rosetta cell lines (Brondyk, 2009). Since neither of these proteins has primary glycosylation sites and *E. coli* offers rapid growth and easy manipulation, it presents an advantageous system for these two proteins (Brondyk, 2009). The Rosetta cell line addresses challenges related to gC1q domain expression by enhancing soluble protein production (Zarkar et al., 2020). However, for glycosylating protein domains such as coiled coil and EGF-like, it may be beneficial to explore alternative expression systems. As discussed previously, the LexSy system offers cost-effectiveness while combining the simplicity of prokaryotic systems and the protein synthesis capabilities of eukaryotic systems, including glycosylation (de Oliveira et al., 2019).

To tackle any potential glycosylation issues, *E. coli* may also be bioengineered to be able to conduct glycosylation. Studies have found that using the pathogen *Campylobacter jejuni* may transfer the glycosylation pathway *in vitro* into strains of *E. coli* that have already expressed the protein (Wayman et al., 2019). This may not mimic the native state of the expressed proteins but offers the chance to test the impact of glycosylated versus non-glycosylated samples. Bacterial bioengineering goes beyond the scope of this project by being able to produce glycosylated proteins but also allows for greater customisation when applying them to treatment and diagnostic strategies (Iwashkiw et al., 2012). The need to boost the yield of glycosylated variants for therapeutics is one of its challenges, though it is predicted that bioengineering *E. coli* will progress relatively quickly given the fact that it may be genetically manipulated at large volumes (Strutton et al., 2018; Harding and Feldman, 2019). Ease of use

and access to equipment may also be a limiting factor for researchers. The structural complexity and heterogeneity of glycans, coupled with their inability to be biosynthesised using recombinant DNA technology, makes this method difficult (Rudd and Dwek, 1997). One study reports a novel cell-free transcription-translation system that may produce glycoproteins in a single-pot reaction, however, in the context of this study, it may be easier to use alternative methods of expressing the MMRN1 glycoprotein (Jaroentomeechai et al., 2018).

4.3 Protein sample analysis

4.3.1 SDS-PAGE optimisation

The SDS-PAGE gel results obtained in this project provided valuable insights into the optimal expression conditions and solubility of the domain proteins of MMRN1. However, improvements and optimisation could enhance the reliability and interpretability of these results by contributing to a more robust interpretation of the experimental outcomes.

To ensure an accurate comparison of concentrations between samples, performing a Bicinchoninic Acid (BCA) assay before loading any sample on the gel would allow for the standardisation of the protein concentrations to minimise variations arising from differences in solubility versus insolubility (Cortes-Rios et al., 2020). To compare like for like, the pellet of the insoluble sample was resuspended in the equivalent volume of the soluble supernatant. However, in some SDS-PAGE results, the insoluble sample appeared to be significantly less concentrated (Fig. 24, 25, 27). This may have occurred due to the insoluble fraction not being fully resuspended or due to slight pipetting variations leading to differences in sample volumes (Kurien and Scofield, 2012). Therefore, an alternative standardisation technique may be beneficial.

4.3.2 Optimising protein processing methods

Interestingly, while assessing western blot outcomes, fluorescence was observed in the whole cell fraction, contrasting with the soluble and insoluble samples (Fig. 26). This may be because of protein damage due to harsh sonication methods on the whole cell culture when acquiring

the soluble and insoluble samples. Although, potential for uneven loading of whole cell compared to insoluble and soluble lysate may have caused the samples to be below the detection limit and should be considered. This exemplifies the previous point of using a BCA assay to ensure proper standardisation of samples.

The same sonication protocol was used throughout the project, yielding good and poor samples (Figs. 24-5, 27-8). Sonication is pivotal for cell breakdown and protein release through sonic waves (Pchelintsev et al., 2016). Given its rigorous nature, optimising sonication parameters, including duration, amplitude, and intervals, through SDS-PAGE analysis, as seen for temperature and incubation times (Section 2.2.14), becomes crucial for efficient cell lysis and enhanced recovery of soluble proteins (Pchelintsev et al., 2016). This may also address the low expression of the purified gC1q domain observed following affinity chromatography (Fig. 28). The forceful sonic waves, however, pose challenges—leading to protein degradation and heat-induced denaturation of both soluble and insoluble proteins (Stathopoulos et al., 2004). Notably, smaller proteins like the EGF-like and EMI domains are more susceptible to degradation due to their higher surface area-to-volume ratio (Liu et al., 2022). In contrast, larger proteins, often possessing a more stable tertiary structure, offer better protection against denaturation. Therefore, adjusting sonication parameters for each protein domain is imperative to minimise protein degradation effects.

Alternative methods of cell lysis, commonly used with bacterial cells, may also be considered, including chemical lysis and French Press. The French Press method disrupts cells by forcing them through a narrow space, using a piston to apply high-pressure, shearing membranes (Goldberg, 2008). French Press has selective advantages to sonication as it is less harsh towards the cells and has minimal heat generation, which may help with protein solubility, however, it is important to note that it requires specialist, expensive equipment that certain laboratories may not be equipped with and when applied for large-scale application, it is not as effective when compared to sonication. Chemical lysis is another method that is less harsh when shearing cell membranes for soluble protein. Detergents are used to break the lipid barrier by disrupting interactions that surround cells (Islam et al., 2017). This is a rapid, efficient method that may provide soluble protein, which is not as costly as French Press but may pose an issue as this method could introduce contaminants to the sample, which, as observed in this project, may lead to downstream effects during purification (Ali et al., 2017).

With considerations, if French Press is available, it may provide an efficient way to lyse cells, otherwise, it may be better to optimise sonication for each protein.

4.3.3 Contamination in SDS-PAGE following protein purification of the gC1q domain construct

The SDS-PAGE analysis following affinity chromatography of the gC1q recombinant protein revealed an unexpected protein at approximately 60kDa, presenting with a higher yield than gC1q protein (Fig. 28). This contamination may have adversely affected the overall yield of the gC1q protein, as the contaminant protein may have outcompeted the target protein for binding sites during the purification process (Potel et al., 2018).

Bartlow et al. (2011) identified several common problematic host-proteins that compete with His-tagged recombinant proteins during immobilised metal affinity chromatography. This included glucosamine fructose-6-phosphate aminotransferase (GlmS) and bifunctional polymyxin resistance protein (ArnA), both around 60kDa, which may explain the contamination observed when trying to purify the gC1q domain. Bolanos-Garcia and Davies (2006) suggest that native *E. coli* contaminants are stress-induced proteins produced when *E. coli* undergo stress conditions such as heat shock, oxidative stress, and nutrient starvation, resulting in the co-purification of these proteins with recombinant proteins. However, while BL21(DE3) is not deficient in either ArnA or GlmS as potential contaminants, other endogenous proteins may have interfered during the purification process. Advanced techniques such as mass spectrometry may be employed to gain a more comprehensive understanding of the purified protein composition and confirm the identity of the contaminant. Mass Spectrometry may provide accurate protein identification, offering insights into the presence of the unexpected protein (Noor et al., 2021). Implementing analytical techniques, such as 2D gel electrophoresis coupled with mass spectrometry, may offer a deeper understanding of protein composition post-purification. While confirming the identity of the contaminant may offer an understanding of the co-purification, it is no obstacle when purifying the gC1q construct as ion-exchange chromatography and SEC may be employed.

4.4 Future directions

Before understanding the structural nature of the MMRN1 domains, protein expression issues would need to be addressed. Using the BL21(DE3) Rosetta cell line may be the easiest, most cost-efficient method to greatly improve expression for sufficient protein (Francis and Page, 2010). Further studies employing X-ray crystallography to provide a 3D outline of the individual domains are essential for gaining more insight into their functional roles, potential binding sites and overall conformational changes in a three-dimensional space (Maveyraud and Mourey, 2020). It could also facilitate structural-based drug design or modifications when testing and optimising the protein's performance in a variety of applications (potential drug-binding sites to modulate platelet function). Not only will information about MMRN1 be uncovered, but valuable insights into the shared domains of the EMILIN family will provide a broader context of how different proteins within the family interact and contribute to biological processes by establishing relationships among family members.

Beyond structural insights, the functional characterisation of MMRN1 and its domains is imperative when assessing its biological importance in coagulation, angiogenesis, or inflammation (Saini et al., 2020). Functional assays, both *in vitro* and *in vivo*, could further expand the physiological significance of MMRN1 while exploring potential diagnostic or therapeutic applications in multiple pathologies. Recent findings have linked MMRN1 to diseases such as cancer and QPD, suggesting its potential involvement in various signalling pathways (Blavignac et al., 2011; Posner, 2022). However, much remains to be discovered regarding its role in other diseases, therefore, a more in-depth analysis is needed to understand how MMRN1 may be implicated.

Exploring the interaction between MMRN1 and Efb through interaction studies (pull-down assays) to explore the mechanism of binding and the downstream effects may also provide insight into MMRN1's importance in its role in coagulation. According to various studies, domains involved in intermolecular interactions, like those in MMRN1, may be crucial when predicting PPIs (Murakami et al., 2017). Efb may be targeting a specific domain, and with the EGF-like domain being specific to MMRN1 across the EMILIN family, this may be the potential binding target (Colombatti et al., 2011). Although, the shared EMI and gC1q domains offer differential functions compared to their family counterparts and, therefore, may also be a binding target (Colombatti et al., 2011). Distinctions between the structural and functional

differences between the EMILIN family must be confirmed, and Efb-MMRN1 interactional studies will aim to reveal this information.

Platelets play a critical role in the pathogenesis of thrombosis and cardiovascular diseases, making them a target for therapeutic intervention (Holinstat, 2017). Medications with anti-thrombotic properties, such as Aspirin, affect platelet aggregation similarly to the action of Efb (Hannachi et al., 2019). Investigating Efb's molecular mechanism in relation to targeting MMRN1 may be useful in anti-thrombotic drug design that offers advantages over existing drugs like Aspirin. Also, with MMRN1 being a platelet carrier protein, exploring the mechanism of infection by *S. aureus* through Efb not only contributes to the understanding of critical physiological pathways but also holds the potential for the development of new drug interventions mitigating the virulence of *S. aureus*. These interventions may provide novel strategies for managing conditions related to abnormal platelet activity but will also help to develop the quality of health and reduce the financial burden associated with hospitalisation due to *S. aureus* infection (Beganovic et al., 2019). Additionally, it is worth exploring MMRN1-VacA interactions as VacA, a virulence factor produced by *H. Pylori*, has been observed to produce opposing effects compared to the impact of Efb (Sato et al., 2013). This will aid in understanding differences in bacterial pathogenicity not only to MMRN1 but in general.

Computational modelling is a new way of investigating proteins that may help predict the impact of mutation or modifications on a protein's behaviour (Dokholyan, 2020). AlphaFold2 may have predicted gC1q and EMI domain with confidence, but the overall structure and how the domains are arranged in 3D space is unclear (low confidence). Therefore, it is essential to determine its structure experimentally to help gain a more comprehensive understanding of its role in cellular networks.

4.4.1 Purification methods for gC1q domain homogeneity

With the gC1q construct having a theoretical isoelectric point (pI) of 8.46, this may be compared with the pI of the 60 kDa protein, and if different, they may be separated using ion-exchange chromatography (Wilkins, 1999). Mass spectrometry may be used to identify the protein which may then be used to find the predicted pI of the contaminant (Fenyo et al.,

2010). If gC1q is positively charged and the other protein negative, cation exchange chromatography should be used, if gC1q is negatively charged, anion exchange chromatography should be used (Stoyanov et al., 2011). Although, trying different pH values in ion exchange may be sufficient to separate the two proteins. Insights into the challenges when expressing MMRN1 may be useful when designing downstream experiments investigating its interactions and domain proteins.

4.4.2 Current methods of detecting Protein-protein interactions (PPI)

Pull-down assays, a form of affinity chromatography, are commonly used to assess interactions between two or more proteins through affinity-tagged bait proteins (Fig. 7) (Louche et al., 2017). In the context of this study, where the goal is to investigate protein-protein interactions between the purified domains of MMRN1 and Efb protein, *in vitro* binding assays may be primarily performed by immobilising purified His-tagged MMRN1 domains onto Nickel beads His-tag binding and incubating the solution with purified Efb to allow for binding (Louche et al., 2017). Any unbound proteins will be washed away while the bound proteins run on an SDS-PAGE gel. If MMRN1 and Efb interact and form a stable complex during the pull-down, two bands should be observed, one corresponding to Efb and the other to MMRN1, as the loading dye is reducing, breaking any non-covalent interactions. When a domain has shown to interact, it may be further analysed via spectrometry and other interaction analytical methods (Lyu et al., 2022; Klumper et al., 2018).

Since a MMRN1-Efb interaction has already been identified, using Co-IP in conjunction with Proximity ligation assay (PLA) may be more beneficial to provide details on specific PPIs. PLA involves detecting and quantifying protein interactions based on proximity, suggesting an interaction or co-localisation (Hegazy et al., 2020). Its ability to provide spatial information is valuable, but proximity does not always imply direct interaction, therefore, the potential for false positive results becomes an issue. With both methods using antibodies to identify interaction partners, Co-IP and PLA offer sensitivity and specificity to detect weak/transient interactions compared to pull-down assays that use affinity tags (Alam, 2022; Free et al., 2009). This is because affinity tags may be inaccessible with complex proteins due to occlusion (Bornhorst and Falke, 2000). This may have been seen with the recombinant gC1q domain

pull-down as the protein was difficult to visualise on an SDS-PAGE gel following incubation with His-beads (Fig. 27). Using two PPI methods may increase the validity of the observed interactions and address any false positive results, therefore it may be worth testing interaction through antibody detecting rather than using tagged bait proteins. Additionally, other biophysical techniques such as surface plasmon resonance (SPR), bio-layer interferometry (BLI), and isothermal titration calorimetry (ITC) play crucial roles when investigating molecular interactions by detecting changes in refractive index, interference patterns, and heat release or absorption to provide insight into binding kinetics, thermodynamics, and structural dynamics (de Mol and Fischer, 2010; Muller-Esparza et al., 2020; Johnson, 2021).

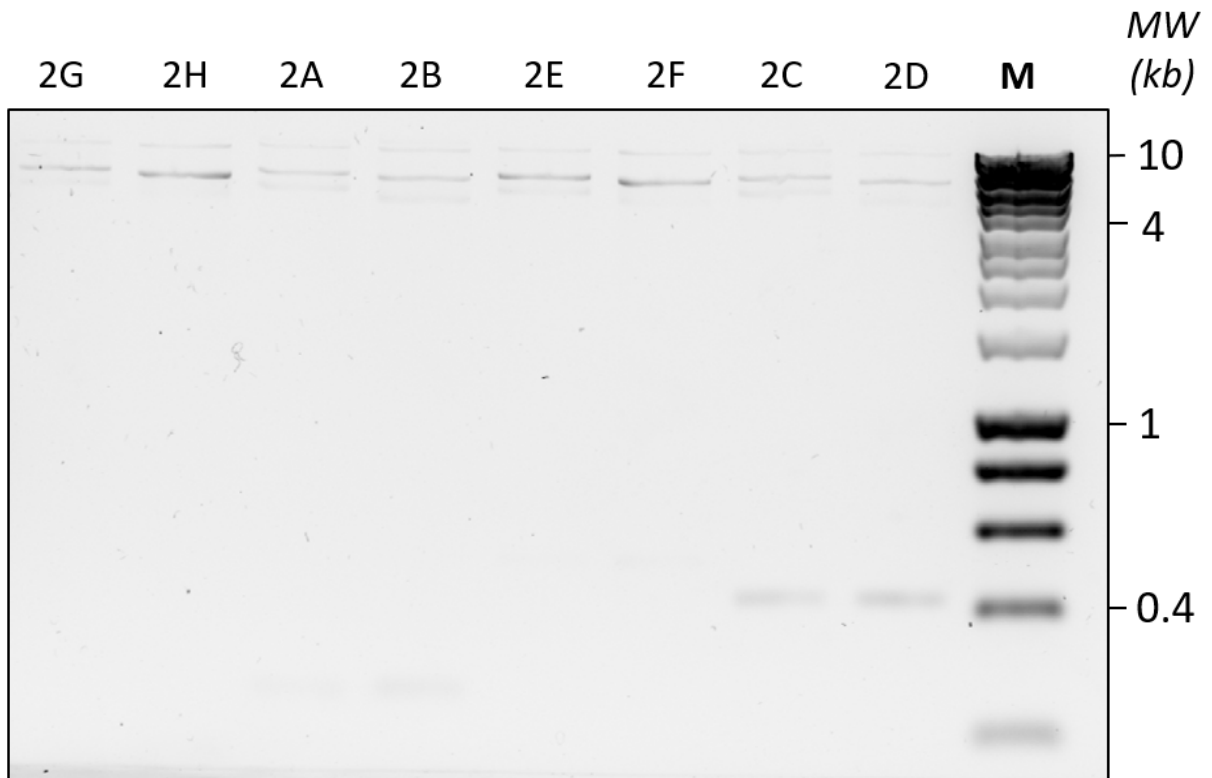
For future analysis, it may be worth investigating this interaction *in vivo*. In-cell interaction studies in the native cellular environment ensure that interactions occur naturally, preserving conditions, modifications, and factors that influence them (Tang et al., 2020; Nishida-Aoki and Gujral, 2019). Observing dynamic interactions in living cells provides insights into regulatory mechanisms and confirms their physiological relevance compared to the potential artifacts that may arise from *in vitro* observations (Bagheri et al., 2022). In-cell studies also enable the exploration of live-cell imaging, which captures real-time cellular processes (Jensen, 2013). However, most in-cell interaction studies use mammalian cell lines, for example, Biomolecular fluorescence complementation (BiFC) and Fluorescence resonance energy transfer (FRET) commonly use HEK cells, therefore, this may be okay when observing interactions with MMRN1 but would be less suitable when testing interactions with the *S. aureus* Efb protein (Schmitz et al., 2021; Zhuo and Knox, 2022). While Efb may be expressed in a mammalian cell line, it is important to consider many shortcomings that may occur with this method, for example, codons would need to be optimised, and mammalian cells may recognise the bacterial protein as foreign, resulting in an immune response and its functions and folding patterns may differ in a eukaryotic environment (Muir et al., 2017). Alternatively, as expression in mammalian or bacterial cells may be complicated, interactions with MMRN1 on platelets may be further investigated through flow cytometry, offering a close approximation to *in vivo* conditions.

5 Conclusion

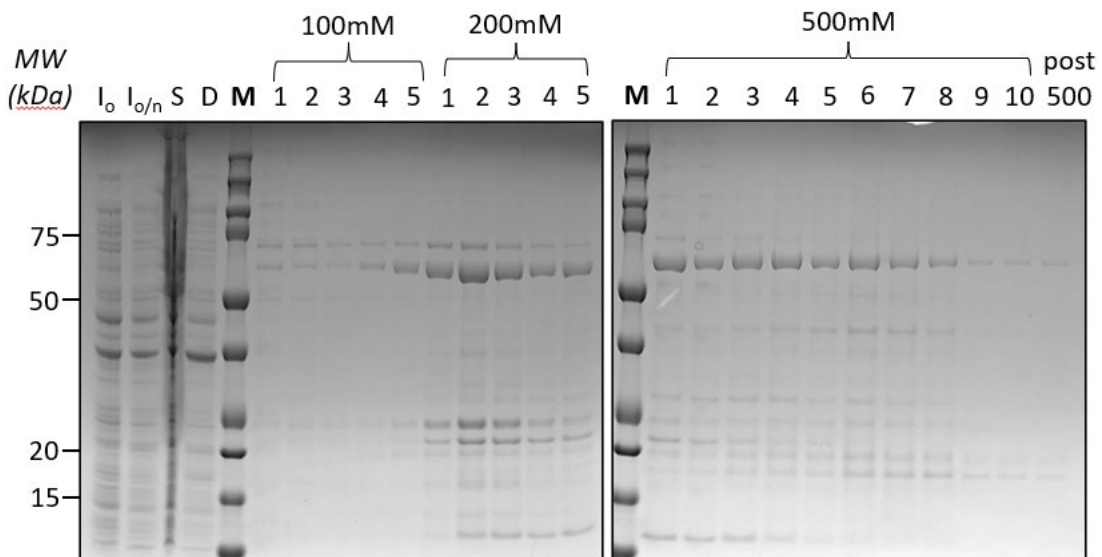
This study aimed to identify expression conditions that will lead to the successful purification of MMRN1 domain proteins to be analysed structurally (X-ray crystallography) and functionally (pull-down assays) to aid in the discovery of MMRN1's physiological roles. Having successfully purified one construct by affinity chromatography, the gC1q domain protein, while investigating the issues surrounding the expression and purification of MMRN1 protein domains, this will allow for further investigation into its structural significance and interactions with other proteins to result in the development of potential novel therapeutics. Given the lack of primary glycosylation sites, EMI and gC1q domain protein expression in *E. coli* is not anticipated to face issues when investigating structure and functionality, however, there is an emphasis on using different expression systems. Where *E. coli* is an easier and more robust system, it is not optimised to express the mammalian domain proteins. To comprehensively understand its physiological roles, focussing on both structural and functional aspects of the MMRN1 and its protein interactions is imperative. The holistic combination of structural, functional, and computational analyses will contribute to a more comprehensive understanding of the importance of MMRN1's role in physiological processes and its implications for health and disease by addressing current gaps in knowledge of this protein.

6 Appendix

6.1 Appendix 1: Unedited ligation DNA agarose gel.

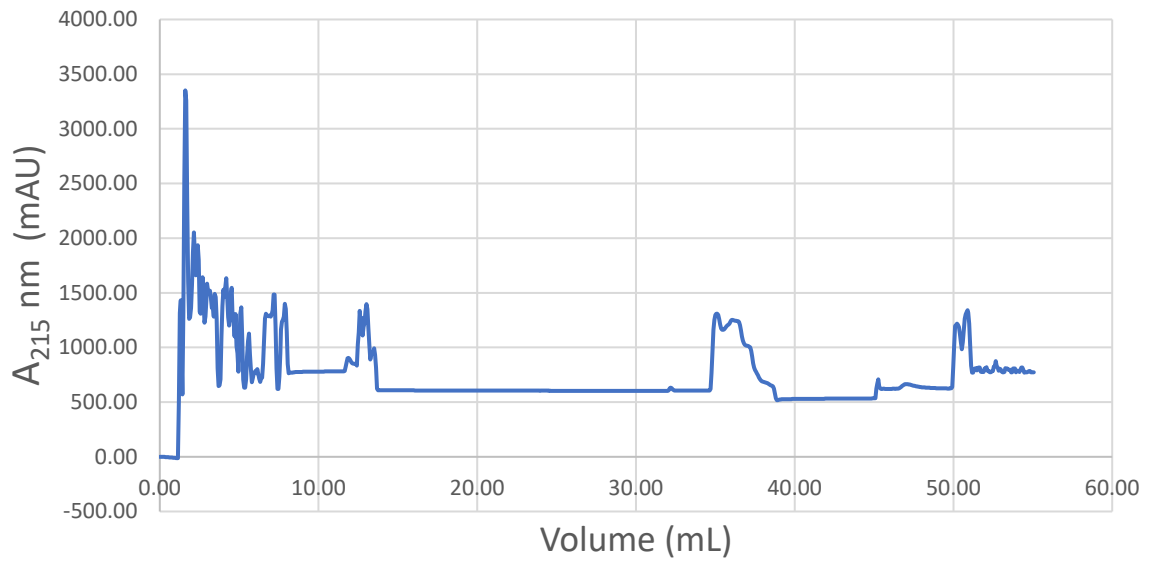


6.2 Appendix 2: Unedited purification SDS-PAGE gels of construct 1C (gC1q).



6.3 Appendix 3: Chromatogram of construct 1C (gC1q) following Äkta purification.

A chromatogram of gC1q domain (construct 1C) purification using the Äkta Purifier measured at a wavelength of 215nm.



7 References

- Adalbert, J. R., Varshney, K., Tobin, R. and Pajaro, R. (2021) 'Clinical outcomes in patients co-infected with COVID-19 and Staphylococcus aureus: a scoping review.' *BMC Infect Dis*, 21(1) p. 985.
- Aeschbach, L. and Dion, V. (2017) 'Minimizing carry-over PCR contamination in expanded CAG/CTG repeat instability applications.' *Sci Rep*, 7(1) p. 18026.
- Alam, M. S. (2022) 'Proximity Ligation Assay (PLA).' *Methods Mol Biol*, 2422, Dec, pp. 191-201.
- Ali, N., Rampazzo, R. C. P., Costa, A. D. T. and Krieger, M. A. (2017) 'Current Nucleic Acid Extraction Methods and Their Implications to Point-of-Care Diagnostics.' *Biomed Res Int*, 2017, Jul, p. 9306564.
- Anand, R. D., Sertil, O. and Lowry, C. V. (2004) 'Restriction digestion monitors facilitate plasmid construction and PCR cloning.' *Biotechniques*, 36(6) pp. 982-985.
- Anastasina, M., Terenin, I., Butcher, S. J. and Kainov, D. E. (2014) 'A technique to increase protein yield in a rabbit reticulocyte lysate translation system.' *Biotechniques*, 56(1) pp. 36-39.
- Baeshen, M. N., Al-Hejin, A. M., Bora, R. S., Ahmed, M. M., Ramadan, H. A., Saini, K. S., Baeshen, N. A. and Redwan, E. M. (2015) 'Production of Biopharmaceuticals in E. coli: Current Scenario and Future Perspectives.' *J Microbiol Biotechnol*, 25(7) pp. 953-962.
- Bagheri, Y., Ali, A. A., Keshri, P., Chambers, J., Gershenson, A. and You, M. (2022) 'Imaging Membrane Order and Dynamic Interactions in Living Cells with a DNA Zipper Probe.' *Angew Chem Int Ed Engl*, 61(6) p. e202112033.
- Bartlow, P., Uechi, G. T., Cardamone, J. J., Jr., Sultana, T., Fruchtl, M., Beitle, R. R. and Ataii, M. M. (2011) 'Identification of native Escherichia coli BL21 (DE3) proteins that bind to immobilized metal affinity chromatography under high imidazole conditions and use of 2D-DIGE to evaluate contamination pools with respect to recombinant protein expression level.' *Protein Expr Purif*, 78(2) pp. 216-224.
- Basak, A., Shervani, N. J., Mbikay, M. and Kolajova, M. (2008) 'Recombinant proprotein convertase 4 (PC4) from Leishmania tarentolae expression system: purification, biochemical study and inhibitor design.' *Protein Expr Purif*, 60(2) pp. 117-126.
- Bastien, P., Chabbert, E. and Lachaud, L. (2003) 'Contamination management of broad-range or specific PCR: is there any difference?' *J Clin Microbiol*, 41(5) p. 2272
- Beganovic, M., Cusumano, J. A., Lopes, V., LaPlante, K. L. and Caffrey, A. R. (2019) 'Comparative Effectiveness of Exclusive Exposure to Nafcillin or Oxacillin, Cefazolin, Piperacillin/Tazobactam, and Fluoroquinolones Among a National Cohort of Veterans With

Methicillin-Susceptible Staphylococcus aureus Bloodstream Infection.' *Open Forum Infect Dis*, 6(7) p. ofz270.

Ben-Abdallah, M., Bondet, V., Fauchereau, F., Beguin, P., Goubran-Botros, H., Pagan, C., Bourgeron, T. and Bellalou, J. (2011) 'Production of soluble, active acetyl serotonin methyl transferase in Leishmania tarentolae.' *Protein Expr Purif*, 75(1) pp. 114-118.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., Bourne, P.E. (2000) 'The Protein Data Bank.' *Nucleic Acids Research*, 28(1) pp. 235-242.

Bessa, D., Pereira, F., Moreira, R., Johansson, B. and Queiros, O. (2012) 'Improved gap repair cloning in yeast: treatment of the gapped vector with Taq DNA polymerase avoids vector self-ligation.' *Yeast*, 29(10) pp. 419-423.

Blavignac, J., Bunimov, N., Rivard, G. E. and Hayward, C. P. (2011) 'Quebec platelet disorder: update on pathogenesis, diagnosis, and treatment.' *Semin Thromb Hemost*, 37(6) pp. 713-720.

Bolanos-Garcia, V. M. and Davies, O. R. (2006) 'Structural analysis and classification of native proteins from E. coli commonly co-purified by immobilised metal affinity chromatography.' *Biochim Biophys Acta*, 1760(9) pp. 1304-1313.

Bollati, M. and Gourlay, L. J. (2022) 'Protein Crystallization of Two Recombinant Lpt Proteins.' *Methods Mol Biol*, 2548, Sep, pp. 249-263.

Bornhorst, J. A. and Falke, J. J. (2000) 'Purification of proteins using polyhistidine affinity tags.' *Methods Enzymol*, 326, Jan, pp. 245-254.

Brady, A. E., Wang, Q. and Limbird, L. E. (2004) 'Study of G-protein-coupled receptor-protein interactions using gel overlay assays and glutathione-S-transferase-fusion protein pull-downs.' *Methods Mol Biol*, 259 pp. 371-378.

Breitling, R., Klingner, S., Callewaert, N., Pietrucha, R., Geyer, A., Ehrlich, G., Hartung, R., Muller, A., et al. (2002) 'Non-pathogenic trypanosomatid protozoa as a platform for protein research and production.' *Protein Expr Purif*, 25(2) pp. 209-218.

Brelford, A., Collin, H., Perrin, N. and Fumagalli, L. (2012) 'Nonspecific PCR amplification by high-fidelity polymerases: implications for next-generation sequencing of AFLP markers.' *Mol Ecol Resour*, 12(1) pp. 123-127.

Briand, L., Marcion, G., Kriznik, A., Heydel, J. M., Artur, Y., Garrido, C., Seigneuric, R. and Neiers, F. (2016) 'A self-inducible heterologous protein expression system in Escherichia coli.' *Sci Rep*, 6, Sep, p. 33037.

Brondyk, W. H. (2009) 'Selecting an appropriate method for expressing a recombinant protein.' *Methods Enzymol*, 463, Nov, pp. 131-147.

Burgess-Brown, N. A., Sharma, S., Sobott, F., Loenarz, C., Oppermann, U. and Gileadi, O. (2008) 'Codon optimization can improve expression of human genes in Escherichia coli: A multi-gene study.' *Protein Expr Purif*, 59(1) pp. 94-102.

Bustin, S. A., Mueller, R. and Nolan, T. (2020) 'Parameters for Successful PCR Primer Design.' *Methods Mol Biol*, 2065, Oct, pp. 5-22.

Carlson, E., Bays, N., David, L. and Skach, W. R. (2005) 'Reticulocyte lysate as a model system to study endoplasmic reticulum membrane protein degradation.' *Methods Mol Biol*, 301 pp. 185-205.

Chokchaichamnankit, D., Watcharatanyatip, K., Subhasitanont, P., Weeraphan, C., Keeratichamroen, S., Sritana, N., Kantathavorn, N., Diskul-Na-Ayudthaya, P., et al. (2019) 'Urinary biomarkers for the diagnosis of cervical cancer by quantitative label-free mass spectrometry analysis.' *Oncol Lett*, 17(6) pp. 5453-5468.

Colombatti, A., Spessotto, P., Doliana, R., Mongiat, M., Bressan, G. M. and Esposito, G. (2011) 'The EMILIN/Multimerin family.' *Front Immunol*, 2, Jan, pp. 93.

Corpet, F. (1988) 'Multiple sequence alignment with hierarchical clustering.' *Nucleic Acids Res*, 16(22), Nov 25, pp. 10881-10890.

Cortes-Rios, J., Zarate, A. M., Figueroa, J. D., Medina, J., Fuentes-Lemus, E., Rodriguez-Fernandez, M., Aliaga, M. and Lopez-Alarcon, C. (2020) 'Protein quantification by bicinchoninic acid (BCA) assay follows complex kinetics and can be performed at short incubation times.' *Anal Biochem*, 608, Nov, p. 113904.

Cromar, G. L., Xiong, X., Chautard, E., Ricard-Blum, S. and Parkinson, J. (2012) 'Toward a systems level view of the ECM and related proteins: a framework for the systematic definition and analysis of biological systems.' *Proteins*, 80(6) pp. 1522-1544.

Dadashpour, M., Fukuta, Y. and Asano, Y. (2011) 'Comparative expression of wild-type and highly soluble mutant His103Leu of hydroxynitrile lyase from *Manihot esculenta* in prokaryotic and eukaryotic expression systems.' *Protein Expr Purif*, 77(1) pp. 92-97.

Dallas-Yang, Q., Jiang, G. and Sladek, F. M. (1998) 'Avoiding false positives in colony PCR.' *Biotechniques*, 24(4) pp. 580-582.

de Mol, N. J. and Fischer, M. J. (2010) 'Surface plasmon resonance: a general introduction.' *Methods Mol Biol*, 627, Jan, pp. 1-14.

de Oliveira, T.A., da Silva, W., da Rocha Torres, N., Badaró de Moraes, J.V., Senra, R.L., de Oliveira Mendes, T.A., Júnior, A.S., Bressan, G.C. and Fietto, J.L.R. (2019) 'Application of the LEXSY *Leishmania tarentolae* system as a recombinant protein expression platform: A review.' *Process Biochemistry*, 87, Dec, pp.164–173.

Dessau, M. A. and Modis, Y. (2011) 'Protein crystallization for X-ray crystallography.' *J Vis Exp*, (47) Jan, p. 2285.

Dokholyan, N. V. (2020) 'Experimentally-driven protein structure modeling.' *J Proteomics*, 220, May, p. 103777.

Doliana, R., Bot, S., Bonaldo, P. and Colombatti, A. (2000) 'EMI, a novel cysteine-rich domain of EMILINs and other extracellular proteins, interacts with the gC1q domains and participates in multimerization.' *FEBS Lett*, 484(2) pp. 164-168.

Ducker, C., Ratnam, M., Shaw, P. E. and Layfield, R. (2023) 'Comparative analysis of protein expression systems and PTM landscape in the study of transcription factor ELK-1.' *Protein Expr Purif*, 203, Mar, p. 106216.

Evans, B. A., Smith, O. L., Pickerill, E. S., York, M. K., Buenconsejo, K. J. P., Chambers, A. E. and Bernstein, D. A. (2018) 'Restriction digest screening facilitates efficient detection of site-directed mutations introduced by CRISPR in *C. albicans* UME6.' *PeerJ*, 6, Jun, p. e4920.

Fenyo, D., Eriksson, J. and Beavis, R. (2010) 'Mass spectrometric protein identification using the global proteome machine.' *Methods Mol Biol*, 673, Aug, pp. 189-202.

Francis, D. M. and Page, R. (2010) 'Strategies to optimize protein expression in *E. coli*.' *Curr Protoc Protein Sci*, Chapter 5(1), Aug, pp. 5 24 21-25 24 29.

Fredriksson-Ahomaa, M. and Korkeala, H. (2003) 'Low occurrence of pathogenic *Yersinia enterocolitica* in clinical, food, and environmental samples: a methodological problem.' *Clin Microbiol Rev*, 16(2) pp. 220-229.

Free, R. B., Hazelwood, L. A. and Sibley, D. R. (2009) 'Identifying novel protein-protein interactions using co-immunoprecipitation and mass spectroscopy.' *Curr Protoc Neurosci*, Chapter 5, Jan, p. Unit 5 28.

Friligou, I., Gassner, J., Knoblauch, D., Kagerer, G., Popp, F., Voit, S., Engel, A. M., Leinenbach, A., et al. (2021) 'Glycosylation of recombinant rabbit immunoglobulins influences protease susceptibility as shown by comprehensive mass spectrometric glycan analysis.' *Glycobiology*, 31(7) pp. 762-771.

Frontini, M. (2020) 'Breaking barriers: Quebec platelet disorder.' *Blood*, 136(23) pp. 2603-2604.

Gagoski, D., Polinkovsky, M. E., Mureev, S., Kunert, A., Johnston, W., Gambin, Y. and Alexandrov, K. (2016) 'Performance benchmarking of four cell-free protein expression systems.' *Biotechnol Bioeng*, 113(2) pp. 292-300.

Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D. and Bairoch, A. (2003) 'ExPASy: The proteomics server for in-depth protein knowledge and analysis.' *Nucleic Acids Res*, 31(13) pp. 3784-3788.

Gkolfinopoulos, S., Jones, R. L. and Constantinidou, A. (2020) 'The Emerging Role of Platelets in the Formation of the Micrometastatic Niche: Current Evidence and Future Perspectives.' *Front Oncol*, 10, Mar, p. 374.

Goddard, T. D., Huang, C. C., Meng, E. C., Pettersen, E. F., Couch, G. S., Morris, J. H. and Ferrin, T. E. (2018) 'UCSF ChimeraX: Meeting modern challenges in visualization and analysis.' *Protein Sci*, 27(1) pp. 14-25.

Goldberg, S. (2008) 'Mechanical/physical methods of cell disruption and tissue homogenization.' *Methods Mol Biol*, 424 pp. 3-22.

Gomes, L., Monteiro, G. and Mergulhao, F. (2020) 'The Impact of IPTG Induction on Plasmid Stability and Heterologous Protein Expression by Escherichia coli Biofilms.' *Int J Mol Sci*, 21(2) p. 576.

Gray, T., Storz, G. and Papenfort, K. (2022) 'Small Proteins; Big Questions.' *J Bacteriol*, 204(1) p. e0034121.

Haemmerle, M., Stone, R. L., Menter, D. G., Afshar-Kharghan, V. and Sood, A. K. (2018) 'The Platelet Lifeline to Cancer: Challenges and Opportunities.' *Cancer Cell*, 33(6) pp. 965-983.

Hamzeh-Cognasse, H., Damien, P., Chabert, A., Pozzetto, B., Cognasse, F., Garraud, O. (2015) 'Platelets and infections – complex interactions with bacteria.' *Frontiers in Immunology*, 6, Feb, p. 82.

Hannachi, N., Habib, G. and Camoin-Jau, L. (2019) 'Aspirin Effect on Staphylococcus aureus-Platelet Interactions During Infectious Endocarditis.' *Front Med (Lausanne)*, 6, Oct, p. 217.

Harding, C. M. and Feldman, M. F. (2019) 'Glycoengineering bioconjugate vaccines, therapeutics, and diagnostics in E. coli.' *Glycobiology*, 29(7) pp. 519-529.

Hayward, C. P., Warkentin, T. E., Horsewood, P. and Kelton, J. G. (1991) 'Multimerin: a series of large disulfide-linked multimeric proteins within platelets.' *Blood*, 77(12) pp. 2556-2560.

Hegazy, M., Cohen-Barak, E., Koetsier, J. L., Najor, N. A., Arvanitis, C., Sprecher, E., Green, K. J. and Godsel, L. M. (2020) 'Proximity Ligation Assay for Detecting Protein-Protein Interactions and Protein Modifications in Cells and Tissues in Situ.' *Curr Protoc Cell Biol*, 89(1) p. e115.

Hemayatkar, M., Mahboudi, F., Majidzadeh, A. K., Davami, F., Vaziri, B., Barkhordari, F., Adeli, A., Mahdian, R., et al. (2010) 'Increased expression of recombinant human tissue plasminogen activator in Leishmania tarentolae.' *Biotechnol J*, 5(11) pp. 1198-1206.

Henke, E., Nandigama, R. and Ergun, S. (2020) 'Extracellular Matrix in the Tumor Microenvironment and Its Impact on Cancer Therapy.' *Front Mol Biosci*, 6, Jan, p. 160.

Henry, N. L. and Hayes, D. F. (2012) 'Cancer biomarkers.' *Mol Oncol*, 6(2) pp. 140-146.

Heptinstall, S. (2012) 'Platelet activation by an extracellular adherence protein from *Staphylococcus aureus* acting via modulation of sulfhydryl groups on platelets.' *Arterioscler Thromb Vasc Biol*, 32(8) pp. 1751-1752.

Heyde, S. A. H. and Norholm, M. H. H. (2021) 'Tailoring the evolution of BL21(DE3) uncovers a key role for RNA stability in gene expression toxicity.' *Commun Biol*, 4(1) p. 963.

Hipolito, C. J., Bashiruddin, N. K. and Suga, H. (2014) 'Protein cocrystallization molecules originating from in vitro selected macrocyclic peptides.' *Curr Opin Struct Biol*, 26, Jun, pp. 24-31.

Holinstat, M. (2017) 'Complement factors (H) into thrombosis.' *Blood*, 129(9) pp. 1065-1066.

Huang, Y., Zhang, X., Jiang, W., Wang, Y., Jin, H., Liu, X. and Xu, C. (2012) 'Discovery of serum biomarkers implicated in the onset and progression of serous ovarian cancer in a rat model using iTRAQ technique.' *Eur J Obstet Gynecol Reprod Biol*, 165(1) pp. 96-103.

Islam, M. S., Aryasomayajula, A. and Selvaganapathy, P. R. (2017) 'A review on macroscale and microscale cell lysis methods.' *Micromachines (Basel)*, 8(3) p. 83.

Iwashkiw, J. A., Fentabil, M. A., Faridmoayer, A., Mills, D. C., Peppler, M., Czibener, C., Ciocchini, A. E., Comerici, D. J., et al. (2012) 'Exploiting the *Campylobacter jejuni* protein glycosylation system for glycoengineering vaccines and diagnostic tools directed against brucellosis.' *Microb Cell Fact*, 11, Jan, p. 13.

Jaroentomeechai, T., Stark, J. C., Natarajan, A., Glasscock, C. J., Yates, L. E., Hsu, K. J., Mrksich, M., Jewett, M. C., et al. (2018) 'Single-pot glycoprotein biosynthesis using a cell-free transcription-translation system enriched with glycosylation machinery.' *Nat Commun*, 9(1) p. 2686.

Jenkins, S. G. and Schuetz, A. N. (2012) 'Current concepts in laboratory testing to guide antimicrobial therapy.' *Mayo Clin Proc*, 87(3) pp. 290-308.

Jensen, E. C. (2013) 'Overview of live-cell imaging: requirements and methods used.' *Anat Rec (Hoboken)*, 296(1) pp. 1-8.

Jia, B. and Jeon, C. O. (2016) 'High-throughput recombinant protein expression in *Escherichia coli*: current status and future perspectives.' *Open Biol*, 6(8) p. 160196

Johnson, C. M. (2021) 'Isothermal Titration Calorimetry.' *Methods Mol Biol*, 2263, Apr, pp. 135-159.

Jongerius, I., von Kockritz-Blickwede, M., Horsburgh, M. J., Ruyken, M., Nizet, V. and Rooijackers, S. H. (2012) '*Staphylococcus aureus* virulence is enhanced by secreted factors that block innate immune defenses.' *J Innate Immun*, 4(3) pp. 301-311.

Kalle, E., Kubista, M. and Rensing, C. (2014) 'Multi-template polymerase chain reaction.' *Biomol Detect Quantif*, 2, Dec, pp. 11-29.

Keeratichamroen, S., Subhasitanont, P., Chokchaichamnankit, D., Weeraphan, C., Saharat, K., Sritana, N., Kantathavorn, N., Wiriyaucharadecha, K., et al. (2020) 'Identification of potential cervical cancer serum biomarkers in Thai patients.' *Oncol Lett*, 19(6) pp. 3815-3826.

Khan, A. H., Bayat, H., Rajabibazl, M., Sabri, S. and Rahimpour, A. (2017) 'Humanizing glycosylation pathways in eukaryotic expression systems.' *World J Microbiol Biotechnol*, 33(1) p. 4.

Kim, J., Luo, H., White, W., Rees, W., Venkat, R. and Albarghouthi, M. (2020) 'Impact of Fc N-linked glycans on in vivo clearance of an immunoglobulin G1 antibody produced by NSO cell line.' *MAbs*, 12(1) p. 1844928.

Klumper, J., Oeljeklaus, S., Warscheid, B., Erdmann, R. and Schliebs, W. (2018) 'Using Pull Down Strategies to Analyze the Interactome of Peroxisomal Membrane Proteins in Human Cells.' *Subcell Biochem*, 89, Oct, pp. 261-285.

Kokkaliaris, K. D. and Scadden, D. T. (2020) 'Cell interactions in the bone marrow microenvironment affecting myeloid malignancies.' *Blood Adv*, 4(15) pp. 3795-3803.

Kostylev, M., Otwell, A. E., Richardson, R. E. and Suzuki, Y. (2015) 'Cloning Should Be Simple: Escherichia coli DH5alpha-Mediated Assembly of Multiple DNA Fragments with Short End Homologies.' *PLoS One*, 10(9) p. e0137466.

Kuipers, A., Stapels, D. A. C., Weerwind, L. T., Ko, Y-P., Ruyken, M. Lee, J. C., van Kessel, K. P. M., Rooijackers, S. H. M. (2016) 'The *Staphylococcus aureus* polysaccharide capsule and Efb-dependent fibrinogen shield act in concert to protect against phagocytosis.' *Microbiology (Reading)*, 162(7) pp. 1185-1194.

Kurien, B. T. and Scofield, R. H. (2012) 'Common artifacts and mistakes made in electrophoresis.' *Methods Mol Biol*, 869, Jan, pp. 633-640.

Laszlo, G. S., Alonzo, T. A., Gudgeon, C. J., Harrington, K. H., Gerbing, R. B., Wang, Y. C., Ries, R. E., Raimondi, S. C., et al. (2015) 'Multimerin-1 (MMRN1) as Novel Adverse Marker in Pediatric Acute Myeloid Leukemia: A Report from the Children's Oncology Group.' *Clin Cancer Res*, 21(14) pp. 3187-3195.

Leatherdale, A., Parker, D., Tasneem, S., Wang, Y., Bihan, D., Bonna, A., Hamaia, S. W., Gross, P. L., et al. (2021) 'Multimerin 1 supports platelet function in vivo and binds to specific GPAGPOGPX motifs in fibrillar collagens that enhance platelet adhesion.' *J Thromb Haemost*, 19(2) pp. 547-561.

Lee, L. Y., Hook, M., Haviland, D., Wetsel, R. A., Yonter, E. O., Syribeys, P., Vernachio, J. and Brown, E. L. (2004) 'Inhibition of complement activation by a secreted *Staphylococcus aureus* protein.' *J Infect Dis*, 190(3) pp. 571-579.

Liu, X., Liu, K., Nie, D., Zhang, J., Zhang, L., Liu, X. and Wang, J. (2022) 'Case report: Biochemical and clinical phenotypes caused by cysteine substitutions in the epidermal growth factor-like domains of fibrillin-1.' *Front Genet*, 13, Aug, p. 928683.

Louche, A., Salcedo, S. P. and Bigot, S. (2017) 'Protein-Protein Interactions: Pull-Down Assays.' *Methods Mol Biol*, 1615, Jul, pp. 247-255.

Lozano Terol, G., Gallego-Jara, J., Sola Martinez, R. A., Martinez Vivancos, A., Canovas Diaz, M. and de Diego Puente, T. (2021) 'Impact of the Expression System on Recombinant Protein Production in Escherichia coli BL21.' *Front Microbiol*, 12, Jun, p. 682001.

Lucotti, S. and Muschel, R. J. (2020) 'Platelets and Metastasis: New Implications of an Old Interplay.' *Front Oncol*, 10, Sep, p. 1350.

Luo, L., King, N. P., Yeo, J. C., Jones, A. and Stow, J. L. (2014) 'Single-step protease cleavage elution for identification of protein-protein interactions from GST pull-down and mass spectrometry.' *Proteomics*, 14(1) pp. 19-23.

Lyu, S., Zhang, C., Hou, X. and Wang, A. (2022) 'Tag-Based Pull-Down Assay.' *Methods Mol Biol*, 2400, Dec, pp. 105-114.

Mason, M., Sweeney, B., Cain, K., Stephens, P. and Sharfstein, S. T. (2014) 'Reduced Culture Temperature Differentially Affects Expression and Biophysical Properties of Monoclonal Antibody Variants.' *Antibodies (Basel)*, 3(3) pp. 253-271.

Maveyraud, L. and Mourey, L. (2020) 'Protein X-ray Crystallography and Drug Discovery.' *Molecules*, 25(5) p. 1030.

Meng, E. C., Goddard, T. D., Pettersen, E. F., Couch, G. S., Pearson, Z. J., Morris, J. H. and Ferrin, T. E. (2023) 'UCSF ChimeraX: Tools for structure building and analysis.' *Protein Sci*, 32(11) p. e4792.

Mirdita, M., Schutze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S. and Steinegger, M. (2022) 'ColabFold: making protein folding accessible to all.' *Nat Methods*, 19(6), Jun, 20220530, pp. 679-682.

Mongiati, M., Mungiguerra, G., Bot, S., Mucignat, M. T., Giacomello, E., Doliana, R. and Colombatti, A. (2000) 'Self-assembly and supramolecular organization of EMILIN.' *J Biol Chem*, 275(33) pp. 25471-25480.

Mongiati, M., Marastoni, S., Ligresti, G., Lorenzon, E., Schiappacassi, M., Perris, R., Frustaci, S. and Colombatti, A. (2010) 'The extracellular matrix glycoprotein elastin microfibril interface located protein 2: a dual role in the tumor microenvironment.' *Neoplasia*, 12(4) pp. 294-304.

- Muir, E., Raza, M., Ellis, C., Burnside, E., Love, F., Heller, S., Elliot, M., Daniell, E., et al. (2017) 'Trafficking and processing of bacterial proteins by mammalian cells: Insights from chondroitinase ABC.' *PLoS One*, 12(11) p. e0186759.
- Muller-Esparza, H., Osorio-Valeriano, M., Steube, N., Thanbichler, M. and Randau, L. (2020) 'Bio-Layer Interferometry Analysis of the Target Binding Activity of CRISPR-Cas Effector Complexes.' *Front Mol Biosci*, 7, May, p. 98.
- Murakami, Y., Tripathi, L. P., Prathipati, P. and Mizuguchi, K. (2017) 'Network analysis and in silico prediction of protein-protein interactions with applications in drug discovery.' *Curr Opin Struct Biol*, 44, Jun, pp. 134-142.
- Murphy, N., Rooney, B., Bhattacharyya, T., Triana-Chavez, O., Krueger, A., Haslam, S. M., O'Rourke, V., Panczuk, M., et al. (2020) 'Glycosylation of Trypanosoma cruzi TcI antigen reveals recognition by chagasic sera.' *Sci Rep*, 10(1) p. 16395.
- Navarro, A., Sanseverino, I., Cappelli, F., Lahm, A., Niegowska, M., Fabbri, M., Paracchini, V., Petrillo, M., et al. (2023) 'Study of antibiotic resistance in freshwater ecosystems with low anthropogenic impact.' *Sci Total Environ*, 857(Pt 3) p. 159378.
- Neelamegham, S., Aoki-Kinoshita, K., Bolton, E., Frank, M., Lisacek, F., Lutteke, T., O'Boyle, N., Packer, N. H., et al. (2019) 'Updates to the Symbol Nomenclature for Glycans guidelines.' *Glycobiology*, 29(9), Aug 20, pp. 620-624.
- New England Biolabs (2023) *NEBioCalculator*. Available at: <https://nebiocalculator.neb.com> (Accessed: 17 December 2023).
- Nguyen, T., Ghebrehiwet, B. and Peerschke, E. I. (2000) 'Staphylococcus aureus protein A recognizes platelet gC1qR/p33: a novel mechanism for staphylococcal interactions with platelets.' *Infect Immun*, 68(4) pp. 2061-2068.
- Nishida-Aoki, N. and Gujral, T. S. (2019) 'Emerging approaches to study cell-cell interactions in tumor microenvironment.' *Oncotarget*, 10(7) pp. 785-797.
- Noor, Z., Ahn, S. B., Baker, M. S., Ranganathan, S. and Mohamedali, A. (2021) 'Mass spectrometry-based protein identification in proteomics-a review.' *Brief Bioinform*, 22(2) pp. 1620-1638.
- Parker, D. N., Tasneem, S., Farndale, R. W., Bihan, D., Sadler, J. E., Sebastian, S., de Groot, P. G. and Hayward, C. P. (2016) 'The functions of the A1A2A3 domains in von Willebrand factor include multimerin 1 binding.' *Thromb Haemost*, 116(1) pp. 87-95.
- Pchelintsev, N. A., Adams, P. D. and Nelson, D. M. (2016) 'Critical Parameters for Efficient Sonication and Improved Chromatin Immunoprecipitation of High Molecular Weight Proteins.' *PLoS One*, 11(1) p. e0148023.

Peleg, M. and Corradini, M. G. (2011) 'Microbial growth curves: what the models tell us and what they cannot.' *Crit Rev Food Sci Nutr*, 51(10) pp. 917-945.

Pellicani, R., Poletto, E., Andreuzzi, E., Paulitti, A., Doliana, R., Bizzotto, D., Braghetta, P., Colladel, R., et al. (2020) 'Multimerin-2 maintains vascular stability and permeability.' *Matrix Biol*, 87, May, pp. 11-25.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Meng, E. C., Couch, G. S., Croll, T. I., Morris, J. H. and Ferrin, T. E. (2021) 'UCSF ChimeraX: Structure visualization for researchers, educators, and developers.' *Protein Sci*, 30(1) pp. 70-82.

Posner, M. G., Upadhyay, A., Abubaker, A. A., Fortunato, T. M., Vara, D., Canobbio, I., Bagby, S. and Pula, G. (2016) 'Extracellular Fibrinogen-binding Protein (Efb) from *Staphylococcus aureus* Inhibits the Formation of Platelet-Leukocyte Complexes.' *J Biol Chem*, 291(6) pp. 2764-2776.

Posner, M. G. (2022) 'Multimerin-1 and cancer: a review.' *Biosci Rep*, 42(2) p. BSR20211248.

Potel, C. M., Lin, M. H., Heck, A. J. R. and Lemeer, S. (2018) 'Defeating Major Contaminants in Fe(3+)- Immobilized Metal Ion Affinity Chromatography (IMAC) Phosphopeptide Enrichment.' *Mol Cell Proteomics*, 17(5) pp. 1028-1034.

Qi, Y., Lv, J., Liu, S., Sun, L., Wang, Y., Li, H., Qi, W. and Qiu, W. (2019) 'TSPAN9 and EMILIN1 synergistically inhibit the migration and invasion of gastric cancer cells by increasing TSPAN9 expression.' *BMC Cancer*, 19(1) p. 630.

Qian, X. and Wen-jun, L. (2013) 'Platelet changes in acute leukemia.' *Cell Biochem Biophys*, 67(3) pp. 1473-1479.

Rabajdova, M., Urban, P., Spakova, I., Saksun, L., Dudic, R., Ostro, A., Caprnda, M., Kruzliak, P., et al. (2016) 'The crucial role of emilin 1 gene expression during progression of tumor growth.' *J Cancer Res Clin Oncol*, 142(11) pp. 2397-2402.

Rudd, P. M. and Dwek, R. A. (1997) 'Glycosylation: heterogeneity and the 3D structure of proteins.' *Crit Rev Biochem Mol Biol*, 32(1) pp. 1-100.

Ruiz-Villalba, A., van Pelt-Verkuil, E., Gunst, Q. D., Ruijter, J. M. and van den Hoff, M. J. (2017) 'Amplification of nonspecific products in quantitative polymerase chain reactions (qPCR).' *Biomol Detect Quantif*, 14, Dec, pp. 7-18.

Saini, A., Chandra, K. B., Kumar, V., Mathur, S. R., Sharma, J. B., Kumar, S. and Yadav, S. (2020) 'Analysis of Multimerin 1 (MMRN1) expression in ovarian cancer.' *Mol Biol Rep*, 47(12) pp. 9459-9468.

Satoh, K., Hirayama, T., Takano, K., Suzuki-Inoue, K., Sato, T., Ohta, M., Nakagomi, J. and Ozaki, Y. (2013) 'VacA, the vacuolating cytotoxin of *Helicobacter pylori*, binds to multimerin 1 on human platelets.' *Thromb J*, 11(1) p. 23.

Scherlinger, M., Richez, C., Tsokos, G. C., Boilard, E. and Blanco, P. (2023) 'The role of platelets in immune-mediated inflammatory diseases.' *Nat Rev Immunol*, 23(8) pp. 495-510.

Schiavinato, A., Becker, A. K., Zanetti, M., Corallo, D., Milanetto, M., Bizzotto, D., Bressan, G., Guljelmovic, M., et al. (2012) 'EMILIN-3, peculiar member of elastin microfibril interface-located protein (EMILIN) family, has distinct expression pattern, forms oligomeric assemblies, and serves as transforming growth factor beta (TGF-beta) antagonist.' *J Biol Chem*, 287(14) pp. 11498-11515.

Schilling, J., Schoppe, J., Sauer, E. and Pluckthun, A. (2014) 'Co-crystallization with conformation-specific designed ankyrin repeat proteins explains the conformational flexibility of BCL-W.' *J Mol Biol*, 426(12), Jun 12, 20140418, pp. 2346-2362.

Schmidt, F. R. (2004) 'Recombinant expression systems in the pharmaceutical industry.' *Appl Microbiol Biotechnol*, 65(4) pp. 363-372.

Schmitz, F., Glas, J., Neutze, R. and Hedfalk, K. (2021) 'A bimolecular fluorescence complementation flow cytometry screen for membrane protein interactions.' *Sci Rep*, 11(1) p. 19232.

Selvadurai, M. V. and Hamilton, J. R. (2018) 'Structure and function of the open canalicular system - the platelet's specialized internal membrane network.' *Platelets*, 29(4) pp. 319-325.

Siboo, I. R., Chambers, H. F. and Sullam, P. M. (2005) 'Role of SraP, a Serine-Rich Surface Protein of *Staphylococcus aureus*, in binding to human platelets.' *Infect Immun*, 73(4) pp. 2273-2280.

Sinha, S., Nevett, C., Shuttleworth, C. A. and Kielty, C. M. (1998) 'Cellular and extracellular biology of the latent transforming growth factor-beta binding proteins.' *Matrix Biol*, 17(8-9) pp. 529-545.

Skarratt, K. K. and Fuller, S. J. (2014) 'Quantitative real-time PCR eliminates false-positives in colony screening PCR.' *J Microbiol Methods*, 96, Jan, pp. 99-100.

Sodoyer, R. (2004) 'Expression systems for the production of recombinant pharmaceuticals.' *BioDrugs*, 18(1) pp. 51-62.

Song, I. J., Ikram, M., Subhan, F., Choi, D. J., Lee, J. R., Kim, H. S., Lim, Y. T. and Yoon, S. (2015) 'Molecular characterization and expression analysis of mouse epidermal growth factor-like domain 8.' *Int J Mol Med*, 36(2) pp. 541-550.

Spessotto, P., Cervi, M., Mucignat, M. T., Mungiguerra, G., Sartoretto, I., Doliana, R. and Colombatti, A. (2003) 'beta 1 Integrin-dependent cell adhesion to EMILIN-1 is mediated by the gC1q domain.' *J Biol Chem*, 278(8) pp. 6160-6167.

Stathopoulos, P. B., Scholz, G. A., Hwang, Y. M., Rumfeldt, J. A., Lepock, J. R. and Meiering, E. M. (2004) 'Sonication of proteins causes formation of aggregates that resemble amyloid.' *Protein Sci*, 13(11) pp. 3017-3027.

Stoyanov, A. V., Rohlfing, C. L., Connolly, S., Roberts, M. L., Nauser, C. L. and Little, R. R. (2011) 'Use of cation exchange chromatography for human C-peptide isotope dilution - mass spectrometric assay.' *J Chromatogr A*, 1218(51) pp. 9244-9249.

Strutton, B., Jaffe, S. R. P., Pandhal, J. and Wright, P. C. (2018) 'Producing a glycosylating Escherichia coli cell factory: The placement of the bacterial oligosaccharyl transferase pgIB onto the genome.' *Biochem Biophys Res Commun*, 495(1) pp. 686-692.

Swartz, J. R. (2009) 'Universal cell-free protein synthesis.' *Nat Biotechnol*, 27(8) pp. 731-732.

Tang, R., Murray, C. W., Linde, I. L., Kramer, N. J., Lyu, Z., Tsai, M. K., Chen, L. C., Cai, H., et al. (2020) 'A versatile system to record cell-cell interactions.' *Elife*, 9, Oct, p. 61080.

Thomas, M. R. and Storey, R. F. (2015) 'The role of platelets in inflammation.' *Thromb Haemost*, 114(3) pp. 449-458.

Tong, S. Y., Davis, J. S., Eichenberger, E., Holland, T. L. and Fowler, V. G., Jr. (2015) 'Staphylococcus aureus infections: epidemiology, pathophysiology, clinical manifestations, and management.' *Clin Microbiol Rev*, 28(3) pp. 603-661.

Ukai, H., Ukai-Tadenuma, M., Ogiu, T. and Tsuji, H. (2002) 'A new technique to prevent self-ligation of DNA.' *J Biotechnol*, 97(3) pp. 233-242.

Uniprot (2023) *Uniprot*. Available at: <https://www.uniprot.org> (Accessed: 17 December 2023).

Varadharajan, B. and Parani, M. (2021) 'DMSO and betaine significantly enhance the PCR amplification of ITS2 DNA barcodes from plants.' *Genome*, 64(3) pp. 165-171.

Varki, A., Cummings, R. D., Aebi, M., Packer, N. H., Seeberger, P. H., Esko, J. D., Stanley, P., Hart, G., et al. (2015) 'Symbol Nomenclature for Graphical Representations of Glycans.' *Glycobiology*, 25(12) pp. 1323-1324.

Viktorinova, I. and Wimmer, E. A. (2007) 'Comparative analysis of binary expression systems for directed gene expression in transgenic insects.' *Insect Biochem Mol Biol*, 37(3) pp. 246-254.

Wallis, S., Wolska, N., Englert, H., Posner, M., Upadhyay, A., Renne, T., Eggleston, I., Bagby, S., et al. (2022) 'A peptide from the staphylococcal protein Efb binds P-selectin and inhibits the interaction of platelets with leukocytes.' *J Thromb Haemost*, 20(3) pp. 729-741.

Wally, N., Schneider, M., Thannesberger, J., Kastner, M. T., Bakonyi, T., Indik, S., Rattei, T., Bedarf, J., et al. (2019) 'Plasmid DNA contaminant in molecular reagents.' *Sci Rep*, 9(1) p. 1652.

- Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M. and Barton, G. J. (2009) 'Jalview Version 2--a multiple sequence alignment editor and analysis workbench.' *Bioinformatics*, 25(9) pp. 1189-1191.
- Wilkins, M. R., Gasteiger, E., Bairoch, A., Sanchez, J. C., Williams, K. L., Appel, R. D. and Hochstrasser, D. F. (1999) 'Protein identification and analysis tools in the ExPASy server.' *Methods Mol Biol*, 112 pp. 531-552.
- Witt, N., Rodger, G., Vandesompele, J., Benes, V., Zumla, A., Rook, G. A. and Huggett, J. F. (2009) 'An assessment of air as a source of DNA contamination encountered when performing PCR.' *J Biomol Tech*, 20(5) pp. 236-240.
- Wu, D. Y., Ugozzoli, L., Pal, B. K., Qian, J. and Wallace, R. B. (1991) 'The effect of temperature and oligonucleotide primer length on the specificity and efficiency of amplification by the polymerase chain reaction.' *DNA Cell Biol*, 10(3) pp. 233-238.
- Xu, J., Sun, H., Huang, G., Liu, G., Li, Z., Yang, H., Jin, L., Cui, X., et al. (2019) 'A fixation method for the optimisation of western blotting.' *Sci Rep*, 9(1) p. 6649.
- Xu, J. M., Wu, Z. S., Zhao, K. J., Xi, Z. J., Wang, L. Y., Cheng, F., Xue, Y. P. and Zheng, Y. G. (2023) 'IPTG-induced high protein expression for whole-cell biosynthesis of L-phosphinothricin.' *Biotechnol J*, 18(9) p. e2300027.
- Yang, S. and Rothman, R. E. (2004) 'PCR-based diagnostics for infectious diseases: uses, limitations, and future applications in acute-care settings.' *Lancet Infect Dis*, 4(6) pp. 337-348.
- Yao, S., Hart, D. J. and An, Y. (2016) 'Recent advances in universal TA cloning methods for use in function studies.' *Protein Eng Des Sel*, 29(11) pp. 551-556.
- Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., Madden, T. (2012) 'Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction.' *BMC Bioinformatics*, 13, Jun, pp. 134.
- Zabelskii, D., Dmitrieva, N., Volkov, O., Shevchenko, V., Kovalev, K., Balandin, T., Soloviov, D., Astashkin, R., et al. (2021) 'Structure-based insights into evolution of rhodopsins.' *Commun Biol*, 4(1) p. 821.
- Zacchigna, L., Vecchione, C., Notte, A., Cordenonsi, M., Dupont, S., Maretto, S., Cifelli, G., Ferrari, A., et al. (2006) 'Emilin1 links TGF-beta maturation to blood pressure homeostasis.' *Cell*, 124(5) pp. 929-942.
- Zarkar, N., Nasiri Khalili, M. A., Khodadadi, S., Zeinoddini, M. and Ahmadpour, F. (2020) 'Expression and purification of soluble and functional fusion protein DAB(389) IL-2 into the E. coli strain Rosetta-gami (DE3).' *Biotechnol Appl Biochem*, 67(2) pp. 206-212.

Zhao, Y., Zhang, X., Yao, J., Jin, Z. and Liu, C. (2020) 'Expression patterns and the prognostic value of the EMILIN/Multimerin family members in low-grade glioma.' *PeerJ*, 8, Mar, p. e8696.

Zhuo, X. and Knox, B. E. (2022) 'Interaction of human CRX and NRL in live HEK293T cells measured using fluorescence resonance energy transfer (FRET).' *Sci Rep*, 12(1) p. 6937.

Acknowledgements

I owe a special thanks to my supervisor, Mareike, and my supervisory team, Chris, Rebecca, James, and Svetlana, for their guidance, expertise, and patience. Their mentorship made this project what it is today. Mareike, you have shaped the trajectory of my career path in many ways and for that I thank you for this amazing opportunity!

I'd like to thank my friends for unwavering support and encouragement throughout my degree. Whether that be words of advice, late night hangouts or general mood boosters. I love you all.

Specifically, I'd like to thank Lucia Gardener, my best friend in the world. University put me in the path of someone that aligns with me in the most unlikely ways. I would like to dedicate a very special thanks to my closest friend, Xiomara. Since I met you during this degree, I feel like this project holds a part of you with it. All those late nights where you had no idea what I was talking about (proteins and domains and such – as per quoted by you) but still nodded away as if you knew. Our times hold a monumental part in my heart, and I've never been so certain about the longevity of a friendship in my life. I would also like to thank Alicita for being my god sent angel throughout this master's degree. You're so inherently good and kind, a literal angel, your heart is truly made of gold. You hold my emotions to such a high regard and you're so patient with me and for that I really appreciate. I can't wait for you to be in my life for a very long time, listening to every story, every emotion and how you speak about me like I put the stars in the sky. To her parents, amo a tu hija.

I would like to thank my family for their much-needed support. To my eldest sibling, you're my biggest inspiration no matter how much you don't feel it in yourself, there's so much for you to live for. To Erion, you're the kindest boy I've ever had the pleasure of knowing. You could never hurt anyone intentionally and I'm so grateful to have you as my brother. To Vanessa and Henry, I love you both so much, you have so much growing to do and I'm excited to see what beautiful people you become. My father, Naim, for being so motivating and loving no matter what choices I make and my mother, Paula, for making sacrifices for me to give me the best no matter the circumstances, I hope you feel better soon.

Lastly, I would like to thank Manchester Metropolitan University and the Science and Engineering department for their care and input throughout this degree.

To all those who played a role, big or small in this project – thank you for being a part of this significant chapter of my life.