

The Structural and Cellular Biology of Endogenous TLR4 Activators in Inflammatory Disease



Jesse Coker

Linacre College, University of Oxford

The Structural Genomics Consortium/Centre for Medicines
Discovery and The Kennedy Institute of Rheumatology

Supervised by:

Dr. Brian Marsden, Dr. Kim Midwood, and Dr. Wyatt Yue

A thesis submitted for the degree of

DPhil Clinical Medicine

Trinity 2021

This thesis is dedicated to my wife and best friend, Julie Brown. I
couldn't have finished it without her—I love you Jules.

Acknowledgements

First and foremost, I want to thank my supervisors **Brian Marsden**, **Wyatt Yue**, and **Kim Midwood**. It was a unique privilege to have three different advisors who were passionate about and supportive of my project. Thank you, Brian, for asking the tough questions, keeping me focused, and providing the motivation I needed to push past discouraging results. Thank you, Wyatt, for allowing me to spend the last four years working in your lab and for creating a positive, collaborative, and team-first environment. Finally, thank you, Kim, for providing the strategic guidance I needed to navigate complicated biology, understand unfamiliar cellular experiments, and position my project in a thoughtful and useful way. All three of you are kind, supportive, and brilliant scientists and I will look back fondly on my time working for you. Finally, thanks to the **Nuffield Department of Medicine** for funding my DPhil and for providing a route for international students to study at Oxford.

I am grateful to the **Lloyd Ruddock Lab** for their hard work on CyDisCo and for their generosity in freely sharing the pMJS class vectors with me, to **Toshiyuki Shimizu** for shipping me TLR4 and MD-2 *Drosophila* expression constructs all the way from Japan, to **Expres²ion Biotechnologies** for supporting the development of FAS2FURIOUS, and to the **Veit Hornung Lab** for freely sharing CRISPaint on Addgene for others to use. Thanks to **Philip Hublitz**, who provided invaluable advice and multiple cups of coffee to help me design GlowPAINT. Finally, an extra special thank you to **Holger Kramer** who provided mass spectrometry services and advice to me—he went above and beyond to help make my proteomics work successful.

Next, I want to personally thank the friends and colleagues who stood by and encouraged me throughout my DPhil. **Jack Kelly**: you are quite simply the best, a wonderful scientist and an even better friend. Thanks

for all the laughs and for letting me score a few points on you in basketball. Ja Morant >> Jayson Tatum. **Henry Bailey**: you were the calm, level-headed, hard-working force I needed to temper my emotions and enjoy the little joys of life. Thanks for always listening, buying extra beer tickets, and for making those corn-hole boards, which were the foundation of the best times I had in Oxford. **Gustavo Arruda Bezzera**: you set an impressive standard for excellence and it was a privilege working with you. Your advice and guidance were imperative for my growth as a scientist—please keep sending me book recommendations and pictures of baby Max. **Fiona Sorrell**: you almost made me like kinases. You are a wonderful scientist and an absolute force at the bench—thanks for the great beer recommendations and the electric Christmas Party outfits. **Will Foster**: I am still convinced you finished your PhD years ago and just never told anyone, your lab skills are matched only by your kindness and grace. **Thomas McCorvie**: I consider you a professor, your knowledge and technical skill are immense and your kindness and advice was always appreciated. Keep practicing frisbee though. **Nick Fox**: you welcomed me into the lab and made the rest of my success possible. You're also a sensational golfer, I'll always aspire to match you. **Sabrina MacKinnon**: you taught me so many nitty-gritty techniques that made this thesis possible. Thanks for introducing me to off-beat music, and please let Drew know I think of you both whenever I see sambuca. **Jennifer Mutisya and Douglas Ferreira**: literally the best summer interns of all time. You didn't know it, but your arrival re-energized my whole project and gave me the boost I needed to press on through a very tough drought. You will both do amazing things. **Alejandra Fernandez-Cid**: what can I say, you bring an unbridled joy and an unflappable dedication to your work that nobody can match. I will always look back fondly on our Spike purification battles—I wouldn't have wanted to do that with anyone else (or maybe ever again, come to think of it).

Thank you to my **Mom** for always being there for me and pushing me to be my best, to my brother **Phillip** for always taking my mind off things, to my in-laws **Bob and Leslie Brown** for sharing their beautiful daughter with me and for supporting my decision to cross an ocean for this DPhil, and to my brother- and sister-in-law **Emily and**

Chris Fairbank for encouraging me and listening through many long-winded rants. Finally, I want to thank my **Dad**, Dr. Kenneth Coker, for setting an example of what it means to be a scientist, a husband, and a father. I think this thesis will make you proud.

Statement of Originality

I declare that all data, words, figures, and results included herein are my own work and have not been submitted for another award, degree, or diploma at any other university or learning institution. I certify that this thesis contains only original work. No other person's work has been included unless explicitly stated in the text.

Abstract

Toll-like receptors (TLRs) are critical to detecting the presence of dangerous infections, initiating a protective innate immune response, and triggering a long-lasting adaptive immune response. TLRs initiate defensive, pro-inflammatory signalling following direct binding to pathogen-associated molecular patterns (PAMPs), which are exogenous molecular fingerprints derived from infectious pathogens like bacteria, viruses, and fungi. In addition to this role, more recent evidence demonstrates that TLRs can also initiate pro-inflammatory responses upon detection of damage-associated molecular patterns (DAMPs), which are endogenous self molecules released at sites of tissue damage or stress. Under normal conditions, TLR activation via DAMPs contributes to tissue repair; however, during autoimmune diseases like rheumatoid arthritis, persistent DAMP accumulation creates a chronically pro-inflammatory extracellular milieu that drives disease progression. Despite the therapeutic importance of the DAMP:TLR axis, the molecular mechanism behind TLR-mediated detection of DAMPs remains unclear: while there are dozens of structures of TLRs bound to PAMPs, there are *no* structures of any TLRs bound to any DAMPs from any organism. In order to clarify these molecular details, herein I employ primarily biophysical techniques to probe a model TLR:DAMP interaction between TLR4 and a class of proteins known as the fibrinogen-related proteins (FRePs). After presenting novel expression and purification methodologies for both TLR4 and the FRePs, I demonstrate with an assortment of biophysical techniques that TLR4 does not directly bind to the FRePs. Next, using an unbiased proteomics workflow to probe the cell-surface FReP interactome, I provide evidence to suggest that integrins and lipoprotein receptors, not TLR4, may be the direct FReP receptors. Based on my data and the existing literature, I synthesize these results into a broader general hypothesis regarding TLR-mediated detection of DAMPs which

I call the “receptorome”. Taken together, this work presents an updated model for DAMP activation of innate immunity: rather than a direct TLR:DAMP binding event, I contest that context-specific receptor mega-complexes are required to indirectly communicate the presence of DAMPs to TLRs. This conclusion clarifies some conflicting reports in the DAMP literature and provides a route to understanding the fundamental molecular mechanism behind the important but mysterious activation of TLRs in the presence of endogenous DAMPs.

Contents

1	Introduction and Background	1
1.1	TLRs and Innate Immunity: A Classical Understanding	1
1.1.1	The History of TLRs	2
1.1.2	Classical PAMP-Activated TLR Signalling	5
1.1.3	Structural Biology of TLR:PAMP Interactions	9
1.2	TLRs and Endogenous DAMPs: Mysteries Remain	13
1.2.1	TLR4-Activating DAMPs: A Key Disconnect	14
1.2.2	S100A8/S100A9	23
1.2.3	Fibronectin	27
1.2.4	Small Leucine-Rich Proteoglycans	31
1.2.5	Summary of TLR:DAMP Literature	35
1.3	TNC and the FReP Family	36
1.3.1	Tenascin-C	36
1.3.2	The Other FRePs	42
1.3.3	FRePs as a Model DAMP	44
1.4	Motivation and Chapter Aims	45
2	Biophysical Characterization of FReP Family Members	48
2.1	Materials and Methods	49
2.1.1	FReP Library and Accession Numbers	49
2.1.2	Transformation of <i>E. coli</i>	50
2.1.3	Ligation Independent (LIC) Cloning	51
2.1.4	1 mL Small-Scale Test Expression in <i>E. coli</i>	52
2.1.5	50 mL Medium-Scale Test Expression of His6-Tagged Proteins	53
2.1.6	Purification Scheme 1 (MBP/SHuffle [®])	54
2.1.7	Purification Scheme 2 (His6/CyDisCo)	55
2.1.8	Crystallization Trials	56
2.1.9	Data Collection and Processing	56

2.1.10	TNC XChem Fragment Screen	57
2.1.11	Differential Scanning Fluorimetry	58
2.1.12	General Mass Spectrometry	59
2.1.13	Disulfide Mapping Using Intact Mass	59
2.1.14	Native MS	59
2.2	Screening Fusion Tags for Solubility	60
2.3	Testing Disulfide-Refolding <i>E. coli</i>	63
2.3.1	Theory of Oxidizing <i>E. coli</i> Strains	63
2.3.2	Evaluation of SHuffle [®] and CyDisCo Cells	64
2.4	Large-Scale FReP Purification Strategies	64
2.4.1	Purification Scheme 1: MBP-tag in SHuffle [®] Cells	65
2.4.2	Purification Scheme 2: His6-tag in CyDisCo Cells	67
2.5	Quality Control of Soluble FRePs	69
2.6	CyDisCo Expression of Truncated Constructs	71
2.7	Structure of TNC	76
2.8	Fragment Screening of TNC Using XChem	80
2.8.1	Motivation for Fragment Screen	80
2.8.2	Results of Fragment Screen	81
2.9	Stabilization of Full-Length ANGPTLs	84
2.9.1	Theory of Thermal Denaturation Assays	84
2.9.2	ANGPTLs Bind Strongly to Calcium	87
2.10	ANGPTL2 and ANGPTL6 Structures with Calcium	88
2.11	Structure of Non-Native FGL1 Dimer	92
2.12	Insights into the FReP Family	95
2.13	Summary and Future Directions	101
3	Expression and Purification of TLR4 via FAS2FURIOUS	104
3.1	Materials and Methods	104
3.1.1	Protein Accession Numbers	105
3.1.2	Small-Scale Eukaryotic Test Expression	105
3.1.3	Medium-Scale Eukaryotic Test Expression	106
3.1.4	Construction of TLR4-c071, TLR4-c072, and MD-2-c014	106
3.1.5	S2 Cells, General Handling	108
3.1.6	Transient Test Expression, ExpreS ² ion Method	108
3.1.7	Selection of Stable S2 Cell Lines	108
3.1.8	Large Scale Purification from Stable Cell Lines	109

3.1.9	Construction of FAS2FURIOUS Vectors	110
3.1.10	HT-Cloning into FAS2FURIOUS Vectors	110
3.1.11	Small-Scale Transient Expression, FAS2FURIOUS	111
3.2	TLR4 Production in Existing SGC Systems	111
3.2.1	TLR4 Ectodomain in Three Expression Systems	111
3.2.2	Membrane-Localized TLR4 in Insect Cells	112
3.3	Expression of TLR4 in <i>Drosophila</i> S2 Cells	114
3.3.1	Introduction to <i>Drosophila</i> S2 Expression Systems	114
3.3.2	Test Expression Results	116
3.4	Large-Scale Purification of TLR4 and MD-2	118
3.5	FAS2FURIOUS: HT Expression in S2	121
3.5.1	Design of F2F Vectors	122
3.5.2	Optimization of F2F Protocol	124
3.5.3	Piloting F2F With Challenging Targets	127
3.5.4	F2F Stable Cell Lines	128
3.6	Chapter 3 Summary	129
4	Characterization of Putative TLR4:FReP Protein-Protein Interac-	
	tion	131
4.1	Materials and Methods	132
4.1.1	TLR4 Binding ELISA	132
4.1.2	Analytical SEC	132
4.1.3	Native-PAGE	133
4.1.4	BLI	133
4.1.5	XL-SDS-PAGE	134
4.1.6	Purification of CD44 Link Domain	134
4.1.7	Purification of S100A8, S100A9, and Calprotectin	135
4.1.8	Purification of ANXA2	136
4.2	Denaturing ELISAs Show a Complex	137
4.2.1	Quality Control of New Reagents	137
4.2.2	Binding of Other FRePs	138
4.2.3	Limitations of TLR4 ELISA	139
4.3	Solution State Methods Show No Complex	141
4.3.1	Co-Secretion in S2	141
4.3.2	Analytical SEC	142
4.3.3	Native-PAGE	144

4.3.4	BLI	146
4.4	XL-Methods Rescue Transient Complex	151
4.5	“Obvious” Co-receptors: No Complex	155
4.5.1	MD-2	155
4.5.2	CD44	157
4.5.3	ANXA2 and LYZ	157
4.5.4	S100 Family Proteins	162
4.6	Chapter 4 Summary	164

5 Exploration of the FReP-Associated Receptorome Using XL-AP-MSMS 167

5.1	Materials and Methods	167
5.1.1	Purification of Biotinylated FRePs	168
5.1.2	Monocyte Stimulation Assay	169
5.1.3	Endotoxin Removal from Biotinylated FRePs	170
5.1.4	Isolation of Primary Human Monocytes via Plating	170
5.1.5	XL-AP-MSMS Pilot 1	171
5.1.6	XL-AP-MSMS General Method for Pilot 2 and 3	172
5.1.7	Silver Staining SDS-PAGE Gels	173
5.1.8	General XL-AP-MSMS Method for Experiment 1 and 2	173
5.1.9	MSMS Data Collection and Analysis	175
5.2	Designing a Method for XL-AP-MSMS	175
5.2.1	Surfaceome Proteomic Methods	175
5.2.2	Final Design of XL-AP-MSMS Method	179
5.3	Purification of Biotinylated FRePs	183
5.4	Pilot Experiments 1, 2 and 3	186
5.5	Full Experiment 1 with DTSSP	189
5.5.1	Experimental Setup and Data Analysis	190
5.5.2	Full Experiment 1 Key Results	192
5.6	Full Experiment 2 with DTBP and DTSSP	197
5.6.1	Improvements from Experiment 1	197
5.6.2	Experimental Setup	199
5.6.3	Bulk Analysis Validates Integrin Hypothesis	200
5.6.4	Time-Dependent Analysis does not Reveal TLR4	204
5.6.5	Speculating on TNC Co-Receptors and Adaptors	209
5.7	Chapter 5 Summary and Next Steps	213

6	CRISPR/Cas9 Functional Epitope Tagging of Endogenous Proteins for Structural Studies and Proteomics using GlowPAINT	216
6.1	Materials and Methods	217
6.1.1	Construction of AAVS1 Targeting Vector	217
6.1.2	Construction of pGlowPAINT GFP ¹¹ -Containing Universal Donors	218
6.1.3	Cloning of eSpCas9(1.1)-sgX Vectors	220
6.1.3.1	Generation of GFP ¹⁻¹⁰ Expressing Cell Lines	222
6.1.3.2	gDNA Isolation and PCR	223
6.1.4	Western Blotting	224
6.1.5	Mismatch Cleavage Assay	224
6.1.6	GlowPAINT Pilot with <i>GDE1</i> Targeting sgRNA	225
6.2	Background and Motivation	226
6.3	GlowPAINT: CRISPaint + Split-GFP	229
6.4	Generation of GFP ¹⁻¹⁰ Cell Lines	234
6.5	Adding GFP ¹¹ to CRISPAINT Universal Donors	238
6.6	Cloning sgRNAs into eSpCas9(1.1)	239
6.7	Checking sgRNA Cleavage Efficiency	243
6.8	Piloting the GlowPAINT Method with <i>GDE1</i>	245
6.9	Conclusions and Future Directions	247
7	The COVID-19 Interlude: Antigen Production for Serological Assays	251
7.1	Materials and Methods	252
7.1.1	Transient Expression and Purification of SARS-CoV-2 S Protein	252
7.1.2	Cloning S-c23	253
7.1.3	Generation of S-Expressing Monoclonal Cell Lines	254
7.2	Introduction to SARS-CoV-2 S and Serology Testing	255
7.3	Method for the Transient Mammalian Expression and Purification of S	258
7.4	Optimization of the Method	262
7.5	Generation and Evaluation of S-Expressing Monoclonal Cell Lines . .	267
7.6	Conclusions and Future Directions	272
8	Conclusion: A New (Old) Hypothesis	275
8.1	Receptorome from First Principles	275
8.2	History of a TLR4 Receptorome	277
8.3	A Unifying Hypothesis	278

A	Supplementary Data and Figures	280
A.1	TNC, ANGPTL2, and ANGPTL6 Complete X-Ray Data Collection and Refinement Statistics	280
A.2	FAS2FURIOUS Manuscript	280
A.3	Raw XL-AP-MSMS Data	280
A.4	List of Abbreviations	280
	Bibliography	281

List of Figures

1.1	TLR4 Signalling in Response to LPS	6
1.2	Structures of TLR:PAMP Complexes	11
1.3	FReP Family Phylogeny	43
2.1	Rescuing FReP Solubility in <i>E. coli</i>	62
2.2	CyDisCo Facilitates His6-FReP Expression	65
2.3	Purification Scheme 1 (MBP/SHuffle®)	66
2.4	Purification Scheme 2 (His6/CyDisCo)	68
2.5	Disulfide Mapping of TNC	70
2.6	Native MS of TNC	72
2.7	Design of Truncated FReP Constructs	74
2.8	Test Expression of Truncated FRePs	75
2.9	Crystallization of TNC-c042	77
2.10	Structure of TNC	78
2.11	Crystallographic Fragment Screen of TNC	82
2.12	TNC:Fragment Structure Depositions	85
2.13	DSF Analysis of FRePs	87
2.14	Crystallization of ANGPTL2 and ANGPTL6	89
2.15	Structures of ANGPTL2 and ANGPTL6	90
2.16	Truncated FGL1 Crystallization	93
2.17	Structure of FGL1 Dimer	94
2.18	FReP Family MSA	96
2.19	Updated FReP Family Phylogeny	98
2.20	FReP Family Structures	99
2.21	Comparison of TNC, ANGPT1, and FFGG	100
2.22	Comparison of TNC, ANGPT1, and FCN1	102
3.1	Attempted TLR4 Ectodomain Expression Strategies	113
3.2	Failure to Scale-Up Membrane-Bound TLR4	114

3.3	Rescuing TLR4 Expression in S2 Cells	117
3.4	TLR4 Purification from S2 Cells	119
3.5	F2F Workflow Overview	122
3.6	F2F Vector Development	125
3.7	F2F Pilot	126
3.8	F2F Stable Cell Lines	129
4.1	TLR4 Binding ELISA QC	138
4.2	FRePs in TLR4 Binding ELISA	140
4.3	Co-Expression and Analytical SEC	143
4.4	Native-PAGE	146
4.5	BLI	150
4.6	XL-SDS-PAGE 1	152
4.7	XL-SDS-PAGE 2	154
4.8	Putative Co-receptors 1	156
4.9	Putative Co-receptors 2	159
4.10	Putative Co-receptors 3	163
5.1	XL-AP-MSMS Method Overview	180
5.2	Purification and QC of Biotinylated FRePs	185
5.3	XL-AP-MSMS Method Development Pilot 2 and Pilot 3	187
5.4	XL-AP-MSMS Experiment 1 Overview	193
5.5	XL-AP-MSMS Experiment 1 Network Analysis	196
5.6	TNC Prey Proteins Across All XL-AP-MSMS Experiments	200
5.7	Top 70 Interactors in Experiment 2-DTBP	202
5.8	Time-Dependent Prey Proteins	206
5.9	Enriched Early and Late Communities	207
6.1	TRiC: A Case Study for Endogenous Tagging	227
6.2	GlowPAINT: Background and Overview	228
6.3	Generation of GFP ¹⁻¹⁰ Expressing Cell Lines	236
6.4	Adding GFP ¹¹ to pCRISPAIN Donors	240
6.5	Cloning Target-Specific sgRNAs	241
6.6	Checking sgRNA Efficiency	246
6.7	GlowPAINT Pilot with <i>GDE1</i>	248
7.1	The SARS-CoV-2 S Protein	257

7.2	Final S-c16 Purification Method	261
7.3	Optimization of S Expression and Purification	264
7.4	S-c17 and S-c19 Monoclonal Cell Lines	269
7.5	S-c23 Monoclonal Cell Line	271

Chapter 1

Introduction and Background

1.1 TLRs and Innate Immunity: A Classical Understanding

The first stage of protective immunity, coined the innate immune response, begins with the detection of an infectious threat by macrophages, monocytes, and dendritic cells. These activated myeloid cells subsequently secrete pro-inflammatory cytokines [particularly interleukin-1 (IL-1), tumor necrosis factor alpha (TNF α), and interleukin-6 (IL-6)] and interferons (IFNs), especially downstream of the master regulatory transcription factor, nuclear factor κ B (NF κ B). These pro-inflammatory effectors induce an immediate, local response in an effort to destroy the invading pathogen, limit potential infectious spread, and repair any damaged tissue. Furthermore, these cytokines and the activated myeloid cells themselves also serve to induce the second stage of the response, known as adaptive immunity, via co-stimulation of lymphocytes (for example via CD80, CD86, and MHCs) during up-regulated antigen presentation. The activation of adaptive immune lymphocytes, especially T- and B-cells, enables slower but more efficient pathogen clearance as well as long-term memory to prevent future invasion of the same pathogen. This classic, two-stage immune response involves direct communication between the innate and adaptive immune systems to ensure both immediate and lasting protection from diverse threats [1].

However, it is only in the last thirty years that innate immunity escaped the shadow of the adaptive immune system and became appreciated as a key driver of human health. According to O’Neal *et al.*, “Before the discovery of Toll-like receptors (TLRs), innate immunity was seen as a crude and unsophisticated part of the im-

immune system; its main purpose was considered by immunologists to be the initiation of the more sophisticated adaptive immune response, which was thought to confer protection on the infected organism” [2]. Prior to the launch of dedicated research programs on TLRs, the molecular basis for pathogen detection was unknown, and the activation, regulation, and signalling of key pro-inflammatory molecules remained mysterious. In fact, despite much historical work on cytokines and their downstream effects, prior to the mapping of TLRs the actual molecular mechanism which activates the innate immune symphony remained unknown—there were cytokines, surely, but how were they induced? Furthermore, the molecular mechanism responsible for innate immune cross-talk with the adaptive immune system, a key step in protective immunity, was also unclear. In many ways, then, the modern understanding of the immune response began with TLRs, and this history provides context and insight for the next phase of study on these critical multi-functional receptors.

1.1.1 The History of TLRs

In 1989, Janeway predicted the existence of innate pattern recognition receptors (PRRs) which could both detect the presence of pathogenic microbes and also initiate a signalling cascade that triggers an adaptive immune response via antigen presentation; such a receptor would provide the missing molecular link between innate and adaptive immunity [3]. The quest to identify these elusive PRRs began with the human IL-1 receptor type I (IL-1R1) and the developmental *D. melanogaster* protein Toll, which were shown to contain a homologous C-terminal domain that drives activation of the NF- κ B family of transcription factors [4, 5, 6]. In 1992, the agonism of NF- κ B by IL-1R1 and Toll was mapped to the cytoplasmic C-terminal Toll-IL-1-resistance (TIR) domain [7], and in 1994 work on the homologous “N” protein from plants showed that the TIR domain was critical to host defense across both plants and animals [8]. In 1993, a constitutively active mutant of *Drosophila* Toll was shown to activate the NF- κ B transcription factor Dif [9], and it was shown in 1996 that Toll-mediated activation of Dif induced the release of antimicrobial peptides in flies and protected flies from lethal fungal infection [10, 11]. Hoffman shared the 2011 Nobel Prize for this work, which conclusively demonstrated that activated Toll induced an innate immune response. With this link between Toll and innate immunity established in *D. melanogaster*, work commenced on the other Toll homologs in humans (outside of the known IL-1R1) which were identified in sequence databases as early as 1994 [12]. In 1997, Medzhitov and Janeway cloned

human Toll (hToll), today known as TLR4, and showed that a constitutively active TLR4 construct activated NF- κ B, promoted the release of pro-inflammatory cytokines, and induced expression of the co-stimulatory molecule CD80 [13]. This landmark work not only showed that TLR4 could activate a pro-inflammatory innate immune response in humans, it also linked innate immune activation to the CD80-mediated stimulation of T-cell activation. Therefore, TLR4 was the first *bona fide* human PRR to be identified, just eight years after Janeway postulated that such a molecule could directly link innate and adaptive immunity.

However, at this point, the link between TLR4 and bacterial cell-wall derived lipopolysaccharide (LPS) was unknown. Rather than stimulating with a ligand, both Hoffman and Medzhitov employed constitutively active Toll constructs in their work—the actual TLR4 agonist therefore remained a mystery. The search for the LPS receptor began all the way back in 1954, when the lipid A moiety of LPS was shown to activate inflammation [14]. LPS-activated inflammation was proven to be mediated by NF κ B in 1986 and mitogen-activated protein kinase (MAPK) in 1993 [15, 16]. Additionally, in the early sixties, the LPS-resistant C3H/HeJ mouse strain was isolated, and in 1978 none other than James Watson mapped the responsible gene, *Lps^d*, to the fourth mouse chromosome. Lipopolysaccharide-binding protein (LBP) was identified in 1986 [17], and subsequent work in 1990 showed that LBP and the GPI-anchored membrane protein CD14 worked together to sensitize immune cells to LPS [18, 19]. Despite the lack of an intracellular signalling domain in CD14, these results led many to hypothesize that CD14 was the sought after LPS receptor. However, in 1998, Yang *et al.* showed that a TLR was required for transducing the LPS signal across the plasma membrane— LPS detection was enhanced by CD14 but in fact depended on a TLR [20]. However, because of contaminating lipoprotein (now known to be a high affinity TLR2 agonist) in their LPS preparations, Yang *et al.* misidentified TLR2 as the LPS receptor. Later in 1998, after positional cloning of murine *Lps^d*, the Beutler group conclusively showed that *Tlr4*, not *Tlr2*, was the responsible gene, a result for which Beutler was included in the 2011 Nobel Prize [21]. This finding was reproduced a year later by the Malo group [22]. In 1999, the Akira group showed that immune cells derived from *Tlr4* knockout mice did not respond to LPS, providing protein-level proof that TLR4 was the LPS receptor [23]. Finally, after the discovery that TLR4 did not by itself confer LPS responsiveness, it was shown in 1999 that LPS-binding myeloid-differentiation factor 2 (MD-2) was required as a co-receptor and that the MD-2:LPS complex was the complete TLR4 agonist [24, 25]. Therefore, after more than 45 years, two research avenues (one looking for PRRs, one looking for the LPS receptor) converged in the late nineties

onto the same protein: TLR4. TLR4 was the first identified PRR, the long sought after LPS receptor, and the first TLR with a known agonist.

The mapping of the CD14:LPS:MD-2:TLR4 axis was a seminal moment that sparked renewed interest in the TLR family and, more broadly, the previously underappreciated innate immune system. A flurry of papers in the early 2000's de-orphanized many of the other TLRs and established the entire family as critical PRRs involved in defense against bacterial and viral infections. These TLR-activating molecules, derived from exogenous infectious sources like bacteria and viruses, are broadly known as pathogen-associated molecular patterns (PAMPs). There are ten TLRs in the human genome (TLR1–TLR10) and twelve in mice (TLR1–TLR9 + TLR11, TLR12, and TLR13; TLR10 is present in the mouse genome but is not functional due to a retrovirus insertion) [26]. In 1999, the second TLR to have its ligand identified was TLR2. TLR2 binds to di- or tri-acylated lipopeptides, a key component of gram-positive bacterial cell walls, via a heterodimer with either TLR1 (triacyl) or TLR6 (diacyl) [27, 28, 29]. Experimentally, TLR2 is often agonized with the synthetic tri-acyl lipopeptide N-palmitoyl-S-dipalmitoylglyceryl Cys-Ser-(Lys)₄ (Pam₃CSK₄). The activation mechanism of TLR2 provided the first evidence that the specificity of TLRs could be modified by heterodimerization. TLR10, which shares approximately 50% sequence identity with and is located within the same chromosomal cluster as TLR1 and TLR6, can also heterodimerize with TLR2 to recognize lipopeptides and other TLR2-binding PAMPs [30, 31]. However, TLR10 seems to exert an anti-inflammatory effect on TLR2 rather than cooperating to activate downstream signalling [32]. Despite a recent report that TLR10 can recognize the HIV-derived structural protein gp41 [33], TLR10 remains the only human TLR without a clear agonist, and for now should be functionally grouped alongside TLR1 and TLR6 as a TLR2 partner. In 2001, Hayashi *et al.* demonstrated that TLR5 recognizes flagellin, the main component of both gram-negative and gram-positive bacterial flagella [34]. In 2003, Zhang *et al.* showed that murine TLR11 prevents infection by uropathogenic bacteria [35], and, as a heterodimer with TLR12, TLR11 can also bind to the protozoan protein profilin and activate an anti-parasitic immune response [36, 37, 38]. Finally, murine TLR13 acts as the receptor for bacterial RNA [39]. Therefore, based on their canonical PAMP-binding abilities, the sub-family of TLRs consisting of TLR1, TLR2, TLR4, TLR5, TLR6, TLR11, TLR12, TLR13, and perhaps TLR10 can be considered the protein- and lipid-sensing TLRs.

The remaining TLRs bind to nucleic acids. In 2000, Hemmi *et al.* showed that bacterial, unmethylated cytidine-phosphate-guanosine (CpG) oligodeoxynucleotides

(ODNs), which are uncommon and often hypermethylated in the mammalian genome, bind to and activate TLR9 [40]. TLR9’s ability to distinguish between host and viral DNA is critical to the host response to herpes simple virus [41, 42], and synthetic CpG ODNs are often used to stimulate TLR9 as a vaccine adjuvant [43]. In 2001, Alexopoulou *et al.* demonstrated that dsRNA, a key intermediate produced by most RNA viruses during their replication cycle, binds to and activates TLR3 [44]. Experimentally, TLR3 is often agonized by the synthetic dsRNA molecule polyinosine–polycytidylic acid [poly(I:C)]. Subsequently, TLR7 was shown to bind to the antiviral small-molecule imiquimod, an imidazoquinoline-containing nucleotide mimetic with a previously unknown mechanism of action [45]. Consistent with its ability to bind the nucleotide analog imiquimod, three groups subsequently reported in 2004 that TLR7 is activated by ssRNA derived from ssRNA-containing viruses like influenza [46, 47, 48]. TLR8, the closest phylogenetic neighbor to TLR7, can also recognize viral ssRNA and trigger pro-inflammatory responses; however, somewhat analogous to TLR10, TLR8 can in some contexts also downregulate TLR7 and TLR9 signalling [49]. In conclusion, based on their canonical PAMP-binding abilities, TLR3, TLR7, TLR8, and TLR9 can be grouped as the nucleic-acid-sensing TLRs.

1.1.2 Classical PAMP-Activated TLR Signalling

PAMP binding to TLRs is only the beginning of a complex downstream signalling network that was elucidated over the fifteen years following the establishment of TLR4 as the *bona fide* PRR for LPS. Broadly speaking, there are two distinct TLR-activated signalling pathways. The first, known as the “MyD88-dependent” pathway (*aka* myeloid differentiation primary-response protein 88 dependent), is triggered at the cell surface predominantly by the protein- and lipid-sensing TLRs and leads to the release of pro-inflammatory cytokines via $\text{NF}\kappa\text{B}$. The second, known as the “MyD88-independent” pathway, is triggered from endosomes predominantly by the nucleic-acid-sensing TLRs and leads primarily to the induction of Type I IFNs via the interferon regulatory factor (IRF) family of transcription factors. Uniquely, TLR4 is able to signal via the MyD88-dependent pathway at the plasma membrane and, following endocytosis, the MyD88-independent pathway from endosomes. Because it has the most diverse downstream signalling network and is most important to my work, in this section I focus on TLR4-activated signalling in response to LPS, with an emphasis on the accessory proteins and co-receptors required for signal transduction. I also provide a mechanistic overview of the pro-inflammatory cytokines and anti-viral interferons released upon TLR4 ligation. For a broader overview, the

entirety of the mammalian TLR signalling landscape has been extensively reviewed elsewhere [2, 26, 50, 51, 52, 53]. As a reference throughout this section, Figure 1.1 provides a simplified schematic of TLR4 signalling.

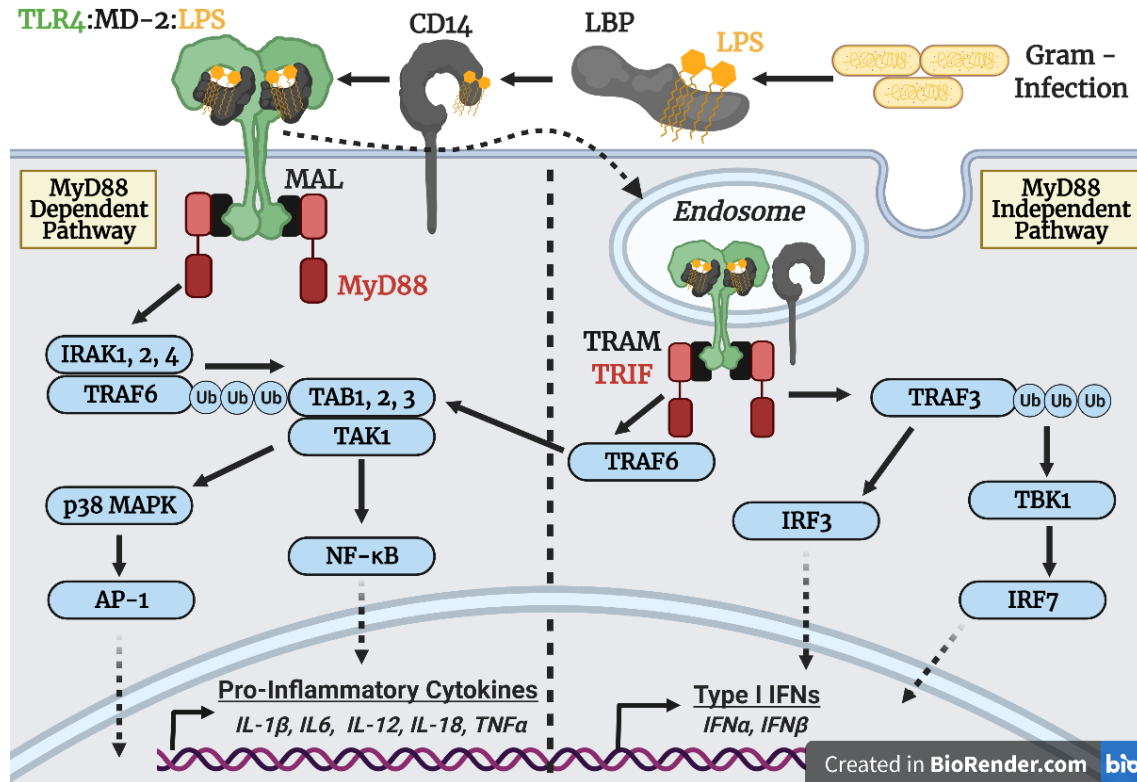


Figure 1.1 TLR4 Signalling in Response to LPS. Simplified signal transduction network involved in the activation of inflammation upon binding of LPS to TLR4.

LPS typically contains two core features: the hydrophobic “Lipid A”, a diglucosamine connected to six acyl chains, and a variable oligosaccharide region that can contain the “O-antigen” side chain [54]. Lipid A is the key recognition motif in LPS; however, LPS must first be processed from bacterial outer membranes prior to presentation to TLR4. First, LBP, which binds to LPS with nanomolar affinity, extracts LPS from gram-negative cell membranes and then acts as a shuttle protein to present LPS to CD14 [17, 55]. Second, CD14 monomerizes LPS and, finally, presents it to a heterodimer consisting of TLR4 and its obligate co-receptor MD-2 [18]. Upon LPS binding, CD14 physically associates with TLR4:MD-2 in a three-way complex that becomes localized to lipid raft microdomains [56]; indeed, CD14 is required for proper endocytosis of the receptor following LPS binding [57], and CD14 remains physically associated with TLR4:MD-2 throughout internalization [58]. While LBP and CD14 combine to greatly increase the sensitivity of cells to LPS, innate immune

activation fundamentally depends on TLR4:MD-2, as knockout of either TLR4 or MD-2 completely abrogates the response to TLR4 in mice [59, 23]. TLR4 itself does not bind to LPS; rather, LPS binds within a deep hydrophobic cup within MD-2 (see Section 1.1.3) [24], and it is the TLR4:MD-2 complex which has the highest affinity for LPS [60]. TLR4:MD-2 exists as a heterodimer at the cell surface [61], and upon LPS-binding the TLR4:MD-2:LPS ternary complex forms an activated, M-shaped, 2:2:2 heterohexamer (a dimer of trimers). The formation of this “TLR4 dimer” brings together the cytosolic TIR domains, which form a symmetrical homodimer via an extensive interface mediated by “BB” and “DD” loops [62, 63], and activates both intracellular signalling cascades [64].

The immediate cytosolic adaptor proteins for TLRs, of which there are five, act as scaffolding proteins via homotypic TIR:TIR interactions [65]. At the cell surface, MyD88 adaptor-like protein (MAL), also known as TIR-domain containing adaptor protein (TIRAP), binds directly to the intracellular TIR dimer formed upon TLR1/2, TLR2/6, or TLR4 dimerization and acts as a bridging adaptor to the central TIR-domain containing regulatory protein, MyD88 [66, 67]. All human IL-1Rs and TLRs, with the exception of TLR3, can use MyD88 as an adaptor; MyD88 can bind directly to the TIR domains of TLR7, TLR8, and TLR9 without requiring the bridging protein MAL [68]. In addition to its C-terminal TIR domain, MyD88 contains an N-terminal “death domain” (DD), which recruits the IL-1R-associated kinases 1 (IRAK1), IRAK2, and IRAK4 via homotypic interactions with the DD domains in the IRAKs [69]. IRAK4, which can also directly activate IRAK1 and IRAK2 via phosphorylation, appears to be the most critical downstream kinase, as IRAK4 KO macrophages and mice are broadly resistant to LPS-induced pro-inflammatory signalling [70]. The megacomplex consisting of MyD88, adaptor proteins, and the IRAKs is a master innate immune regulator known as the “Myddosome” that is key to integrating the TLR signal [52, 71]. The activated IRAKs subsequently interact with and activate the RING E3 ubiquitin ligase TNF receptor associated factor 6 (TRAF6), which becomes polyubiquitinated and recruits a platform consisting of TGF β activated kinase 1 binding protein 1 (TAB1), TAB2, and TAB3 [72, 73]. Finally, the ubiquitin-dependent kinase TGF β activated kinase 1 (TAK1, also known as MAP3K7) is recruited to the TRAF6:TAB1:TAB2:TAB3 platform, activated, and then triggers the rest of the downstream signalling network [74]. TAK1 activation leads to the degradation of the NF- κ B inhibitor I κ B, which leads to translocation of the active transcription factor NF- κ B (also known as p50/p56) into the nucleus and the subsequent activation of pro-inflammatory genes [75]. Simultaneously, through an orthogonal phosphorylation cascade, activated TAK1 activates the kinases JNK,

p38 MAPK, and ERK1/2, which broadly activate the activating-protein 1 (AP-1) family of transcription factors, including Jun and Fos [76]. The concomitant activation of NF- κ B and AP-1 induces a broad inflammatory response and leads to the release of the pro-inflammatory cytokines IL-1, IL-6, IL-8, IL-12, IL-18, and TNF- α [68]. These processes also “prime” the inflammasome by increasing expression of NOD-, LRR- and pyrin domain-containing protein 3 (NLRP3); following activation by a second stimuli, the NLRP3 inflammasome is fundamental to the processing of IL-1 β and IL-18 [77, 78]. In conclusion, the “MyD88-dependent”, cell surface TLR4 signalling pathway relies on the adaptors MAL and MyD88 and leads to the broad induction of pro-inflammatory signalling via NF- κ B and AP-1.

The MyD88-independent pathway is triggered from endosomes and depends on TIR-containing adaptor-inducing IFN (TRIF) and TIR related adaptor molecule (TRAM) [79, 80]. The MyD88-independent pathway is activated downstream of the resident endosomal nucleic-acid-sensing receptors TLR3, TLR7, TLR8, TLR9, and, following endocytosis from the cell surface and vesicular trafficking, TLR4. In the context of TLR4 signalling, TRAM acts analogously to MAL. TRAM binds to the TIR domain of TLR4 and acts as a bridge to bind TRIF which, as the master upstream regulator of this pathway, recruits the E3 ubiquitin ligase TRAF3. TRAF3 activates the IRF transcription factor IRF7 and, via phosphorylation of the regulatory serine/threonine kinase TANK-binding kinase 1 (TBK1), IRF3 [81, 82, 2]. The resulting activation of IRF3 and IRF7 leads to the induction of antiviral Type I interferons, IFN α and IFN β , as well as the chemokines CXCL9, CXCL10, and CCL5 [83]. The fifth TIR-containing adaptor protein, sterile-alpha and armadillo-motif containing protein (SARM), is anti-inflammatory and downregulates the MyD88-independent pathway by competing with downstream proteins for binding to TRIF [84]. Importantly, the MyD88-independent pathway can also activate NF- κ B and AP-1 (via cross-talk with TRAF6), although less strongly and with slower kinetics than the MyD88-dependent pathway; this explains why LPS can trigger slow NF- κ B activation in MyD88 KO models and why the strictly endosomal TLRs can still activate NF- κ B [79, 2]. Somewhat confusingly, endosomal TLR7, TLR8, and TLR9 use MyD88 as an adaptor to activate the MyD88-independent pathway [2, 85], so it is easiest to categorize the terms “MyD88-dependent” and “MyD88-independent” in the context of TLR4 signalling. However, the most important point to understand is that, from the endosomal compartment, TLR4 can activate IRF transcription factors via the bridging adaptor TRIF and induce the expression of Type I IFNs.

Via these two complementary pathways, LPS binding to TLR4 triggers the expression of pro-inflammatory cytokines and anti-viral interferons which have diverse immunomodulatory function. First, the pro-inflammatory cytokines IL-1 (IL-1 α and IL-1 β) and IL-18 belong to the IL-1 super-family and have multiple effects on immune functions, including the stimulation of NF- κ B and MAPK in adjacent, un-activated immune cells, especially macrophages [86, 87]. IL-1 leads to the increased expression of adhesion molecules to promote extravastion (the infiltration of immune cells from the circulatory system into the infected tissue), and it potently induces IL-6 secretion. IL-1 drives the expression of genes which contribute to fever, vasodilation, and a lower pain threshold, and it leads to the proliferation, survival, activation, and differentiation of T cells towards a T helper (T_H) phenotype. IL-18 also modifies the adaptive immune response and can induce both T_H1 and T_H2 T-cell responses. Second, IL-6 contributes to host defense primarily by activating the systemic secretion of defensive acute phase proteins by the liver, which broadly control the early response to infection [88]. IL-6 is also required for T_H17 differentiation from naive CD4+ T-cells, cytotoxic T-cell differentiation from CD8+ precursors, and the activation of B-cells to promote an antibody response [89]. Third, TNF α , a key member of the TNF super-family, promotes strong local inflammation via activation of endothelial and immune cells and, if released systemically, induces lethal septic shock and the associated “cytokine storm” [90]. TNF α is the master inducer of over a hundred genes (including many other cytokines) via NF κ B and other pathways, causes tissue degeneration, and leads to the recruitment and survival of immune cells [91]. Because of its strong and diverse pro-inflammatory properties, especially at sites of chronic inflammation, TNF α is a major therapeutic target in autoimmune diseases and cancer [92]. Finally, via the activation of many interferon-stimulated genes (ISGs), IFN α and IFN β regulate host antiviral responses and are critical to the clearance of HIV, influenza, and other viruses [93]. Critically, IFN α and IFN β promote antigen presentation by dendritic cells (DCs), regulate CD4+ and CD8+ T-cell responses, and enhance B-cell activation and antibody responses [94]. In summary, TLR4 activation eventually changes the expression pattern of hundreds of genes, leading to both a diverse and systemic pro-inflammatory innate immune response and the induction of a protective adaptive immune response.

1.1.3 Structural Biology of TLR:PAMP Interactions

The unique structural fold of the TLRs facilitates the detection of diverse ligands. The ectodomains of TLRs contain 20–28 tandem leucine-rich repeats (LRRs), which

are 20–30 amino acids in length, contain the conserved sequence LxxLxLxxN, and form a horseshoe-like solenoid [95]. There are thousands of LRR-containing proteins in the broader super-family, from viruses on up to eukaryotes, and more than 60 LRR-containing structures in the PDB [96]. The LRR architecture, defined by a large, stable, parallel β -sheet on the concave face and variable structural motifs on the convex face of the horseshoe, is uniquely tuned for protein-protein interactions and ligand binding [97]. Most LRRs interact with other proteins via their concave face, which can wrap around interactors to maximize contact surface area [96]; atypically, all of the TLRs primarily use their hypervariable convex regions for recognition of PAMPs, diversifying the PAMP landscape and explaining TLR specificity [85]. The TLRs also contain the atypical N-terminal (LRRNT) and C-terminal (LRRCT) “capping” modules, which do not conform to the standard LRR sequence motif, often contain stabilizing disulfide bonds, and seem to protect the core of the TLR fold from the extracellular environment. PAMP-binding results in TLR hetero- or homodimerization, a step that is strictly required to bring the intracellular TIR domains into proximity for initiation of the downstream signalling cascades reviewed in Section 1.1.2. My goal in this section is not to review the structure of TLRs in detail, which has been extensively done elsewhere [98], but rather to emphasize that all known PAMP:TLR interactions are direct, high-affinity, and therefore amenable to conventional structural studies. While there may be accessory proteins involved in delivering PAMPs to TLRs and fine-tuning the downstream signalling, there exist definitive, stable complexes between TLRs and PAMPs.

The structure of mouse TLR4:MD-2 bound with a synthetic antagonist was solved in 2007 with crystallography by Kim *et al.* [99], and the structure of activated mouse TLR4:MD-2:LPS was solved in 2012 by Ohto *et al.* [100]. Human TLR4:MD-2:LPS was first solved in 2009 by Park *et al.* [64], human TLR4:MD-2:Lipid IVa (a tetra-acylated antagonist of human TLR4:MD-2) was solved in 2012 [100], and the hyporesponsive TLR4 D299G + T399I double mutant was solved in complex with MD-2 and LPS in 2012 [101]. The structure of human TLR4:MD-2:LPS (PDB: 3FXI) is shown in Figure 1.2A and demonstrates the m-shaped TLR4 dimer formed upon LPS binding. Note that the acyl chains in LPS bind in the hydrophobic β -cup of MD-2 while the exposed, phosphorylated glucosamine backbone of LPS protrudes from the pocket and interacts with the convex face of the adjacent TLR4 molecule to drive dimerization. The sixth acyl chain in LPS also partially points outside of the MD-2 pocket and interacts with the adjacent TLR4 via two conserved salt bridges (Lys388 and Gln436) to enhance dimerization. Both Ohto *et al.* and Kim *et al.* show that the addition of LPS to recombinant mouse or human TLR4:MD-2 leads

to the formation of a stable 2:2:2 activated complex that remains associated during size-exclusion chromatography (SEC). Using Native-PAGE, Ohto *et al.* estimated the K_D of LPS for TLR4:MD-2 to be approximately 200 pM [101], which is broadly consistent with other reports using orthogonal methods that report an affinity in the single-digit nanomolar range [102, 60]. Without question, TLR4:MD-2 forms a stable, high-affinity, direct, and structurally tractable complex with its cognate PAMP, LPS.

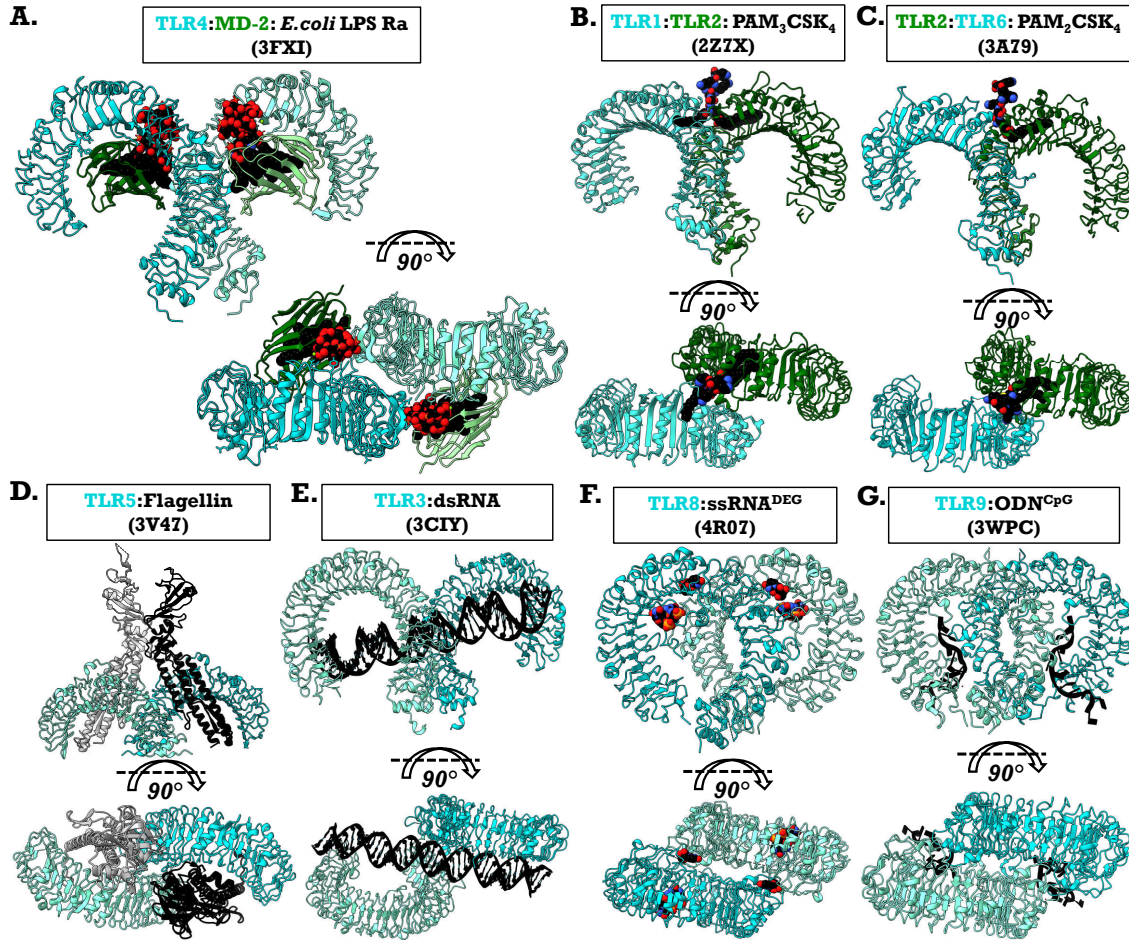


Figure 1.2 Structures of TLR:PAMP Complexes. **A.**The structure of human TLR4:MD-2:LPS (PDB: 3FXI). **B.**The structure of human TLR1:TLR2:PAM₃CSK₄ (PDB: 2Z7X). **C.**The structure of mouse TLR2:TLR6:PAM₂CSK₄ (PDB: 3A79). **D.**The structure of zebrafish TLR5 bound to *S. enterica* flagellin (PDB: 3V47). **E.**The structure of mouse TLR3 bound to dsRNA (PBD: 3CIY). **F.**The structure of human TLR8 bound to degradation products from ssRNA (uridine and a di-nucleotide) (PDB: 4R07). **G.**The structure of horse TLR9 bound to a CpG-containing ODN 12-mer (PDB: 3WPC). TLRs depicted in blue/turquoise, PAMPs depicted in black/grey; PAMPs colored by heteroatom, red oxygens, blue nitrogens).

This pattern holds true for many other TLRs. The structure of human TLR1:TLR2:PAM₃CSK₄ (PDB: 2Z7X) was solved in 2007 by Jin *et al.* and is shown in Figure 1.2B, while the structure of TLR1:TLR2:Diprovocim (a high affinity agonist that binds in the same pocket) was solved by Su *et al.* in 2019 [103]. PAM₃CSK₄ binds exclusively to the convex face of TLR1 and TLR2, with two acyl chains penetrating into a deep hydrophobic pocket in TLR2 and the third acyl chain driving dimerization via binding to a shallower hydrophobic channel in the adjacent TLR1. Both Jin *et al.* and Su *et al.* report that the TLR1:TLR2:PAM₃CSK₄ complex remains associated during SEC; this is consistent with the single-digit nanomolar EC₅₀ of PAM₃CSK₄ and the sub-nanomolar EC₅₀ of Diprovocim in cell stimulation assays [104]. The structure of mouse TLR2:TLR6:PAM₂CSK₄ (PDB: 3A79) was solved in 2009 by Kang *et al.* and is presented in Figure 1.2C. Analogously to PAM₃CSK₄, the two acyl chains in PAM₂CSK₄ bind in the hydrophobic pocket on TLR2; however, in the absence of a third acyl chain, a key hydrogen bond between TLR6 Phe319 and the peptide backbone in the PAM₂CSK₄ head-group drives heterodimerization. Kang *et al.* show that the TLR2:TLR6:PAM₂CSK₄ complex is stable during SEC, consistent with an EC₅₀ of 8 nM in a cell stimulation assay [105]. The structure of zebrafish TLR5 bound to *S. enterica* flagellin was solved in 2012 by Yoon *et al.* and is presented in Figure 1.2D [106]. At the primary binding site, flagellin forms a V-shaped interaction with the sides of both TLR5 subunits; homodimerization at the secondary interface is driven by interactions between flagellin and the convex face of the adjacent TLR5. Yoon *et al.* report that flagellin triggers the formation of a stable, 2:2 homodimer during SEC, consistent with a cellular EC₅₀ of 77 pM for the related flagellin analog CBLB502 in a human TLR5 NF- κ B reporter assay [106]. When reconstituted into liposomes, membrane-bound human TLR5 has also been used in a flagellin detection assay, with TLR5 displaying a direct, surface-plasmon-resonance (SPR)-derived binding affinity of 14 nM for *S. typhimurium* flagellin and 841 nM for *E. coli* flagellin [107]. The structure of mouse TLR3 bound to dsRNA was solved in 2008 by Liu *et al.* (PDB: 3CIY) and is presented in Figure 1.2E [108]. Uniquely, a single piece of dsRNA binds in-between two TLR3 molecules (2:1 stoichiometry) and forms two distal interfaces with each TLR3 subunit, one at the N-terminus with LRR3 and one at the C-terminus with LRR19–LRR21. TLR3 homodimerization is driven by a cooperative protein-protein interface that forms between the TLR3 LRRCT domains that are brought into close proximity by dsRNA ligation. Liu *et al.* report that the mouse TLR3:dsRNA activated complex remains stably associated during SEC, a result which has been reproduced with human TLR3

[109]. Recombinant human TLR3 binds to dsRNA with an affinity of 5–2250 nM depending on the dsRNA length and the prevailing pH [109], and under optimal conditions the standard TLR3 agonist poly(I:C) binds with a K_D of approximately 10 nM [110]. The structure of human TLR8 in complex with two degradation products of ssRNA (uridine mononucleoside and a di-nucleotide) was solved in 2015 by Ohto *et al.* (PDB: 4R07) and is presented in Figure 1.2F [111]. Binding of the ssRNA degradation products occurs at two sites: one containing uridine mononucleoside at the convex face of LRR11–14 which drives homodimerization, the other containing a di-nucleotide bound at the concave face of LRR10–13. Binding of short RNA oligos at the second site increases the affinity of TLR8 for uridine, with isothermal titration calorimetry (ITC) demonstrating an increase in affinity from 55 μ M to 1 μ M when an oligo is added. Tanji *et al.* also report an ITC-derived affinity of 200 nM for the interaction between TLR8 and the nucleotide analog resiquimod, and further demonstrate using small-angle x-ray scattering (SAXS) that resiquimod induces dimerization of TLR8 in solution [111]. Finally, the structure of horse TLR9 bound to an agonistic ODN 12-mer containing a CpG motif was solved in 2015 by Tanji *et al.* (PDB: 3WPC) and is presented in Figure 1.2G [112]. The CpG motif in the 12-mer is recognized by the N-terminal, concave face of TLR9 (primarily at LRRNT, LRR1, and LRR2) via direct interactions with amino acids or with water-mediated hydrogen bonds, while homodimerization is mediated by an interaction between the convex face of the adjacent TLR9 and the rest of the 12-mer. Ohto *et al.* show that the activated TLR9:ODN complex remains associated during SEC, and, using ITC, they report a K_D of 20 nM for this interaction [112]. In summary, in every reported case, TLRs bind directly to exogenous PAMPs and form stable, high affinity (i.e. nanomolar or below), dimeric complexes which survive SEC and are amenable to structural characterization with X-ray crystallography.

1.2 TLRs and Endogenous DAMPs: Mysteries Remain

Section 1.1 discussed the rise of TLRs and innate immunity during the 90’s and early 2000’s, with cellular and structural biology combining to map the TLR:PAMP landscape, deconvolute the intricate downstream signalling network following PAMP detection, and establish direct, high-affinity TLR:PAMP complexes as essential drivers of host defense. In line with the canonical understanding of TLRs as PRRs, all of

the work reviewed until this point focused on *exogenous* PAMPs derived from bacteria or viruses. However, a “canary in the coal mine” report from von Bernuth *et al.* in 2008 suggested that there remains much to be learned about the TLRs. In nine children with inherited MyD88 deficiency, von Bernuth *et al.* found that MyD88, and therefore the majority of TLRs and IL-1Rs which require this ubiquitous adaptor protein, was only required for protection from a small subset of pyogenic (i.e. pus forming) bacterial infections like pneumococcus [113]. Despite a permanent lack of MyD88, which causes at least 35 experimental infections in mice, the children became healthier over time (likely due to the compensatory effects of the adaptive immune system) and were immune to most other non-pyogenic microbes. These data challenged the dogma that TLRs were absolutely indispensable for host defense; rather, von Bernuth *et al.* argued, “The MyD88-dependent TLRs and IL-1Rs are therefore essential for protective immunity to a small number of pyogenic bacteria, but redundant for host defense to most natural infections” [113]. This result is supported by subsequent work on the three other non-TLR super-families of PRRs which contribute to PRR redundancy: the nucleotide-binding oligomerization domain LRR-containing receptors (NLRs), the retinoic acid-inducible gene 1-like receptors (RLRs), and the C-type lectin receptors (CLRs) [114]. As a result, O’Neill *et al.* speculated in a 2013 review that “TLRs have a more important role in repairing damaged tissue than in anti-microbial defense” [2]. Therefore, the next phase of TLR inquiry has been and will remain focused on non-canonical TLR activity during tissue repair and other “sterile” (i.e. no ongoing infection) processes. As it turns out, these non-canonical TLR functions rely on the detection of endogenous molecules broadly coined damage-associated molecular patterns (DAMPs). TLR-mediated detection of DAMPs, which now include well over three dozen “self” molecules, drives tissue repair during homeostasis and, when dysregulated, promotes the progression of many human diseases; however, the molecular details of TLR4:DAMP interactions remain mysterious [115].

1.2.1 TLR4-Activating DAMPs: A Key Disconnect

Initially, the observation that TLRs could trigger inflammation following stimulation by endogenous host molecules was surprising and controversial—why would there be an inherent self-reactivity in the innate immune system? In fact, after the first reports of TLR activation by “self” ligands, many blamed contamination of the ligands with bacterial lipoproteins or LPS for spurious TLR activation [115]. However, after many years and careful controls to rule out artefacts, it is now clear that

TLRs do indeed trigger inflammation upon the detection of endogenous DAMPs [116, 117]. The current consensus is that DAMPs alert the innate immune system to tissue damage or cellular stress, with the resulting inflammation mediating the clearance of debris and the induction of tissue repair. For this reason, DAMPs are oftentimes called “alarmins”, because they serve as the earliest warning of local tissue trauma [118]. There are a variety of negative feedback mechanisms in place to prevent runaway DAMP-driven inflammation, especially with regards to a spatiotemporal separation between PRRs and DAMPs. A host of examples show that DAMPs are often transiently released at sites of tissue injury, transcriptionally up-regulated following detection of an exogenous PAMP, confined intracellularly until cell lysis, released as smaller fragments from extracellular matrix (ECM) polymers, or maintained in an inactive form prior to a specific post-translational modification [119]. These checkpoints, while varied in form, all function to ensure that PRRs only encounter DAMPs and trigger inflammation under specific, well-regulated circumstances. However, during chronic autoimmune conditions, DAMPs can cause a spiralling pro-inflammatory feedback loop which Midwood and Piccinini refer to as the “damage chain reaction” [115]. In this model, an initiating harmful stimulus (perhaps via infection, local tissue damage, or a tumor) causes the release of DAMPs and the up-regulation of pro-inflammatory cytokines, which induce the release of more DAMPs via further tissue damage or transcriptional activation; if left unresolved, DAMPs increasingly accumulate in the tissue and create a chronically pro-inflammatory extracellular milieu. This damage chain reaction has been implicated in pathogenesis of atherosclerosis, rheumatoid arthritis (RA), inflammatory bowel disease (IBS), systemic lupus erythematosus (SLE), systemic sclerosis (SS), psoriasis, multiple sclerosis (MS), Sjogren’s syndrome (SS), myocardial infarction, and cancer (especially breast, lung, and pancreatic) [115].

There are now dozens of DAMPs with a link to human disease that have been reported to activate diverse PRRs, including TLRs [117], NLRs [120, 121], and RLRs [122]. However, especially for non-nucleic acid DAMPs (proteins, peptides, lipoproteins, and lipids), the TLRs seem to be the predominant PRR, with TLR4 and TLR2 being the most common [116]. Herein, I focus exclusively on TLR4-activating DAMPs, which are both the most numerous and the most well characterized. Table 1.1 provides an overview of DAMPs reported to activate TLR4-dependent signalling in cellular and/or animal models—these are listed in Column 1. Column 2 details the biophysical evidence, if any, showing direct binding of DAMPs to TLR4, with sub-optimal data to be taken as preliminary annotated with an asterisk. Data

are primarily marked with an asterisk due to artefactual, non-specific, or super-stoichiometric binding in biosensor assays like SPR; examples of these type of problems are discussed in detail in Section 1.2.2–1.2.4. Finally, Column 3 details any evidence regarding co-receptor or accessory protein requirements for the specified DAMP. Table 1.1 is not a comprehensive list of all TLR4-activating DAMPs but is rather designed to highlight an interesting selection of fifteen diverse DAMPs with the most comprehensive characterization in the literature. Finally, rather than discussing every DAMP in detail, which is beyond the scope of this thesis and has already been accomplished in the excellent review articles cited above, in the following Sections 1.2.2–1.2.4 I focus on three specific anecdotes to highlight what is, and still is not, known about TLR4-activating DAMPs.

Table 1.1 TLR4-Activating DAMPs. List of TLR4-activating DAMPs, with details regarding any biophysical or other data suggesting a direct interaction and/or data implying the involvement of co-receptors. Asterisks denote data with limitations, based on my interpretation from the original literature.

Table 1.1		
DAMP	Biophysical and Other Data	Co-Receptor Requirements
S100A8 and S100A9 [123]	3 nM binding of S100A9 to immobilized TLR4:MD-2 in SPR, dependent on Zn^{2+} and Ca^{2+} [124].* Direct binding of S100A8 to TLR4:MD-2 in SPR [123].* Immobilized TLR4:MD-2 pulls down a specific peptide from S100A9.* S100A9 binds to immobilized TLR4:MD-2 in ELISA [125].	Approximately 150 nM binding of S100A9 to CD14 in SPR. CD14 required for S100A9-mediated cell activation [126]. During cardiac inflammation, RAGE is the receptor driving $\text{NF}\kappa\text{B}$ [127]. CD14 and MD-2 are required for HEK293 reporter activation [125].
Fibronectin (FN) [128, 129]	FN-EDA binds surface of HEK293 expressing TLR4:MD-2:CD-14 via FACS, blocks anti-TLR4 antibody binding to these cells, and promotes adhesion of these cells [130].* FN-EDA can IP TLR4 and vice-versa [131].	MD-2 is required for FN-EDA [128]. Co-treatment with $\text{TGF-}\beta$ increases amount of EDA that TLR4 can IP [131]. MAPK activation, $\text{TGF-}\beta$ secretion, and fibroblast differentiation in response to FN-EDA blocked by anti-integrin $\alpha 4\beta 7$ antibodies [132]. FN-EDA can also bind to integrin $\alpha 9\beta 1$ and integrin $\alpha 4\beta 1$ to control cell adhesion, MAPK activation, and filopodial protrusions [133, 134]. Integrin $\alpha 4\beta 1$ and TLR4 cooperate to induce fibrosis upon FN-EDA treatment, and FN-EDA binding to $\alpha 4\beta 1$ sensitizes cells to TLR4 stimulation [135].
Fibrinogen [136, 137]	<i>None found</i>	Antibody blockade of CD14 renders monocytes unresponsive to fibrinogen [138]. D299G/T399I mutations in TLR4 cause hyper-responsiveness to fibrinogen (but hypo-responsiveness to LPS) [139]. Integrin $\alpha \text{M}\beta 2$ is required for fibrinogen-derived peptides to maximally activate TLR4; TLR4 is engaged indirectly via $\alpha \text{M}\beta 2$ [140]. TLR4 and $\text{Fc-}\gamma$ receptor both required for maximal macrophage activation by citrullinated fibrinogen [141].

Table 1.1 Continued		
DAMP	Biophysical and Other Data	Co-Receptor Requirements
Tenascin-C (TNC) [142, 143]	Soluble TLR4 binds ELISA plates coated with TNC ^{FBG} ($K_D = 58$ nM) or with peptides derived thereof [144]. By IF, TNC and TLR4 co-localize in murine heart tissue following myocardial infarction [145]. Anti-TLR4 antibodies can co-IP TNC from skin fibroblasts [146]	Neither MD-2 or CD14 is required for TLR4 activation [142]. CD36 is the major scavenger receptor required for TNC-induced foam cell formation [147]. LPS and TNC ^{FBG} result in different activated macrophage phenotypes [148]. TNC, but not LPS, induces residual release of TNF- α in TLR4 KO microglia, TNC-induced CCL2, CCL3, and CCL4 is independent of TLR4 expression, and TLR4 KO microglia still chemotactically respond to TNC but not LPS [149].
SLRPs Biglycan [150] and Decorin [151]	Biglycan IPs TLR4 and TLR2; biglycan binds TLR4 with $K_D = 23$ nM in MST [152].* Anti-decorin antibodies IP TLR4 and TLR2; decorin binds TLR4:MD-2 with $K_D = 37$ nM in MST [151]. In the presence of cross-linker, anti-biglycan antibodies co-IP TLR2 and TLR4 [150].	Decorin and biglycan are promiscuous binders of diverse receptor classes [153, 154]. Concomitant TLR2/TLR4 KO is required to obliterate activation [152]. Purinogenic receptors P2X ₇ and P2X ₄ cooperate to form a complex with TLR4, TLR2, and biglycan; anti-P2X antibodies IP all three of TLR2, TLR4, and biglycan [155]. Biglycan physically interacts with the Wnt ligand Wnt3a and the receptor LRP6 to drive Wnt activation [156]. Cytokine profiles differ between biglycan, decorin, and LPS [157, 158]. Biglycan IPs CD14 from macrophages, binds CD14 with $K_D = 67$ nM in MST, and requires CD14 for cytokine release [159]. Biglycan uses co-receptor CD14 to activate TLR2/TLR4, but co-receptor CD44 ($K_D = 9$ nM in MST) to activate TLR4-dependent macrophage autophagy [160].

Table 1.1 Continued		
DAMP	Biophysical and Other Data	Co-Receptor Requirements
Oxidized and minimally modified low-density lipoprotein (oxLDL and mmLDL) [161, 162, 163]	<i>None found</i>	CD36 is the immediate binding event for oxLDL, triggering a TLR4:TLR6 heterodimer [164]. CD14 is the immediate binding event for mmLDL, which is presented to TLR4:MD-2 [161]. Anti-TLR4 antibodies IP a 37 kDa co-receptor bound to oxidized phospholipids derived from mmLDL [162]. Cell surface protein DC-SIGN co-localizes with and interacts with TLR4 to drive cytokine release in response to oxLDL [165]. TLR2, TLR4, and CD36 all required for oxLDL-induced inflammation, but only TLR4 and CD36 required for oxLDL-induced foam cell formation [166].
Amyloid β ($A\beta$) [167]	Peptide derived from related serum amyloid A3 binds immobilized TLR4:MD-2 in SPR [168].*	TLR2 is key receptor in immune cells; TLR1 co-expression increases response to $A\beta$ aggregates, while TLR4 6 co-expression decreases response [169]. Anti-CD36 co-IPs TLR4 and TLR6 upon $A\beta$ treatment [164]. Concomitant TLR2, TLR4, and CD14 inhibition required to destroy $A\beta$ -induced cytokines [170]. $A\beta$ is a promiscuous binder to dozens of receptors [171].
HMGB1 [172, 173]	Macrophage cell-surface FRET shows TLR4:HMGB1 co-localization; anti-HMGB1 antibodies co-IP TLR4 [174]. Fibroblast proximity ligation assay shows cell-surface HMGB1:TLR4 co-localization; HMGB1 analyte binds to TLR4:MD-2 ligand with $K_D = 1.5 \mu\text{M}$ in SPR [175].* SPR shows oxidized HMGB1 binds TLR4:MD-2 ($K_D = 0.42 \mu\text{M}$) and TLR4 ($K_D = 0.64 \mu\text{M}$) [176].	TLR2, TLR4, and RAGE are all required for maximal cell activation by HMGB1 [174]. Mammalian-cell derived HMGB1 activates TLR4 in primary macrophages but not in HEK293 reporter cells; HMGB1 activates TLR2 in HEK293 reporters but not in macrophages [173].

Table 1.1 Continued		
DAMP	Biophysical and Other Data	Co-Receptor Requirements
Hyaluronic Acid Fragments (HA) [177, 178, 179, 180]	HA competitively prevents LPS stimulation of TLR4:MD-2 [181].	Accessory molecule CD44, which binds HA directly, forms a physical complex with TLR4:MD-2 (by co-IP) and is required for maximal detection of HA [182]. TLR2, not TLR4, may be the receptor for HA [183]. Concomitant TLR2/TLR4 KO required to abolish chemokine release from murine macrophages [179]. CD136 can act as an HA receptor [184]. HA may not be a direct TLR binder; regulates TLRs by modifying receptor access via a physical “jelly-like” barrier in mesangial cells [185]. HA can promiscuously bind diverse context-dependent receptors [180].
Histones [186, 187, 188]	Histone H4 binds TLR4:MD-2 with $K_D = 6 \text{ nM}$ in MST [186].* Anti-H2B can co-IP TLR4 and vice-versa from over-expressing ganglion cells [188]. Histone H3 can co-IP TLR4 from over-expressing HEK293 reporters and vice-versa [189].	Concomitant TLR4/TLR2 KO required to abolish signalling, H4 can activate both receptors [186, 190]. H4 also activates TLR9 and can co-IP TLR9 from hepatic cells [191]. H4-mediated platelet activation is not completely destroyed by concomitant TLR2/TLR4 blockade [192], and whole histone-mediate endothelial cell activation is only partially abrogated by TLR2 or TLR4 blockade [193].
Fetuin (AHSG) [194, 195]	Recombinant AHSG and TLR4 pull-down each other after mixing; AHSG binds TLR4:MD-2 in SPR with $K_D = 18 \text{ nM}$ when AHSG is the analyte and 23 nM when TLR4 is the analyte;* AHSG binds TLR4 ectodomain in intracellular yeast-two-hybrid system;* Native-PAGE shows TLR4:AHSG:Fatty Acid (FA) complex;* ELISA shows dose-dependent capture of AHSG by TLR4 and vice-versa [194]. AHSG and TLR4 co-localize both intra- and extra-cellularly in breast cancer cells by IF [196].	AHSG binds to and antagonizes insulin receptor via interactions with FN3 repeat domains [197]. AHSG binds to TGF- β and antagonizes binding to TGF- β receptor II [198]. AHSG is itself a co-receptor to present FAs to TLR4 [194]. In perivascular fat cells, AHSG-mediate inflammation and angiogenesis are only partially mediated by TLR4 [199]. AHSG is pleiotropic, being both pro- and anti-inflammatory depending on the context [200].

Table 1.1 Continued		
DAMP	Biophysical and Other Data	Co-Receptor Requirements
Nicotinate and Nicotinamide Phosphoribosyl Transferases (NAPRT and NAMPT) [201, 202, 203, 204]	In SPR with anti-NAMRT antibody as ligand, higher RU observed with TLR4 and NAMRT as analyte than with NAMRT alone;* NAMPT shares 30% identity with MD-2, with <i>in silico</i> modelling suggesting homology at the TLR4-binding region of MD-2 [203]. In SPR with anti-NAPRT antibody as ligand, slightly higher RU with TLR4 and NAPRT as analyte than with NAPRT alone;* deletion of a cationic peptide in NAPRT prevents TLR4 activation in macrophages [202]. Recombinant NAMPT is prone to aggregation in the absence of stabilizing SUMO tag, implying the existence of aggregation artefacts in reported binding and stimulation experiments [205].	NAMPT is historically considered both pro- and anti-inflammatory depending on the context [206]. Both NAMPT and NAPRT drive monocyte differentiation into macrophages via up-regulation of M-CSF, something LPS cannot achieve [204].
Lysozyme (LYZ) and Annexin A2 (ANXA2) [207, 208]	Anti-TLR4 antibody co-IPs LYZ from primary human spinal cord tissue [207]. Anti-ANXA2 antibodies co-IP TLR4, TRAM, and TRIF upon LPS stimulation of macrophages; ANXA2 co-localizes with LPS-activated TLR4 by IF [209]. Anti-TLR4 and anti-ANXA2 antibodies co-IP the other protein [210].	TLR4 is necessary, but not sufficient, for ANXA2 activation, since HEK293s over-expressing TLR4 or TLR4:MD-2 are unresponsive [208]. LYZ does not require MD-2, but does require the adaptor ANXA2, to which LYZ binds with $K_D = 3 \text{ nM}$ in SPR; ANXA2:LYZ complex is the true TLR4 ligand [207]. A lipid-raft-localized complex consisting of TLR4, calreticulin, nucleolin, and ANXA2 is required for anti-phospholipid antibodies to activate endothelial cells [210].

Table 1.1 Continued		
DAMP	Biophysical and Other Data	Co-Receptor Requirements
Surfactant Protein A (SP-A) and D (SP-D) [211, 212]	SP-D binds to ELISA plates coated with TLR4 in Ca^{2+} -dependent manner, estimated $K_D = 16 \text{ nM}$; SP-D binds to glycosylated and de-glycosylated TLR4 transferred to PVDF membrane [211]. TLR4 and SP-A can co-IP each other from baboon lung tissue and over-expressing HEK293s; in ELISA, SP-A can bind to MD-2 and TLR4:MD-2 complex; computational docking suggests a TLR4:MD-2:SP-A complex and reveals an SP-A peptide that binds TLR4 in ELISA* [213]. SP-A binds to TLR4 and MD-2 (ELISA) and pulls-down TLR4 and MD-2 in Ca^{2+} -dependent manner; in SPR, immobilized SP-A binds TLR4 with $K_D = 20 \text{ nM}$ and MD-2 with $K_D = 33 \text{ nM}^*$ [214].	MD-2 is not required for SP-A activation of NF- κ B in HEK293 reporter [212]. SP-A and SP-D both interact with and down-regulate TLR2 [211, 215, 216]. SP-A and SP-D both bind to CD14 (ELISA) and can modulate LPS responses; AUC evidence of a CD14:LPS:SP-A ternary complex [217, 218]. SP-A-derived peptides are anti-inflammatory, not pro-inflammatory [213, 219]. SP-A reduces TLR4-mediated cytokine release in murine intestine [220]. SP-A does not co-localize with TLR4 in alveolar macrophages but rather down-regulates LPS-activated TLR4 via β -arrestin [221]. Consensus is that SP-A can activate TLR4 but also inhibit LPS-mediated activation of TLR4.
Pancreatic (PE) and Neutrophil Elastase (NE) [222, 223, 224, 225, 226]	NE proteolytically degrades TLR4 and MD-2 both <i>in vitro</i> and <i>in vivo</i> , consistent with a direct interaction [223, 227].	MD-2 is not required in HEK293 reporter [222]. In murine macrophages, TLR4 deletion does not completely destroy NE-induced cytokine release [223]. NE degradation of TLR2, TLR4, MD-2, and cytokines prevents, rather than activates, an innate immune response during pneumococcal pneumonia [227]. NE is a “double-edged sword” with both host defense and tissue destruction phenotypes observed in different contexts [227, 228]. Integrin α M and NE co-localize in murine macrophages by IF [229]. NE trans-activates IL-12 release upon TLR4 ligation with LPS via EGFR [230]. NE activation of TLR4 is likely indirect, driven by activation of the matrix metalloprotease meprin which subsequently induces EGFR/TLR4 heterocomplexes (by co-IP and IF) [231]. Elastases can generically degrade the ECM to generate orthogonal cryptic DAMPs, including FN-EDA [232]. TLR4 co-operates and physically associates (co-IP) with the GPCR PAR2 to activate cytokine release in macrophages [233, 234].
End of Table 1.1		

1.2.2 S100A8/S100A9

Protein S100A8 (*aka* MRP8) and protein S100A9 (*aka* MRP14) are small (11 and 13 kDa respectively), antibacterial, EF-hand containing, metal-binding proteins belonging to the super-family of S100 proteins [235]. S100A8 and S100A9 are specifically expressed in circulating monocytes, macrophages, and neutrophils and, during inflammation, in keratinocytes and epithelial cells [236, 237, 238]. *In vivo*, S100A8 and S100A9 exist as heterodimer called Calprotectin (CP) which, in the presence of Ca^{2+} and other divalent cations like Zn^{2+} , can form a “CP dimer”, which is really a 2:2 heterotetramer [239, 240]; for clarity, I will refer to the higher order CP oligomer (S100A8:S100A9)₂ as a heterotetramer and the lower order CP oligomer (S100A8:S100A9) as a heterodimer. Note that CP binding to transition metals results in tetramerization but relies on different sites than classical Ca^{2+} binding to the EF-hand subdomains [241]. *In vitro*, S100A8 and S100A9 can also form obligate homodimers, which are often used experimentally to model the CP heterodimer due to the inability of the homodimers to tetramerize in the presence of divalent cations [241, 125]. S100A8 and S100A9 are usually cytosolic, representing up to half of the intracellular protein content in neutrophils and 5% in monocytes [242], where they contribute primarily to maintenance of the actin cytoskeleton and regulation of cell mobility [243, 244]. However, upon cell activation during infection, S100A8 and S100A9 are actively secreted extracellularly via a non-canonical pathway involving microtubules [245], resulting in a 100-fold increase in serum levels during inflammation [246]. Because of its strong induction during infection, CP is a *bona fide* biomarker for inflammation [247]. CP is the most abundant protein in the synovial fluid from patients with RA, where it correlates with disease progression, and CP can reach local concentrations of up to 1 mg mL^{-1} at sites of infection or tissue damage [248, 249]. Upon extracellular release, CP directly contributes to microbial clearance via the high-affinity, EDTA-mimicking chelation of metal nutrients including Zn, Mn, Cu, Fe, and Ni [250]. In addition to its direct role in bacterial clearance, extracellular CP acts as a DAMP to drive pro-inflammatory cytokine release via activation of TLR4 and receptor for advanced glycation endproducts (RAGE) [123, 127]. Chronic up-regulation of CP has been extensively linked to maintenance of autoimmune diseases including RA, psoriasis, SLE, SS, and osteoarthritis (OA) [243], and CP directly mediates both joint inflammation and tissue destruction during antigen-induced arthritis [251]. Primarily because of its early identification and contribution to diverse human diseases, CP is one of the most well characterized DAMPs in the literature.

Vogl *et al.* first reported in 2007 that S100A8 interacts with TLR4:MD-2 to activate NF- κ B and IRAK1 [123]. In S100A9 KO murine bone-marrow derived mononuclear cells (BMCs), which lack both S100A8 and S100A9, supplementation with S100A8 or CP, but not S100A9, resulted in the strong release of TNF- α . In fact, $5\text{ }\mu\text{g mL}^{-1}$ CP resulted in higher TNF- α levels than stimulation with $1\text{ }\mu\text{g mL}^{-1}$ of LPS. The addition of Polymixin B, an LPS binding antagonist, could not reverse this phenotype, but heat-denaturation of S100A8 rendered the protein inactive, confirming that pro-inflammatory cytokine release was due to the S100A8 protein and not LPS contamination. In BMCs from *TLR4* mutant C3H/HeJ mice, S100A8 was completely unable to induce pro-inflammatory cytokine release, while over-expression of TLR4, MD-2, and CD14 rendered HEK293 reporter cells responsive to S100A8, confirming that S100A8 activates TLR4-dependent signalling. Similarly, Grevers *et al.* reported that S100A8 could induce differentiation in wild-type, but not TLR4-deficient, murine osteoclasts. These cellular results have been independently reproduced in human patients by Lee *et al.* [252]. In healthy human peripheral blood mononuclear cells (PBMCs), Lee *et al.* reported that [S100A8] as low as $0.2\text{ }\mu\text{g mL}^{-1}$, but not S100A9, induce the secretion of IL-1 β , IL-6, IL-17, and TNF- α . Addition of S100A8 to patient-derived RA fibroblast-like synoviocytes (FLS) results in the time- and concentration-dependent release of IL-6, and this effect could be reversed by an anti-TLR4 antibody or small-molecule antagonists of PI3K, c-Jun, ERK1/2, MAPK, or NF- κ B. In contrast to these reports, which argue that S100A8 is the dominant TLR4 activator, Riva *et al.* showed that recombinant, LPS-free S100A9 homodimers could also activate NF- κ B and drive pro-inflammatory cytokine release in both THP-1 monocytes and bone-marrow derived murine dendritic cells (BM-DCs) [253]. These phenotypes were fully TLR4-dependent—in BM-DCs from TLR4 KO mice, S100A9 could not induce IL-6, IL-1 β , TNF α , or nitric oxide secretion. In 2018, Vogl *et al.* provided further validation and nuance to these results by elucidating a negative feedback mechanism based on Ca^{2+} binding [125]. In human monocytes cultured in Ca^{2+} -containing media, only S100A8 and S100A9 homodimers, but not CP, could induce TNF α secretion. The activity of CP could be rescued by introduction of mutations in S100A9 (N69A or E78A) which prevent calcium binding and the resulting tetramerization. These results were reproduced in a TLR4-expressing HEK293 reporter system and with *in vivo* mouse models: only tetramerization-deficient CP^{mut} could induce expression of IL-8 and TNF α to drive inflammation. Vogl *et al.* proposed that secreted CP is initially able to activate TLR4 but, due to the high concentration of extracellular Ca^{2+} , becomes rapidly

sequestered into an inactive tetramer. This explains why S100A8 and S100A9 homodimers, which do not tetramerize, could still induce pro-inflammatory cytokine release in Ca^{2+} -containing media. Therefore, across many studies, the cellular biology is clear: homodimeric S100A8, homodimeric S100A9, and heterodimeric CP activate inflammation in mice and humans in a TLR4-dependent manner.

However, the biophysical data supporting a direct interaction between TLR4 and S100A8/A9 are much less clear. For all of the biosensor data (primarily SPR) I discuss, and often critique, in this and the following Sections, I rely heavily on the work of Rich and Myszka, who in 2010 published the definitive guide to conducting, analyzing, and reporting optical biosensor-based binding experiments [254]. In their original 2007 paper, Vogl *et al.* report one representative SPR trace from three experiments with TLR4:MD-2 as the analyte (i.e. the soluble phase molecule flowed over the chip) and S100A8 as the ligand (i.e. the immobilized molecule on the surface of the chip). The authors use a mouse TLR4:MD-2 fusion construct (known as “LPS-Trap” and containing MD-2 fused to the C-terminus of TLR4 with a medium-length GlySer linker [255]) expressed in S2 insect cells as the liquid-phase analyte, and covalently immobilized recombinant mouse S100A8 as the solid-phase ligand. As negative controls, the authors immobilize RAGE-Fc and mouse serum albumin (MSA). The authors do not report the ligand loading density, model fits, or any calculated binding constants, but they do show a response of over 500 RU with 121 nM TLR4:MD-2 analyte and 300 RU with 67 nM; they observe no response with immobilized RAGE or MSA. In addition to the non-physiological TLR4:MD-2 construct used as the analyte and the lack of affinity constants or model fits, there are pathologies in both the shape and the magnitude of Vogl *et al.*’s sensorgrams which suggest artefactual binding. First, the association phases of the sensorgrams never reach an equilibrium plateau, and during the dissociation phase the response does not return to baseline, both of which indicate the non-specific accumulation of analyte on the biosensor surface. Furthermore, lower immobilization densities are more accurate, as super-saturation of the chip can lead to artefactual binding. Based on a physical ratio of the molecular weights of the ligand and analyte, it is possible to calculate a theoretical maximum response based on the initial response during ligand loading (for details, see <https://www.sprpages.nl/immobilization/immobilization-procedures>). To reach a maximum response of 500 RU, the authors would have needed to immobilize S100A8 to a level of 112 RU, which is relatively high; if a lower immobilization density were used, then this response of 500 RU would be super-stoichiometric and indicative of protein aggregation.

Similar pathologies are present in other S100:TLR4 binding data. S100A9 reportedly binds to and activates TLR4:MD-2, but only in the presence of Ca^{2+} and Zn^{2+} [124]; this is unexpected, since S100A9 does not change oligomeric states in the presence of metal ions. The authors use an SPR assay with 100 nM mouse S100A9 or CP as the analyte and mouse TLR4:MD-2 (fusion construct, LPS-Trap) as the immobilized ligand in a running buffer containing Ca^{2+} and Zn^{2+} . For S100A9, the authors report a max response of 1,270 RU for S100A9 and a K_D of 2.1 nM, and for CP, a maximum response of 280 RU and an affinity of 3.8 nM. They also show that the response increases as a function of $[\text{Zn}^{2+}]$. The authors do not report the model fits, but they do report an initial loading density of 3,000 RU for TLR4:MD-2 which is exceedingly dense. Even with this overloaded chip, the maximum theoretical response for S100A9 dimer as the analyte is 673 RU, which is half of the maximum response reported in their sensorgram. This is a clear sign of aggregation or non-specific binding. Additionally, the high affinity of CP for TLR4:MD-2 in the presence of divalent metal ions is concerning, because tetramerized CP cannot activate TLR4:MD-2 as discussed above. The pathologies I identify in this data are supported by a 2019 report from Noes *et al.* which showed, via detailed ITC, circular dichromism (CD), and inductively coupled plasma-mass spectrometry experiments, that S100A9 is prone to Zn^{2+} -dependent *aggregation* and that, when the Zn^{2+} binding residues are removed by mutation, S100A9^{mut} is still able to activate TLR4 in a HEK293 reporter assay [256]. Therefore, the binding reported by Bjork *et al.* is likely due to Zn^{2+} -dependent aggregation of S100A9 and not a specific interaction.

A final piece of biophysical data in the 2018 Vogl *et al.* report shows that immobilized human TLR4:MD-2 (anti-His antibody immobilized on beads and then used to capture His-tagged TLR4:MD-2) can pull-down a specific S100A9 peptide that is detected by MSMS [125]. Using an ELISA with immobilized TLR4:MD-2, the authors also show that S100A9 mutations within this peptide prevent adsorption of S100A9^{mut}. These results allow them to propose a structural model for TLR4:MD-2:S100A9 using previously reported crystal structures. I discuss the limitations of anti-TLR4 ELISAs in Section 4.2, so I will not belabor the point here, but the MS assay is also incomplete in my view. The authors do not show pull-down of whole S100A9 with immobilized TLR4:MD-2, but rather show pull-down of just a peptide from a S100A9 pre-digest or a specific synthesized S100A9 peptide (63-79). Furthermore, in a control pull-down with a scrambled peptide, the scrambled peptide is present in the elution (although at a lower relative intensity than the correct peptide 63-79), and the authors do not report the results from a negative control pull-down of the 63-79 peptide over beads without immobilized TLR4:MD-2 (although they do

report this negative control for the scrambled peptide and show that the beads themselves pull-down a detectable amount of the scrambled peptide). Yet again, then, these biophysical data showing a direct TLR4:S100A9 interaction are incomplete.

In contrast, there now exists more reliable SPR data showing that CD14 can bind directly to S100A9 [126]. The authors use human CD14 as the analyte and either human S100A8 or S100A9 as the immobilized ligand in SPR and show that S100A9, but not S100A8, binds CD14. Although they do not report ligand loading densities or model fitting parameters, and they only show one of two experiments, the maximum reported response is less than 20 RU which is reasonable and does not immediately indicate super-stoichiometric binding. The authors also show that blockade of CD14 prevents S100A9-activated pro-inflammatory cytokine release in THP-1 monocytes and that knock-out of either TLR4 or CD14 in murine BM-DCs prevents S100A9-induced TNF α release and S100A9 internalization. In addition to CD14, RAGE has also been indicated as the primary receptor for CP in a murine post-ischemic heart failure model [127]. Despite reported SPR data showing no interaction between S100A8 and RAGE (Vogl *et al.* report this as a negative control [123]) or CP and RAGE (Bjork *et al.* report that S100A9 but not CP can bind to RAGE [124]), Volz *et al.* show that RAGE is absolutely required to sustain CP-induced NF- κ B and inflammation *in vivo* [127].

In summary then, the cellular evidence that dimeric S100A8, S100A9, and CP can activate TLR4 both *in vitro* and *in vivo* is conclusive, but the biophysical data often cited to justify a direct 1:1 interaction is incomplete, sub-optimal, and contradictory. Furthermore, there is solid evidence that CD14 binds directly to S100A9 and that RAGE is required for CP activity *in vivo*—these data hint at co-receptor involvement rather than, or perhaps in addition to, a direct TLR4 binding event. This pattern of data will re-appear throughout the remaining anecdotes in this Section.

1.2.3 Fibronectin

Fibronectin (FN) is a polymeric, core ECM glycoprotein containing 12 fibronectin type I (FNI) repeats, two FNII repeats, 15 constitutively expressed FNIII repeats, two alternatively spliced FNIII domains called extra domains A and B (FN-EDA and FN-EDB), a variable FNIII connecting region (V), and a C-terminal di-cysteine motif that controls dimerization [257]. FN has diverse biological roles, including detecting and maintaining extracellular tension and tissue networks, regulating cell

adherence via interaction with integrin $\alpha 5 \beta 1$, and promoting tissue regeneration [258, 259]. FN’s pro-inflammatory activity is controlled by cryptic sites exposed by alternative splicing, proteolysis, and/or domain unfolding. Constitutively expressed FN in adult tissues does not contain the EDA domain, but at sites of tissue injury, remodelling, or chronic inflammation, EDA-containing FN is strongly up-regulated [257]. The EDA domain is the minimal DAMP responsible for inflammation, as either the isolated domain itself or the EDA-containing polymer (FN^{EDA}) activates TLR4-dependent cytokine release [128, 260, 261]. FN-EDA was first demonstrated to activate TLR4:MD-2 in 2001, making it one of the oldest known DAMPs [128], and its pro-inflammatory properties have been implicated in tumor metastasis, preterm birth, fibrosis, and other autoimmune conditions [131, 262, 263]. Additionally, a peptide derived from the first FNIII repeat (known as FNIII-1c) that mimics the unfolded FNIII1 domain under conditions of matrix rigidification can activate both TLR2 and TLR4 [264, 265]. Indeed, FNIII-1c and FN-EDA can synergize to enhance pro-inflammatory cytokine release via TLR4, demonstrating that both the splicing status and structure of FN are directly communicated to cells via two distinct pro-inflammatory regions [265]. Finally, the 13th and 14th FNIII repeat domains (FNIII13–14), which can be proteolytically released upon degradation of full-length FN, can activate the secretion of aggrecanase from wild-type murine hip explants but not from TLR4 KO explants [266]. In summary, rather than being controlled by extracellular release like CP, the pro-inflammatory properties of FN are modulated by physical changes which activate or reveal cryptic pro-inflammatory epitopes.

In their original 2001 paper, Okamura *et al.* showed that recombinant FN-EDA can activate MMP-9 production, NF- κ B nuclear translocation, and IL-10 release in HEK293 reporter cells, THP-1 monocytes, and mouse splenocytes in a TLR4:MD-2-dependent manner [128]. The authors show that TLR4 activation is dependent on the complete FN-EDA protein and not due to LPS contamination, as heat denaturation of FN-EDA prevents activation but pre-incubation with the LPS antagonist Polymixin B does not. Gondokaryono *et al.* showed that FN-EDA at concentrations as low as $1 \mu\text{g mL}^{-1}$, but not other FNIII repeats, can stimulate the release of TNF α , IL-1 β , and IL-6 in wild type murine bone-marrow derived mast cells (BMMCs) but not in BMMCs from TLR4 KO mice [261]. Furthermore, Gondokaryono *et al.* demonstrated that TLR4 is required for joint swelling and cytokine release *in vivo* upon FN-EDA injection into murine joints. Lefebvre *et al.* found that FN-EDA can prime the biosynthesis of leukotriene B₄ in and activate migration of polymorphonuclear neutrophils (PMNs) both *in vitro* and *in vivo* in a TLR4-dependent manner [260]. Just 300 nM FN-EDA induces strong PMN activation, a phenotype

which can be reversed by small molecule TLR4 inhibition but not by pre-incubation with Polymixin B. With regards to the pro-inflammatory peptide FNIII-1c, You *et al.* showed that FNIII-1c stimulates a strong transcriptional upregulation of TNF α , IL-8, and IFN- β in human dermal fibroblasts and that IL-8 release following stimulation with FNIII-1c can be inhibited by anti-TLR4 antibodies, but not anti-TLR2 or IgG control antibodies, in a dose-dependent manner [129]. Finally, a comprehensive reproduction of these results was published in 2014 by Kelsh *et al.* [265]. Kelsh *et al.* not only showed that FNIII-1c and FN-EDA can activate TLR4-dependent NF- κ B signalling and pro-inflammatory cytokine release in human dermal fibroblasts, but also that the two DAMPs can synergize to activate even stronger and higher affinity cytokine release. When used in combination, just 50–100 nM of FNIII-1c + FN-EDA can activate significant release of IL-8, and FN-EDA shifts the cellular EC50 of FNIII-1c by ten-fold from 9 μ M to 900 nM. In conclusion, in both human and murine models, FN-EDA, FNIII-1c, and to a lesser extent FNIII13–14 drive potent and reproducible pro-inflammatory signalling through TLR4.

However, over the twenty years since the FN:TLR4 axis was proposed by Okamura *et al.*, there remains very little evidence to support direct binding of either FN-EDA or FNIII-1c to TLR4. Using FACS, Lasarte *et al.* reported that recombinant FN-EDA binds to the cell surface of HEK293 cells expressing TLR4:MD-2:CD14 but not to LacZ-expressing HEK293s [130]. Similarly, Lasarte *et al.* showed that TLR4:MD-2:CD14-expressing HEK293 cells, but not LacZ-expressing cells, can adhere to FN-EDA-coated microtitre plates, although this could be indicative of TLR4-integrin cross-talk rather than direct binding, especially given FN-EDA’s promiscuous integrin binding activity (see Table 1.1). Furthermore, the authors found that increasing amounts of FN-EDA inhibit the binding of anti-TLR4 antibodies to the HEK293 cell surface in a dose-dependent manner. However, the maximum binding inhibition is under 50%, and the authors use up to 0.6 mM FN-EDA to achieve it, bringing into question the specificity of the observed antibody blockade. In whole-cell lysates from human skin fibroblasts, Bhattacharyya *et al.* reported that anti-TLR4 antibodies can co-IP FN^{EDA} and that anti-FN^{EDA} antibodies can co-IP TLR4 [131]. In a similar experiment in murine liver-derived megakaryocytes stimulated with FN^{EDA}, Malara *et al.* reported that an anti-TLR4 antibody could co-IP FN^{EDA} [267]. Outside of these indirect indications of binding, I could find no other biophysical evidence in support of a direct FN:TLR4 interaction.

In alignment with the dearth of biophysical data, there is increasing evidence to suggest the involvement of co-receptors in FN detection. At a basic level, Okamura *et*

al. showed in their original 2001 report that MD-2 is required to sensitize HEK293 cells to FN-EDA [128]. Furthermore, Kohan *et al.* reported that MAPK activation, TGF- β secretion, and fibroblast differentiation in response to FN-EDA (all of which are TLR4-dependent in other reports) can be inhibited by anti-integrin $\alpha 4\beta 7$ antibodies, suggesting cross-talk between TLR4 and integrin $\alpha 4\beta 7$ [132]. Along the same line, Kelsh-Lasher *et al.* demonstrated cooperation between TLR4 and integrin $\alpha 4\beta 1$ in the FN-EDA-dependent induction of fibrosis: maximal fibroblast responses required both receptors and pre-stimulation of $\alpha 4\beta 1$ with FN-EDA primed the cells for subsequent TLR4 activation [135]. Bhattacharyya *et al.* showed that treatment of skin fibroblasts with TGF- β *increases* the amount of TLR4 that can be co-IP'd with an anti-FN^{EDA} antibody and the amount of FN^{EDA} that can be co-IP'd with an anti-TLR4 antibody [131]. This persuasively suggests the involvement of other proteins, perhaps members of the TGF receptor super-family itself or downstream protein targets transcriptionally activated by TGF- β . Irrespective of the exact mechanism, however, the best explanation for a secondary stimulus changing the apparent affinity of FN^{EDA} for TLR4 is the physical involvement of a context-specific accessory protein(s). Interestingly, in human lung fibroblasts, Zheng *et al.* found that anti-TLR2 antibodies, but not anti-TLR4 antibodies or IgG controls, abrogate FNIII-1c-induced IL-8 secretion and NF- κ B nuclear translocation [264]. This result implies that FNIII-1c can use different TLRs in different contexts, as dermal fibroblasts rely on TLR4 for pro-inflammatory signalling [265]. This is indicative of co-receptor involvement, since TLR4 is also expressed by lung fibroblasts and TLR2 is also expressed in dermal fibroblasts [268, 269]; therefore, FNIII-1c's TLR flexibility is likely controlled by context-specific adaptor proteins that can only present the peptide to one or the other of TLR2 and TLR4. An additional experiment from Kelsh *et al.* supports this notion. Kelsh *et al.* reported that FN-EDA and FNIII-1c activate IL-8 secretion from both monocytes and dermal fibroblasts, while Tenascin-C (TNC; another TLR4 activating DAMP and the focus of this thesis; see Section 1.3) could only activate IL-8 release in monocytes and not dermal fibroblasts [265]. All of these molecules rely on TLR4 to drive inflammation, and both cell types express TLR4—therefore, if a direct TLR4 binding event were sufficient for cell activation, why can FN but not TNC stimulate cytokine release from dermal fibroblasts? The most parsimonious explanation for these data is the involvement of an FN co-receptor(s) that is expressed in both cell types and a separate TNC co-receptor that is expressed in only monocytes. To conclude, in my assessment the evidence for co-receptor involvement in the TLR4-mediated detection of FN is stronger than the biophysical data suggesting a direct interaction.

1.2.4 Small Leucine-Rich Proteoglycans

Biglycan and decorin are class I small leucine-rich proteoglycans (SLRPs) which share 60% sequence homology, are composed of tandem LRRs accompanied by a unique CX₃CXCX₆C motif at the N-terminal Cys-cluster, bear glycosaminoglycans containing chondroitin and dermatan sulfate disaccharide side chains, and are localized within the ECM [270]. Because of their structural complexity and glycosaminoglycan heterogeneity, decorin and biglycan have been reported to bind to dozens of matrix proteins, cytokines, and receptors with diverse impacts on collagen fibrillar organization and matrix assembly, cell adhesion and migration, cell proliferation, and fibrosis [271, 272]. During normal conditions, these multi-functional proteins remain sequestered in the ECM at collagen-enriched areas; however, during tissue injury or stress, decorin and biglycan are proteolytically liberated from the ECM by matrix metalloproteases (MMPs) and/or synthesized and secreted as soluble proteins *de novo*, allowing them to act as DAMPs via activation of TLRs and other receptors [273, 274]. For this reason, the pro-inflammatory incarnations of SLRPs are often termed “soluble” to differentiate them from the solid-phase matrix-bound species that dominate during homeostasis. As with FN-EDA, SLRPs were some of the first identified DAMPs—Schaefer *et al.* demonstrated biglycan activation of TLR4 and TLR2 as early as 2005 [150]. Because of their diversity, multifunctionality, abundance, and ubiquity, SLRPs orchestrate cross-talk between numerous signalling pathways, including the feed-forward amplification of pro-inflammatory signals from diverse sources [155, 151, 154, 119]. The literature surrounding decorin and biglycan is immense, and interested readers are pointed towards comprehensive reviews which detail the involvement of these SLRPs in renal disease, carcinogenesis, diabetes, autoimmune diseases, fibrosis, tissue regeneration, and more [273, 274, 271, 275, 276, 277, 278, 272, 158]. In this section, I briefly review the evidence surrounding biglycan and decorin’s interaction with TLR4 and other immune receptors.

In their original 2005 paper, Schaefer *et al.* demonstrated that soluble biglycan activates MAPK and NF- κ B and the release of TNF- α , IL-6, and IL-1 β from macrophages and HEK293 reporter cells [150]. Biglycan-mediated cytokine secretion is reduced by knockout of either TLR2 or TLR4 and completely abrogated by concomitant TLR2/TLR4 knockout or MyD88 deletion. Babelova *et al.* reproduced reduced IL-1 β secretion from TLR2/TLR4 KO murine peritoneal macrophages upon treatment with 4 μ g μ L⁻¹ biglycan [155]. Biglycan activation of TLR2 and TLR4 also triggers the expression of B-cell chemoattractant CXCL13 in murine macrophages

and DCs, implying that biglycan also communicates with the adaptive immune system [279]. As with the other cytokines, Moreth *et al.* reported that biglycan-triggered CXCL13 levels are reduced in TLR2 or TLR4 knockout cells and completely abolished in macrophages and DCs lacking both receptors; these results also hold true in *in vivo* mouse models, with TLR2 or TLR4 KO reducing plasma levels of pro-inflammatory cytokines following biglycan injection. [279]. Similarly, in a constitutive mouse model where biglycan is overproduced in hepatocytes and released systemically, Zeng-Brouwers reported a robust up-regulation of CXCL1, CXCL2, and CCL2 which could be reversed with MyD88 deletion, and an up-regulation of CCL5, which could be reversed by TRIF deletion, providing strong *in vivo* evidence for biglycan’s ability to activate both the MyD88-dependent and MyD88-independent signalling pathways downstream of TLR2/TLR4 [280]. Broadly, the cellular results are quite similar for decorin. Merline *et al.* reported in 2011 that decorin activates NF- κ B and the release of TNF α and IL-12 from macrophages, and that concomitant knockout of TLR2 and TLR4 is required to completely reverse decorin-induced cytokine release [151]. Additionally, in an *in vivo* murine model of SS, splenocytes secrete TNF α upon decorin stimulation but not in the presence of the small molecule TLR4 inhibitor TAK242 [157]. Yet again, the cell biology here is robust: across diverse human and murine models, decorin and biglycan can activate pro-inflammatory cytokine release via TLR2 and TLR4.

Of all the DAMPs discussed in Section 1.2.1, biglycan and decorin have the strongest biophysical evidence for a direct interaction with TLR4, allowing the authors of a 2020 review to conclude that “biglycan binds to TLR2 and TLR4 with an affinity comparable to respective pathogen derived ligands” [158]. However, careful review of the evidence suggests this might be a premature conclusion, both in terms of known co-receptor requirements and the quality of the biophysical data itself. The ability of decorin and biglycan to IP TLR4 is clear: Moreth *et al.* reported that anti-biglycan antibodies can co-IP TLR4 in HEK293 reporter cells [152], a result that has been directly reproduced in the same cells using anti-decorin antibodies to co-IP TLR4 [151]. In their original report, Schaefer *et al.* further demonstrated that, in the presence of a cross-linker, anti-biglycan antibodies can co-IP TLR4 from macrophages; Schaefer *et al.* used both western blot and MSMS to validate this pull-down, removing any doubt about antibody specificity [150]. However, IP experiments are always indirect—indeed, all three of these papers demonstrate that decorin and biglycan can also co-IP TLR2, consistent with the requirement for both TLR4 and TLR2 to achieve maximal response to these SLRPs. These IPs cannot,

therefore, distinguish between separate binding events to TLR2 and TLR4 or a single binding event to a larger complex that includes both TLRs. In addition to these IP data, Moreth *et al.* published a microscale thermophoresis (MST) experiment purporting to show binding of fluorescently labelled human TLR4 to human biglycan with a $K_D = 23$ nM [152]. However, the data are quite noisy (and are presented not as averages with error bars, which would be appropriate for an experiment with $n = 3$, but rather as a scatter plot, which misrepresents the spread of the data), and despite their software package placing a binding curve, the maximum average response actually occurs *at* the reported K_D , where the response should be half maximal. This is a severe pathology in the data that renders the curve uninterpretable in my opinion. The only reported negative control is albumin, which does not appear to bind biglycan; however, this is a poor control, especially since both TLR4 and biglycan are heavily glycosylated LRRs, features which can complicate the behavior of proteins in narrow glass MST capillary tubes due to non-specific adsorption. A final problem with these data is the use of TLR4 and not TLR4:MD-2, especially since Schaefer *et al.* reported a clear dependence on MD-2 for cell activation in the presence of biglycan [150]. An improved MST experiment using human TLR4:MD-2 and human fluorescently labelled decorin is present in the Supplementary Information from Merline *et al.* and shows an apparent $K_D = 38$ nM [151]. These data are the best evidence of a direct interaction between any DAMP and TLR4 that I have seen in the literature—there are no obvious pathologies in the data, and the biologically relevant TLR4:MD2 complex is used as the binding partner. However, just like Moreth *et al.*, Merline *et al.* use Tween-20 rather than the preferred BSA to prevent non-specific adsorption to the glass capillaries, and they do not report or mention pre-test scans which rule out this artefact [281, 282]; furthermore, there is no report of a positive control (LPS being an obvious choice) which could orient the data by showing what a “true”, specific response looks like. There is a final, more generic observation that makes me suspicious of non-specific binding in these MST experiments. A low-nanomolar interaction would imply both the stability of the complex during SEC, which is a common experiment presented for the PAMPs as discussed in Section 1.1.3, and also that binding could be trivially observed with SPR, ITC, cross-linking SDS-PAGE, or other orthogonal techniques. However, despite careful searching, I could not find any of these simple follow-up experiments in the literature for decorin or biglycan. Therefore, at best, there is preliminary evidence from just MST to suggest that just decorin binds to TLR4:MD-2, but in the absence of orthogonal validation these data remain quite preliminary and do not justify the consensus in the literature of a direct interaction.

This conclusion is further underscored by the now overwhelming evidence that decorin and biglycan rely on co-receptors. First, as mentioned previously, both TLR2 and TLR4 are required for maximal responses to decorin and biglycan, suggesting the involvement of multiple receptors. Second, Babelova *et al.* reported that inflammasome activation and IL-1 β release in response to biglycan is dependent on the purinogenic receptors P2X₄ and P2X₇ in addition to TLR2/TLR4 [155]. They show that knockout or knockdown of these purinogenic receptors renders murine peritoneal macrophages unresponsive to biglycan, and go further by demonstrating that anti-biglycan antibodies simultaneously co-IP TLR2, TLR4, P2X₄, and P2X₇. Although the authors are careful to state that independent receptor engagement is possible, the bulk of their data allows them to conclude that “biglycan mediates clustering of TLR2/4 and purinogenic P2X₄/P2X₇ receptors, inducing receptor cooperativity within these newly formed multi-receptor complexes” [155]. Third, Berendsen *et al.* demonstrated that biglycan is necessary for canonical activation of Wnt signalling through a complex with Wnt-agonist Wnt3a and the lipoprotein receptor LRP6, as biglycan knockout abrogates Wnt activation both *in vitro* and *in vivo* [156]. Berendsen *et al.* reported that anti-LRP6 and anti-biglycan antibodies co-IP all three of biglycan, LRP6, and Wnt3a from over-expressing HEK293 cells, allowing the authors to hypothesize a role for biglycan as a co-receptor that presents Wnt3a to the receptors Wnt and LRP6 to drive Wnt activation. Fourth, Roedig *et al.* reported a direct interaction between biglycan and CD14 in MST, with high quality data showing an apparent $K_D = 67$ nM, and showed that anti-biglycan antibodies can co-IP CD14 [159]. The observed interaction between biglycan and CD14 is functionally important, as knockout of CD14 in murine peritoneal macrophages renders them wholly unable to secrete TNF α , CCL2, or CCL5 in response to biglycan stimulation. Roedig *et al.*’s data, taken in full, argue that CD14 is the master regulator for biglycan-induced pro-inflammatory signalling through both TLR2 and TLR4, since CD14 KO reproduces the phenotype of concomitant TLR2/TLR4 KO both *in vitro* and *in vivo*. Poluzzi *et al.* expanded these results further to show that biglycan can “select” for different co-receptors to drive a particular phenotype: biglycan is presented to TLR4 by CD44 to drive autophagy in human and murine macrophages, but it is presented to TLR4 and TLR2 by CD14 to drive pro-inflammatory cytokine release [160]. The authors published good quality MST data showing an interaction between human biglycan and human CD44 with an apparent $K_D = 8.5$ nM and demonstrated co-IP of biglycan by anti-CD44 antibodies directly from murine macrophage lysates. Poluzzi *et al.*’s results bring clarity to conflicting

evidence that biglycan can be anti-inflammatory: the CD44-dependent pathway results in autophagy-induced M2 macrophage polarization and tissue repair, while the CD14-dependent pathway results in pro-inflammatory M1 macrophage infiltration, cytokine release, and tissue damage. Taken together, these reports demonstrate that the SLRPs biglycan and decorin rely on a variety of context-dependent, cell-surface co-receptors and likely do not form a direct 1:1 complex with TLR4 or TLR2. Particularly, the observation that co-receptor selection by the *same* ligand results in *different* phenotypes through the *same* receptor cannot be explained by direct, high-affinity binding. To conclude, although these explanations are not mutually exclusive, as with CP and FN there is much stronger evidence for SLRP co-receptors than there is for a direct interaction between SLRPs and TLR4.

1.2.5 Summary of TLR:DAMP Literature

My goal in Section 1.2 was to first introduce the concept of endogenous inflammatory signalling and then highlight the disparity between the cellular and biophysical data surrounding TLR:DAMP interactions. Hundreds of experiments in cellular and animal models support the conclusion that DAMPs activate pro-inflammatory signalling via TLRs—knockdown, knockout, antibody blockade, mutation, or small-molecule inhibition of specific TLRs (especially TLR2 and, of primary interest here, TLR4) renders cells and animals non-responsive to DAMP stimulation. However, the biophysical evidence supporting a direct interaction between TLRs and DAMPs is much sparser. In fact, much of the biophysical data in support of the relatively common hypothesis that DAMPs form a stable complex with TLRs is sub-optimal and, in a few exemplar cases, has been challenged by data suggesting direct binding to essential co-receptors. Indeed, the literature is scattered with low quality SPR and MST data characterized by non-specific or super-stoichiometric binding artefacts. Furthermore, unlike for TLR:PAMP interactions (reviewed in Section 1.1.3), there are no structures of any DAMP bound to any TLR from any organism, and I could find no evidence of TLR4 co-eluting with any DAMP during SEC; the lack of these data creates a gap in the literature that obscures the true molecular mechanism behind DAMP detection and makes therapeutic targeting of DAMPs more challenging. TLRs are certainly essential for DAMP activation of inflammation—but do direct, high-affinity complexes between TLRs and DAMPs exist? Are co-receptors or accessory proteins the direct DAMP binders? Are receptor mega-complexes involved in fine-tuning DAMP signalling? What are the fundamental molecular events behind DAMP activation of the innate immune system? These questions drive the rest of

my thesis and lead into Section 1.3, wherein I introduce the so-called “FReP” family of TLR4-activating DAMPs as an exemplar test-case to probe these unanswered questions surrounding DAMP biology.

1.3 TNC and the FReP Family

Section 1.2 provided an overview of the literature surrounding TLR4’s interaction with various DAMPs to emphasize the mystery surrounding the molecular mechanism behind DAMP activation of the innate immune system. In this Section 1.3, I introduce the DAMPs of interest in the rest of my thesis, namely TNC and the other fibrinogen-related proteins (FRePs), which serve as a model system for interrogating the TLR4:DAMP axis more broadly.

1.3.1 Tenascin-C

TNC is a large, multi-modular ECM glycoprotein whose name is derived from the two tissues where it is predominantly expressed: tendons (Latin “tenere”, meaning to hold) and embryos (Latin “nasci”, meaning to be born) [283]. The human *TNC* gene contains 30 exons and encodes for a 190–330 kDa full-length protein (glycosylation at 23 different sites being responsible for the range) containing up to 2201 residues [284, 285]. At the N-terminus, TNC contains multiple heptad repeats which facilitate homotrimerization; subsequently, TNC homotrimers can dimerize to create homohexameric TNC “hexabrachions”, which are large, structural ECM polymers. Moving from the N-terminus, TNC contains 14.5 EGF-like repeats and then a long string of up to 17 FNIII repeats, eight of which are constitutively expressed and nine of which are alternatively spliced. At the very C-terminus (Ile1974–Ala2201), TNC contains a fibrinogen-like globe (FBG) domain, which in Chapter 1 I refer to as TNC^{FBG} to differentiate it from the full-length protein. Unlike most ECM constituents, which are often broadly expressed both spatially across tissues and temporally across lifespans, TNC is sparsely expressed in just a small subset of adult tissues including tensile-bearing connective tendons, underneath some epithelial cells, and in certain stem cell niches like hair follicles and bone marrow [286, 287, 288]. Uniquely, TNC is transiently and specifically up-regulated only at sites of inflammation and/or tissue damage and contributes to the re-shaping of the ECM towards tissue repair [289, 290, 291], including via direct modulation of cell adhesion via

binding to FN [283, 292, 293]. Broadly, TNC promotes tissue rebuilding, ECM synthesis and assembly, wound contraction, cell clearance, and angiogenesis, all of which are important functions in the global inflammatory response that promote repair following infection or physical damage [294]. Consistent with these biological roles, TNC expression can be induced by a wide range of inflammation-related signals, including cytokines, growth factors, hypoxia, mechanical stress, and reactive oxygen species, which activate *TNC* via diverse transcriptional pathways including NF- κ B and c-Jun [291, 289]. Indeed, because of its potent and ubiquitous up-regulation at the earliest stages of inflammation, TNC is a reliable systemic biomarker for inflammation and tissue damage [294]. Importantly, TNC can drive its own transcription through activation of NF- κ B [295], a property which explains why TNC is often referred to as a “double-edged sword”. Under normal conditions, TNC expression ceases following resolution of the prevailing damage or infection; however, at sites of chronic inflammation TNC expression is never resolved, and accumulating TNC drives a spiralling autoimmune positive feedback loop that shapes a perpetually pro-inflammatory extracellular milieu [295, 296, 297, 298]. As a result, TNC has been implicated as a pathogenic driver of RA, cancer progression and metastasis, atherosclerosis, vascular and cardiac diseases, organ fibrosis, asthma, IBS, MS, and more [287, 291, 299, 289, 115, 298].

My goal in this section is not to review TNC biology in detail, which has been covered in the extensive reviews cited in the preceding paragraph (most recently reviewed excellently in 2018 by Marzeda and Midwood [298]), but rather to focus specifically on TNC^{FBG}, the key pro-inflammatory region within TNC that will be the focus of my thesis. TNC^{FBG} is a globular, 27 kDa domain stabilized by two disulfide bonds that is related to the fibrinogen globe domain first resolved structurally in 1997 [300]. I discuss the structural biology of this domain in detail in Chapter 2, particularly in Section 2.7 and 2.12, so I will focus here on the cellular and molecular biology of TNC^{FBG}. Primarily using a murine model of arthritic joint inflammation, Midwood *et al.* identified TNC^{FBG} as the TLR4-activating, pro-inflammatory DAMP within TNC in a seminal paper in 2009 [142]. Midwood *et al.* first showed that TNC KO mice are protected from both methylated-BSA-induced RA and zymosan-induced acute synovitis. Next, using primary human macrophages, Midwood *et al.* demonstrated that exogenous addition of LPS-free TNC induces the release of pro-inflammatory TNF- α , IL-6, and IL-8; in synovial fibroblasts, the authors reported a context-dependent release of only IL-8. Using purified, LPS-free recombinant domains of TNC (including the N-terminal heptad oligomerization domain, the EGF repeats, and multiple FNIII repeats), Midwood *et al.* reported

that TNC^{FBG} is the only domain that can stimulate pro-inflammatory cytokine release *in vitro* and inflammation *in vivo*—therefore, TNC^{FBG} is both necessary and sufficient for TNC’s pro-inflammatory properties, making this subdomain the *bona fide* DAMP of interest. To confirm TLR4-dependency, the authors go on to show that anti-TLR4 antibodies, but not anti-TLR2 antibodies, fully abrogate TNC^{FBG}-induced cytokine release from macrophages and synovial fibroblasts. Furthermore, Midwood *et al.* showed that neither CD14 (via anti-CD14 antibodies) nor MD-2 (via an antagonistic MD-2 ligand) are required for TNC^{FBG}-induced TNF- α release from primary human macrophages, a piece of evidence suggesting that TNC^{FBG} detection might not require a co-receptor. Many of these results were orthogonally reproduced in human cardiac myofibroblasts, in which only recombinant TNC^{FBG}, but not other TNC sub-domains, could induce pro-inflammatory cytokine and MMP-3 release, a phenotype that could be reversed with anti-TLR4 polyclonal antibodies [145].

These results were extended by Piccinini *et al.* in a 2016 report showing that LPS and TNC^{FBG} activate surprisingly different pro-inflammatory phenotypes in macrophages [148]. As expected, Piccinini *et al.* found that both LPS and TNC^{FBG} can activate NF- κ B, MAPK, and p38 α and induce TNF- α and IL-6 release; however, only LPS can induce IL-12 and IL-23 and TNC^{FBG} often induces statistically significantly lower cytokine levels than LPS. Importantly, all of these cytokine profiles are TLR4-dependent and can be inhibited by addition of the TLR4 signalling inhibitor TAK242 or an anti-TLR4 polyclonal antibody. Consistent with differences in the observed cytokine profiles, Piccinini *et al.* reported a clear divergence in the phosphoproteome of LPS- and TNC^{FBG}-stimulated human M-CSF-differentiated macrophages. By MSMS, 45 phosphoproteins are enriched upon LPS treatment and 25 phosphoproteins upon TNC^{FBG} treatment—but, strikingly, only two of these phosphoproteins (LMNA and DCD-2) are shared in both conditions, suggesting a fundamentally different phenotype in response to these TLR4 agonists. These differences have functional consequences, as TNC^{FBG} induces a macrophage phenotype characterized by ECM synthesis and phosphorylation of specific collagens, while LPS stimulates a matrix-degrading phenotype via collagenases. This study clearly demonstrates that macrophages can tell the difference between exogenous and endogenous TLR4 stimuli in order to initiate divergent (indeed, almost opposite) phenotypes; however, the authors are careful to highlight in the conclusion that the molecular mechanism behind these differences remains unclear, although they speculate that “The answer may lie in the recruitment of distinct adaptor molecules early in the signalling cascade” [148]. It is important to note here that in both Piccinini *et al.* and Midwood *et al.*, the concentration of TNC^{FBG} used for cell

stimulation (usually, at or around 1 μM) is consistent with TNC levels reported in actual human tissues, which range from 0.37 to 1.2 μM [301, 302].

Further work by Zuliani-Alvarez *et al.* identified the specific peptides within TNC^{FBG} that control TLR4 activation and binding [303]; this work is the key foundation of and motivation for my choice to pursue TNC^{FBG} as a model DAMP. First, Zuliani-Alvarez *et al.* deployed a solid-phase binding ELISA with 96-well plates coated with recombinant, *E.coli*-derived TNC^{FBG} to show dose-dependent binding to recombinant, commercially available human TLR4 ectodomain ($K_D = 58 \text{ nM}$; a schematic of this assay is presented in Figure 4.1A). Next, the authors defined nine overlapping peptides (P1–P9) stretching the entire length of TNC^{FBG} and showed that only three of them (P5, P6, and P7) induce dose-dependent binding of TLR4 to the immobilized peptides in the aforementioned ELISA. They also showed that these three peptides can competitively block binding of TLR4 to TNC^{FBG} in the ELISA if the peptides are pre-incubated with TLR4. In a THP-1 NF- κ B reporter system, Zuliani-Alvarez *et al.* showed correspondence between the peptide ELISAs and cell activation: the TLR4-binding P5 and P6 (at concentrations ranging from 20 to 100 μM) can dose-dependently activate NF- κ B. Next, Zuliani-Alvarez *et al.* proceeded with careful mutagenesis across the length of P5, P6, P7, and P9 (which itself does not bind to or activate TLR4 but was interrogated in detail by the authors because of its basic properties matching P5 and P6) to drill down on the key residues involved in TLR4 binding and TLR4 activation. Consistently, the cell activation and ELISA binding data correlate, with non-binding mutated peptides failing to stimulate either TLR4 reporters or primary human macrophages. Zuliani-Alvarez *et al.*’s big picture conclusion is that a region within P5 and P6 (nicknamed the “cationic ridge” and containing the critical sequence “**KTRYKLK**”) is required for TLR4 activation and binding, while regions in P7 (“loop 7 hydrophobic patch”, critical sequence “**DSAITN**”) and P9 (“C-terminal cationic loop”, critical sequence **NLEGRRKRA**) enhance binding affinity and activation level. For visualization of these key regions on TNC^{FBG} itself, please refer to Figure 2.10. To drive home their conclusions, Zuliani-Alvarez *et al.* reported that re-introduction of these three critical regions into the FBG domain from the homolog Tenascin X (TNX^{FBG}), which itself does not bind to or activate TLR4, *rescues* TLR4 binding and activation. The TNX^{FBG} construct containing the cationic ridge, hydrophobic patch, and C-terminal loop from TNC^{FBG} achieves equivalent TLR4 binding in the ELISA, NF- κ B activation in THP-1 reporters, and TNF- α release from primary human macrophages as the wild-type TNC^{FBG} . Finally, the authors showed the *in vivo* relevance of these findings by demonstrating that wild-type TNC^{FBG} , but not TNC^{FBG} containing point

mutations in the cationic ridge, can cause inflammation and tissue destruction when injected into murine joints.

These solid-phase ELISAs represent the best biophysical evidence purporting to show a direct interaction between TLR4 and TNC^{FBG} , and the correlation between the solid-phase binding data and the cell activation data is persuasive and does suggest a direct 1:1 binding event might be occurring. This is especially true in light of the “up assay” rescue experiment with TNX^{FBG} , which shows that installation of direct TLR4 binding properties (inferred from the ELISA) can impart TLR4 activation properties in both cell lines and primary cells. This conclusion is buttressed by the fact that neither CD14 nor MD-2 is required for TNC^{FBG} to bind to TLR4 in an ELISA or activate TLR4 in cells and that anti-TLR4 antibodies can co-IP full-length TNC from skin fibroblasts [142, 146]. However, like many of the other DAMPs discussed in Section 1.2, many of the “gold standard” complexation assays, including SPR and SEC co-elution experiments, have never been reported for TNC^{FBG} and TLR4. Furthermore, the phenotypic difference between LPS- and TNC^{FBG} -activated macrophages reported by Piccinini *et al.* is also puzzling in the context of a direct interaction. As Piccinini *et al.* noted, if a single binding event between either $\text{TNC}^{\text{FBG}}:\text{TLR4}$ or $\text{LPS}:\text{MD-2}:\text{TLR4}$ is sufficient to trigger receptor dimerization and downstream $\text{NF}\kappa\text{B}$ activation, what explains the divergence in the resultant effects of these two agonists? Which adaptors or co-receptors are required to impart this agonist-specific phenotype via an identical receptor? Yet again, therefore, the pattern of unambiguous cellular data coupled with incomplete biophysical data emerges with TNC^{FBG} much like the other DAMPs discussed in Section 1.2.

This gap in our understanding of TNC^{FBG} biology provides a strong impetus for detailed molecular studies, especially since TNC is a *bona fide* therapeutic target in human disease. In 2016, Bhattacharyya *et al.* showed that TNC (in a TLR4-dependent mechanism, thereby indirectly implicating TNC^{FBG}) is a key driver of skin and lung fibrosis during SS, a rare and sometimes deadly chronic disease characterized by organism-wide inflammation and fibrotic lesions [146]. Bhattacharyya *et al.* showed that TNC is strongly up-regulated in biopsies from SS patients and that exogenous addition of TNC to skin fibroblasts or 3D organoids results in fibrotic responses characterized by pro-inflammatory signalling, collagen expression, and reduced myofibroblast differentiation, phenotypes which can all be abrogated by TLR4 KO. Indeed, TNC KO protects mice from both bleomycin-induced skin and lung fibrosis, suggesting that inhibition of TLR4 activation via TNC^{FBG} could

systemically reduce fibrosis and provide therapeutic benefit in a challenging, complex disease like SS. However, follow-up work from the same group demonstrated that a small-molecule MD-2 antagonist could block pro-fibrotic responses *in vitro* and *ex vivo* and both prevent and reverse organ fibrosis in murine models [304]. Since the prevailing hypothesis is that TNC^{FBG} does not require MD-2 to activate TLR4, it is surprising that the authors also reported that their MD-2 antagonist can inhibit TNC-induced collagen secretion in skin fibroblasts. While this suggests that there remains controversy regarding the co-receptor requirements for and the pathogenic mechanism of TNC, it does again show that the TLR4:TNC^{FBG} axis is a viable drug target in SS. As a second example of the disease relevance, built upon Midwood *et al.*'s original finding that TNC^{FBG} drives murine arthritic joint disease through TLR4 [291], a biotech company called Nascent Ltd. developed and published anti-TNC^{FBG} monoclonal antibodies with therapeutic potential in RA [305]. The best antibody binds TNC^{FBG} with 70 pM affinity and could inhibit TNC^{FBG}-induced cytokine release by patient-derived RA synovial cells and prevent disease progression in a murine model of collagen-induced arthritis. Using the monoclonal antibody for staining experiments, the authors also showed that TNC levels are highest in patients with early RA versus patients with spontaneously resolving joint inflammation or patients with established RA—this suggests that early induction of the TNC^{FBG}:TLR4 axis may indeed be the initiating event, and therefore a critical therapeutic checkpoint, in RA.

However, despite these promising pre-clinical results in two different inflammatory diseases, the lack of clarity surrounding the exact molecular mechanism behind TNC^{FBG}-mediated activation of TLR4 makes designing targeted therapies challenging. Bhattacharyya *et al.* captured this truth quite succinctly: “DAMPs have vital protective functions required for tissue repair following injury...accordingly, caution will be needed in the pursuit of therapeutic approaches that target TLR4 activation by its endogenous ligands...” [146]. In this context, is a broad anti-TNC^{FBG} antibody, which likely neutralizes the small domain across multiple functions, the right approach? Are TLR4 or MD-2 inhibitors sensible approaches to “DAMPening” endogenous activation, since these molecules might also compromise protective innate immunity during bacterial infections? Is there a way, perhaps via small-molecule PPI inhibitors, to block pathogenic DAMP signalling but retain protective DAMP signalling? In conclusion, then, the pre-clinical evidence indicating TNC^{FBG} in the pathogenesis of human diseases is clear, but therapeutic interventions are limited and additional insight into the structure and molecular mechanism of TNC^{FBG}-mediated TLR4 activation could provide safer and more targeted strategies.

1.3.2 The Other FRePs

TNC^{FBG} belongs to a larger family of 24 human proteins, called FRePs, which all contain an FBG domain. The architecture of the FBG domain, characterized by at least two stabilizing disulfide bonds, a core β -sheet backbone, and three defined sub-domains, is strongly conserved both structurally (see Figure 2.20) and evolutionarily—even invertebrate marine sea squirts express FBG domains [306]. For a detailed structural comparison of FBG domains, please refer to Chapter 2, particularly Section 2.7 and 2.12. Based on the key TLR4-activating role of the cationic ridge in TNC^{FBG}, Zuliani-Alvarez *et al.* deployed homology modelling and bioinformatics to hypothesize that 12 other human FRePs should activate TLR4 via a conserved positively charged epitope, while the remaining 11 family members should not activate TLR4 due to the absence of this peptide [303]. To validate this hypothesis, Zuliani-Alvarez *et al.* showed that the recombinant FBG domains from fibrinogen and ficolin-1 (FCN1), which both contain a cationic ridge, can bind to TLR4 in an ELISA, stimulate IL-6 and IL-8 release from primary human macrophages, and drive synovial inflammation in mice. Importantly, cytokine release induced by these FBG domains can be inhibited by the TLR4 antagonist TAK242, suggesting that TLR4 is the required receptor across all FRePs. By contrast, the FBG domain from angiopoietin-related protein 4 (ANGPTL4), which does not contain a cationic ridge, cannot bind TLR4 in an ELISA, cannot stimulate cytokine release from macrophages, and cannot drive inflammation *in vivo*. These results validate Zuliani-Alvarez *et al.*'s hypothesis and demonstrate that the FReP family can be clearly segregated into TLR4 activators and non-activators via the presence or absence of a cationic ridge, respectively. In Figure 1.3, I provide a phylogenetic representation of the entire human FReP family annotated with proteins predicted to contain a cationic ridge and activate TLR4 (green box) and those with a known structure deposited in the PDB (grey circle; for PDB IDs, please refer to Section 2.1.1).

The angiopoietin-related proteins (ANGPTL1–ANGPTL7) are an interesting FReP sub-family which I will explore in some detail throughout this thesis. ANGPTL1, ANGPTL2, and ANGPTL6 are predicted TLR4 activators containing a cationic ridge, while ANGPTL3, ANGPTL4, ANGPTL5, and ANGPTL7 are non-activators. Note that ANGPTL8 exists, but it does not have an FBG domain so is not considered a FReP. The ANGPTLs contain an N-terminal coiled-coil domain and a C-terminal FBG domain connected by an unstructured linker. All ANGPTLs contain a signal peptide for secretion and are therefore extracellular and often glycosylated;

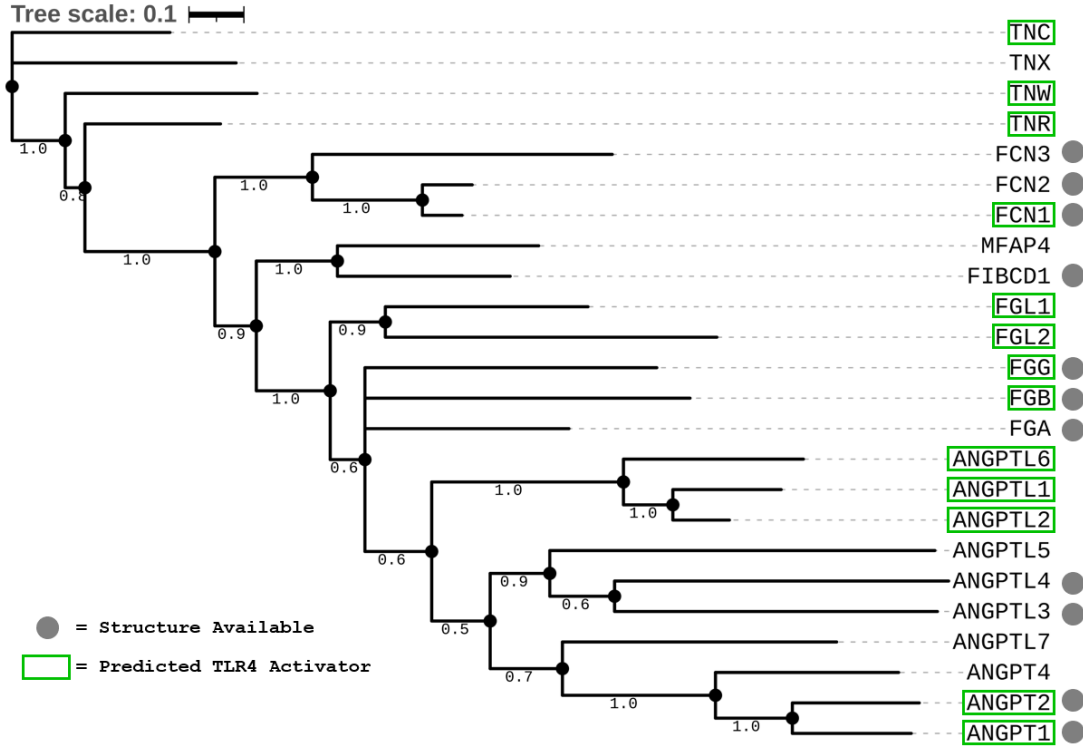


Figure 1.3 FReP Family Phylogeny. Phylogeny of human FRePs (protein names denote just the FBG domain). The MSA was generated in ICMPPro, the phylogenetic tree was calculated using MrBayes v3.2, and the figure was generated using iTOL [307, 308]. Predicted TLR4 activators are boxed in green, and FRePs with a previously reported structure are marked with a grey circle.

ANGPTLs are broadly expressed in the liver, hematopoietic system, and vasculature, where they often circulate systemically [309]. ANGPTLs can be proteolytically processed within the linker region to separate the N-terminal coiled-coil from the C-terminal FBG domain, representing a potential mechanism for the presentation of just the ANGPTL FBG domains as DAMPs [310, 311, 312, 313]. While structurally similar proteins from the angiopoietin (ANGPT) family (including the FRePs ANGPT1, ANGPT2, and ANGPT4) bind to and activate pro-angiogenic tyrosine kinase receptors Tie1 and Tie2, the ANGPTLs cannot bind Tie1 or Tie2 and have traditionally been considered orphan ligands, making the new hypothesis surrounding ANGPTL-mediated activation of TLR4 intriguing [314]. This is especially true since the ANGPTLs have been recognized as inflammatory mediators for some time—in fact, all ANGPTLs except for ANGPTL1 and ANGPTL8 have been reported as pro-inflammatory in some context [309]. ANGPTL2 is broadly pro-inflammatory and contributes to diabetes, atherosclerosis, and other chronic diseases and can drive synovial inflammation in RA [315, 316, 317]. ANGPTL4 stimulates intravasation of immune cells into lung tissue and can be up-regulated by $TGF\beta$ within the

pro-inflammatory cancer matrix [318, 319]. Finally, ANGPTL6 can directly stimulate wound healing [320], and up-regulation of ANGPTL6 in murine keratinocytes enhances psoriasis *in vivo* [321]. Many of these phenotypic effects could be mediated by TLR4; however, evidence suggests that ANGPTL1, ANGPTL2, ANGPTL5, and ANGPTL7 (two TLR4 activators and two non-activators) can bind leukocyte immunoglobulin-like receptor B2 (LILRB2; traditionally recognized as an immune-inhibitory receptor), and ANGPTL2, ANGPTL3, and ANGPTL4 can bind different integrins [322, 309], so the potential for receptor cross-talk is present. While the ANGPTLs were originally identified due to their role in modifying angiogenesis, their Tie1/Tie2-independent mechanism of action remains somewhat mysterious and their biological roles and contribution to disease are complex and often contradictory [309]. For example, ANGPTL1 is considered anti-angiogenic and anti-tumorigenic [323], while ANGPTL2, ANGPTL4, and ANGPTL6 are considered pro-angiogenic [309]. ANGPTL2 and ANGPTL4 especially are therapeutic targets and biomarkers in diverse cancers because of their pro-angiogenic, pro-metastatic, and pro-inflammatory properties across different cell types and tissues [324, 325, 326, 319]. Yet, despite these opposite phenotypes, ANGPTL1/2/6 are TLR4 activators and ANGPTL4 is not: clearly, then, there remains much to be learned about the ANGPTLs, particularly their regulation, context-dependent phenotypic effects, and receptor binding properties. In summary, then, the ANGPTLs are an intriguing family of disease-relevant, mysterious, and orphan DAMPs with opposing TLR4 binding properties and divergent biological roles, making them interesting orthogonal FRePs to study alongside the canonical FReP family member TNC^{FBG}.

1.3.3 FRePs as a Model DAMP

Based on a review of the FRePs in the context of the broader DAMP field, a few defining features encouraged me to focus on the FRePs as a model system to interrogate the molecular mechanism behind DAMP activation of TLR4:

- The FRePs are a large, structurally conserved, and newly identified family of both TLR4-activating and non-activating proteins, which are segregated by specific peptide-level differences that explain both TLR4 binding and TLR4 activation.
- The FRePs themselves can be purified as recombinant proteins from *E. coli*, simplifying protein production for biophysical assays.

- The 11 FReP structures in the PDB demonstrate that the FRePs are clearly amenable to structural characterization with X-ray crystallography. However, many of the TLR4-activating DAMPs, including the canonical family member TNC^{FBG} along with ANGPTL1, ANGPTL2, and ANGPTL6, remain unsolved, creating an opportunity to enrich the available structural data.
- The FRePs seem to engage TLR4 directly in the absence of a co-receptor, specifically based on ELISA binding data. However, the literature surrounding the TLR4:FReP interaction is missing gold standard biophysical experiments, meaning that new biophysical data or even structures of TLR4:FReP complexes would provide much needed clarity.
- The FRePs are fascinating from a biological standpoint, especially given their diverse, context-dependent roles in promoting tissue repair while also driving autoimmune diseases and cancers. However, therapeutic options are limited because of uncertainty surrounding the exact mechanism of action and the proper way to target pathological versus beneficial FReP functions. Therefore, biophysical and/or structural insight into the molecular mechanism of FReP action could provide new therapeutic angles, including perhaps the structure-guided design of a TLR4:FReP PPI inhibitor.

1.4 Motivation and Chapter Aims

My goal in Chapter 1 has been to highlight five fundamental conclusions that motivate and explain the rest of my thesis. First, TLRs, and particularly TLR4, are fundamental to both the activity and our understanding of the innate immune system. TLRs initiate a pro-inflammatory innate immune response in the presence of infection (via detection of PAMPs) or tissue damage (via detection of DAMPs) by activating master regulatory pathways like NF- κ B, MAPK, and IRF. Second, TLR4 forms direct, high-affinity, stable, and structurally tractable complexes with PAMPs. Third, the cellular data demonstrating the activation of TLR4 in the presence of diverse DAMPs is unequivocal. Fourth, across the entire DAMP landscape, biophysical data supporting direct, high affinity TLR4:DAMP complexes are sparse and incomplete, and there are no structures of any TLR bound to any DAMP. The biosensor data in the literature is scattered with non-specific artefacts, and I could find no evidence of co-elution of DAMPs and TLR4 during SEC. Put simply, there is a chasm between the biophysical evidence supporting TLR4:PAMP complexes

versus TLR4:DAMP complexes. Finally, the FRePs, particularly TNC^{FBG} and the ANGPTLs, are exemplar and well-characterized TLR4-activating DAMPs with a clear role in the pathogenesis of autoimmune diseases, making them an intriguing specific case to probe the general molecular mechanism behind DAMP detection.

Therefore, the stated aim of my four year DPhil project was to clarify the molecular mechanism of DAMP-activated TLR4 signalling by applying biophysical techniques to probe the interaction between TLR4 and FRePs. At the outset, I hypothesized that TLR4 could form a high-affinity, structurally tractable, and perhaps even therapeutically manipulable complex with one or more of the FRePs. However, as data emerged demonstrating the lack of a high-affinity interaction between TLR4 and the FRePs, I updated my priors and moved forward with cellular proteomics and gene-editing experiments based on the hypothesis that co-receptors and other accessory proteins are required for FReP-mediated activation of TLR4 in the absence of a 1:1 complex. In total, then, the work presented herein provides important insights into the mysterious interaction between TLR4 and FRePs in general and TLR4 and DAMPs more broadly.

In **Chapter 2**, I present the expression and purification of the FReP family of proteins using a dedicated disulfide-refolding *E. coli* cell line called CyDisCo. My work represents the first reported method for the soluble and high-yielding purification of the FRePs, which allowed me to solve three novel FReP structures by X-ray crystallography. In this Chapter, I also present a crystallographic fragment screen against TNC^{FBG} as well as a bioinformatic interrogation of the FRePs in the context of my new structures.

In **Chapter 3**, I present the expression and purification of TLR4, MD-2, and TLR4:MD-2 using *Drosophila* S2 insect cells. Driven by my success using S2 cells to purify TLR4, which proved challenging in other in-house systems, I modified previously reported methodology to generate FAS2FURIOUS, a medium-throughput, rapid, and robust platform for the parallel expression of up to 46 constructs in S2 cells.

In **Chapter 4**, I present a battery of biophysical assays which demonstrate no high-affinity complex between recombinant, bacteria-derived FRePs and recombinant, S2-derived TLR4 or TLR4:MD-2. I also present preliminary data showing that previously reported TLR4 co-receptors cannot rescue a TLR4:FReP complex.

In **Chapter 5**, I pivot away from reductionist biophysical assays and present an

optimized cross-linking proteomics workflow with primary human monocytes to explore the FReP-associated interactome at the cell surface in an unbiased manner. These experiments again do not support a direct interaction with TLR4, but instead reveal integrin-activating properties of the FRePs and suggest intriguing new hypotheses about potential FReP co-receptors.

In **Chapter 6**, I explore the potential for introducing an affinity tag directly onto TLR4 for more targeted studies using CRISPR-Cas9. Motivated by the success of others with similar knock-in approaches, I combine two previously reported technologies to generate GlowPAINT, a fast, moderate-throughput, and easy to deploy pipeline for the knock-in of affinity tags to endogenous genomic loci.

Penultimately, I present the tangentially related **Chapter 7** under the University's COVID-19 Student Policy. This Chapter details my efforts to express and purify large quantities of the COVID-19 Spike protein for serological testing during the pandemic. The protocol I helped to develop and deploy has since been used to supply over 400 mg of Spike protein and made an important impact on Oxford's COVID-19 response.

Finally, in **Chapter 8** I provide a high-level overview of my findings and present a principles-first hypothesis regarding the requirement of accessory proteins to fine-tune DAMP-mediated activation of TLR4.

Chapter 2

Biophysical Characterization of FReP Family Members

This chapter presents my work on individual members of the human FReP family, specifically focusing on the TLR4-activating members without a previously reported structure (TNC, ANGPTL1, ANGPTL2, ANGPTL6, FGL1, and FGL2). First, I describe a much improved strategy for soluble expression of the FRePs in *E. coli*, which yielded reasonable quantities of homogeneous, properly folded, and stable protein. Using a domain truncation approach, I report both the novel crystal structure of TNC and a crystallographic fragment screening campaign which revealed the first reported TNC-binding chemical matter. I next show the structure of a non-native FGL1 dimer, followed by the novel crystal structures of ANGPTL2 and ANGPTL6 using a thermal stabilization approach. These new structures facilitate a structural comparison of the human FReP family, containing twenty-four members and fourteen structures, which concludes the chapter and provides new insights into the structural and functional diversity of FBG domains. To clarify up-front, throughout this and subsequent Chapters, gene names indicate *only* the FBG domain from each target rather than the entire polypeptide; please refer below to Section 3.1.1 for target and sequence information for the FRePs I discuss herein.

2.1 Materials and Methods

2.1.1 FReP Library and Accession Numbers

Below are the UniProtID accession numbers, residue boundaries, and PDB ID's (if solved) for the full-length FBG domains used herein, either for purification or bioinformatics.

FReP Proteins Purified (UniProt, Residue, PDB):

- **Tenascin-C (TNC)**: P24821 Isoform 1, Ile1974–Ala2201, 6QNV. In-house cDNA contains Glu2008Gln variant (dbSNP:rs13321). For simplicity when discussing the truncated constructs in Section 2.6, I refer to the domain boundaries as Ile1–Ala228.
- **Angiopoietin-Related Protein 1 (ANGPTL1)**: O95841, Phe271–Asp491, N/A
- **Angiopoietin-Related Protein 2 (ANGPTL2)**: Q9UKU9, Asp269–His493, 6Y41
- **Angiopoietin-Related Protein 4 (ANGPTL4)**: Q9BY76, Ser179–Ser406, 6EUB
- **Angiopoietin-Related Protein 6 (ANGPTL6)**: Q8NI99, Thr251–Leu470, 6Y43
- **Fibrinogen-Like Protein 1 (FGL1)**: Q08830, Leu74–Ile312, N/A
- **Fibroleukin (FGL2)**: Q14314, Pro204–Pro439, N/A
- **Ficolin 1 (FCN1)**: O00602, Gln109–Ala326, 2JHM.

Other FReP Family Members (UniProt, Residue, PDB):

- **Tenascin-X (TNX)**: P22105, Gly4021–Asn4236, N/A
- **Tenascin-W (TNW) aka Tenascin-N (TNN)**: Q9UQP3, Gly1061–Gly1278, N/A
- **Tenascin-R (TNR)**: Q92752, Gly1129–Asn1344, N/A

- **Ficolin-2 (FCN2)**: Q15485, Gln96–Ala313, 49RJ
- **Ficolin-3 (FCN3)**: O75636, Leu84–Arg299, 2J5Z
- **Microfibril-associated glycoprotein 4 (MFAP4)**: P55083, Arg32–Ala255, N/A
- **Fibrinogen C-domain containing protein 1 (FIBCD1)**: Q8N539, Cys235–Arg458, 4M7H
- **Fibrinogen alpha chain (FGA)**: P02671, His623–Val864, 1FzD
- **Fibrinogen beta chain (FGB)**: P02675, Asn232–Phe488, 1FZA
- **Fibrinogen gamma chain (FGG)**: P02679, Gln170–Asn416, 1FID
- **Angiopoietin-Related Protein 3 (ANGPTL3)**: Q9Y5C1, Val237–Asp455, 6EUA
- **Angiopoietin-Related Protein 5 (ANGPTL5)**: Q86XS5, Arg141–Tyr383, N/A
- **Angiopoietin-Related Protein 7 (ANGPTL7)**: O43827, Gln122–Asp343, N/A
- **Angiopoietin 1 (ANGPT1)**: Q15389, Arg277–Asp497, 4EPU
- **Angiopoietin 2 (ANGPT2)**: O15123, Lys275–Asp495, 1Z3U
- **Angiopoietin 4 (ANGPT4)**: Q9Y264, Met282–Asp502, N/A

2.1.2 Transformation of *E. coli*

All *E. coli* cell lines used herein, including SHuffle[®], CyDisCo, BL21(DE3) Rosetta, BL21(DE3) Tuner, and Mach 1, were transformed using heat shock unless otherwise indicated. For each transformation, 40 μL of chemically competent *E. coli* were added to 1 μL of plasmid DNA at approximately $100 \text{ ng } \mu\text{L}^{-1}$ and incubated for 30 min on ice. The mixture was then shocked at 42°C for 45 seconds, returned to ice for 2 minutes, diluted with approximately 200 μL of 2X LB, and allowed to recover for 1 hour at 37°C . 75 μL of outgrowth was then plated onto dishes containing 1.5% LB-Agar and the appropriate antibiotic. For long (10 kb or higher), difficult to transform, or low-concentration plasmids, the outgrowth was centrifuged for 5

min at 4000 x g and the entire pellet was resuspended in 75 μ L of 2X LB before plating. Transformants were allowed to grow overnight at 37 °C before downstream use. During cloning, single colonies were picked; during expression, a streak of colonies were used.

2.1.3 Ligation Independent (LIC) Cloning

The SGC's LIC cloning methodology has been reported previously and is based on the generation of short, complementary, single-stranded 5'-tails by the 3'-to-5' exonuclease activity of T4 DNA polymerase [327, 328]. Briefly, in a 96-well plate, the targets of interest were amplified by Touchdown PCR using either Q5 DNA Polymerase (NEB) or Herculase II (Agilent) according to the manufacturer's recommendations, using primers containing a destination-vector-specific 12-16 bp overhang [329]. PCR products were visualized with agarose-gel electrophoresis and incubated with 1 μ L of *DpnI* (diluted 1:10 in NEB Buffer 2.1) for 1 hour at 37 °C before purification using the GenElute™ 96-well PCR Clean-Up Kit (Sigma) according to the manufacturer's instructions. The cleaned PCR products were digested with T4 DNA polymerase in the presence of either dGTP or dCTP, which defines the length of the ssDNA overhang generated by the digestion. The destination vectors were linearized with the appropriate restriction endonuclease (either *BsaI*, *BfuRI*, or *BseRI*) and the drop-out of a 2 kb *SacB* stuffer fragment was confirmed with agarose-gel electrophoresis. The linearized vectors were spin-purified on a mini-column according to the manufacturer's recommendations, and then digested with T4 DNA polymerase in the presence of the complementary nucleotide to the PCR product (either dGTP or dCTP). 1 μ L of T4 digested vector was incubated with 3 μ L of T4 digested PCR product for 30 min at room temperature to anneal before transformation into Mach 1 cells as before (see Section 2.1.2), except for the addition of 5% sucrose to the plates, which serves to remove any traces of the circular parental vector. The presence of cloned inserts was confirmed using the MyTaq™ Red Colony PCR Kit (Bioline) using vector-specific primers which sit outside of the open reading frame, followed by agarose-gel electrophoresis. Colony PCR was repeated until a colony with an insert of the expected length was found. 1 mL overnight cultures of the correct constructs were grown in deep-well 96-well blocks at 37 °C in LB supplemented with 50 μ g mL⁻¹ kanamycin (Kan). Glycerol stocks were prepared by 1:1 dilution of turbid overnight cultures with sterile 50% glycerol and frozen at -80 °C. Plasmid DNA was purified using a 96-well plate and vacuum manifold according to

the manufacturer’s instructions (Merck Millipore), followed by long-term storage at -20°C .

2.1.4 1 mL Small-Scale Test Expression in *E. coli*

The SGC’s small-scale test expression protocol has been reported previously and is based on the work of Page *et al.* [328, 330]. Briefly, constructs were transformed into BL21(DE3)-Rosetta-pRARE2 cells by heat shock and selected with $50\text{ }\mu\text{g mL}^{-1}$ Kan and $34\text{ }\mu\text{g mL}^{-1}$ chloramphenicol (Chlor). A streak of colonies from each construct was used to inoculate a 1 mL overnight culture in 2X LB with kan/chlor in a deep-well 96-well block. 20 μL of overnight culture was inoculated into 1 mL of TB with kan/chlor in 96-well blocks and grown at 37°C and 700 RPM for approximately 5 hours. Once an OD600 of 2-3 was achieved, the cultures were rested for 30 min at room temperature, induced with 0.1 mM IPTG, and expressed overnight at 18°C and 700 RPM. Pellets were harvested the next morning via centrifugation at $3000\times g$ for 20 min, mixed with 200 μL of lysis buffer containing 1 mg mL^{-1} hen egg-white lysozyme and 1X benzonase, incubated for 5 minutes at 1500 RPM on a MixMate plate shaker, and stored frozen at -80°C .

For test purification, the lysates were thawed, spun down at $3000\times g$ for 15 min, and the supernatants were incubated for one hour with 25 μL bed volume of pre-equilibrated Ni-NTA Agarose (BioRad). The bead/supernatant mixture was transferred to a 96-well deep-well filter plate (Merck Millipore) and beads were harvested via centrifugation for 1 min at $200\times g$. Flow-through was discarded and the resin was washed three times using wash buffer containing 40 mM imidazole and gentle centrifugation as before. Bound proteins were eluted by 10 min incubation with 50 μL elution buffer containing 500 mM imidazole and centrifugation for 2 min at $200\times g$, with eluate collected in a 96-well V-bottom plate. 15 μL sample of eluates were analyzed via reducing SDS-PAGE, with putative target bands confirmed by in-gel tryptic digest MSMS. The base buffer for the purification process contained 50 mM HEPES $\text{pH} = 7.5$, 500 mM NaCl, 5% glycerol, 10 mM imidazole, and 0.5 mM TCEP.

2.1.5 50 mL Medium-Scale Test Expression of His6-Tagged Proteins

The following buffers were prepared as follows and filter sterilized before use. **Base Buffer** = 50 mM HEPES pH = 7.5, 300 mM NaCl, 5% glycerol. **Lysis Buffer** = Base Buffer + 1 mg mL⁻¹ hen egg-white lysozyme + 1X benzonase + 1X Merck EDTA-free protease inhibitors. **Wash Buffer** = Base Buffer + 10 mM imidazole. **Elution Buffer** = Base Buffer + 150 mM imidazole.

My 50 mL test expression protocol is modified from an unpublished method called “PREPX” developed by Dr. Michael Fairhead and further optimized by Dr. Sabrina MacKinnon. FReP constructs were transformed into chemically competent BL21(DE3) Tuner cells containing CyDisCo plasmid pMJS226 or pMJS205, or SHuffle[®] cells containing the pRARE2 plasmid, and selected on kan/chlor plates as before. 2 mL overnight cultures were used to inoculate 50 mL of TB-PLUS (1X TB + 15 mL L⁻¹ glycerol + 0.01% Antifoam + 1 mM MgSO₄ + 10 mM (NH₄)₂SO₄ + 0.5% D-glucose + 1X *E. coli* trace metals) spiked with 50 µg mL⁻¹ Kan and 34 µg mL⁻¹ Chlor in 250 mL sterilized baffled flasks. Cultures were grown at 30 °C and 250 RPM in an InforsHT incubator until OD600 reached 2-3, at which point the temperature was dropped to 18°C and expression was induced with 0.5 mM IPTG and allowed to continue overnight.

Pellets were harvested by centrifugation at 4000 x *g* for 15 min, re-suspended in lysis buffer containing 1 mg mL⁻¹ lysozyme, and incubated for 30 min at room temperature with gentle agitation. Lysates were briefly vortexed and 1% Triton X-100 (Sigma) was added before storage at -80°C overnight. Lysates were thawed in a room temperature water bath and clarified by centrifugation at 4°C for 45 minutes at 4000 x *g*. Clarified supernatants were applied to pre-packed 1 mL pre-equilibrated TALON columns (TakaraBio) by gravity flow and washed with 70 CV of base buffer followed by 70 CV of wash buffer. Bound protein was eluted with 2.5 mL of elution buffer directly onto a PD-10 desalting column (Sephadex G-25 Resin, BioRad), which was eluted with 3.5 mL of base buffer. Note that I observed precipitation of the FRePs in high [imidazole] and therefore always desalted the protein immediately following elution. 20 µL of the desalted eluates were analyzed with reducing SDS-PAGE followed by MSMS, and the A280 was checked with a NanoDrop (ThermoFisher) to estimate yield.

2.1.6 Purification Scheme 1 (MBP/SHuffle[®])

The following buffers were prepared as follows and filter sterilized before use. **Base Buffer** = 50 mM HEPES pH = 7.5, 300 mM NaCl, 5% glycerol. **Lysis Buffer** = Base Buffer + 1X benzonase + 1X Merck EDTA-free protease inhibitors. **Ion-Exchange Buffer A** = 20 mM HEPES pH = 8.0. **Ion-Exchange Buffer B** = 20 mM HEPES pH = 8.0, 1 M NaCl. **Storage Buffer** = 20 mM HEPES pH = 7.5, 250 mM NaCl, 2% glycerol.

MBP-FReP constructs were transformed into chemically competent SHuffle[®] cells containing pRARE2 via heat shock. Overnight cultures were grown in 1X TB + 50 $\mu\text{g mL}^{-1}$ Kan + 34 $\mu\text{g mL}^{-1}$ Chlor in 250 mL baffled flasks at 37°C with 250 RPM shaking. 12 mL of overnight culture was used to inoculate each of up-to 12X 3L sterilized baffled flasks containing 1L of TB-PLUS (prepared as in Section 2.1.5) with 50 $\mu\text{g mL}^{-1}$ Kan + 34 $\mu\text{g mL}^{-1}$ Chlor. The flasks were covered with air-porous seals (AirOtop[™], Thomson) and grown at 37°C with 200 RPM shaking to an OD600 between 2-3. The temperature was dropped to 18°C and expression was induced overnight with the addition of 0.5 mM IPTG. Pellets were harvested by centrifugation for 15 minutes at 4000 x *g* (Beckman Coulter Avanti JXN-26 with JA 8.1 rotor) and then stored indefinitely in plastic bags at -80°C.

For purification, pellets were re-suspended in lysis buffer (3 mL buffer per 1 g wet cell weight) and broken by sonication (15 minutes, 5 seconds on, 10 seconds off, 35% amplitude). Lysates were clarified by centrifugation for 1 hour at 25,000 x *g* in a JA25.5 rotor followed by application to a 2 mL bed volume amylose affinity column (NEB). Column was rinsed with at least 50 CV of base buffer and the re-suspended in 2 mL base buffer containing 2 mg of Tobacco Etch Virus nuclear-inclusion-a endopeptidase (TEV) protease. For on-column cleavage, the resin/TEV mixture was incubated overnight at 4°C with end-over-end mixing. The following morning, the resin/protein mixture was re-applied to an empty column and the flow-through plus a 10 mL wash with base buffer was collected. The cleaved FReP was diluted with ion-exchange buffer to less than 30 mM NaCl and the pH was adjusted to 8.0 as necessary. For a final polishing step, the protein was applied to a 1 mL ion-exchange column (either MonoQ or MonoS 5/50, GE Healthcare) using an AKTA Express (GE Healthcare). The FReP was eluted with a increasing gradient of NaCl in ion-exchange buffer B, and the peak fractions were analyzed using SDS-PAGE. FReP-containing fractions were pooled and buffer exchanged into storage buffer by sequential dilution using a 10K MWCO Amicon centrifugal concentrator. Finally,

the FReP was concentrated down to approximately 3 mg mL^{-1} for biophysical experiments or up to 10 mg mL^{-1} for crystallization. FReP identity was consistently confirmed via in-gel tryptic digest MSMS and intact mass analysis. Any protein not used immediately was flash frozen in liquid nitrogen and stored at -80°C .

2.1.7 Purification Scheme 2 (His6/CyDisCo)

The following buffers were prepared as follows and filter sterilized before use. **Base Buffer** = 50 mM HEPES pH = 7.5, 300 mM NaCl, 5% glycerol. **Lysis Buffer** = Base Buffer + 1X benzonase + 1X Merck EDTA-free protease inhibitors. **Wash Buffer** = Base Buffer + 10 mM imidazole. **Gel Filtration Buffer** = 20 mM HEPES pH = 7.5, 250 mM NaCl, 2% glycerol.

His6- or His6-Twin-Strep-tagged FRePs were transformed into chemically competent BL21(DE3) Tuner cells containing the pMJS226 CyDisCo plasmid. Protein expression and cell harvest was carried out exactly as described in Section 2.1.6 except for maintaining the initial growth temperature at 30°C to preserve the folding of CyDisCo components.

For purification, pellets were re-suspended in lysis buffer (3 mL buffer per 1 g wet cell weight) and broken by sonication (15 minutes, 5 seconds on, 10 seconds off, 35% amplitude). Lysates were clarified by centrifugation for 1 hour at $25,000 \times g$ in a JA25.5 rotor followed by application to a 2 mL bed volume TALON column. The column was washed with 50 CV base buffer and 50 CV wash buffer before a 10 min incubation with 7.5 mL elution buffer. Eluate was immediately desalted back into base buffer using 3X PD-10 columns to remove imidazole. To cleave the tag, the FReP was incubated with 1:10 (w:w) GST-tagged TEV protease at room temperature with gentle agitation. Cleavage was periodically checked using intact mass, with complete cleavage typically being observed after 1.5 to 2 hours. The cleaved FReP was concentrated to less than 5 mL using a 10K MWCO Amicon centrifugal concentrator and applied using an AKTA Express (GE Healthcare) to a 124 mL SEC column (Superdex 75 HiLoad 16/60, GE Healthcare) pre-equilibrated in gel filtration buffer. FRePs were loaded and eluted at a constant 1 mL/min and collected in 1.8 mL fractions, which were analyzed by reducing SDS-PAGE. Monodisperse, FReP-containing fractions were pooled and concentrated using a 10K MWCO Amicon centrifugal concentrator to approximately 3 mg mL^{-1} for biophysical experiments or up to 30 mg mL^{-1} for crystallization. FReP identity was consistently

confirmed via in-gel tryptic digest MSMS and intact mass analysis. Any protein not used immediately was flash frozen in liquid nitrogen and stored at -80°C without any loss in activity or precipitation upon thawing.

2.1.8 Crystallization Trials

Throughout, sitting-drop vapour diffusion experiments were performed in 96-well, three-sub-well SWISSCI plates (SWISSCI AG). 20 μL of mother liquor from commercially available coarse screens was added to the SWISSCI master wells, followed by the preparation of 2:1, 1:1, and 1:2 (protein:mother liquor) dilutions within the three sub-wells by TTP Labtech's Mosquito[®] liquid handling robot. Plates were incubated at either 4°C or 20°C and regularly imaged using a Formulatrix[®] automatic inspection system (Rigaku). Crystals were managed, scored, and evaluated using the MATLAB application TeXRank [331]. A variety of coarse screens were regularly used, most notably Morpheus [332], MidasPLUS [333], JCSG/PACT [334], LMB [335], and the Ligand Friendly Screen [336].

The structure of TNC was solved using a crystal from the Morpheus Screen, which appeared after 28 days at 20°C with TNC at 17 mg mL^{-1} mixed with screen condition C9: 0.09 M NPS + 0.1 M Buffer System 3 pH = 8.5 + 50% Precipitant Mix 1. The structure of ANGPTL2 was solved using a crystal from the Morpheus Screen, which appeared after 14 days at 20°C with ANGPTL2 at 6 mg mL^{-1} mixed with 25 mM CaCl_2 and screen condition A12: 0.09 M Divalents + 0.1 M Buffer System 3 pH = 8.5 + 50% Precipitant Mix 4. The structure of ANGPTL6 was solved using a crystal from the Morpheus Screen, which appeared after 7 days at 20°C with ANGPTL6 at 10 mg mL^{-1} mixed with 25 mM CaCl_2 and screen condition B5: 0.09 M Halogens + 0.1 M Buffer System 2 pH = 8.5 + 50% Precipitant Mix 1. The structure of the FGL1-c011 dimer was solved using a crystal from a homemade follow-up screen based on Morpheus Conditions D05 and F09. The crystal appeared after 7 days at 20°C with FGL1 at 29 mg mL^{-1} mixed with screen condition D10: 0.12 M Monosaccharides + 0.1 M Buffer System 3 pH = 8.5 + 50% Precipitant Mix 4.

2.1.9 Data Collection and Processing

All crystallographic data was collected at Diamond Light Source using beamline i04, i04-1, and i03. Diffraction data were automatically indexed, integrated, and scaled

using the Xia2 pipeline [337]. All downstream analysis was performed using the *PHENIX* software suite [338]. All structures were phased by molecular replacement using *Phaser* [339], with FGG (PDB: 1FIC) used as the search model for TNC. The final structure of TNC (PDB: 6QNV) was subsequently used as the search model to phase ANGPTL2 and ANGPTL6. The final structure of ANGPTL6 (PDB: 6Y43) was used as the search model to phase the structure of the FGL1 dimer. Structures were refined as appropriate with cycles of automatic refinement in *phenix.refine* coupled to manual rebuilding in Coot [340, 341]. Structural models were validated and improved with guidance from *MolProbity* and the wwPDB Validation Report [342, 343]. All structures were deposited to the PDB after internal peer-review by senior SGC crystallographers, either Dr. Gustavo Bezerra or Dr. Tobias Krojer.

2.1.10 TNC XChem Fragment Screen

TNC-c042 was concentrated to 30 mg mL⁻¹ and mixed 100 nL:100 nL in all sub-wells of SWISSCI plates using the custom fine screen JAC-MorpheusC09-z001, which sampled the chemical space around Morpheus C09 by varying the pH of Buffer System 3 from 7.0 to 9.0 and the concentration of Precipitant Mix 1 from 44 to 56%. After 14 days, approximately 150 crystals per plate were obtained. These crystals were used to complete a full fragment screen at the XChem Facility, Diamond Light Source Beamline i04-1, based on standard methodologies freely shared on Diamond’s website [<https://www.diamond.ac.uk/Instruments/Mx/Fragment-Screening.html>].

Briefly, using an Echo[®] Acoustic Liquid Handler, 20 nL of the DSI-Poised 1 Fragment Library (Enamine) was added to crystal-containing drops, for a final concentration of 10% DMSO. The fragment/crystal mixture was incubated for 2 hours at 20 °C before mounting and flash freezing in liquid nitrogen using the Shifter, a robotic crystal-mounting platform developed by Wright *et al.* [344]. Diffraction data were collected at Diamond Light Source Beamline i04-1 in automatic mode followed by indexing, integration, and scaling using the automated Xia2 pipeline [337]. Density maps were imported into the XChemExplorer software suite developed by Krojer *et al.*, which provided support for all downstream data processing [345]. Molecular replacement and initial refinement was completed by DIMPLE using the apo-TNC structure (6QNV) as a search model [346]. Fragment binding events were identified by PanDDA and manually annotated [347], followed by refinement using Coot and REFMAC within XChemExplorer [341, 348]. Refined, fragment-bound structures were exported, deposited to the PDB, and visualized using Chimera [349].

2.1.11 Differential Scanning Fluorimetry

DSF experiments were carried out essentially as reported by Niesen *et.al.* and Boivin *et al.* with a few modifications [350, 351]. For additive screening, in a 96-well v-bottom white qPCR plate, 10 μ L of 2X HEPES-buffered saline (HBS) was mixed with 4 μ L of ddH₂O, 2 μ L of 10X additives from the Morpheus Additive Screen, and 2 μ L of the target protein at a stock concentration of 3 mg mL⁻¹. The protein/additive mixture was incubated for 15 min at room temperature with gentle agitation, followed by the addition of 2 μ L 10X SYPROTM Orange (ThermoFisher) freshly prepared from a 5,000X stock in ddH₂O. The samples were equilibrated for 10 min in the dark before collection of thermal denaturation curves using a Mx3005p RT-PCR machine (Stratagene). The samples were heated from 25 °C to 95 °C over 70 minutes, with a ramp of 1 °C/ min. T_m and R² values were calculated by non-linear fitting in GraphPad Prism using the ideal monophasic ProteoPlex model from Chari *et al.* [352].

Custom equation in Prism (can be copied and pasted directly into the User-Defined Equation editor):

$$Y = \frac{\text{maxflour} * (\exp(-((\text{DeltaH}(x - \text{MeltT})) / (8.3145 * x * \text{MeltT}))))}{[1 + (\exp(-((\text{DeltaH}(x - \text{MeltT})) / (8.3145 * x * \text{MeltT}))))]}$$

Where maxflour is the maximum normalized fluorescence, DeltaH is the unfolding enthalpy, MeltT is the melting temperature expressed in Kelvin, and x is the assay temperature expressed in Kelvin. The pre-specific rules for initial values and fitting were as follows:

maxflour = 1000

DeltaH = 100,000 (to be fit; unconstrained)

MeltT = 330 (to be fit; must be between 0 and 400)

For Ca²⁺ binding curves, an 11-point serial dilution of CaCl₂ was first prepared in ddH₂O at 10-fold the desired final concentration. In a 96-well v-bottom white qPCR plate, 10 μ L of 2X HEPES-buffered saline (HBS) was mixed with 4 μ L of ddH₂O, 2 μ L of the 10X CaCl₂ serial dilution, and 2 μ L of the target protein at 3 mg mL⁻¹. The curves were prepared in technical triplicate. After a 15 min incubation, SYPROTM Orange was added and melting curves were collected as above.

T_m values were determined using non-linear fitting as above, followed by subtraction from the calcium-free T_m to determine ΔT_m values, which were plotted against $\log[Ca^{2+}]$ and fit using a one-site saturation binding model in GraphPad Prism.

2.1.12 General Mass Spectrometry

Overnight, in-gel tryptic digest MSMS analyses were routinely performed by Dr. Rod Chalk, Tiago Moreira, and Oktawia Borkowska according to published methods on a Bruker *Esquire* HTC ESI-ion trap mass spectrometer [353], with peptide identification and protein matching completed with the MASCOT search engine [354]. Intact mass analyses were completed essentially as reported in Chalk *et al.* [353], with 1 μ g of target protein diluted in 60 μ L of 0.1% formic acid, passed through an HPLC, and injected onto an Agilent 6530 ESI-QTOF mass spectrometer, followed by spectral deconvolution and analysis using MassHunter (Agilent).

2.1.13 Disulfide Mapping Using Intact Mass

To analyze the number of disulfide bonds present in the purified FRePs, three samples were prepared for intact mass analysis. The first, “as purified”, sample, was prepared and analyzed exactly as in Section 2.1.12. The second, “+IA”, sample was prepared by denaturing 5 μ g of protein for 1 hour in 50 μ L denaturing buffer (8 M urea in 50 mM ammonium bicarbonate pH = 8.0), followed by the addition of 25 mM freshly prepared iodoacetamide and another 1 hour incubation. 5 μ L of the sample was used for intact mass analysis as before. The final sample, “+IA/TCEP”, was prepared like +IA, except for the presence of 10 mM TCEP in the denaturing buffer. Each alkylation event corresponds to a mass shift of 57-58 Da, representing the capping of a single Cys. Cysteines involved in disulfide bonds will only be capped in the +IA/TCEP sample, while free cysteines will be capped in both the +IA and +IA/TCEP samples. Only FReP batches with two disulfide bonds were used for downstream biophysical studies.

2.1.14 Native MS

75 μ g of purified protein at 1 mg mL⁻¹ was desalted into 50 mM ammonium acetate pH = 6.5 using three consecutive exchanges on ZEBRA Micro Spin Desalting Columns

(Thermo Fisher) according to the manufacturer’s recommendations. Native MS spectra collection was performed by either Dr. Rod Chalk, Tiago Moreira, or Oktawia Borkowska, followed by manual data analysis using MassHunter and custom *Excel* scripts. For collection, desalted samples were directly injected into an Agilent 6530 ESI-QTOF at a constant rate of 6 $\mu\text{L}/\text{min}$ using a 1.0 mL gas-tight displacement syringe (Hamilton) connected to the ionization source with 0.005 inch PEEK capillary tubing. The QTOF was operated in positive ion mode using the 1 GHz detector mode, a scan range of 100-20,000 m/z , and a fragmentation voltage of 430 V.

2.2 Screening Fusion Tags for Solubility

The previously reported expression strategy for FRePs relied on insoluble expression of His6-tagged, full-length FBG domains as inclusion bodies in *E. coli*, followed by a tedious eight-day refolding protocol [142]. Disulfide-bonded proteins often express as insoluble aggregates in *E. coli* due to the reducing nature of the *E. coli* cytosol, which prevents disulfide-bond formation [355]. Despite successful implementation of refolding protocols to purify many disulfide-bonded proteins, these approaches require target-specific optimization and often yield small amounts of functional final product [356]. This proved true in my hands, with in-house refolded yields for the FRePs never exceeding a meager 0.20 mg L^{-1} of culture. Given that my primary objective was large-scale production of multiple FRePs for structural characterization, this methodology was insufficient; therefore, I started my work on the FReP family by exploring approaches for rescuing disulfide bond formation to facilitate soluble expression of the FRePs in *E. coli*.

Using ligation-independent cloning methodology (LIC; see 2.1) and guided by Costa *et al.*’s detailed review of recombinant tags [357], Ms. Leela Shrestha, a scientist within the SGC’s Biotechnology Group, and I cloned full-length FRePs (accession numbers and boundaries in Section 2.1.1) in-frame with fusion tags reported to enhance protein solubility and/or catalyze disulfide bond formation. All vectors also incorporated a His6-tag for pull-down using immobilized metal affinity chromatography (IMAC), as well as a TEV cleavage site for eventual tag removal. This initial construct library screened the following six tags, all placed at the N-terminus to avoid interference with C-terminal residues predicted to interact with TLR4 (see Section 2.7):

- **OmpA Fusion** [358, 359]

The OmpA fusion contains an N-terminal signal peptide derived from *E. coli* outer membrane protein A, which directs translated polypeptides from the cytosol into the oxidizing periplasm of *E. coli*.

- **DsbC Fusion** [360]

In addition to serving as a native disulfide isomerase in the periplasm, *E. coli* disulfide bond isomerase C (DsbC) is frequently employed as a fusion partner to promote folding of disulfide-rich proteins (DRPs). My vector for DsbC fusion also directs the nascent polypeptide to the periplasm, essentially giving the target a double-dose of dedicated disulfide-refolding enzymes.

- **TRX Fusion** [361]

The thioredoxin (TRX) tag is a 11.7 kDa protein derived from *E. coli* thioredoxin A. The TRX tag directly catalyzes the formation of disulfide bonds via a redox-active catalytic dithiol in the active site of the enzyme, and also provides a stabilizing affect due to its high thermal stability (over 80 °C).

- **SUMO Fusion** [362, 363]

The small ubiquitin-like modifier (SUMO) protein is a small (11.5 kDa), covalent, reversible PTM added to specific eukaryotic proteins to direct cell processes including apoptosis and cell cycle progression. When used as a recombinant fusion partner, the SUMO tag can enhance protein stability and solubility.

- **MBP Fusion** [364, 365, 366]

The 42.5 kDa *E. coli* maltose binding protein (MBP) enhances the solubility of fusion partners, although the precise mechanism remains unclear. Reports suggest that MBP-tags both actively recruit endogenous chaperones like GroEL and passively interact with partially folded fusion partners to prevent aggregation. MBP-tags also interact strongly with amylose, allowing simple purification on amylose-coupled resins.

- **SPB SET1 Fusion**

The streptavidin peptide binding (SPB) and solubility-enhancing tag 1 (SET1) is a proprietary fusion partner (VeriFlex™, Agilent). The SET1 portion of the tag is based on the C-terminal portion of the T7 phage gene 10B sequence and introduces a solubilizing net-charge of -6, while the SPB portion can be used for affinity purification on streptavidin-conjugated resin.

To see if any of these fusion partners improved FReP solubility, Ms. Leela Shrestha and I first completed a 1 mL test expression of the construct library in the standard BL21(DE3) *E. coli* expression strain, followed by a one-step IMAC purification and analysis by SDS-PAGE and in-gel tryptic digest MSMS (see Section 2.1 for details). As Figure 2.1A demonstrates, only MBP or DsbC tags rescued soluble expression, and even this improvement was highly target dependent. FCN1, TNC, and ANGPTL2 were solubilized by MBP, while FCN1, ANGPTL1, and ANGPTL6 were solubilized by DsbC. Given that the MBP constructs expressed at a higher level than the DsbC constructs and that the MBP tag facilitated the soluble expression of the canonical family member TNC, I initially moved forward with optimizing the expression of MBP-tagged FRePs.

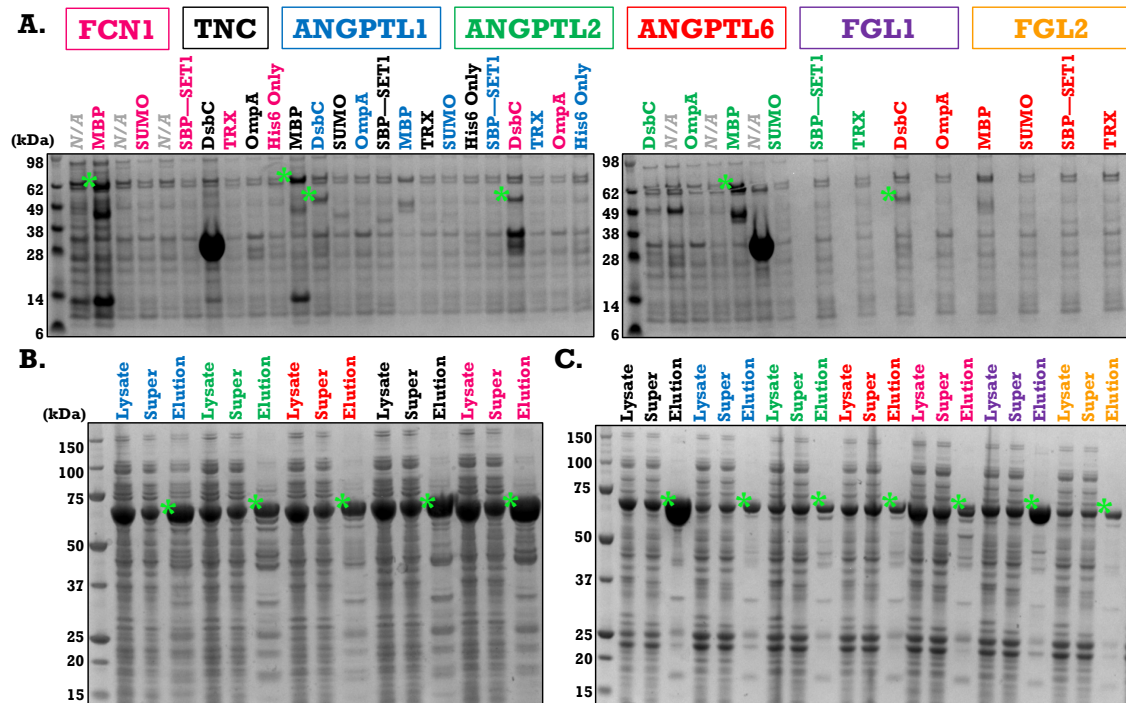


Figure 2.1 Rescuing FReP Solubility in *E. coli*. **A.** 1 mL test expression of tagged FRePs in standard BL21(DE3) *E. coli* expression strain. **B.** 50 mL test expression of MBP-tagged FRePs in SHuffle[®] cells with an oxidizing cytosol. **C.** 50 mL test expression of MBP-tagged FRePs co-expressed with first generation CyDisCo components from pFH255 in BL21-DE3 cells. Throughout, green asterisks denote bands which had their identity confirmed by in-gel tryptic digest MSMS, while N/A denotes bands from non-FReP targets which will not be discussed herein.

2.3 Testing Disulfide-Refolding *E. coli*

2.3.1 Theory of Oxidizing *E. coli* Strains

Although the MBP-tag solubilizes the FRePs in BL21(DE3) cells, I hypothesized that transitioning into engineered strains designed to enhance disulfide bond formation could increase expression levels. These specialized cell lines eventually proved transformative for the FRePs (see Section 2.6), so a brief detour from my experimental work and into the history of these systems will clarify subsequent sections.

Two common commercially available disulfide-refolding strains, both containing an oxidizing cytosol due to the $\Delta trxB \Delta gor ahpC^*$ genotype [367, 368, 369], are OrigamiTM (Merck Millipore) and SHuffle[®] (New England Biolabs). SHuffle[®] and OrigamiTM cells are identical except for the genetic insertion of a signal-sequence-free DsbC in SHuffle[®], which improves final yields [370].

Despite representing a massive step forward for the production of DRPs in *E. coli*, neither SHuffle[®] nor OrigamiTM cells are perfect. Because of significant genetic manipulations and the disruption of core reducing pathways, these lines grow slower and express less protein than BL21(DE3) strains, meaning significant target-specific optimization, including the co-expression of additional chaperones, is often required [371]. Therefore, a system for cytosolic folding of DRPs in reducing BL21(DE3) lines was developed and coined CyDisCo (**CY**toplasmic **D**isulfide bond formation in *E. CO*li) [372]. The term “CyDisCo” encompasses a variety of vectors that facilitate the over-expression of specific disulfide-refolding co-chaperones that efficiently fold and solubilize DRPs even in an otherwise reducing cytosol.

The optimized CyDisCo approach relies on over-expression of the sulfhydryl oxidase Erv1p and a disulfide isomerase (either PDI or DsbC) from the Chlor^R pLysS backbone [373]. CyDisCo plasmids utilize a p15A origin of replication, making them compatible for co-expression with a targets cloned into standard backbones with f1 origins [374]. The first generation CyDisCo plasmid, called pFH255, contains Erv1p and DbsC under a titratable, arabinose-inducible promoter. The next generation plasmids, pMJS205 and pMJS226, contain Erv1p and PDI under the control of an IPTG-inducible T7 lacO promoter. pMJS226 additionally contains the *lac* repressor (*lacI*), which reduces background expression of CyDisCo components before induction.

2.3.2 Evaluation of SHuffle[®] and CyDisCo Cells

As for the deployment of these specialized strains in my work, I initially tested just MBP-tagged FRePs using both SHuffle[®] and CyDisCo pFH255 cells. Figure 2.1B and Figure 2.1C demonstrate that both SHuffle[®] and CyDisCo cells rescued the expression of all MBP-tagged FRePs, including targets like ANGPTL1 and ANGPTL6 which failed in the original test expression (see Figure 2.1A). I also observed that SHuffle[®] outperformed CyDisCo pFH255 cells based on the total yield of protein in the elution fractions. Therefore, MBP-tagged FRePs expressed in SHuffle[®] cells represented a tenable approach for preliminary scale-up experiments.

More broadly, the results in Figure 2.1 demonstrate that solubilizing tags can synergize with disulfide-refolding cell lines, suggesting that disulfide-containing proteins should always be tested in dedicated cell lines like SHuffle[®] or CyDisCo. In fact, these results called into question my original selection of the MBP-tag based on observed expression levels in reducing BL21(DE3) cells that are demonstrably inferior for DRPs. Therefore, I repeated a similar experiment to that presented in Figure 2.1C, but this time using the next generation CyDisCo plasmids pMJS205 and pMJS226. This experiment allowed me to both test for alternatives to the MBP tag and also to compare pMJS205 and pMJS226 head-to-head.

Figure 2.2 demonstrates that CyDisCo cells can solubly express FReP constructs with tags other than MBP, including TRX, SBP-SET1, MBP and DsbC. More impressively, for every FReP tested, CyDisCo cells efficiently express just His6-tagged constructs, a somewhat surprising result given my initial hypothesis about the requirement for a solubilizing tag. This experiment also demonstrates that CyDisCo pMJS226 outperforms pMJS205 across all targets and tags, a result which is particularly pronounced for His6-TNC, His6-ANGPTL6, and MBP-ANGPTL2. Therefore, for all downstream CyDisCo expressions, I used pMJS226. The ability to express His6-tagged FRePs with CyDisCo pMJS226 greatly simplified purification (see Section 2.4) and eventually led to novel FReP crystal structures (see Section 2.6 and 2.9).

2.4 Large-Scale FReP Purification Strategies

After screening my full-length FReP construct library using six different fusion partners in five different *E. coli* cell lines, I identified (1) an MBP-tag in SHuffle[®] cells

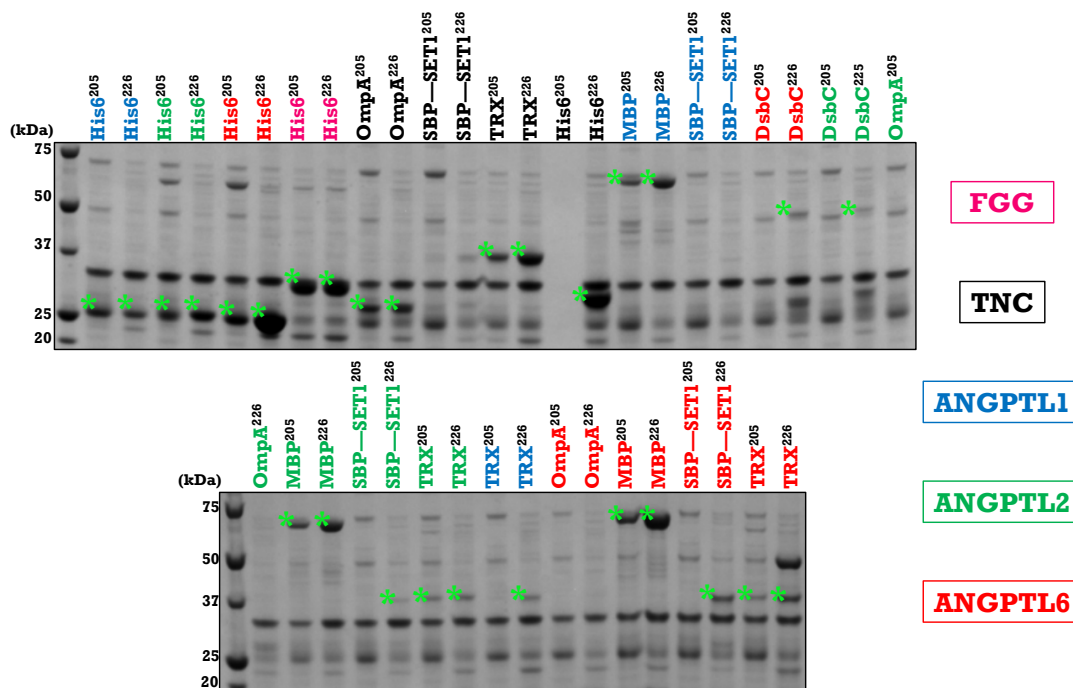


Figure 2.2 CyDisCo Facilitates His6-FReP Expression. IMAC elution fractions from a 30 mL test expression and purification of fusion-tagged FRePs in BL21(DE3) cells co-expressing CyDisCo components from either pMJS205 or pMJS226. Green asterisks denote bands which were confirmed by in-gel tryptic digest MSMS.

and (2) a His6-tag in CyDisCo pMJS226 cells as viable strategies for soluble FReP expression. This section details my optimization of large-scale FReP purification using either Scheme 1 (MBP/SHuffle[®]) or Scheme 2 (His6/CyDisCo), leading to the eventual conclusion that Scheme 2 is superior for quickly purifying FRePs in large quantities.

2.4.1 Purification Scheme 1: MBP-tag in SHuffle[®] Cells

My first dozen large-scale FReP preps used the MBP/SHuffle[®] approach detailed in Figure 2.3A. Briefly, I settled on an amylose affinity capture step followed by on-column cleavage to remove the MBP-tag, with a final polishing step using ion-exchange chromatography (IEX). However, despite successfully obtaining pure protein with Scheme 1 (Figure 2.3D), the approach had some fundamental limitations:

1. The MBP tag stabilizes misfolded protein.

Figure 2.3B demonstrates a generic problem uncovered for all FRePs purified

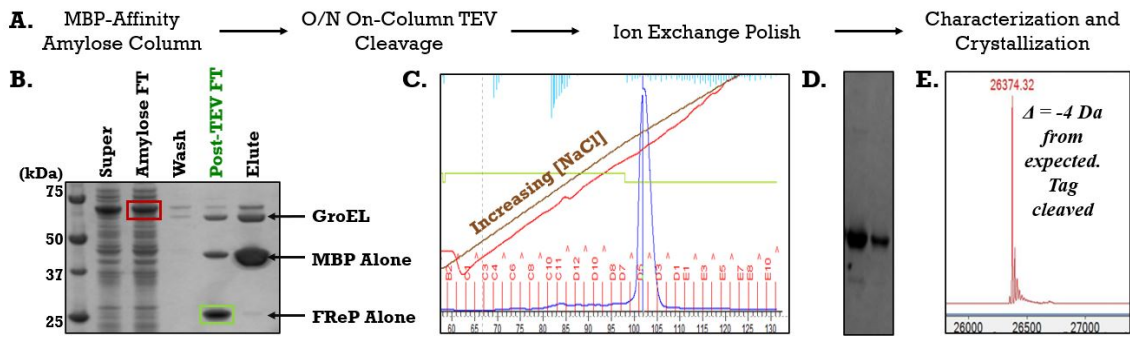


Figure 2.3 Purification Scheme 1 (MBP/SHuffle®). **A.** Optimized purification Scheme 1, using TNC as an example. **B.** SDS-PAGE showing on-column TEV cleavage to remove the MBP tag. The red box highlights unbound protein in the flow-through from the amylose affinity step, while the green box identifies the cleaved target of interest. **C.** A280 Chromatogram from a representative ion-exchange polish step, showing that a high-performance MonoS column can remove MBP and GroEL contaminants. **D.** SDS-PAGE analysis of the MonoS peak from C., showing the representative final purity of Scheme 1. **E.** Representative intact mass analysis which conclusively identifies the target and implies the presence of two disulfide bonds.

with Scheme 1: almost the entirety of the soluble MBP-tagged protein escapes the first step affinity column. Therefore, I hypothesize that most of the soluble MBP-tagged protein is misfolded or aggregated and therefore unsuitable for use. The data indicate that the MBP-tag is so effective at solubilizing its fusion partners that small-scale test expressions are prone to false-positives. As a result, my final purified yields were significantly lower than the test expression results implied. Despite many attempts, including screening for stabilizing additives in the lysis buffer and massively increasing the solid-phase bed volume of the affinity resin, this flow-through problem held true for all targets, suggesting a fundamental pathology with the MBP-tag itself.

2. Untagged FRePs themselves bind to IMAC columns.

Likely because of a conserved cation-binding loop (see Section 2.10 for the architecture of this loop bound to Ca^{2+}), I observed that all FRePs bound relatively tightly to both Nickel and Cobalt affinity matrices, even without a His6-tag, preventing the use of an IMAC rebind step to remove the His6-MBP tag following TEV cleavage. Amylose rebind steps did work; however, because of the high affinity of maltose for MBP, the MBP-tag can only be used one time during a purification [375]. As a result, I settled on a high-purity amylose affinity first step followed by an on-column overnight digest with TEV to release the tag-free FReP. This gave higher purity than an IMAC first step

followed by an amylose rebind to remove the tag, especially with regards to GroEL contamination which is common when purifying from *E. coli* using IMAC [376].

3. The MBP-tag is sticky and difficult to remove.

As Figure 2.3B demonstrates, even after an on-column cleavage of the MBP-tag, some of the free MBP (which should bind to the resin) escaped and co-eluted with the target protein. This held true even after size-exclusion chromatography—despite being double the size of the FRePs, I consistently observed significant cross-contamination of the FRePs with MBP (data not shown). I eventually overcame this challenge by implementing a high-performance IEX step using MonoS/MonoQ columns. This worked exceptionally well for positively charged targets (like TNC) and yielded very pure protein (Figure 2.3D). However, this approach was less successful for negatively charged targets like ANGPTL6, necessitating significant target-specific optimization of elution gradients to obtain reasonable purity.

As a result of the aforementioned challenges, consistently low final yields (less than 0.25 mg L^{-1}) restricted the number of crystallization trials I could complete; after 11 FReP purifications across five different targets using Scheme 1, I only set-up 15 different crystallization plates, and no crystals appeared in any condition. Given that successful crystallization often requires dozens of plates per target, even the fully optimized Scheme 1 represented a bottleneck to solving FReP structures. Motivated to increase final yields and simplify a relatively challenging purification scheme, I decided to pursue a second purification scheme that did not rely on the cumbersome MBP-tag.

2.4.2 Purification Scheme 2: His6-tag in CyDisCo Cells

As demonstrated previously in Figure 2.2, next-generation CyDisCo cells can express FRePs containing only an N-terminal, TEV-cleavable His6-tag, allowing a simpler and higher yielding purification Scheme 2 (Figure 2.4A). Instead of the 72–96 hour Scheme 1, Scheme 2 takes less than 36 hours to complete: the first affinity step and a two-hour room temperature TEV cleavage are finished on the first day, followed by an overnight SEC run on a Superdex 75 column to remove aggregates, impurities, and trace TEV protease. I used Scheme 2 to purify the entire FReP library, including the previously intransigent FGL2, with representative purification results

for each target found in Figure 2.4B. Furthermore, Figure 2.4C demonstrates that the purified FRePs are monodisperse and monomeric during SEC, except for FGL2, which tends to give a broader peak in SEC, indicative of intrinsic disorder or partial unfolding.

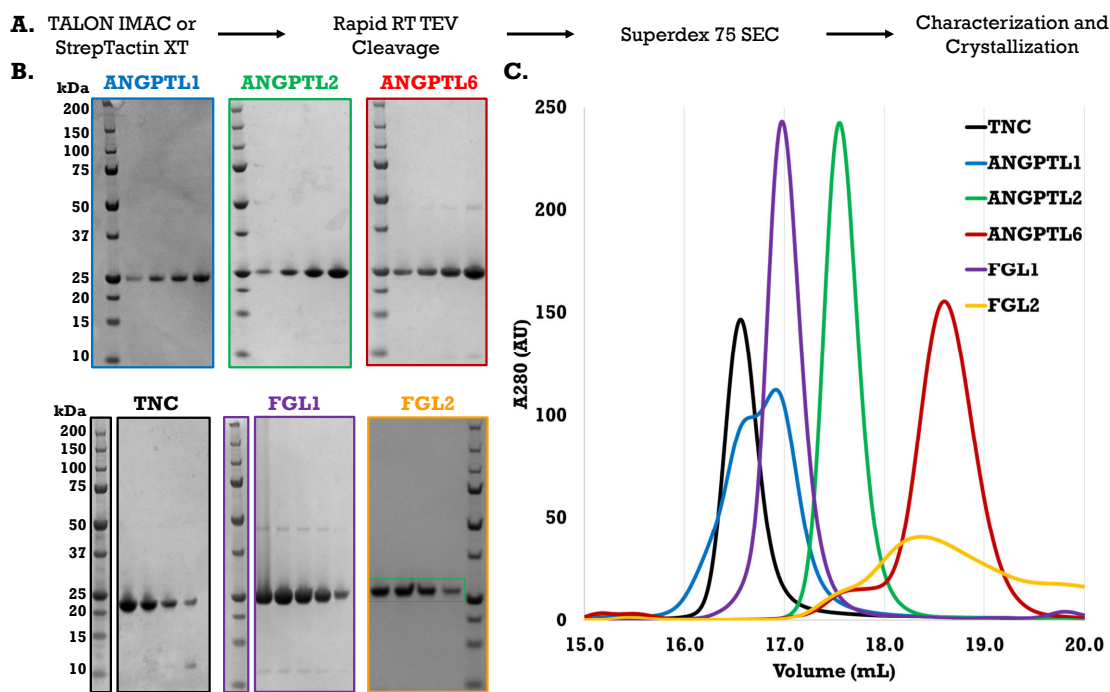


Figure 2.4 Purification Scheme 2 (His6/CyDisCo). **A.** Optimized purification Scheme 2. **B.** SDS-PAGE showing purified FRePs after final Superdex 75 step. Samples taken from across a monodisperse SEC peak. Protein identities confirmed by intact mass analysis and MSMS. **C.** Analytical SEC on a Superdex 200 Increase column. Approximate elution volume of 17 mL corresponds to a monomeric species as judged by comparison to calibration standards.

I observed that CyDisCo cells grow slower and to a lower final density than BL21(DE3) or SHuffle[®] cells, even accounting for the fact that CyDisCo cells must grow at 30 °C to ensure proper folding of the co-chaperones. These observations were echoed by Dr. Lloyd Ruddock, the CyDisCo inventor, in a personal communication. To increase biomass during the expression stage, I switched from using standard Terrific Broth (TB) to a customized recipe developed in-house by Dr. Michael Fairhead called “TB-PLUS”. In addition to all other standard TB ingredients, TB-PLUS contains a source of carbohydrate (in the form of glucose), magnesium, ammonium, carbon (in the form of excess glycerol), and trace metals, along with antifoam to prevent bubble formation at high shaking speeds (250 RPM). The switch to TB-PLUS essentially doubled the final biomass, allowing me to culture reasonable volumes at a

time (never more than 12 L).

On average for the full-length FRePs, Scheme 2 yielded 2.2 mg L^{-1} , with TNC giving the highest (5.0 mg L^{-1}) and FGL2 giving the lowest (0.5 mg L^{-1}) yields. These improved yields, up to 10-fold higher than with Scheme 1, allowed me to set additional crystallization trials with different screens and higher protein concentrations. However, after four additional plates for ANGPTL2, four for ANGPTL6, and two for FGL1, I again obtained no crystals for the full-length FRePs, consistent with the negative crystallization results post-Scheme 1.

Taken together, my data clearly show that Scheme 2 is superior to Scheme 1 in terms of time, purity, and final yields, so I transitioned completely away from Scheme 1 during the subsequent stages of the project (especially important in Section 2.6). Unless otherwise noted, I purified all FRePs for structural or biophysical characterization using the optimized Scheme 2.

2.5 Quality Control of Soluble FRePs

Throughout the project, I relied on stringent quality control experiments to ensure that my soluble FRePs were the correct target, properly folded, homogeneous, and biophysically equivalent to FRePs purified using the old refolding protocol. Representative data from these validation experiments are presented and discussed below.

Using intact mass and in-gel tryptic digest MSMS, I confirmed the identity of each FReP. Furthermore, I used a simple “disulfide mapping” assay that allowed experimental validation of the presence of two disulfide bonds in the expressed FRePs. In these experiments, I measured the intact mass of the protein three times—once as purified, once after incubation with iodoacetamide (IA), and once after reduction with TCEP and incubation with IA. IA reacts with any free thiol (i.e. reduced Cys residues) to form a covalent acetylated adduct with a fingerprint mass shift of +57 Da. Cysteine residues involved in a disulfide bond will only be acetylated following reduction with TCEP, while reduced free cysteines will undergo acetylation by IA without reduction. Because the in-house time-of-flight mass spectrometer which I used only had a resolution of ± 1.0 Da, intact mass alone could not unambiguously confirm the presence of disulfide bonds. On higher resolution instruments, disulfides can be directly measured via the oxidation-dependent loss of two protons, and therefore 2 Da, per disulfide bond.

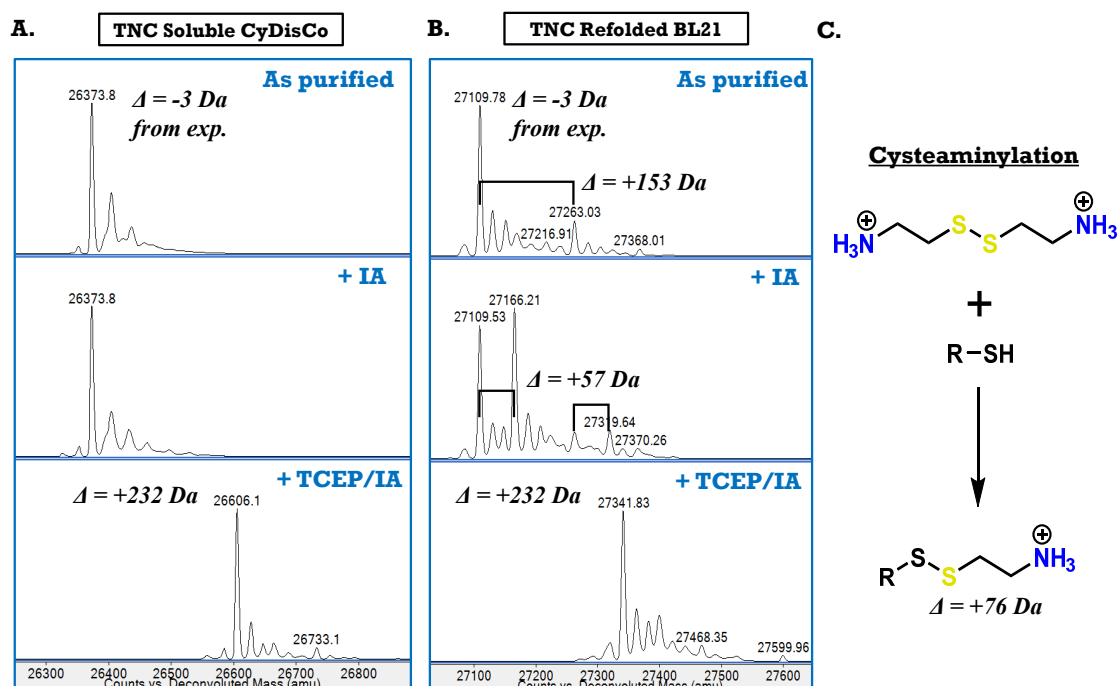


Figure 2.5 Disulfide Mapping of TNC. **A.** Disulfide mapping of soluble TNC expressed in CyDisCo cells using Scheme 2. Protein contains two disulfide bonds. **B.** Disulfide mapping of refolded TNC expressed in standard BL21(DE3) using previously reported purification scheme. Protein contains incorrectly folded minor species containing free cysteines or cysteaminylation. **C.** Schematic of cysteaminylation event caused by excess cystamine used in the refolding step. The +153 Da adduct in B. arises from two cysteaminylation events

Figure 2.5 shows a representative result of this disulfide mapping assay, used to directly compare solubly expressed and refolded TNC (purified by and received from Anna Marzeda). The solubly expressed TNC, shown in Figure 2.5A, contains only one species ($m = 26,374$) with the expected two disulfide bonds. This is deduced by the lack of a mass shift in the “+IA” sample, implying no free cysteines, and the +232 Da mass shift in the “+TCEP/IA” sample, implying four oxidized cysteines and therefore two disulfide bonds. The refolded TNC, shown in Figure 2.5B, contains a major species that behaves as expected, with two disulfide bonds ($m = 27,110$); however, the refolded TNC also contains multiple improperly folded minor species. The +153 Da adduct in the “As purified” sample corresponds to two cysteaminylation events, detailed in Figure 2.5C and resulting from excess cystamine used in the refolding buffer. Furthermore, a portion of the purified protein also contains one free cysteine, demonstrated by the +57 Da shifts in the “+IA” sample, showing that both disulfide bonds are not always present after refolding. The

relative abundance of these minor species varied from prep to prep, but all refolded TNC I tested contained some improperly folded protein. Therefore, the solubly expressed FRePs not only match but in fact outperform the refolded FRePs in terms of homogeneity and disulfide bond status.

In addition to verifying the monodispersity of the soluble FRePs during every purification via SEC (see Figure 2.4), I also verified the folding and oligomeric state of the FRePs using Native MS. Unlike intact mass analysis, where the protein is denatured by strong acids like TFA before injection, Native MS detects proteins in their solution state, preserving protein-protein interactions, oligomers, fragile PTMs, and ligand or co-factor binding [377]. As Figure 2.6 demonstrates, both refolded and solubly expressed TNC exist predominately as a folded monomer in solution, which aligns with expectations and with the SEC results (note that the small difference in the modal m/z is due to the presence of a His6-tag on the refolded sample). However, refolded TNC also contains a number of low m/z species, corresponding to unfolded or denatured protein, which are absent in the solubly expressed sample. In summary, refolded and solubly expressed FRePs are broadly equivalent biophysically, with the solubly expressed samples displaying superior homogeneity and disulfide bond formation.

2.6 CyDisCo Expression of Truncated Constructs

Despite the success of Scheme 2 and the ability to rapidly purify homogeneous, properly folded full-length FRePs, all of my initial crystallization trials failed to yield even crystalline precipitates for any target after 25 different crystallization plates. This result, common amongst all of the FRePs I tested, suggested a pathological resistance to crystallization in the full-length constructs. Because crystallization requires proteins to condense into a repetitive lattice via the formation of conserved, reproducible contacts between adjacent molecules [378], it is not uncommon for full-length proteins to resist crystallization; in fact, evidence suggests that only 25-30% of recombinantly expressed proteins can form crystals [379]. Proteins with highly flexible regions, especially solvent exposed loops or disordered subdomain linkers, are particularly intransigent to crystallization due to a high degree of conformational freedom which reduces the probability of forming stable crystal contacts. Intuitively, then, evidence demonstrates that truncating targets down to a minimal crystallizable construct, either by isolating individual domains or removing flexible loops, can

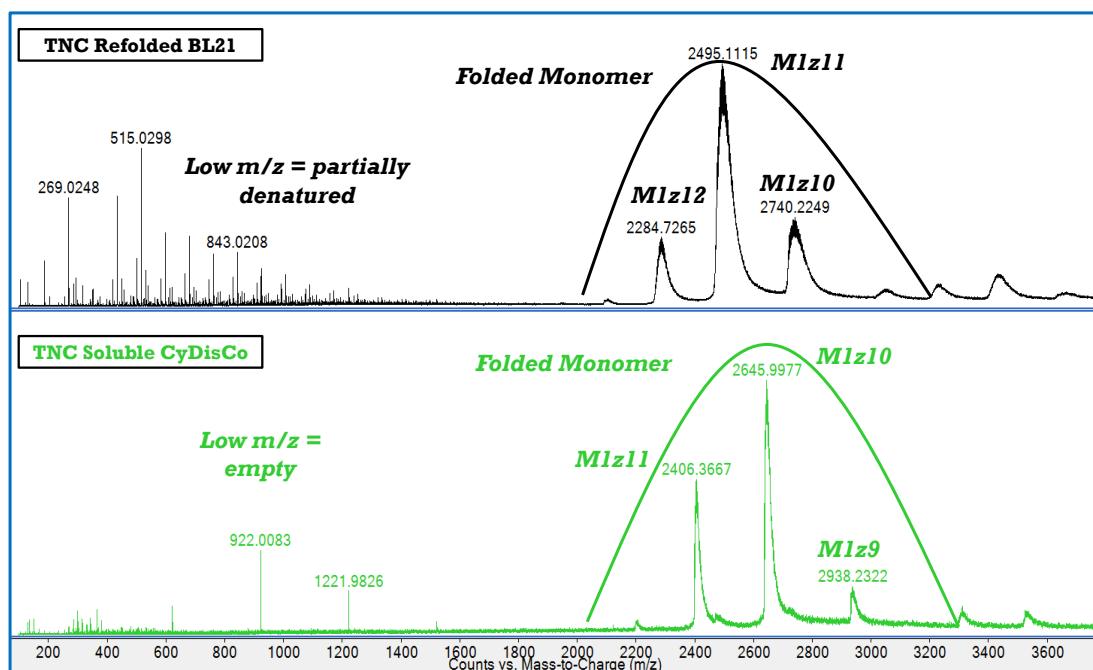


Figure 2.6 Native MS of TNC. Native MS analysis of solubly expressed and refolded TNC. Both samples display a single charge-state distribution corresponding to a folded monomer; however, the refolded TNC also contains denatured or poorly folded protein in the low m/z region.

improve the success rate of crystallization [328, 380, 381, 382]. Therefore, after failed attempts to crystallize the full-length FRePs, I moved forward with the systematic design and testing of a truncated FReP library.

I relied on a variety of bioinformatic approaches to identify potentially disordered regions and rationally design truncated constructs, including secondary structure and disorder prediction using the PSIPRED Workbench [383], structure-guided multiple sequence alignments using T-Coffee [384], and comparative homology modelling using ICM-PRO [385]. I used the following structures of human FRePs, available at the time, as templates to guide my construct design: ANGPTL3 (PDB: 6EUA), ANGPTL4 (PDB: 6EUB), ANG1 (PDB: 4JYO), ANG2 (PDB: 1Z36), FCN1 (PDB: 2JHH), and FGG (PDB: 1FIC).

Because the full-length FRePs already contain just a single FBG domain, there existed no opportunities for further subdomain truncation or removal of inter-domain linkers. Additionally, I hesitated to delete internal loops which might compromise the biological relevance of the resultant structures, especially since the loop-filled P-subdomain of the FRePs contributes to receptor and calcium binding (see Section

2.10). However, my analysis did reveal the presence of N- and C-terminal loops in each of the FRePs which are predicted to be disordered and are not resolved in the crystal structures of the closest orthologs, suggesting a high degree of conformational flexibility. This result is most pronounced in the alignment of TNC with the structures of ANG1 and ANG2, seen in Figure 2.7A; the intact FBG domain of TNC contains a hydrophobic five residue insertion at the N-terminus and a strongly cationic, ten residue insertion at the C-terminus. Disordered, extended terminal loops are also present at the N-terminus of both the ANGPTLs and the FGLs (see Figure 2.7B and 2.7C). I hypothesized that these flexible regions at the protein termini were preventing crystallization but were likely non-essential for the folding of the FBG domain.

Therefore, I focused my construct truncations on the protein termini, including 12 different domain boundaries for TNC and six boundaries for each of ANGPTL1, ANGPTL2, ANGPTL6, FGL1, and FGL2. Since crystal structures of the FBG domains from FCN1 and FGG were already publicly available, I did not design truncated mutants for these targets. Figure 2.7 shows multiple sequence alignments annotated with my designed N-terminal and C-terminal truncations.

Using LIC methodology, I cloned each of the 42 truncated FRePs into both the pNIC28-Bsa4 vector, which adds an N-terminal TEV-cleavable His6-tag, and the pNIC-NHStIIT vector, which adds an N-terminal TEV-cleavable His6-Twin-Strep-tag. The Twin-Strep-tag, purified using StrepTactinXT[®] resin, is especially useful for purifying poorly expressed or secreted proteins where the target is present at low concentrations before the affinity step [386]. Given the small yields for some of the FRePs (as low as 0.5 mg L⁻¹; see Section 2.4) and the need to maximize recovery of potentially poorly expressed truncated constructs, I took the opportunity to pilot the Twin-Strep-tag in parallel with testing terminal truncations.

I expressed each of the 84 truncated constructs at a 50 mL scale in TB-PLUS, followed by a one-step affinity purification using TALON (pNIC28-Bsa4 constructs) or StrepTactinXT[®] (pNIC-NHStIIT constructs) resin. Figure 2.8A and 2.8B show the test expression and purification results, which illustrate several important conclusions. First, Twin-strep-tagged constructs outperform His6-tagged constructs, as the higher-affinity tag allows pull-down of eight additional constructs versus the TALON pull-down. This appears especially important for ANGPTL1 and ANGPTL2, which express so poorly that they are undetectable when pulled down with the lower-affinity, lower purity TALON resin. Second, ANGPTL1, ANGPTL2, ANGPTL6,

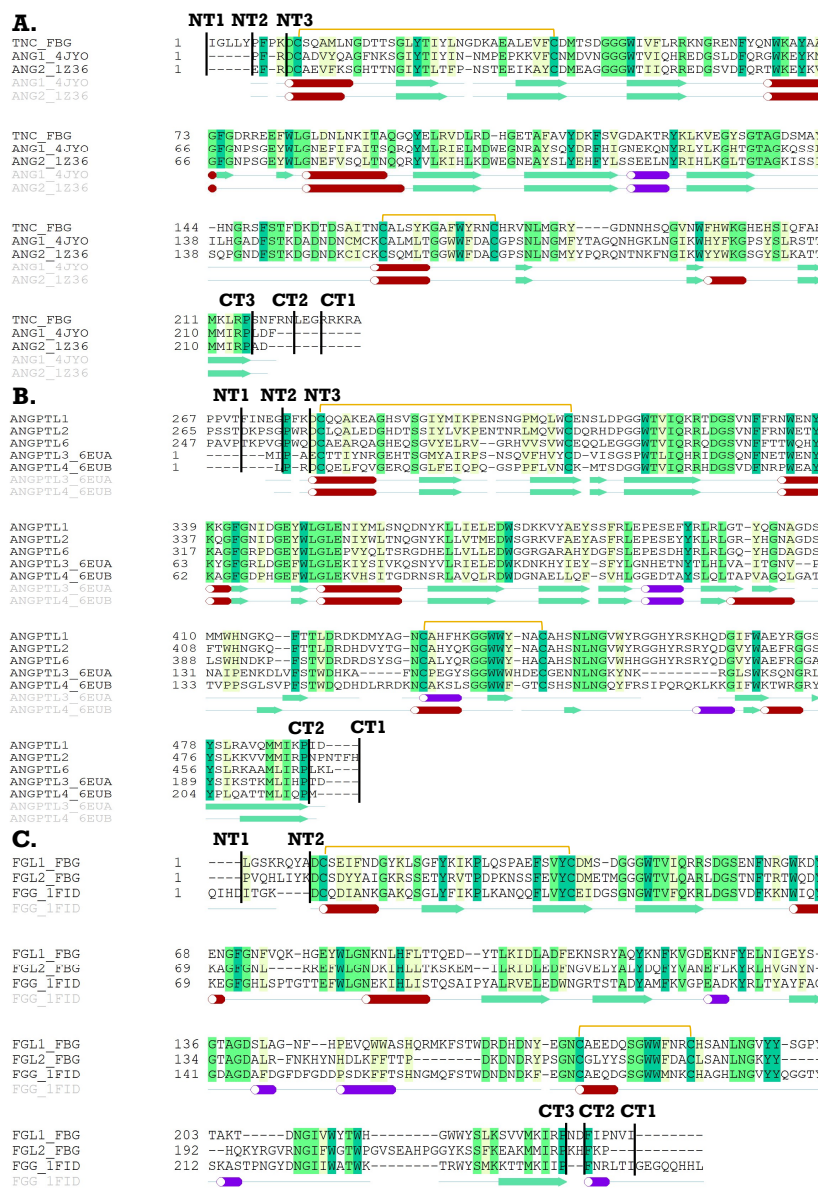


Figure 2.7 Design of Truncated FReP Constructs. **A.** Rational design of truncated TNC constructs via comparison to ANG1 and ANG2 structures. **B.** Rational design of truncated ANGPTLs via comparison to ANGPTL3 and ANGPTL4 structures. **C.** Rational design of truncated FGLs via comparison to FGG structure. In all panels, structure-guided MSAs were generated by ICM-PRO and colored by ICM's “consensus strength” algorithm, which emphasizes the key conserved residues that anchor the alignment. Secondary structures from the comparator structures are shown underneath. Conserved disulfides are marked with a yellow bracket.

and FGL2 are not amenable to domain truncation, as only the full-length constructs expressed. This suggests that even small changes can alter the solubility of particular FBG domains and that the overall solubility of the FRePs varies from target to target. Finally, and most importantly, both TNC and FGL1 are in fact amenable to truncation, and I uncovered five new TNC boundaries and two new FGL1 boundaries which expressed well.

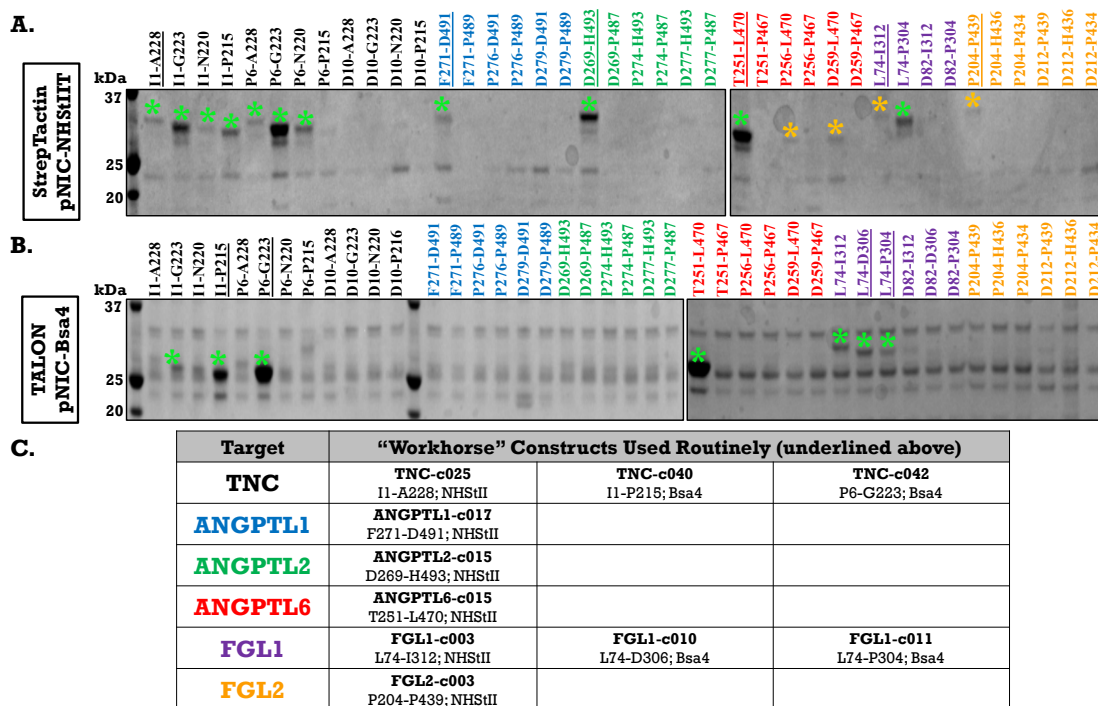


Figure 2.8 Test Expression of Truncated FRePs. **A.** 50 mL test expression of FReP truncations with an N-terminal His6-Twin-strep-tag cloned into pNIC-NHStiIT and purified with StrepTactinXT[®] resin. **B.** 50 mL test expression of FReP truncations with an N-terminal His6-tag cloned into pNIC28-Bsa4 and purified with TALON resin. Green asterisks indicate bands confirmed by in-gel tryptic digest MSMS, while orange asterisks indicate putative target bands not verified by MSMS. **C.** Overview of “workhouse” FReP constructs, underlined in A. and B., used for downstream biophysical experiments throughout the duration of the project. For simplicity, Ile1974 of TNC, corresponding to the first residue of the FBG-domain, is labelled as Ile1.

My results demonstrate that C-terminal truncations out-perform N-terminal truncations for FGL1, and the overall expression level does not change between the Ile312, Asp306, and Leu304 boundaries. In contrast, both N- and C-terminal truncations work for TNC, with deletion of the cationic C-terminus seeming particularly beneficial for protein solubility. TNC constructs ending in Gly223, Asn220, or Pro215 express better than the full-length TNC, an effect which synergizes with a short

N-terminal truncation to Pro6. The Pro6-Gly223 TNC construct proved to be the best expressed construct across the entire truncated library, followed closely by TNC Ile1-Pro215 and FGL1 Leu74-Pro304. As Sections 2.7 and 2.11 demonstrate, these truncated constructs are crystallizable.

Overall, while the expression of truncated FRePs varied wildly from target to target, my truncated construct library did reveal eight new domain boundaries for two different FRePs which proved useful for crystallization. Furthermore, this experiment clearly demonstrated the utility of the Twin-strep-tag for isolating poorly expressed FRePs, particularly full-length ANGPTL1, ANGPTL2, and FGL2, which are not detected from a 50 mL expression when purified using only the His6-tag. Figure 2.8C provides an overview of the “workhorse” FReP constructs uncovered by the truncated library which I routinely scaled-up using purification Scheme 2.

2.7 Structure of TNC

The highest yielding FReP construct identified in Figure 2.8 was TNC-c042, which is truncated by five residues at both the N- and C-termini to contain Pro1979–Gly2196 (labelled for simplicity as Pro6–Gly223 in Figure 2.8) of TNC. Using Scheme 2, I obtained greater than 10 mg L^{-1} of purified, un-tagged TNC-c042, allowing me to set over a dozen crystallization trials with different random matrix screens (JCSG7, Morpheus, LMB, MidasPlus), protein concentrations (17 and 30 mg mL^{-1}), and temperatures (4°C and 20°C). I obtained crystals in two different conditions at 20°C , one from the Morpheus coarse screen at 17 mg mL^{-1} and one from the LMB coarse screen at 30 mg mL^{-1} . While crystals from both conditions routinely diffracted to better than $2 \text{ \AA}$, the highest resolution dataset came from a crystal in the Morpheus condition which diffracted to $1.4 \text{ \AA}$.

These crystals belong to the tetragonal spacegroup $P4_12_12$ with one molecule in the asymmetric unit. Using molecular replacement with the FGG structure as the search model (PDB: 1FIC), I solved the structure of the FBG domain of TNC, which I deposited as PDB ID: 6QNV. Figure 2.9 provides an overview of the purification and crystallization of TNC-c042, including data processing and refinement statistics. The final model converged to $R_{\text{work}}/R_{\text{free}}$ of $0.17/0.19$ and resolved Pro1978–Leu2914 of TNC. Residues Asp2128–Thr2134, corresponding to a putative Ca^{2+} binding loop (see below) were not modelled due to poorly defined electron density, and the final two C-terminal residues (Glu2195, Gly2196) were not observed.

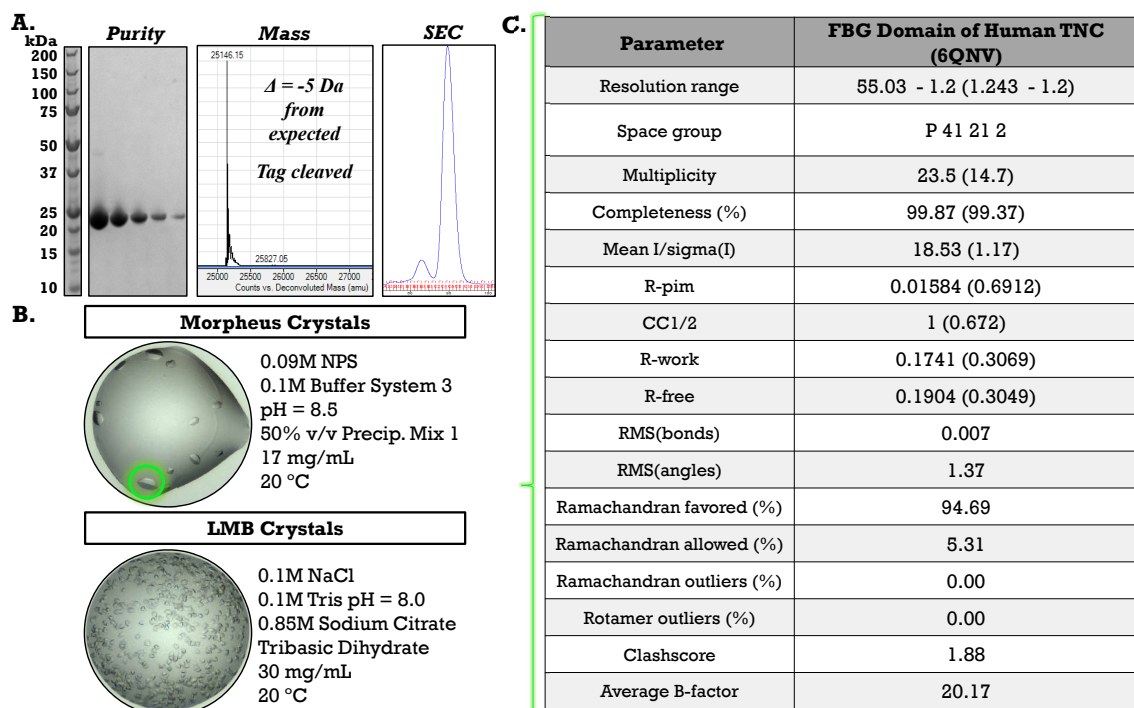


Figure 2.9 Crystallization of TNC-c042. **A.** 2L purification of TNC-c042 using Scheme 2. Final purity, intact mass, and SEC chromatogram shown. **B.** Conditions with crystals of TNC-c042. The best crystal, which became 6QNV, is highlighted with a green circle. **C.** Select data processing and refinement statistics for the structure of TNC-c042. Full statistics in the Appendix.

The relative ease with which I crystallized TNC-c042, after unsuccessful attempts with the full-length domain, highlights the importance of designing and testing truncated constructs and confirms my hypothesis about disordered termini preventing crystallization (see Section 2.6). I also obtained crystals of TNC-c040, which contains just a C-terminal truncation to Pro2188 while maintaining the full-length N-terminus, that diffracted to around 2.5 Å (data not shown). I was able to phase and solve the structure from TNC-c040; however, the lower resolution and larger R-factors led me to deposit the higher quality structure solved using TNC-c042. Therefore, my data demonstrate that the cationic C-terminus in TNC is likely responsible for preventing crystallization of the full-length domain. The crystals of TNC-c042 actually pack with adjacent C-termini in contact, meaning that the flexible, positively charged C-terminus in the full-length FBG domain both electrostatically and sterically clashes to prevent crystallization.

The structure of TNC adopts the topology typically seen in other FBG domains [300], including a three sub-domain architecture, a core structural seven-stranded β -sheet (composed of β -sheets 3-8 and 11), and two disulfide bonds (Cys1984–Cys2014

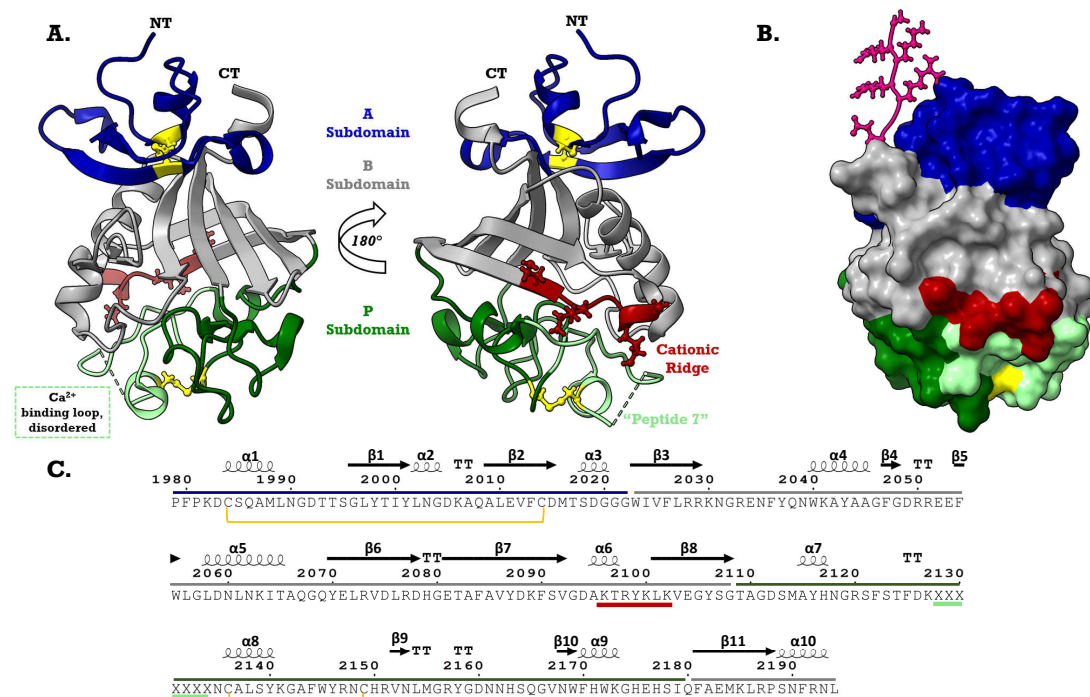


Figure 2.10 Structure of TNC. **A.** Structure of TNC, colored by sub-domain and highlighting the disulfide bonds (yellow), TLR4-binding cationic ridge (red), and the disordered calcium binding loop (light green). **B.** Surface view of TNC as in A., but with the C-terminal truncation "EGRRKRA" (in pink) modelled back into the crystal structure **C.** Sequence level depiction, colored as in A. Figure generated with EndScript [387].

and Cys2136–Cys2149). This structure also allows visualization of key structural and functional epitopes previously identified biochemically. Figure 2.10 highlights these regions of interest, including:

1. The Canonical ABP-subdomain Architecture

Yee *et al.* solved the original structure of the FG C-terminal globe domain and coined ABP as the nomenclature for the FRePs [300]. In TNC, the small A-subdomain at the N-terminus (Pro2979–Gly2022), shown in blue, contains the first disulfide bond in the protein (Cys1984–Cys2104), which anchors the N-terminal helix $\alpha 1$ to strand $\beta 2$. The core B-subdomain (residues Trp2023–Gly2108 and Phe2181–Leu2194), shown in grey, contains a conserved seven-stranded β -sheet which provides a rigid scaffold for the rest of the domain. The B-subdomain also encompasses the distal C-terminal $\beta 11$ (Phe2181–Pro2188) that, despite being non-contiguous with the first six strands in the sequence, folds back with the core β -sheet and is therefore included in the B-subdomain. Two loops divide the B-subdomain from the P-subdomain, shown in green.

The P-subdomain (residues Thr2109–Gln2180) is the least conserved across family members and contains limited secondary structure, along with the second disulfide bond (residues Cys2136–Cys2149). This variable, loop-filled P-subdomain contains functional moieties, including ligand binding sites and PPI interfaces, in other homologs that differentiate the biological activity of the FRePs [388, 389].

2. The Presence of Two Disulfide Bonds

There was strong and unambiguous electron density for both disulfide bonds, colored in yellow, in agreement with the mass spectrometry experiments in Section 2.5. The disulfide bonds in TNC complement the seven-stranded β -sheet and provide structural integrity by tethering either the flexible N-terminal $\alpha 1$ or the Ca^{2+} -binding-loop-adjacent $\alpha 8$ to the rest of the domain.

3. The Calcium Binding Loop

The P-subdomain contains a putative solvent exposed Ca^{2+} binding loop (residues Asp2128–Thr2134) which is the only disordered region in my TNC structure (see Section 2.10 for a resolved Ca^{2+} binding loop). The lack of observed electron density for the loop or a metal ion is not unexpected, given the lack of calcium in the crystallization condition and the flexibility of metal-binding loops in the absence of their ligands.

4. The TLR4 Binding Epitopes

Using carefully designed overlapping peptides and both cellular and biophysical assays, Zuliana-Alvarez *et al.* mapped the TLR4-interacting region on TNC to a conserved cationic ridge in the center of the domain, displayed here in red [303]. In fact, as mentioned in Section 1.3, the presence of this cationic ridge classifies FRePs into TLR4 activators and non-activators. In TNC, the core cationic ridge consists of a **KTRYKLK** motif (residues Lys2096–Lys2102) in the B-subdomain which sits just above the Ca^{2+} binding loop. Zuliana-Alvarez *et al.* also found that the loop-helix-loop motif surrounding the Ca^{2+} binding motif, dubbed Peptide 7, enhances the TLR4-binding properties of the cationic ridge. This region (residues Arg2120–Asn2148), shown in light green, sits within the P-subdomain directly underneath the cationic ridge. The proximity of these TLR4-activating epitopes hints at the potential for a calcium-dependent interaction between TLR4 and TNC, since calcium-binding would significantly alter the TLR4-binding region within TNC. The combination of my structure with Zuliana-Alvarez *et al.*'s results allows clear visualization of

the core TLR4-associated PPI interface in TNC and provides a target region for neutralizing antibodies or inhibitory small molecules.

5. The Disordered C-Terminus

Although I truncated the final five residues of TNC to enhance crystallization, I can model-in the truncated residues to the core structure using simple energy minimization. This region, designated Loop 10 in the work of Zuliana-Alvarez *et al.*, cooperates with the cationic ridge to enhance TLR4-binding. While this approach is strictly suggestive given the high degree of conformational flexibility at the C-terminus, the modelling results in Figure 2.10B show that the C-terminus likely protrudes outward from the core of the domain into the solvent. Yee *et al.* actually solved the structure of FGG with three different conformations of the C-terminal extension, with the O1 conformation closely matching my energy minimized model [300]. It is intriguing that the positive charge of the C-terminal extension matches the positive charge in the cationic ridge; however, given the distance between the cationic ridge and the C-terminus (greater than 25 Å) and the flexibility in the loop, the exact nature of the entire TNC:TLR4 binding interface is impossible to predict in the absence of structural data for the complex.

2.8 Fragment Screening of TNC Using XChem

2.8.1 Motivation for Fragment Screen

As discussed in Chapter 1, inhibitors of the TLR4:TNC PPI hold promise in a variety of autoimmune conditions. In face, a start-up biotech company called Nascient Ltd. developed high-affinity neutralizing anti-TNC-FBG monoclonal antibodies which showed promise in animal models of RA and are continuing to clinical development [305]. Because TNC is an extracellular target, Nascient’s work on neutralizing antibodies follows a traditional approach. However, small molecules can also effectively target extracellular proteins and provide an orthogonal approach to more expensive antibody injections. Additionally, my in-house data and the work of others suggests that TNC interacts with a multitude of proteins at the cell surface to form a “receptorome”, with varying functional consequences depending on the ligated receptors (see Chapter 5). Given the small size and compact nature of TNC, neutralizing antibodies are likely to inhibit *all* TNC-containing PPIs, providing a

rationale for the development of small molecules which selectively target only a subset of TNC-mediated PPIs. These interaction-specific inhibitors could also serve as useful chemical probes for elucidating the complex biology surrounding TNC.

The reproducible resolution and crystallizability of the P41212 TNC-c042 crystal system made it a perfect candidate for crystallographic fragment screening [390]. Crystallography is a powerful tool for fragment-based drug discovery because it detects very weak binders, accommodates target proteins of any size, allows simultaneous sampling of the entire target surface, and couples hit-identification with structural elucidation to accelerate structure-guided drug discovery [391]. Recent advances in automated beamline technology for data collection and algorithms for the identification of weak hits during data analysis (particularly PanDDA, developed by Pearce *et al.*) make it possible to routinely use crystallography as a primary fragment screen [347, 392, 393]. The XChem facility at Diamond Light Source, beamline I04-1, is particularly equipped for high-throughput crystallographic fragment screening, with innovative solutions for crystal mounting and data collection allowing users to screen up to five-hundred fragments in just two to three days [394].

2.8.2 Results of Fragment Screen

Based on the potential for small-molecule TNC binders, I completed a full crystallographic fragment screen against TNC-c042 in collaboration with the XChem team at Diamond. An overview of this screen, consisting of 300 fragment-soaked datasets with an average resolution of 1.7 Å, is shown in Figure 2.11. Based on a PanDDA analysis [347], the screen revealed 15 fragments bound (i.e. visible upon manual inspection in the enriched PanDDA maps) in five different sites (named A–E). Based on the observed electron density in the final, refined 2mFo-DFc maps, 11 of these hits, across three sites (A, D, and E), were deposited to the PDB. These eleven structures are summarized in Figure 2.12 alongside the electron density for each identified fragment. I chose not to publicly deposit the other four co-crystal structures, containing fragments **7**, **10**, **11**, and **12**, because the fragment electron density was only clear in the enriched PanDDA maps (i.e. in standard 2mFo-DFc density maps, the fragment could not be visualized or refined). This was a conscious choice to avoid pollution of the PDB with dubious ligands, which is a recognized problem [395]. However, for the sake of an exploratory thesis, I discuss all fifteen of the hits below, but I want to make it clear that fragments **7**, **10**, **11**, and **12**

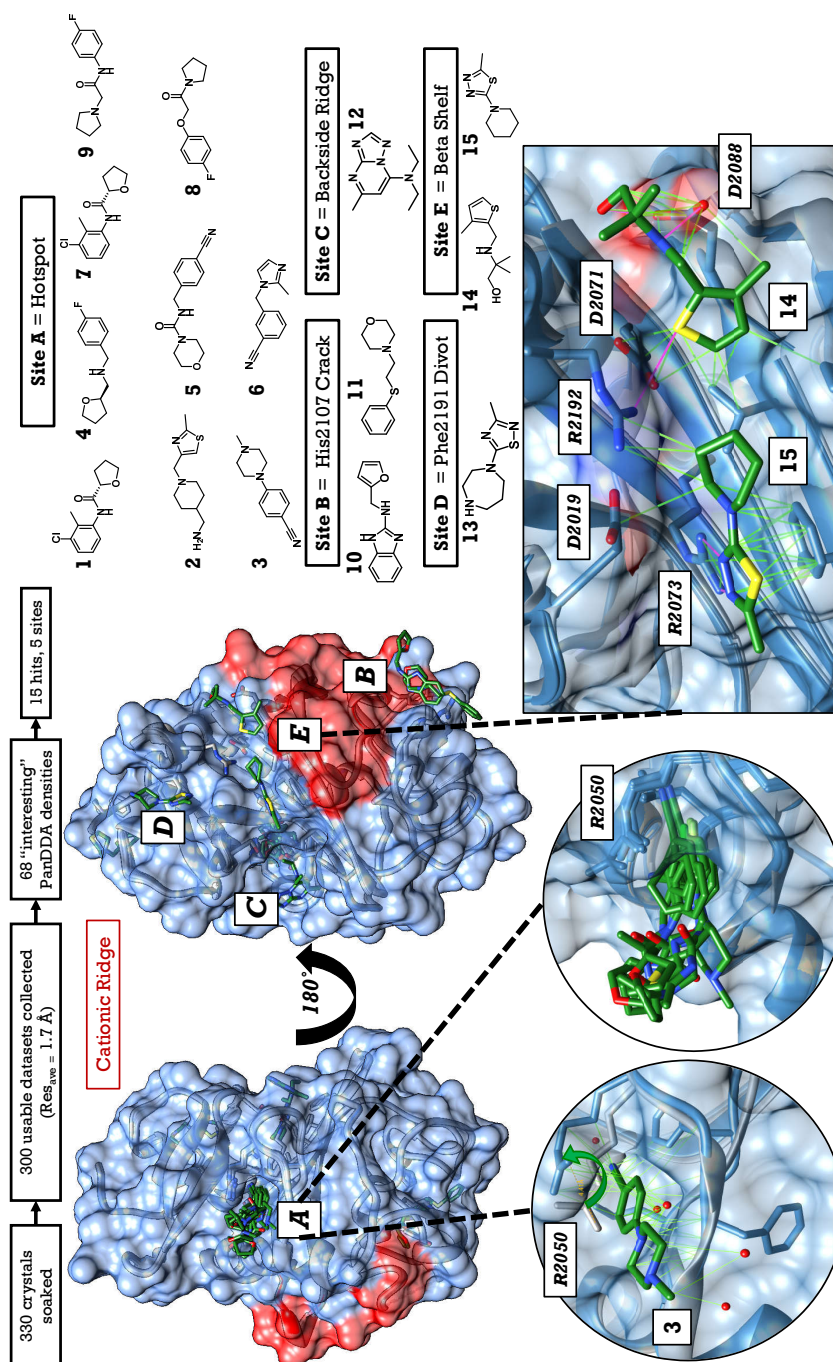


Figure 2.11 Crystallographic Fragment Screen of TNC. Overview of the crystallographic screening campaign completed against TNC-c042 using the XChem I04-1 Beamline at Diamond Light Source. Hydrogen bonds are shown with pink lines, Van der Waals interactions are displayed with bright green lines. The TLR4 binding interface is displayed as a red surface. All fragments hits are colored by heteroatom with carbons in dark green.

are weaker than the other hits, which are clearly visible in standard maps as Figure 2.12 demonstrates. The results presented herein form the basis of a “Target Enabling Package” (TEP) for TNC, which was freely released online in December 2020 (<https://www.thesgc.org/tep/tnc>) to support drug discovery efforts by other groups.

The site with the most hits, Site A, contains fragments **1–9** and is a previously unidentified cryptic hot-spot on the opposite face from the cationic ridge. This hot-spot forms due to the movement of the Arg2050 side chain, which is outwardly displaced by 4.4 Å upon fragment binding to reveal a small hydrophobic pocket. The hot-spot accommodates a phenyl ring, with electron-withdrawing *para*-substituents, including nitrile and fluoro groups, seeming particularly preferred. The strongest electron density observed in the screen, a weak proxy for the affinity of the interaction, came from fragment **3**, which contains a benzonitrile that penetrates to the back wall of the hot-spot formed by the core beta-sheet (see Figure 2.11, circular inset). The rest of the hot-spot fragments adopt a similar conformation, with the phenyl moiety of the fragment bound underneath R2050 and the rest exposed to the solvent.

The Site A hot-spot is the only concave binding pocket in the protein, while the other fragments are found in shallow solvent-exposed grooves. Because of its proximity to the target TLR4-binding interface, Site E, nicknamed the “beta-shelf”, is intriguing. Site E contains fragments **14** and **15** bound directly above the cationic ridge along a surface cleft. Site E forms along the core TNC β -sheet via interactions with nonpolar residues at the base of the shelf and charged residues on the upper part of the shelf. Fragment **14** is stabilized by two hydrogen bonds, one between Arg2192 and a thiophene sulfur, the other between Asp2088 and a secondary amine. Fragment **15** is also stabilized by two hydrogen bonds, in this case between Arg2073 and both thiadiazole nitrogens (see Figure 2.11, rectangular inset). Like the Site E beta-shelf fragments, the Site B “His2107 crack” fragments are proximal to the cationic ridge; however, these two hits are at least partially mediated by crystal contacts with the adjacent symmetry mate, suggesting they are unlikely to bind in solution and are therefore sub-optimal starting points for medicinal chemistry. As a result, **14** and **15** are the only high confidence hits in close proximity to the TLR4 binding interface. Finally, the screen uncovered two additional sites each with one binding event: fragment **12** binds to Site C, aka the “backside ride”, and fragment **13** binds to Site D, aka the “Phe2191 Divot”. Neither of these sites are predicted to be functionally relevant, and neither of the fragments are tethered by hydrogen bonds

in the shallow binding sites. Therefore, the most intriguing hits for follow-up are in the site A hot-spot and the Site E beta-shelf.

On one hand, my screen revealed the benefits of crystallographic fragment screening, while on the other hand it demonstrated that TNC is likely to be difficult to target with small molecules. My screen suggests that only one concave pocket exists in the protein, and to its credit the crystallographic screen revealed this cryptic hot-spot. However, the hot-spot only accommodates up to a six-member ring, leaving no obvious room for extending the fragment to increase affinity. The other observed sites adequately sample much of the surface of TNC, highlighting again the strength of the approach, but no clear pockets appear. Rather, the fragments bind in shallow, extended, solvent-exposed grooves, which are much more challenging to drug with small molecules. Based on current knowledge, Site E is likely the most functionally relevant due to its proximity to the cationic ridge; however, this interface is wholly solvent exposed, again making it difficult to target with a small molecule. As the biology surrounding the TNC-associated receptorome (see Chapter 5) is untangled and other binding partners and functional interfaces are mapped, other sites could become more intriguing, especially if they sit at a currently unknown PPI interface. A PPI involving the hot-spot, for example, could create a druggable bi-functional binding interface; molecules binding to the partner protein but protruding into the small hot-spot could theoretically block the interaction. However, based on current data, my fragment screen served primarily to illustrate the validity of Nascient’s approach to target TNC with antibodies, which can bind the shallower, more extended surface epitopes which predominate in TNC. As a result of these observations, I did not proceed with further validation of fragment binding or any additional medicinal chemistry. However, the data are freely available in the PDB and online to jump-start efforts in other groups, especially as additional TNC biology emerges in the future.

2.9 Stabilization of Full-Length ANGPTLs

2.9.1 Theory of Thermal Denaturation Assays

As Sections 2.7 and 2.11 demonstrate, small C-terminal truncations rescued the crystallization of TNC and FGL1. Unfortunately, however, only some FRePs can tolerate these truncations, with the ANGPTLs seeming particularly destabilized

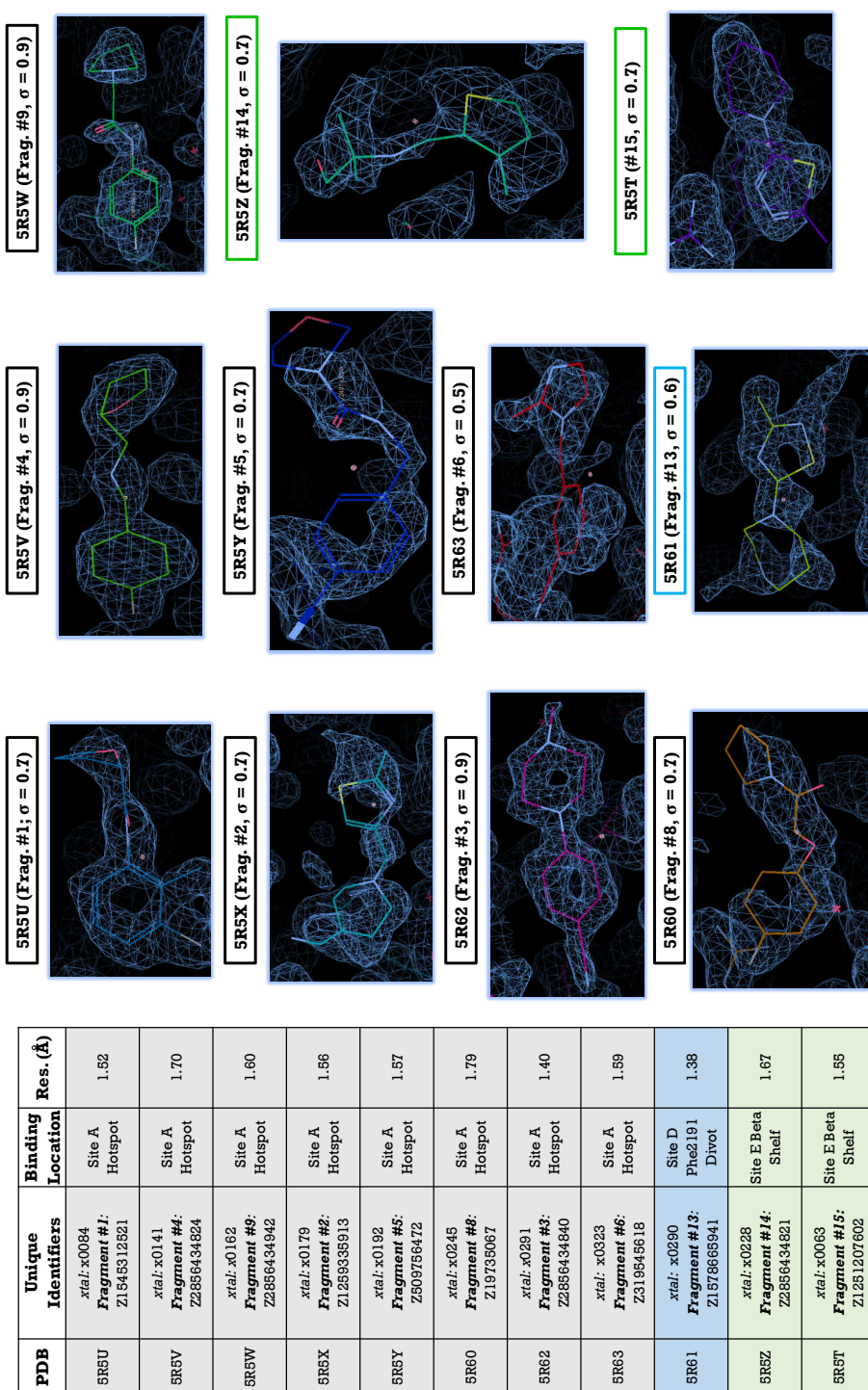


Figure 2.12 TNC:Fragment Structure Depositions. The 11 fragment-bound TNC structures, containing unambiguous electron density for the ligands, which were deposited to the PDB. Insets contain the indicated fragment and associated 2mFo-DFc electron density contoured at the indicated σ level (rmsd) using Coot.

by even small deletions (see Figure 2.8). To overcome this challenge and achieve crystallization of full-length ANGPTLs, I searched for stabilizing additives using thermal denaturation.

Thermal denaturation assays, often called “thermal shift” or, for fluorescence-based experiments, differential scanning fluorimetry (DSF), measure protein unfolding upon step-wise increases in temperature. For DSF, the protein is mixed with a dye (e.g., SYPROTM Orange) that fluoresces upon binding to hydrophobic patches on the surface of the protein. As the protein denatures, core hydrophobic regions become exposed, leading to a temperature-dependent increase in fluorescence [350]. The mid-point of the melting curve represents the melting temperature (T_m) and describes the temperature at which 50% of the protein denatures. Upon incubation with ligands, co-factors, salts, or buffers, the melting temperature of the protein often shifts; an increase in T_m indicates thermal stabilization and can imply both direct binding or indirect stabilization. Multiple groups report that simple 96-well DSF experiments can identify stabilizing buffer and additive conditions, which not only can improve purification yields but often enhance crystallizability [351, 396].

Chari *et al.* reported an important extension of the standard DSF experiment in 2015, finding that the *slope* of the melting curve is more important than the *shift* in the melting temperature when selecting stabilizing conditions [352]. Rather than picking the conditions which give the highest T_m , Chari *et al.* recommend selecting conditions which yield a smooth, monophasic melt curve even if the T_m is lower. Chari *et al.* develop a theoretical framework, called ProteoPlex, that uses nonlinear regression to estimate the number of “transition states” along the melting curve. The closer the protein sample is to an ideal two-state unfolding profile (i.e., folded to unfolded with no transition states in between), the better.

In summary, the literature contains evidence that both the T_m and the slope of the melt curve impact stability and crystallizability. To capture both aspects of DSF experiments, I implemented a GraphPad Prism workflow which calculates the coefficient of determination (R^2) between the observed melting curve and the optimal two-phase melt suggested by ProteoPlex. An R^2 of 1 implies an ideal, monophasic, two-state thermal denaturation. By plotting the ProteoPlex R^2 against the T_m , I can select conditions which stabilize the target across two criterion. Mathematically, this implies conditions at the farthest linear distance away from the origin. This simple analysis pipeline unifies the sometimes conflicting suggestions in the literature and is, to my knowledge, novel.

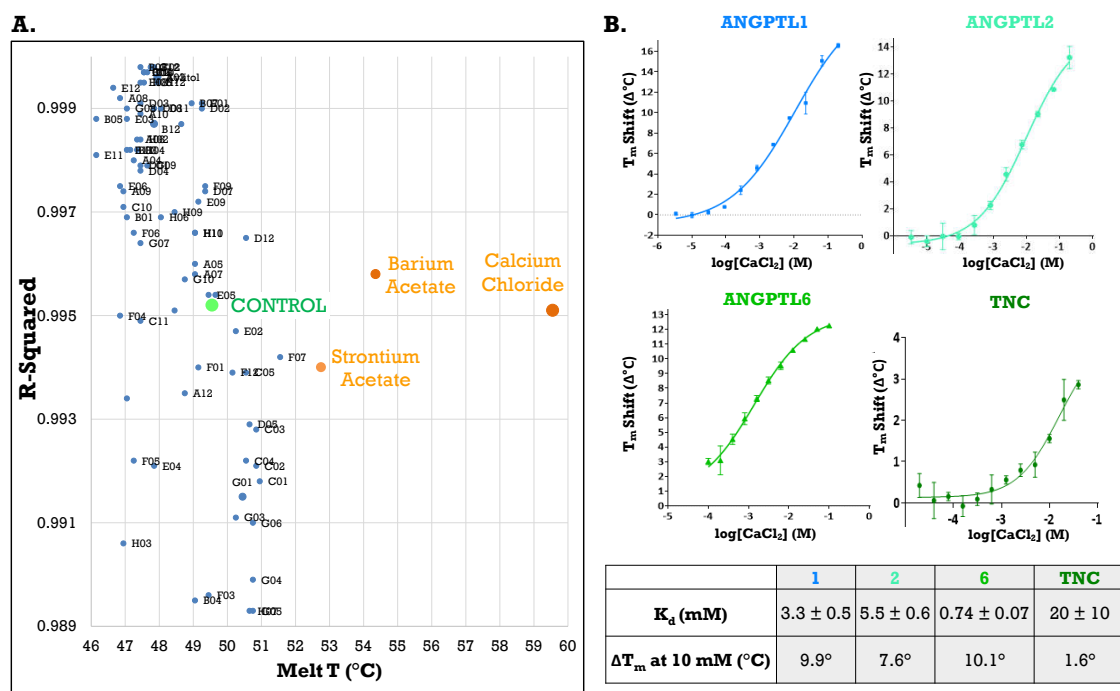


Figure 2.13 DSF Analysis of FRePs. **A.** Additive screen with ANGPTL6, shown as scatter plot of ProteoPlex R^2 versus T_m . Conditions farthest from the origin, all divalent metals, highlighted in orange. **B.** Dose-dependence of Ca^{2+} binding to ANGPTLs and TNC. Data are triplicate curves from one biological replicate.

2.9.2 ANGPTLs Bind Strongly to Calcium

Using the 96-well plate screen created by Boivin *et al.* and my two-parameter analysis protocol, I used DSF to screen the full-length ANGPTLs for stabilizing buffer conditions and additives. The buffer screen revealed that my standard purification buffer (HEPES pH = 7.5) was already optimal (data not shown). However, the additive screen, shown in Figure 2.13A, revealed that the ANGPTLs are strongly stabilized by divalent group II metals such as strontium, barium, and calcium. The discovery of these stabilizing metals is unsurprising given the predicted calcium-binding loop in the ANGPTLs; however, the *degree* of stabilization is quite remarkable. As the dose-response curves in Figure 2.13B demonstrate, 10 mM Ca^{2+} stabilizes the ANGPTLs by upwards of 10 °C, and even μM amounts of Ca^{2+} give large thermal shifts.

Given that just a 2 °C shift is often considered significant, these results imply that engaging the Ca^{2+} binding loop not only stabilizes the loop itself, but likely stabilizes the entire P-subdomain, making Ca^{2+} an ideal additive for crystallization of the

FRePs. This is clear when the ANGPTL and TNC calcium-binding curves are compared. As Figure 2.13B shows, the affinity (20 mM) and the degree of thermal stabilization (less than 4 °C) are much more modest for TNC in the presence of Ca^{2+} . This provides at least a partial explanation for why I could crystallize TNC, but not the ANGPTLs, with an unoccupied, disordered Ca^{2+} binding loop (see Section 2.7). My data indicate that, in the absence of divalent metals, the P-subdomain in the ANGPTLs is more disordered than TNC's.

2.10 ANGPTL2 and ANGPTL6 Structures with Calcium

To explore the crystallizability of the ANGPTLs in the presence of stabilizing Ca^{2+} , I purified 4 L cultures of the full-length, Twin-strep-tagged constructs using Scheme 2. I maintained 25 mM calcium throughout the purification, including in the crystallization buffer, to ensure complete saturation of the Ca^{2+} -binding loop in the ANGPTLs. The expression trend observed at a 50 mL scale (see Figure 2.8) continued at a 4 L scale, with ANGPTL2 giving the smallest final yield (0.2 mg), ANGPTL1 the next highest (0.6 mg), and ANGPTL6 giving the largest yield (2.2 mg). While these yields are quite low, especially relative to TNC-c042 (see Section 2.7), the 4 L scale-ups yielded enough protein to set two crystal plates for ANGPTL1, one plate for ANGPTL2, and three plates for ANGPTL6. Based on my successful use of the Morphueus screen at 20 °C for the crystallization of TNC-c042, I used these conditions for the ANGPTLs as well. The high $[\text{Ca}^{2+}]$ did consistently yield salt crystals in phosphate- and sulfate-containing conditions, so I ignored the (oftentimes beautiful cubic) crystals appearing in these drops.

Using this setup, I obtained well diffracting crystals for ANGPTL2 and ANGPTL6 on my first try and directly out of coarse screens, resulting in a 1.8 Å structure of ANGPTL2 and a 1.6 Å structure of ANGPTL6. These two structures represent the first ever structures of a TLR4-activating ANGPTL. As an aside, I only observed salt crystals for ANGPTL1. The purification and crystallization of ANGPTL2 and ANGPTL6, including data collection and refinement statistics, are detailed in Figure 2.14. Despite their structural similarity and propensity to crystallize in orthorhombic space groups, ANGPTL6 crystallized with only one copy in a small asymmetric unit (ASU) while ANGPTL2 crystallized with sixteen monomers in a massive ASU. In fact, the unit cell volume of the ANGPTL2 ASU is 21-fold larger than

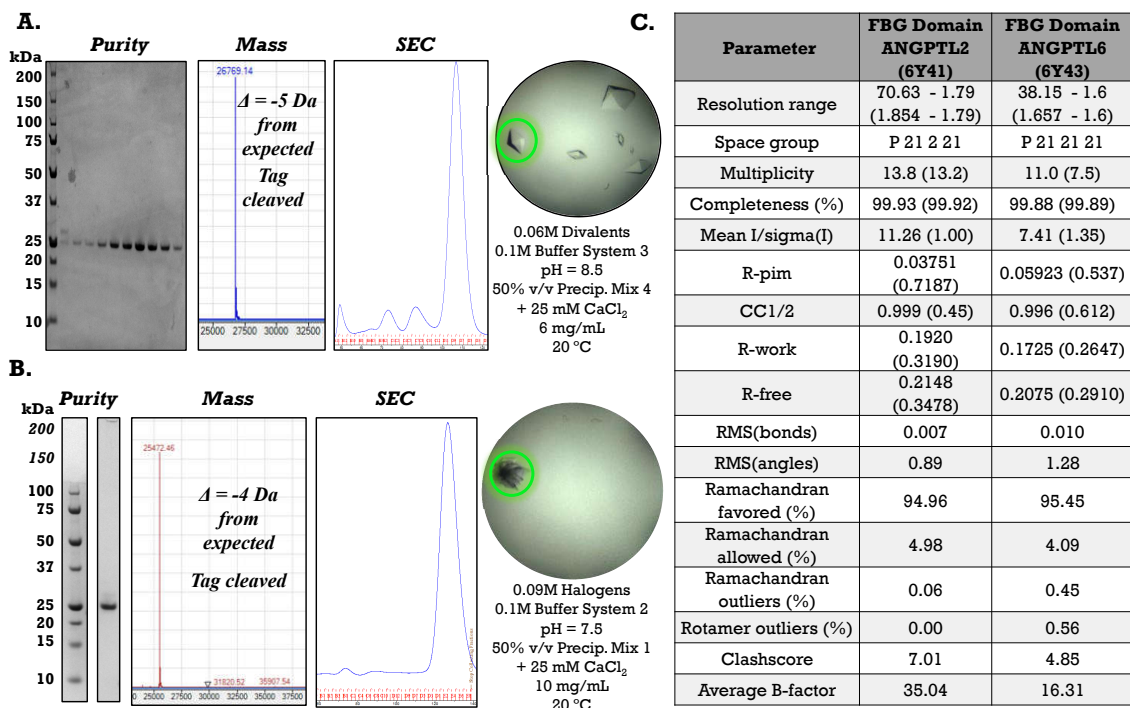


Figure 2.14 Crystallization of ANGPTL2 and ANGPTL6. **A.** 4L purification and crystallization of full-length ANGPTL2-c015 using Scheme 2. **B.** 4L purification and crystallization of full-length ANGPTL6-c015 using Scheme 2. **C.** Selected merging and refinement summary statistics for the structure of ANGPTL2 and ANGPTL6, from the crystals highlighted in green circles. Complete statistics in the Appendix.

the ANGPTL6 ASU. This unique crystal lattice resulted in strong translational non-crystallographic symmetry (tNCS) for ANGPTL2, making data merging and processing more challenging. Fortunately, the high resolution of the dataset allowed me to iteratively build all 16 copies within the ASU with reasonable R-factors in spite of the tNCS. However, because some of the copies are only weakly resolved in the density, the average B-factors as well as the R_{free} and R_{work} statistics are slightly higher for ANGPTL2. For simplicity, I labelled the best resolved copy “Chain A” in the PDB coordinate file, and from herein will use just the ANGPTL2 Chain A for analysis.

The ease with which I crystallized ANGPTL2 and ANGPTL6 at high resolution in the presence of Ca^{2+} provides strong evidence for the importance of screening for stabilizing conditions with DSF before crystallization. It also supports my hypothesis that the P-subdomain in TNC is basally more stable than in the ANGPTLs, which seem to require Ca^{2+} binding to sufficiently order the domain for crystallization.

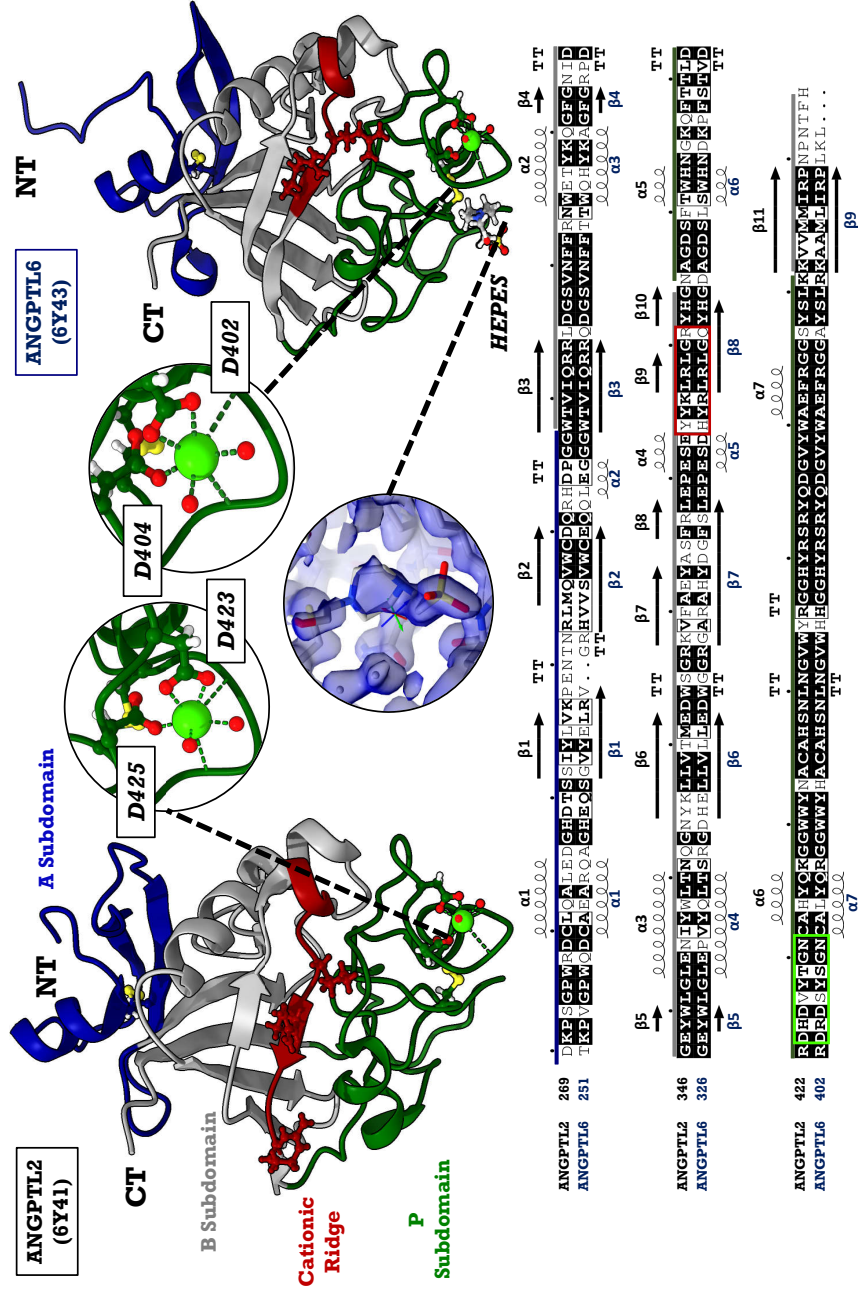


Figure 2.15 Structures of ANGPTL2 and ANGPTL6. Structures of ANGPTL2 (PDB: 6Y41) and ANGPTL6 (PDB: 6Y43) bound to Ca^{2+} , with subdomains colored as in Figure 2.10. Basic residues within the cationic ridge are shown as sticks. Circular insets highlight the Ca^{2+} binding loops and the bound HEPES molecule (2mFo-DFc electron density visualized using Clipper from the ChimeraX Toolshed [397]).

As visualized in Figure 2.15, the overall architecture of ANGPTL2 and ANGPTL6 matches the canonical “ABP” FBG domain arrangement described for TNC and the other FRePs (see Section 2.12 for a superposition), including the core seven-strand β -sheet and the P-subdomain with little secondary structure. Note that sheets marked $\beta 7$ and $\beta 8$ in ANGPTL6 are actually split in half by a short, two residue loop into two shorter β -strands in the structure of ANGPTL2. So, technically, ANGPTL2 contains a *nine* stranded core β -sheet, but as the structures show this is purely a semantic difference as the division does not change the orientation or positioning of the strands within the global β -sheet core. Additionally, ANGPTL6 contains the helix $\alpha 2$ in the A-subdomain rather than a simple turn which is present in ANGPTL2. Besides these two small differences, the domains contain identical secondary structure features.

The main difference in my ANGPTL structures is the ordering of the Ca^{2+} binding loop (Figure 2.15, circular inset), which I did not observe in the TNC structure. Yee *et al.* first characterized the FReP cation-binding loop in 1997 with the structure of human FGG bound to Ca^{2+} ; at the time, this represented a novel cation binding motif, quite distinct from canonical EF-hands [300]. The Ca^{2+} -binding loop in ANGPTL2 and ANGPTL6 is identical to FGG and other FRePs, with a total of seven oxygen atoms coordinating a single Ca^{2+} ion with near-perfect octahedral symmetry. The binding site consists of two conserved Asp residues (Asp423/Asp425 in ANGPTL2 and Asp402/Asp404 in ANGPTL6), which together provide three coordinating oxygens, along with two backbone oxygens (Tyr427/Gly429 in ANGPTL2 and Tyr406/Gly408 in ANGPTL6) and two water molecules. The entire cation binding loop is just eight residues and is stabilized by a disulfide bond immediately after the loop (Cys430/Cys443 in ANGPTL2 Cys410/Cys423 in ANGPTL6) that tethers the short helix $\alpha 6$ or $\alpha 7$ to the rest of the P-subdomain. This disulfide-containing helix is one of only two chunks of secondary structure in the entire P-subdomain, suggesting it is critical for correctly orienting the short Ca^{2+} binding loop.

Furthermore, I resolved a HEPES buffer molecule bound to ANGPTL6 directly adjacent to the Ca^{2+} binding loop (electron density shown in the circular inset of Figure 2.15). Although I expect this has no functional relevance, it could explain why HEPES best stabilized the protein in my DSF-based buffer screen. Because I used the full-length FBG-domain, I also observed the entirety of the C-terminal loop in the ANGPTLs, which is significantly shorter than the cationic extension I deleted from TNC-c042. Interestingly, the C-terminus of ANGPTL2 and ANGPTL6 actually

bends back towards the cationic ridge rather than protruding directly upward into the solvent as I modelled in Figure 2.11C.

2.11 Structure of Non-Native FGL1 Dimer

In addition to the crystallizable TNC-c040 and TNC-c042, I uncovered two highly expressed constructs of FGL1 with C-terminal truncations: FGL1-c010 (Leu74–Asp306; $\Delta 6$ residues) and FGL1-c011 (Leu74–Pro304; $\Delta 8$ residues) (see Figure 2.8). While I only obtained 0.25 mg L^{-1} of purified FGL1 using the full-length FGL1-c003, FGL1-c010 yielded 5.25 mg L^{-1} and FGL1-c011 yielded 2.75 mg L^{-1} when purified using Scheme 2 (see Figure 2.16). Similarly to TNC, truncations to remove short, flexible regions at the C-terminus stabilized FGL1 and increased purified yields. Unlike TNC, however, FGL1 exists as both a monomer and a dimer in solution and resolved into two distinct peaks during SEC as shown in Figure 2.16. FGL1 is the only FReP I purified that displayed a preference for dimerization, so I attempted to crystallize both the monomeric and dimeric forms of each truncated FGL1 construct. In 15 preliminary experiments with four different coarse screens and at $[\text{FGL1}]$ ranging from 9 to 30 mg mL^{-1} , I obtained initial crystals for only the dimer of the shorter FGL1-c011. One crystal from a MIDAS condition diffracted to $5 \text{ \AA}$ with obvious twinning, and a second crystal from a Morpheus condition diffracted to $3.5 \text{ \AA}$. An additional fan-like proteinaceous crystal from a similar Morpheus condition shattered upon mounting but appeared promising. These results are summarized in Figure 2.16C.

With promising conditions in hand, I scaled up just FGL1-c011 and optimized the crystallization using follow-up fine screens. I designed one screen sampling the chemical space surrounding the Midas condition and a second screen narrowing in on the two Morpheus conditions. Using a fresh prep of the FGL1-c011 dimer, I tested 29 mg mL^{-1} protein in both the presence and absence of Ca^{2+} . I also used some small crystal shards that appeared in a few days to prepare seeds for random matrix seeding back into the follow-up screens and the Morpheus coarse screen, this time with FGL1 at 20 mg mL^{-1} . In total, I set 14 follow-up screens, representing over 4,000 unique conditions and resulting in hundreds of crystals. I mounted 30 crystals for data collection, which yielded 13 high resolution datasets. All datasets without significant twinning or severe tNCS came from crystals grown in the absence of Ca^{2+} or seeds and in the Morpheus-based follow-up screen. The best crystal diffracted to $1.8 \text{ \AA}$ in the monoclinic space group P1211.

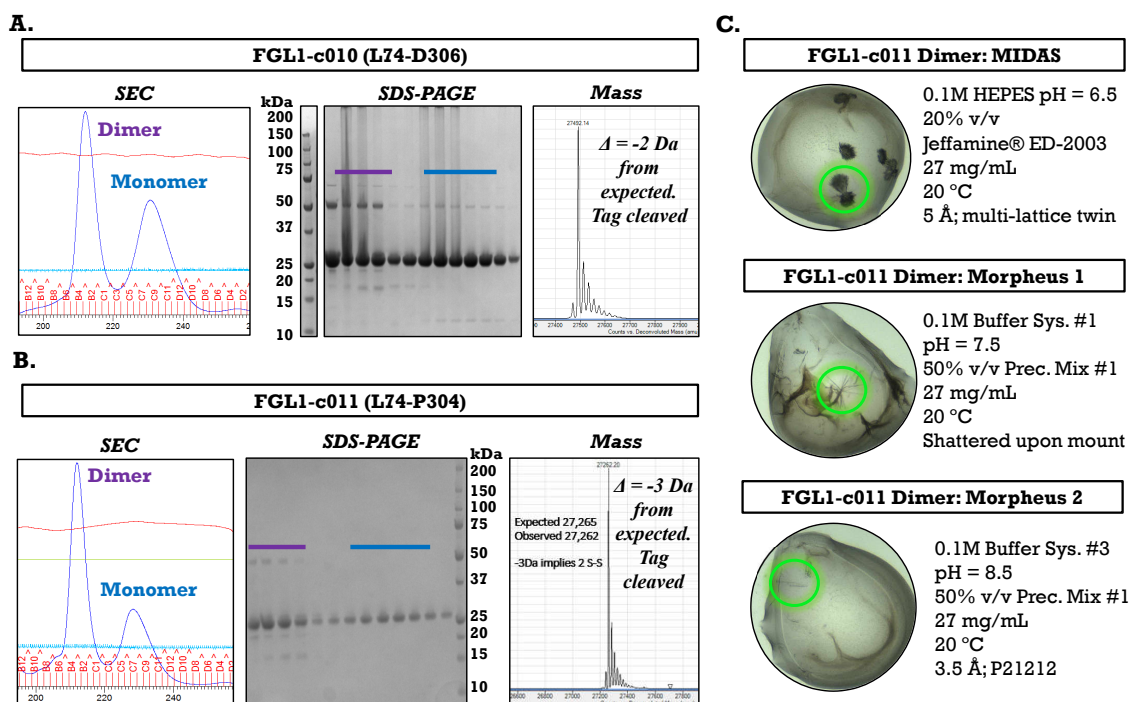


Figure 2.16 Truncated FGL1 Crystallization. **A.** Purification of 2 L of FGL1-c010 using Scheme 2. **B.** Purification of 2 L of FGL1-c011 using Scheme 2. In both A. and B., oligomeric state was assigned via comparison to known molecular weight standards. **C.** Promising initial crystallization conditions, all from the dimeric peak of the shorter construct FGL1-c011.

To solve the structure, I started by phasing the dataset using the homologous structure of ANGPTL6 followed by a round of AutoBuild and refinement in Phenix. While the best solution contained two copies in the ASU (a dimer as expected), after manual inspection it became clear that the refined solution was accurate only in the A- and B-subdomains, while the loop-filled P-subdomain did not fit in the density. Therefore, I phased the dataset a second time, this time using the partial solution containing just the A- and B-subdomains of FGL1 as a search model. Phaser confidently placed just this portion of the protein, and strong unoccupied electron density appeared for the missing P-subdomain. This region was built manually in Coot, and subsequent refinement in Phenix yielded the structure of the FGL1 dimer ($R_{\text{work}}/R_{\text{free}} = 0.24/0.28$). Additional methodological details can be found in Section 2.1.8.

Unfortunately, the additional challenge of manually building the P-subdomain arose from a non-native artefact in the structure. Instead of containing a conserved intramolecular disulfide bond in the P-subdomain like the other FRePs, the FGL1 dimer is stabilized by *intermolecular* disulfide bonds stretching between the sub-

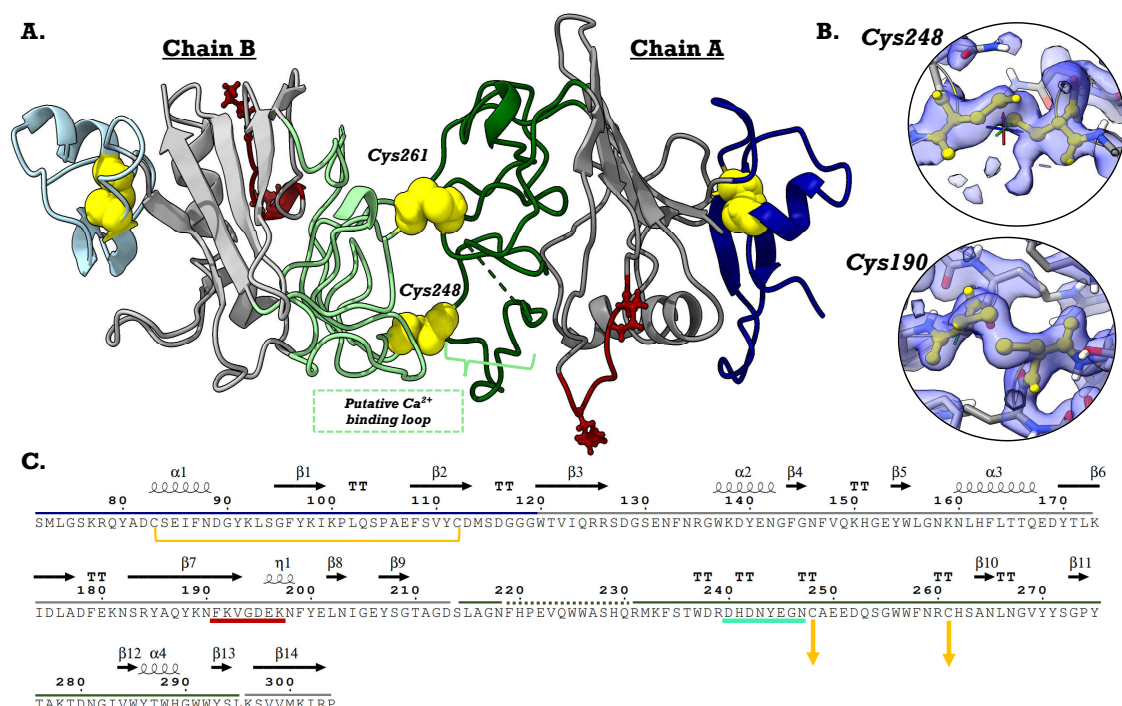


Figure 2.17 Structure of FGL1 Dimer. **A.** Structure of the FGL1 dimer, colored by subdomain as in Figure 2.10. **B.** Inset of electron density for non-native intermolecular disulfide bonds, generated with Clipper from the ChimeraX Toolshed [397]. **C.** Sequence level depiction, colored as in A. Figure generated with EndScript [387]. Intermolecular disulfide bonds at Cys248 and Cys261 denoted with yellow arrows.

units and formed between Cys248–Cys248 and Cys261–Cys261 (see Figure 2.17A). The electron density, shown in Figure 2.17C, supports the manual placement of both non-native disulfide bonds, which strongly distort the structure of the domain and explain why using ANGPTL6, with the correct disulfide bond architecture, as a search model generated such a poor solution for the P-subdomain of FGL1. Using the refined FGL1 dimer, I went on to phase the other datasets and found that they all contained the intermolecular disulfide artefact. This suggests that the incorrect disulfide bond exists prior to crystallization, rather than being formed *in situ* in specific crystallization conditions. Furthermore, there exists no evidence for a biologically relevant, disulfide-bonded FGL1 dimer in the literature. Therefore, the most parsimonious explanation is that FGL1 preferentially forms a non-native dimer that is covalently tethered during expression in CyDisCo cells and retained throughout purification and crystallization. The minor monomeric species, matching the other FRePs, likely represents the biologically relevant species—however, since the monomer did not crystallize in preliminary experiments, I unwittingly solved a

high resolution but biologically dubious dimer. Therefore, I chose not to deposit the structure presented in Figure 2.17 into the PDB.

Figure 2.17 does demonstrate that FGL1 adopts the canonical “ABP” subdomain architecture, and the A- and B- subdomains, largely unaffected by the incorrect disulfide bonds, match the structures of TNC, ANGPTL2, and ANGPTL6. The cationic ridge of FGL1, displayed in red, is resolved in the structure and does protrude noticeably. Progressing towards the C-terminus through the P-subdomain, however, the structure becomes uninterpretable. Firstly, an eleven residue segment (Phe219–Gln230) is missing from the density, indicative of unusual disorder in a region that is resolved in the other FRePs structures. Secondly, the putative Ca^{2+} -binding loop is unoccupied and not positioned to accept a metal ion because the native loop conformation is supported in the other FRePs by the missing intramolecular disulfide bond (see Section 2.10 for comparison to an occupied metal-binding loop). And, finally, there are no additional interactions between adjacent subunits, which seem to be held in a dimer exclusively by the disulfide bonds. Since the poorly conserved P-subdomain drives the biological differences between FReP family members, these observations render the structure relatively useless unless unexpected evidence suggesting a native FGL1 dimer appears. Therefore, the FGL1 structure will not be discussed further, and after repeated failures to crystallize the FGL1 monomer, it represents the extent of my efforts towards resolving the novel structure of FGL1.

2.12 Insights into the FReP Family

Subtle structural differences can help explain the diverse functional activities of the 24-member FReP family. Fourteen members now have structures deposited in the PDB; the addition of my TNC, ANGPTL2, and ANGPTL6 structures brings the total to over 50% of the family members. This high coverage provides an opportunity to structurally compare the different family members, with an eye towards integrating previously reported biochemical evidence with structural explanations, which I present in this section. I limit the analysis to big picture conclusions, with a focus on regions in specific proteins which strongly diverge from the rest of the family, in an effort to spark additional interest in the surprisingly multi-functional FBG domain. Interested readers are pointed to a more detailed review published elsewhere by Zuliani-Alvarez and Midwood [398].

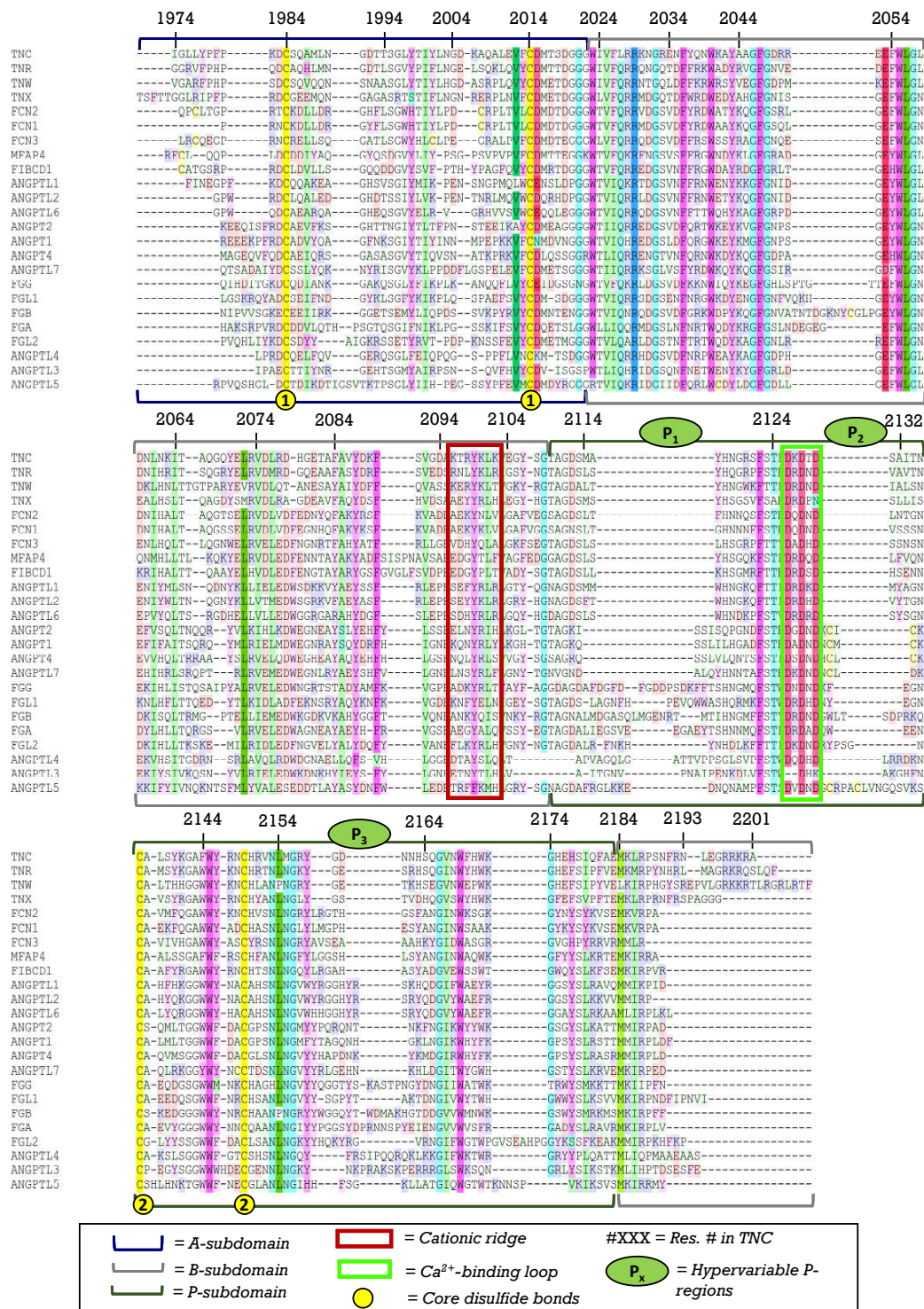


Figure 2.18 FReP Family MSA. MSA of the FReP family, generated by ICMPro and colored by residue type (color) and conservation (intensity). Numbered relative to TNC.

To orient this discussion, I provide a multiple-sequence alignment (MSA) for the family in Figure 2.18, built in ICMPro and colored according to both residue identity (color itself) and conservation across the family (intensity of the color). As before, I highlighted the A-, B-, and P- subdomains, the conserved disulfide bonds, the cationic ridge, and the Ca^{2+} -binding loop on the alignment. To maintain a frame of reference with the high-resolution TNC structure discussed in Section 2.7, and since TNC is the first protein in the alignment, I numbered the alignment according to the TNC residue position. Next, I provide a phylogenetic tree for the family in Figure 2.19, generated from the MSA using the gold-standard MrBayes v3.2 algorithm with 10,000 generations of Monte Carlo simulation [307]. On the phylogenetic tree, I annotated the FRePs with publicly available structures and I placed a box around the FRePs predicted to activate TLR4 by Zuliani-Alvarez *et al.* [303]. Note that predicted TLR4-binding activity is spread evenly throughout the family and does not seem to define clustering within the phylogeny. Finally, I overlaid one chain from each of the 14 FReP structures and colored them with a rainbow gradient according to residue number; the result can be found in Figure 2.20. These high-level, family-wide analyses provide the foundation for subsequent analyses and will be referenced throughout this section.

Already, the general patterns alluded to in Section 2.7 and 2.10 emerge when the entire family is considered. The overall fold of the FBG domain is strongly conserved, especially in the A- and B-subdomains. Not only are the two structural disulfide-bonds fully conserved in all members (Figure 2.18, yellow circles), the core beta-sheet in the B-subdomain strongly aligns ($\text{RMSD} < 1.0 \text{ \AA}$) even with 14 overlaid structures (see Figure 2.20). After a bit of variation at the N-terminus prior to the first disulfide-involved Cys1984, just a few extended loops in the fibrinogen clade prevent the A- and B-subdomains from aligning essentially without gaps across all 24 members. However, this conservation noticeably breaks down around Met2114, where the P-subdomain begins to vary extensively. Interestingly, the 96% conserved cation binding loop (green box) and the 100% conserved second disulfide bond act as anchors separating three hyper-variable regions (2114–2122; 2130–2135; and 2158–2174) in the P-subdomains. I nicknamed these regions P_1 , P_2 , and P_3 , and it is often within these hyper-variable regions in the P-subdomain that FReP-specific activities are defined structurally. In fact, for 10 of the 11 family members with known ligand or receptor binding properties, the interacting region has been mapped to the P-subdomain [398]. In the structural overlay presented in Figure 2.20, this sequence level divergence is even more obvious; in the light-green through orange P-subdomain region, the backbones cease to overlay well and instead form a tangle of loops with

few conserved secondary structure motifs. Similar in principle, if not form, to an antibody, the core FBG domain can be viewed as a rigid architectural scaffold that stabilizes and presents the loop-filled, irregular, and unique P-subdomain to achieve diverse functions.

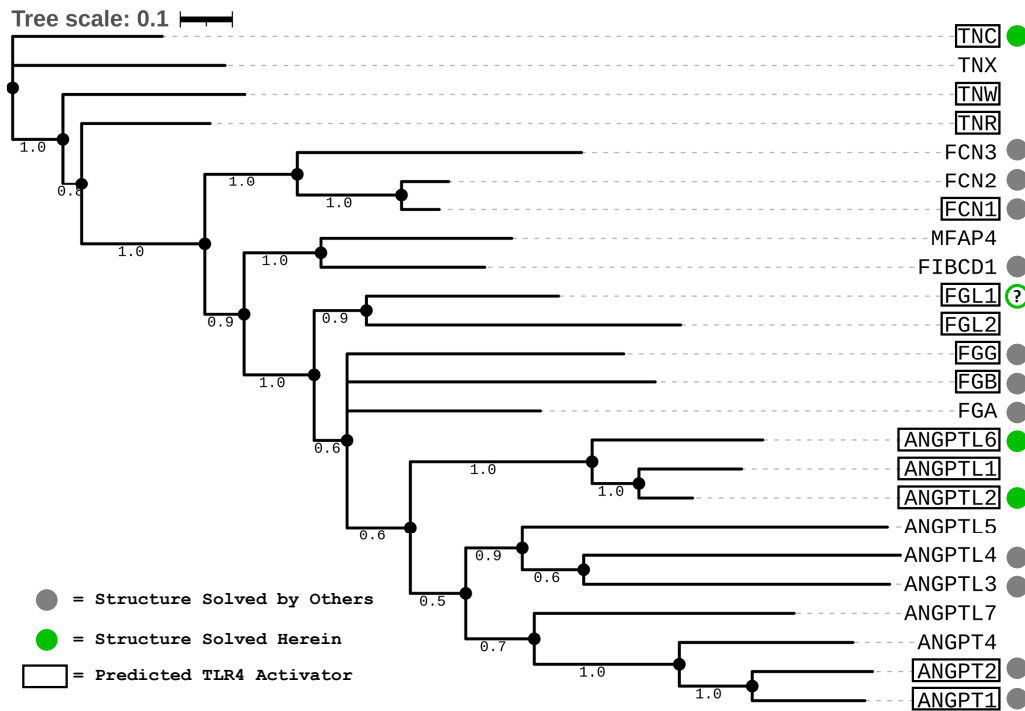


Figure 2.19 Updated FReP Family Phylogeny. Phylogeny of human FRePs reproduced from Figure 1.3 and updated with my novel structures. The MSA was generated in ICMPPro, the phylogenetic tree was calculated using MrBayes v3.2, and the figure was generated using iTOL [307, 308].

A few notable sequence-level features are worth discussing before jumping into a structural comparison. First, the cationic C-terminal extension, a key feature in TNC (4 basic residues; starting at Arg2196) that I deleted from the crystallizable construct TNC-c042, is additionally present only in the tenascin family members TNW (3 basic residues) and TNR (7 basic residues), both of which are predicted TLR4 activators. I hypothesize that these residues drive the formation of tenascin-specific interactions at either the cell surface or ECM; furthermore, given that this epitope contributes to increased TLR4 activation in TNC, it is likely that TNC, TNW, and TNR are the highest affinity TLR4 activators relative to the rest of the FReP family. This difference also suggests that truncation of the C-terminal loop of TNR and TNW could yield these structures, both of which remain unsolved.

Secondly, all of the FRePs contain a cation-binding region and are predicted to bind Ca^{2+} with the exception of ANGPTL3, which is missing the second Asp residue required for divalent metal ion chelation. Thirdly, the P_3 region has been mapped as the region responsible for integrin $\alpha V\beta 3$ binding in both TNC and FG [399]; this is also the region responsible for Tie2 binding by ANGPT1 and ANGPT2 [400]. This suggests that the P_3 region, positioned within a flexible loop, is key to defining the receptor binding properties of the FRePs; given the diversity in this region, both in the length and the residue chemistry, I expect that future work will implicate this region in the receptor-binding behavior of the 13 orphan FRePs without a mapped receptor [398]. The methodology I develop in Chapter 5 could prove crucial to de-orphanizing these 13 family members.

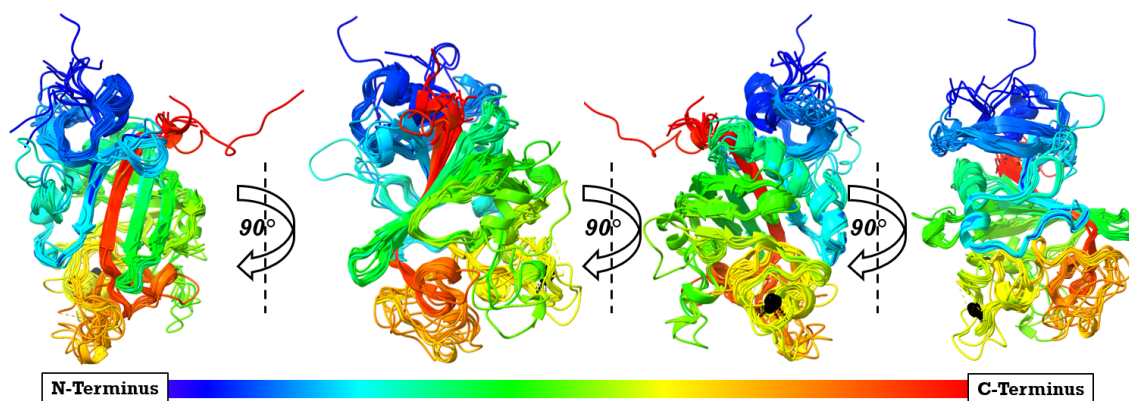


Figure 2.20 FReP Family Structures Overlay of all 14 FReP structures available as of March 2021.

Pairwise comparison of a few FReP structures can provide a structural rationale for how variations in the P_1 , P_2 , and P_3 regions differentiate FReP behavior. One intriguing comparison is between ANGPTL2, which lacks ligand binding properties and is a TLR4 activator, and ANGPTL4, which does not activate TLR4 but contains a ligand binding pocket that has been resolved structurally occupied by pentaethylene glycol (PE; PDB: 6EUB), glycerol (PDB: 6UOA), palmitic acid (PDB: 6U1U), and myristic acid (PDB: 6U73). The difference in ligand binding properties stems directly from a conformational difference in the P_1 variable domain. Figure 2.21A displays an overlay of ANGPTL2 (green) and ANGPTL4 (blue), with the P_1 loops highlighted in bright colors. The P_1 loop in ANGPTL4 protrudes outward from the domain, creating a hydrophobic ligand binding pocket underneath the loop (occupied in this structure by PE). However, in ANGPTL2, the P_1 region is packed closely to the rest of the domain and fully blocks the region where the hydrophobic pocket is located in ANGPTL4. This is not driven by the length of the region,

which is the same, but rather by secondary structure: ANGPTL2 contains a short additional β -strand that incorporates into the core β -sheet and keeps the P₁ region closely packed. However, ANGPTL4 contains a unique α -helix that, through steric hindrance, must protrude outward from the core domain to expose the hydrophobic pocket beneath. In addition to modifying ligand binding, differences in the hyper-variable P regions can also dictate differences in PPI formation. Take, for example, a comparison of TNC and FGG (PDB: 1FID), which is presented in Figure 2.21B. FGG (blue) contains extended and protruding loops (green) in both the P₁ and P₃ variable regions. The P₃ loop has been mapped as the integrin α V β 3 binding region in FGG [399], while the P₁ region is responsible for accessible presentation of a plasmin proteolytic cut site (“SDK/FFT”) [401]. Therefore, the relatively rigid core FBG structure allows presentation of these variable loop extensions in the P-subdomain of FGG to facilitate both receptor binding and post-translational processing. These two pairwise comparisons are meant to be illustrative examples of how relatively significant local structural differences in the FReP family serve to functionalize an otherwise strongly conserved domain.

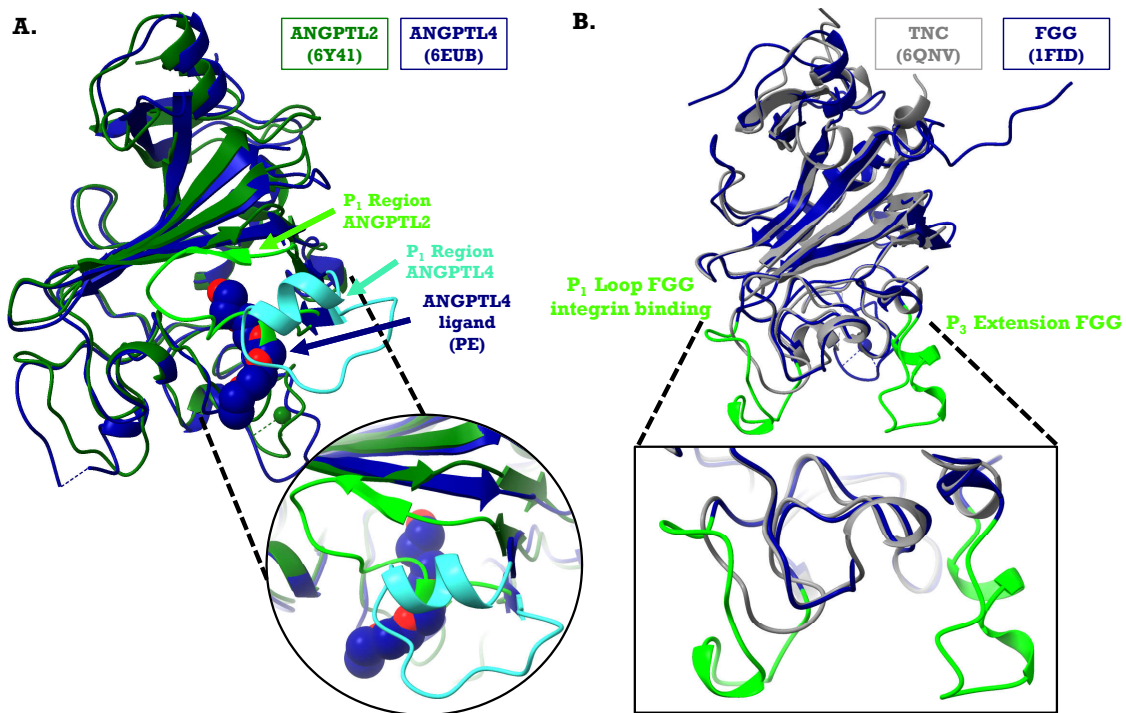


Figure 2.21 Comparison of TNC, ANGPTL1, and FGG. **A.** Overlay of ANGPTL2 (PDB: 6Y41; green) and ANGPTL4 bound to polyethylene glycol (PDB: 6EUB, blue). Inset highlights the ligand binding site. **B.** Overlay of TNC and FGG (PDB: 1FID, turquoise). Inset highlights the P₁ and P₃ extensions in FGG.

However, not all of the differences in the FReP activity can be as easily explained by large conformational differences. An interesting comparison is between the two most distantly related FReP family members, TNC (a TLR4 ligand) and ANGPT1 (a Tie 2 ligand). An overlay of these structures is presented in Figure 2.22A, with the Tie2-binding epitope in ANGPT1 marked in green and the TLR4-binding epitope in TNC marked in red. Strikingly, there are no remarkable conformational differences between these two proteins at their key receptor-binding regions, despite being the most distantly related FReP family members. Although the Tie-2-binding-region proximal loop in ANGPT1 is longer than in TNC, the actual interacting residues align neatly at the backbone level. This comparison provides an example of the subtlety of the differences, driven more by residue identity than by the structural fold, which define FReP activity. A similar observation holds elsewhere when comparing ligand binding activity. FCN1, also known as M-Ficolin, is known to bind to acetylated carbohydrates like GalNAc and Neu5Ac [402]. An overlay of TNC (grey) with the Neu5Ac-bound FCN1 (turquoise) is provided in Figure 2.22B. Yet again, despite representing a defining functional epitope, it is the side-chains of the binding site which differentiate the two FRePs; the peptide backbone aligns perfectly at the Neu5Ac-binding site. Apparently, the functional regions which differentiate the FRePs are not always simple to discern structurally, highlighting the importance of peptide-level experiments in biochemical and cellular assays. This approach was deployed by Zuliani-Alvarez *et al.* to define the TLR4-binding epitope on TNC and, while tedious, my structural analysis suggests that these kind of experiments are critical for delineating important features within the FReP family [303].

2.13 Summary and Future Directions

In this Chapter, I presented the optimized expression and purification of a small library of human FRePs, relying on the disulfide refolding CyDisCo *E. coli* cell line. My methodology, which only takes 36 hours from lysis to freezer, greatly improves upon the previously reported refolding protocol, which requires 8 days and yields significantly less protein. My robust purification workflow facilitated the generation and evaluation of a truncated construct library, which uncovered the highly expressed and crystallizable TNC-c042. Subsequent efforts yielded the novel crystal structure of TNC-c042 and 15 co-crystal structures with small molecule fragments. The truncated library also revealed the highly expressed FGL1-c010 and FGL1-c011, which seemed to preferentially form a dimer. However, crystallography revealed this

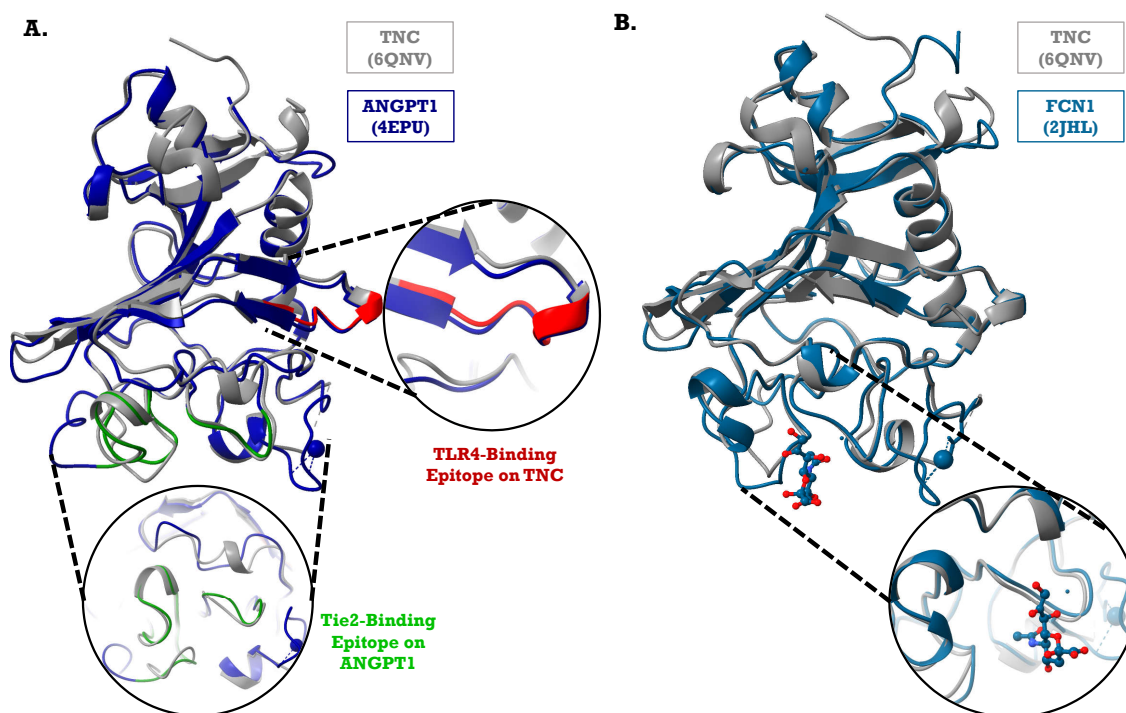


Figure 2.22 Comparison of TNC, ANGPT1, and FCN1. **A.** Overlay of ANGPT1 (PDB: 4EPU; blue) and TNC (PDB: 6QNV, grey). Insets highlight the Tie-2 binding site in ANGPT1 and the TLR4 binding site in TNC. **B.** Overlay of TNC and FCN1 (PDB: 2JHL, turquoise). Inset highlights the Neu5Ac-binding site in FCN1.

“dimer” to be an artefact due to the formation of non-native disulfide bonds. Next, DSF additive screening demonstrated the FRePs’ ability to bind Ca^{2+} , leading to the novel structures of ANGPTL2 and ANGPTL6 in complex with a Ca^{2+} ion that occupies and stabilizes a cation-binding loop. Taken together, these three novel crystal structures provide critical missing information about the TLR4-activating members of the FReP family and will prove useful for downstream drug discovery and computational efforts. Finally, I concluded the Chapter by exploring the FReP family bioinformatically, placing my new structures in context and highlighting interesting FReP-specific features that differentiate their biological activity.

It is tempting to speculate on the biological relevance of my observation that the FRePs all bind to Ca^{2+} . Extracellular Ca^{2+} is strictly maintained at 2.5 mM, with about 1.0–1.4 mM in the free ionic form unbound to carrier proteins [403]. My results imply that ANGPTL6 is mostly calcium-bound in serum, while only a fraction of ANGPTL1 and ANGPTL2 and a very small amount of TNC are calcium-bound. Furthermore, extracellular calcium levels are known to increase during specific processes, including at sites of infection and tissue repair; in fact, high Ca^{2+} levels act as

a homing signal for monocytes, and changes in extracellular Ca^{2+} levels are detected by dozens of calcium sensors with diverse and tissue-specific effects on cell behavior [404]. Extracellular increases in Ca^{2+} lead to calcification at sites of chronic inflammation [405], and Ca^{2+} itself can act as a danger signal that activates the NLRP3 inflammasome [406]. Therefore, it is worth considering that some or all of the FRePs act as Ca^{2+} sensors during inflammation, with Ca^{2+} binding variably modulating the functional activity of the FRePs. This avenue merits further investigation.

Another interesting avenue for future study seems to be the biology and activity of FGL1 and FGL2. Solving the structure of the native FGL1 monomer is obvious, and I would recommend more comprehensive additive screening and crystallization trials with the monomer fractions of both the full-length and truncated FGL1 constructs. I, as it turned out, mistakenly put most of my efforts into the dimer, and I think that a few additional months of dedicated efforts with the monomer constructs will yield the structure; in fact, in a non-COVID world, I feel confident that I would have solved the FGL1 monomer structure as well. Secondly, I am interested in both the potential biological rationale for the FGL1 dimer and the instability of the FGL2 domain. Could FGL2 require a partner for stability, and if so what is that binding partner? Why is FGL1 the only FReP in my library that formed this dimer? Could FGL1 form this strange dimer to prevent Ca^{2+} binding or to avoid proteolysis in the ECM following secretion? I still think that the dimer is artefactual, but I am optimistic that a simple purification route will encourage further work on the FGL sub-family, especially since very little literature exists regarding the innate immune activity of the FGLs. A third route for future work focuses on the pharmacological inhibition of FReP-mediated innate immune activation. My fragment screen suggests that, at least for TNC, small molecule inhibitors will be challenging to develop, especially until additional functionally relevant interactors and PPI interfaces are mapped. Therefore, I would put effort into generating biological binders for the FRePs, especially since TLR4-activation occurs extracellularly. Any biological probes could be invaluable tools for unlocking the biology of these proteins; in fact, at one point in the project I wrote a small grant (ultimately rejected) for the generation of nanobody binders for the FRePs towards this exact purpose. Yet again, my reliable soluble purification strategy will simplify such an effort, which I believe should be pursued in earnest due to the catalytic effect that reliable biological probes can have in under-explored areas like FReP biology.

Chapter 3

Expression and Purification of TLR4 via FAS2FURIOUS

This chapter focuses on the expression and purification of human TLR4 which, until this work, had never been successfully purified at a large scale without co-expression of MD-2. Initial failures using all standard in-house expression systems led me to implement secreted protein expression in *Drosophila melanogaster* Schneider 2 (S2) insect cells, which eventually facilitated the large-scale purification of TLR4, MD-2, and the TLR4:MD-2 complex. Motivated by initial successes with *Drosophila* S2 cells, I transformed a commercially available low-throughput S2 system, marketed by ExpreS²ion Biotechnologies, into a rapid, simple, and medium-throughput pipeline nicknamed FAS2FURIOUS. After careful method development, I report evidence that FAS2FURIOUS represents a general solution for expressing challenging proteins, including receptor ectodomains, secreted glycoproteins, and even intransigent intracellular enzymes. FAS2FURIOUS stands out as the fastest eukaryotic expression pipeline in use at the SGC and represents a methodological step forward for secreted protein purification.

3.1 Materials and Methods

Note that the entirety of the FAS2FURIOUS method in step-by-step detail will be published elsewhere alongside a manuscript describing key aspects of the work presented in Chapter 3. Therefore, the F2F-relevant methods described herein provide just an overview of the approach. For convenience, I have attached the

FAS2FURIOUS manuscript and the accompanying detailed methods as an Appendix.

3.1.1 Protein Accession Numbers

Below are the UniProtID accession numbers and residue boundaries for the human proteins purified in this chapter.

- **Toll-Like Receptor Four (TLR4):** UniProt O00206. Near-full length, membrane localized construct TLR4-c052 = Pro23–Gly818. Ectodomain, used in TLR4-c056, TLR4-c071, and TLR4-c072 = Pro23–Met629.
- **Lymphocyte Antigen 96 (MD-2):** UniProt Q9Y6Y9, Glu17–Asn160.
- **Monocyte Differentiation Antigen CD14 (CD14):** UniProt P08571, Thr20–Asn345.
- **CD44 Antigen (CD44):** UniProt P16070, Isoform 11, Ala20–Pro335.

3.1.2 Small-Scale Eukaryotic Test Expression

This section summarizes the methods used for the failed small-scale expression of TLR4 in standard SGC expression systems. For small-scale test expression of constructs in insect cells, the Bac-to-Bac[®] system was used [407]. For small-scale test expression of constructs in human Expi293[™] cells, the BacMam system was used [408]. These baculoviral approaches have been extensively reported by the SGC previously [328]. Briefly, LIC-cloning methods (see Section 2.1.3) were used to introduce TLR4 into either pFB-class (insect cell) or pHTBV-class (mammalian cell) vectors bearing diverse purification tags. Recombinant “bacmids” were created by recombination in DH10Bac[™] (insect) or DH10EMBacY[™] (mammalian) *E. coli* cell lines, which were generously obtained from Imre Berger and Geneva Biotech. Blue-white screening on Blue-Gal (Thermo) containing agar plates was used to identify bacmids with inserted gene cargo, and the bacmids were purified using 96-well filter plates and precipitated for concentration and sterilization. P0 baculoviruses were generated by transfection of 2 x 10⁵ Sf9 cells per mL with Insect Genejuice (Merck-Sigma) in Sf-900 II SFM medium in 24-well deep well blocks. After 3 days of growth, sterile P0 baculovirus (in the cell culture supernatant) was harvested by

centrifugation. For test expressions, 120 μ L of the P0 baculoviral stock was used to infect either Sf9, High FiveTM, or Expi293TM cells in 24-well deep well blocks. Expi293TM cells were routinely cultured in FreeStyleTM 293 media (Thermo). After three days of expression under standard culture conditions, the P1 baculovirus was harvested and stored at 4 °C while the cell pellets were purified. Section 2.1.4 describes standard small scale purification in 96-well plates, the only difference in this case being a shorter (4 min) sonication step for eukaryotic cells. Eluates from the small scale IMAC were analyzed by SDS-PAGE and putative target bands were checked using in-gel tryptic digest MSMS.

3.1.3 Medium-Scale Eukaryotic Test Expression

This section summarizes the medium-scale test expressions completed for the original TLR4 constructs in standard SGC systems, focusing specifically on the near-full-length membrane protein TLR4-c052. For medium scale test expression in eukaryotic cells, 120 μ L of P1 baculovirus was amplified further by infecting 50 mL cultures of Sf9 cells, which were cultured for three days as standard. The P2 baculovirus was collected by centrifugation and 120 μ L was used in a test expression to infect 50 mL cultures of either Sf9, High FiveTM, or Expi293TM cells grown as standard in flat-bottomed flasks. Cell pellets were harvested by centrifugation (15 min at 900 x *g*) and lysed in 5 mL of lysis buffer containing 50 mM HEPES pH = 7.5, 300 mM NaCl, 5% glycerol, and 2X Merck Protease Inhibitors. For membrane protein construct TLR4-c052, 1% detergent was used to extract the membrane fraction after sonication, and detergent was subsequently maintained at 3X critical micellar concentration (CMC) throughout the test purification. Detergents included DDM, DM, OGNG, LMNG, LDAO, and C12E8. Clarified lysates were pulled down with 250 μ L of Ni-NTA resin and washed and eluted as before. Eluates were visualized by SDS-PAGE and MSMS. Occasionally, 100 μ L of eluates were also analyzed via analytical size-exclusion chromatography on a DionexTM (Thermo Fischer) HPLC system coupled to a fluorescence detector.

3.1.4 Construction of TLR4-c071, TLR4-c072, and MD-2-c014

For clarity, please refer to Figure 3.3 for a schematic of these constructs. *Drosophila* codon-optimized human TLR4 ectodomain (UniProtKB: O00206, Pro23-Met629)

and protein A tagged MD-2 (UniProtKB: Q9Y6Y9, Gln19-Asn160) were received from Dr. Toshiyuki Shimizu (University of Tokyo) in the pMT/BiP/V5 vector backbone. These entry constructs were named TLR4-c056 and MD-2-c004, respectively. 2 µg of pExpreS²-1 and pExpreS²-2 were linearized at the multiple-cloning site according to NEB's recommendations for 3 hours at 37 °C with *Bam*HI-*HF* and *Age*I in the presence of calf intestinal alkaline phosphatase (CIP). TLR4-c056 and MD-2-c004 were double digested with the same enzymes in the absence of CIP, which liberated the entire open reading frames, hereafter called TLR4^{Untagged} and MD-2^{Tagged}. The linearized destination vectors and inserts were visualized by agarose gel electrophoresis and target bands were extracted with the QIAQuick Gel Extraction Kit. 3 µL of purified insert and 1 µL of purified vector were ligated for 30 min at room temperature with QuickLigase (NEB) and then transformed into Mach 1 cells before plating onto either Kan^R or Zeo^R LB-agar plates. Colonies were screened with the MyTaqTM Red Colony PCR Kit (Bioline) using the following sequencing primers:

pExpreS2-1-for (5'-ACAAGACAGGTTTAAGGAGAC; upstream synthetic intron)
pExpreS2-2-for (5'-CCGGAGTATAAATAGAGGCGC; upstream HSP70 promoter)
pExpreS2-rev (5'-GCGCTTGAAAGGAGTGTGTA; downstream opie pAA signal)

The new constructs were confirmed by sequencing and named TLR4-c071 (containing TLR4^{Untagged} in pExpreS²-2) and MD-2-c014 (MD-2^{Tagged} in pExpreS²-1).

TLR4-c072, containing TLR4 fused to a protein A tag, was created with a HiFi assembly. MD-2-c014 was linearized by PCR with the high-fidelity Q5[®] Polymerase (NEB) using the following primers, which remove the MD-2 portion of the construct but leave behind the C-terminal protein A tag and N-terminal BiP signal peptide:

S21-LIN-for (5'-CTGGTGCCACGCGGATC)
S21-LIN-rev (5'-CCCGAGCGAGAGGCCAACA)

TLR4^{Untagged} insert was PCR amplified from TLR4-c071 with Q5[®] Polymerase and primers containing 20 bp 5'-HiFi overhangs homologous to the destination vector. 50 ng of the linearized pExpreS²-1-MD-2-c014 destination vector was mixed at a 1:2 molar ratio with the TLR4^{Untagged} PCR-product and assembled with a HiFi assembly according to NEB's recommendations before transformation into Mach 1 cells as before. Colonies were screened with the MyTaqTM Red Colony PCR Kit (Bioline), and a single clone with the correctly sized insert was minipreped and

confirmed by sequencing. This construct was named TLR4-c072 and contains the TLR4^{Tagged} insert in the pExpres²-1 backbone.

3.1.5 S2 Cells, General Handling

Expres² cells (herein referred to simply as “S2”) were received from Expres²ion Biotechnologies and regularly maintained in EX-CELL[®] 420 (Sigma-Aldrich) supplemented with 1X Pen/Strep. S2 cells were routinely split every Monday and Friday to a density of 8×10^6 cells per mL in round bottom flasks and grown at 25 °C and 110 RPM. Cells were maintained for up to eight weeks before thawing and reviving a fresh batch, which were stored as DMSO stocks in liquid nitrogen according to Expres²ion’s recommended protocol.

3.1.6 Transient Test Expression, Expres²ion Method

S2 cells were split to 8×10^6 cells per mL by centrifugation ($400 \times g$ for 3 min) and 10 mL were seeded into 100 mL round-bottom flasks. For transient transfection, 25 µg of vector DNA and 125 µL of Expres²ion Biotechnologies’ proprietary Insect-TRx5 transfection reagent (TRx5) were mixed with the cells and rested for 15 min at room temperature. Transfected cells were incubated for 68 hours at 25 °C and 100 RPM before the supernatants were harvested by centrifugation ($4,000 \times g$ for 15 min) and purified using 100 µL bed volume of IgG Sepharose (GE). IgG columns were washed with 1X PBS and eluted with either 0.5 M Acetic Acid pH = 3.4 or 100 mM Glycine pH = 3.5 before visualization by SDS-PAGE and confirmation of putative target bands with MSMS.

3.1.7 Selection of Stable S2 Cell Lines

To select stable cell lines of well-expressing constructs, S2 cells were split to 2×10^6 cells per mL by centrifugation ($400 \times g$ for 3 min) and 5 mL were seeded into T25 TC-grade flasks. Cells were transfected with 50 µL of TRx5 and 12.5 µg of construct DNA and rested at 25 °C for 4 hours before the addition of 1 mL FBS. For a positive control, one flask was transfected with pExpres²-1-GFP. For a negative control, one flask was transfected with empty TE buffer. After 24 hours, appropriate selection reagent was added. For pExpres²-1 class vectors, 2 mg mL^{-1} Zeocin was used. For

pExpreS²-2 class vectors, 4 mg mL⁻¹ G418 was used. For co-selection of complexes, 1.5 mg mL⁻¹ Zeocin + 4 mg mL⁻¹ G418 was used. Cells were selected for 24 total days, during which every 3–4 days the cells were counted and split back to 1 x 10⁶ cells per mL in fresh EX-CELL[®] 420 + 10% FBS + selection reagent. After 24 days, all cells in the positive control were GFP-positive and all cells in the negative control were dead. Cells were subsequently transferred in 15 mL to a T75 flask with 10% FBS (without selection marker), and then scaled up again to 25 mL in a 125 mL shake flask without FBS. Stable cell lines were then scaled-up, passaged, and frozen down as described above for the wild type S2 cells.

3.1.8 Large Scale Purification from Stable Cell Lines

Every 3–4 days, up to 1 L of S2 stable cell lines growing at 25 °C and 110 RPM in flat-bottomed flasks were centrifuged in sterile containers for 10 min at 900 x *g*. The cells were resuspended in fresh EX-CELL[®] 420 with 1X Pen/Strep and returned to the incubator/shaker, while the clarified supernatant was filtered through a 0.22 µm filter before being used for purification. Stable cell lines could be harvested twice a week for up to six weeks before expression levels decreased. Supernatants were first purified by standard affinity chromatography, using a 5 mL bed volume of IgG Sepharose for protein A constructs, Strep-Tactin[®] XT High Capacity Resin (IBA) for Twin-strep-tagged constructs, or Ni-NTA (GE) resin for His-tagged constructs. Note that, before Strep-Tactin[®] XT pull-down, 3 mL L⁻¹ of BioLock biotin blocking solution (IBA) was added. IgG columns were eluted with either 0.5 M Acetic Acid pH = 3.4 or 100 mM Glycine pH = 3.5 before immediate neutralization by the addition of 1:10 (v:v) of 1 M Tris pH = 8.0 and 1:20 (v:v) of 20X PBS. Strep-Tactin[®] XT columns were eluted with 50 mM D-Biotin in 1X PBS, while Ni-NTA columns were eluted with 500 mM imidazole in 1X PBS. Eluates were concentrated in centrifugal concentrators (Amicon) and then incubated for at least 4 hours at room temperature with 10U human thrombin protease (EMD Millipore) per mg of target protein (except for MD-2, which was left tagged). The cleaved protein was subsequently polished on a Superdex 200 16/600 column (GE) pre-equilibrated in either 1X PBS (for TLR4, TLR4:MD-2, and CD14) or 25 mM bicine (pH = 8.5) + 250 mM NaCl (for MD-2). Monodisperse target peak was pooled and concentrated to approximately 5 mg mL⁻¹ before flash freezing in LN₂ and storage at -80 °C. All targets were confirmed by in-gel tryptic digest MSMS and, where possible, intact mass analysis.

3.1.9 Construction of FAS2FURIOUS Vectors

To create pExpreS²-1-CRPA, the cassette encoding BiP-MD-2-thrombin-protein A (from MD-2-c004, the original MD-2 construct from Dr. Shimizu) was excised from pMT/BiP/V5/His via *Bam*HI/*Age*I restriction sites and cloned into the MCS of pExpreS²-1 linearized with the same restriction enzymes. This approach maintained a *Bgl*III cut site in between the BiP signal peptide and MD-2 open reading frame which was present in the original MD-2-c004. The MD-2 insert, along with accessory restriction cloning scars from the original cassette, was removed via PCR and the Q5[®] Site Directed Mutagenesis Kit (NEB), which relies on blunt-end ligation of PCR-linearized vectors incubated with “KLD” Master Mix, yielding the final vector pExpreS²-1-CRPA. To create pExpreS²-1-CRPA-10H-SIII, the pExpreS²-1-MD-2-thrombin-protein-A intermediate vector (herein also referred to as MD-2-c014; see Section 3.1.4) was used as a template to add a 10His-Twin-Strep tag[®] via one cycle of PCR mutagenesis as before, and then the MD-2 insert was removed with a second cycle of PCR mutagenesis to yield pExpreS²-1-CRPA-10H-SIII. pExpreS²-1-C3PA and pExpreS²-1-C3PA-10H-SIII were created exactly as described above for their thrombin-cleavable counterparts, except the thrombin cut site was simultaneously mutated to a C3 cut site during PCR amplification. Finally, to generate pExpreS²-C3-10H-SIII, both the MD-2 insert and the protein A tag were removed from the pExpreS²-1-MD-2-Protein-A-10H-SIII intermediate vector via PCR mutagenesis. All F2F vectors were validated by Sanger sequencing. Dr. Michael Fairhead assisted me in the design and preparation of these F2F vectors.

3.1.10 HT-Cloning into FAS2FURIOUS Vectors

All cloning steps were completed in 96-well plates. 5 µg of F2F vectors were first linearized with *Bgl*III according to the manufacturer’s recommendations (NEB) and purified with the QIAQuick Gel Extraction Kit. Recall that *Bgl*III linearization opens the F2F vectors in between N-terminal BiP and C-terminal tags. Target inserts were amplified with PCR using Q5[®] polymerase, with primers containing the following vector-specific HiFi compatible 5’ overhangs:

Forward Overhang: 5’-CCTTTGTTGGCCTCTCGCTCGGG

C3 Reverse Overhang: 5’-CCCTGAAACAGAACCTCCAA

Thrombin Reverse Overhang: 5’-GCCGATCCGCGTGGCACCAG

1 μ L of purified, linearized F2F vectors were fused with 2 μ L of purified PCR-amplicons using the NEBuilder[®] HiFi Assembly according to the manufacturer’s recommendation, except that the reaction volume was decreased by 50% to just 10 μ L. Constructs were transformed and validated by colony PCR as before and one clone per construct was minipreped on a spin-column. DNA was heat sterilized for 15 min at 95 °C before being used for transfection.

3.1.11 Small-Scale Transient Expression, FAS2FURIOUS

S2 cells were split to 8×10^6 cells per mL by centrifugation ($400 \times g$ for 3 min) and 3 mL were seeded into deep-well 24-well blocks. Cells were transfected with 37.5 μ L of TRx5 and 7.5 μ g of target-containing F2F construct at a concentration above 200 ng μ L⁻¹. Transfected cells were allowed to grow at 500 RPM for 68-72 hours at 25 °C. Supernatants were harvested via centrifugation ($1,500 \times g$ for 20 min) and purified using 100 μ L bed volume of Ni-NTA resin. Using 96-well deep-well filter blocks (Nunc[®]; Merck), proteins were washed twice with 2 mL of 1X PBS + 40 mM imidazole and eluted with 100 μ L of 1X PBS + 500 mM imidazole. See Section 2.1.4 for additional purification details. Eluates were analyzed by SDS-PAGE and putative target bands were confirmed using in-gel tryptic digest MSMS.

3.2 TLR4 Production in Existing SGC Systems

3.2.1 TLR4 Ectodomain in Three Expression Systems

At the start of this project, the purification strategy and structural architecture of the TLR4:MD-2 complex was well characterized. In 2009, Park *et al.* reported the structure of the human TLR4:MD-2 heterodimer in complex with LPS [64]. Park *et al.* co-secreted the complex in High Five[™] insect cells using the TLR4 ectodomain (Glu27–Lys631), along with full-length MD-2 fused to a C-terminal protein A tag. In 2012, Ohto *et al.* published the structure of the mouse TLR4:MD-2 heterodimer in complex with both LPS and lipid IVa [409]. In a separate paper from the same group, Ohto *et al.* also characterized two SNPs in human TLR4 via crystallization of TLR4^{SNP}:MD-2:LPS complexes [101]. In both papers, Ohto *et al.* relied on secretion of the proteins from *Drosophila* S2 cells using the Invitrogen DES[™] system. Ohto *et al.* used a similar construct to Park *et al.*, relying on a tag-free TLR4 ectodomain

(Pro23–Met629) co-secreted with full-length MD-2 fused to a C-terminal protein A tag.

Based on these reports, I designed new human TLR4 constructs to facilitate the purification of TLR4 by itself and/or in complex with novel binding partners like the FRePs. Given that the TLR4 ectodomain is heavily glycosylated, contains five disulfide bonds, and had never before be purified in isolation, I expected this to be challenging *a priori*, so I cast a wide net during construct design. Figure 3.1A outlines the three TLR4 ectodomain boundaries I tested: Pro23–Met629, Glu27–Met629, and Glu27–Lys631. Next, I selected a wide variety of expression vectors, encompassing bacterial, insect, and mammalian systems, encoding a variety of affinity tags, and directing expression to multiple cellular compartments. Using standard LIC cloning methodologies (see Section 2.1.3) Leela Shrestha of the SGC Biotech Team and I cloned and tested these TLR4 constructs in either a small-scale (3 mL) or medium-scale (50 mL) expression. Throughout, I leveraged well established high-throughput SGC-modified baculoviral systems Bac-to-Bac[®] and BacMam [407, 408, 328]. Section 3.1.2 outlines the test expression protocols I implemented across the different eukaryotic systems, while Section 2.1.4 covers the protocol for bacterial test expressions. In total, I completed over 170 TLR4 expression experiments encompassing fourteen different strategies which systematically varied constructs, systems, and co-expression partners. As documented in Figure 3.1B, in all cases I failed to observe soluble expression of the TLR4 ectodomain, meaning that after a small- or medium-scale test expression I did not observe TLR4 in the elution fraction.

3.2.2 Membrane-Localized TLR4 in Insect Cells

As highlighted in Figure 3.1A, I also designed TLR4 constructs which encoded not only the TLR4 ectodomain, but also the single-pass transmembrane helix and intracellular TIR domain. One of these constructs, TLR4-c052, encoding near full-length TLR4 (Pro23–Gly818) with an N-terminal 8His-GFP tag (see Figure 3.2A), gave the only positive test expression result of any construct in any pre-established expression system. The results from a 50 mL test expression using Strategy 14 are shown in Figure 3.2B and C; the target TLR4 band is denoted with a green asterisk in the gel, and the analytical size exclusion chromatogram from analysis on a DionexTM HPLC system demonstrates the protein is monodisperse in the detergent DDM. However, despite solid expression levels at a 1 L preparatory scale as visualized by GFP-positive transduced cells in Figure 3.2D, I was unable to recover TLR4 after

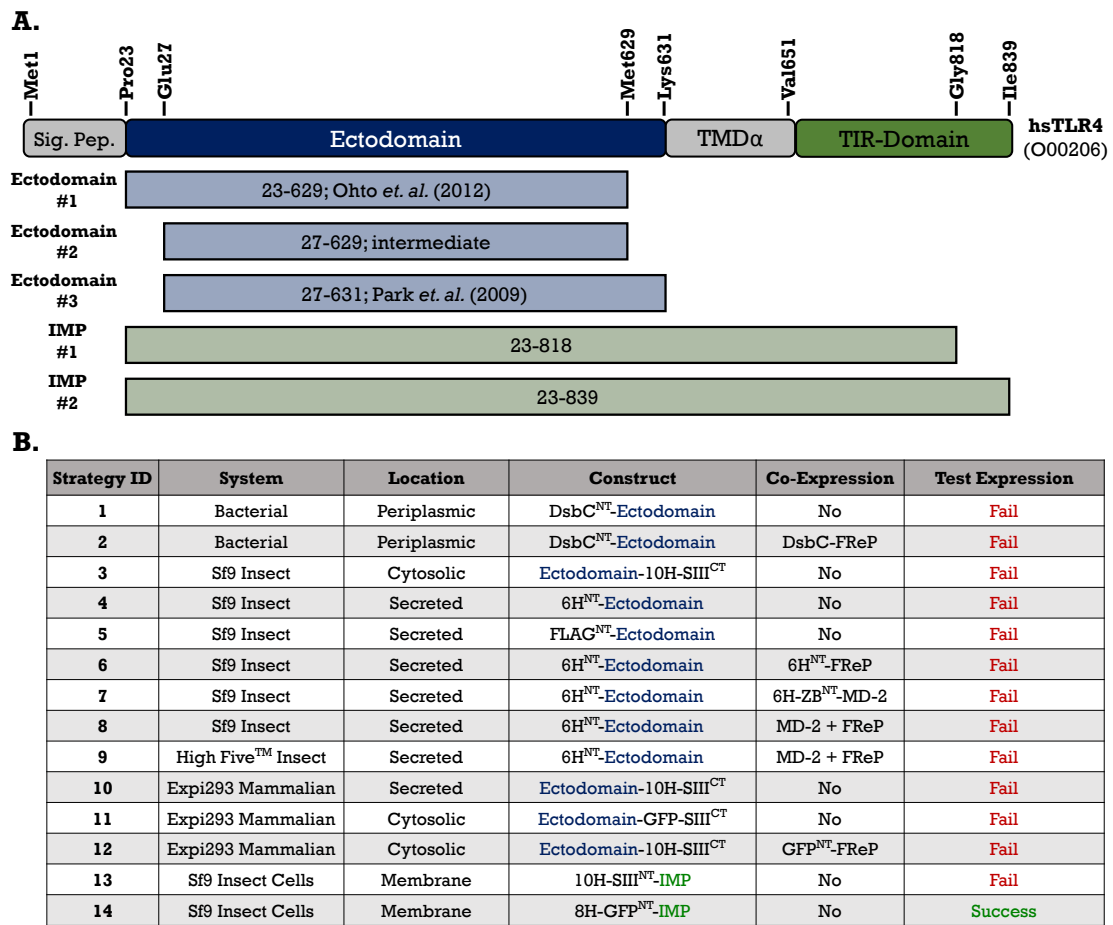


Figure 3.1 Attempted TLR4 Ectodomain Expression Strategies. **A.** Overview of TLR4 domain architecture and different construct boundaries tested. **B.** Summary of all attempted test expression strategies, totalling over 170 individual experiments.

IMAC purification despite screening ten different detergents (see the empty gel in Figure 3.2E). I repeatedly observed that TLR4-c052, containing a Type I transmembrane protein with just a single transmembrane helix, was unstable and precipitated after longer than six hours in solution, no matter the detergent used for extraction. Ultimately, therefore, the membrane-bound construct also failed to yield useful levels of TLR4 since I could not reproduce the small-scale results or achieve a practical scale-up.

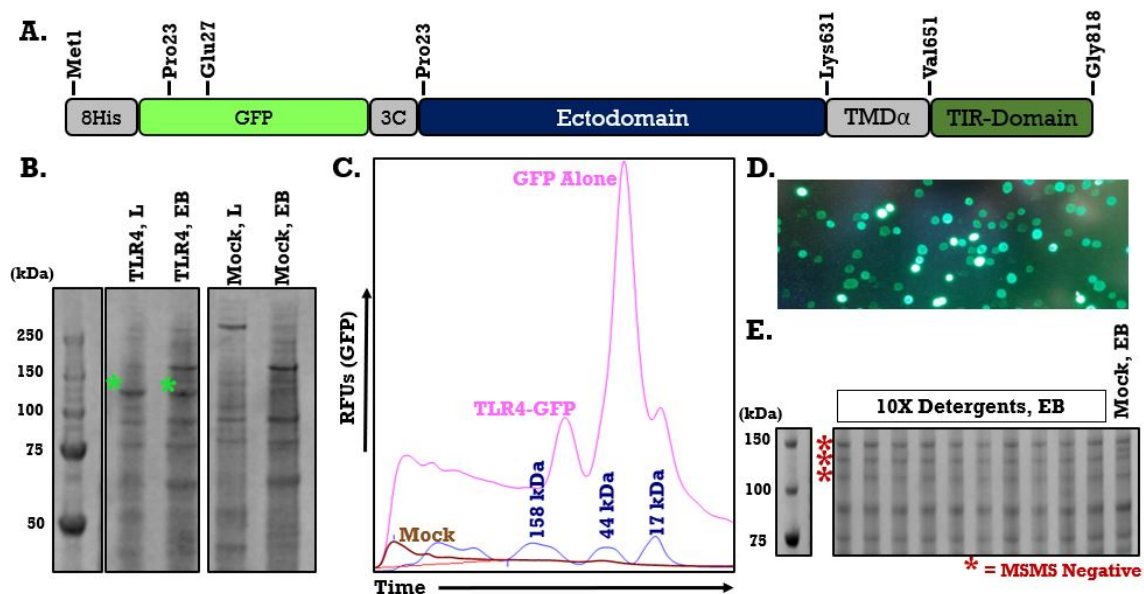


Figure 3.2 Failure to Scale-Up Membrane-Bound TLR4. **A.** Near full-length construct TLR4-c052. **B.** SDS-PAGE from 50 mL test expression of TLR4-c052 in Sf9 cells. Green asterisks denote TLR4-containing bands confirmed by MSMS. L = lysate, EB = eluate. **C.** Analytical SEC traces collected from the eluates in B on a Dionex™ HPLC system. GFP-fluorescence chromatograms shown. **D.** Over-expression of TLR4-c052 in Sf9 cells at the 1 L scale, indicated by GFP-positive cells. **E.** SDS-PAGE showing a ten detergent screen to stabilize TLR4-c052 during 1L scale-ups. Red asterisks indicate putative target bands not identified as TLR4 by MSMS.

3.3 Expression of TLR4 in *Drosophila* S2 Cells

The efforts detailed in Section 3.2 show that I failed to express or purify TLR4 with all fourteen different strategies in standard SGC expression systems. Particularly noteworthy was my inability to reproduce the previously reported co-secretion of TLR4:MD-2 from insect cells, despite trying both Sf9 (Strategy 7 and 8) and High Five™ (Strategy 9) cells. These results led me to re-consider the approach at a fundamental level and, motivated by Ohto *et.al.*'s repeatedly successful use of *Drosophila* insect cells to express TLR4, I decided to implement a *Drosophila*-based expression system for the first time at the SGC.

3.3.1 Introduction to *Drosophila* S2 Expression Systems

Because the *Drosophila*-based expression system eventually proved successful for TLR4, MD-2, and a variety of other secreted targets (see Section 3.5), it is worth

discussing in some detail the advantages of *Drosophila* S2 expression systems. Most insect cell systems rely on baculoviral transduction for target expression, with innovative solutions like Bac-to-Bac[®], *flashBAC*[™], MultiBac[™], FlexiBac, and biG-Bac facilitating the simple and high-throughput expression of challenging, multi-domain, and post-translationally modified protein targets and multi-protein complexes [407, 410, 411, 412, 413]. However, the *Autographa californica* nuclear polyhedrosis virus (AcNPV) traditionally used by baculoviral systems can only propagate in lepidopteran insect cells, meaning that these innovative expression platforms are incompatible with *Drosophila*-derived cell lines which belong to the order diptera. Despite some evidence in the literature that promoter optimization can rescue baculoviral-mediated gene expression in *Drosophila* lines, nearly all marketed insect cell systems rely on the incompatible polyhedron promoter [414, 415]

As a result of the incompatibility of *Drosophila* cell lines with common AcNPV systems, these lines are used relatively rarely, especially by academic laboratories; in fact, until recently there existed just one major commercially available *Drosophila* expression platform. Marketed by Invitrogen, the DES[™] system employs a low-throughput two-plasmid approach to transfect and select stable cell lines expressing the target(s) of interest. Recently, however, a Danish company called ExpreS²ion Biotechnologies disrupted this market with a *Drosophila* platform based on a single “pExpreS²”-class vector, which places targets under the control of a strong, constitutive fused Actin-Hsp70 core promoter with either a Zeocin^R or G418^R backbone. This one-plasmid approach simplifies the cloning, transfection, and selection stages of the workflow [416], so I selected the ExpreS²ion platform, instead of the more ubiquitous DES[™] platform, for my *Drosophila* experiments. All of the results in this and the following sections were supported by an active collaboration with ExpreS²ion Biotechnologies, who facilitated my efforts to acquire, implement, and eventually improve the ExpreS²ion system.

The lack of innovation surrounding *Drosophila* expression is unfortunate, since *Drosophila* lines possess many advantages over more ubiquitous Sf9 and High Five[™] lines. The most common laboratory *Drosophila* cell line is Schneider 2 (S2), which was derived from *Drosophila* macrophages in the early 1970’s [417]. Due in part to their small diameter, S2 cells grow much faster and to much higher densities (greater than 100 million cells per mL) than other insect cells, greatly reducing requisite culture volumes [418]. S2 cells are naturally suspension adapted, tolerate diverse serum-free and serum-containing medias, grow in all standard containers including T-flasks, shake flasks, and bio-reactors, and readily propagate anywhere

between 22 °C to 30 °C [416]. S2 cells are powerhouse protein secretors and routinely outperformed the allegedly secretion-tuned High FiveTM cells in my hands.

Furthermore, secreted expression itself has many systematic advantages because secreted proteins undergo packaging in the endoplasmic reticulum and Golgi apparatus, where resident chaperones and glycosylating enzymes help to stabilize flexible, disulfide-bonded, partially disordered, multi-domain, and/or post-translationally modified proteins. Recombinant secreted yields from S2 cells routinely surpass 30 mg L⁻¹, facilitating industrial scale production of diverse and oftentimes challenging targets [419, 420]. As a result of these advantages, S2-derived protein has led to the deposition of over 60 structures in the PDB, including receptor ectodomains (multiple TLRs), viral glycoproteins, hormones, growth factors, and more.

3.3.2 Test Expression Results

Motivated by the aforementioned advantages and the evidence that other groups could express TLRs with S2 cells, I first attempted to reproduce Ohto *et al.*'s purification of TLR4:MD-2. I obtained the original TLR4 and MD-2 constructs as a gift from Professor Toshiyuki Shimizu of the University of Tokyo, which arrived cloned into the pMT/BiP/V5 backbone for inducible expression in the DESTM system. The Shimizu constructs contained *Drosophila* codon-optimized inserts, which herein I will refer to as TLR4^{Untagged} and MD-2^{Tagged}; these coding regions are detailed for clarity in Figure 3.3A.

Using a simple restriction/ligation cloning approach with *Bam*HI and *Age*I, I transferred the TLR4^{Untagged} and MD-2^{Tagged} inserts into the multiple-cloning site (MCS) of constitutively expressed pExpres² backbones for co-expression with the improved Expres²ion platform. Figure 3.3C shows the architecture of the pExpres²-1 backbone, while Figure 3.3D shows a DNA gel from the restriction digest step of the insert shuffling (described in detail in Section 3.1.4). Note that TLR4-c071 contains TLR4^{Untagged} in a G418^R backbone, while MD-2-c014 contains MD-2^{Tagged} in a Zeocin^R backbone, facilitating the eventual double selection of a complex-expressing stable cell line. Additionally, through a second-round of cloning using PCR and a HiFi assembly (also detailed in Section 3.1.4), I created TLR4-c072, containing the TLR4^{Tagged} insert within the pExpres²-1 backbone, for expression of TLR4 by itself. To help the reader track these constructs, Figure 3.3B summarizes both the

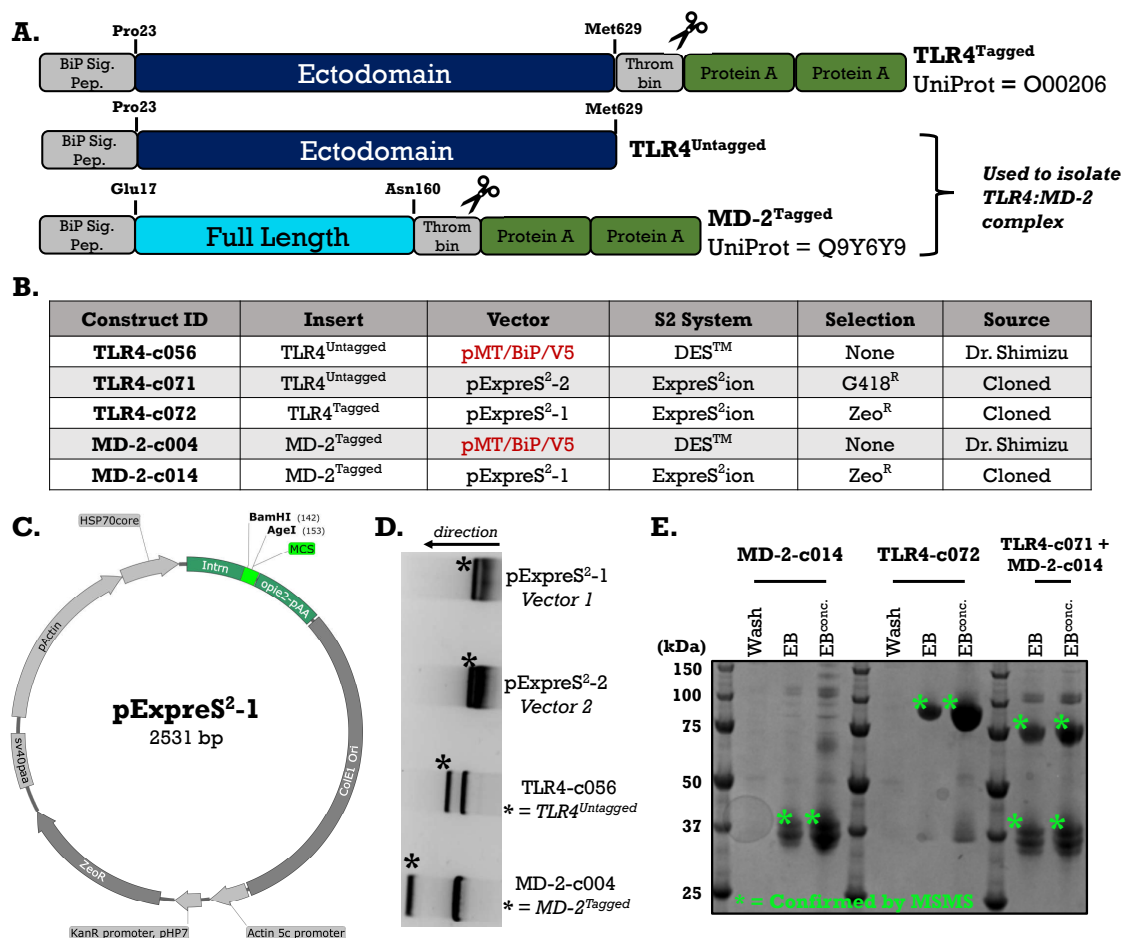


Figure 3.3 Rescuing TLR4 Expression in S2 Cells. **A.** Cartoon representation of TLR4 and MD-2 coding regions. **B.** Summary of key TLR4 and MD-2 constructs. **C.** The pExpres^S-1 vector backbone. **D.** Agarose gel showing the restriction digest step of my “insert shuttling”, wherein the coding regions from the TLR4-c056 and MD-2-c004 were transferred into pExpres² backbones. **E.** SDS-PAGE results from 10 mL transient test expressions in S2 cells.

constructs which I obtained from Dr. Shimizu and the constructs I prepared for use with the Expres²ion system.

With these new constructs in hand, I completed a transient 10 mL test expression in S2 cells according to Expres²ion’s recommended protocol (detailed in Section 3.1.6). After a three day expression, I pulled down the cell culture media with anti-protein A IgG resin and checked both diluted and concentrated eluates by SDS-PAGE. As Figure 3.3E shows, I finally reproduced the isolation of the TLR4:MD-2 complex using a co-transfection of TLR4-c071 with MD-2-c014. Furthermore, I demonstrated that both MD-2 (MD-2-c014) and TLR4 (TLR4-c072) could be

expressed and isolated as single components from S2 cells using a protein A tag. Therefore, in my hands and with these targets, *Drosophila* S2 cells outperformed all other tested systems and allowed me to reliably express TLR4 and MD-2.

3.4 Large-Scale Purification of TLR4 and MD-2

After confirming that S2 cells expressed TLR4 and MD-2 in a transient test expression, I selected polyclonal stable cell lines for downstream scale-ups according to ExpreS²ion’s recommended protocol (see Section 3.1.7). I selected and froze three lines, expressing TLR4-c072 alone (Zeo^R), MD-2-c014 alone (Zeo^R), or the complex of TLR4-c071 and MD-2-c014 (Zeo^R and G418^R). The polyclonal cell lines maintained high levels of protein secretion for up to six weeks, allowing purification “runs” with two harvests per cell line per week.

I purified these targets directly from the cell culture media using an anti-protein A pull-down on IgG Sepharose, followed by a size-exclusion clean-up step (see Section 3.1.8 for details). Because secreted proteins float in dilute culture media at elevated temperatures for 3–4 days during expression, the SEC step is imperative to remove misfolded and aggregated protein. Although it varied target-to-target, in my hands up to 50% of the affinity-captured protein proved to be aggregated or otherwise misfolded.

Throughout, I used optimized buffers selected from the same DSF-based buffer screen deployed in Section 2.9—TLR4 and TLR4:MD-2 were stored in 1X PBS while MD-2 was stored in 25 mM bicine (pH = 8.5) plus 250 mM NaCl. Before SEC, I treated TLR4 and TLR4:MD-2 with recombinant human thrombin protease to remove the protein A tag; MD-2 proved to be aggregation-prone without the tag when purified by itself, so I retained the protein A tag on MD-2 in the final protocol. I routinely observed final purified yields of approximately 1 to 3 mg L⁻¹, allowing purification of more than 20 mg of protein per six-week purification run. Figure 3.4A, B, and C show an exemplar purification from each of the three polyclonal stable cell lines, confirming that the final products are pure and monodisperse.

Next, I performed analytical SEC experiments to interrogate the oligomeric state of the purified proteins, detailed in Figure 3.4D. TLR4 exists as a monomer in solution as expected, with an implied molecular weight of 78 kDa. Interestingly, MD-2 by itself elutes at an implied molecular weight of 122 kDa, which is approximately 4-fold

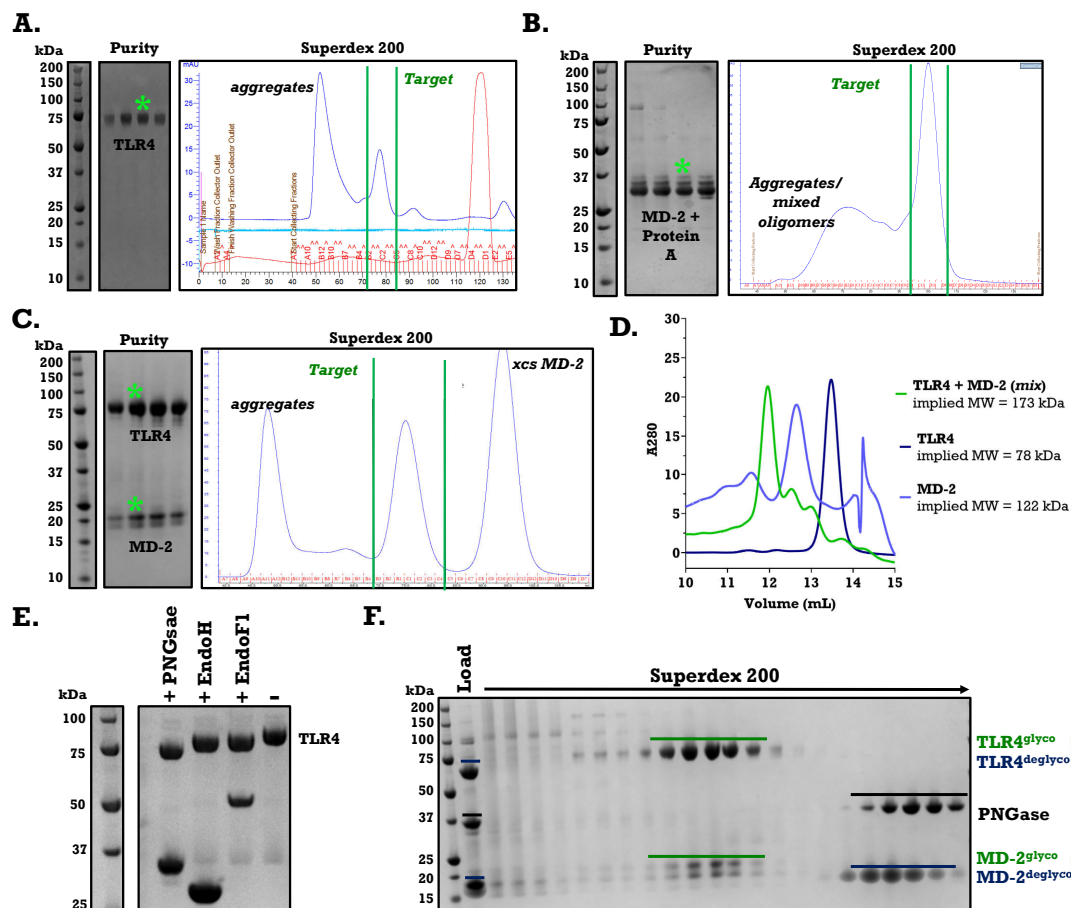


Figure 3.4 TLR4 Purification from S2 Cells. A-C Large-scale purification of TLR4, MD-2, and TLR4:MD-2 respectively. Green asterisks denote bands confirmed by MSMS. D. Analytical SEC showing that TLR4 and MD-2 expressed separately still form a complex. E. De-glycosylation analysis of TLR4 using PNGase, EndoH, and EndoF1. F. Gel showing that TLR4 and MD-2 are only de-glycosylated during SDS-PAGE. Once the PNGase is removed via SEC, both proteins elute in their “de-glyco” state.

higher than the expected monomeric weight of 32 kDa. As the extended left-hand tail of the chromatogram in Figure 3.4B suggests, MD-2 forms mixed oligomers and larger molecular weight aggregates when in solution by itself, even when the solubilizing protein A tag is retained. However, when TLR4 and MD-2 purified separately are *mixed*, I observed a larger complex with an implied molecular weight of 173 kDa, confirming that MD-2 and TLR4 retain the ability to bind even when expressed alone.

The implied molecular weight of the complex in Figure 3.4D is larger than the predicted TLR4:MD-2 dimer (approximately 110 kDa) but smaller than the predicted TLR4:MD-2 heterotetramer (approximately 220 kDa). Analytical SEC work by Ohto *et al.* shows that both the human and murine TLR4:MD-2 complexes exist in the inactive heterodimeric state in the absence of ligand [101, 100]. Furthermore, when co-expressed as in Figure 3.4C, the TLR4:MD-2 complex (without the protein A tag on MD-2) elutes as a heterodimer. I hypothesized that the protein A tag might create a biologically irrelevant 3-way complex (TLR4:MD-2:MD-2; expected molecular weight of about 140 kDa) when the separately purified proteins are mixed as in this experiment. In any case, especially considered in light of the cross-linking SDS-PAGE experiments in Chapter 4, the data show that TLR4 exists as a monomer, MD-2 as higher molecular weight oligomers, and TLR4:MD-2 exists as a ligand-free heterodimer when expressed with my S2 system.

Finally, I interrogated the glycosylation state of the proteins. As expected, Figure 3.4E shows that TLR4 is heavily but homogeneously glycosylated when secreted from S2 cells, as treatment with a variety of endoglycosidases results in a downward shift of the single TLR4 band during SDS-PAGE. Figure 3.4B shows, in contrast, that MD-2 is heterogeneously glycosylated, with multiple glycosylation states (likely 2–4) resolved as discrete bands during SDS-PAGE.

The results in Figure 3.4E suggest that treatment with endoglycosidases, particularly the most promiscuous enzyme PNGase, can trim the glycans during purification; however, this hypothesis is shown to be untrue in Figure 3.4F. The first lane shows the TLR4:MD-2 complex after incubation with PNGase but before injection onto a Superdex 200 column. In this “load” sample, both TLR4 and MD-2 resolve in their “deglyco” state, as expected in the presence of PNGase. However, once the PNGase is removed during SEC, both proteins *return* to their “glyco” form, demonstrated by an upward shift in the bands (i.e. a failure to remain downshifted) relative to the “deglyco” load sample. This counter-intuitive result suggests that

PNGase is only cleaving the glycans once the sample is diluted in SDS-containing loading buffer. I validated this result numerous times, including by mixing PNGase directly into the SDS-PAGE loading buffer and observing de-glycosylated protein on the gel (data not shown). Therefore, I was unable to replicate previous reports in the literature that endoglycosidases are capable of trimming away the attached glycans from the TLR4:MD-2 complex in solution [101, 64]. This can rightly be interpreted as a weakness of my data which requires follow-up work. All TLR4, MD-2, and TLR4:MD-2 protein used herein retains a *Drosophila*-like glycosylation state, containing mostly simple Asn-linked paucimannosidic glycans that might not reflect complex mammalian glycosylation patterns [421]. I could never convincingly remove the glycans from TLR4 or MD-2 in non-denaturing conditions despite numerous efforts with different enzymes, pH's, temperatures, and incubation times. I discuss the possible implications of this observation, and future opportunities to obtain mammalian-like glycosylation, in Section 3.6

3.5 FAS2FURIOUS: HT Expression in S2

ExpreS²ion's *Drosophila* system rapidly allowed me to express and purify TLR4, MD-2, and TLR4:MD-2. Motivated by this success, I hypothesized that protein secretion in S2 cells represented a general technological solution for the expression of challenging, disulfide-bonded, and post-translationally modified protein targets. Therefore, I transformed the existing low-throughput methods of the ExpreS²ion system into an SGC-style high-throughput expression pipeline, which I nicknamed "FAS2FURIOUS" or F2F. Herein, I describe the re-design of the pExpreS² class vectors, the parallelization of cloning and transfection protocols, and the successful pilot of the system with diverse targets submitted by multiple groups. FAS2FURIOUS is now the fastest eukaryotic expression pipeline deployed at the SGC, facilitating the cloning and transient test expression of up to forty-six constructs in just two weeks, and it performs particularly well with secreted and "secretable" targets. In an effort to spark the uptake of F2F in other groups, I plan to publish the detailed method in *Frontiers in Molecular Biosciences* and *Bio-Protocol*; for temporary convenience, the manuscript is also provided as an Appendix. Some of the text and figures in the next sections were adapted from these reports. I also completed this work in direct collaboration with ExpreS²ion Biotechnologies, who provided discounted reagents and scientific advice throughout the project. A high-level schematic overview of the method is provided in Figure 3.5A.

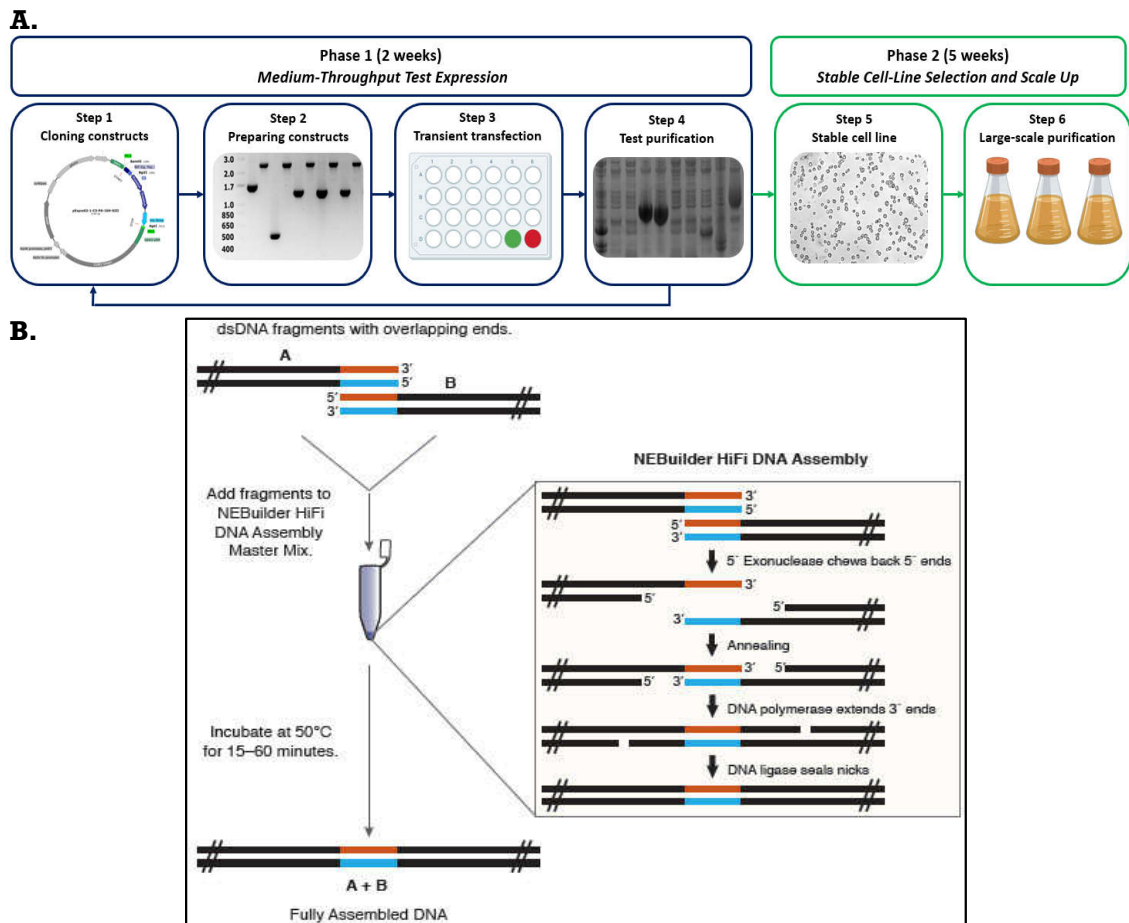


Figure 3.5 F2F Workflow Overview. **A.** Schematic of the optimized FAS2FURIOUS workflow designed and deployed herein. **B.** Overview of the chemistry of the HiFi reaction, adapted from NEB’s website (<https://www.neb-online.de/en/cloning-synthetic-biology/dna-assembly/nebuilder-dna-assembly/>).

3.5.1 Design of F2F Vectors

The original pExpres²-1 backbone (see Figure 3.3A) contains only a restriction-endonuclease-based MCS, which is unsuitable for high-throughput cloning. Furthermore, Expres²ion Biotechnologies reported in a personal communication that cloning scars between the signal peptide and the protein of interest lead to markedly decreased secretion efficiencies of certain targets, necessitating the use of a strictly seamless cloning strategy. Therefore, I made all F2F vectors compatible with the NEBuilder[®] HiFi DNA Assembly (HiFi), which uses a modified Gibson Assembly to facilitate seamless, one-step, and high efficiency assembly of DNA fragments [422]. For reference, the detailed HiFi reaction chemistry is diagrammed in Figure 3.5B. In this approach, a DNA region is amplified by PCR with primers containing 20

bp 5'-overhangs which are complementary to the desired insertion region within the destination vector. During a 30 minute reaction between a PCR product and the linearized vector of interest, T5 exonuclease creates 3'-ssDNA overhangs at the ends of the vector and insert. This frees the homologous ssDNA regions to anneal prior to extension by a thermophilic DNA polymerase to fill the gaps and subsequent nick ligation by a DNA ligase. During the HiFi reaction, short mismatches (up to 10 bp) between the homologous regions can be trivially deleted from the final construct; as discussed in detail later, this allows the removal of restriction sites which may otherwise lead to scars. HiFi assemblies are compatible with un-purified PCR products, allowing me to bypass a tedious PCR-clean-up step and assemble constructs from PCR to transformation in just one day. The main drawback of HiFi assemblies is the cost (£10 per reaction), which I partially mitigated by halving the recommended reaction volume without compromising assembly efficiency. Furthermore, the recent development of a cheap, homemade, and efficacious HiFi Master Mix should increase uptake of the HiFi approach which, in my hands, convincingly and reproducibly outperforms all other cloning strategies [423].

In addition to introducing HiFi-compatible regions, I also functionalized the pExpreS²-1 backbone with secretion and purification epitopes. All of my F2F vectors include at the N-terminus a *Drosophila* immunoglobulin heavy chain binding protein (BiP) signal sequence, which is commonly used for efficient secretion in S2 cells [424, 425, 426]; this is the same signal peptide I used in the original constructs MD-2-c014, TLR4-c071, and TLR4-c072. At the C-terminus, I added a selection of affinity purification epitopes, including a protein A tag and/or the tandem-affinity purification (TAP) 10His-Twin-Strep-tag[®]. The protein A tag and Twin-Strep-tag[®] were specifically chosen for their high affinity interaction with solid-phase resins, allowing even poorly expressed targets to be captured from dilute cell culture media. The protein A tag has a K_D of just 10 nM for IgG Sepharose and is further reported to increase the solubility and proteolytic stability of fusion partners (consistent with my observations with MD-2) [427]. The Twin-Strep-tag[®] was specifically designed for the purification of targets from cell culture media, and gentle elution from Strep-Tactin[®] affinity resin relies on excess free biotin rather than the violent pH change required for elution from IgG Sepharose. Finally, to facilitate downstream tag removal, I also introduced either a thrombin or human rhinovirus 3C-protease (3C) site before the fusion tag(s). 3C, also marketed as PreScission[™] Protease, markedly out-performs TEV protease, particularly in imidazole-containing elution buffers and at 4 °C [428]. Thrombin protease has a much higher specific activity than 3C, allowing tag cleavage with very small amounts of protease, and it performs very well in my hands for

TLR4 and MD-2. However, thrombin sometimes suffers from off-target activity and is not commercially available with an affinity tag for simple tag removal [429], which explains my choice to add 3C-compatible vectors to the F2F suite.

In between the N-terminal BiP signal peptide and the cleavable C-terminal affinity tag(s), I introduced a six base pair *Bgl*III cut site. This site facilitates F2F vector linearization and is fully removed during the HiFi assembly, resulting in a perfectly seamless fusion of the target to BiP at the N-terminus and the purification tags at the C-terminus. In total, I constructed five F2F vectors, summarized in Figure 3.6A, using standard PCR-based approaches (described in detail in Section 3.1.9) recommend by the NEB Q5[®] Site Directed Mutagenesis Kit. I validated all cloning sites and tags in all vectors using Sanger sequencing. Figure 3.6B provides a detailed look at one F2F vector, pExpreS²-1-C3-PA-10H-SIII, focusing specifically on the architecture of the cloning site. Notice that the HiFi overhangs, marked in bright blue, *exclude* the *Bgl*III cut site, leading to the removal of the “AGATCT” scar in the final construct. Together, the F2F vector set I created facilitates diverse tagging and purification strategies that will apply across disparate secreted targets.

3.5.2 Optimization of F2F Protocol

After improving the pExpreS²-class vectors, I then worked to miniaturize and parallelize the reported transient transfection protocol to facilitate medium-throughput screening of multiple constructs in a single experiment. Before I started developing F2F, ExpreS²ion Biotechnologies recommended two approaches for the transient test transfection: (1) an 8 mL transient transfection in 50 mL Falcon tubes using 20 μ g of DNA at a concentration greater than 0.6 μ g μ L⁻¹ or (2) a 10 mL transient transfection in 100 mL round bottom flasks using 25 μ g of DNA at a concentration greater than 0.6 μ g μ L⁻¹. While these protocols work well, and I successfully used the 10 mL method in preliminary experiments with TLR4 and MD-2, they are not compatible with parallel expression testing. The recommended approaches require midi- or maxi-prepping construct DNA and the cumbersome manual handling of dozens of individual culture containers. As a result, before FAS2FURIOUS I could only test 3–4 constructs per week, preventing me from deploying even a medium-throughput pipeline.

Therefore, I miniaturized the existing transfection methodology by implementing three improvements into the F2F method. Firstly, I reduced the cell volume to just

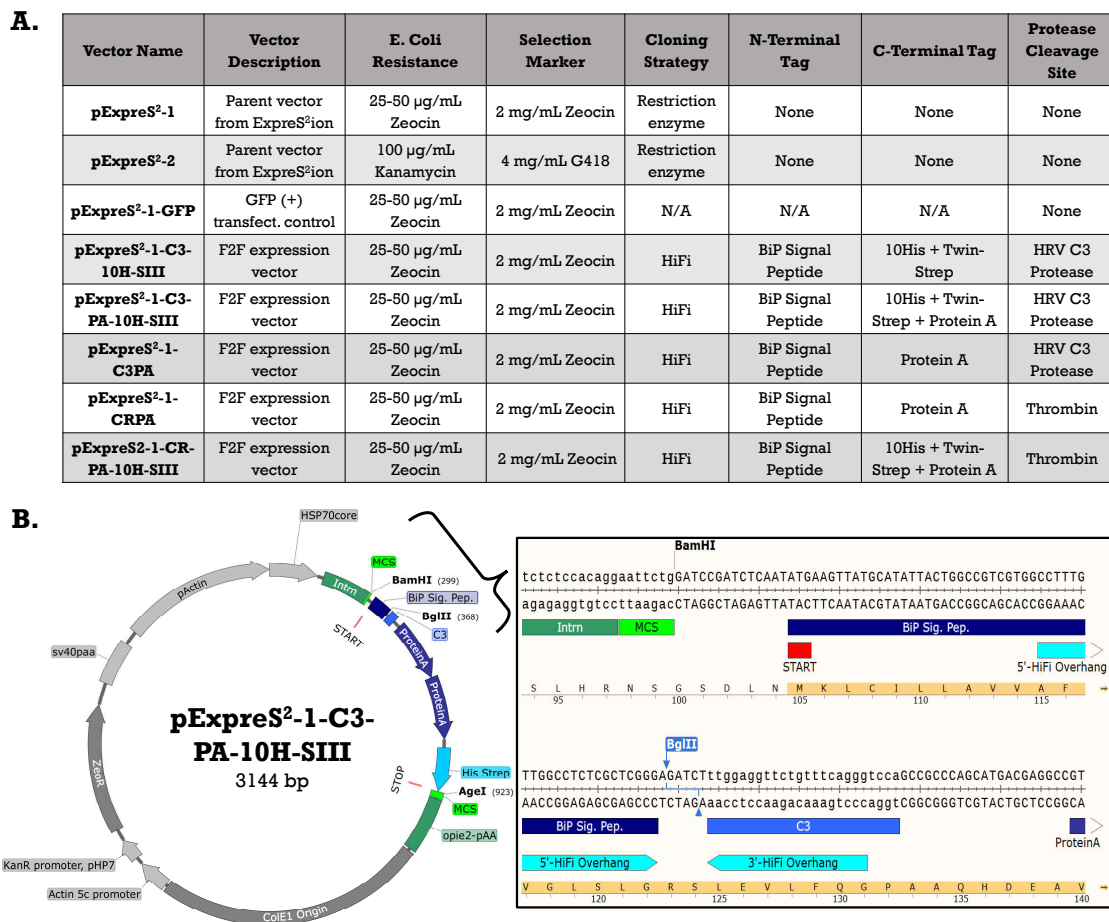


Figure 3.6 F2F Vector Development. **A.** Summary of F2F vector set. **B.** Architecture of an exemplar F2F vector, pExpreS²-1-C3-PA-10H-SIII. The rectangular inset highlights the seamless HiFi compatible cloning site.

3 mL per construct, allowing transient test expressions to be completed in 24-well deep-well blocks rather than large, individual containers. Secondly, I reduced the requisite amount and concentration of DNA by 63% and 66% respectively, down to just 7.5 µg at a concentration greater than 200 ng µL⁻¹. I routinely achieved this quantity of DNA with just a standard, spin-column-based miniprep of the F2F vectors. I also demonstrated that a brief heat-sterilization of the miniprep DNA is sufficient to avoid contamination during the transient transfection, making F2F compatible with any commercially available non-sterile miniprep kit. Finally, I implemented a 96-well filter-plate-based test purification protocol, allowing rapid pull-down of transiently expressed constructs in parallel. When combined with the high-throughput HiFi assembly, these improvements allow either twenty-two (one mini-prep cycle, one test expression plate) or forty-six (two mini-prep cycles, two test expression plates) constructs to be tested in parallel alongside appropriate controls.

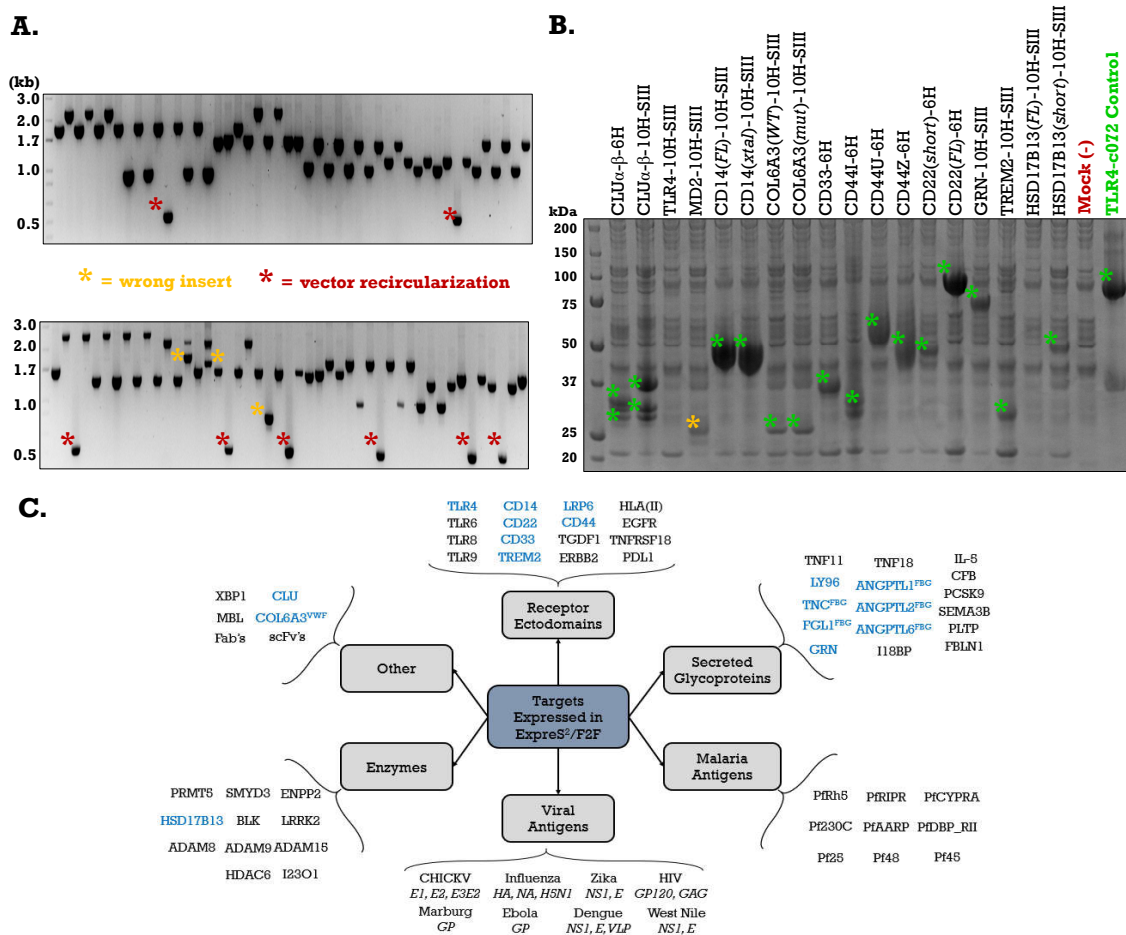


Figure 3.7 F2F Pilot. **A.** Colony PCR after HiFi assembly and transformation, demonstrating over 90% insertion efficiency. **B.** Small-scale IMAC eluates from a 3 mL transient test transfection. Green asterisks denote bands confirmed by MSMS; yellow asterisks denote putative target bands not confirmed by MSMS. **C.** Overview of targets expressed with ExpreS² cells either in-house (blue) or by ExpreS²ion Biotechnologies.

As a positive control, I used the high expressing TLR4-c072. For both visualization of transfection efficiency and as a negative control for SDS-PAGE, I used pExpreS²-1-GFP, a commercially available vector from ExpreS²ion Biotechnologies that encodes a cytosolic, tag-free GFP.

Taken together, these improvements mean the cloning and transient test transfection stages of F2F take only two weeks, critically without the tedious bacmid purification and virus amplification steps that baculoviral systems demand. Therefore, in terms of manual handling time, F2F is most similar to *E. coli* expression protocols, making it by far the most rapid eukaryotic expression pipeline deployed at the SGC.

3.5.3 Piloting F2F With Challenging Targets

To demonstrate the robustness of the new F2F method, I assembled a set of eighteen different constructs containing eleven challenging targets, many of which expressed poorly or not at all in other systems. Three groups from the SGC and one group from the Kennedy Institute suggested targets. The final “pilot” genes included receptor ectodomains like toll-like receptor 4 (TLR4), monocyte differentiation antigen CD14 (CD14), B-cell receptor CD22 (CD22), myeloid surface antigen CD33 (CD33), CD44 antigen (CD44), and triggering receptor expressed on myeloid cells 2 (TREM2), naturally secreted proteins like MD-2, collagen alpha-3(VI) (COL6A3) VWFA domain, clusterin (CLU), and progranulin (GRN), and the cytosolic enzyme 17-beta-hydroxysteroid dehydrogenase 13 (HSD17B13). I will re-visit CD14 in Chapter 4; the biological details and research interest of the other targets are outside the scope of this thesis, and I mean for this and the following section to serve as validation of the F2F method rather than as an introduction to new targets. Note that, in this pilot, I tested new TLR4 and MD-2 constructs with a 10His-Twin-Strep-tag[®] rather than simply replicating the previously successful protein A constructs. Note also that, based on literature reports, some targets (specifically CD22, CD33, and CD44) preferred only an un-cleavable His6-tag; this was achieved by the inclusion of the His6-tag in the PCR primers followed by a stop codon to prevent read through of the tags supplied by the F2F backbone. This represents yet another extension of my vectors and shows that customized C-terminal tags can be trivially added to the F2F backbones for specific applications.

Figure 3.7A shows a colony PCR following insert amplification, HiFi assembly with *BglIII*-linearized F2F destination vectors, and transformation into *E.coli*. To estimate the efficiency of HiFi assembly with the F2F vectors, I screened up to six colonies for each of the eighteen constructs. Of the 95 colonies screened, only eight (or 8.4%) contained re-circularized vector (marked with a red asterisk) without insertion of the target-containing PCR product. This is a necessary consequence of my HiFi approach which relies on a single *BglIII* cut site to linearize the vector; however, I reasoned that the incorporation of a scar-free signal peptide was more important than maximizing the insertion efficiency with a doubly-digested linear vector. In any case, a 92.6% HiFi insertion efficiency implies that screening just two colonies per construct will yield an insert-containing clone on the first try in over 99% of cases.

Figure 3.7B shows the results from a 3 mL transient test transfection of the constructs in 24-well deep well blocks using the F2F method. After expression for 72 hours, I pulled down the supernatants with 100 μ L of Ni-NTA resin and analyzed the eluates via SDS-PAGE. The positive control, TLR4-c072, was pulled down with IgG Sepharose. MSMS confirmed the expression of 15/18 constructs from 9/11 genes. In one case, HSD17B13, only a truncated “short” construct expressed. In the case of TLR4 and MD-2, the other two genes which did not conclusively express in this pilot experiment, I previously demonstrated that a protein A tag facilitates expression. Therefore, taken holistically, I was able to express all eleven pilot proteins using F2F, an impressive result given that the nominated targets were intentionally selected to be challenging. To further emphasize the diversity of targets amenable to secretion in S2 cells, I combined data from myself and ExpreS²ion Biotechnologies to create Figure 3.7C, which provides an overview of over 70 targets which have been successfully expressed in the ExpreS²ion system.

3.5.4 F2F Stable Cell Lines

To demonstrate the final step of the F2F method, I selected stable cell lines for high priority targets CD14, COL6A3^{WT}, COL6A3^{mut}, and HSD17B13. Additionally, Dr. Vittorio Katis of the Gileadi Group at the SGC independently selected stable cell lines expressing CD33 and CD44U (a particular splice variant of CD44, again outside the scope of this thesis). Note that all lines were selected essentially according to ExpreS²ion’s recommended protocol and I did not spend time optimizing this step of the process. At most, I could manage the selection of four or fewer stable cell lines at once, meaning this step is still low-throughput and represents a potential weakness in the F2F methodology that could be improved. Nevertheless, only constructs that express well in a small-scale transient test need to be transformed into a stable cell line, making the time investment worthwhile at this stage.

In all cases, the stable cell lines expressed the correct target protein on the first try, with relative expression levels implied at the small-scale reproduced at the large scale. For example, full-length CD14, which expressed very strongly at the small scale (see Lane 5, Figure 3.7B), yielded 26 mg L⁻¹ as a stable cell line. COL6A3^{WT}, which expressed much more weakly during the transient transfection (see Lane 7, Figure 3.7B), yielded just 5.4 mg L⁻¹ as a stable cell line. For large-scale purification, I often relied on an affinity pull-down with Strep-Tactin[®] Sepharose rather than Ni-NTA, which yielded purer protein. Figure 3.8 illustrates a representative large scale

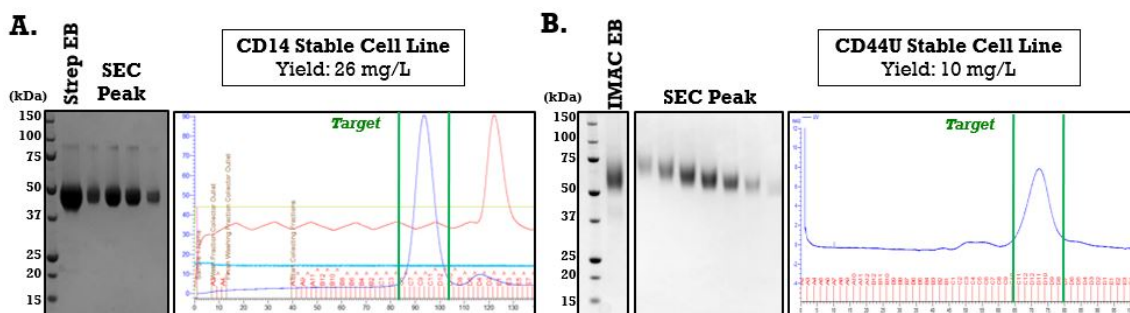


Figure 3.8 F2F Stable Cell Lines. A. Large scale purification of CD14. **B.** Large-scale purification of CD44U ectodomain, courtesy of Dr. Vittorio Katis.

purification of CD14 and CD44U from the stable cell lines to demonstrate that the proteins are pure and monodisperse. The CD44U purification was completed by Dr. Katis and is reproduced with permission.

3.6 Chapter 3 Summary

In conclusion, after failing to obtain TLR4 from existing in-house expression systems, I optimized and deployed a novel *Drosophila* S2 system nicknamed FAS2FURIOUS. The most important result from this chapter is the expression of TLR4 and MD-2 using FAS2FURIOUS. The purification of TLR4 presented herein is the first ever reported to generate structural quantities of the protein, and the ability to quickly generate large amounts of TLR4 was paramount for the experiments presented in Chapter 4. Future work should focus on de-glycosylating recombinant TLR4, which proved difficult in my hands; orthogonally, expression in “glyco-engineered” cell lines, which provide more human-like glycan patterns, should be explored. Although S2 glycosylation is mostly paucimannosidic and therefore relatively standard, a weakness in my data is the presence of non-human glycan structures on the secreted TLR4, which could modify the protein’s activity. I also hope that the ability to cheaply generate recombinant TLR4, previously only commercially available at prohibitive prices, will lead to detailed *in vitro* characterization of the protein and its activity. These type of biochemical and biophysical experiments are conspicuously missing from the literature but are vitally important to provide context to TLR4’s cellular function. The results of Chapter 4 will highlight just how important and informative simple experiments with recombinant TLR4 can be for elucidating the activity of this multi-functional innate immune receptor.

Furthermore, I extended the impact of FAS2FURIOUS beyond TLR4 to demonstrate that protein secretion in S2 cells represents a powerful method for expressing many challenging targets, particularly disulfide-bonded, post-translationally modified, or partially disordered proteins as well as receptor ectodomains and naturally secreted proteins. Over the course of my D. Phil, I or Dr. Katis created stable S2 cell lines expressing TLR4, MD-2, CD14, CD33, CD44, COL6A3, and HSD17B13, all of which yielded monodisperse protein in reasonable to excellent yields. Furthermore, the miniaturized and now moderate-throughput FAS2FURIOUS method expressed approximately 90% of the targets I tested, which is consistent with ExpreS²ion Biotechnologies' internal results (personal communication). The F2F method represents a significant improvement in S2-compatible technologies, especially in terms of throughput. F2F allows cloning and test expression of up to forty-six constructs in just two weeks and stands as the only medium-throughput method ever reported for S2 cells. To increase the uptake of my F2F system, I plan to release a detailed step-by-step protocol and make the F2F vectors freely available on AddGene upon acceptance of the F2F manuscript. I hope the work presented herein will encourage other groups to deploy the FAS2FURIOUS system, which is the fastest insect-cell expression platform currently deployed at the SGC, to express challenging secreted targets in the future.

Chapter 4

Characterization of Putative TLR4:FReP Protein-Protein Interaction

This chapter describes my attempts to re-constitute the predicted TLR4:FReP complex using recombinant proteins. Leveraging diverse experimental techniques, including enzyme-linked immunosorbent assays (ELISAs), co-expression, Native-PAGE, cross-linking SDS-PAGE (XL-SDS-PAGE), analytical SEC, and bio-layer interferometry (BLI), my goal was to assemble a TLR4:FReP complex with an eye towards structural characterization. However, the data demonstrate that a TLR4:FReP complex cannot be assembled under native conditions in solution from recombinant components. While a set of XL-SDS-PAGE experiments hint at the presence of a highly transient complex, the results presented in this Chapter halted my efforts to solve the structure of TLR4:FReP complexes, which are not amenable to structural studies. In an effort to explain this unexpected observation, I extend the analysis to show that “obvious”, previously reported TLR4-associated co-receptors do not rescue the predicted TLR4:FReP PPI. Therefore, while the results in this Chapter are all “negative”, they hint at truly unknown biology at play in the innate immune recognition of endogenous ligands, a conclusion which motivated the unbiased MS-based work to be discussed in Chapter 5.

4.1 Materials and Methods

4.1.1 TLR4 Binding ELISA

The TLR4 binding ELISA was implemented essentially as previously reported [303]. Briefly, $1\text{ }\mu\text{g mL}^{-1}$ of recombinant TNC (as in Chapter 2, FReP protein names denote just the FBG domain) was coated overnight onto Microlon Medium Binding Plates (Greiner Bio One 655080) before blocking for two hours with 10% BSA in PBST. A concentration series of TLR4, dissolved in PBST + 2% BSA, was prepared for both in-house purified *Drosophila* derived TLR4 and commercially available TLR4 from mammalian cells (Research and Diagnostic Systems). After incubating for two hours at $37\text{ }^{\circ}\text{C}$, the plate was probed for TLR4 binding using primary antibody (Invivogen, mAbg-hTLR4) in PBST + 2% BSA for two hours at room temperature. After three ten minute washes in PBST, anti-rabbit HRP-coupled secondary antibody (BioRad STAR13B) was incubated for one hour at room temperature before colorimetric readout using TMB Substrate Solution (Thermo) according to manufacturer's recommendations. OD450 was measured using a FluoStar Omega plate reader.

4.1.2 Analytical SEC

Analytical SEC experiments, such as in Figure 4.3, were completed using a Superdex 200 Increase 30/100 column connected to a DionexTM Ultimate 3000 UHPLC using PBS as the running buffer. A280 traces were collected from 100 μL injections containing 200 μg TLR4, 400 μg of the binding partner (FReP or MD-2), or 200 μg TLR4 mixed with 400 μg of the binding partner. To correct for any column drift over the course of the run, the peak corresponding to the free binding partner (smallest MW, and therefore the largest retention volume) was used to align the chromatograms. Implied molecular weight was estimated by comparison to a standard curve generated with gel filtration standards (BioRad) collected on the same day and with the same buffer.

For the ANXA2/LYZ/TNC experiment (Figure 4.9E), a Unix-C SEC 300 UHPLC Column (Sepax) was used with the DionexTM Ultimate 3000 UHPLC connected to an in-line fluorescence detector. Recombinant TNC was first labelled with NHS-AlexaFluor555 (becoming TNC-555) (Thermo Fisher) to a density of 0.1% (1/1000 free amines covalently labelled with the dye) for 30 min at $4\text{ }^{\circ}\text{C}$ before quenching any

excess dye with 10 mM Tris. 110 μ L samples were prepared in PBS containing 10 μ g of either ANXA2 or LYZ and an increasing concentration (10 to 40 μ g) of TNC-555. Fluorescent chromatograms tracking 575 nm emission upon 555 nm excitation were collected to track just the TNC-555 species.

4.1.3 Native-PAGE

1X HBS: 20 mM HEPES pH = 7.5, 150 mM NaCl.

Native-PAGE experiments were conducted with the NativePAGETM System (Novex, Life Technologies) using precast 4-16% Bis-Tris Gels according to the manufacturer's recommendations. Samples were prepared in 1X HBS and gels were run for 120 min at 150V. Samples containing 4 μ g of TLR4 and either 4 or 8 μ g (denoted in the text as "+" or "++" respectively) of the putative binding partner were prepared in a 16 μ L reaction; if indicated, CaCl₂ was added at this stage. After a 10 min incubation at 25 °C, 4 μ L of stain-free loading dye was added and then 7 μ L of the sample (containing no more than 4 μ g total protein) was loaded into the gel and analyzed.

4.1.4 BLI

Running Buffer 1: 20 mM HEPES pH = 7.5, 150 mM NaCl, 0.05% Tween-20

Running Buffer 2: 20 mM HEPES pH = 7.5, 150 mM NaCl, 0.05% Tween-20

Wash Buffer 1: 100 mM Bis Tris pH = 5.5, 1 M NaCl, 0.05% Tween-20

All BLI experiments were completed on the OctedRED384 16-Channel High Throughput Octet System (ForteBio, now Sartorius). All plates used for dipping (either 96-well or 384-well) were black (Greiner). All response curves were background subtracted using a matched reference sensor with no immobilized ligand. For biotinylated ligands (FRePs), SA Chips were equilibrated in Running Buffer 1 or Running Buffer 2 and then coated for 300 sec by dipping into 200 μ L of ligand at the requisite concentration (between 1 and 5 μ g mL⁻¹) to obtain a low and recommended loading density between 1 and 5 nm. Coated chips were dipped sequentially for 300 sec into a 3-fold increasing concentration of analyte (150 μ L of analyte at each concentration) diluted in matching Running Buffer, followed by a 300 sec dissociation step in Running Buffer. Between each association/dissociation step at each concentration of analyte, the chips were washed for 60 sec in Wash Buffer 1 to ensure return to

baseline. For covalent immobilization of un-tagged ligands (TLR4), Amine Reactive 2nd Generation Chips (AR2G; ForteBio) were used to covalently immobilize the ligands via coupling to free amines. Loading of the chip was performed according to the manufacturer's recommendations using EDC/NHS to activate, $20\text{ }\mu\text{g mL}^{-1}$ analyte at $\text{pH} = 4.0$ to load, and excess ethanolamine to quench. Curves were collected as before, with 300 sec association/dissociation phases followed by a 60 sec wash between each concentration point. Data analysis, background subtraction, and equilibrium curve fitting (assuming a standard 1:1 binding stoichiometry and steady-state affinity kinetics), were performed with the Octet Data Analysis Software.

4.1.5 XL-SDS-PAGE

$5\text{ }\mu\text{g}$ of each protein was mixed in a $19\text{ }\mu\text{L}$ reaction in 1X HBS. If indicated, between 1 and 25 mM CaCl_2 or EDTA was also added from 1 M stocks before a 15 min equilibration with gentle rocking at 25°C . The indicated cross-linker (either BS^3 , $\text{BS}(\text{PEG})_9$, or LC-SDA) was added to a final concentration of 1 mM by adding $1\text{ }\mu\text{L}$ from a freshly prepared 20 mM stock, and the cross-linking reaction was allowed to proceed for 30 min at 25°C with gentle rocking. The excess cross-linker was quenched with $1\text{ }\mu\text{L}$ of 1 M Tris $\text{pH} = 8.0$ prior to the addition of $5\text{ }\mu\text{L}$ of 4X SDS-PAGE Loading Dye. For the LC-SDA cross-linked samples, the reactions were irradiated with 8 W, 254 nm UV-light for 15 min to cross-link the UV-reactive portion of the heterobifunctional cross-linker before adding the Loading Dye. $15\text{ }\mu\text{L}$ of the samples were analyzed via SDS-PAGE, and putative complex bands of interest were validated with in-gel tryptic digest MSMS.

4.1.6 Purification of CD44 Link Domain

This work was completed by Dr. Vittorio Katis and his protocol is reproduced here with permission. The ectodomain of human CD44 containing only the Link domain (UniProtID: P16070 Isoform 1, Ala20-Asn172), referred to hereafter as "CD44-Link", was cloned by LIC into pNIC28-Bsa4. This places CD44-Link downstream of a TEV-cleavable 6X-His tag. CD44-Link was expressed as inclusion bodies in 1 L of Terrific Broth from BL21(DE3) Rosetta cells. After harvest, the inclusion bodies were resuspended into 20 mL of 20 mM Tris $\text{pH} = 8.0$. Washing of the inclusion bodies (up to 4 times) was performed by sonication (10 min), centrifugation (9000 x

g, 20 min, 4 °C) and resuspension into 30 mL of 50 mM Tris pH = 8.0, 100 mM NaCl, 0.5% Triton X-100. A final wash in 50 mM Tris pH = 8.0, 100 mM NaCl was performed to remove Triton X-100. The protein was solubilized in 10 mL of 8 M Urea, 50 mM MES pH 6.5, 0.1 mM EDTA, 0.1 mM DTT at 4 °C overnight on a rotating wheel. CD44-Link was refolded over 2 days at 4 °C by rapid dilution into 250 mM L-arginine, 100 mM Tris-HCl pH = 8.5, 2.5 mM reduced glutathione, 2.5 mM oxidized glutathione, and a 1:200 dilution of Complete EDTA-free protease inhibitors (Merck) to give a final protein concentration of 50 $\mu\text{g mL}^{-1}$. The refolded protein was passed through a 0.2 μm filter before concentrating using a Vivaflow 200 (Sartorius) tangential flow filtration unit (to 50 mL), followed by an Amicon centrifugal filtration unit. Further purification and removal of aggregates was performed on a Superdex 75 16/600 column into 20 mM Tris pH = 8.0, 150 mM NaCl. Peak fractions were concentrated to 7 mg mL^{-1} prior to flash freezing and storage. Protein identity was validated by intact mass spectrometry, revealing correct formation of the 3 disulfide bonds within CD44-Link.

4.1.7 Purification of S100A8, S100A9, and Calprotectin

Lysis Buffer 1: 50 mM HEPES pH = 7.5, 500 mM NaCl, 5% glycerol, 1 mM TCEP, 1 mg mL^{-1} hen egg-white lysozyme, 1X benzonase, 1X Merck Protease Inhibitor Cocktail. **Lysis Buffer 2:** 50 mM Tris pH = 8.0, 100 mM NaCl, 2 mM TCEP, 1 mM EDTA, 0.5% Triton X-100, 1X Merck Protease Inhibitor Cocktail. **Solubilization Buffer 1:** 50 mM Tris pH = 8.0, 100 mM NaCl, 2 mM TCEP, 1 mM EDTA, 4 M Guanidine HCl. **Dialysis Buffer 1:** 20 mM HEPES pH = 8.0, 1 mM EDTA. **IEX Elution Buffer:** 20 mM HEPES pH = 8.0, 1 mM EDTA, 1 M NaCl. **Gel Filtration Buffer:** 20 mM HEPES pH = 8.0, 100 mM NaCl, 1 mM TCEP, 0.5 mM EDTA.

S100A8, S100A9, and CP were purified essentially as previously reported [240]. NT-His6-tagged, full-length human S100A8 (UniProtID: P05109) and S100A9 (UniProtID: P06702) expression constructs were obtained from Dr. Ritika Sethi (SGC Oxford) as a gift. The constructs were transformed into BL21(DE3) Rosetta *E.coli* strain and expressed in LB to OD = 0.8 before 0.5 mM IPTG induction overnight at 18 °C. For S100A8, 1 L of cell pellet was harvested and lysed in 25 mL of Lysis Buffer 1 for 30 min at 25 °C, followed by the addition of 1% Triton X-100 and overnight freezing at −80 °C. The supernatant was clarified with centrifugation after thawing and soluble S100A8 was then purified via IMAC with NiNTA resin before a final polishing step on a Superdex 75 16/600 column into Gel Filtration Buffer.

The monodisperse, dimeric peak was collected, pooled, and concentrated in a 3.5 MWCO Amicon centrifugal concentrator before storage at -80°C . The final yield of S100A8 was 31 mg L^{-1} .

To prepare S100A9 and CP (heterodimer of S100A8:S100A9), 1 L of S100A9 cell pellet and 1 L of S100A8 cell pellet were mixed and lysed via sonication in Lysis Buffer 2. Following centrifugation at $14,000 \times g$ for 20 min, the soluble fraction was discarded and the inclusion bodies were further purified. The inclusion bodies were washed twice with Lysis Buffer 2 by cycles of sonication (5 min, 30 sec on, 10 sec off) and centrifugation ($14,000 \times g$ for 20 min). After washing, the white pellet was resuspended until clear in 35 mL of Solubilization Buffer 1 prior to sonication and centrifugation as before. The clarified supernatant, containing denatured S100A8 and S100A9, was refolded by serial dialysis (3 x 4 L, 24 hours each) into Dialysis Buffer 1. After refolding, the protein was filtered and loaded onto a 1 mL HiTrap Q column equilibrated in Dialysis Buffer 1. CP was separated from excess S100A9 with a 90 mL gradient elution from 0–40% of IEX Elution Buffer. The first eluted peak (in low salt) containing CP was pooled separately from the second peak (in high salt) containing just S100A9. The two samples were polished on a Superdex 75 column and stored exactly as above. The final yield of S100A9 was 12.5 mg and the final yield of CP was 4.8 mg. The identity of all three proteins was confirmed by intact mass analysis and in-gel tryptic digest MSMS.

4.1.8 Purification of ANXA2

Lysis Buffer 1: 20 mM HEPES pH = 6.5, 0.1 mM TCEP, 0.1 mM EDTA, 1X Merck Protease Inhibitor Cocktail. **IEX Elution Buffer:** 20 mM HEPES pH = 6.5, 0.1 mM TCEP, 1 M NaCl. **Gel Filtration Buffer:** 20 mM HEPES pH = 8.0, 300 mM NaCl, 0.1 mM TCEP.

The ANXA2 expression construct, containing full-length, un-tagged human ANXA2 (UniProtID: P07355) with the point mutant A66E was obtained as a gift from Dr. Peter Ecsedi at Etvos Lorand University in Hungary. The protein was purified essentially as described previously [430]. The construct was transformed into the BL21(DE3) Rosetta *E. coli* strain and grown in LB to $\text{OD}_{600} = 0.75$ before induction with 0.5 mM IPTG and expression for 3 hours at 37°C . 3 L of cell pellet was lysed with sonication in Lysis Buffer 1 followed by centrifugation for 50 min at 20,000 RPM in a JA 25.5 rotor. Clarified supernatant was re-adjusted to pH = 6.5

by the addition of concentrated HCl, filtered through a 0.8 μm filter, and applied to a tandem 5 mL HiTrap Q column connected to a 5 mL HiTrap S column. The sample first passed over the Q column, where negatively charged *E. coli* contaminants were removed, and then passed over the S column, where the positively charged ANXA2 bound. After sample application, the Q column was removed and ANXA2 was eluted from the S column with a 180 mL gradient from 0-50% with IEX Elution Buffer. The fractions containing ANXA2 were pooled and loaded directly onto a Superdex 200 16/600 column equilibrated in Gel Filtration Buffer. The monodisperse peak containing ANXA2 was pooled and concentrated with a 10,000 MWCO Amicon Concentrator prior to flash freezing and storage at -80°C . The final yield was 1.1 mg L^{-1} , and the protein identity was confirmed with intact mass and in-gel tryptic digest MSMS.

4.2 Denaturing ELISAs Show a Complex

4.2.1 Quality Control of New Reagents

As described in detail in Chapter 2 and Chapter 3, I employed novel methods for the expression of both the FRePs (CyDisCo) and TLR4 (*Drosophila* S2). Therefore, before embarking on downstream interaction assays, I first validated my own reagents in an established TLR4-binding assay. In this ELISA, diagrammed in Figure 4.1A and previously reported by Zuliani-Alvarez *et al.* [303], the FReP of interest is coated onto the surface of a 96-well plate and, after blocking with BSA, a concentration series of TLR4 is added over the top. Following thorough washing, any remaining TLR4 bound to the FReP of interest is quantified using an anti-TLR4 primary antibody and an HRP-coupled secondary antibody.

Using this assay, I compared the performance of refolded TNC and commercially available mammalian-cell-derived TLR4 (replicating the results of Zuliani-Alvarez *et al.*) with in-house, CyDisCo-derived TNC and S2-derived TLR4. As demonstrated in Figure 4.1B, my homemade proteins performed as well as or better than the previously reported reagents with respect to binding affinity; in fact, my homemade combination of S2-derived TLR4 and CyDisCo-derived TNC showing the strongest apparent binding affinity ($K_D = 160\text{ nM}$). The commercially available TLR4 binds to TNC with $K_D \approx 435\text{ nM}$, while while S2-derived in-house TLR4 binds more strongly

with a $K_D \approx 175$ nM. The refolded and CyDisCo-derived TNC samples behave similarly, although there is a less pronounced trend towards tighter K_D 's using CyDisCo derived TNC. The affinity I measured broadly corresponds to previously reported values; Zuliani-Alvarez *et al.* reported a $K_D \approx 50$ nM in an identical assay. Furthermore, when fit using an unconstrained saturation binding model, all reagent combinations had a Hill Slope ≈ 1.0 , suggesting a 1:1 binding stoichiometry between TLR4 and TNC. Therefore, this experiment certified the use of the TLR4-binding ELISA in my hands, reproduced the observed 1:1 TLR4:TNC interaction predicted in the literature, and validated the utility of my CyDisCo-derived FRePs and S2-derived TLR4 for downstream experiments.

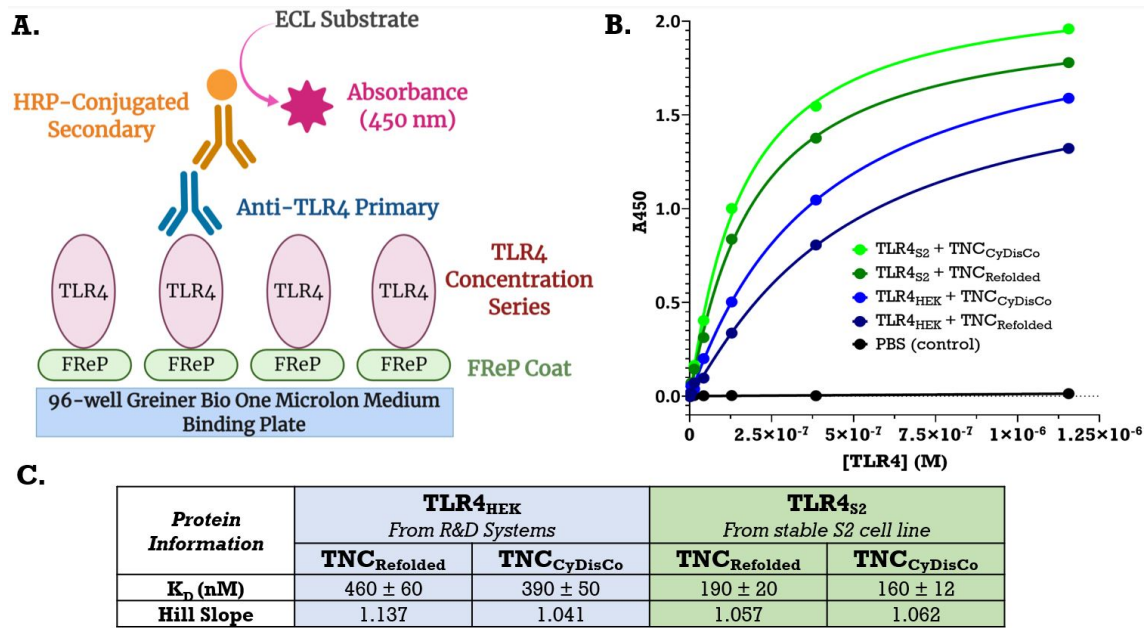


Figure 4.1 TLR4 Binding ELISA QC. **A.** Schematic of the TLR4-binding ELISA. **B.** Comparison of homemade reagents (S2-derived TLR4; CyDisCo-derived TNC) to previously reported reagents (mammalian-cell-derived TLR4; refolded TNC). $n=1$ due to cost of mammalian-cell-derived TLR4 from RD Systems. **C.** Estimate of K_D 's from B. Fit in GraphPad using a variable slope saturation binding model.

4.2.2 Binding of Other FRePs

After demonstrating that my new reagents behaved well in the TLR4-binding ELISA, I proceeded to test more members of the FReP family, particularly focusing on the ANGPTLs. As discussed in Chapter 1, Zuliani-Alvarez *et al.* predicted that ANGPTL1, ANGPTL2, and ANGPTL6 should bind to and activate TLR4, while

ANGPTL4 should not [303]. In fact, Zuliani-Alvarez *et al.* analyzed ANGPTL4 in the TLR4-binding ELISA and detected no interaction. Their hypothesis is consistent with unpublished results from our group demonstrating that ANGPTL1, ANGPTL2, and ANGPTL6, but not ANGPTL4, can stimulate the release of pro-inflammatory cytokines from primary human macrophages in a TLR4-dependent manner.

I purified the exact ANGPTL4 construct used by Zuliani-Alvarez *et al.*, containing the full-length ANGPTL4 FBG-domain with an uncleavable His6-tag, using Scheme 2 (see Section 2.4). I then analyzed this protein alongside TLR4-activating FRePs ANGPTL1, ANGPTL2, ANGPTL6, and TNC. In addition to the standard PBS-coated reference, I also used an unrelated protein (*cdClpP*, a bacterial protease purified for different applications in our group) as an orthogonal negative control. Figure 4.2A shows a representative ELISA, completed in technical triplicate and fit using a simple saturation binding model, while Figure 4.2B shows a comparison of the calculated K_D 's from at least three biological replicates.

All K_D 's fall between 0.19–1.1 μM , with a consistent trend in affinities across all replicates: $\text{TNC} > \text{ANGPTL2} > \text{ANGPTL6} > \text{ANGPTL1}$. However, a simple one-way ANOVA demonstrates only one statistically significant difference in the affinities for TLR4, between ANGPTL1 and TNC. Therefore, I concluded that, in this ELISA, TNC binds the strongest to TLR4, while the ANGPTLs all bind with a similar and weaker affinity. Surprisingly, I could not reproduce the previously reported negative result for ANGPTL4; in fact, ANGPTL4 was the tightest non-TNC TLR4 binder in my assay. In subsequent experiments, I even analyzed refolded ANGPTL4 directly from the lab of Zuliani-Alvarez *et al.* and still observed strong binding to TLR4 ($K_D = 0.3 \mu\text{M}$). Therefore, in my hands, the TLR4 ELISA could not differentiate between activating and non-activating FRePs. This is the first of what will be many hints at the complexity of TLR4 signalling at the cell surface, and the difficulties of recapitulating cellular phenotypes with simple recombinant assays. The results also suggest some limitations inherent in the TLR4 ELISA itself, which I discuss below.

4.2.3 Limitations of TLR4 ELISA

The TLR4-binding ELISA will turn out in the latter sections of this chapter to be the only non-cross-linking assay that demonstrated a TLR4:FReP PPI, so it is worth exploring in some detail the underlying limitations with the approach, especially

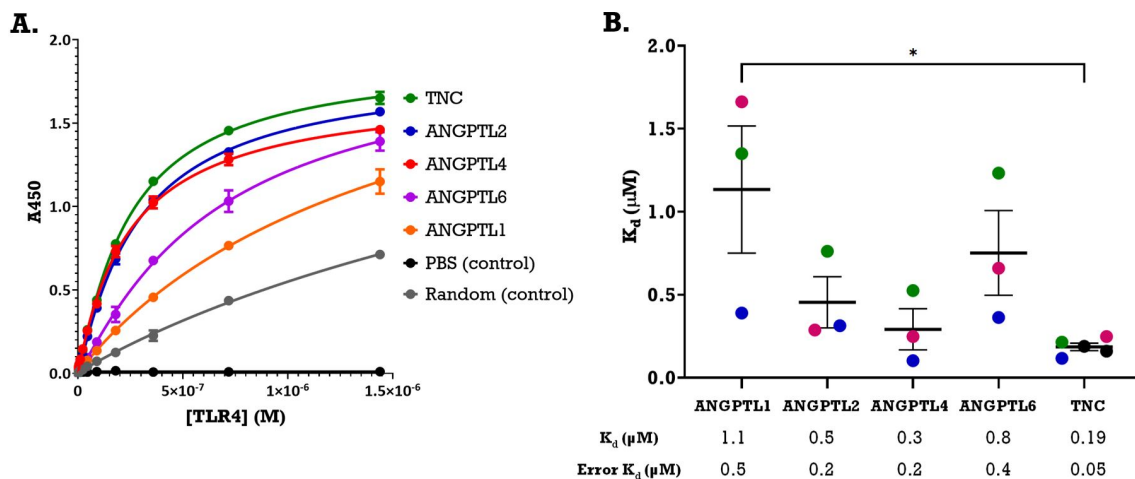


Figure 4.2 FRePs in TLR4 Binding ELISA. **A.** Representative TLR4-binding ELISA with diverse FRePs. The “random” control uses an unrelated protein to coat the plate. Data fit in GraphPad using a variable slope saturation binding model (Hill Slope ≈ 1 in all cases). **B.** K_D's for the ANGPTLs (n = 3) and TNC (n = 5). Means were compared with a one-way ANOVA; only significant comparisons (p < .05) are annotated. Biological replicates are color coded.

given its ubiquity in the TLR4 literature. First, ELISAs are fundamentally denaturing, as coating the FReP onto the surface of the 96-well plate distorts the native FBG fold. This feature explains the reliability of ELISAs for antibody-antigen-like interactions, where the binding epitope is contiguous, small, and well defined; however, for more complex PPI's, denaturation of the bait protein on the plate is problematic. Previous reports suggest that spatially disparate regions of the FBG domain contribute to TLR4 binding (see Section 2.10), meaning denaturation during the ELISA could compromise the biological relevancy of the assay. Along the same lines, TLR4 is known to require oligomerization for signal transduction. Does this ELISA allow ternary complex formation when the FReP is denatured and displayed on the surface of the plate? Is a Hill Slope ≈ 1 , corresponding to a 1:1 interaction, a reasonable result for an interaction known to require oligomerization at the cell surface? These are unanswered questions; however, I assert that this ELISA is likely not an accurate representation of the true biologically relevant binding event.

Secondly, the ELISA is irreproducible in my hands despite incredibly careful experimental optimization; the assay is not precise enough across replicates to rank strong binding affinities, it cannot distinguish between activating and non-activating FRePs, and it is sensitive to the highest concentration of TLR4 used for the dilution series (which violates a core assumption of saturation binding equilibrium analysis).

The observed irreproducibility also appears between experimenters, with subtle experimental differences such as washing cycles, incubation times, reagent suppliers, and protein preparations vastly changing the final results (personal communication, Midwood Lab). Finally, the assay has high background binding, which is likely a result of the large size, complexity, and glycosylation envelope of TLR4 (both S2-derived and commercially available TLR4 are purified via secretion, resulting in significant glycosylation as discussed in Section 3.4). Coating the plate with a random protein gives a maximum $A_{450} > 0.5$ in my experiments, suggesting that recombinant TLR4 is prone to non-specific “sticking” in this ELISA and reducing the accuracy of the calculated affinities.

In summary, the TLR4 ELISA is irreproducible, incapable of distinguishing activating and non-activating FRePs, and likely not biologically relevant. My experience with this assay, and the cumulative weight of the additional data in this Chapter, led me to the conclusion that TLR4-associated ELISA results, including those reported elsewhere in the literature, should be carefully reproduced with orthogonal binding assays. This observation led me to explore more robust, solution-state methods to probe the TLR4:FReP interaction, which I discuss for the remainder of the Chapter.

4.3 Solution State Methods Show No Complex

4.3.1 Co-Secretion in S2

Because of my success with co-expressing TLR4 and MD-2 to isolate the TLR4:MD-2 complex (see Figure 3.3), I tested co-secretion of the FRePs with TLR4 from S2 cells. I completed a 10 mL transient co-transfection in S2 cells using TLR4-c071 (protocol in Section 3.1.6), containing TLR4^{Untagged} in the pExpreS²-2 backbone, along with multiple FRePs containing a C-terminal protein A tag in the pEpxreS²-1-CR-PA backbone. Figure 4.3A shows the PCR-amplification step from the construction of the pEpxreS²-1-CR-PA FReP expression vectors, which were subsequently cloned as before with a HiFi assembly and confirmed by Sanger sequencing. As a positive control, I used MD-2-c014, also containing a protein A tag (see Section 3.1.4). Figure 4.3B shows the eluted protein following a one-step affinity purification with IgG Sepharose directly from the cell culture supernatant sixty-eight hours after transfection. The gel demonstrates that *only* MD-2 can bind TLR4. Untagged TLR4-c071, despite being transfected in all samples, is pulled down only when co-expressed with

tagged MD-2-c014 but not with any of the tagged FRePs. All of the FRePs, excluding FGL2, are expressed in S2 cells as confirmed by SDS-PAGE and MSMS, but none of them form a complex with TLR4. Note that TNC expressed at a very low level and instead seemed prone to degradation, with the strongest band containing just protein A (Lane 2). This is a weakness of the experiment, since TNC was the strongest TLR4-binder in the ELISA. Nevertheless, my inability to form a multi-protein complex during co-expression provided the first piece of evidence that the TLR4:FReP complex is weak or nonexistent when using recombinant proteins secreted from S2 cells.

4.3.2 Analytical SEC

A potential weakness of the co-expression approach is the glycosylation pattern on the FRePs themselves; as the multiple banding pattern in the gel in Figure 4.3B demonstrates, the FRePs express in S2 cells with heterogeneous glycosylation states. Therefore, I next completed analytical SEC experiments using *E. coli* derived (fully glycan free) FRePs purified using the standard Scheme 2. Using a DionexTM 3000 HPLC with a Superdex 200 Increase 10/300 analytical grade column, I collected chromatograms for TLR4 both alone and after incubation with a variety of putative binding partners, based on the hypothesis that an upward shift in the TLR4 peak elution volume would indicate the formation of a complex. I injected 200 μ g of TLR4-c072, purified as described in Section 3.3, and 400 μ g of the binding partner, representing an $\approx$ 4-fold molar excess of the putative ligand. As a positive control, I again used MD-2-c014.

For each binding partner (MD-2, TNC, ANGPTL1, ANGPTL2, ANGPTL6, FGL1, and FGL2), I collected a chromatogram alone and after incubation with TLR4. Figure 4.3C shows a representative set of chromatograms from the analysis of TNC. I used the elution volume of “Peak 2” to align the different chromatograms, which were collected over a span of 24 hours, to correct for any drift in the column performance over time. I based this small correction on the assumption that the excess free FReP elutes at the same volume when analyzed alone or when mixed with TLR4. I then used the aligned chromatograms to analyze the elution volume and implied molecular weight of the TLR4-containing “Peak 1”. The implied molecular weight was calculated by comparison to a standard curve ($R^2 = 0.982$) generated from gel filtration standards run during the same experiment (standards displayed in grey in Figure 4.3C). The extracted elution volume (black diamond) and implied

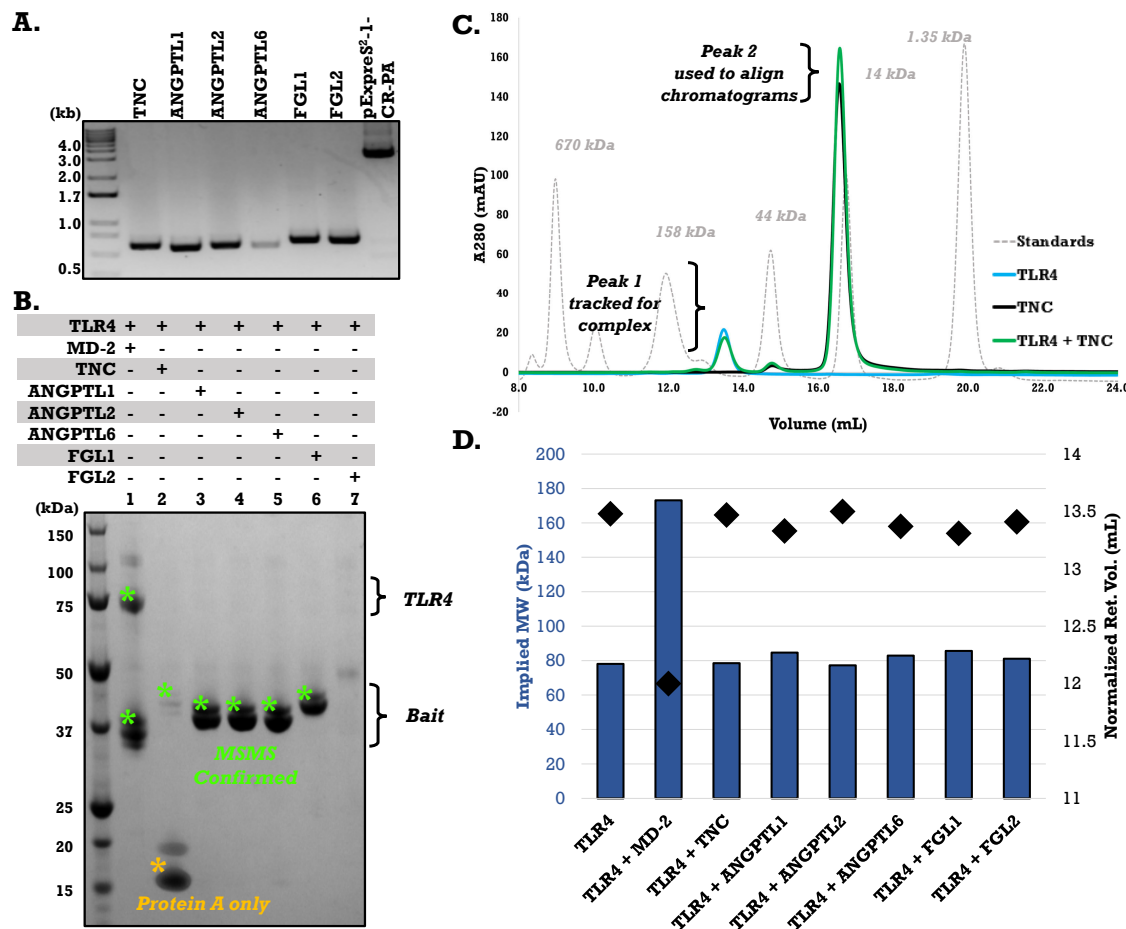


Figure 4.3 Co-Expression and Analytical SEC. **A.** PCR amplification step during construction of pExpres²-1-CR-PA FReP expression constructs. **B.** IgG eluates after a 10 mL transient co-secretion of TLR4-c071 and assorted FRePs from S2 cells. **C.** Representative analytical SEC trace (from TLR4 and TNC) generated via HPLC. The FReP-containing Peak 2 was used to align the chromatogram, while the TLR4-containing Peak 1 was tracked for signs of complexation (e.g. any up-shift of peak 1). **D.** Quantification of multiple analytical SEC runs, analyzed just as described for the TLR4:TNC sample in C. The implied molecular weight (blue bars) along with the normalized retention volume (black diamonds) of the TLR4-associated Peak 1 are shown.

molecular weight (blue bar) from the TLR4-containing Peak 1 are plotted for all chromatography runs in Figure 4.3D.

Consistent with the co-expression experiment, only MD-2 formed a stable complex with TLR4 during SEC, suggested by the decrease in retention volume (−1.48 mL) and the increase in implied molecular weight (121%) of Peak 1 in the presence of MD-2. This more than doubling in the implied molecular weight upon addition of MD-2, to approximately 175 kDa, is consistent with the formation of an active

heterotetramer. This broadly aligns with similar SEC experiments conducted with murine TLR4 and MD-2 in the presence of LPS [100]; although I did not add LPS directly, the HPLC system was used routinely by others and was not prepared to be endotoxin free prior to the experiment, which explains why I likely observed the active heterotetramer rather than just the TLR4:MD-2 heterodimer. The more important result is that no other ligand caused a comparable up-shift in the elution volume of Peak 1. The “largest” shift, observed with FGL1, was just 0.17 mL, corresponding to only a 7.1 kDa increase in the implied molecular weight of Peak 1. The difference between a positive and negative result becomes even more obvious when the “no-complex” TNC chromatogram in Figure 4.3C is compared to the positive control MD-2 chromatogram previously presented in Figure 3.4D. For complexes to be tractable for structural studies, it is virtually essential that the complex stays together during gel filtration to facilitate separation of the complex from individual components. Therefore, this result represented the end of my attempts to solve the structure of a TLR4:FReP complex, and I instead moved forward with other methods capable of interrogating transient interactions that do not survive SEC.

4.3.3 Native-PAGE

After failing to detect a stable complex with co-secretion or SEC, I next moved towards orthogonal biophysical techniques with the potential to measure the formation of *any* TLR4:FReP complex. Low-affinity transient complexes dissociate during chromatography, either SEC or affinity, due to a fast off-rate. This represents a potential limitation of my previous experiments which were focused on detecting only stable, structurally tractable complexes. Therefore, I next tried to detect a TLR4:FReP complex using Native-PAGE, with the rationale that the method is capable of separating bound and unbound species existing in an equilibrium.

As before, in the Native-PAGE experiments I mixed individually purified putative binders with TLR4 at a moderate (2- to 4-fold) molar excess of the binding partner. After a short room temperature incubation to establish equilibrium, I analyzed the samples via Native-PAGE to track the appearance of any higher molecular weight complexes. Because the samples were not subjected to chromatography before analysis, this experimental set-up should theoretically reveal the presence of low-affinity or low-abundance complexes.

In Figure 4.4A, I looked for complexation between TLR4 and either MD-2, ANGPTL6,

or TNC in the presence of a 2- or 4-fold molar excess of the binding partner and in the presence or absence of 25 mM CaCl_2 . Given the known Ca^{2+} -binding properties of the FRePs, but unknown underlying biology, I routinely checked to see if exogenous Ca^{2+} could stimulate the formation of a complex. Notice that TLR4 resolves as a clear single band (Lane 1), as expected, while MD-2 (Lane 2) resolves as a higher molecular weight streak. Both of these observations are consistent with the SEC data shown in Section 3.4, in that MD-2 exists not as a monomer but rather as a mixture of higher-molecular weight, heavily glycosylated species when present alone in solution. Neither ANGPTL6 nor TNC resolve well during Native-PAGE, perhaps due to their small size; TNC disappears completely (Lane 3), while ANGPTL6 appears as a streak (Lane 4). In any case, the key result from this experiment is unambiguous: the target TLR4 band only shifts upwards in the presence of MD-2 (Lanes 5 and 6). Neither doubling the amount of the FReP nor spiking in Ca^{2+} rescues the formation of a TLR4:FReP complex. As desired, Native-PAGE can separate “bound” and “unbound” species existing in an equilibrium; this is clear when comparing Lanes 5 and 6, which differ only in the amount of MD-2 added. Both the TLR4 and the larger TLR4:MD-2 bands are present in Lane 5, while only the complex band is present in Lane 6 when more MD-2 is added. This implies that should even a low abundance TLR4:FReP complex exist in a dynamic equilibrium, it would resolve as a higher molecular weight species during Native-PAGE *a la* the incompletely formed TLR4:MD-2 complex in Lane 5.

Next, I repeated a similar experiment, this time testing every TLR4-activating FReP in my small library (presented in Figure 4.4B). Based on the results in Figure 4.4A, I tested a 2-fold molar excess of the FRePs in the absence of Ca^{2+} . As before, I only observed a complex in the presence of MD-2 (Lane 6). Figure 4.4B demonstrates that neither TNC, ANGPTL1, ANGPTL2, ANGPTL6, FGL1, nor FGL2 complexes with TLR4 during Native-PAGE. As an added observation, FGL1 is the only FReP to resolve as two distinct bands during Native-PAGE. This is remarkably consistent with the two FGL1 oligomers observed in Figure 2.16 and provides additional evidence that FGL1 is unique amongst the FRePs tested in its ability to form a disulfide stabilized dimer when purified from CyDisCo cells. Nevertheless, the key takeaway again is that no TLR4:FReP complex appeared during Native-PAGE.

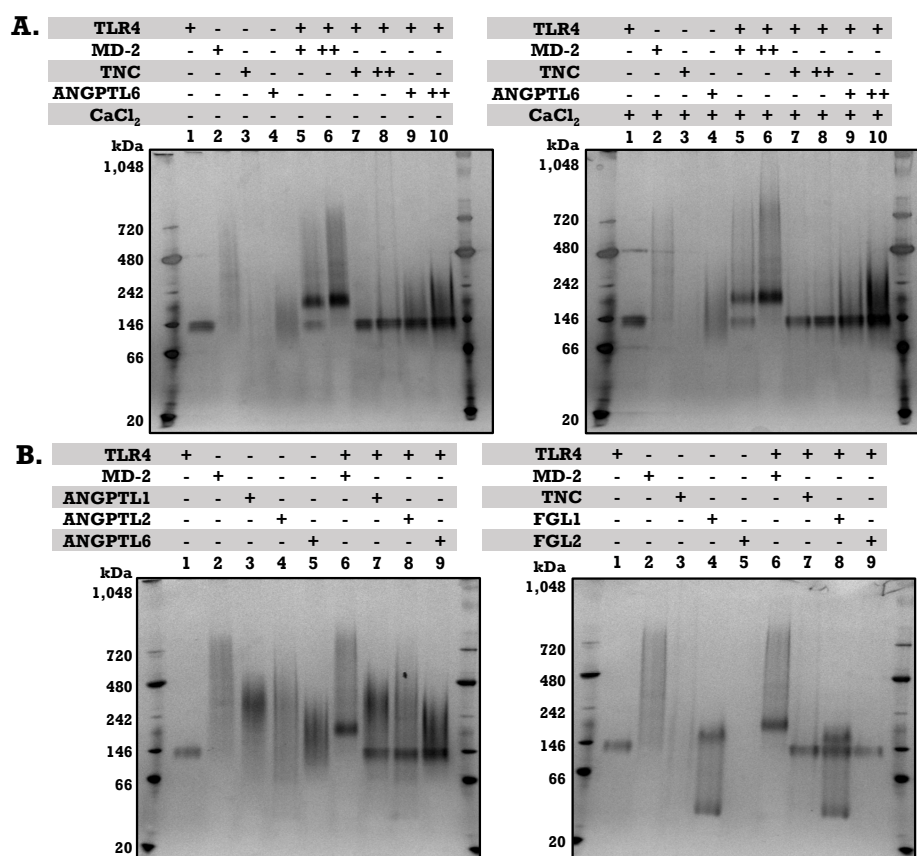


Figure 4.4 Native-PAGE. **A.** Native-PAGE analysis with either a 2- (+) or 4-fold (++) molar excess of putative binders mixed with TLR4, in either the presence or absence of 25 mM CaCl₂. **B.** Native-PAGE analysis with a 2-fold molar excess of the entire FReP family mixed with TLR4 in the absence of CaCl₂.

4.3.4 BLI

BLI is an orthogonal biosensor technology that provides a real-time, label-free method for detecting and quantifying key kinetic and affinity parameters (including k_{on} , k_{off} , and K_D) of biological interactions. In a BLI experiment on the Octet[®] system, a “ligand” molecule is immobilized to the surface of a biosensor chip, either via covalent coupling or affinity capture. The loaded chip is then sequentially dipped into solutions (prepared in either 96- or 384-well plates) containing different concentrations of “analyte” molecules. Analyte binding to the immobilized ligand causes a measurable and concentration-depending change in the optical parameters of the biosensor (technically, an increase in the wavelength of the interference pattern of reflected white light) which can be read-out in real-time. To account for

background binding of analytes to the chip, the measured response for the ligand-loaded biosensor (in nm) is always corrected by subtraction of the response from a ligand-free negative control biosensor before analysis. BLI is quite similar in principle to surface-plasmon resonance (SPR), wherein ligand binding induces changes in the optical parameters at the surface of a solid-phase chip; however, BLI instruments are more high-throughput and BLI biosensors are cheaper, making BLI a simple and reliable way to detect and quantify diverse binding interactions. BLI is reliable for moderate-affinity interactions (μM range and below) and performs especially well with PPIs because of the significant size change at the biosensor surface upon binding of large protein analytes. Detailed reviews and assay optimization results provided the foundation for the design and execution of the BLI experiments presented herein [431, 432, 254, 433].

Key variables in a BLI experiment include the choice of ligand and immobilization chemistry, the magnitude of ligand loading, the concentration of the analyte, and the washing conditions for removing bound analyte and regenerating the biosensor surface before a subsequent dip into a new analyte sample. Throughout the following experiments, I endeavoured to use best practices for BLI experiments, including (1) testing the assay in both “orientations” by switching the ligand and analyte (2) never saturating the surface of the biosensor and instead selecting a moderate-to-low immobilization level to avoid artefactual avidity-driven binding and (3) using gentle regeneration conditions like 1 M NaCl. Below, I briefly describe three different BLI experiments, with subtle but important experimental differences, that demonstrate yet again no strong, specific binding of the FRePs to TLR4. To paraphrase Tolstoy in *Anna Karenina*, all good biosensor data are alike; all bad biosensor data are unreliable in their own way. Therefore, I carefully discuss the unreliable aspects of the data which highlight both why the data do not support a true interaction and also where my experimental approach could be improved in the future.

In the first experiment, I immobilized the smaller protein (TNC) on the surface of the biosensor, a standard approach for maximizing the response upon binding of the largest molecule (in this case, TLR4). I immobilized C-terminally biotinylated full-length TNC (see Section 5.3 for purification details) onto the surface of streptavidin-coated SA chips to a response level of ≈ 5 nm. Please refer to Section 5.3 for details about the purification of biotinylated FRePs, which I also used for MSMS experiments detailed in Chapter 5. Figure 4.5A shows the background corrected response after dipping the TNC-loaded chips into a concentration series of TLR4, up to a maximum concentration of 6 μM . I observed no concentration-dependent

increase in the measured response, even at the highest concentration of TLR4. Note that BLI biosensors are sensitive enough to detect binding at concentrations below 10% of the actual K_D . Therefore, these results suggest that, if a direct interaction between TNC and TLR4 does exist, the affinity is weaker than 60 μM .

Next, I reversed the orientation of the assay, this time via covalent immobilization of TLR4 onto AR2G chips. In this approach, TLR4 is coupled directly to the biosensor surface via a covalent bond to free amines (either the N-terminus or Lys residues) using standard coupling chemistry (see Section 4.1.4). After optimizing the coupling $\text{pH} = 4.0$, I immobilized TLR4 to the level of $\approx 3.0 \text{ nm}$. Then, I dipped the TLR4-loaded chips into a concentration series of either TNC or MD-2, prepared at the highest concentration of 100 μM and 5 μM , respectively. In between each analyte concentration, I compared regeneration of the biosensor surface by dipping into either assay buffer or 1 M NaCl.

In all cases, I observed no specific binding of TLR4 to TNC or MD-2. Rather than showing the background-corrected traces as before, Figure 4.5B shows the raw reaction traces over the course of the experiment. These response curves suffer from severe non-specific binding of the analytes, both to the reference sensor and to the TLR4-loaded sensor. The large signal in the reference sensors demonstrates non-specific analyte binding to the empty AR2G chip surface; this is most pronounced for the analyte MD-2, where the reference sensor actually gives a *larger* response than the TLR4-loaded sensor. This is annotated as “Ref > analyte” in Figure 4.5B. Non-specific binding of the analyte to the TLR4-loaded surface is implied by the curvature of the TNC-association step. Note that, across every cycle in the concentration series, the association stage never reaches a plateau, but rather continuously increases. This curvature, corresponding to a lack of equilibrium, is a hallmark of non-specific binding and implies that the binding is an artefact; these kind of traces can often be mistakenly used to prove binding because the response still increases in a concentration-dependent manner [254]. The curvature problem is annotated as “No equil. during assoc.” in Figure 4.5B. Finally, I observed “background drift” over the course of the experiment, wherein the curves do not return to the previous baseline even after the dissociation and regeneration phases. This implies a non-specific deposition of the analyte onto the surface of the biosensor and further complicates the reliable extraction of binding parameters. I annotated this problem as “Baseline Drift” in Figure 4.5B.

Especially with heavily glycosylated proteins like TLR4 and MD-2, non-specific

binding artefacts are common and must be carefully tracked. Furthermore, in my hands the covalent capture of TLR4 onto the biosensor seems particularly prone to non-specific binding, highlighting the importance of testing biosensor-based assays in both orientations. The lack of confident binding to MD-2 in Figure 4.5B, however, is troubling, given that the high affinity TLR4:MD-2 interaction routinely served as a positive control in my experiments. I have two hypotheses for this result: (1) covalent immobilization at low $\text{pH} = 4.0$ somehow compromised the structural integrity of TLR4, preventing binding to MD-2 or (2) covalent capture of contaminating LPS (which contains a free amine and is a high-affinity ligand of MD-2) on the reference sensor swamps the TLR4-specific binding signal. Either of these explanations could rationalize the relatively larger response in the reference sensor channels. The lack of a true positive control, no matter the reason, is a weakness in my data. Follow-up experiments could focus on systematic removal of trace LPS from the buffers and BLI system and/or the introduction of an orthogonal high-affinity TLR4-binder (perhaps an antibody) as the positive control.

For my final BLI experiment, I returned to affinity immobilization of biotinylated ligands, namely all of the biotinylated FRePs I could purify, followed by dipping into $6\text{ }\mu\text{M}$ TLR4. The goal of this experiment was to determine if any FReP showed any binding to TLR4, up to an implied affinity of $60\text{ }\mu\text{M}$. I immobilized each FReP at a concentration of 1 and $5\text{ }\mu\text{g mL}^{-1}$ to a density of ≈ 2 or 8 nm , respectively. I dipped each sensor twice into $6\text{ }\mu\text{M}$ TLR4, with a regeneration step in-between using 1 M NaCl as before. Figure 4.5C shows the background-subtracted responses for all FRePs at both ligand densities. In all cases, I observed no specific binding of TLR4 analyte to the FRePs. Most association phases have no curvature and very low response, while the association phases for ANGPTL2 and ANGPTL6 actually start curving downwards, two common pathologies of non-specific binders in biosensor experiments. Again, because of the lack of a biotinylated MD-2 construct that could express in *E. coli* (see Section 5.3), I had no reliable positive control in this experiment. My data could be improved by checking each FReP against a known binder and/or by developing a method for the biotinylation of recombinant MD-2 from S2 cells.

In conclusion, across three very different BLI experiments, I never observed FReP binding to TLR4, up to an implied K_D of $60\text{ }\mu\text{M}$. Many of the collected response curves, like Tolstoy’s unhappy families, contained diverse manifestations of artefacts and non-specific binding. Nevertheless, the data allowed me to hypothesize that if a direct TLR4:FReP PPI does exist, it is quite transient and not amenable to

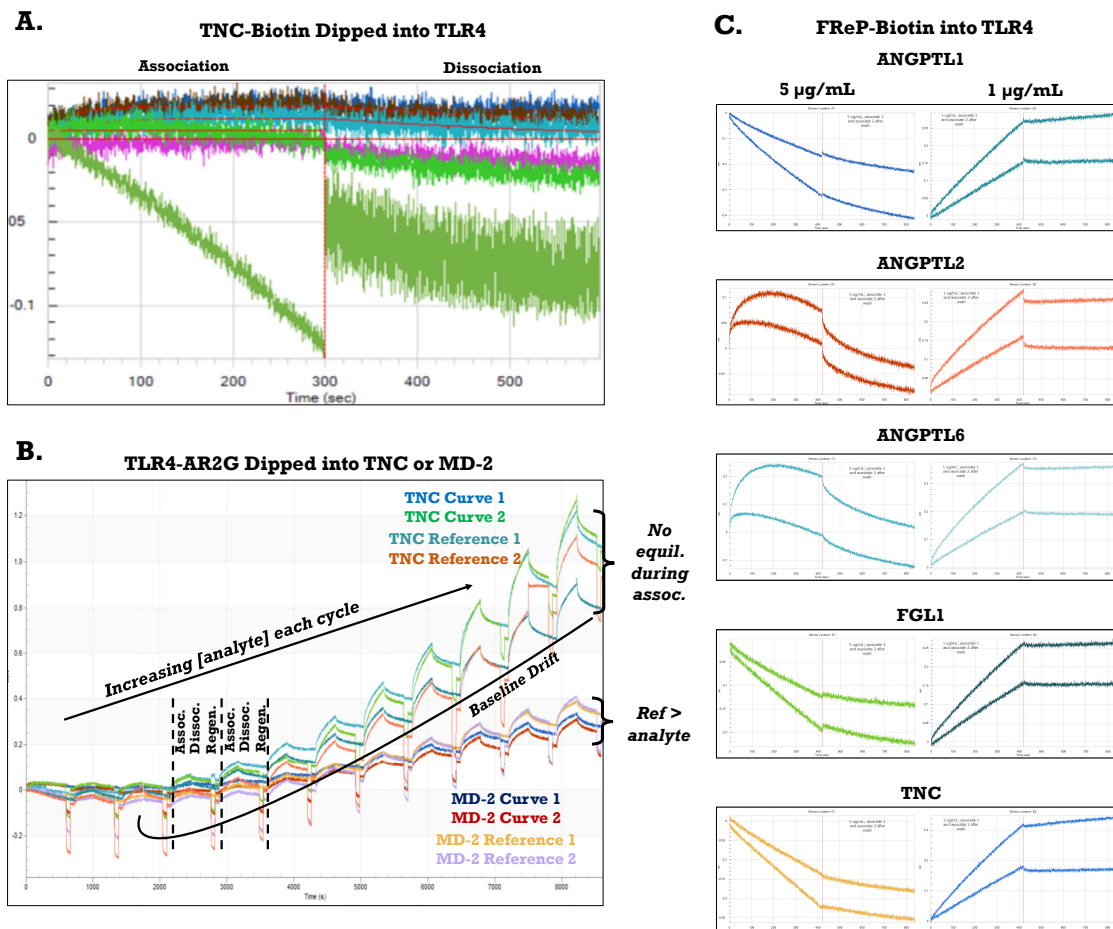


Figure 4.5 BLI. **A.** Background subtracted response from TNC loaded SA chips dipped into a TLR4 analyte concentration series. **B.** Raw response traces from covalently TLR4 loaded AR2G chips dipped sequentially (following the marked association, dissociation, regeneration cycle) into either TNC or MD-2 analyte concentration series. The pathologies in the data, discussed in detail in the main text, are annotated for clarity. **C.** All background-corrected response traces generated from FReP ligands immobilized at two different densities (1 and $5\text{ }\mu\text{g mL}^{-1}$) onto SA chips and subsequently dipped twice into $6\text{ }\mu\text{M}$ TLR4 analyte. Each trace corresponds to the response over time during a single association-dissociation cycle. In all panels, the x-axis represents time (in sec) while the y-axis represents response units (in nm).

biophysical characterization using BLI. However, I do not deny that there is room for improvement in these experiments, particularly with regards to the generation and characterization of well behaved positive controls. Absence of evidence is not evidence of absence, particularly in complex biosensor experiments, so the BLI data should only be considered alongside the totality of the negative data I present in Chapter 4.

4.4 XL-Methods Rescue Transient Complex

Based on all of the data presented above, I hypothesized that the TLR4:FReP complex, if it exists at all, is of low affinity and therefore transient. Therefore, in order to detect *any* complex, I completed a set of cross-linking (XL) experiments. XL experiments are ideal for transient interactions because addition of a covalent cross-linker prevents dissociation of short-lived complexes. The cross-linker enriches the proportion of the sample that exists as a complex during the experiment, as any transient interaction is trapped before falling apart. XL experiments do have some drawbacks, including the enrichment of non-specific (or, at least, non-native) complexes, and the difficulty of translating the results into a proper binding affinity. Nevertheless, the XL-SDS-PAGE experiments presented below hint at the presence of a low-affinity, transient complex between TLR4 and many of the TLR4-activating FRePs, especially TNC and ANGPTL6. As suggested by other techniques, the complexes are weak and always represent a small minority population.

Cross-linked proteins remain together during SDS-PAGE, allowing simple read-out of cross-linked species on a gel. In the experiments which follow, I employed multiple cross-linkers, including the “short” BS³ (bis(sulfosuccinimidyl)suberate; 11.4 Å hydrocarbon spacer between the warheads) and the “long” BS(PEG)₉ (PEGylated bis(sulfosuccinimidyl)suberate; 35.8 Å spacer). Both are homobifunctional, with primary amine-reactive N-hydroxysuccinimide (NHS) warheads on both sides of a hydrocarbon linker; this means BS³ and BS(PEG)₉ react preferentially with the N-terminus and Lys residues. For many FRePs, notably TNC, the known TLR4-activating epitope(s) contain lysine (see Section 2.7), making amine-reactive cross-linkers sensible. Additionally, I also tested LC-SDA (succinimidyl 6-(4,4'-azipentanamido)hexanoate; 12.5 Å spacer), a heterobifunctional cross-linker containing an amine-reactive NHS warhead on one side and a non-specific UV-activatable warhead on the other. LC-SDA is a useful orthogonal cross-linker because it reacts first to amines under ambient light, followed by the opposite side reacting non-specifically only after activation with strong UV light. This allows a step-wise cross-linking, which can be beneficial if intramolecular cross-links dominate over intermolecular cross-links during the experiment.

In all cases, the reactions were incubated for 15-30 min at room temperature before cross-linking for 30 min, quenching with excess free amine (usually Tris pH = 8.0), and visualizing with SDS-PAGE. The LC-SDA reactions were irradiated with 8 W, 254 nm UV-light for 15 min before quenching and analysis. For a detailed protocol,

see Section 4.1.5. Because these gels are visually quite busy, I have only annotated bands of interest, particularly focusing on the putative TLR4:FR_{EP} bands which were confirmed by MSMS.

Figure 4.6A shows a preliminary experiment with TLR4, MD-2, TLR4:MD-2, and TNC cross-linked with BS(PEG)₉ or LC-SDA. Firstly, regardless of whether TLR4 and MD-2 are co-purified (as in Lane 1 and 14) or separately purified and then mixed (as in Lane 6 and 19), a strong two-way TLR4:MD-2 cross-linked complex band appears in the gel. As a reminder, the mass shift is larger when the components are separately purified only because the protein A tag is retained when MD-2 is

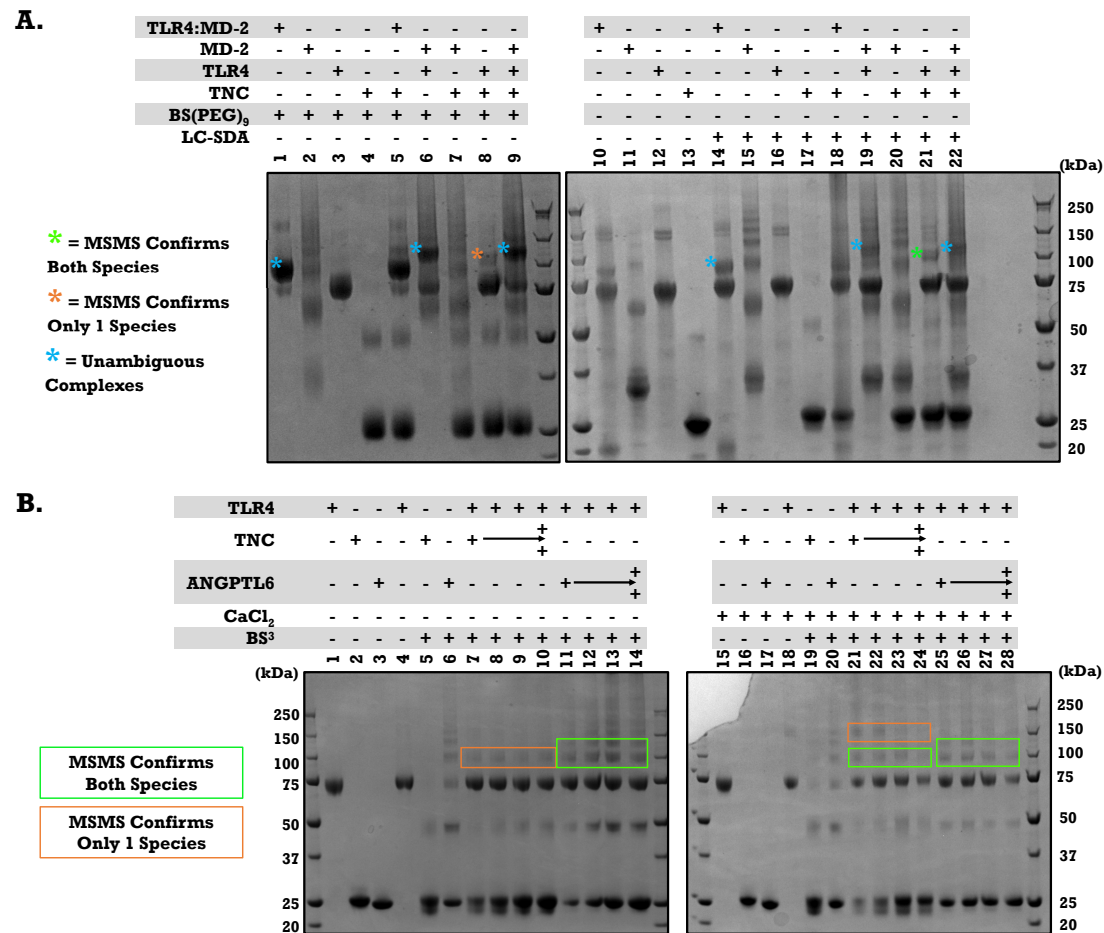


Figure 4.6 XL-SDS-PAGE 1. **A.** XL-SDS-PAGE with TLR4, MD-2, and TNC; “TLR4:MD-2” samples contain the co-purified dimeric complex. Equal mass of each protein added and 1 mM cross-linker used throughout. Note especially the strong TLR4:MD-2 complexes (Lane 1, 6, 9, 14, 19) and the faint TLR4:TNC complexes (Lane 8, 21) **B.** Titration of TNC and ANGPTL6 (from 1.25 µg to 7 µg, cross-linked with BS³ in the presence of 5 µg TLR4 with and without 60 mM CaCl₂. Note especially the appearance of TLR4:TNC and TLR4:ANGPTL6 complexes (boxed).

purified alone, whereas in the co-purified complex the tag is removed (see Section 3.4 for details). In both cases the shift in molecular weight is consistent with a 1:1 heterodimer, and further demonstrates the reliability of the TLR4:MD-2 complex as a positive control for these XL-SDS-PAGE experiments. Secondly, for the first time outside of the ELISA, I resolved a TLR4:TNC complex in the presence of both cross-linkers (Lane 8 and Lane 21). However, the TLR4:TNC complex remains a minority species as judged by band intensity, especially relative to the strong TLR4:MD-2 complex. Thirdly, note the presence of multiple bands when MD-2 is cross-linked by itself (Lane 2 and 15). This aligns with my previous observations regarding multiple oligomers when MD-2 is expressed alone. Finally, note the similarities between Lane 6 and Lane 9 as well as Lane 19 and 22; this is the first indication that a TLR4:MD-2:TNC complex does not exist, an observation I will explore in much more detail in Section 4.5.

Next, I focused on the TLR4:FReP interaction, this time using increasing concentrations of TNC and ANGPTL6 in the presence of the shortest cross-linker BS³ and in the presence of Ca²⁺. Figure 4.6B shows the results of this experiment, which are broadly consistent. I observed a weak but MSMS confirmed TLR4:TNC and TLR4:ANGPTL6 complex, highlighted in green boxes. Interestingly, ANGPTL6 seems to form more complex than TNC, which contradicts the ELISA which identified TNC as the strongest TLR4 binder (see Section 4.2). Upon close inspection, the TLR4:TNC complex might be stronger in the presence of Ca²⁺, and some higher-MW species do appear (Lanes 21-24 relative to Lanes 7-10). Conversely, the TLR4:ANGPTL6 complex seems slightly weaker in the presence of Ca²⁺ (Lanes 25-28 relative to Lanes 11-14). In full, the results show that Ca²⁺ cannot rescue a strong TLR4:FReP complex.

Finally, I extended the analysis to include all of the other FRePs cross-linked with TLR4 in the presence of either BS³ or BS(PEG)₉. The results are presented in Figure 4.7; I excluded TNC from this experiment since I repeatedly demonstrated the weak TLR4:TNC complex in other experiments (as in Figure 4.6). In this experiment, I detected an MSMS-confirmed TLR4:MD-2 complex (Lanes 17, 18, 33, 34), TLR4:ANGPTL2 complex (Lane 36), and TLR4:ANGPTL6 complex (Lane 37). I also observed weaker putative complexes containing TLR4:ANGPTL1 (Lane 19 and 35), although I did not detect both proteins during MSMS. FGL1, FGL2, and FCN1 *perhaps* showed a low-abundance complex with TLR4; however, I did not detect either protein with MSMS, consistent with the weak intensity of the complex band. The behavior of the FRePs cross-linked with themselves also provided some

the observation that recombinant TLR4:FReP complexes, as presently purified, are highly transient and intractable for structural studies.

4.5 “Obvious” Co-receptors: No Complex

After multiple strategies did not provide any evidence to support a strong TLR4:FReP complex in solution with recombinant components, I proceeded with an extensive literature review to identify additional targets, reported to bind to TLR4, which might act as a co-receptor to facilitate the TLR4:FReP interaction. I purified or purchased a variety of putative co-receptors and tested for the formation of a complex with TLR4 and/or TNC. Throughout this section, I focus exclusively on TNC, the exemplar FReP family member and the highest affinity binder based on the solid-phase ELISA and the monocyte activation data discussed previously. All experiments herein should be considered preliminary pilots; I only tested each target in one or two assays, reasoning that a true high-affinity interactor would appear unambiguously (much like the TLR4:MD-2 complex). After learning my lesson from the TLR4:TNC experiments, I started with XL-SDS-PAGE, and only moved to more native experiments if the cross-linking approach supported the formation of any complex. This section presents data across four different novel co-receptor hypotheses, with no evidence to suggest that any of these proteins can rescue a TLR4:FReP interaction.

4.5.1 MD-2

MD-2 is the *bona fide* co-receptor for LPS and forms a high-affinity complex with TLR4, making it a front-runner as the co-receptor for other TLR4 ligands like the FRePs. I first tested this hypothesis using XL-SDS-PAGE with recombinant TLR4, TNC and MD-2. Returning to Figure 4.6A, Lane 8 and Lane 20 show no complex between TNC and MD-2; with the addition of TLR4 in Lane 9 or 22, the TLR4:MD-2 complex appears as expected, but no larger complex appears. This result is unchanged even when starting with the pre-purified TLR4:MD-2 complex and then adding TNC, as in Lane 5 and Lane 18. A biological replicate of this experiment is shown in Figure 4.8A. Yet again, the TLR4:MD-2 complex forms in Lane 9, but no higher molecular weight species appears upon the addition of TNC in Lane 14.

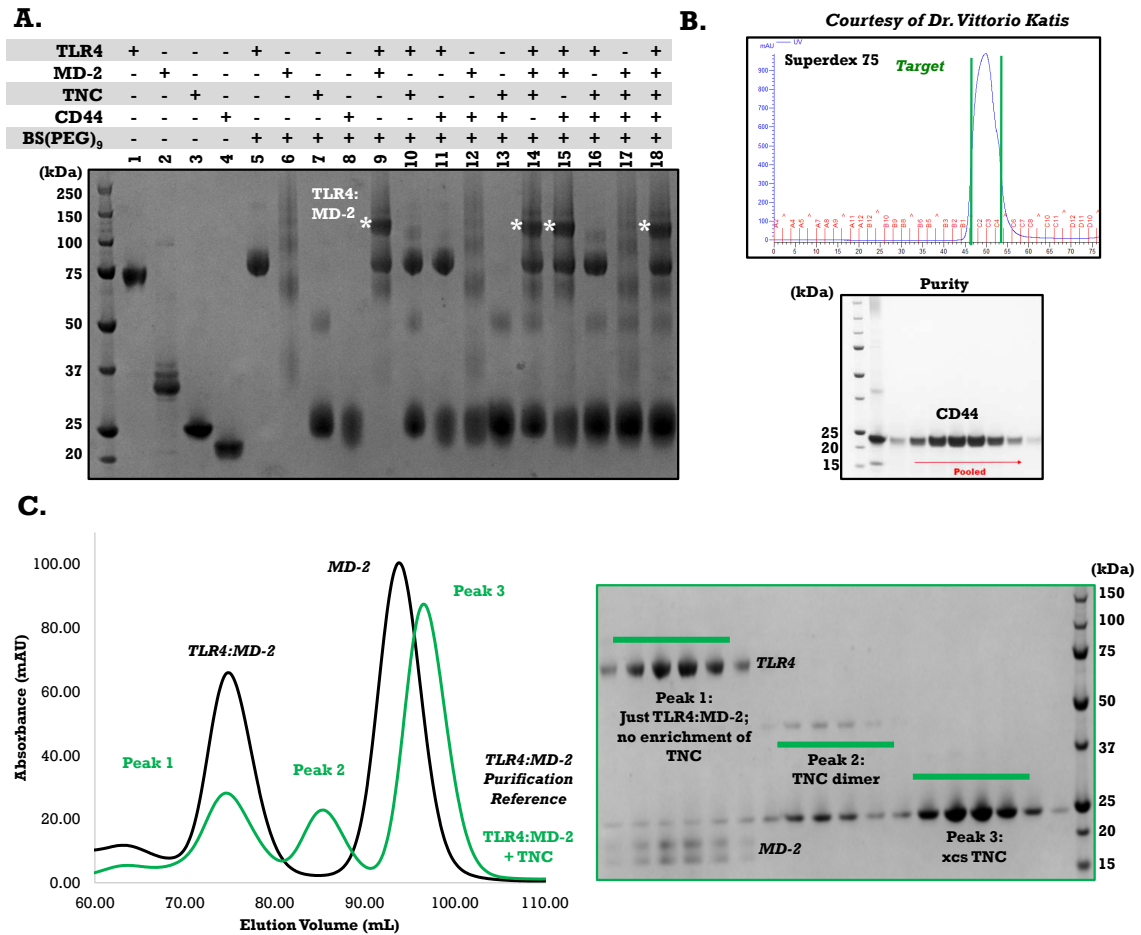


Figure 4.8 Putative Co-receptors 1. **A.** TLR4, MD-2, TNC, and the CD44 Link domain mixed and cross-linked with BS(PEG)₉. **B.** Purification of CD44 Link domain from *E. coli* inclusion bodies by SEC, courtesy of Dr. Vittorio Katis. Final gel shows the high purity of CD44-Link. **C.** SEC runs, generated from purification of the TLR4:MD-2 complex (black) and after the TLR4:MD-2 complex was mixed with excess TNC (green). Accompanying SDS-PAGE corresponds to the TLR4:MD-2 + TNC green trace. No enrichment of TNC in Peak 1 and no shift in the Peak 1 elution volume points to the absence of a TLR4:MD-2:TNC complex.

I also checked for the formation of a tertiary TLR4:MD-2:TNC complex using SEC, the results of which are shown in Figure 4.8C from a preparatory grade Superdex 200 16/600 column. Addition of excess TNC (green trace) to the purified, pre-formed TLR4:MD-2 dimer (black trace, Peak 1) does not lead to a change in the Peak 1 elution volume, and TNC is not enriched in Peak 1 as visualized by SDS-PAGE. Therefore, my biophysical data support the cellular data showing that MD-2 is dispensable for FReP activation of TLR4 and is not the sought after co-receptor [142].

4.5.2 CD44

CD44 is the co-receptor for another TLR4 ligand, hyaluronic acid (HA), and is known to physically interact with TLR4 and MD-2 at the cell surface to drive pro-inflammatory cytokine release [182]. Furthermore, Poluzzi *et al.* recently showed that CD44 is the co-receptor for biglycan, a small leucine-rich proteoglycan and known TLR4-activating DAMP [160]. To test the hypothesis that CD44 might act analogously as the co-receptor for TNC, Dr. Vittorio Katis purified the HA-binding Link domain of CD44 (UniProtID: P16070; Ala20–Asn172) to high purity via refolding of inclusion bodies from *E. coli*, based on a previously reported protocol [434]. Dr. Katis also independently demonstrated that the refolded CD44-Link could bind directly to HA in a biochemical assay with the same affinity as eukaryotic-cell-derived CD44-Link (data not shown). Dr. Katis generously shared some of this protein with me, and the purification shown in Figure 4.8B was completed by him in full and reproduced with permission.

Figure 4.8A shows the results from an XL-SDS-PAGE experiment with CD44, TLR4, MD-2, and TNC. No binary TNC:CD44 (Lane 13), tertiary TLR4:CD44:TNC (Lane 16), or quaternary TLR4:MD-2:CD44:TNC (Lane 18) complexes were detected. Throughout, only the control TLR4:MD-2 complex appeared (Lane 9, 14, 15, and 18) no matter the additional proteins present. Interestingly, I detected no formation of a TLR4:CD44 (Lane 11) or TLR4:CD44:MD-2 (Lane 15) complex, consistent with the notion that HA-binding is required to form this signalling complex. An obvious limitation of this experiment is the use of only the HA-binding Link domain of CD44; at the time, full-length CD44 ectodomain (purified as an exemplar application of the FAS2FURIOUS system, see Figure 3.8) had yet to be produced. At this stage, my data only allows me to hypothesize that the CD44-Link domain is not sufficient to act as a TNC co-receptor. Before CD44 is completely ruled out as a FReP co-receptor, this experiment should be repeated in the future with the full-length CD44 ectodomain.

4.5.3 ANXA2 and LYZ

In 2019, Yadav and Surolia published a fascinating study showing that neuronal lysozyme (LYZ) is a DAMP which activates TLR4 signalling during nerve injury, leading to neuroinflammation via NF- κ B and increased pain sensitivity [207]. Using anti-LYZ IP-MSMS experiments, the authors show that the co-receptor ANXA2

is absolutely required for LYZ-mediated TLR4 activation, and they further show a direct and high affinity interaction between ANXA2 and LYZ using SPR. TLR4 activation in the presence of LYZ is independent of MD-2, suggesting the possibility of a novel LYZ:ANXA2:TLR4 signalling axis. This work builds upon previous reports of cross-talk between TLR4 and ANXA2. Zhang *et al.* showed that an interaction between TLR4 and ANXA2 in endosomes negatively regulates inflammatory responses to LPS; in fact, using speculative structural modelling, the authors hypothesized the existence of a quaternary complex of TLR4:ANXA2:TRAM:TRIF [209]. Furthermore, Swisher *et al.* showed that the heterotetramer between ANXA2 and S100A10 activates TLR4-dependent pro-inflammatory signalling in both human and murine monocytes and macrophages [208]. Taken together, the literature suggests that ANXA2 can act as both a direct TLR4 DAMP and an indirect TLR4 co-receptor, making it an intriguing candidate as the FReP co-receptor. Therefore, I biophysically explored the hypothesis that ANXA2 and/or LYZ can act as a co-receptor for TNC.

ANXA2 is a known Ca^{2+} binding protein with a short, disordered regulatory N-terminal domain (Met1–Phe33) subject to phosphorylation at Ser26, tethered to the core C-terminal domain (Asp34–Asp339) made up of four helical “annexin repeats” [435]. Upon Ca^{2+} binding, ANXA2 binds directly to EF-hand containing S100 family members, especially S100A4 and S100A10, and can associate directly with the plasma membrane via binding to phospholipids. The Ca^{2+} binding properties of ANXA2 are especially interesting given my results showing FReP binding to Ca^{2+} . While human LYZ was cheaply available commercially, ANXA2 required purification in-house. In 2017, Ecsedi *et al.* reported multiple crystal structures of ANXA2, including a complex with S100A4, and a detailed protocol for the purification of ANXA2 from *E. coli* [430]. I reached out to Paul Ecsedi, who generously shared a full-length, un-tagged ANXA2 expression construct (UniProtID: P07355). Since the disordered N-terminus is required for S100 effector binding, I decided to focus on this construct for complex studies, rather than ΔNT constructs which have also been reported. Using a cation exchange column directly from the clarified lysate followed by an SEC clean-up step, I purified ANXA2 to high purity with reasonable yields (approximately 1 mg L^{-1}). The purity of the protein is highlighted in Figure 4.9A and the purification method is reported in Section 4.1.8.

To begin, I first tested ANXA2 and LYZ in the XL-SDS-PAGE assay alongside TLR4 and TNC in both the presence and absence of Ca^{2+} . For clarity of the results in Figure 4.9B, MSMS-confirmed complexes of interest are highlighted with colored

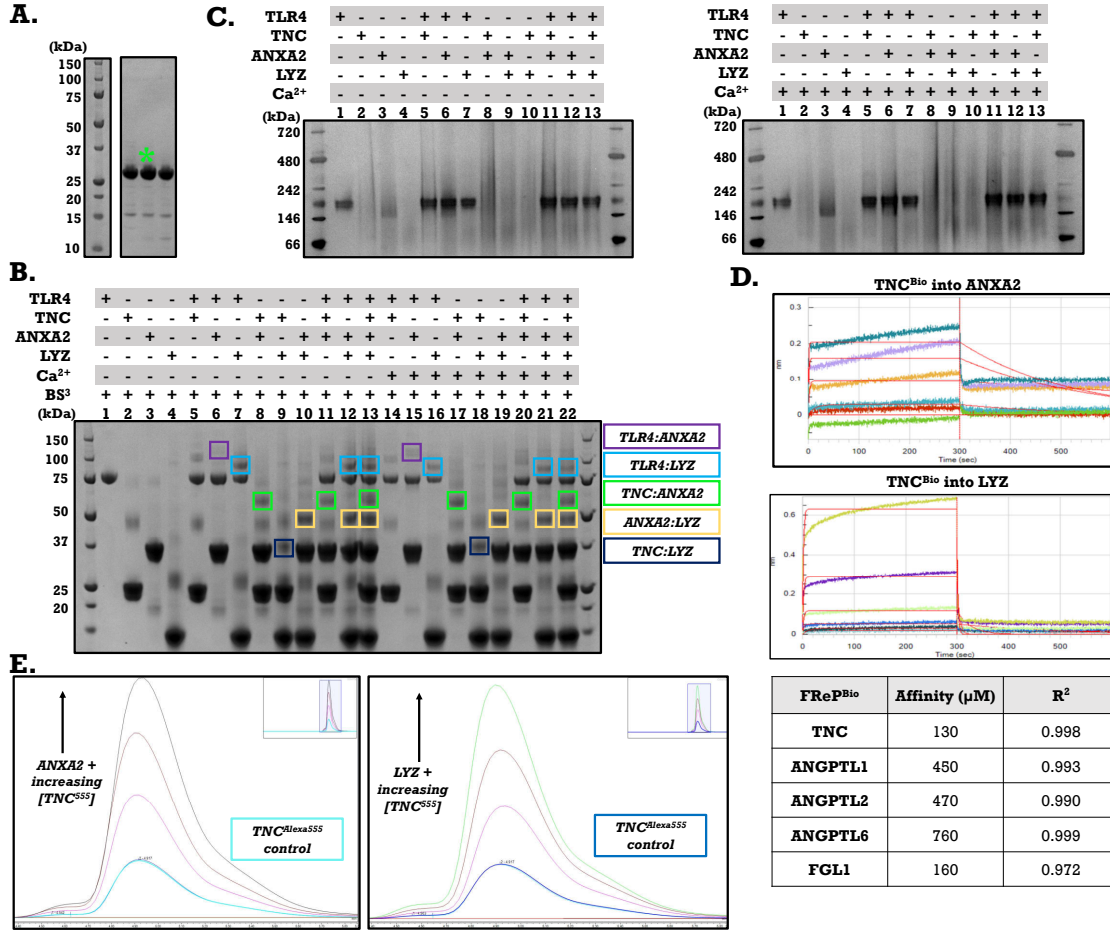


Figure 4.9 Putative Co-receptors 2. **A.** Purity of full-length ANXA2 confirmed by MSMS (green asterisk) after SEC. **B.** XL-SDS-PAGE experiment using BS³. Complex bands positively identified with MSMS are boxed and colored according to the constituent proteins. **C.** Native-PAGE showing no multiprotein complexes at a higher molecular weight than the TLR4 control. **D.** BLI experiments with biotinylated TNC dipped into a concentration series of ANXA2 (top) or LYZ (middle). The x-axis is time (sec) while the y-axis denotes response units (in nm). Lines of different colors correspond to different concentrations of the analyte. Table at bottom shows the dissociation constant (K_D) and fit to a one-to-one binding model between LYZ and all FReP ligands tested in a replicate experiment. **E.** FSEC experiment tracking increasing concentrations of Alexa555-TNC in the presence of either ANXA2 (left) or LYZ (right). No change in the elution volume (relative to the control trace, colored in either cyan or blue) or appearance of earlier peak implies the absence of two-way complexes.

boxes. As a control, the weak TLR4:TNC complex, discussed in detail in Section 4.4, appears in Lane 5 and 14. Furthermore, the high affinity complex between ANXA2 and LYZ, reported by Yadav and Surolia, appears throughout (yellow box), most notably in Lane 10 (without Ca²⁺) and 19 (with Ca²⁺). Additional aspects of Yadav

and Surolia's cellular work are supported, including a weak complex between TLR4 and ANXA2 (purple box; Lane 6 and 15) and between TLR4 and LYZ (light blue box, especially Lane 7 and 16). Of particular interest to my project, I observed novel complexes between TNC and LYZ (navy blue box, Lane 9 and 18) and TNC and ANXA2 (green box, especially Lane 8 and 17). Neither of these interactions depend on Ca^{2+} . However, I did not observe any three-way complexes, either those suggested by the literature (TLR4:ANXA2:LYZ) or hypothesized by myself (TLR4:LYZ:TNC or TLR4:ANXA2:TNC); all TLR4-containing complexes have just one additional protein. Lane 22 suggests that LYZ is the preferred TLR4 binder in the experiment; when all four proteins were mixed, the TLR4:LYZ band predominated. This suggests that LYZ is a tighter TLR4 binder than TNC in this experiment, although follow-up analytical SEC experiments did not support a stable TLR4:LYZ interaction (data not shown). In conclusion, this preliminary XL-SDS-PAGE experiment validated a smattering of previously reported complexes and offered evidence that TNC can directly bind to the hypothesized co-receptors ANXA2 and LYZ, but that neither of these proteins can rescue a TLR4-containing signalling complex.

So, I moved forward to more native experiments. First, I validated the lack of a TLR4-containing multiprotein complex with Native-PAGE. The results in Figure 4.9C are strikingly consistent with the XL-results above. In either the presence or absence of Ca^{2+} , none of the tested proteins in any combination form a larger complex with TLR4. Unfortunately, because of their small size and positive charge, ANXA2 and LYZ (much like TNC and the other FRePs) do not resolve well on Native-PAGE, meaning the results in Figure 4.9C are inconclusive regarding two-way complexes involving TNC. Therefore, I next moved to a fluorescent analytical SEC experiment (FSEC), which relied on tracking fluorescent TNC (recombinant TNC covalently labelled with the fluorescent small molecule Alexa555) in the presence of ANXA2 or LYZ. The FSEC approach, with detection at 575 nm, simplifies interpretation of the chromatograms, since *only* the labelled protein is measured (rather than all proteins in the sample giving an absorbance reading at 280 nm). In Figure 4.9E, I added an increasing amount of Alexa555-labelled TNC to a fixed amount of either ANXA2 (left) or LYZ (right), and looked for the appearance of a second, larger peak upstream containing Alexa555-TNC that would correspond to a two-way complex. I observed no larger complexes in the presence of ANXA2 or LYZ.

Taken together, the experiments suggest a weak interaction between TNC:ANXA2 and TNC:LYZ that can be resolved only with a cross-linker. Therefore, in an at-

tempt to estimate the affinity of these complexes, I proceeded with BLI, looking for binding of ANXA2 or LYZ to biotinylated TNC immobilized on an SA chip. In preliminary experiments, I observed only nonspecific binding with analyte ANXA2 (Figure 4.9D, top). However, in the presence of high concentrations of LYZ (up to 100 μM), I observed relatively specific binding, with a top response of 0.6 nm and a clear equilibrium curvature in the response (Figure 4.9D, middle). Therefore, I repeated the experiment with a larger concentration series of LYZ (up to 150 μM), this time using all biotinylated FRePs as immobilized ligands. The derived affinities and fit to a standard one-to-one binding model are shown in the bottom panel of Figure 4.7D. All FRePs tested seemed to bind to LYZ, but with incredibly weak affinity (average K_D of 0.4 mM). These affinities should also be taken as approximations only, as for all ligands except TNC the calculated affinity exceeds the highest concentration of analyte tested. TNC bound most tightly to LYZ amongst the FReP family; however, at an affinity of 130 μM , this interaction classifies as transient.

In the BLI experiments presented in Section 4.3.4, I demonstrated no discernible binding between TNC and TLR4 up to an implied affinity of 60 μM . In the current section, due to the cheap availability of highly concentrated LYZ, I estimated a TNC:LYZ binding affinity of approximately 130 μM . Comparing the XL-SDS-PAGE results in Figure 4.6A with the results in Figure 4.9B, I noted the faint intensity of the TLR4:TNC complex band relative to the much more intense TNC:LYZ complex band. This comparison suggests that the putative TLR4:TNC interaction is *highly* transient, with an affinity much weaker than even the previously speculated 60 μM , since the cross-linked band is fainter than a related complex with a known affinity of 130 μM .

To conclude, the data presented herein, while interesting, do not support the hypothesis that ANXA2 and/or LYZ act as a co-receptor to present TNC to TLR4, as no three-way complex was detected under the tested experimental conditions. While a transient and, to my knowledge, novel interaction between TNC:ANXA2 and especially between TNC:LYZ is supported by the data, these two-component complexes do not translate into a larger complex with TLR4 that would be indicative of a true co-receptor. Therefore, I filed ANXA2 and LYZ away as interesting potential TNC interactors with completely unknown biology but outside the scope of the project.

4.5.4 S100 Family Proteins

A full discussion of S100A8 and S100A9 can be found in Section 1.2.2. As a reminder, S100A8 (*aka* MRP8) and S100A9 (*aka* MRP14) are metal binding, dimeric, multifunctional, antibiotic “alarmin” proteins known to be secreted by monocytes and macrophages in response to infection and at sites of active inflammation [436]. In addition to their direct antibacterial activity, S100A8 and S100A9 can act as DAMPs to activate inflammation via the TLR4:MD-2 complex [123]. In fact, S100A8 is the most prevalent protein in the sera from patients with RA and is known to stimulate pro-inflammatory IL6 release [252, 248]. S100A8 and S100A9 naturally form a heterodimer, known as Calprotectin (CP), which itself can dimerize into a heterotetramer in the presence of divalent metals like Ca^{2+} and Zn^{2+} [241]. Ca^{2+} -induced tetramerization *prevents* CP from activating TLR4 and therefore represents a negative feedback loop to prevent chronic inflammation as a result of CP release [125]. Recent data, generated using a pull-down assay with different peptides derived from S100A9, hints at a direct interaction between S100A9 and the TLR4:MD-2 complex, implying a possible TLR4:MD-2:S100A9 complex [125]. Finally, in preliminary pull-down experiments (see Chapter 5), I detected an interaction between S100A8 and TNC. Based on these literature data, I decided to use my *in vitro* biophysical assays to test two hypotheses. First, I sought to validate results from the aforementioned cellular assays and show that either S100A8, S100A9, or CP could bind to the TLR4:MD-2 complex. Second, I tested to see if either S100A8, S100A9, or CP could rescue a TLR4:TNC interaction.

Purification protocols for both S100A8, S100A9, and the heterodimeric CP are well known [240], and former SGC scientist Dr. Ritika Sethi had already cloned His6-tagged constructs of full-length S100A8 and S100A9. As a result, I was able to quickly express and purify S100A8, S100A9, and CP to high purity (Figure 4.10A), and intact mass analysis confirmed the identity of the proteins and the 1:1 stoichiometry of the CP heterodimer (Figure 4.10B). Using these proteins, I tested for the formation of complexes involving TLR4 and/or TNC in multiple assays, and results from a representative XL-SDS-PAGE experiment are shown in Figure 4.10C. As a quality control check, the expected S100 dimers (either A8:A8, A9:A9, or A8:A9), boxed in yellow, and the Ca^{2+} -dependent tetramerization of CP, boxed in blue, were observed. I also observed the weak but MSMS-confirmed presence of S100A8:TNC and S100A9:TNC complexes, highlighted in green. Based on the intensity of the bands, the S100A8:TNC interaction (Lanes 15 and 37) is stronger, which is consistent with my preliminary pull-down results suggesting a direct S100A8:TNC

interaction. However, quite clearly, in either the presence or absence of 2 mM Ca^{2+} , I did not detect *any* complexes involving any S100 protein and either TLR4 or TLR4:MD-2. In my hands, neither S100A8, S100A9, nor CP could form a complex with TLR4, TLR4:MD-2, or TLR4:MD-2:TNC; as before, the only higher MW complex observed in the experiment is the positive control TLR4:MD-2 dimer, boxed in purple.

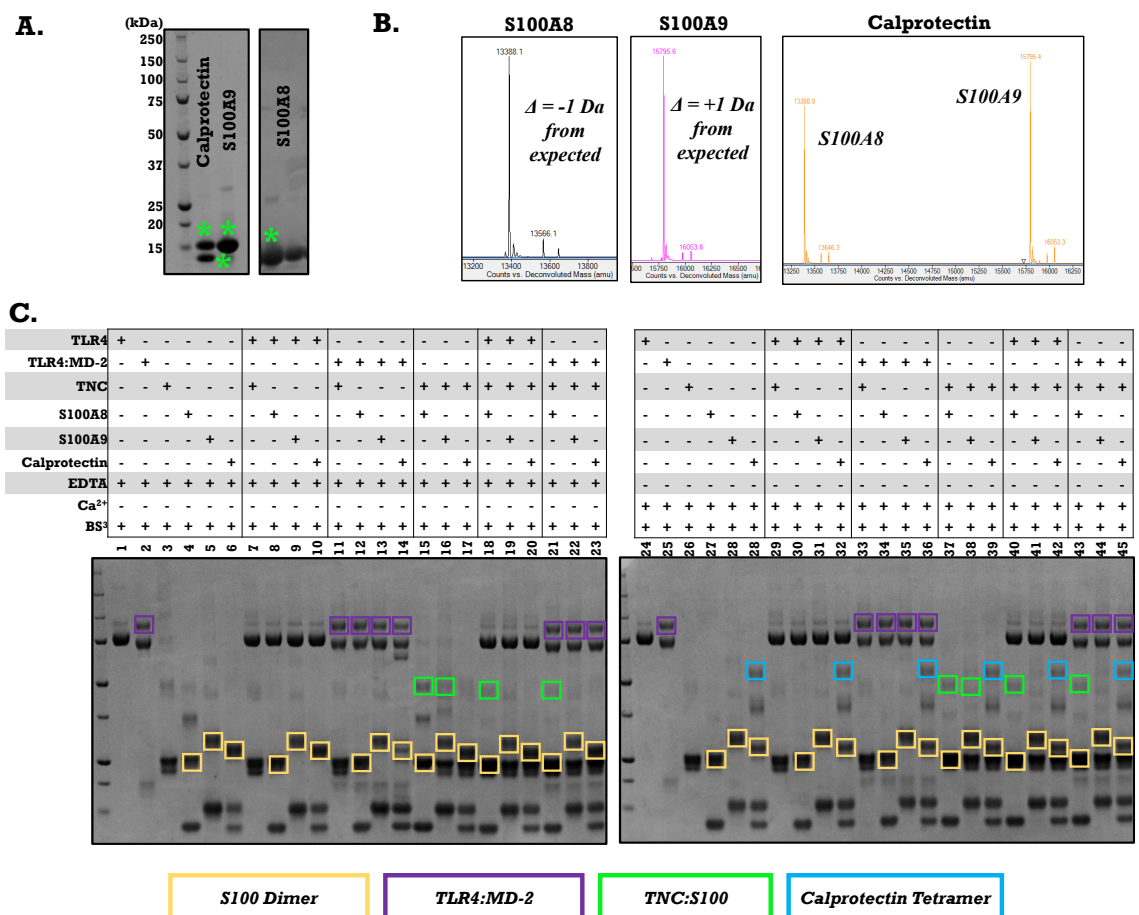


Figure 4.10 Putative Co-receptors 3. **A.** Final purity of S100A8, S100A9, and CP. Bands marked with a green asterisk were confirmed by MSMS. **B.** Intact mass analysis confirms the protein identity and the 1:1 stoichiometry of the CP heterodimer. **C.** XL-SDS-PAGE analysis involving different mixtures of TLR4, MD-2, S100A8, S100A9, CP, and TNC in the presence of either EDTA or Ca^{2+} and cross-linked with BS³. MSMS confirmed bands of interest are highlighted with colored boxes.

Therefore, I could not validate the literature results which suggest a direct interaction between S100 family proteins and TLR4, and I did not find any evidence to suggest that these S100 family members can act as a co-receptor for TNC. Especially with regards to S100A9, an interaction with TLR4 is well established in the litera-

ture, and Bjork *et al.* report SPR data showing direct binding between TLR4:MD-2 and S100A9 [124]. However, as I discussed in Section 1.2.2 based on the warnings in Rich and Myszkowski's review about biosensor data [254], I find the overall shape (no equilibrium) and height (over 1,200 RU) of the SPR sensorgrams reported by Bjork *et al.* to be indicative of non-specific binding. Bjork *et al.* report a K_D of 3 nM between S100A9 analyte and TLR4:MD-2 ligand; if the complex were that stable, I would expect to see it during the cross-linking experiments (especially Lanes 14 and 35 in Figure 4.10). Therefore, I remain unconvinced about a direct S100A9:TLR4 complex. As will become a theme in Chapter 5, I instead hypothesize that I did not observe a TLR4:S100A9 interaction in these experiments because other adaptor proteins are required to stabilize the complex. As a second reminder from Section 1.2.2, He *et al.* reported that CD14 is required as a co-receptor for S100A9, and they showed direct binding of S100A9 to CD14 in SPR [126]. He *et al.*'s sensorgrams are both qualitatively (clear curvature in the response) and quantitatively (maximum response of just 25 RU) more consistent with specific binding between CD14:S100A9 than Bjork *et al.*'s data showing S100A9:TLR4:MD-2 binding. Therefore, yet again, the biophysical data for the S100 family (both generated herein and in the literature) does not achieve the standard of evidence required to prove the existence of a direct complex with TLR4. As a final observation, while the putative and novel TNC:S100A8 interaction is intriguing, especially given the shared Ca^{2+} -binding activity of both proteins, this observation is outside the scope of my project and I did not pursue it further.

4.6 Chapter 4 Summary

In this Chapter, I presented comprehensive biophysical characterization of the putative TLR4:TNC complex which demonstrated that the complex is highly transient or non-existent, depending on the experimental conditions and method of detection. Neither analytical SEC, Native-PAGE, XL-SDS-PAGE, or BLI facilitated the detection of the high-affinity complex I hypothesized to exist at the outset based on previous cellular and ELISA data. These results prevented efforts towards further structural work on the TLR4:FReP interaction, since at least a moderate affinity interaction is required for crystallography or cryo-EM. Furthermore, these data represented a turning point in my project, as I subsequently abandoned the notion of working with a TLR4:TNC complex and instead turned my attention towards explaining *why* the complex does not seem to exist with recombinant components.

As a first effort towards answering this question, I showed preliminary evidence that “obvious” co-receptors derived from the literature, including MD-2, CD44, ANXA2, LYZ, S100A8, and S100A9, cannot rescue a TLR4:TNC interaction.

However, the experiments presented herein do offer a glimpse at other TNC-associated complexes worth exploring in the future. XL-SDS-PAGE and BLI experiments demonstrated a direct and novel interaction between TNC:ANXA2, TNC:LYZ, and TNC:S100A8, suggesting that this small domain could be acting as a true scaffolding hub in the extracellular milieu. Additionally, although I never demonstrated it experimentally in complexation assays, I expect that the Ca^{2+} binding properties of TNC are biologically relevant. Are these novel interactions present *in vivo*, and do they contribute to disease? Does metal-ion binding modulate the TNC interactome? Are other TNC-involved complexes structurally tractable? All of these questions merit further investigation. I also would like to draw attention to the seemingly tighter complex between TLR4:LYZ detected in the XL-SDS-PAGE studies, even though I could not validate this result in a single follow-up FSEC experiment. With more time, I would have focused additional energy on exploring this interaction, since out of all the DAMPs I tested, it was LYZ (not any of the FRePs) which gave the clearest complex band with TLR4 during XL-SDS-PAGE.

A key weakness of my data is the non-native glycosylation pattern on TLR4 purified from S2 cells, which I also discussed in Section 3.4. Knockout of certain N-linked glycans via single point Asn to Ala mutations demonstrated that some of the N-linked glycan sites on TLR4 and MD-2 are important for the function of the receptor [437]. As shown in Figure 3.4, the TLR4 I purified was heavily glycosylated, but I could only remove the glycans in the presence a denaturant like SDS—the soluble, recombinant TLR4 used for all experiments in this Chapter 4 still bear the glycans added by the S2 cells. That being said, S2 cells do confer eukaryotic, mostly paucimannosidic glycosylation, and the TLR4 construct I used contained the native N-linked glycan sites (i.e. the glycans on my TLR4 are different in character from the mammalian glycans but not fully removed like in an Asn to Ala mutant protein). Finally, as demonstrated in Figure 4.1, the S2-derived TLR4 matched the performance of mammalian-derived TLR4 in an ELISA which, despite the limitations of the assay, does suggest that the glycans do not abolish activity. That being said, my results should be validated with TLR4 secreted and purified from human cells to ensure the S2 glycosylation is not responsible for the lack of a TLR4:FReP complex. Towards this end, I present experiments in Chapter 5 using primary human monocytes which support my contention that TLR4 does not directly bind to the FRePs.

To conclude, the totality of the data in this Chapter suggest that truly novel biology may be at play here. As discussed in Chapter 1, TNC activates inflammatory signalling in a TLR4-dependent manner; however, I could never detect a direct interaction between TLR4:TNC using purified proteins. In fact, I *never* reproducibly detected the formation of a high-affinity complex between TLR4 and *any* DAMP outside of the problematic ELISA. These results form the basis of my thinking around the TLR4 receptorome, which I experimentally probe in Chapter 5 using unbiased proteomics. I expect these results to challenge the existing model of a 1:1 interaction between TLR4 and DAMPs and provide the impetus for important experiments probing the true nature of innate immune activation in response to endogenous ligands.

Chapter 5

Exploration of the FReP-Associated Receptorome Using XL-AP-MSMS

This chapter details my efforts to explain, using unbiased proteomics, why the FRePs do not directly engage TLR4 to form a high-affinity complex. Based on the biophysical evidence presented in Chapter 4 and the literature reviewed in Chapter 1, my new hypothesis, which I coined the “receptorome”, is that the FRePs do not directly activate TLR4 without context-dependent assistance from co-receptors, adaptors, or other scaffolding proteins. I probe this hypothesis using a new and well validated cross-linking cell-surface pull down proteomics assay (XL-AP-MSMS) directly from primary human monocytes. Results from these experiments demonstrate that integrins, rather than TLR4, are the dominant feature of the FReP-associated innate immune interactome; in fact, even on the highest resolution mass spectrometer, I again find no evidence of an interaction between TLR4 and the FRePs. I then present other highlights from the XL-AP-MSMS data, including identification of potential cell-surface accessory proteins that interact with TNC, and conclude by discussing the future implications of these data for the FReP receptorome specifically and on our understanding of innate immune biology more generally.

5.1 Materials and Methods

In this Chapter, I report three optimization pilot experiments (1, 2, and 3) and two full experiments (1 and 2) using the XL-AP-MSMS workflow. More details will be

provided in the main text, but for clarity before the methods are introduced I will highlight some key differences between experiments that emerged as I worked up the method. In Pilot 1, I used G-MCSF-differentiated macrophages for the pull-down, but for all subsequent experiments I used the easier-to-isolate monocytes. Starting with Pilot 2, I switched to using monocytes isolated via plating rather than via magnetic cell sorting (MACS) to achieve maximal cell yields. All of the TNC used in Experiment 1 and Experiment 2 was ultra-pure, LPS-free, and TEV-cleaved to remove the His6-tag (see Figure 5.2 for purification details); the other bait proteins used in the pilots and Experiment 1 contained LPS-contaminated, His6-tagged proteins. The use of LPS-contaminated proteins could obscure the TLR4-binding properties of the baits, but because of time limitations, difficulty with concentrating some of the non-TNC bait proteins, and in-line with the goal of the pilots as method optimization, I did use LPS-contaminated protein in this Chapter. I am explicit about these differences throughout the Chapter, but I want to emphasize here that the TNC bait protein used in the large Experiments 1 and 2 was of the highest possible grade.

In the pilots and Experiment 1, I use the cell-impermeable, disulfide cleavable cross-linker 3,3'-dithiobis(sulfosuccinimidylpropionate) (DTSSP). In Experiment 2, I also employed the cell-permeable, disulfide cleavable cross-linker DTBP (dimethyl dithiobispropionimidate). Both of these cross-linkers are homobifunctional and create a covalent bond between adjacent primary amines on either Lys residues or the N-terminus; I clearly state which cross-linker was used throughout this Chapter. Finally, because of restrictions on human blood handling in my building and then cross-building travel restrictions during COVID-19, *all* of the cell culture work (isolation and plating of monocytes as well as LPS testing and ELISAs) was completed on my behalf by Anna Marzeda. Prior to the lockdown, I shadowed all of the isolation protocols at least once to ensure I understood the method, but I never completed these steps by myself. Therefore, I am grateful to Anna for her help with obtaining primary human monocytes, without which I could not have finished these experiments.

5.1.1 Purification of Biotinylated FRePs

Full-length FBG domains (accession numbers in Section 2.1.1) and the truncated TNC-c042 construct were cloned by Dong Wang into pNIC-Bio3 using LIC-cloning as previously reported (see Section 2.1.3), which places the FRePs downstream of

a NT-His6-TEV tag and upstream of an uncleavable CT-AviTagTM. These constructs were co-transformed alongside a Strep^R pCDF-LIC plasmid containing the un-tagged biotin ligase BirA into CyDisCo pMJS226 competent *E. coli* cells. These triple-component cell lines were selected and grown on plates and in medium containing kanamycin, chloramphenicol, and spectinomycin to ensure that the FReP, CyDisCo, and BirA proteins, respectively, were present. The proteins were expressed as previously reported (Section 2.1.6), except that a small scoop (approximately 1 g L⁻¹) of D-biotin (FluoroChem) was added to the cultures at the time of IPTG induction. Medium-scale test purifications from 50 mL cultures were performed essentially as described in Section 2.1.5, except that 4 mM D-biotin was added to the lysate to push the equilibrium towards 100% biotinylation. For large-scale purification of biotinylated, full-length TNC, Scheme 2 was employed exactly as described in Section 2.1.7 except for the addition of D-biotin upon induction and in the lysis buffer. The protein identity, degree of biotinylation (always greater than 95%), and disulfide-bond status were confirmed with intact mass spectrometry.

5.1.2 Monocyte Stimulation Assay

Primary peripheral blood mononuclear cells (PBMCs) were isolated from a leukocyte cone (NHSBT at the John Radcliffe Hospital) using ficoll-gradient centrifugation (Histopaque-1077; Sigma-Aldrich) according to the manufacturer's recommendations. Please refer to Section 5.1.4 for more details on this step. Monocytes were then enriched using negative selection with MACS using the EasySepTM Human Monocyte Enrichment Kit without CD16 Depletion (STEMCELL) as recommended and subsequently seeded into 96-well TC-treated plates (100,000 cells per well) before resting overnight. The monocytes were stimulated with endotoxin-free TNC at 1 μ M or LPS (*E. coli* O55:B5, Sigma-Aldrich, cat no. L4524, lot 057M4027) at 100 ng mL⁻¹ for 24 hours. Prior to stimulation, the TNC samples were pre-incubated with polymixin-B (PMB) for 30 min on ice to exclude the impact of residual endotoxin contamination. Vehicle controls containing matched protein sample buffer with PMB were also included. Following stimulation, the culture medium was collected and used to measure cytokine release by ELISA.

5.1.3 Endotoxin Removal from Biotinylated FRePs

Before using the FRePs for XL-AP-MSMS experiments, they were treated with PierceTM High Capacity Endotoxin Removal Resin (PERR) according to the manufacturer’s recommendations. In short, 500 μ L of PERR was cleaned with 0.2 M NaOH overnight at 25 °C with gentle end-over-end rotation. The PERR was then washed with LPS-free 2 M NaCl before equilibration in Base Buffer (same for all subsequent experiments; 50 mM HEPES + 250 mM NaCl prepared with strictly LPS-free reagents and water). The concentrated FReP sample, ranging between 250 and 500 μ L in volume, was incubated with the resin for 1 hour at 4 °C with gentle mixing, and the resin was removed with centrifugation for 1 min at 500 x *g*. The concentration of the protein sample was checked, and in all cases the recovery was greater than 95% from the PERR. Endotoxin levels in the treated protein were checked with Recombinant Factor C Assay Kit (Lonza) by Anna Marzeda at 1 μ M; exact LPS levels are given below. In all cases, treatment with PERR reduced the LPS levels by over 74%; for the TNC protein used in Experiment 1 and Experiment 2, the levels reached the threshold considered to be LPS-free for *in vitro* assays (below 10 pg mL^{-1}). A key step to getting endotoxin-free protein was concentrating the TNC to 10 mg mL^{-1} , which reduces the volume that must be added to the monocytes and simplifies reaching the endotoxin-free threshold. The final LPS levels recorded for each protein when polished with PERR were as follows:

- **TNC:** 5.1 pg mL^{-1}
- **ANGPTL2:** 1347 pg mL^{-1}
- **ANGPTL6:** 2125 pg mL^{-1}
- **FGL1:** 3176 pg mL^{-1}

5.1.4 Isolation of Primary Human Monocytes via Plating

This plating method was the core method used to prepare monocytes for the majority of the experiments presented herein, including Pilot 2 and 3 and Experiment 1 and 2, in an effort to maximize cell yields. For Pilot 1 and the monocyte stimulation assay described in Section 5.1.2, MACS-isolated monocytes were used. Primary PBMCs were isolated from a leukocyte cone (NHSBT at the John Radcliffe Hospital) using ficoll-gradient centrifugation (Histopaque-1077; Sigma-Aldrich) according

to the manufacturer’s recommendations. In short, the 8 mL leukocyte cone was diluted to 25 mL with PBS and then layered on-top of 20 mL of Histopaque-1077 prior to centrifugation at $600 \times g$ for 30 min. The opaque middle layer, containing PBMCs, was collected with a Pasteur pipette and washed twice with 50 mL PBS via centrifugation for 10 min at $300 \times g$. A final wash with 50 mL PBS followed by centrifugation for 10 min at $200 \times g$ was used to remove contaminating platelets. To remove contaminating red blood cells, the pellet was resuspended in 2 mL RBC Lysis Buffer (0.15 M NH_4Cl , 10 mM KHCO_3 , 0.1 mM EDTA, pH = 7.3) and incubated for 5 min at 25 °C. 10 mL of PBS was added, and then the cells were washed three total times with PBS via centrifugation for 5 min at $400 \times g$. After the last wash, the purified PBMCs were resuspended in complete medium (CM; RPMI1640 + 3% FBS + 1X Pen/Strep) and 100×10^6 PBMCs were plated onto 10 cm TC-treated petri dishes. Following a two-hour incubation at 37 °C to allow the monocytes to adhere, non-adherent cells were removed with stringent washing. 10 mL of fresh CM, without FBS, was added to each plate and the now-enriched monocytes were rested overnight at 37 °C, 5% CO_2 prior to use.

5.1.5 XL-AP-MSMS Pilot 1

For Pilot 1, PBMCs were isolated using ficoll-gradient centrifugation exactly as described in Section 5.1.4. For Pilot 1 only, monocytes were enriched from the bulk PBMCs with negative selection MACS using the EasySep™ Human Monocyte Enrichment Kit without CD16 Depletion (STEMCELL). 20×10^6 MACS-purified monocytes were seeded at a density of 1×10^6 cells per mL into TC-treated bottles containing RPMI1640 with 5% FBS and 1X Pen/Strep. For Pilot 1 only, $100 \text{ ng } \mu\text{L}^{-1}$ M-CSF (Peprotech) was added and the monocytes were allowed to differentiate into macrophages for five days. After removing the dead cells and spent medium, the differentiated macrophages were harvested with scraping, gently centrifuged, and finally resuspended in CM. Two T75 flasks were seeded with 4×10^6 macrophages and rested overnight before continuing with the pull-down.

The next morning, $1.8 \mu\text{M}$ LPS-contaminated, His6-tagged, full-length TNC purified from the 50 mL test expression was added to one flask, and the corresponding amount of buffer was added to the other as a negative control. After a 30 min pre-incubation, 2 mM DTSSP (Thermo Fisher) was added to cross-link for an additional 30 min before quenching of the excess DTSSP with 20 mM Tris pH = 8.0. The cross-linked macrophages were harvested, centrifuged for 5 min at $500 \times g$, and resuspended

in 100 μ L of Mammalian Protein Extraction Reagent (M-PERTM; Thermo Fisher). After a 10 min incubation for lysis, 25 μ L of NeutrAvidinTM agarose, pre-equilibrated in PBS, was added and the samples were incubated for 75 min at 4 °C with end-over-end mixing. The beads were then washed via centrifugation (1 min, 500 x *g* each time) three times with 400 μ L PBS and three times with 400 μ L of PBS + 0.75 M NaCl prior to elution of the interactors with 2 x 25 μ L of PBS + 50 mM DTT. Eluates were concentrated via evaporation and the entire eluate was run on an SDS-PAGE gel stained with Coomassie. As discussed in the main text, no bands were visible; therefore, I do not recommend using this method.

5.1.6 XL-AP-MSMS General Method for Pilot 2 and 3

For Pilot 2, PERR-treated, full-length, biotinylated TNC was added to 200 x 10⁶ monocytes in a 15 cm TC-treated dish, isolated as described in Section 5.1.4, at a final concentration of 4.5 μ M. An identical number of monocytes were treated with Base Buffer as a negative control. Note that the RPMI medium was reduced to 5 mL, the minimum amount to cover the cells in the dish, before adding the FReP. After a 30 minute pre-incubation, the samples were cross-linked with 2 mM DTSSP for 30 minutes and the reaction was quenched with the addition of 10 mM Tris pH = 8.5. The monocytes were harvested, lysed with 5 mL of M-PERTM containing 1X benzonase and 1X Protease Inhibitors (Merck), and incubated with 50 μ L of NeutrAvidinTM agarose for two hours at 4 °C with end-over-end rotation. The resin was washed 3 times with 1 mL PBS, 3 times with 1 mL PBS + 0.75 mM NaCl, and eluted with 100 μ L of PBS + 50 mM DTT. For a final elution, 100 μ L of 6 M guanidine hydrochloride was added to completely strip everything from the agarose. The eluates were analyzed with SDS-PAGE followed by silver staining.

For Pilot 3, endotoxin contaminated, truncated, biotinylated TNC (matching the boundaries of TNC-c042) was added to five TC-treated petri dishes containing 100 x 10⁶ monocytes to a final concentration of 1.4 μ M and cross-linked as in Pilot 2 with either 0, 0.1, 0.5, 1, or 2 mM DTSSP. Both the cell lysate, prepared in 2.5 mL of M-PERTM + 1X benzonase + 1X protease inhibitors, and the cell culture medium were pulled down with 25 μ L NeutrAvidinTM agarose as in Pilot 2, except that the 0 mM sample was (mistakenly) not washed with high salt. The interactors were eluted with 100 μ L of 50 mM DTT in PBS, except for the 0 mM sample which was eluted with 50 mM DTT + 1 M NaCl. The eluates were concentrated to 20 μ L *in vacuo* and then analyzed with a silver-stained SDS-PAGE gel as described in Section 5.1.7.

5.1.7 Silver Staining SDS-PAGE Gels

SDS-PAGE on purified bait proteins and the eluates from the pilot experiments was completed with a 4–12% NuPAGE™ Bis-Tris gel (Invitrogen). For the purified FRePs, 2 µg protein was used as input; for the eluates, 20 µL was typically used. The gel was then fixed with an overnight incubation in 50% MeOH + 5% acetic acid followed by two 10 min washes with 50% MeOH and then water. The gel was then sensitized for 1.5 min at room temperature in 0.02% sodium thiosulfate followed by two washes in water. To stain, the gel was then incubated for 1 hour at 4 °C with ice cold 0.1% silver nitrate, followed by two washes with water. Finally, the gel was developed for 5–15 min with 0.04% formalin + 2% sodium carbonate at room temperature. Prior to imaging, the development was stopped with excess 50% acetic acid and the stained gel was stored in water.

5.1.8 General XL-AP-MSMS Method for Experiment 1 and 2

For the full Experiment 1, 10 TC-treated petri dishes containing 100×10^6 monocytes were treated with 2 µM of the designated FReP. Note that eight plates came from Donor 88 and two plates came from Donor 87. Four plates, two from Donor 88 and two from Donor 87, were used to analyze full-length, LPS-free, biotinylated TNC. Two plates were used to analyze PERR-treated (but not strictly LPS-free), biotinylated, His6-tagged ANGPTL6, two plates to analyze PERR-treated, biotinylated, His6-tagged ANGPTL2, and the final two plates to analyze PERR-treated, biotinylated, His6-tagged FGL1, all with protein purified from the medium-scale test expression. A matched negative control, with an equal amount of Base Buffer, was prepared for each pull-down. Note that, before adding the FRePs, the medium was reduced to 2.5 mL, just enough to coat the plate, in an effort to minimize the absolute mass of cross-linker that needed to be added. After a 30 min pre-incubation, 1 mM DTSSP was added to each plate, incubated for 30 minutes at 37 °C to cross-link, and then quenched with 10 mM Tris pH = 8.5. The medium, containing any dislodged monocytes, was centrifuged for 3 min at $500 \times g$ and lysed alongside the cells still attached to the plate using 2.5 mL of M-PER™ + 1X benzonase + 1X protease inhibitors. 50 µL NeutrAvidin™ agarose was added to each sample and incubated for two hours at 4 °C with gentle end-over-end rotation. Using sequential rounds of centrifugation for 2 min at $500 \times g$, the agarose was washed twice with 5 mL PBS, three times with PBS + 0.75 M NaCl + 1% Triton X-100, and then twice back into

just PBS. The washed, bound resin was then transferred to PierceTM Micro-Spin Columns (Thermo Fisher), incubated for 10 min with 100 μ L PBS + 50 mM DTT, and eluted directly into low-protein binding eppendorf tubes. The approximate protein concentration was estimated with a NanoDrop (usually between 17 and 20 μ g per sample) and then the eluates were stored at -20°C prior to preparation of the peptides for MSMS.

Experiment 2 was carried out essentially as reported for Experiment 1, except for the following changes. First, 12 plates with 70×10^6 monocytes were used, and the monocytes were pre-incubated over a time series for either 30 sec, 10 min, or 30 min after the addition of 2 μ M full-length, LPS-free, biotinylated TNC. Six plates (3 with TNC, 3 matched controls, one set for each time point) were cross-linked with 1 mM DTSSP and the other six plates were cross-linked with 3 mM DTBP for 30 min (regardless of the pre-incubation time prior to the addition of the cross-linker, the cross-linking reaction proceeded for a full 30 minutes). Pull-down on NeutrAvidinTM agarose was completed as before, except that the Triton X-100 was omitted from the wash buffer. For elution, PBS + 50 mM DTT was used for the DTSSP samples and PBS + 100 mM DTT was used for the DTBP samples. The resin was incubated with the elution buffer for 30 min at 50°C prior to collection of the eluted proteins.

To prepare the eluate samples for MSMS, the samples were first dried *in vacuo* and then re-suspended in 25 μ L of 100 mM NH_4CO_3 + 7.2 M Urea, which denatures the proteins. The eluates were incubated for 1 hr at 51°C to ensure the DTT in the elution buffer completely reduced all Cys residues, and then the disulfides were alkylated for 1 hour in the dark at 25°C with a final concentration of 20 mM IA. Note that the residual DTT from the elution buffer must first be reacted (via a 1:1 reaction with IA), so the amount of IA added was sufficient to quench the remaining DTT and give excess of about 20 mM to alkylate the Cys residues. For DTSSP cross-linked samples, this implied adding a nominal concentration of 170 mM freshly prepared IA; for DTBP cross-linked samples this implied adding a nominal concentration of 340 mM freshly prepared IA. After alkylation, the samples were diluted with 265 μ L of 50 mM NH_4CO_3 to reduce the [Urea] below 1 M, and the samples were digested overnight at 37°C with a 1:20 w:w ratio of sequencing grade trypsin protease (Sigma Aldrich). The next morning, the trypsin digest was quenched by the addition of 2.5% acetic acid, acidified to approximately 1% TFA by addition from a 5% TFA stock, and then the peptides were desalted into 100 μ L of 60% acetonitrile + 0.1% formic acid using C18 OMIX Stage Tips (Agilent). The peptides were completely dried *in vacuo* and shipped dry at room temperature to Dr. Holger Kramer for analysis.

5.1.9 MSMS Data Collection and Analysis

All LC/MSMS data collection was completed by Dr. Holger Kramer according to standard methods as previously reported [438, 439]. During Experiment 1, Dr. Kramer analyzed the samples in technical duplicate on a Q ExactiveTM Orbitrap Mass Spectrometer (Thermo Fisher). Following deconvolution and processing in MaxQuant [440], Dr. Kramer provided me with both protein hit lists filtered at a 1% false-discovery rate (FDR) as well as raw spectral counts (SPC). For separation of true and false interactors, I deployed the SAINT algorithm as implemented on the CRAPome web portal [441, 442]. Prior to SAINT analysis, I averaged the raw SPC from the technical duplicates for each pull-down, and I specify the SAINT summary probability (SP) score used during analysis throughout the text (always 0.9 or greater). For Experiment 2, Dr. Kramer analyzed the peptides on the more sensitive Q ExactiveTM HF-XI Mass Spectrometer (Thermo Fisher) as before, except with the addition of 5% DMSO to the LC solvents to improve signal intensity [443], and the raw SPC data was processed via SAINT exactly as in Experiment 1. For more specific details regarding the choice of the SAINT algorithm and the probability thresholds, please refer to the main text.

5.2 Designing a Method for XL-AP-MSMS

5.2.1 Surfaceome Proteomic Methods

In Chapter 4, using recombinant, non-natively glycosylated TLR4 and biophysical methods, I could not detect the expected high-affinity interaction between TLR4 and the FRePs. However, while comprehensive, my use of recombinant components was reductionist—important cellular variables, including the PTMs on TLR4, the TLR4 transmembrane domain and the membrane environment itself, and the interplay with accessory proteins were missing in the biophysical experiments. Most importantly, I hypothesized that as-yet unknown accessory proteins mediate the TLR4:FReP interaction at the cell-surface, explaining the divergence between the cellular and biophysical data. Therefore, to probe the potential of a TLR4-regulating receptorome and address the fundamental limitations of my recombinant work, I required an unbiased method to explore the cell-surface FReP-associated interactome—these requirements led me directly to bottom-up MSMS approaches. My stated

objective in selecting a proteomics method was to use a human cell system to identify the native cell-surface FReP-associated receptorome, hopefully including endogenous TLR4, which could explain why TLR4 is required for signalling but not sufficient to form a high affinity complex.

Before re-inventing the wheel, I probed the literature for best practice methodology in so-called “cell-surface proteomics”, which poses unique challenges relative to cytosolic proteomics [444]. First, nearly all cell-surface proteins are heterogeneously glycosylated, which can complicate identification of proteins due to the unpredictable mass shift(s) in glycosylated peptides. Second, cell-surface proteins are, by definition, either transmembrane or membrane-associated, meaning that standard lysis conditions, either mechanical or detergent-based, destroy the integrity of cell-surface interactions by disrupting the cell membrane. This also renders most antibody-based enrichment strategies, which are foundational for cytosolic proteins in IP-MSMS experiments, impossible; the disruption of the membrane during lysis often sufficiently denatures the antigen to cause dissociation of the antibody:antigen complex and prevent pull-down of the putative target. Finally, cell-surface proteins, particularly receptors, are much less abundant than intracellular proteins, due in part to the physical volume difference between the membrane and the cytosol. All of the technological solutions discussed briefly here and reviewed in full by Bausch-Fluck *et al.* [444], directly address these challenges with careful chemical-biology-based solutions.

The most comprehensive Cell-Surface Protein Atlas was built on cell-surface capturing (CSC) technology developed by Wollscheid *et al.* [445, 446]. In this approach, cell-surface proteins containing N-glycans are first activated and chemically tagged with a bi-functional cross-linker that covalently binds to N-glycan moieties under native conditions directly on the surface of living cells. Following cell homogenization, peptide digestion, and affinity enrichment of the labelled peptides using the other half of the cross-linker containing a biotin group, captured peptides are eluted by enzymatic cleavage at site of N-linked glycosylation before MSMS analysis. This method elegantly takes advantage of the N-linked glycans to label cell-surface proteins *prior* to lysis and digestion, but removes the glycans upon elution to simplify peptide identification. In this workflow, only “formerly N-glycosylated” peptides from cell-surface proteins are analyzed by MSMS. This method can detect the majority of the surfaceome, up to 1,492 human proteins and 1,296 mouse proteins.

While the CSC method highlights important tenants of cell-surface proteomics, it is

not amenable to directly studying specific receptor:ligand interactions. In 2012, Frei *et al.* reported a relevant extension of the CSC method that is broadly referred to as ligand-receptor capture (LRC) [447]. Frei *et al.*'s workhorse LRC method relies on a trifunctional cross-linker called "TRICEPS", which contains (1) an amine-reactive NHS ester for coupling to the ligand of interest (2) an aldehyde-reactive hydrazine for coupling the ligand of interest to glycosylated receptors after a binding event and (3) a biotin handle for affinity purification of cross-linked peptides [448]. In this approach, the ligand of interest is first coupled to TRICEPS with the NHS moiety and is then added, under native conditions, to relevant cells of interest, where it binds to its receptor. Following mild oxidation the hydrazine warhead covalently binds proximal receptor glycans, directly linking ligand and receptor peptides. The cross-linking predominately occurs only between the bait and directly bound receptors; to further increase reliability, the authors recommend comparison to an unrelated negative control bait protein to remove non-specific interactors. Following lysis, peptide digestion with trypsin protease, and enrichment of just the cross-linked peptides using the biotin handle from TRICEPS, *just* the receptor-derived portion of the cross-linked peptides are released via enzymatic cleavage of the N-linked glycans. In this method, therefore, only formerly N-glycosylated peptides from proximal receptors bound to the ligand of interest are analyzed by MSMS. By comparison to the aforementioned matched negative control, receptor-ligand binding can be quantified and unknown receptors for ligands of interest can be unambiguously identified.

Although quite nifty, LRC-TRICEPS can only detect interactions between ligands and N-glycosylated peptides in the receptor, which can reduce signal especially for receptors with few glycosylation sites or with predominantly O-linked glycosylation. Therefore, a further extension of the LRC-TRICEPS method was reported in 2017 by Tremblay and Hill [449]. Tremblay and Hill use a modified trifunctional cross-linker, called ASB biotin, which results in the *transfer* a biotin moiety from a ligand to a receptor(s). The ligand-reactive portion of ASB biotin is coupled via a cleavable disulfide bond, and the receptor-reactive portion again conjugates to N-glycans (with an aminooxy warhead). However, once the labelled ligand binds to the receptor and cross-linking occurs, the ligand can be released via reduction, resulting in a complete transfer of the biotin moiety from the ligand to the receptor. Essentially, then, this assay is an enzyme-free proximity biotinylation approach. Upon anti-biotin pull-down, the entirety of the biotinylated receptors, rather than just the glycosylated peptides, are enriched, and peptides covering the entire receptor sequence can be analyzed in the MSMS following post-pull-down digestion. This greatly increases the

coverage of receptor peptides and therefore the confidence of assigning interactors, and it means that receptors with few N-linked glycan sites can be captured.

I highlight the development of these two LRC methods because they are the gold standard for mapping unknown receptors to a purified ligand; in fact, the company Dualsystems Biotech will perform the LRC-TRICEPS analysis as a service with any ligand and any cell line. Unfortunately, this service is prohibitively expensive (over \$10,000 per ligand), and at-home implementation of either the TRICEPS or ASB biotin methods requires bespoke synthesis of the specialized trifunctional cross-linker. Therefore, I looked for cheaper options that built on the framework of the gold-standard LRC methods but utilized more readily available reagents. A 2007 paper from Freed *et al.*, used a much simpler assay to isolate TNF α associated receptor complexes [450]. Their approach relied on site-specifically adding a biotin molecule to the N-terminus of recombinant TNF α , which is then added to rat vascular endothelial cells. After incubation to ensure receptor engagement, Freed *et al.* added a cheap, commercially available, membrane impermeable, homobifunctional, lysine reactive, cleavable cross-linker, dithiobis(succinimidyl propionate) (DSP), to covalently link TNF α and its receptor(s) between two adjacent Lys residues. DSP contains a cleavable disulfide bond in the center; therefore, following harsh lysis, a stringent anti-biotin pull down, and denaturing elution in SDS, the cross-linker can be broken with reduction, releasing the two halves of the cross-linked peptides for MSMS analysis. A very similar method was employed by Byron *et al.* to map integrin-associated “adhesome” complexes [451]. Using beads coated with integrin-binding antibodies and the generic, cell-permeable, cleavable cross-linker dimethyl dithiobispropionimide (DTBP), Byron *et al.* could isolate and quantify massive mega-complexes composed of integrins and their extensive intra- and extra-cellular interactors.

The simpler methods from Freed *et al.* and Byron *et al.* are broadly similar, with just small differences in the bait (recombinant protein vs. antibody) and the cross-linker (impermeable vs. permeable), and they formed the foundation of the method I implemented herein. The choice of an impermeable or permeable cross-linker controls whether extracellular or extracellular and intracellular interactors are captured. Like the LRC methods, these cheaper methods maintain the use of a cross-linker to covalently tether the bait protein to cell-surface membrane proteins under native conditions prior to denaturing lysis, ensuring that PPIs are captured in their native lipid environment. The use of a generic cross-linker, while messier than the ligand-targeted trifunctional LRC reagents, can actually be viewed as a strength for

dissecting mega-complexes. Upon ligand binding, the proximal receptor is covalently cross-linked to the ligand, and *all associated co-receptors and accessory proteins are covalently cross-linked to the receptor*. By covalently tethering the labelled ligand of interest to both the immediate receptor and any associated neighbors, entire ligand-associated protein communities can be pulled-down and measured. In the LRC methods, just the immediate receptor(s) are cross-linked and captured in the analysis, and the “transitive” cross-linking features of DSP or DTBP are lost. Therefore, based on both cost and applicability to my FReP receptorome hypothesis, I selected the methods reported by Freed *et al.* and Byron *et al.* as a template, rather than further pursuing the gold-standard LRC methods.

5.2.2 Final Design of XL-AP-MSMS Method

Even after selecting the big-picture strategy for the unbiased proteomics, I still had to make a variety of choices regarding the details. To maximize the specificity of the pull-down while maintaining simplicity, I selected a biotin-streptavidin affinity system, which achieves near-covalent immobilization strength, allows stringent washing conditions, and is broadly considered best practice as an affinity system in the mass spectrometry community [452]. Biotinylation of recombinant proteins can be achieved in *E. coli* using a C-terminal AviTagTM on the protein of interest and co-expression with the biotin ligase BirA [453]. I quickly showed that the CyDisCo expression system could be adapted to facilitate biotinylation of the FRePs, discussed in detail in Section 5.3, allowing me to generate large quantities of biotinylated bait protein. Using biotinylated FRePs as the bait circumvented the use of antibodies, which can be unreliable for pull-downs especially for the understudied FReP family members. Next, I selected primary human monocytes as the target cell for the pull-down, based on the extensive characterization of TLR4-dependent, FReP-mediated activation of inflammation in these cells in the literature and by others in my group. Working with primary cells made the method more challenging than using immortalized monocytic cell-lines like THP1. However, this choice ensured the identified interactors were biologically relevant to inflammation and *a priori* maximized the chance of picking out TLR4, which is expressed relatively highly on these innate immune cells. Fortunately for me, an excellent student colleague of mine, Anna Marzeda, worked up a robust method for isolating huge numbers of monocytes ($>1.2 \times 10^9$ cells) from a single donor, which greatly facilitated the use of primary cells in these experiments. Finally, I decided to focus initially on

impermeable cross-linkers *a la* Freed *et al.*, to capture just extracellular interactions with immediate co-receptors.

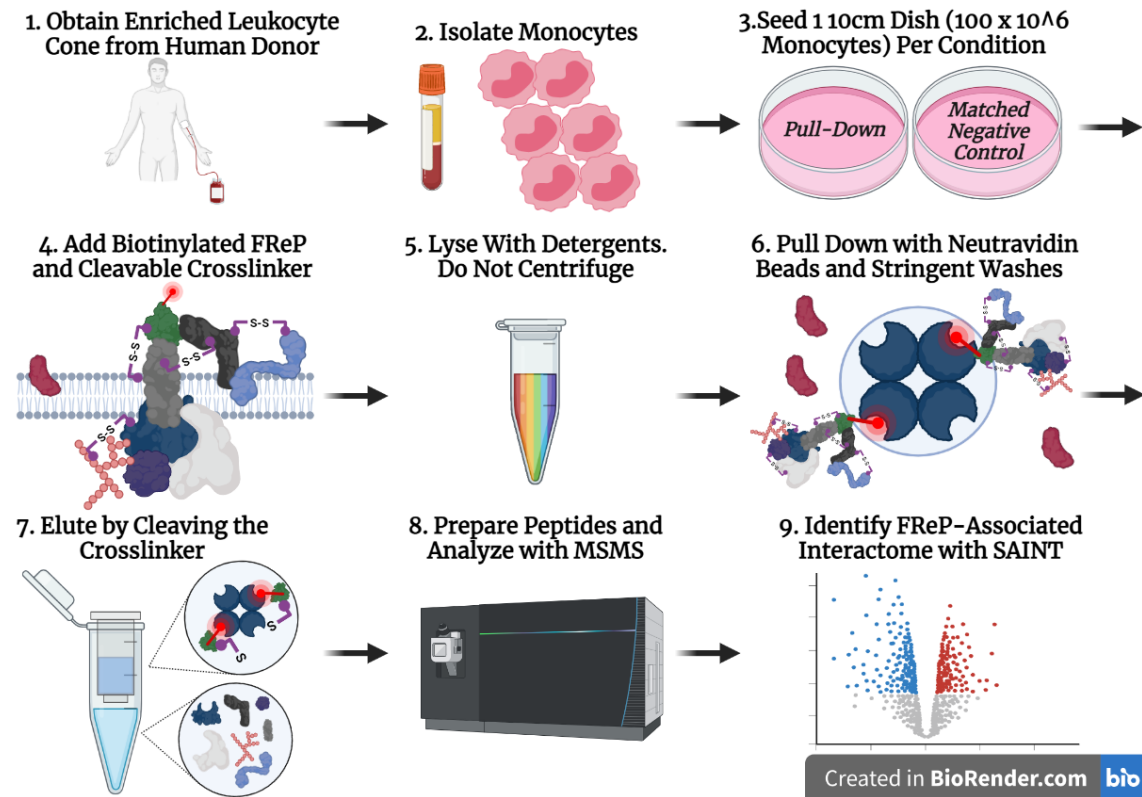


Figure 5.1 XL-AP-MSMS Method Overview. A. Schematic of the XL-AP-MSMS method developed and used throughout Chapter 5 to probe the FReP-associated interactome.

With all of this in mind, I settled on a final method for my XL-AP-MSMS workflow which is diagrammed in full in Figure 5.1. For clarity, and to highlight useful features in the method, I describe each step of the method below:

1. Obtain an enriched leukocyte cone from a human donor.

Through the Kennedy Institute's existing channels, I obtained leukocyte cones directly from human donors which contained concentrated and enriched white-blood cells isolated by leukapheresis.

2. Isolate primary human monocytes.

To isolate monocytes, contaminating red-blood cells and other immune cells (T-cells, B-cells, neutrophils, ect.) must be removed. Many groups use commercially available magnetic cell sorting (MACS) kits to isolate monocytes with marker-specific antibodies. While the final cell population is highly pure

following MACS, the process recovers only a fraction of the monocytes present in the cone. For this XL-AP-MSMS method, I required hundreds of millions to billions of monocytes, making the MACS method inefficient. In short, I needed an isolation method that favored cell count over cell purity. Therefore, with the help of Anna Marzeda, I implemented an “old-school” plating method to isolate monocytes. In this approach, red blood cells are lysed and removed from the sample and the entire white-blood cell fraction is plated onto TC-treated dishes. After a two hour incubation, monocytes preferentially adhere, and contaminating cells can be washed away with aggressive PBS washes. Using this simpler method, Anna helped me to routinely isolate over a billion monocytes per cone. By counting under a microscope, the cells were over 90% monocytes, with a minority population of contaminant T-cells.

3. Seed 10 cm dishes with 100 million monocytes per dish.

To maximize the amount of material captured in the pull-down, Anna and I seeded 70–100 million monocytes per 10 cm dish. From a single cone we could usually seed twelve dishes, with enough cells to complete six pull-downs (one with bait added, one with a directly matched negative control treated with an equal volume of the bait protein buffer).

4. Add biotinylated FReP and cleavable cross-linker.

After resting the seeded monocytes overnight, I added the biotinylated FReP of interest directly to the cell culture medium at a final concentration between 1 and 5 μM . This concentration is sufficient to strongly activate pro-inflammatory signalling in primary human monocytes, meaning the TLR4-associated signalling complex must form at this concentration. I hesitated to push the concentration further to prevent non-specific interactions and to maintain the LPS levels at acceptable levels for *in vitro* assays. After a brief incubation, I then added the water soluble analog of DSP, dithio-bis(sulfosuccinimidylpropionate) (DTSSP), to covalently cross-link the FReP bait to the extracellular receptorome. After thirty minutes, any un-reacted cross-linker was quenched with excess Tris to ensure that all interactions were trapped prior to cell lysis.

5. Lyse with detergents.

After cross-linking, I then collected the monocytes, which required centrifuging the cell culture medium (to capture detached monocytes) to pool with the cells still adhered to the plate. I completed this tedious two-step harvest by hand for each plate to ensure maximal input material to the pull-down. Note that

the cell culture medium containing excess unbound FReP was discarded at this stage, ensuring that only cell-associated bait was pulled down. For lysis, I used the commercially available Mammalian Protein Extraction Reagent (MPERTM; Thermo Fisher), which contains a mixture of gentle detergents to solubilize both cytosolic and membrane proteins. I did not centrifuge prior to affinity enrichment to ensure that both soluble, insoluble, and/or micelle-enriched proteins were included in the input. To reduce the viscosity in the absence of a centrifugation step, I added a small amount of benzonase to degrade contaminating gDNA.

6. Pull down with NeutrAvidinTM beads and stringent washes.

To pull-down the biotinylated FRePs and the cross-linked interactors, I incubated the lysates for at least two hours with PierceTM High Capacity NeutrAvidinTM Agarose (Thermo Fisher), which is the highest capacity and most specific anti-biotin resin available. For washing, I employed harsh buffers, including high salt (0.9 M NaCl) or high detergent (1% Triton X-100). These harsh washes were facilitated by the remarkably stable biotin:NeutrAvidinTM interaction and the covalent linkage between bait and prey proteins.

7. Elute interactors by cleaving the cross-linker.

This step represents the main improvement I made in the previously reported methods. To elute the interacting proteins, I added a high concentration of reducing agent directly to the resin. This step breaks the DTSSP in half and causes prey proteins, but *not* the FReP bait, to dissociate and elute. This overcomes a challenge of AP experiments wherein the prey itself is by far the most prevalent protein in the eluate. This can swamp the signal of other low abundance peptides in the MSMS and prevent detection of important but less abundant interactors. This step is especially important in my case because I relied on direct solution digestion of the eluates prior to MSMS analysis, rather than first purifying the samples on a gel and then cutting prey gel bands for MSMS.

8. Prepare peptides and submit for MSMS analysis.

After elution, I prepared tryptic peptides for MSMS using standard methods. In short, I removed the elution buffer *in vacuo* prior to denaturation in Urea, reduction and alkylation of Cys residues, and overnight digestion with trypsin protease. Finally, I desalted the digested peptides using C18 Stage Tips and then shipped dried peptides to my mass spectrometry collaborator Dr. Holger

Kramer at the MRC London Institute of Medical Sciences at Imperial College for analysis on high-performance Orbitrap mass spectrometers.

9. Identify FReP-associated interactome with SAINT.

The data provided by Dr. Kramer included the proteins identified at a 1% FDR threshold and the number of spectral-peptide counts (SPCs), which is a standard quantitative measure of the absolute abundance of a particular protein in the sample. I then compared the FReP-containing pull-downs with the matched negative controls using a gold-standard statistical method, called SAINT (Significance Analysis of INTeractome), to identify high-confidence “true” interactors and remove non-specific or low-confidence background contaminants [442]. The final target list, containing the FReP interactome from primary human monocytes, was then subjected to additional bioinformatic analyses.

Throughout the experiments I present next, I will highlight some small changes to this core method, including switching to a cell permeable cross-linker and removing detergent from the wash buffer, which were made to address specific emergent challenges. As I will demonstrate, this XL-AP-MSMS method proved robust, reproducible, and informative, and it represents an approachable and relatively inexpensive generic solution for mapping multi-protein receptoromes in the presence of a biotinylated ligand of interest.

5.3 Purification of Biotinylated FRePs

The first step towards executing the XL-AP-MSMS workflow was to purify biotinylated FReP proteins to serve as bait. Dong Wang from the SGC Biotechnology Group generously re-cloned full-length ANGPTL1, ANGPTL2, ANGPTL6, FCN1, FGL1, FGL2, MD-2, TNC, and the truncated crystallizable construct TNC-c042 (which herein I call TNC^{XTAL}), into the vector pNIC-Bio3, which encodes for a TEV-cleavable N-terminal His6-tag and an uncleavable C-terminal AviTagTM. For *in situ* biotinylation during expression, I co-transformed CyDisCo competent cells (already Chlor^R) with the FReP constructs (Kan^R) and a helper vector containing an IPTG-inducible BirA biotin ligase (Strep^R) previously prepared by Dr. Michael Fairhead. Selection on triple antibiotic plates yielded a custom *E. coli* strain capable of simultaneously co-expressing the FReP of interest, the CyDisCo co-chaperones

for folding, and BirA for biotinylation. I performed a 50 mL test expression of this system exactly as reported in Section 2.1.5, except with the addition of excess free biotin to both the cell culture medium prior to induction and the lysis buffer. After a one-step IMAC purification and intact mass analysis, I demonstrated that ANGPTL1, ANGPTL2, ANGPTL6, FGL1, TNC, and TNC^{XTAL} could be solubly expressed with a single C-terminal biotinylation in CyDisCo cells (see Figure 5.2A). FCN1, FGL2, and MD-2 could not be expressed in this system. After confirming expression of the six biotinylated constructs, I then scaled-up biotinylated TNC using Purification Scheme 2 (see Section 2.1.7). From a 2 L purification, I obtained over 10 mg of monodisperse, pure protein (Figure 5.2B) that was 100% singly biotinylated and contained the expected two disulfide bonds based on intact mass analysis (Figure 5.2C). Therefore, biotinylated TNC was biophysically identical to the un-tagged TNC I relied on for the assays discussed in Chapter 4.

However, the biotinylated TNC needed to pass a second, more stringent step of quality control prior to use in cellular assays. First, I polished the purified protein using commercially available PierceTM High Capacity Endotoxin Removal Resin (Thermo) to remove contaminating LPS prior to adding the protein to the monocytes. I reasoned that high concentrations of LPS could saturate TLR4 and out-compete the FReP-mediated binding event of interest; therefore, I tried wherever possible to use endotoxin-free protein that met a stringent threshold of less than 10 pg mL⁻¹ LPS. All of the biotinylated TNC used in Section 5.5 and Section 5.6 contained 5.1 pg mL⁻¹ LPS when tested at 1 μ M in a standard commercially available assay (see Section 5.1.3). This level is considered endotoxin-free for cellular assays and is below the threshold for TLR4-mediated detection of LPS. However, LPS removal is non-trivial and expensive; in fact, based on the experience of others in my group, I was exceedingly lucky that TNC could be concentrated easily and survived treatment with the Endotoxin Removal Resin. In the portions of this chapter where I used LPS-contaminated protein for pilot experiments as a result of these challenges, including FGL1, ANGPTL6, and TNC^{XTAL}, I will explicitly specify so in the text.

Finally, Anna Marzeda graciously completed a standard monocyte stimulation assay to directly compare un-tagged TNC and biotinylated TNC. In this assay, reported in Figure 5.2D, primary human monocytes were incubated with endotoxin-free protein for 24 hours prior to readout of pro-inflammatory cytokine levels with a standard ELISA. Just like un-tagged TNC, biotinylated TNC stimulated the strong release of IL-6 and TNF α at a concentration of 1 μ M. The response was similar to that elicited by directly adding 100 ng mL⁻¹ LPS to the monocytes. Interestingly, biotinylated

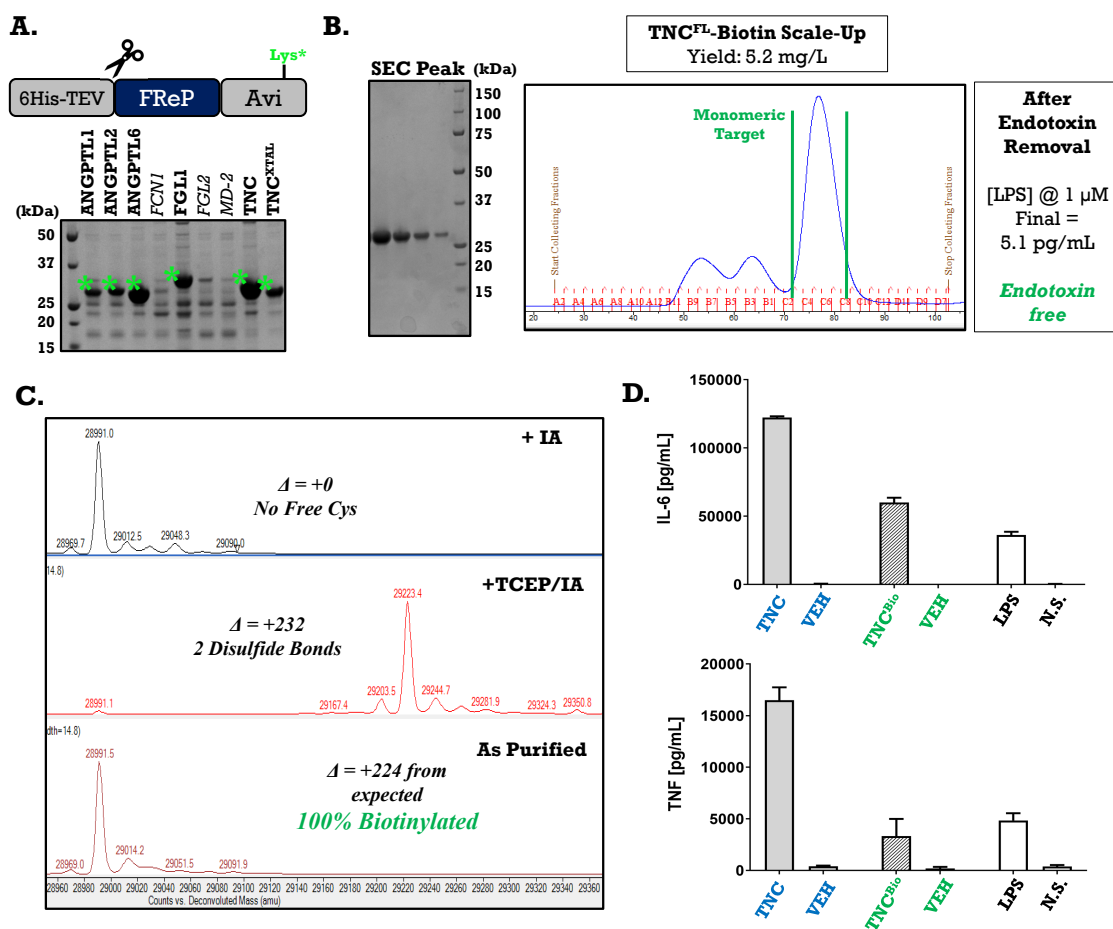


Figure 5.2 Purification and QC of Biotinylated FRePs. **A.** 50 mL test expression results showing that most FRePs can be expressed C-terminally biotinylated in CyDisCo cells. Green asterisks denote bands with MSMS confirmed identity and intact mass confirmed biotinylation. **B.** Large-scale purification of biotinylated TNC, showing that the protein is pure, monodisperse, and endotoxin-free. **C.** Disulfide mapping assay showing that biotinylated TNC retains two disulfide bonds. **D.** Stimulation of primary human monocytes shows that 1 μ M biotinylated TNC elicits pro-inflammatory cytokine release (IL-6, top; TNF α bottom) just like un-tagged TNC. N.S = unstimulated; VEH = buffer only.

TNC induced lower cytokine levels than un-tagged TNC, which could be due to the C-terminal AviTagTM interfering with the activity of the C-terminal cationic tail (see Section 2.7). However, despite being slightly less active than un-tagged TNC, biotinylated TNC was clearly pro-inflammatory in primary human monocytes, proving that the biological activity and critical receptor binding properties were retained. Taken together, these results demonstrate that biotinylated TNC is a reliable affinity bait for the XL-AP-MSMS workflow that reasonably recapitulates both the biophysical and biological properties of un-tagged TNC.

5.4 Pilot Experiments 1, 2 and 3

With a reliable biotinylated reagent in hand, I moved forward with three pilot experiments to validate the XL-AP-MSMS method and work out quirks in the protocol before proceeding with full MSMS analysis. The first Pilot 1 was essentially a disaster, so I only briefly discuss it here to highlight the lessons I learned. I used MACS to isolate the monocytes, limiting the overall cell numbers, and then I differentiated the monocytes into macrophages over five days with GM-CSF, which kills a significant fraction of the precursor monocytes. I originally thought macrophages would be more reliable than monocytes because macrophages express more TLR4 based on data from the Human Protein Atlas [454]. However, by the time I finished isolating the macrophages, I only had 8×10^6 cells, enough for one pull-down and one matched control with just 4×10^6 macrophages per condition. I completed the cross-linking pull-down with DTSSP using $1.8 \mu\text{M}$ LPS-contaminated biotinylated TNC as the bait and ran the eluates on SDS-PAGE followed by coomassie staining. As expected, I essentially saw nothing; the eluate bands, with or without the bait, contained no protein, and even the undiluted lysate samples were faint on the gel.

In short, Pilot 1 suggested that I needed significantly more cells and that I likely could not rely on coomassie-stained gels to visualize the eluates. Therefore, based on the excellent suggestion of Anna Marzeda, for Pilot 2 I switched to monocytes isolated via plating rather than MACS. This became the core isolation step in the protocol moving forward. Monocytes respond well to biotinylated TNC (Figure 5.2D), they are routinely used for stimulation with FRePs and other DAMPs, and they do not require additional differentiation after isolation. Furthermore, publicly available mass spectrometry data from pax-db shows that TLR4 can be detected from monocytes (<https://pax-db.org/>). The old-school plating method, which bypasses MACS and relies on the propensity of monocytes to strongly adhere to TC-treated dishes, simplified the protocol and greatly increased cell yields. During Pilot 2, from just half of a donor blood cone, I was able to isolate 400×10^6 cells, enough for one pull-down and a matched negative control with 200×10^6 monocytes. I completed the cross-linking pull-down with 2 mM of the impermeable cross-linker DTSSP and using $4.5 \mu\text{M}$ biotinylated TNC as the bait. After multiple, 100 CV washes with PBS containing 0.9 M NaCl, I eluted the interactors and visualized the eluates with SDS-PAGE. This time, however, I silver-stained the gel to increase the sensitivity of the readout.

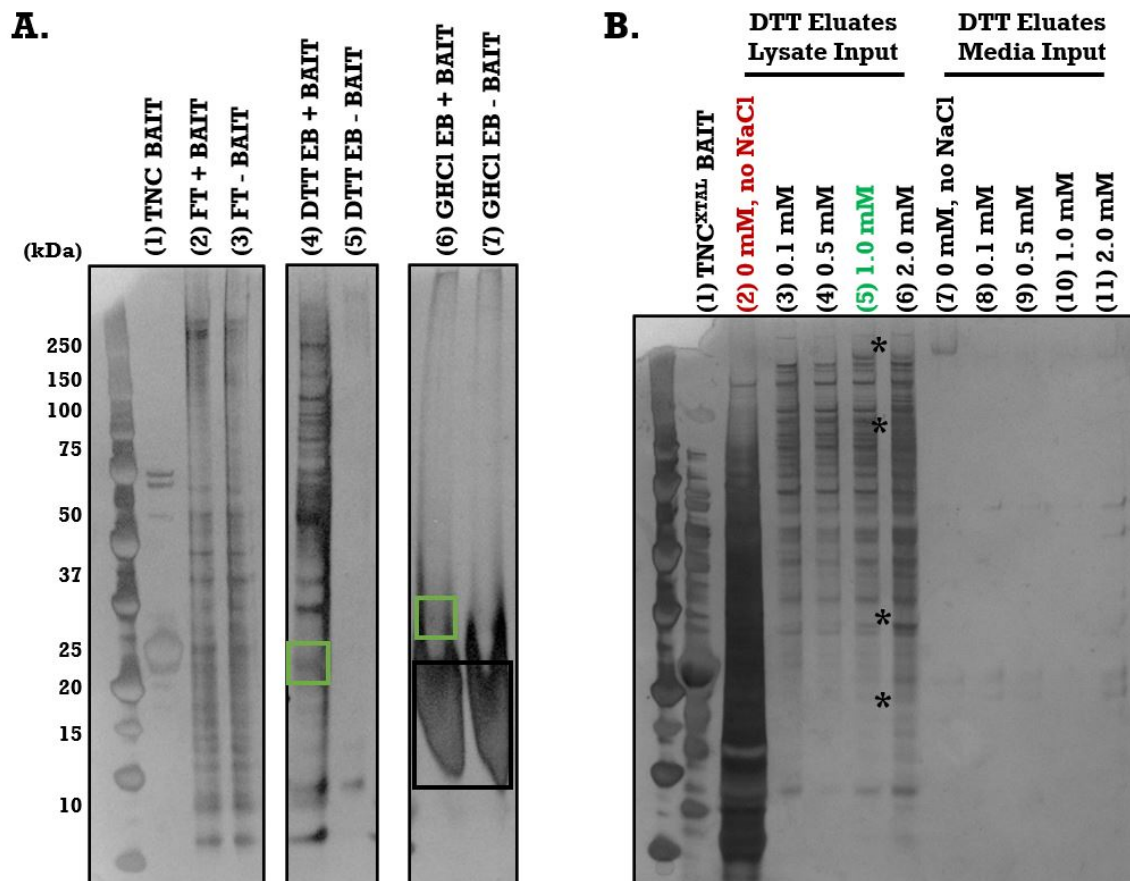


Figure 5.3 XL-AP-MSMS Method Development Pilot 2 and Pilot 3. **A.** Silver-stained gel from Pilot 2. Putative TNC bands are highlighted with green boxes, while contaminating NeutrAvidinTM bands are marked in black. FT = flow-through, DTT = first eluate after the addition of 50 mM DTT to cleave DTSSP, GHCl = second eluate after the denaturing addition of 6 M guanidine hydrochloride. **B.** Silver-stained gel with DTT eluates from Pilot 3. Five different [DTSSP] were used to pull-down TNC interactors from either lysate (left) or cell culture medium (right). Putative interactors missing in the lower [DTSSP] are marked with black asterisk. The uninterpretable negative control, washed incorrectly, is colored in red.

This time the final gel from Pilot 2, shown in Figure 5.3A, demonstrated that the pull-down worked exactly as I had hoped. Lane 1 shows that, even after a sensitive silver stain, the biotinylated TNC used as bait was highly pure. Lane 2, containing the flow-through after a two-hour incubation with the anti-biotin NeutrAvidinTM resin, matches the negative control Lane 3, showing that all of the bait protein was captured. If too little resin had been used or the lysate had not been incubated for long enough with the lysate, a strong un-captured TNC band near the 25 kDa marker would appear in Lane 2. Lanes 4 and 5 contain the final eluates following the addition of 50 mM DTT to cleave the DTSSP cross-linker and release the TNC

receptorome from the resin. Critically, comparison of Lanes 4 and 5 shows that the pull-down is very specific, and the presence of protein in the eluate is TNC-dependent; the bait-free negative control eluate, Lane 5, contains no visible protein bands, whilst the bait-containing eluate Lane 4 contains dozens of proteins. Furthermore, the TNC bait band at 25 kDa, highlighted with a green box in Lane 4, does not overwhelm the sample. This confirms that cleavage of the DTSSP releases predominantly interacting prey proteins and leaves behind the biotinylated TNC bait. To further prove this assertion, I also completed a second elution with denaturing guanidine hydrochloride, which destroys the NeutrAvidinTM agarose and releases the resin protein and anything else bound. A band corresponding to retained TNC appears in Lane 6 (green box) but not in the negative control Lane 7. Note that the sample is distorted due the large NeutrAvidinTM contaminant, marked with a black box. This is a key advantage of my method—by eluting the pull-down with non-denaturing cleavage of the cross-linker, the eluates contain primarily the interacting proteins without contamination with excess bait or NeutrAvidinTM. This maximizes the likelihood of detecting interactors during MSMS, since mass spectrometers preferentially detect the most abundant species.

In summary, Pilot 2 demonstrated that specific TNC-dependent interactors could be pulled down directly from primary human monocytes using the cleavable cross-linker DTSSP. However, before proceeding with full MSMS characterization of the interactors, I wanted to check a few additional variables. In Pilot 3, I optimized the concentration of DTSSP (previously at 1 mM) used during the cross-linking step, checked to ensure that I was not missing any critical soluble interactors from the cell culture medium by pulling-down only from the cell pellet, and experimented with using fewer cells per condition. During Pilot 3, I seeded five dishes each with 100×10^6 monocytes, half the cell number used in Pilot 2. Due to protein limitations at the time, I used 1.8 μ M of the truncated biotinylated TNC^{XTAL} construct (matching the crystallizable TNC-c042, containing a five residue deletion at each terminus) as bait; as Lane 1 in Figure 5.3B shows, this protein was significantly less pure than the bait used in Pilot 2. After adding the bait, I treated the monocytes with five different concentrations of DTSSP: 0 mM, 0.1 mM, 0.5 mM, 1 mM, and 2 mM. After the cross-linking, I then harvested the cells as in Pilot 2, but I also retained the conditioned cell culture medium, theoretically containing soluble TNC interactors. I then pulled down both the lysate and the medium from each of the five [DTSSP] using NeutrAvidinTM agarose, followed by a single reducing elution with DTT. I analyzed the eluates with a silver-stained SDS-PAGE as before, the results of which are presented in Figure 5.3B.

During the washing stage of the pull-down, I made a mistake in Pilot 3. Out of a desire to look for TNC interactors present *without* requiring a cross-linker to stabilize them, I chose to use just PBS (rather than PBS with 0.9 M NaCl) as the wash buffer for the 0 mM DTSSP samples. As it turned out, PBS is a woefully insufficient wash buffer, and the eluate from the 0 mM DTSSP sample (Lane 2) contained a smear of non-specific bands. This means I have no proper negative control in this experiment for comparison, because I did not include a “no bait” comparator like in Pilot 2. In some ways, the messiness of Lane 2 highlights the critical value of the large 0.9 M NaCl washes. Despite the glaring omission of a bait-free, properly washed negative control, Pilot 3 still provided important information about the method. First, as Lanes 7–11 shows, there are essentially no TNC-specific interactors in the cell culture medium, which validated the use of only the cellular fraction for the pull-down. Secondly, the cross-linker titration spanning Lanes 3–6 showed that 1 mM DTSSP is appropriate. At higher [DTSSP], non-specific cross-linking appears, indicated by the smeary, indistinct bands in Lane 6. Furthermore, at lower [DTSSP] in Lanes 3 and 4, some bands (indicated with black asterisks) disappear relative to Lane 5. Therefore, 1 mM DTSSP represents the optimal balance between capturing the most interactors and avoiding non-specific cross-linking. Pilot 3 also demonstrated that 100×10^6 monocytes is sufficient input material to observe bands in the eluate, so I maintained approximately this cell density in later experiments. Finally, on a more general level, Pilot 3 represented an independent biological replicate of Pilot 2. Using cells from a different donor, with the dirtier and different TNC^{XTAL} construct as bait, I still detected TNC interactors that could be specifically eluted by cleaving the DTSSP cross-linker. Therefore, after these three pilots, I felt confident enough to move forward with full MSMS experiments to confidently identify the specific TNC interactors coming through the pull-down.

5.5 Full Experiment 1 with DTSSP

With a validated method in hand, I moved forward with two full XL-AP-MSMS experiments, complete with sample analysis on high-end mass spectrometers in collaboration with Dr. Holger Kramer at Imperial College London. The experiments were completed independently almost three months apart, and the results from the first experiment proposed a hypothesis confirmed by the second experiment. Therefore, I present the results separately and chronologically over the rest of Chapter 5. Also, at this point, COVID-19 restrictions prevented me from working in the

Kennedy Institute, and I was not allowed to process human blood at the SGC. So, all of the primary human monocytes used in these experiments were prepared by Anna Marzeda and given to me.

5.5.1 Experimental Setup and Data Analysis

For Experiment 1, Anna Marzeda processed an entire leukocyte cone from “Donor 88” and seeded eight plates each containing approximately 100×10^6 monocytes. From spare cells derived from a separate donor, Donor 87, Anna also seeded two additional plates, giving me a total of ten plates for this experiment and allowing me to complete five pull-downs with five matched controls. I used four plates, two from Donor 88 and two from Donor 87, to analyze full-length, LPS-free, biotinylated TNC from the large-scale purification presented in Figure 5.2B. I used two plates to analyze LPS-contaminated, biotinylated, His6-tagged ANGPTL6, two plates to analyze LPS-contaminated, biotinylated, His6-tagged ANGPTL2, and the final two plates to analyze LPS-contaminated, biotinylated, His6-tagged FGL1, all with protein purified from the medium-scale test expression presented in Figure 5.2A. I attempted to polish ANGPTL2, ANGPTL6, and FGL1 to remove contaminating endotoxin; however, while the treatment reduced the LPS levels by more than 75%, the levels were not low enough to be considered truly LPS-free (exact levels reported in Section 5.1.3). All of the baits used were of high purity, as demonstrated in the silver-stained gel shown in Figure 5.4A. I decided to perform the pull-down with three different FRePs, from three different subfamilies, to explore FReP-specific interactors. Throughout, I used $2 \mu\text{M}$ of the bait protein (or the equivalent volume of buffer for the negative controls) and incubated the bait with the monocytes for 30 minutes before cross-linking with 1 mM DTSSP. As before, I incubated the cell lysate for two hours with NeutrAvidinTM agarose and eluted with 50 mM DTT. However, based on a freshly published protocol from Cho *et al.* and my desire to maximize the specificity of the pull-down, I used a wash buffer containing 0.9 M NaCl and 1% Triton X-100 detergent [455]. Because of the covalent cross-linking and the essentially covalent interaction between NeutrAvidinTM and biotin, incredibly harsh wash buffers are compatible with the method; in fact, Cho *et al.* used 0.1% SDS and 1% Triton X-100 in their lysis and wash buffers for anti-biotin pull-downs. I will return to this choice, which I ended up reversing for the next experiment, later in Section 5.6.

For the first time during Experiment 1, I prepared the eluates for MSMS analysis rather than running them on a gel. First, I dried the eluates *in vacuo*, and next denatured them with 8 M Urea prior to reduction and alkylation of free Cys residues with IA. As described in Section 5.1, I considered the large amount of DTT present in the eluate and added sufficient IA to quench the DTT and still give a standard final concentration of 20 mM IA for alkylation. After diluting out the Urea with ammonium bicarbonate, I digested the proteins overnight with a 1:20 w:w ratio of sequencing-grade trypsin protease (Sigma Aldrich). The generated peptides were desalted the next morning with C18 stage tips according to the manufacturer’s recommendation and then dried *in vacuo* before shipping to Dr. Kramer for analysis. Dr. Kramer analyzed the samples in technical duplicate on a Q ExactiveTM Orbitrap Mass Spectrometer (Thermo Fisher) and provided both protein hit lists filtered at a 1% FDR rate as well as raw spectral counts (SPC) for quantification.

A key step in any AP-MSMS workflow is filtering the data to separate high-confidence, true interactors from non-specific ones, which can pollute the hit list and cloud interpretation of the data. Although many algorithmic methods exist to accomplish this task, I implemented a gold-standard method known as SAINT [442]. Upon supplying the program with raw spectral counts for each of the identified prey proteins in the pull-downs and matched negative controls, SAINT calculates the posterior probability of a true interaction between each bait and each prey protein in the dataset using a Bayesian statistical modeling approach. SAINT generates two probability distributions for each unique bait-prey interaction, one for the probability of a true interaction based on the spectral counts from the bait-containing pull-downs, and one for the probability of a “false” (i.e. non-specific) interaction derived from the spectral counts from all of the negative control bait-free pull-downs. Note that SAINT *pools* all of the negative controls to generate a higher-confidence estimation of the false interaction probability distribution, which increases the sensitivity during MSMS workflows with a small number of biological replicates (usually between $n = 3$ and $n = 5$). SAINT also normalizes the spectral counts for the parent protein length (i.e. the larger the protein, the more peptides that must be detected and vice-versa). Comparison of the distance between the “true” and “false” probability distributions generates a summary probability (SP) for each bait-prey pair, ranging from 0 (absolutely a false interaction) to 1 (absolutely a true interaction), allowing specific, high-confidence interactors to be identified across the dataset. In real terms, more spectral counts in the bait-containing pull-down(s) and/or fewer spectral counts in the negative controls increases the likelihood of a true interaction. SAINT results are often filtered at an SP level of either 0.9 or 0.95, representing a 90% or 95%

chance of a true interaction with a specific prey. SAINT is considered a rigorous and relatively stringent way to identify true interactors in pull-down experiments: it out-performs many other algorithms in the detection of low-abundance (i.e. few SPC) but still high-confidence interactors, and it is commonly used to identify interactors across a variety of AP-MSMS workflows [456, 457]. Conveniently, the SAINT algorithm is implemented in the CRAPome web portal, which facilitates the analysis of AP-MSMS data for free and without any local installation [441]. The CRAPome portal also outputs simplistic log-fold changes in the spectral counts (i.e. prey SPC with bait divided by the background prey SPC in the negative controls) alongside the more complex SAINT-derived SP scores [441]. For these reasons, I used SAINT as deployed on the CRAPome server as the workhorse analytical software in this Chapter to identify and score FReP-interactors.

5.5.2 Full Experiment 1 Key Results

The raw MSMS ion intensities per sample, a relative indicator of the total amount of peptides present, are presented for Experiment 1 in Figure 5.4B. Note that, except for the TNC pull-down from Donor 88, the matched negative control contains fewer peptides than the bait-containing pull-down. I expected this to be the case *a priori*, especially based on the degree of enrichment observed with silver-staining in Pilot 2 (Figure 5.3A, Lanes 4 and 5). The reversal of this pattern from the TNC Donor 88 sample was puzzling—it is my hypothesis that I either forgot to add TNC to the monocytes or, more likely, I switched the samples at some point. The workflow requires transitioning from TC plates to Falcon tubes to spin columns to eppendorfs to desalting columns and finally to fresh eppendorfs over the course of a two-day experiment, creating ample opportunity to switch the samples. However, with no concrete proof of a sample mix-up, I decided to exclude TNC Donor 88 from the analysis and use only the TNC Donor 87 sample, accepting the caveat that I would not be able to compare TNC to other FRePS in exactly the same donor.

Next, I analyzed the raw SPC data with SAINT, but first I would like to emphasize a critical point. Throughout this Chapter, I err on the side of caution when identifying hits. The SAINT authors consider scores above 0.8 to be informative [442], and scores above 0.9 are considered stringent. I apply at least this level of stringency (and for network analysis to follow, I use 0.95 as a cut-off) for a reason—the cross-linking step can enrich transient interactions and lead to the pull-down of entire protein interaction networks. Therefore, neighbors of neighbors of the direct binding

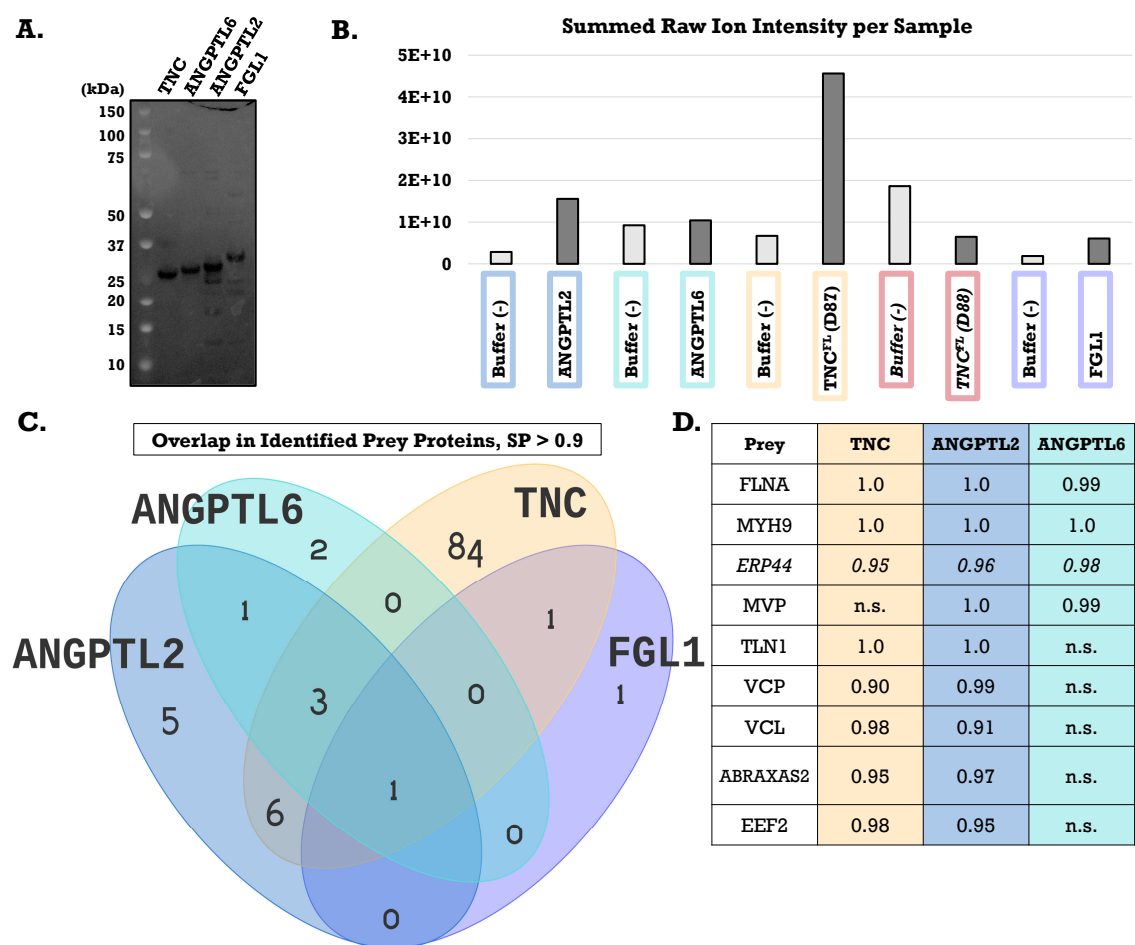


Figure 5.4 XL-AP-MSMS Experiment 1 Overview. **A.** 2 μ g of each bait protein analyzed for purity with silver-stained SDS-PAGE. **B.** Raw MSMS ion intensity per sample. Except for TNC from Donor 88 (marked in red), the intensity in the matched negative control is lower than the bait-containing pull-down. **C.** Overlap of identified interactors (SP > 0.9) in Experiment 1, colored as in **B.** The unreliable TNC Donor 88 data is not included. **D.** Gene ID's for preys identified from at least two of the baits (TNC, ANGPTL2, or ANGPTL6).

event can be detected; this is in fact the whole point of the method focusing on receptoromes. Nevertheless, this feature can lead to many peripherally associated proteins of dubious biological relevance appearing as interactors. For example, if a receptor which is bound to the actin cytoskeleton is ligated, it is reasonable to expect *a priori* that *all* other proteins bound to the actin cytoskeleton (for reasons related to the bait or not) could be enriched in the pull-down. Therefore, in an effort to drill down to the most dominant and functionally relevant protein communities enriched in the data, I used quite stringent thresholds to weed out some of the noise.

Now, back to the data. After excluding the baits themselves, I identified 103 unique,

high-confidence interactors with an SP score greater than 0.9. The flower-plot shown in Figure 5.4C shows the overlap between interactors identified for each bait. Interestingly, 82% of the interactors were unique to TNC, which aligns well with the raw ion intensity data; TNC Donor 87 had by the far the most peptides, and therefore SAINT can detect more interactors with higher confidence. Since an equal amount of each bait protein was used in the pull-down, the over-representation of TNC interactors could imply that TNC binds to and modulates the monocytes most potently. FGL1 only pulled-down three proteins with high confidence, two of which (ERP44 and P4HB) were CyDisCo component homologs that almost surely represent contaminants rather than physiological partners; in fact, the ERP44 contaminant was the only protein detected in all four samples. All prey proteins appearing as high-confidence interactors of at least two of ANGPTL2, ANGPTL6, or TNC are listed in Figure 5.4D.

To better explore the biological relevance of the interactors, I next completed a network analysis with the highest confidence identified prey proteins that interact with ANGPTL2, ANGPTL6, and TNC. I considered any prey with an SP score above 0.95 and a $\log_2$ fold-change greater than 1.5, and I integrated previously reported interaction information directly from the BioGrid database [458]. The results of this analysis are presented in Figure 5.5. The most striking feature of the network is the central hub, containing the actin cytoskeleton and cytoskeleton-associated proteins, which is present across all three baits and is highly interconnected based on previously reported data. Because cytoskeletal proteins are often contaminants in MSMS work-flows, I want to emphasize here that these proteins were identified as very high confidence interactors by SAINT, and that the raw peptide counts confirm that these proteins were pulled-down in a FReP-dependent manner. For example, across all negative control pull-downs, I detected only a single myosin heavy chain 9 (MYH9) peptide; however, I observed 297 SPC in the TNC pull-down, 11 in the ANGPTL6 pull-down, and 20 in the ANGPTL2 pull-down. Therefore, this cytoskeletal hub is not a contaminant—it is a biologically relevant feature of the FRePs.

The next questions, then, became obvious. *What* biologically relevant feature? How can an extracellular ligand, cross-linked to extracellular membrane proteins, specifically pull-down a fully intracellular cytoskeletal network? It turns out that many of the proteins in the central hub are known to be associated with activated integrin adhesomes. In the network analysis presented in Figure 5.5, proteins annotated in the adhesome.org database are marked with a black asterisk [459], and proteins identified by Byron *et al.* as being enriched specifically in activated adhesomes are

marked with a purple asterisk [451]. Integrins are heterodimeric transmembrane receptors which facilitate communication between cells and the ECM by integrating diverse extracellular signals, especially via a direct and physical connection between integrins and the cytoskeleton. Integrin signalling relies on the assembly and coordination of up to two hundred diverse effectors, which together are termed the adhesome [460]. Adhesome constituents change depending on the ligand binding event and on the existing intracellular environment, allowing Byron *et al.* to map differences in the adhesome upon the binding of activating and inactivating anti-integrin antibodies [451]. Eleven of the high confidence interactors I identified are enriched in activated integrin adhesomes in the Byron *et al.* proteomics dataset; integrin-linked protein kinase (ILK) is the only protein identified as an interactor that is preferentially enriched in inactive adhesomes. Furthermore, four of most common, “core” adhesome components [filamin A (FLNA), MYH9, talin-1 (TLN1), and vinculin (VCL)] are pulled down with at least two different baits, and FLNA and MYH9 are the only two prey proteins (outside of the contaminant ERP44) detected in all three of the TNC, ANGPTL2, and ANGPTL6 pull-downs. Therefore, the data strongly support the hypothesis that FRePs directly engage integrins and lead to the formation of activated adhesome complexes in primary human monocytes.

Another important feature of the identified adhesome components is their known proximity to the integrin receptor itself. There are five proteins which serve as direct, single component bridges physically linking integrins to the cytoskeleton: TLN1, FLNA, plectin (PLEC), tensin-1 (TNS1), and α -actinin-1 (ACTN1) [459]. Of those, TLN1, FLNA, and PLEC are enriched in activated adhesomes. I identified *all three* of these as high confidence interactors in Experiment 1. This is a crucially important point, because DTSSP is impermeable—the only covalent bond formed with the adhesome in this experiment is between the biotinylated bait and the integrin ectodomain. This means that the intracellular interactors must have survived stringent washing without the help of a cross-linker to stabilize them. *A priori*, this would imply that the *immediate* integrin adaptors would be most likely to come through the pull-down, and this is exactly what I observed. Furthermore, many of the other identified adhesome components, including the cytoskeleton itself, actin cytoplasmic 1 (ACTB), and known actin-binding proteins, including MYH9 and VCL, are at most two “steps” (i.e. bridging protein-protein interactions) away from the receptor itself in the Byron *et al.* dataset [451]. So, the network is not only enriched in activated adhesome components, it is enriched in *integrin proximal* activated adhesome components, quite consistent with the notion that the most

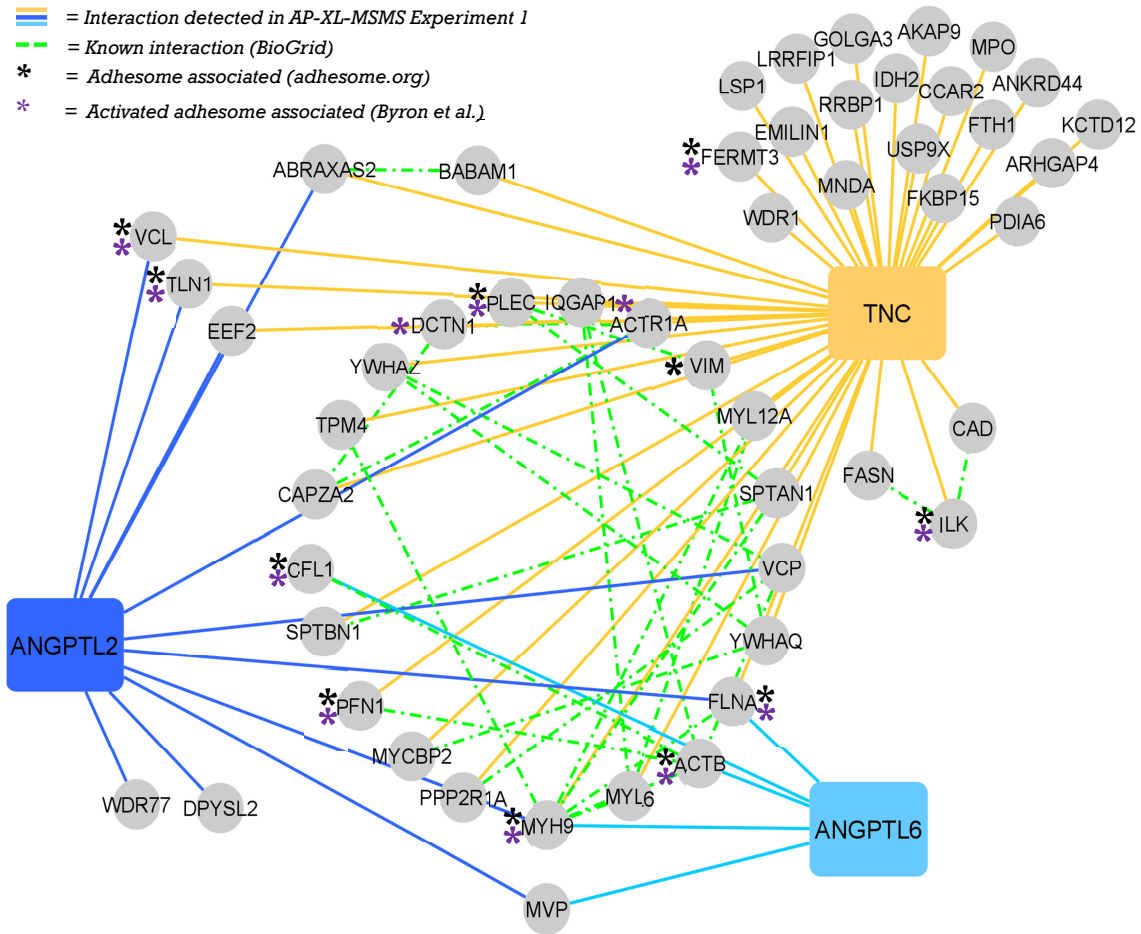


Figure 5.5 XL-AP-MSMS Experiment 1 Network Analysis. Network analysis of identified interactors of TNC, ANGPTL2, and ANGPTL6 with SP >0.95 and a log₂ fold-change >1.5. All preys are in grey circles. Interactions found with ANGPTL2, ANGPTL6, and/or TNC in Experiment 1 are marked with solid edges. Known interactions, annotated at least twice in BioGrid, are marked with green edges [458]. Proteins annotated in adhesome.org are marked with a black asterisk [459], while proteins identified as *activated* adhesome associated by Byron *et al.* are marked with a purple asterisk [451]. The network was constructed and displayed using Cytoscape [461].

strongly associated, nearby interacting proteins survived the pull-down despite not being cross-linked.

As a final check to ensure that my focus on the integrin aspect of the dataset was reasonable, I performed standard gene-ontology (GO) enrichment on the hit-list of interactors using the online tool PANTHER [462, 463], and pathway enrichment analysis using the Reactome platform [464]. The top GO molecular functions enriched in the dataset include cell adhesion molecule binding ($p = 2.4 \times 10^{-15}$), actin binding ($p = 2.6 \times 10^{-15}$), cytoskeletal protein binding ($p = 4.3 \times 10^{-14}$), structural

molecule activity ($p = 2.5 \times 10^{-6}$), and integrin binding ($p = 5.4 \times 10^{-5}$). The top over-represented Reactome pathways in the hit-list include smooth muscle contraction ($p = 1.57 \times 10^{-8}$), cell-cell communication ($p = 4.33 \times 10^{-6}$), RHO GTPase effectors ($p = 4.58 \times 10^{-6}$), and cell-extracellular matrix interactions ($p = 1.2 \times 10^{-4}$). Finally, the only two pathways enriched via PANTHER Pathway analysis were cytoskeletal regulation by Rho GTPase ($p = 8.13 \times 10^{-5}$) and integrin signalling pathway ($p = 1.8 \times 10^{-4}$). This unbiased enrichment of integrin- and cytoskeleton-associated terms and pathways validates that adhesome formation in response to the FRePs is truly the core conclusion of Experiment 1, and I have not cherry picked this feature from the network by manually comparing my data to a few papers from the literature.

The final, and somewhat disappointing, result was the lack of TLR4 or any directly TLR4-associated proteins in any of the pull-downs. Like integrin activation, TLR4 activation triggers the formation of large intracellular protein networks, including the Myddosome [52], and often requires the participation of extracellular co-receptors like MD-2. However, I detected none of these proteins in Experiment 1, and no GO terms relating to TLRs or innate immune activation appeared. I will return to the interactor leucine-rich repeat flightless-interacting protein 1 (LRRFIP1) in Section 5.6.2; however, excluding *perhaps* this protein, I found no evidence of any TLR4 binding during Experiment 1, and certainly no evidence that a TLR4-associated protein network was enriched alongside the adhesome network.

5.6 Full Experiment 2 with DTBP and DTSSP

5.6.1 Improvements from Experiment 1

Although the identification of integrin-associated interactors was exciting, there were a few features of the data from Experiment 1 that made me cautious. First, I did not actually detect the presence of any integrins—rather, I observed that intracellular, integrin-associated proteins were pulled-down. The strong and statistically confident enrichment of these intracellular interactors demonstrated that FReP-bound integrin adhesomes, cross-linked extracellularly but containing proximal intracellular interactors, survived the harsh washes and eluted upon DTSSP cleavage. However, to validate my hypothesis, I wanted to both detect the mediator integrin and validate the intracellular interactors using a cell-permeable cross-linker. Therefore, in Experiment 2 I compared the use of DTSSP with a new, cell permeable cross-linker:

DTBP. DTBP is, in fact, the exact same cross-linker used by Byron *et al.* to generate the reference set of adhesome-associated proteins, making it an excellent and well validated choice [451]. DTBP cross-linking additionally has the potential to cross-link and pull-down more distantly connected members of the intracellular adhesome network, rather than just the immediately proximal proteins identified in Experiment 1.

Secondly, I did not detect TLR4 or any known TLR4-associated signalling proteins in Experiment 1. Therefore, the experiment contained essentially no data to support or disprove my hypothesis regarding a TLR4-containing, FReP-activated receptorome. One possible explanation is that the FReP-bound TLR4 was internalized during the 30 min pre-incubation with the bait prior to the addition of DTSSP. This is consistent with data from Han *et al.* showing that, at least for LPS, TLR4 is activated and internalized within 5 to 15 minutes [465]. If this were true for the FRePs, with a 30 minute pre-incubation I would have missed the TLR4-ligation phase completely; the TLR4:FReP complex would be fully internalized by the time the DTSSP was added, completely shielding it from cross-linking and therefore detection. To address this problem in Experiment 2, I tested three different pre-incubation time points to try to rescue a fast-forming, fast-disappearing TLR4-containing complex.

Thirdly, in addition to not detecting TLR4 or the putative integrins themselves during Experiment 1, I did not confidently detect any known cell-surface receptors. Therefore, Dr. Kramer agreed to run the samples from Experiment 2 on a more sensitive mass spectrometer (Q ExactiveTM HF-XI from Thermo Fisher) to increase the likelihood of detecting low-abundance peptides. Furthermore, Dr. Kramer added 5% DMSO to the liquid chromatography solvents, a method reported by Hahne *et al.* to boost the signal intensity and the number of identified peptides in bottom-up proteomics experiments [443]. As an additional measure to increase sensitivity, Dr. Kramer looked at the raw data from Experiment 1 and detected significant detergent contamination in the mass spectra. Detergents act as ion suppressants in mass spectrometers and can broadly reduce sensitivity across the entire LC-MSMS run. The detergent contamination appeared despite multiple large washes with empty PBS between the washing and elution steps of the pull-down, suggesting that it would be difficult to remove the detergent with a simple protocol change. Therefore, in Experiment 2, I reversed my decision to include Triton X-100 and returned to the simple wash buffer containing just 0.9 M NaCl.

Fourthly, it is best practice in AP-MSMS experiments to perform at least three, and

preferably five, biological replicates of the same bait protein in parallel to ensure the reproducibility of the identified prey proteins. Therefore, instead of testing multiple FRePs, I used only LPS-free, biotinylated, full-length TNC as the bait across both cross-linkers and all three time points in Experiment 2. I made this choice based primarily on the observation that TNC clearly pulled down the most interactors in Experiment 1; as an orthogonal reason, TNC is the canonical FReP family member and the FReP with the clearest link to disease biology. Finally, I chose to incubate the lysate with the NeutrAvidinTM agarose overnight, rather than for just two hours, partially to maximize recovery and partially to simplify the experiment. The XL-AP-MSMS workflow used for Experiment 1 takes over 14 hours to complete on the first day; incubating with the resin overnight splits the work more manageably and makes the method less daunting.

5.6.2 Experimental Setup

With the limitations in Experiment 1 identified and a strategy for improving the experiment developed, I moved forward with executing Experiment 2 essentially as previously described. I obtained 12 plates with 70×10^6 monocytes per plate from Anna Marzeda, enough for six pull-downs and six matched controls. I added LPS-free, biotinylated full-length TNC to six dishes at a final concentration of 2 μ M, while a matching amount of empty buffer was added to six negative controls. However, this time I pre-incubated TNC with the cells for either 30 sec, 10 min, or 30 min before the addition of the cross-linker. I used a final concentration of 1 mM DTSSP as before; for the permeable cross-linker DTBP, I used 3 mM, exactly the concentration used by Byron *et al.* [451]. Again, I allowed the cross-linking reaction to proceed for 30 minutes (no matter how long the pre-incubation lasted) before quenching with excess Tris, harvesting the cells, and pulling down on NeutrAvidinTM agarose overnight. Following stringent washes with a wash buffer containing only 0.9 M NaCl, I eluted the samples. For DTSSP, the elution buffer contained 50 mM DTT; for DTBP, the elution buffer contained 100 mM DTT. As recommended for DTBP, which has a harder-to-cleave disulfide bond in the middle, the beads were incubated for 30 min at 50 °C with the elution buffer to promote complete cleavage of the cross-linker. I also added this incubation step to the DTSSP samples to ensure all samples were treated equivalently; the NeutrAvidinTM:biotin interaction is stable even up to the point of boiling, so this heated elution strategy still only released the interacting proteins. The eluate samples were alkylated, digested, and desalted as before and sent to Dr. Kramer, who collected MSMS data on each sample in

technical duplicate using the top-end Q ExactiveTM HF-XI mass spectrometer. The raw SPC were averaged across the two technical replicates and used as the input to SAINT for identification of interactors, exactly as in Experiment 1. As discussed further in Section 5.6.3 below, I looked for proteins that were confidently enriched both across and within the three time points using SAINT.

5.6.3 Bulk Analysis Validates Integrin Hypothesis

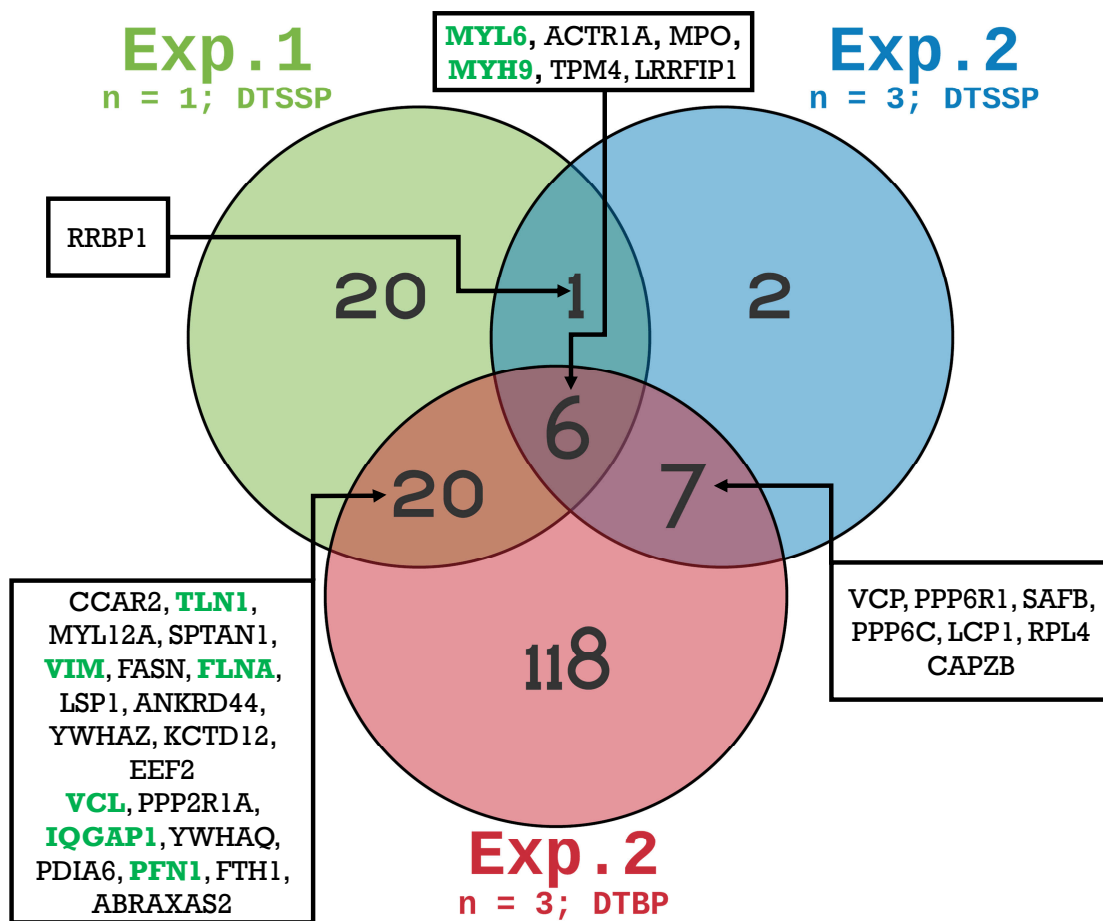


Figure 5.6 TNC Prey Proteins Across All XL-AP-MSMS Experiments. TNC-dependent prey proteins confidently detected in Experiment 1 and/or Experiment 2 with SP > 0.95. The proteins in the overlapping regions are specified; any protein known to be associated with adhesomes (as annotated in Figure 5.5) is displayed in green text.

To get a broad picture of the data, I first looked for prey proteins broadly enriched (SP > 0.95) within each cross-linker. With this analysis, each time point (along with the matched negative control) is treated as a single biological replicate by

SAINT, meaning that proteins consistently enriched across all three time points will be identified as confident interactors. This approach provides a top-down view of the data in order to compare broad differences between the cross-linkers used in Experiment 2 as well as the differences between Experiment 1 and Experiment 2.

Figure 5.6 presents a Venn diagram visualizing the overlap between the TNC-dependent prey proteins enriched in Experiment 1, Experiment 2-DTSSP, and Experiment 2-DTBP. There are five conclusions to be drawn from this data. First, note that more TNC interactors (47) were identified in Experiment 1 than in Experiment 2-DTSSP (16). This could be due to the omission of detergent in the Experiment 2 wash buffer or stochastic differences between the cells used; manual inspection of the raw peptide counts confirmed that there were simply less peptides detected in Experiment 2-DTSSP than in Experiment 1. Second, Experiment 2-DTBP identified by far the most prey proteins (131), consistent with DTBP's permeability and my expectations. DTBP, after crossing the membrane, can cross-link farther downstream the TNC-induced network, covalently stabilizing these more distant interactors and allowing them to be pulled down. As a result, there were 118 preys detected in Experiment 2-DTBP that were not detected in either DTSSP pull-down. Third, the increased number of interactors in the DTBP pull-down supports my previous hypothesis that the DTSSP interactors represent a strongly associated subset of the most receptor-proximal proteins and validates the use of a permeable cross-linker to replicate Experiment 1. Since there were so many unique interactors identified in the DTBP pull-down, I provide a summary of the top 70 hits in Figure 5.7. Fourth, there was good agreement between Experiment 1, Experiment 2-DTSSP, and Experiment 2-DTBP: 57% of the preys identified in Experiment 1 were validated in Experiment 2, and 34 different preys were identified in at least two of the three datasets. These overlapping, reproducible preys are specifically enumerated in the inset black boxes in Figure 5.6.

The fifth, and most important, conclusion is that Experiment 2 conclusively reproduced the enrichment of integrin-associated adhesome components during Experiment 1. In Figure 5.6, protein names in bold green font are the TNC-dependent adhesome components identified in Experiment 1 and validated in Experiment 2. These included the immediately integrin-proximal adaptors TLN1 and FLNA, the scaffolding protein Ras GTPase-activating-like protein (IQGAP1), the actin-binding regulators profilin-1 (PFN1) and VCL, and the myosins light polypeptide 6 (MYL6) and MYH9, all of which are reported to be up-regulated at activated integrin adhesomes by Byron *et al.* In addition to MYL6 and MYH9, the cytoskeletal pro-

Protein	log(FC)	SP	Sum SPC + TNC	Sum SPC Neg. Control
VCP	14.14	1	389	0
MYH9	13.22	1	384	0
FLNA	8.67	1	228	0
TLN1	4.15	1	97	0
ACTN4	4.11	1	83	0
LCP1	3.41	1	70	0
MYL6	3.12	1	72	0
PRKDC	3.11	1	57	0
TPM4	2.96	1	60	0
VCL	2.25	1	34	0
TPP2	2.25	1	36	0
ACTN1	2.24	1	32	0
DDB1	2.14	1	28	0
AHNAK	2.04	1	35	0
EEF2	2	1	30	0
HNRNPU	1.94	1	29	0
MPO	1.88	1	27	0
ESYT1	1.87	1	23	0
SNRPD2	1.65	1	20	0
PPP6C	2.59	0.99	44	0
HNRNPM	2.33	0.99	41	0
YWHAZ	2.2	0.99	34	0
RNH1	2.16	0.99	29	0
DDX17	1.86	0.99	26	0
MSN	1.86	0.99	26	0
LRRFIP1	1.85	0.99	24	0
DPYSL2	1.83	0.99	25	0
PRPF19	1.73	0.99	23	0
MVP	1.7	0.99	20	0
IQGAP1	1.64	0.99	21	0
PAICS	1.61	0.99	15	0
FASN	1.61	0.99	15	0
HNRNPA3	1.58	0.99	19	0
PCCA	1.48	0.99	15	0
HSP90AA1	1.48	0.99	15	0

Protein	log(FC)	SP	Sum SPC + TNC	Sum SPC Neg. Control
ARHGEF2	1.48	0.99	15	0
SAFB	1.45	0.99	14	0
STIP1	1.45	0.99	12	0
P4HB	1.45	0.99	14	0
HCFC1	1.43	0.99	13	0
TKT	1.4	0.99	12	0
MYO1G	1.32	0.99	9	0
HCLS1	1.84	0.98	27	0
ACTR3	1.81	0.98	26	0
SRSF1	1.8	0.98	24	0
ALDOA	1.77	0.98	21	0
MDH2	1.7	0.98	20	0
KHSRP	1.68	0.98	21	0
PCBP1	1.65	0.98	20	0
HNRNPH1	1.63	0.98	21	0
ACAA2	1.61	0.98	18	0
ACTR1A	1.58	0.98	19	0
HNRNPK	1.56	0.98	18	0
SAMHD1	1.55	0.98	16	0
SRSF3	1.53	0.98	17	0
ACTR2	1.5	0.98	14	0
DHX9	1.49	0.98	17	0
PPP2R1A	1.45	0.98	14	0
PDIA3	1.44	0.98	15	0
CAPN2	1.39	0.98	10	0
TRIM28	1.38	0.98	11	0
ALDH16A1	1.38	0.98	13	0
EIF3A	1.36	0.98	12	0
SART1	1.35	0.98	12	0
GSR	1.3	0.98	8	0
ATP6V1A	1.25	0.98	8	0
VIM	2.2	0.97	139	4
CAPZB	1.78	0.97	25	0
RPL4	1.73	0.97	23	0
PCMT1	1.67	0.97	19	0

Figure 5.7 Top 70 Interactors in Experiment 2-DTBP. List of top 70 proteins identified in the pooled analysis ($n = 3$) of the Experiment 2-DTBP pull-downs. In order, the $\log_2$ -fold change, SAINT calculated SP score, and raw sPC summed from the three +TNC and three negative control pull-downs are provided.

teins alpha-actinin (ACTR1A) and tropomyosin alpha-4 chain (TPM4), calcium-binding defense protein myeloperoxidase (MPO), and the previously mentioned LRRFIP1 were identified as TNC interactors in all three experiments. The use of DTBP and the more sensitive mass spectrometer also allowed me to detect peptides from three-different integrins: integrin α M (ITGAM; 3 total peptides), integrin β 2 (ITGB2; two peptides), and integrin β 3 (ITGB3; two peptides). While these low peptide counts only resulted in a statistically significant enrichment of ITGB2 at the 30 second time point ($SP = 0.98$; see below for discussion of time points), there was not a single integrin-derived peptide in any of the six negative control pull-downs. Therefore, Experiment 2 not only reproduced Experiment 1, it also allowed me to speculatively hypothesize, based on previously reported integrin super-family dimerization preferences [466], that the integrin heterodimers α M β 2 and α V β 3 could be

the (co-)receptors for TNC.

Integrin $\alpha V\beta 3$ is the fibrinogen and vitronectin receptor which typically binds to RGD-containing peptides; however, it can also bind to proteins without this motif [467]. Furthermore, FGG (the first FReP family member to have its structure solved, see Section 2.7) is an $\alpha V\beta 3$ ligand [468], and it has been reported that TNC can also bind to integrin $\alpha V\beta 3$ via the peptide Trp2145–Gly2714 [399]. This region contains the pseudo-canonical RGD motif “RYGD” and is present in a loop-sheet-loop motif in the P-subdomain of TNC (see structure and sequence in Figure 2.10). My detection of ITGB3, which preferentially associates only with integrin αV (ITGAV), as a TNC interactor in Experiment 2 represents a biophysical validation of these previous reports that TNC binds to $\alpha V\beta 3$. Integrin $\alpha V\beta 3$ is also known to associate with and regulate TLRs, suggesting that it could be involved in a TLR4-associated receptorome. Integrin $\alpha V\beta 3$ ligation with an RGD-containing protein inhibits LPS-induced cytokine production, MAPK activation, and monocyte migration in both *in vitro* and *in vivo* assays [469]. Conversely, in a mouse model of lung sepsis, Chen *et al.* showed that LPS-activated TLR4 drives increased expression of integrin $\alpha V\beta 3$, which in turn enhances the release of the pro-inflammatory cytokine TNF α . Similarly, Chen *et al.* also demonstrated that TLR4 and integrin $\beta 3$ can co-localize after LPS stimulation, and that deletion of integrin $\beta 3$ suppresses LPS-induced cytokine release in murine peritoneal macrophages [470]. Antonov *et al.* also suggested that integrin $\alpha V\beta 3$ drives inflammation, reporting that integrin $\alpha V\beta 3$ maintains chronic inflammatory processes and synergizes with LPS-induced *in vitro* activation of human monocytes and macrophages [471]. Partially alongside TLR2, integrin $\alpha V\beta 3$ can also itself act as a non-canonical PRR and trigger an anti-viral innate immune response [472, 473].

Based on multiple PubMed searches, TNC has never been discussed in a paper alongside integrin $\alpha M\beta 2$, meaning this interaction could be novel; yet again, however, there is some evidence in the literature that $\alpha M\beta 2$ is a reasonable candidate for a TNC co-receptor. Integrin $\alpha M\beta 2$ is a phagocytic receptor also known as complement receptor 3 (CR3), macrophage-1 antigen (Mac-1), or CD11b/CD18. Additionally, based on proteomics data in pax-db, both ITGAM and ITGB2 are expressed at their highest level in monocytes, the cells used throughout Chapter 5. Signalling through $\alpha M\beta 2$ (and, indeed, $\alpha V\beta 3$) is driven primarily through the adaptors TLN1 [474], which was significantly enriched in both Experiment 1 and Experiment 2, and Kindlin-3 *aka* FERMT3, which was significantly enriched in Experiment 1. Just like $\alpha V\beta 3$, FGG has been reported to bind directly to $\alpha M\beta 2$, likely using the cationic

peptide 190–GWTVFQKRLDGSV [475]. During non-classical signalling, integrin $\alpha M\beta 2$ can cluster with other receptors and stimulate the formation of membrane ruffles in LPS-simulated macrophages [476], and ITGAM has been reported to positively regulate LPS-activation of TLR4 and to be essential for endocytosis of TLR4 in dendritic cells [477]. $\beta 2$ integrins can also directly stimulate cytokine expression in macrophages and neutrophils, and they are key regulators of leukocyte recruitment at sites of infection [478, 479].

Therefore, I hypothesize that integrin $\alpha M\beta 2$ and/or $\alpha V\beta 3$ could be direct TNC-interactors that might also contribute to a TLR4-associated receptorome. Another FReP, FGG, is known to directly bind to both heterodimers, and TNC has itself been shown to bind to $\alpha V\beta 3$. Furthermore, both integrin heterodimers have clear innate immune roles, either alongside or independent of TLRs, and the bulk of the evidence seems to suggest that they are pro-inflammatory in myeloid cells and leukocytes. Although this remains a speculative hypothesis, the formation of activated integrin adhesomes in the presence of TNC, via $\alpha M\beta 2$ and/or $\alpha V\beta 3$, is clear and noteworthy, and the biological consequences of this integrin ligation event merit further inquiry.

5.6.4 Time-Dependent Analysis does not Reveal TLR4

Like in Experiment 1, I detected no enrichment of TLR4 using the pooled time points in Experiment 2. However, to explore the potential of an “early” TLR4 ligation and subsequent internalization, I processed each time point independently. I treated each cross-linker/time point combination as a single biological replicate, and I used all six negative control pull-downs as the reference for SAINT to detect enriched interactors. Using all of the negative control data increases the robustness of SAINT, provides a stringent evaluation of background contaminants (in that detection of the protein in *any* negative control reduces the SP score), and is reasonable given that all negative control cells were treated with the same volume of empty buffer prior to cross-linking and pull-down. Because SAINT assigns priority to prey proteins which appear reproducibly across all biological replicates, running the analysis with single time point pull-downs results in a larger number of proteins being identified as “significantly” enriched relative to the bulk analysis presented in Section 5.6.2. For example, a protein with SPC of eight, zero, and five across the time points would not be counted as significant in a bulk analysis due to the complete lack of peptides in the 2nd time point. However, in a single-replicate analysis, the protein would be scored as significantly enriched in the time points with SPC of eight and

five. Therefore, because I only had a single replicate for each time point, the hits in this section are notably less reliable and more speculative than the previous section discussing the pooled data. However, digging into these single replicate time points is the only way to look for the transient enrichment of TLR4, so it is a worthwhile trade-off.

With this caveat in mind, the results of this time point analysis are presented as a Venn diagram in Figure 5.8. I detected 68 proteins that were enriched in at least one of the DTSSP time points (Figure 5.8A), and 444 proteins that were enriched in at least one of the DTBP time points (Figure 5.8B). However, not a single one of these proteins was TLR4. Note that this is not because TLR4 was wholly absent from the dataset, a problem which could suggest that the method is inadequately sensitive for detection of a low abundance protein. Rather, there were six TLR4 peptides in the DTSSP negative controls and only one in the TNC pull-downs, while there were seven TLR4 peptides in the DTBP negative controls and ten in the TNC pull-downs. Therefore, TLR4 was detected but not significantly enriched at any time point with either cross-linker. As in Experiment 1, I cannot draw meaningful conclusions about a TLR4-associated receptorome with the data from Experiment 2, despite making every effort with the methodological changes discussed above to increase sensitivity and identify a suitable time point.

However, even though TLR4 was not identified, the time-dependent enrichment of certain proteins does suggest that there are distinct “early” and “late” phases within the thirty minute window following stimulation with TNC in human monocytes. In the DTSSP pull-down, there were five proteins identified only in the 30 sec “early” time point (listed in the green box in Figure 5.8A) and 24 proteins identified only in the 30 min “late” time point. The single most highly enriched GO term at the late time point was “RNA processing” ($n = 11$; $p = 6.4 \times 10^{-9}$), and the second highest was “gene expression” ($n = 15$, $p = 6.2 \times 10^{-9}$). This means that more than 50% of the late proteins were involved in transcriptional processes, suggesting that by the 30 min time point, the signal from TNC was fully integrated and the entire downstream signalling cascade was formed. This is borne out by comparison to the DTBP dataset, in which 105 proteins were specifically enriched in the early time point and 94 proteins were specifically enriched in the late time point. GO enrichment of the early time point proteins, provided in detail in Figure 5.8B, yielded terms including “cellular response to cytokine stimulus”, “innate immune response”, and “establishment of localization in cell”. However, at the late time point, GO enrichment again

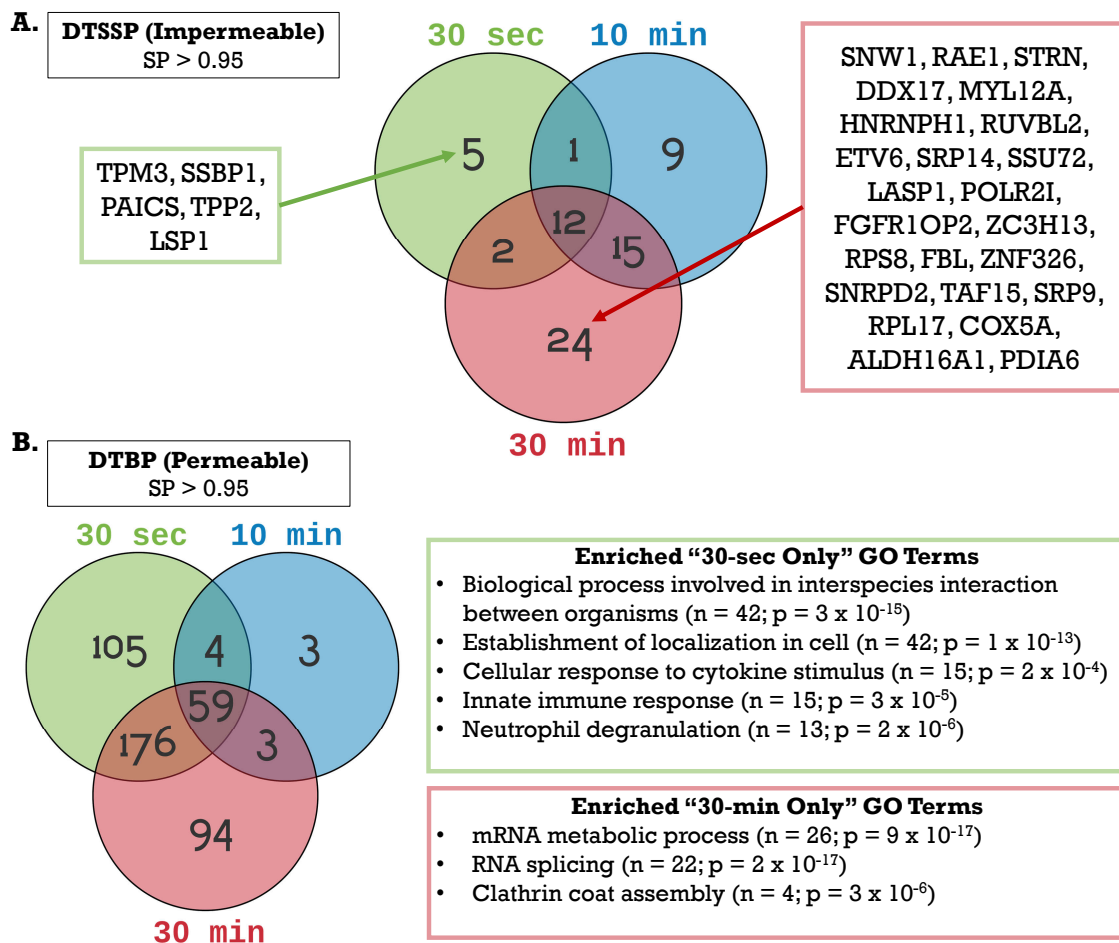


Figure 5.8 Time-Dependent Prey Proteins. **A.** Identified interactors (SP > 0.95) across the three DTSSP time points. Proteins unique to the 30 sec time point in the green box, proteins unique to 30 min time point in the red box. **B.** Same format as in A., except for the DTBP pull-downs. Significantly enriched biological process GO terms from the unique “early” 30 sec and “late” 30 min proteins are provided. GO enrichment was performed with PANTHER [462, 463].

picked out more transcriptional terms, including “mRNA metabolic processes” and “regulation of RNA splicing”.

To further emphasize this kinetic difference, in Figure 5.9 I compare the specific genes enriched at the 30 second DTBP time point under the GO term “innate immune response” against the genes enriched at the 30 min DTBP time point under the GO term “mRNA metabolic processes”. The early interactors (in Figure 5.9A) include multiple proteins known to be involved with TLR4-triggered immune responses. First, PRKCD, a member of the protein kinase C (PKC) family also known as PKC- δ , is a known downstream regulator of many TLRs, including TLR4. LPS (and other

A.

Innate immune response (n = 15; p = 3 x 10 ⁻⁵)		
UniProt	Gene ID	Gene Name
P12724	RNASE3	Eosinophil cationic protein
P23246	SFPQ	Splicing factor, proline- and glutamine-rich
P19878	NCF2	Neutrophil cytosol factor 2
P07814	EPRS	Bifunctional glutamate/proline-tRNA ligase
P04839	CYBB	Cytochrome b-245 heavy chain
P13010	KRCC5	X-ray repair cross-complementing protein 5
P10321	HLA-C	HLA class I histocompatibility antigen, C alpha chain
Q05655	PRKCD	Protein kinase C delta type
Q9UM07	PADI4	Protein-arginine deiminase type-4
P28062	PSMB8	Proteasome subunit beta type-8
Q15233	NONO	Non-POU domain-containing octamer-binding protein
P17931	LGALS3	Galectin-3
P02671	FGA	Fibrinogen alpha chain
Q9Y2Y1	POLR3K	DNA-directed RNA polymerase III subunit RPC10
P20592	MX2	Interferon-induced GTP-binding protein Mx2

B.

mRNA metabolic processes (n = 26; p = 9 x 10 ⁻¹⁷)		
UniProt	Gene ID	Gene Name
O95400	CD2BP2	CD2 antigen cytoplasmic tail-binding protein 2
Q99459	CDC5L	Cell division cycle 5-like protein
P62314	SNRPD1	Small nuclear ribonucleoprotein Sm D1
Q9UKV3	ACIN1	Apoptotic chromatin condensation inducer in the nucleus
O43143	DHX15	Pre-mRNA-splicing factor ATP-dependent RNA helicase DHX15
O43172	PRPF4	U4/U6 small nuclear ribonucleoprotein Prp4
Q13315	ATM	Serine-protein kinase ATM
Q15637	SF1	Splicing factor 1
Q96FV9	THOC1	THO complex subunit 1
P62316	SNRPD1	Small nuclear ribonucleoprotein Sm D2
Q13285	SF1	Steroidogenic factor 1
Q99590	SCAF11	Protein SCAF11
Q9Y5S9	RBM8A	RNA-binding protein 8A
Q9Y3F4	STRAP	Serine-threonine kinase receptor-associated protein
Q01130	SRSF2	Serine/arginine-rich splicing factor 2
P49756	RBM25	RNA-binding protein 25;RBM25;ortholog
P67775	PPP2CA	Serine/threonine-protein phosphatase 2A catalytic subunit alpha isoform
P63151	PPP2R2A	Serine/threonine-protein phosphatase 2A 55 kDa regulatory subunit B alpha isoform
O75934	BCAS2	Pre-mRNA-splicing factor SPF27
Q9P013	CWC15	Spliceosome-associated protein CWC15 homolog
P19387	POLR2C	DNA-directed RNA polymerase II subunit RPB3
P62310	LSM3	U6 snRNA-associated Sm-like protein Lsm3
Q8IX12	CCAR1	Cell division cycle and apoptosis regulator protein 1
P15880	RPS2	40S ribosomal protein S2
P52597	HNRNPF	Heterogeneous nuclear ribonucleoprotein F
Q07666	KHDRBS1	KH domain-containing, RNA-binding, signal transduction-associated protein 1

Figure 5.9 Enriched Early and Late Communities. **A.**Prey proteins enriched at the early DTBP time point and classified under the GO term “innate immune response”. **B.**Prey proteins enriched at the late DTBP time point and classified under the GO term “mRNA metabolic processes”. These groups come from the GO enrichment presented in Figure 5.8.

TLR ligands) broadly activate PKC family members in monocytes and lymphocytes, and PKC inhibitors reduce the activation of NF κ B upon LPS stimulation [480]. In fact, PRKCD is known to directly interact with the TLR4 adaptor TIRAP, and downregulation of PRKCD sharply reduces downstream activation of MAPK and NF κ B following stimulation with either TLR2 or TLR4 ligands [481]. Secondly, LGALS3, also known as galectin-3, is a β -galactoside-binding lectin that also binds directly to LPS, and it has diverse and sometime opposing roles in the innate immune response. LGALS3 is known to downregulate the LPS-induced pro-inflammatory response in immune cells [482]; however, LGALS3 also increases inflammation in arthritic joints, likely by activating the release of IL-6 downstream of TLR2, TLR3, and TLR4 [483]. Intriguingly, secreted LGALS3 has been reported to bind directly to TLR4 with a biophysically determined (via MST) K_D of 1 μ M, an interaction which

activates microglia during neuroinflammation and is required to elicit the maximal response to LPS [484]. Next, enrichment of HLA-C, part of the MHC class I family of antigen presenters, along with the proteasome subunit PSMB8 suggests that antigen presentation, a key link between the innate and adaptive immune responses, has been up-regulated at the early time point [485]. This could also be TLR-dependent: MHC class II molecules, acting in a non-canonical function, are direct adaptors for TLR3, TLR4, and TLR9 and link TLRs to Bruton's tyrosine kinase (BTK) via Myd88 and TRIF, resulting in the enhanced release of pro-inflammatory cytokines [486]. Therefore, the early time point interactors from Experiment 2 are consistent with the activation of an innate immune response almost immediately upon addition of TNC, which is consistent with expectations and with the cytokine release I observed in Figure 5.2D. Furthermore, while I did not directly detect the enrichment of TLR4, a handful of these innate immune interactors are known to be TLR4-associated in relevant models. Broadly, then, these MSMS results suggest a model whereby TNC activates a TLR4-dependent signalling mechanism, but *not* via a direct ligation of TLR4 itself.

At the 30 minute time point, proteins involved in mRNA metabolic processes are enriched (see Figure 5.9B), suggesting that the signal from TNC is integrated into a transcriptional response during the pre-incubation step. These proteins include RNA binders and regulators, ribosomal subunits, transcriptional regulators, splicing factors, and even members of the spliceosome. CCAR1 is a known regulator of the Mediator megacomplex that controls transcription by RNA polymerase II [487], and POL2RC is also enriched at this time point. The up-regulation of transcriptional and splicing proteins implies that DTBP cross-linking stabilizes a TNC-associated signal transduction network all the way into the nucleus, although the involvement of non-specific cytoskeleton-associated proteins complicates deconvolution of the specific cascade. Additionally, PPP2CA (aka RP-C) and PPP2R2A are members of the serine-threonine protein phosphatase 2 complex, which is a tumor suppressor that downregulates signalling by dephosphorylating active kinases [486]; this suggests that the monocytes might already be dampening their pro-inflammatory response to TNC after just 30 minutes.

Taken together, the time point analyses imply a model whereby, almost immediately upon addition, TNC binds to cell-surface receptors (including integrins, a TLR4-involved complex, and unknown other components) and triggers an innate immune response; however, this response is not fully integrated or communicated to downstream transcriptional regulators until at least 30 minutes later. As mentioned

in Section 5.6.3, it was only in the 30 sec time point that ITGB2 was significantly enriched, consistent with a model wherein receptor ligation occurs nearly instantaneously upon TNC addition. Based on this data and the LPS time course from Han *et al.* discussed previously, I believe that similar MSMS experiments with DAMPs should be conducted with a very short pre-incubation, rather than with the standard 30 min pre-incubation often used and which I deployed in Experiment 1. These data provide an impetus for repeating the earlier time point with a full set (at least five) of biological replicates for both TNC and the other FRePs, an experiment I had planned but which was stopped by COVID-19. If additional replicates were completed at the 30-sec time point, providing more robust statistical confidence in the identified preys, Experiment 2 implies that FReP-specific receptoromes could be deeply probed with my XL-AP-MSMS method.

As a final point, I identified over 440 “interactors” during Experiment 2-DTBP. However, because TNC ligated cytoskeleton-bound integrins and the sample was cross-linked with the permeable DTBP, it is impossible to distinguish between proteins involved in a direct signalling network downstream of TNC and proteins which are simply non-specifically associated in some fashion with the cytoskeleton. The covalent linking of neighboring proteins is both the strength and the weakness of the method, and it explains why the number of “interactors” is an order of magnitude higher than would be expected in a standard IP-MSMS experiment. This fact is why I limited my analysis to generic conclusions, focused primarily on the integrin hypothesis and qualitative differences between the time points, rather than drilling down to focus on individual identified proteins. This kind of MSMS experiment should be treated strictly as a hypothesis generation exercise, especially in the absence of the *a priori* hypothesized receptor (in this case TLR4) which could anchor a downstream network analysis based on prior knowledge. However, in line with the goal of the experiment to clarify the uncharacterized FReP-associated receptorome, I abandon this high-level viewpoint in the following Section 5.6.5 and speculate about specific potential TNC co-receptors.

5.6.5 Speculating on TNC Co-Receptors and Adaptors

Despite the potential for false-positives due to the non-specific covalent capture of “interactors” that are simply connected to the cytoskeleton, in the spirit of hypothesis generation I discuss five different putative co-receptors or adaptors in this section: LRRFIP1, MPO, KCTD12, APOBR, and LRP1. These proteins are all known to

be integral membrane proteins, secreted, or directly membrane proximal, and they were detected in at least one of the analyses presented previously above. These proteins require careful validation, and I provide recommendations for downstream experiments in Section 5.7 which could provide additional evidence to support the role of these proteins in FReP activation of monocytes.

LRRFIP1 (also known as Flap-1, Fli-I, or GCF2) is one of the six proteins (see Figure 5.6) that was pulled down by TNC in all three of Experiment 1, Experiment 2-DTBP, and Experiment 2-DTSSP. LRRFIP1 is a nucleotide binding protein that is canonically viewed as a nuclear transcriptional repressor for genes including $\text{TNF}\alpha$, EGFR, and platelet-derived growth factor (PDGF) [488, 489]. However, more recent evidence shows that cytosolic LRRFIP1, in conjunction with the related LRRFIP2, is a direct Myd88 interactor which activates TLR4 signalling in response to LPS by competing for the Myddosome inhibitor FLII. [490]. Furthermore, LRRFIP1 has been shown to co-activate the release of $\text{IFN}\beta$ by acting as a direct, intracellular dsDNA sensor during *L monocytogenes* and VSV infection [491]. This effect stems from the activation of β -catenin and subsequent induction of the key $\text{IFN}\beta$ transcriptional regulator IRF3, which is known to act downstream of many PRRs including TLRs [492]. Therefore, LRRFIP1 can either directly or indirectly activate a pro-inflammatory response to PAMPs. Even though LRRFIP1 is cytosolic, it was enriched in the extracellularly cross-linked DTSSP samples; this suggests that LRRFIP1 is not a direct TNC binder, but rather a key, immediately proximal downstream regulator (much like FLNA or TLN1 just downstream of the integrins). LRRFIP1 enrichment could serve as mediator downstream of TLR4, as reported for LRRFIP1 regulation of LPS responses, or an orthogonal receptor.

Along with LRRFIP1, MPO was the other non-cytoskeletal protein enriched in all three pull-downs. MPO is a heme-containing peroxidase that, in either the phagolysosomal compartment or when secreted by neutrophils and monocytes, catalyzes the formation of antibacterial reactive oxygen species [493]. Because of its upregulation at sites of inflammation, MPO has been implicated in the progression of RA, pulmonary inflammation, and skin inflammation; however, MPO can be both pro- and anti-inflammatory depending on the model system. MPO inhibitors have been actively developed, primarily due to the observation that MPO-derived oxidants contribute to tissue damage and increased MPO levels predict major cardiac events [494]. In fact, MPO is a well recognized and highly accurate biomarker for inflammation in ischemic heart disease and other coronary diseases [495]. Although most previous work has focused on MPO metabolites or anti-MPO auto-antibodies,

the MPO protein itself can also have immune regulatory functions [496]. In neutrophils, MPO has twice been reported to bind directly to $\beta 2$ integrins and act as a survival signal to prolong inflammation [497, 498]. Therefore, a reasonable hypothesis is that extracellular MPO acts as a co-receptor for TNC along the integrin axis, either as a direct TNC binder or as a $\beta 2$ integrin adaptor.

KCTD12, *aka* potassium channel tetramerization domain containing protein 12, is a relatively understudied (only 62 publications in PubMed), cytosolic, membrane-associated protein that was enriched in both Experiment 1 and Experiment 2-DTBP. At the N-terminus, KCTD12 contains a T1 domain that controls receptor oligomerization in other KCTD family potassium channels [499]; however, KCTD12 is not itself a transmembrane channel. In the brain, KCTD12 is a key auxiliary activator of the G-protein-coupled receptor GABA_B that increases agonist potency and accelerates signalling [500]. KCTD12 has been reported as an oncoprotein that facilitates the G2/M transition and promotes tumorigenesis by directly interacting with cyclin-dependent kinase 1 (CDK1) and activating Aurora kinase A [501]. There are no reports of immune-modulating activities of KCTD12, but, according to pax-db, KCTD12 is the 28th most highly expressed protein in human monocytes. Therefore, like LRRFIP1, it is likely that KCTD12 is a receptor-proximal adaptor in the cytosol that regulates the response to TNC. The potential involvement of GPCRs in TNC detection by monocytes is intriguing, especially given KCTD12's direct association with GABA_B. However, given the lack of immune-relevant data on KCTD12, it is impossible to speculate as to whether KCTD12 is acting along the GABA_B, CDK1, or a completely unknown axis in this context.

APOBR, also known as apolipoprotein receptor B or apoB48R, is an ApoE-independent receptor that binds to apolipoprotein B48 (ApoB48) in either dietary triglyceride-rich lipoproteins (TRLs) or very low density lipoproteins (VLDLs) [502]. APOBR does not contain a transmembrane domain and is likely localized to the cell-surface via interaction with an as yet unknown adaptor protein of approximately 35 kDa [503]. APOBR is expressed highly on monocytes according to the Human Protein Atlas and was statistically significantly enriched in the Experiment 2-DTBP pull-down. Additionally, in Experiment 1, there were 38 APOBR peptides in the TNC pull-down and only nine in the negative control, a more-than 3-fold increase that, however, did not reach statistical significance (SP = 0.58). ApoB48, the ligand for APOBR, is the N-terminal portion of the larger apolipoprotein B100 (apoB100, often called simply ApoB). ApoB48 naturally occurs in circulating chylomicrons and

TRLs and, upon internalization, triggers macrophage foam cell formation via binding to APOBR [504]. Full-length ApoB100 can bind to the classic super-family of low-density lipoprotein receptors (LDLRs), but ApoB48 cannot, explaining the need for the orthogonal receptor APOBR. APOBR has no known orthologs in the human genome based on a BLAST analysis, and it is greatly understudied relative to the ApoE- and ApoB100-binding LDLR super-family (which includes LDLR, VLDLR, ApoER2, LRP1, LRP2, and LRP6) [505]. A PubMed search with “apoB48R OR APOBR” yields only 24 papers, while the query “LDLR” retrieves over 5,800. Interestingly, according to UniProtKB (P04114) the LDLR receptor binding regions in ApoB100 are cationic (3174–KAQYKKNKHRH and 3386–RLTRKRGLK), just like the mapped TLR4-activating region on TNC (see Section 2.10). There has also been many reports of cross-talk between TLR4 and lipoprotein accumulation in macrophages, including a direct LDL:CD14:MD-2:TLR4 complex which activates cytoskeletal rearrangements and a dependence on TLR4 for macrophage mast cell formation [506]. Based on the evidence in the literature, then, a reasonable hypothesis is that APOBR acts as a direct co-receptor which binds to and presents TNC to TLR4.

In a similar vein, LRP1, also known as prolow-density lipoprotein receptor-related protein 1, APOER, or CD91, was enriched at 30 minutes in Experiment 2-DTBP. LRP1 was a member of the enriched GO term “receptor metabolic process” ($p = 1.5 \times 10^{-4}$), which also included the clathrin proteins CLTA and CLTC. LRP1 contains a massive, multi-functional 515 kDa “ α -chain” ectodomain coupled via a single transmembrane domain to an 85 kDa intracellular “ β -chain” [507]. Quoting directly from a 2014 review, “Among the LDLR family members, LRP1 is the most studied receptor...LRP1 is a large multi-functional receptor that regulates the endocytosis of diverse ligands and transduces several cell signalling pathways by coupling with other cell-surface receptors.” [508]. As a known scavenger receptor, LRP1 binds to diverse ligands including canonical ApoE, matrix metalloproteases, growth factors including PDGF and TGF β , matrix proteins like fibronectin, and other proteins like calreticulin and complement C3; these ligands are subsequently internalized [509]. LRP1 is known to mediate interactions with amyloid and tau during neurodegenerative conditions, and the direct interaction between tau and LRP1 involves key lysine residues within tau [510]. Activated LRP1 is also a known negative regulator of LPS-induced NF- κ B signalling and has been reported to rely on PKC α as a downstream adaptor [509]. Therefore, one hypothesis is that LRP1 is involved in dampening the inflammatory response to TNC, perhaps explaining its enrichment only at the late time point. However, based on the enrichment of both APOBR and

LRP1, two lipoprotein binders which seem to respond to cationic peptides and may act as scavenger receptors for multiple ligands, my favorite hypothesis is that either APOBR and/or LRP1 are a co-receptor for TNC. Along with ITGB2, these two proteins were the only interactors I detected which are reported to act as bona fide transmembrane receptors, and the established cross-talk between lipid metabolism and TLR4 makes them interesting candidates.

5.7 Chapter 5 Summary and Next Steps

In conclusion, motivated by the lack of a direct interaction between TLR4 and TNC (explored in Chapter 4), I optimized and deployed an XL-AP-MSMS method to explore the TNC interactome in primary human monocytes in an unbiased fashion using proteomics. Using two different cross-linkers and based on a stringent enrichment analysis using the gold-standard SAINT algorithm, I demonstrated that TNC, and perhaps ANGPTL2 and ANGPTL6, trigger the formation of activated integrin adhesomes. Based on the data from Experiment 2-DTBP, I hypothesized that this is due to a direct interaction between TNC and the integrins $\alpha M\beta 2$ and/or $\alpha V\beta 3$. Although I did not detect an enrichment of TLR4 in any experiment, the upregulation of innate immune proteins at the 30 second time point in Experiment 2 is consistent with a model whereby TNC nearly instantaneously activates pro-inflammatory signalling which is integrated into a transcriptional response over the subsequent 30 minutes. Finally, I speculated about potential TNC adaptors which appeared (KCTD12, LRRFIP1, and MPO), and I identified the lipoprotein receptors APOBR and LRP1 as candidates for a direct TNC co-receptor. Broadly speaking, the data were consistent with TNC activating a TLR4-dependent signalling pathway but without a direct TNC:TLR4 interaction. Much like CD14's presentation of LPS to TLR4:MD-2, I hypothesize that TNC must first bind to an up-stream co-receptor (my personal favorite candidate being APOBR, mostly because of how understudied it is relative to LRP1), which then displays TNC to a receptor complex which minimally includes TLR4 and integrins. It is my assertion, based on the data in Chapter 4 and Chapter 5, that the high-affinity TNC activation of TLR4 is wholly dependent on adaptor proteins, and that, unlike for LPS and other PAMPs, the simple "one receptor, one ligand" model is inadequate to explain the biological activity of endogenous DAMPs.

However, there are key weaknesses in my data which require validation in future experiments. First, I never detected TLR4, and many of the peptide counts for

identified membrane proteins (including the integrins) were low. I expect this is due to glycosylation on the ectodomains of membrane proteins, which produces heterogeneous peptides that are difficult to deconvolute when analyzed with MSMS. In the future, replication of the XL-AP-MSMS workflow should include a de-glycosylation step with PNGase, which will truncate all of the N-linked glycans and generate a specific mass-shift from the remaining GlcNAc. I expect this to increase sensitivity, especially for the membrane proteins which contain significant glycosylation, and provide more extracellular interactors. Secondly, the use of a covalent cross-linker to reveal neighboring interactors is both the most interesting and most challenging aspect of the workflow. Because TNC ligates integrins, the entire cytoskeletal network is pulled-down in a TNC-dependent manner, leading to the detection of a huge number of “interactors” that are likely non-specifically associated with the cytoskeleton. This makes it impossible to separate truly TNC-dependent interactors, especially as the signal is propagated to downstream effectors. Therefore, I recommend that these results be validated with one of the chemoproteomic approaches discussed in Section 5.2.1, particularly the expensive LRC-TRICEPS method. This experiment should unambiguously identify the immediate, direct TNC receptor(s) and validate the integrin binding I observed. Confident identification of the direct receptor(s) also provides an anchor point for deconvolution of the downstream signalling network determined via XL-AP-MSMS by integrating previous knowledge about the receptor itself.

Thirdly, at the outset I hoped to enrich TLR4 and identify a TLR4-associated receptorome; without confident detection of TLR4, I can only speculate, rather than prove, that TLR4 is directly involved in TNC detection during the experiment. Therefore, an important and orthogonal experiment would be to reverse the orientation of the pull-down via insertion of an affinity tag directly onto TLR4. This would ensure that TLR4 is detected in the pull-down and could provide clarity about which of the TNC interactors identified herein are also associated with TLR4; any proteins which were identified as both TNC and TLR4 interactors would be excellent co-receptor candidates. In the following Chapter 6, I work-up a CRISPR/Cas9 endogenous tagging pipeline which could be used for such a TLR4-focused pull-down experiment.

Finally, MSMS experiments should fundamentally be treated as hypothesis generation exercises, especially when the initial motivation for the experiment is hypothesis free (i.e. “let me see what I pull out” not “let me see if I pull out Protein X”). Therefore, all of my conclusions, especially the integrin and lipoprotein receptor binding,

should be validated in hypothesis-driven experiments. Besides validating specific interactors with a targeted pull-down and subsequent western blot, a relatively simple validation experiment would be to pre-incubate monocytes with siRNAs targeting key interactors prior to TNC stimulation and a cytokine ELISA. Reduction in the release of IL-6 or TNF α upon siRNA treatment would implicate that protein in TNC-activated, TLR4-dependent signalling. A lower throughput but more reliable approach would be the generation of knockout THP1 lines for each of the candidate interactors. In a non-COVID world, I would have liked to finish at least one of these validation experiments to provide support for the hypotheses formulated in this Chapter.

However, even with all of these caveats in mind, the XL-AP-MSMS method I developed is clearly functional, reliable, and suitable for interrogating complicated signalling pathways involving multiple receptors and adaptors. I expect the approach, which already generated intriguing new ideas about TNC, to be useful for studying diverse DAMPs—all that is required is a recombinant, biotinylated bait and a suitable cell line. Given DAMP-to-DAMP variability in signalling profiles, it is reasonable to expect that the composition of the receptor complex changes depending on the ligated DAMP. Therefore, this method might prove most useful in cross-DAMP comparison studies, which could clarify the mystery surrounding innate immune detection of endogenous proteins and shed light on the specific proteins worth targeting to abrogate pathological, DAMP-induced inflammation.

Chapter 6

CRISPR/Cas9 Functional Epitope Tagging of Endogenous Proteins for Structural Studies and Proteomics using GlowPAINT

The XL-AP-MSMS results presented in Chapter 5 suggest that the FRePs, especially TNC, are *bona fide* outside-in integrin activators that can additionally bind multiple cell-surface receptors. However, because I never detected a FReP-dependent enrichment of TLR4, my original hypothesis surrounding a TLR4-dependent receptorome remained essentially unexplored. To guarantee capture of TLR4 in MSMS experiments, the tag must be placed on TLR4, rather than on the FRePs, a fact which led me to explore gene editing methods to introduce affinity epitopes directly onto endogenous TLR4. Furthermore, motivated by the exciting work of my colleague Dr. Jack Kelly on the endogenous isolation of a chaperonin complex for cryo-EM and proteomics, I viewed this as a general opportunity to explore methods for the routine and parallelizable generation of affinity-tagged cell lines for diverse applications. Towards this end, Chapter 6 presents my efforts to create a rapid, simple, and moderate-throughput CRISPR-Cas9 pipeline based on the combination of two technologies: split-GFP and CRISPaint. Although I did not finish the project by generating a tagged TLR4 cell line, the method lays the foundation for others to generate this and other endogenously tagged cell lines, and it is currently being applied to other projects within my group. While the work ultimately does not directly speak to TLR4 or FReP biology, the underlying advancement will simplify studying the system in the future, and the work is of general utility to many experimental workflows. Therefore, I include it herein as the first of two “method development

vignettes” (along with the COVID-19-related work presented in Chapter 7) to conclude this thesis.

6.1 Materials and Methods

6.1.1 Construction of AAVS1 Targeting Vector

The pAAVS1-ZEO-TETON-GFP1-10 vector (diagrammed in Figure 6.3 and used to knock in GFP¹⁻¹⁰ to the *AAVS1* safe harbor locus) was constructed in four chunks using a HiFi assembly, with primers generated using the online tool NEBuilder (<https://nebuilder.neb.com/>). Chunk 2, containing the bacterial Amp^R cassette, the origin of replication, and both *AAVS1*-targeting homology arms, was generated by restriction digest. 2.5 µg of the parental vector pAAVS1-Puro-Tet3G-3xFLAG-Twin-Strep (Addgene #92099) was sequentially digested with *PmeI* and *BmgBI* in NEB Buffer 3.1 according to the manufacturer’s recommendations (NEB). The Chunk 2 digestion product was purified by gel extraction. The other three chunks were generated via touchdown PCR using Q5[®] Polymerase and primers containing HiFi compatible overhangs; the PCR products were spin purified prior to assembly. Chunk 1, containing the GFP¹⁻¹⁰ insert, was prepared using the template pcDNA3.1-GFP1-10 (Addgene #70219) and the primers Chunk1-f1 and Chunk1-r1. Chunk 3, containing the eukaryotic Zeocin^R cassette, was amplified from pExpres²-1 using the primers Chunk3-f1 and Chunk3-r1. Chunk 4, containing the bi-directional Tet-ON promoter system, was amplified from pAAVS1-Puro-Tet3G-3xFLAG-Twin-Strep using the primers Chunk4-f1 and Chunk4-r1. The four chunks were assembled in a 1:1:1:1 molar ratio into 50 ng of the Chunk 2 vector backbone using a 60 min HiFi reaction at 50 °C. The assembled plasmids were transformed into Mach 1 cells and selected on Amp^R plates. Multiple colonies were checked for the correct product via colony PCR with two primer sets: TREG3-seq-f1 + GFP-seq-r1 and HA-L-seq-f1 + ZeoR-seq-r1. Two colonies containing both of the expected colony PCR amplicons were minipreped and submitted for Sanger sequencing using HA-L-seq-f1 and TREG3-seq-f1, which confirmed that both the Zeo^R and GFP¹⁻¹⁰ inserts were cloned in-frame as expected. The final pAAVS1-ZEO-TETON-GFP1-10 plasmid was stocked as miniprep DNA at –20 °C and as a glycerol stock at –80 °C prior to use in downstream knock-in experiments.

Primer sequences used to assemble pAAVS1-ZEO-TETON-GFP1-10 (overhangs in

lowercase, annealing regions in uppercase font):

Chunk1-f1 (5'-taaggcggccgcATGTCCAAAGGAGAAGAAC)

Chunk1-r1 (5'-ctagaaggcacagcctgcagggtttTCATTTTTCATTTGGATCTTTG)

Chunk3-f1 (5'-aggaagtcttctaacaatgcggtgacgtggaggagaatcccggccctagc-
GCCAAGTTGACCAGTGCC)

Chunk3-r1 (5-atgtctgtctagaTCAGTCCTGCTCCTCGGC)

Chunk4-f1 (5'-gagcaggactga-
TCTAGACAGACATGATAAGATAACATTGATGAGTTTGGAC)

Chunk4-r1 (5'-ctccttttgacatGCGGCCGCCTTAAGGCTG)

Sequencing primers to check pAAVS1-ZEO-TETON-GFP1-10 in colony PCR and with Sanger Sequencing:

TREG3-seq-f1 (5'-GGGCGCCTATAAAAGCAGAG; sits in Tet-ON promoter upstream of GFP¹⁻¹⁰)

GFP-seq-r1 (5'-TGGATCTTTGCTCAGGACTGT; sits at the 3' end of GFP¹⁻¹⁰)

HA-L-seq-f1 (5'-CATCTTCCAGGGGTCCGAG; sits in homology arm left upstream of the ZeoR cassette)

ZeoR-seq-r1 (5'-TCAGTCCTGCTCCTCGGC; sits at the 3' end of the ZeoR marker)

6.1.2 Construction of pGlowPAINT GFP¹¹-Containing Universal Donors

The new vectors generated in this section include:

- pGlowPAINT-3xFLAG-PuroR
- pGlowPAINT-AviTag-PuroR
- pGlowPAINT-NanoLuc-PuroR
- pGlowPAINT-Proteo-TunerDD-PuroR
- pGlowPAINT-SpyTag-PuroR
- pGlowPAINT-StrepTagII-PuroR

They are built from the original pCRISPAIN-TAG-PuroR Universal Donor Vectors (Addgene Kit #1000000086), to which I added GFP¹¹ (the final peptide from GFP used for reconstitution with split-GFP technology) downstream of the tag and upstream of the T2A-PuroR cassette. The subset of pCRISPAIN-TAG-PuroR vectors I chose to modify with GFP¹¹ are designed to encompass a diverse set of applications including affinity purification, immobilization, knockout, microscopy, or cellular assays. The added sequence encodes the following peptide (GFP¹¹ in bold): GGGGSGGGGSR**DHMLHEYVNAAGIT**GGGG.

To add this peptide, a common forward primer (GFP11-Insert-f1) was ordered that annealed to the T2A region and contained half of the GFP¹¹ peptide as a 5'-overhang. Reverse primers (3XFLAG-r1, AVI-r1, HALO-r1, NANOLUC-r1, PROTUNER-r1, SPY-r1, and STREPII-r1), unique to each tag, were ordered that annealed to the specific tag region and contained the other half of the GFP¹¹ peptide as a 5'-overhang. The primer sequences can be found below. Along with the two primers, the corresponding pCRISPAIN-TAG-PuroR vector was used as a template for amplification using Q5[®] polymerase and touchdown-PCR with a 2 min 30 sec extension time. The PCR amplicons, containing the linearized pCRISPAIN vectors with the GFP¹¹ added as two partial overhangs at the ends, were treated with DpnI for 1 hour at 37 °C prior to spin purification. The purified linearized vectors were phosphorylated and re-circularized using the following reaction:

5 µL of 2X RapidLigation Buffer (Promega)
 1.5 µL of purified PCR amplicon, containing between 60 and 75 µg of DNA
 1 µL of T4 DNA Ligase (Promega)
 1 µL of Polynucleotide DNA Kinase (Promega)
 1.5 µL of Nuclease-Free Water

The ligation reaction was incubated for 1.5 hours at room temperature prior to transformation of 2.5 µL of the reaction into Mach 1 cells. Transformants were recovered on Carb^R plates and checked by colony PCR using a common set of sequencing primers (PSITE-seq-f1 + GFP11-seq-r1), which amplify the tag region only if the GFP¹¹ sequence has been inserted. Two clones for each tag with the correctly sized amplicon in the colony PCR were submitted for Sanger Sequencing with PSITE-seq-f1. At least one plasmid for each pGlowPAINT vector contained the correct in-frame insertion of GFP¹¹. The new pGlowPAINT vectors were stored as miniprep DNA at -20 °C and a glycerol stock at -80 °C prior to use in knock-in experiments.

Primers used to add GFP¹¹ to the original pCRISPAIN-TAG-PuroR vectors (annealing region in uppercase, GFP¹¹-containing overhang in lowercase):

GFP11-Insert-f1 (5'-cttcatgagtatgttaatgctgcagggtatcacaggaggcggtggc-TCCGGAGGAGAGGGCAGA)

3xFLAG-r1 (5'-aaccatatggtctctactccctccccctccagaaccgcctccacc-CTTATCGTCATCGTCCTTATAATCAATATCATG)

AVI-r1 (5'-aaccatatggtctctactccctccccctccagaaccgcctccacc-CTCGTGCCACTCGATCTTCTGGG)

HALO-r1 (5'-aaccatatggtctctactccctccccctccagaaccgcctccacc-ACCGGAAATCTCCAGAGTAGACAGCC)

NANOLUC-r1 (5'-aaccatatggtctctactccctccccctccagaaccgcctccacc-CGCCAGAATGCGTTCGCACAG)

PROTUNER-r1 (5'-aaccatatggtctctactccctccccctccagaaccgcctccacc-TTCCAGTTCTAGAAGCTCCACATCG)

SPY-r1 (5'-aaccatatggtctctactccctccccctccagaaccgcctccacc-CTTGGTGGGCTTGTTAGGCGTC)

STREP11-r1 (5'-aaccatatggtctctactccctccccctccagaaccgcctccacc-CTTCTCGAACTGGGGGTGGCTC)

Sequencing primers used to check all pGlowPAINT vectors in colony PCR and Sanger Sequencing:

PSITE-seq-f1 (5'-CGCTCTCCATCAAAACAAAACG; sits upstream of the tag in the common pause site)

GFP11-seq-r1 (5'-CTCATGAAGAACCATATGGTCTCTACTCCC; sits within the common GFP¹¹ addition downstream of the tag)

6.1.3 Cloning of eSpCas9(1.1)-sgX Vectors

To generate eSpCas9(1.1)-sgX vectors for CRISPR/Cas9-mediated cleavage of gDNA, the targeting protospacer regions, as close to the stop codon as possible within protospacer adjacent motif (PAM) restrictions, were taken directly from Schmid-Burgk *et al.* [511]. These regions are reported for each of the 26 genes tested in Figure 6.5D. Two complementary oligos were ordered for each sgRNA. The oligos contained the requisite *BbsI* cloning sites and 19 bp of protospacer (denoted N¹⁹). The sequence of the forward oligo was 5'-CACCG-N¹⁹, and the sequence of the reverse oligo was 5'-AAAC-N¹⁹-CC. Note that the reserve oligo contains the reverse complement of

the N¹⁹ region in the forward oligo to ensure proper annealing. The oligos were resuspended to 100 μ M and annealed to each other in the following reaction:

10 μ L Oligo-F aka “top” oligo
10 μ L Oligo-R aka “bottom” oligo
10 μ L NEB Buffer 2.1
70 μ L Nuclease-Free Water

Using a thermal cycler, the oligo reaction was denatured for 10 min at 95 °C and then annealed by cooling from 95 °C to 25 °C in increments of 10 °C every 1.5 min. The annealed oligos were precipitated by the addition of 30 μ L of 3 M NaAcetate (pH = 5.2) and 300 μ L of 100% EtOH prior to centrifugation for 45 min at 4 °C. The precipitated, annealed oligos were washed once with 70% EtOH and then allowed to air dry for 2 hours to remove trace EtOH. The annealed oligos were then resuspended in nuclease-free water to an average concentration of 220 ng μ L⁻¹. Finally, the oligos were phosphorylated for 1 hour at 37 °C in the following reaction:

2 μ L 10X T4 PNK Buffer (Thermo Fisher)
1 μ L Polynucleotide DNA Kinase (Thermo Fisher)
5 μ L of annealed oligo, approximately 1 μ g total DNA
12 μ L Nuclease-Free Water

The destination vector eSpCas9(1.1) (Addgene #71814), which contains a U6-promoter-driven open reading frame for sgRNA expression alongside the eSpCas9(1.1) enzyme itself, was linearized with *BbsI*-HF (NEB) for two hours at 37 °C prior to gel extraction and purification. 20 ng of linearized eSpCas9(1.1) was ligated with 25 to 50 ng of the phosphorylated, annealed oligos using T4 DNA Ligase according to the manufacturer’s recommendations (Thermo Fisher). The circularized vectors were transformed into Mach 1 cells and selected on Carb^R plates.

To confirm insertion of the targeting region into the sgRNA scaffold, colony PCR was completed using a standard forward primer (hU6-f1) and the target-specific reverse oligo as the reverse primer; a strong band at approximately 300 bp implied insertion of the targeting region. At least one colony-PCR-positive clone was miniprepmed and validated by Sanger Sequencing using hU6-f1. As shown in Figure

6.5C, the Colony PCR strategy was 100% effective at identifying clones with the correct sgRNA sequence, suggesting that this method can be used in a HT-manner without requiring sequencing of every eSpCas9(1.1)-sg vector.

Sequencing primer used for Colony PCR and Sanger Sequencing:

hU6-f1 (5'-GAGGGCCTATTTCCCATGATT; sits upstream of the sgRNA scaffold in the U6 promoter)

6.1.3.1 Generation of GFP¹⁻¹⁰ Expressing Cell Lines

Expi293 cells were maintained primarily in suspension using FreeStyleTM 293 Expression Medium (Thermo Fisher) containing Pen/Strep. HeLa cells were maintained primarily as adherent cultures in Dulbecco's Modified Eagle Medium (DMEM) + 10% heat-inactivated fetal bovine serum (FBS) containing Pen/Strep and split using TrypLE (Gibco) according to the manufacturer's recommendations. Prior to starting the experiment, the sequence of the *AAVS1* targeting region was confirmed via PCR and Sanger Sequencing as described in Section 6.1.3.2. For transfection in 6-well plates, 2 mL of HeLa cells were seeded at a density of 0.5×10^6 cells per mL or 2 mL of Expi293 cells were seeded at a density of 0.4×10^6 cells per mL. The cells were rested for 24 hours prior to transfection according to the manufacturer's recommendations (Thermo Fisher) with 7.5 μ L LipofectamineTM 3000, 2 μ L μ g⁻¹ P3000 reagent, 2 μ g eSpCas9(1.1)-AAVS1-sg1, and 4 μ g pAAVS1-ZEO-TETON-GFP1-10. A mock transfection containing no DNA was prepared similarly as a negative control. After 48 hours to allow cleavage and insertion, fresh medium containing 100 μ g mL⁻¹ Zeocin was added. 48 hours later, the Zeocin was increased to 250 μ g mL⁻¹ for the Expi293 cells. Zeocin selection was maintained for 8 additional days until the negative control cells appeared fully dead by manual inspection. The Zeocin-selected bulk pools were washed with PBS and given fresh medium with no Zeocin (containing 5% added FBS for Expi293, 10% FBS as standard for HeLa), recovered for two days, and then expanded as standard into T75 flasks (HeLa). The bulk pool cell lines were used for gDNA extraction and, after induction with 1 μ g mL⁻¹ doxycycline, for western blotting as described below. The bulk pools were frozen in 10% DMSO and stored long-term in LN₂.

To generate the monoclonal Expi293^{GFP10} cell line, the recovered, Zeocin-selected bulk pool was serially diluted to 0.5 cells per 100 μ L in FreeStyleTM + 5% FBS and 100 μ L was seeded into each well of two 96-well plates. After three weeks, I observed surviving monoclonal cell lines in three wells, which were expanded into a 6-well

dish prior to induction of the GFP¹⁻¹⁰ with 1 $\mu\text{g mL}^{-1}$ doxycycline for 48 hours. Two monoclonal lines (mcB2 and mcG2) expressed GFP¹⁻¹⁰ by western blot and contained the correct insert via PCR. These correct lines were suspension-adapted from 6-well dishes into small shake flasks containing empty FreeStyleTM medium prior to long-term storage in 10% DMSO in LN₂. The suspension-adapted mcB2 line was re-named Expi293^{GFP10} and used for subsequent experiments.

6.1.3.2 gDNA Isolation and PCR

gDNA was extracted using high-salt ethanol precipitation essentially as previously described [512]. Between 500,000 and 5×10^6 cells were lysed with 300 μL of Lysis Buffer (10 mM Tris pH = 8.0, 10 mM EDTA, 150 mM NaCl, 0.2% SDS) containing 1:25 dilution of 10 mg mL^{-1} Proteinase K (NEB). The lysates were incubated overnight at 55 °C. The next morning, the samples were incubated for 15 min at room temperature with gentle shaking after the addition of 100 μL of saturated NaCl. The precipitated proteins were removed by centrifugation at 21,000 $\times g$ at room temperature for 10 min, the supernatant was transferred to a fresh tube, and 600 μL of ice cold 100% EtOH was added to precipitate the gDNA. The samples were centrifuged as before, the pellet was washed once with 70% EtOH, and the gDNA pellet was resuspended with 50 μL of TE. The concentration and purity of the gDNA was estimated using a NanoDrop.

Please refer to Figure 6.3B for a schematic of these PCR verification reactions. For IN-IN PCR of the *AAVS1* locus, the primer set TREG3-seq-f1 + GFP-seq-r1 (see Section 6.1.1 for sequences) was used; for IN-OUT PCR, the primer set TREG3-seq-f1 + AAVS1-gen-r2 was used. For amplification of the *AAVS1* locus for genotyping prior to starting the GFP¹⁻¹⁰ experiment described in Section 6.1.3.1, the primer pairs AAVS1-gen-f1 + AAVS-gen-r1 and AAVS-gen-f2 + AAVS1-gen-r2 were used. 200 ng of gDNA was used as a template for PCR amplification using Q5[®] Polymerase with 40 cycles using a 30 sec denaturing step at 98 °C, a 30 sec annealing step at 55 °C, and a 1.5 min elongation step at 72 °C. PCR amplicons were subjected to agarose gel electrophoresis for analysis.

New primers used for PCR analysis of the *AAVS1* locus:

AAVS1-gen-f1 (5'-CTAGTCTGTGCTAGCTCTTCCAG; sits upstream of the cut site induced by eSpCas9(1.1)-AAVS1-sg1 in the left gDNA homology arm)

AAVS1-gen-r1 (5'-TGACGCACGGAGGAACAATATAA; sits downstream of the cut site induced by eSpCas9(1.1)-AAVS1-sg1 in the right gDNA homology arm)

AAVS1-gen-f2 (5'-GAGCTCAGCTAGTCTTCTTCCTC; sits upstream of the cut site induced by eSpCas9(1.1)-AAVS1-sg1 in the left gDNA homology arm)

AAVS1-gen-r2 (5'-TCTCCTCTTGGGAAGTGTAAGGA; sits downstream of the cut site induced by eSpCas9(1.1)-AAVS1-sg1 in the right gDNA homology arm)

6.1.4 Western Blotting

Cell lysates were prepared using M-PER according to the manufacturer's recommendations (Thermo Fisher), and the approximate concentration of the lysates was estimated on a NanoDrop. Between 10 and 50 µg total input protein, standardized between experimental and control samples, was used for SDS-PAGE. SeeBluePlusTM Pre-Stained Standards (Thermo Fisher) were used as molecular weight references. The gels were semi-dry transferred onto PVDF membranes using the TransBlot[®] TurboTM system according to the manufacturer's recommendations (Bio-Rad) and then blocked for 1 hour with PBST (1X PBS + 0.05% Tween-20) + 5% BSA. The blots were probed with 1:2000 anti-GFP primary antibody (Chromotek, Rat anti-GFP, [3H9]) prepared in PBST + 5% BSA and incubated overnight at 4 °C. The blots were washed 3 times for 5 minutes with PBST and then incubated for 1 hour at room temperature with 1:5000 anti-rat secondary antibody (Invitrogen, Goat anti-Rabbit IgG HRP, Cat. 31470) prepared in PBST + 5% BSA. The blot was washed 3 times for 5 minutes with PBST, developed with the PierceTM ECL Substrate according to the manufacturer's recommendations (Thermo Fisher), and then imaged for chemiluminescence using the LAS-3000 Imaging System (Fuji). For anti-3xFLAG western-blotting, a directly HRP-coupled primary antibody (Sigma A8592, Clone M2) was used exactly as above.

6.1.5 Mismatch Cleavage Assay

The eSpCas9(1.1)-sgX vectors were minipreped and heat sterilized (10 min at 90 °C) prior to use. 800,000 Expi293^{GFP10} cells were seeded in 1 mL FreeStyleTM medium into each well of a 24-well plate. Each well was transfected with a different sgRNA as recommended by the manufacturer using 3 µL of LipofectamineTM 3000, 1 µg eSpCas9(1.1)-sgX, and 2 µL of P3000 reagent. Transfected cells were allowed to grow for 48 hours with 170 RPM shaking at 37 °C and 8% CO₂ prior to harvesting of the cells via centrifugation (500 x *g* for 15 min). gDNA was extracted using 100 µL

of the commercially available reagent DNAzol™ according to the manufacturer’s recommendations (Thermo Fisher). The gDNA was resuspended to a final volume of 35 μL and the concentrations, ranging from 20 to 400 $\text{ng } \mu\text{L}^{-1}$, were estimated with a NanoDrop. gDNA was also extracted from a bulk sample of 30×10^6 untransfected Expi293^{GFP10} cells as a negative control.

To amplify the sgRNA targeting regions, two sets of gene-specific primers per sgRNA were designed as recommended using Primer3 [513]. 200 ng of gDNA was used as the template to amplify the target regions (for each primer set, from either the transfected or un-transfected cells) using Touchdown PCR with Q5® Polymerase and the Q5® GC-Enhancer additive. For each sgRNA tested, the strongest amplicon as judged by agarose gel electrophoresis (along with the matched negative control) was spin purified and the concentration was estimated with a NanoDrop. The purified amplicons were then denatured and re-annealed (5 min at 95 °C, 95 to 85 °C over 30 sec, then 85 to 25 °C over 10 min) in the following reaction:

250 ng spin-purified amplicon
2 μL 10X NEB Buffer 2
Up to 19 μL with nuclease-free water

After denaturing and re-annealing, 1 μL of T7 Endonuclease I (NEB) was added and the samples were incubated for 15-30 min at 37 °C. T7-induced cleavage of the heteroduplex DNA was analyzed via agarose gel electrophoresis.

6.1.6 GlowPAINT Pilot with *GDE1* Targeting sgRNA

To pilot the entire GlowPAINT method, targeting the human *GDE1* locus for knock-in, 1×10^6 Expi293^{GFP10} cells were seeded in 2.5 mL FreeStyle™ 293 medium in 6-well suspension plates. The cells were transfected as before with Lipofectamine™ 3000 according to the manufacturer’s recommendations with the following: 1 μg eSpCas9(1.1)-GDE1-sg1 + 1 μg Frame Selector Plasmid (tested #0, #1, and #2) + 2 μg pGlowPAINT-3xFLAG-PuroR. For negative controls, one well was transfected with just eSpCas9(1.1)-GDE1-sg1, one well with 1 μg of each Frame Selector Plasmid and 2 μg of pGlowPAINT-3xFLAG-PuroR, and one well with empty transfection reagent. The cells were grown for 72 hours at 37 °C and 120 RPM to allow cleavage and insertion. The un-selected bulk pools were then divided in half into two 6-well

plates and $2\text{ }\mu\text{g }\mu\text{L}^{-1}$ puromycin was added to one of the 6-well plates. The cells were selected for eight days with puromycin prior to recovery in FreeStyle™ 293 medium to generate polyclonal selected lines, which were frozen with 10% DMSO prior to downstream experiments. The samples were subjected to western blot as described in Section 6.1.4 except using a single anti-FLAG HRP-conjugated primary antibody (Sigma A8592). gDNA was also extracted and the targeted region was amplified using OUT-OUT PCR according to the protocol reported in Section 6.1.5 using the primers GDE1-gen-f1 + GDE1-gen-r1. To generate monoclonal cell lines, the polyclonal Puro^R pools were subjected to limiting dilution and, after allowing the cells to re-grow, a second selection with puromycin. Lysates from five monoclonal cell lines (mc1–mc5) were analyzed with an anti-FLAG western blot as before, and 120×10^6 cells from the best expressing line mc4 were subjected to an anti-FLAG purification followed by anti-FLAG western blotting to validate expression of the correct GDE1-3xFLAG-GFP¹¹ protein product.

Primers used for PCR analysis of the *GDE1* locus:

GDE1-gen-f1 (5'-GATTCCTGAGACAGCTGCTT; sits approximately 200 bp upstream of the target site)

GDE1-gen-r1 (5'-GTGTTGAACTCTGGCATGCC; sits approximately 200 bp upstream of the target site)

6.2 Background and Motivation

In Chapter 5, using biotinylated FRePs and MSMS, I derived new insights into FReP biology; however, I did not observe TLR4, meaning the results did not conclusively address my core questions about the TLR4:FReP interaction. The only way to guarantee detection of a specific protein in a proteomics experiment is to install an affinity tag directly on that protein; therefore, my next experimental aim focused on the generation of affinity-tagged TLR4 cell lines. Around the same time I decided to pursue such a knock-in, Dr. Jack Kelly from the Yue Lab demonstrated the potential of leveraging affinity-tagged cell lines to isolate an endogenous multi-protein chaperonin for structural studies. The tailless complex polypeptide 1 ring chaperonin complex (TRiC) is composed of two vertically stacked rings made from eight proteins each (CCT1–CCT8) and is responsible for refolding 10 percent of the human proteome. Jack decided to introduce an affinity tag directly into the CCT5

genomic locus in HEK293 cells via a standard CRISPR-Cas9 homology-directed repair (HDR) workflow which is summarized in Figure 6.1A.

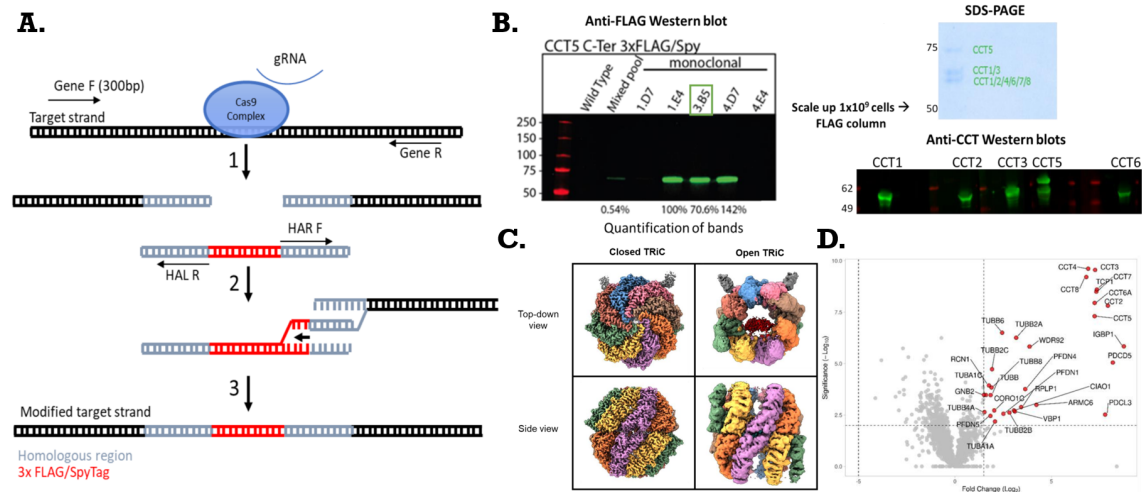


Figure 6.1 TRiC: A Case Study for Endogenous Tagging. **A.** Standard HDR repair approach for knock-in of 3xFLAG-Spy Tag to CCT5 C-terminus. **B.** (left) Anti-flag western blot for monoclonal tagged CCT5 cell line, (right top) SDS-PAGE of purified TRiC complex after anti-FLAG column, (right bottom) anti-CCT_x western blot showing that the entire TRiC complex has been isolated. **C.** High-resolution structures of the human TRiC complex. **D.** TRiC interactors identified from the tagged cell line via AP-MSMS. A–D. are reproduced from Dr. Jack Kelly’s thesis with permission.

First, the Cas9 enzyme, bound to a CCT5 targeting sgRNA, introduces a dsDNA break in the CCT5 gene near the C-terminus. The cell then attempts to repair the break via non-homologous end joining (NHEJ) or HDR. If an “HDR template”, containing 5’ and 3’ homology arms flanking the desired insertion, is supplied, a subset of cells repair the break by inserting the novel sequence. HDR templates often contain homology arms up to 1 kb in length on each side of the cut and must be assembled *ad hoc* for every desired edit. Finally, once the dsDNA break is repaired via HDR and the novel sequence is knocked-in, the bulk pool of edited cells are subjected to limiting dilution and monoclonal cell lines bearing the desired edit are identified, usually by western blot. This laborious process of cell dilution and monoclonal line generation can take up to eight weeks to complete for one gene; however, the results were truly dramatic for TRiC. As Figure 6.1B shows, Jack was able to select a monoclonal line containing a 3xFLAG/SpyTag on the C-terminus of CCT5 and scale up the line to isolate the entire TRiC complex via anti-FLAG affinity purification. Using endogenously purified TRiC, Jack was able to both solve multiple high-resolution structures of TRiC in various conformations using Cryo-

EM (see Figure 6.1C) and define the endogenous TRiC interactome using MSMS proteomics (see Figure 6.1D).

I present these results, which are reproduced with permission from Dr. Kelly's thesis and a March 2021 pre-print [439], to highlight the power of the approach developed herein. Edited cell lines, generated with CRISPR-Cas9, can form the bedrock of truly integrative structural biology projects and facilitate experiments that are impractical or impossible with standard plasmid-based over-expression. This is especially true for multiprotein complexes, which can be challenging to reconstitute recombinantly. Jack's successful pilot energized me to look beyond the one-off generation of a tagged TLR4 cell line and instead consider methodology for the rapid (one to two week), routine knock-in of functional epitopes to any gene for any reason at any time.

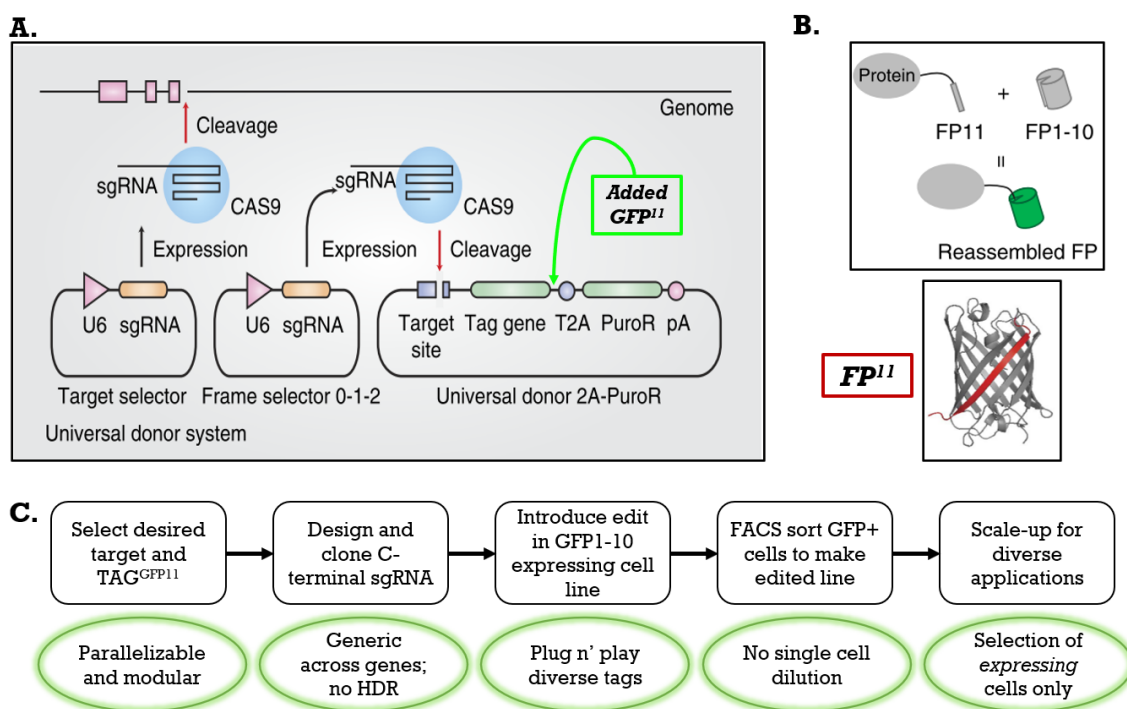


Figure 6.2 GlowPAINT: Background and Overview. **A.** Schematic of the CRISPAINT method, modified from Schmid-Burgk *et al.* [511]. Insertion point for GFP¹¹ is annotated. **B.** Cartoon of split-FP technology, modified from Kamiyama *et al.* [514]. **C.** Overview of the GlowPAINT workflow (black squares) and the advantages of the approach (green circles).

6.3 GlowPAINT: CRISPaint + Split-GFP

To facilitate uptake by others, I sought to minimize the molecular biology required for each edit, and much like FAS2FURIOUS (see Chapter 3) I wanted the method to be moderate-throughput to allow multiple cell lines to be generated in parallel. After defining the requirements for the system, I looked to the literature for current best-practice methodologies; I also met twice with Dr. Philip Hublitz, the leader of the WIMM Genome Engineering Centre, who provided invaluable advice about potential solutions. Rather than review the CRISPR literature in detail, which is outside the scope of this thesis, I focus only on the cutting-edge CRISPR-Cas9 knock-in technologies which I considered for the new workflow.

CRISPR-Cas9 as a platform has matured rapidly in recent years, and methods are available to generate insertions and deletions essentially of any kind at essentially any genomic locus [515]. Ran *et al.*'s 2013 report outlined the key aspects of a standard CRISPR-Cas9 workflow [516], and interested readers are pointed to detailed reviews covering the history of CRISPR-Cas9 technology and its various applications [517, 518, 519]. Throughout, I discuss and use the original, most common SpCas9 enzyme derived from *S. pyogenes* [520], even though many other enzymes with diverse specificity and sequence restrictions are now available [515].

All CRISPR experiments begin with the design of a target-specific sgRNA, which contains a scaffolding region to load the sgRNA into the Cas9 enzyme and a gene-specific segment to direct the enzyme to cleave a specific genomic DNA (gDNA) sequence. Oftentimes, the variable gene-specific targeting region of the sgRNA is cloned into a plasmid in-frame with the scaffolding region and downstream of a U6 promoter, which promotes high levels of transcription. The sgRNA must be carefully selected to maximize on-target cutting while minimizing off-target activity; a variety of standard computational tools for predicting sgRNA efficiency *in silico* are widely available [521]. Upon co-delivery of the sgRNA and the Cas9 enzyme to the cells of interest (either as a plasmid via transfection or as a pre-packaged ribonucleoprotein, *aka* RNP, via electroporation), the sgRNA-loaded Cas9 enzyme scans the DNA and introduces a dsDNA break 3-bp upstream of the protospacer adjacent motif (PAM). The PAM, which for SpCas9 is 5'-NGG, is strictly required for cleavage and therefore defines which regions are amenable to editing. While the NGG motif is quite common across the genome, other enzymes can be used in AT-rich regions [522]. After the dsDNA break is introduced, the cell can repair the break via either NHEJ or HDR, both of which can be leveraged to knock-in new genetic

material at the cut locus. At this point in the experiment, the cells bearing the desired edit (representing a minority of the total population among un-transfected cells and cells bearing indels rather than the knock-in) must be isolated, often via antibiotic selection or limiting dilution. The polyclonal line, generated via selection, or monoclonal line, generated via single-cell dilution, can then be scaled up and used for tag-specific applications.

The oldest and most common approach to generating a knock-in relies on a long dsDNA HDR template containing the edit of interest and flanking homology arms; this is the approach that Jack used for CCT5 (described in Section 6.2). This approach is low throughput and slow, primarily because a bespoke HDR template must be manually generated via PCR and Gibson assembly for each edit; amplification of the long homology arms directly from gDNA is also especially finicky. Furthermore, the knock-in efficiency is often quite low, requiring many clones to be checked after a tedious single-cell dilution step. The efficiency can be improved by coupling the edit to a marker, either via direct incorporation of a fluorescent or antibiotic-resistant protein in the HDR template or via orthogonal editing of the *ATP1A1* locus and co-selection with ouabain [523]. However, even with these improvements, the use of long dsDNA HDR repair templates is fundamentally low-throughput.

An interesting simplification of the HDR approach allows the user to bypass HDR template generation. Instead of editing the target genetic locus, many groups instead re-introduce their gene of interest (fused to the appropriate tag) into the safe-harbor *AAVS1* locus [524]. The *AAVS1* gene, *aka* the “Adeno-Associated Virus Integration Site 1”, is routinely used as a destination locus for exogenous genetic cargo [524]. The locus lies within the non-essential *PPP1R12C* gene and facilitates the robust (but still physiological, i.e. not over-expressed) expression of transgenes. Because the same loci is targeted in every experiment, only one *AAVS1*-targeting sgRNA is required, and the *AAVS1*-specific homology arms can be included in a generic plasmid flanking an MCS for insertion of the gene of interest. For each edit, all that is required is standard sub-cloning of the gene of interest into the HDR template plasmid, which often contains a selection marker alongside the desired tag and the *AAVS1*-specific homology arms. Dalvai *et al.* deployed this *AAVS1*-knock-in approach to great effect to isolate and study multiprotein complexes using a TAP 3xFLAG-TwinStrep tag [525]. Dalvai *et al.* also shared the required reagents and detailed methods for this approach, which I relied on heavily when designing my experiments [526]. However, despite its advantages, the *AAVS1*-knock-in method still requires bespoke sub-cloning of each gene of interest to generate the HDR template,

and the knocked-in gene is not expressed at native levels (but rather at the level of the *PPP1R12C* gene, whose promoter the knocked-in gene co-opts). Furthermore, the un-tagged allele (from the native locus, which is not edited) is still present and is expressed alongside the tagged version; especially for cellular and functional assays, this represents a severe drawback and can complicate interpretation of the knock-in phenotype. Therefore, since I wanted a generic method for both endogenous purification and cellular assays, I instead looked for next generation techniques which result in modification of the native genomic loci.

One alternative involves the use of linear, single-stranded oligodeoxyribonucleotide (ssODN) or linear dsDNA HDR repair templates. Very short homology arms (always less than 100 bp, sometimes as short as 15 bp) are sufficient to engage the HDR mechanism and achieve knock-in. Furthermore, since the HDR templates are short, they can be trivially synthesized for each edit, removing all manual intervention in the construction of the HDR template. For larger inserts (i.e. an entire fluorescent protein) which would be expensive to synthesize each time, the short homology arms can even be included as overhangs in PCR primers, allowing the HDR template to be generated with just a single PCR. In either case, the use of linear DNA with short homology arms simplifies the generation of a repair template versus traditional approaches. Paix *et al.* used synthetic ssODN donors with short homology arms (35 bp) to knock-in inserts (including full-length GFP) up to 1,000 bp in length in human cell lines [527]. Similarly, Kanca *et al.* used homemade, PCR-generated ssODN repair templates, containing 100 bp homology arms, to “drop-in” large inserts of up to 2 kb in *Drosophila* cell lines [528]. The approach also works *in vivo*: Miura *et al.* deployed “Easi-CRISPR”, with long ssDNA donors and short homology arms (55-105 bp), to generate knock-in mice containing up to 1.5 kb of added cargo with up to 100% efficiency. A final variation, coined “CHoP-IN” and co-developed with GlaxoSmithKline, removes the requirement for HDR altogether [529]. Using just two cleverly designed primers, which are unique to each desired edit and are used to amplify the tag of interest (for example, GFP) with compatible homology arms supplied as overhangs, Manna *et al.* showed that targeted, large insertions (up to 1 kb) can be made with only NHEJ. All of these methods are “next-generation” in the sense that they bypass the tedious generation of large HDR templates, making the process faster and simpler, and result in the modification of the native genetic locus.

For a one-off edit, CHoP-IN is the best choice in my opinion, and I would choose CHoP-IN for bespoke edits moving forward. However, in the context of a moderate-

throughput system, all of the next-generation methods still have drawbacks. First, the approaches described above require *both* a target-specific sgRNA and an edit-specific repair template. Even though the repair templates are easier to make than classic HDR approaches, they still must be designed, synthesized, and prepared for each desired gene/tag combination. As the number of cell lines to be made in parallel increases, the design and handling becomes cumbersome; in fact, just the design stage of the CHoP-IN workflow takes me up to an hour per edit, irrespective of the requisite lab time to generate and purify the required edit-specific PCR products. Therefore, for the work I present herein, I decided to deploy the only truly generic solution for endogenous tagging, called CRISPAINT [511], which is diagrammed in Figure 6.2A. CRISPAINT is fully gene agnostic—only a target-specific sgRNA is required to introduce a dsDNA break, at which point a generic NHEJ repair template, containing the functional tag and a selection marker, is introduced. CRISPAINT requires three plasmids at transfection: the “target selector” containing the target-specific sgRNA and Cas9 enzyme to cut the gDNA locus, the “universal donor” containing the desired tag and a coupled Puro^R selection marker, and the so-called “frame selector”. The frame selector plasmid contains a second sgRNA which can induce a dsDNA break *within the universal donor*, creating a linear dsDNA fragment that can be inserted via NHEJ at the site of the genomic dsDNA break. There are three frame selector plasmids (+0, +1, and +2), which linearize the universal donor at 3 different positions just up-stream of the tag. This ensures that the selected tag is incorporated in-frame with the target gene’s C-terminus even if the PAM is off-set from the stop codon. CRISPAINT can only introduce tags at the C-terminus due to the 3’ junk plasmid backbone DNA that is knocked in along with the tag itself; however, this is a worthwhile trade-off for an approach that does not require the creation of an edit-specific repair template of any kind. With CRISPAINT, *only* a target-specific sgRNA is required; the NHEJ repair template in the universal donor plasmid can be used over and over again at different loci. Furthermore, Schmid-Burgk *et al.* generously shared the entire “pCRISPAINT” plasmid library on Addgene, including dozens of different tags for diverse applications, which facilitates rapid uptake by other labs. Therefore, CRISPAINT was the simplest (i.e. the least molecular biology work-up per target) technology that met my stated requirements, so I chose it over the others for the application presented herein.

However, even with CRISPAINT, the editing workflow still requires a single-cell dilution and the manual screening and expansion of correct clones. This creates a severe bottleneck and limits the number of cell lines that can be simultaneously generated. In my experience, any more than two edits becomes impractical because

of the intensive dilution, screening, and propagation steps. Therefore, I decided to combine the CRISPAIN T platform with a second technology to facilitate rapid selection of the edited clones: split-GFP. The eleven-stranded GFP β -barrel can be structurally split into two parts: one large fragment containing the first 10 β -strands (GFP¹⁻¹⁰) and a second small fragment containing just the final strand (GFP¹¹). Neither of these fragments are fluorescent alone; however, if both portions are present, full-length, fluorescent GFP is reconstituted (see Figure 6.2B for a schematic) [530]. Split-FP technologies (which work not only for GFP but for other related proteins like sfCherry) provide a foundation for a variety of complementation assays and microscopy applications [514, 531]. More importantly for this application, split-GFP allows a small tag (GFP¹¹, just 16 residues) to fluorescently label a gene of interest. My idea, then, was to combine split-GFP with CRISPAIN T by adding the short GFP¹¹ sequence to the universal donor plasmids with the tags of interest; then, edited cells (already expressing GFP¹⁻¹⁰) could be identified and sorted at scale using fluorescence-activated cell sorting (FACS). Selection of GFP¹¹-containing edited cells via FACS is strict and direct—only cells containing an in-frame insertion of the desired tag at an actively transcribed locus which generates a soluble fusion protein will fluoresce. This is much stricter than coupled knock-in of the entire GFP cassette (for example, downstream of the edit and separated by a self-cleaving T2A peptide), which provides no indication that the target fusion protein is stable (or, in fact, if the fusion protein is present at all). Furthermore, if the base cell line contains an *inducible* GFP¹⁻¹⁰ cassette, then fluorescence can be turned on for FACS-selection but off for functional experiments, leaving behind just a small, 16 residue GFP¹¹ scar on the fusion protein. This is superior to directly fusing the large GFP molecule directly to the target protein, which could compromise its functional activity.

The combination of split-GFP with CRISPAIN T leads to a moderate-throughput and simple workflow, which is summarized in Figure 6.2C and nicknamed “GlowPAINT”. The user simply selects an edit to be made and designs and clones a target-specific sgRNA that induces a dsDNA break just upstream of the native stop codon. Cloning of the sgRNA is required in all plasmid-based CRISPR workflows, and it is the *only* target-specific molecular biology that must be completed for GlowPAINT. Furthermore, because the tags themselves are included on the universal donor plasmid, a single sgRNA allows any tag to be added to the target gene, providing flexibility across experimental modalities without requiring additional sub-cloning for each application. The only other requirement for GlowPAINT is that

the cell-line to be edited must express GFP¹⁻¹⁰, either constitutively or from a transiently transfected helper plasmid. Since modern FACS instruments can complete dozens of sorts in a single afternoon, this facilitates isolation of either polyclonal (all green cells into a flask) or monoclonal (single green cells into 96-well plates) edited cell lines in a moderate-throughput manner. GlowPAINT minimizes the required molecular biology, bypasses manual single-cell dilution and screening of the clones, couples the selection phenotype directly to the target protein, works with dozens of tags, and parallelizes to allow multiple edits to be made simultaneously. The rest of Chapter 6 describes my efforts to create, set-up, and pilot this GlowPAINT method.

6.4 Generation of GFP¹⁻¹⁰ Cell Lines

The first step towards deploying GlowPAINT was the generation of a GFP¹⁻¹⁰ expressing cell line to be used as the foundational line for editing. To achieve this, I used the *AAVS1*-targeting CRISPR/Cas9 knock-in methodology discussed briefly in Section 6.3 [526]. I started by ordering the plasmid AAVS1-Puro-Tet3G-3xFLAG-Twin-Strep from Addgene, which allows a Puro^R cassette and a Dox-inducible open-reading frame to be added to the safe-harbor *AAVS1* locus using CRISPR/Cas9. However, I needed to make some changes to the vector before use. First, and most obviously, I needed to clone GFP¹⁻¹⁰ into the Dox-inducible open reading frame; less obviously, I also needed to remove the 3xFLAG-TwinStrep tag from the plasmid backbone. Even with an inducible GFP¹⁻¹⁰ bearing these tags, I feared that it could complicate the use of edited cells lines containing a targeted knock-in of a 3xFLAG or TwinStrep tag, so I felt it best to remove these epitopes. Secondly, the CRISPAINT universal donor plasmids contain a Puro^R cassette downstream of the inserted tag to enrich for edited cells; therefore, I needed to change the resistance marker in pAAVS1-Puro-Tet3G-3xFLAG-Twin-Strep. I selected Zeocin, which was reported to be the best selection marker for human cells in a head-to-head comparison [532]; furthermore, I had successfully used Zeo^R as a marker in FAS2FURIOUS, so I had the required coding sequence and the reagent itself on-hand. These changes required me to manually construct a new vector, called pAAVS1-ZEO-TETON-GFP1-10, which is diagrammed in Figure 6.3A. When used in conjunction with a standard *AAVS1*-targeting sgRNA, the pAAVS1-ZEO-TETON-GFP1-10 vector results in the Zeo^R, chromosomal insertion of a Dox-inducible, un-tagged GFP¹⁻¹⁰ into the safe harbor *AAVS1* locus. A schematic of the final edited locus contained in

the GFP¹⁻¹⁰ expressing cell lines generated with pAAVS1-ZEO-TETON-GFP1-10 as the HDR repair template is provided in Figure 6.3B.

I constructed pAAVS1-ZEO-TETON-GFP1-10 in four “chunks” using a HiFi assembly. I describe the method in detail in Section 6.1.1 and summarize the generation of the chunks in Figure 6.3C; the chunks themselves are also annotated directly on the plasmid map in Figure 6.3A. In short, I excised the backbone from the original pAAVS1-Puro-Tet3G-3xFLAG-Twin-Strep (containing Amp^R, origin of replication, and the *AAVS1*-targeting homology arms) using a restriction digest; I called this Chunk 2. I then prepared the GFP¹⁻¹⁰ insert (Chunk 1), the Zeo^R cassette (Chunk 3), and the internal TetON system (also from pAAVS1-Puro-Tet3G-3xFLAG-Twin-Strep, Chunk 4) using PCR with primers containing HiFi compatible overhangs. I then assembled these four components to generate the final plasmid pAAVS1-ZEO-TETON-GFP1-10, which I confirmed with colony PCR and then Sanger Sequencing. With this newly constructed vector in hand, all I needed in order to generate the GFP¹⁻¹⁰ cell line was a plasmid containing the previously reported *AAVS1*-specific sgRNA. I describe the procedure for cloning sgRNAs into the plasmid eSpCas9(1.1), which also contains the high efficiency SpCas9 enzyme eSpCas9(1.1), in more detail in Section 6.6 [533]. For vectors containing a cloned sgRNA, I use the nomenclature eSpCas9(1.1)-GENEID-sgX; therefore, the *AAVS1*-targeting plasmid was called eSpCas9(1.1)-AAVS1-sg1.

For pilot experiments, I decided to focus on Expi293 cells (which are suspension-adapted, engineered HEK293 cells from Thermo Fisher that are commonly used for mammalian protein expression) and wild type HeLa cells. Note that neither of these cell lines express the innate immune receptor TLR4; however, Expi293 and HeLa cells are highly amenable to transfection, editing, and expansion, making them reliable cell lines for piloting the GlowPAINT method before moving into more challenging immune cell lines such as THP1. As described in detail in Section 6.1.3.1, I transfected 1×10^6 cells in 6-well dishes with both eSpCas9(1.1)-AAVS1-sg1 and pAAVS1-ZEO-TETON-GFP1-10; as negative controls, I also transfected cells with each plasmid by itself or with empty transfection reagent. After 48 hours to allow time for dsDNA break generation and HDR repair to occur, I added Zeocin and maintained the selection pressure for two weeks.

At this point, all cells in the negative control wells were dead, whilst the cells transfected with both plasmid components were resistant to Zeocin as expected. At this stage, using the bulk selected pools, I extracted gDNA and ran a standard

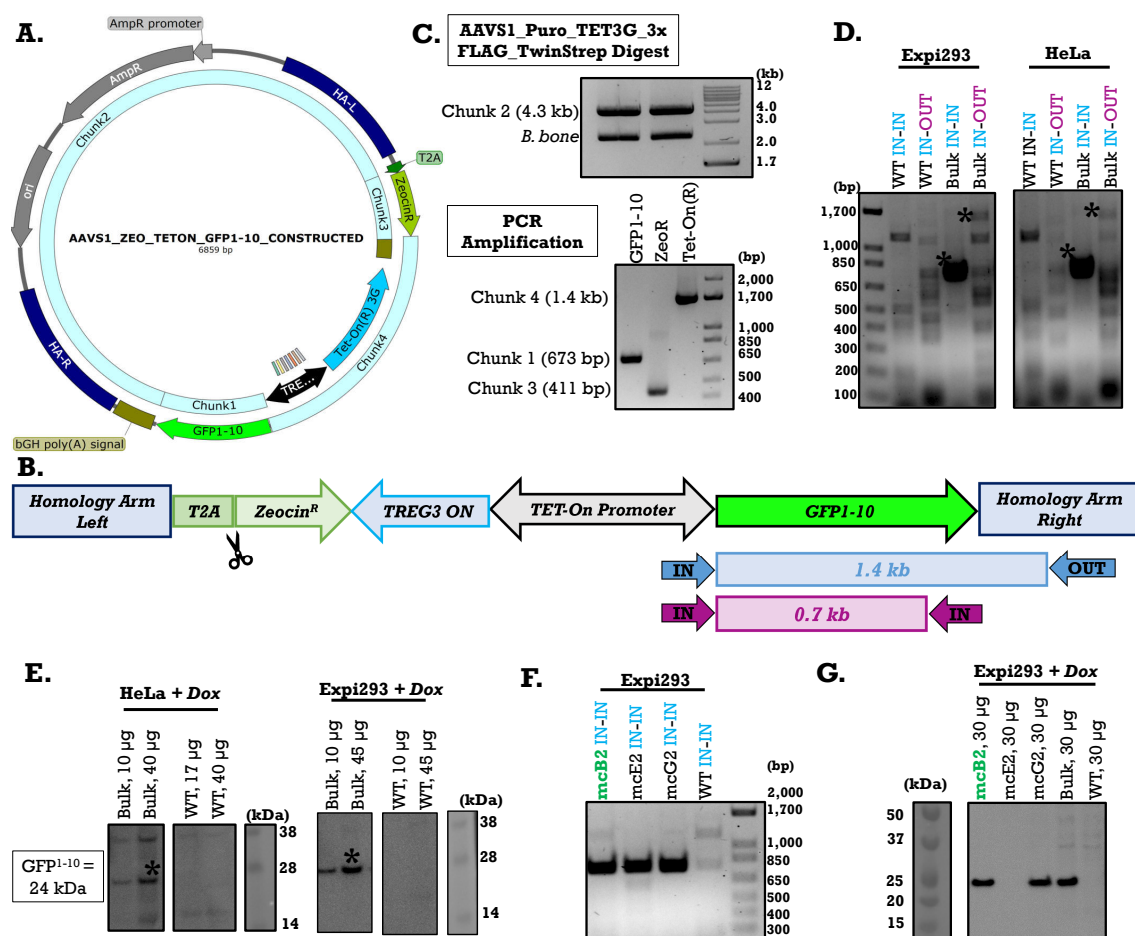


Figure 6.3 Generation of GFP¹⁻¹⁰ Expressing Cell Lines. **A.** pAAVS1-ZEO-TETON-GFP1-10 plasmid map. **B.** AAVS1 genetic locus after editing with CRISPR/Cas9 editing using pAAVS1-ZEO-TETON-GFP1-10 as the HDR template. **C.** Agarose gel showing the generation of the four “chunks” used to assemble the homemade pAAVS1-ZEO-TETON-GFP1-10. **D.** IN-IN and IN-OUT PCR using template gDNA from the GFP¹⁻¹⁰ bulk pools. Correct amplicons, corresponding to the predicted amplicons included in B., are marked with an asterisk. **E.** Anti-GFP western blot from the Dox-induced bulk pools. Specific GFP¹⁻¹⁰ band at 24 kDa marked with black asterisk. **F.** IN-IN PCR from the monoclonal Expi293 GFP¹⁻¹⁰ cell lines following limiting dilution and suspension adaptation. **G.** Anti-GFP western blot from the monoclonal lines. mcB2, marked in green, became the core cell line Expi293^{GFP10}.

PCR using two sets of insert-specific primers, which are annotated in Figure 6.3B. One set, called “IN-IN”, contains a forward primer which anneals in the TET-ON promoter and another which anneals at the 3'-end of the GFP¹⁻¹⁰ insert. For so-called “IN-OUT” PCR, the reverse primer anneals to the genomic homology arm, wholly outside of the inserted cargo. For shorter insertions, “OUT-OUT” PCR (with primers annealing inside both homology arms, wholly outside the inserted region)

is often used; however, with such a large insert (almost 3 kb), OUT-OUT PCR was impractical and failed to give an amplicon in my hands (data not shown). As a negative control, I also extracted gDNA from wild-type, unedited cells. As Figure 6.3D demonstrates, both the Expi293 and HeLa bulk pools contained the expected insert, indicated by an IN-IN band at 700 bp and an IN-OUT band at 1.4 kb. The IN-OUT band was much fainter, as PCR amplification from gDNA becomes more challenging as the length of the amplicon increases; however, comparison to the negative control WT amplification shows that this 1.4 kb IN-OUT band was indeed insert-specific. As a further confirmation, I added Dox to the bulk pools to induce the expression of GFP¹⁻¹⁰ and, after 72 hours, ran a western blot with an anti-GFP antibody (described in detail in Section 6.1.4) to look for expression of the GFP¹⁻¹⁰ protein. In both bulk pools, I observed a specific GFP¹⁻¹⁰ band at the expected size; importantly, this band was not present in an equal mass of lysate derived from the unedited wild type cells (Figure 6.3E).

Therefore, at the gDNA and protein level, both the HeLa and Expi293 polyclonal bulk pools contained GFP¹⁻¹⁰ inserted into the *AAVS1*-locus. This represented a successful replication of previously reported efforts to insert cargo into the *AAVS1* safe harbor; comfortably, it also validated that the homemade pAAVS1-ZEO-TETON-GFP1-10 performed exactly as desired. At this point, since the HeLa cells contained less pronounced GFP¹⁻¹⁰ expression (the HeLa blot in Figure 6.3E was exposed for significantly longer than the Expi293 blot), I froze down the HeLa bulk pool and moved forward with selecting a monoclonal, GFP¹⁻¹⁰ Expi293 cell line. I subjected the bulk pool to limiting dilution into two 96-well plates and, after two weeks, obtained three monoclonal cell lines (named mcB2, mcE2, and mcG2) which were scaled up and suspension adapted (both to simplify splitting and because Expi293 cells are engineered to prefer suspension cultures). I repeated the IN-IN PCR, which showed that all three monoclonal lines contained the GFP¹⁻¹⁰ insert (Figure 6.3F). However, only two of the lines (mcB2 and mcG2) contained GFP¹⁻¹⁰ protein upon Dox induction (Figure 6.3G). It is likely that mcE2 contained a frame-shift mutation or that the locus was transcriptionally silenced—this underscores the importance of validating knock-in at both the genetic and protein level. It also provided a nice example of how indirect selection markers (the Zeo^R, in this case) can be imperfect, and why the direct incorporation of GFP¹¹ into the target protein is a more reliable way to search for edited cells. In any case, I moved forward with mcB2 for subsequent editing experiments; hereafter, I will refer to this cell line as Expi293^{GFP10}.

6.5 Adding GFP¹¹ to CRISPAIN Universal Donors

The next step in the development of GlowPAINT was the addition of GFP¹¹ into the universal donor plasmids. The original CRISPAIN plasmid set contained 19 different universal donors with diverse functional tags (Addgene Kit #1000000086) [511]. For GlowPAINT, I selected seven universal donors for incorporation of GFP¹¹: HaloTag[®] (an engineered haloalkane dehalogenase which can bind a number of ligands for different applications)[534], NanoLuc[®] (a bright luciferase for bioluminescence-based assays)[535], ProteoTuner[™] (a degron which can be used to induce degradation of the tagged protein)[536], 3xFLAG and StrepII (affinity epitopes for purification), AviTag[™] (for biotinylation), and SpyTag (a small adhesin-derived peptide that can form a covalent isopeptide bond with a partner protein)[537]. These seven tags cover a diverse set of applications, including cellular proximity and interaction assays, western blotting, affinity enrichment, immobilization, and microscopy. Four of the seven (3xFLAG, StrepII, AviTag[™], and SpyTag) can be used for endogenous isolation of the tagged protein; the SpyTag is especially versatile, as the choice of the binding partner can facilitate either covalent immobilization or affinity-enrichment and elution [538]. ProteoTuner[™] facilitates inducible knock-out experiments, while the combination of the HaloTag[®] and NanoLuc[®] can be used to set up the increasingly popular NanoBRET assay [539]. The GlowPAINT method is theoretically compatible with any universal donor vector; this subset was chosen to be diverse and useful for pilot experiments.

The architecture of the universal donors merits detailed consideration, so the CRISPAIN schematic is reproduced for clarity in Figure 6.4A with an arrow denoting the GFP¹¹ insertion point. Recall that the frame selector contains an sgRNA that leads to linearization of the universal donor just upstream of the tag itself, creating a linear dsDNA NHEJ repair template. Downstream of the tag itself, the universal donor contains a T2A self-cleaving peptide and then the Puro^R protein, which is released from the translated fusion protein and used for selective enrichment of edited clones. Therefore, I decided to add the GFP¹¹ peptide after (3') the functional tag but before (5') the T2A site. This ensures that the GFP¹¹ peptide is retained directly on the target fusion protein along with the functional tag. As an example, I provide the nucleotide-level architecture of the pGlowPAINT-StrepII-PuroR vector in Figure 6.4; note that a short GlySer linker separates the 16 residue GFP¹¹ tag from both the StrepII region and the T2A peptide to ensure sufficient flexibility for complementation with the core GFP¹⁻¹⁰ protein.

All six of the other pGlowPAINT vectors are identical in terms of the placement of the GFP¹¹ and the GlySer linker. I constructed the vectors using PCR as described in detail in Section 6.1.2. Essentially, I designed primers containing overhangs corresponding to the GFP¹¹ insert that annealed to the pCRISPAINT template vector. The forward primer annealed in the T2A peptide and contained the 3'-half of the GFP¹¹ insert; the reverse primer annealed in the tag region and contained the 5'-half of the GFP¹¹ insert. A single PCR reaction created an amplicon containing the entire, linearized CRISPAINT vector with half of the GFP¹¹ inserted at the 5'-end and the other half at the 3'-end. Circularization of the amplicon using T4 Polynucleotide Kinase and T4 DNA Ligase generated the final pGlowPAINT product. To check for GFP¹¹ insertion via colony PCR, I ordered a reverse primer (GFP¹¹-seq-r1) annealing within GFP¹¹ and a forward primer annealing upstream of the tag in the vector backbone (PSITE-seq-f1). These primers lead to the amplification of the entire tag (which controls the amplicon size) during colony PCR, but only if GFP¹¹ is added downstream. As Figure 6.4B demonstrates, the approach was highly efficient, and I obtained correctly sized clones (annotated with an asterisk) for all seven tags on my first attempt. I miniprepmed and sequenced two clones for each tag and in all cases validated the in-frame insertion of the GFP¹¹ tag in the correct location. A representative sequencing chromatogram, from the generation of pGlowPAINT-StrepII-PuroR, is aligned with the predicted sequence in Figure 6.4C.

6.6 Cloning sgRNAs into eSpCas9(1.1)

At this point, I had successfully generated the Expi293^{GFP10} cell line and constructed the GFP¹¹-containing pGlowPAINT universal donor plasmids. The last step before piloting GlowPAINT in full was the cloning of target-specific sgRNAs. As a destination vector for the sgRNAs, I used eSpCas9(1.1) (Addgene #71814), which contains the high-efficiency eSpCas9(1.1) enzyme along with a U6-promoter driven sgRNA scaffold. This plasmid is diagrammed in Figure 6.5A. Before use, the target-specific 20 bp protospacer, which is specific to the gene of interest and defines the precise location of the induced dsDNA break, must be added to the vector. To achieve this, two complementary ssDNA oligos are ordered, containing the 20 bp protospacer along with 5'-overhangs corresponding to the *BbsI* cut site in the eSpCas9(1.1) vector. The oligos are annealed to each other with a simple thermal annealing reaction and then directly ligated into the *BbsI*-linearized eSpCas9(1.1) destination vector to generate the final target-cutting CRISPR/Cas9 plasmid. The architecture of the

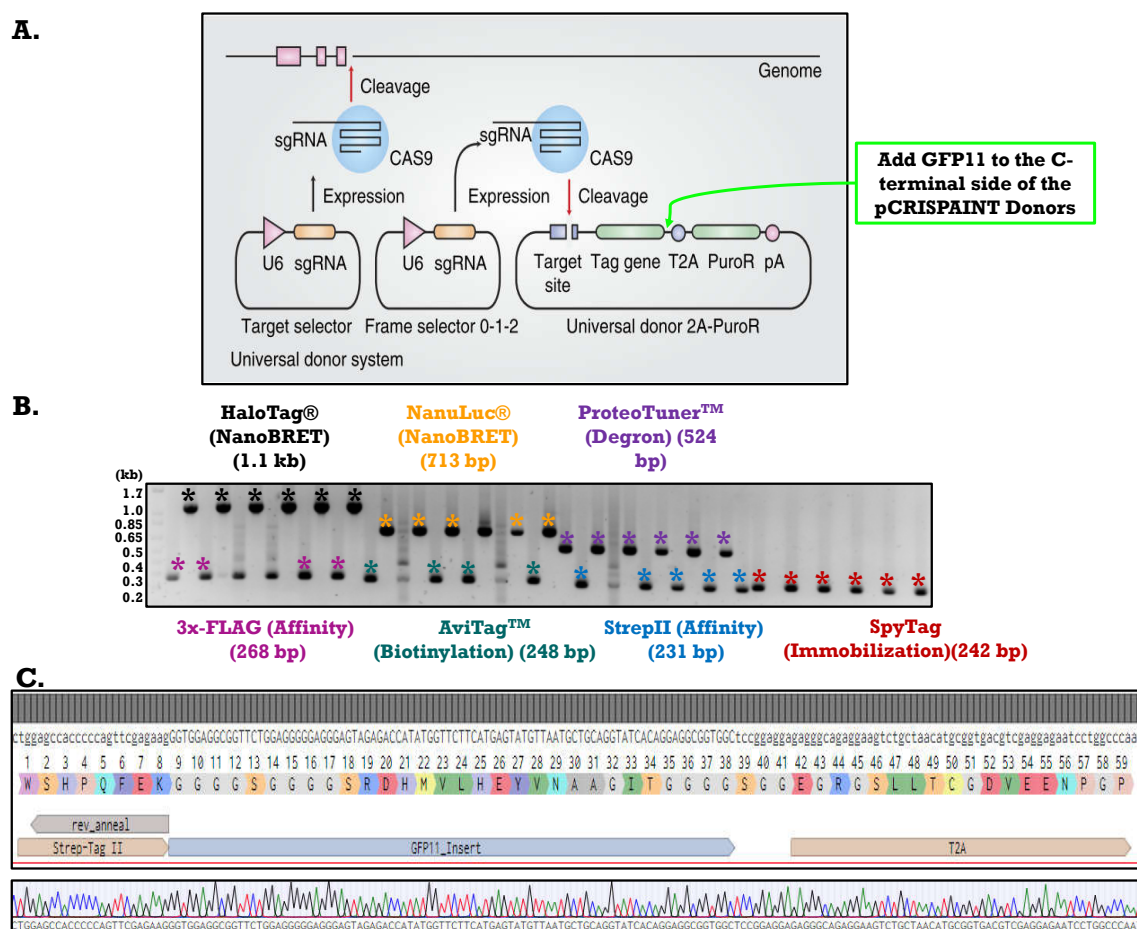


Figure 6.4 Adding GFP¹¹ to pCRISPAINT Donors. **A.** CRISPAINT schematic reproduced from Figure 6.2A. The arrow indicates the region downstream of the tag where the GFP¹¹ peptide was inserted. **B.** Colony PCR following the addition of GFP¹¹ to seven different CRISPAINT donors to generate the pGlowPAINT donor library. Clones marked with an asterisk were the expected size. Seven different tag moieties were modified with GFP¹¹. **C.** Representative sequencing chromatogram for the newly created pGlowPAINT-StrepII-PuroR. The GFP¹¹ was inserted 3' to the tag and 5' to the self-cleaving T2A sequence (which releases the PuroR marker). A short GlySer linker was added before and after the GFP¹¹ peptide. All seven GFP¹¹-containing pGlowPAINT donors were similarly confirmed with Sanger sequencing.

U6-promoter driven sgRNA within eSpCas9(1.1) is displayed in Figure 6.5B; the protospacer targeting region is highlighted in dark green, along with the *BbsI*-cut sites and the 5', 4 bp oligo overhangs used for ligation. The details of this standard sgRNA cloning method, which were provided to me by Dr. Hublitz, are provided in Section 6.1.3.

To exemplify the potential of pGlowPAINT to be used in a moderate-throughput

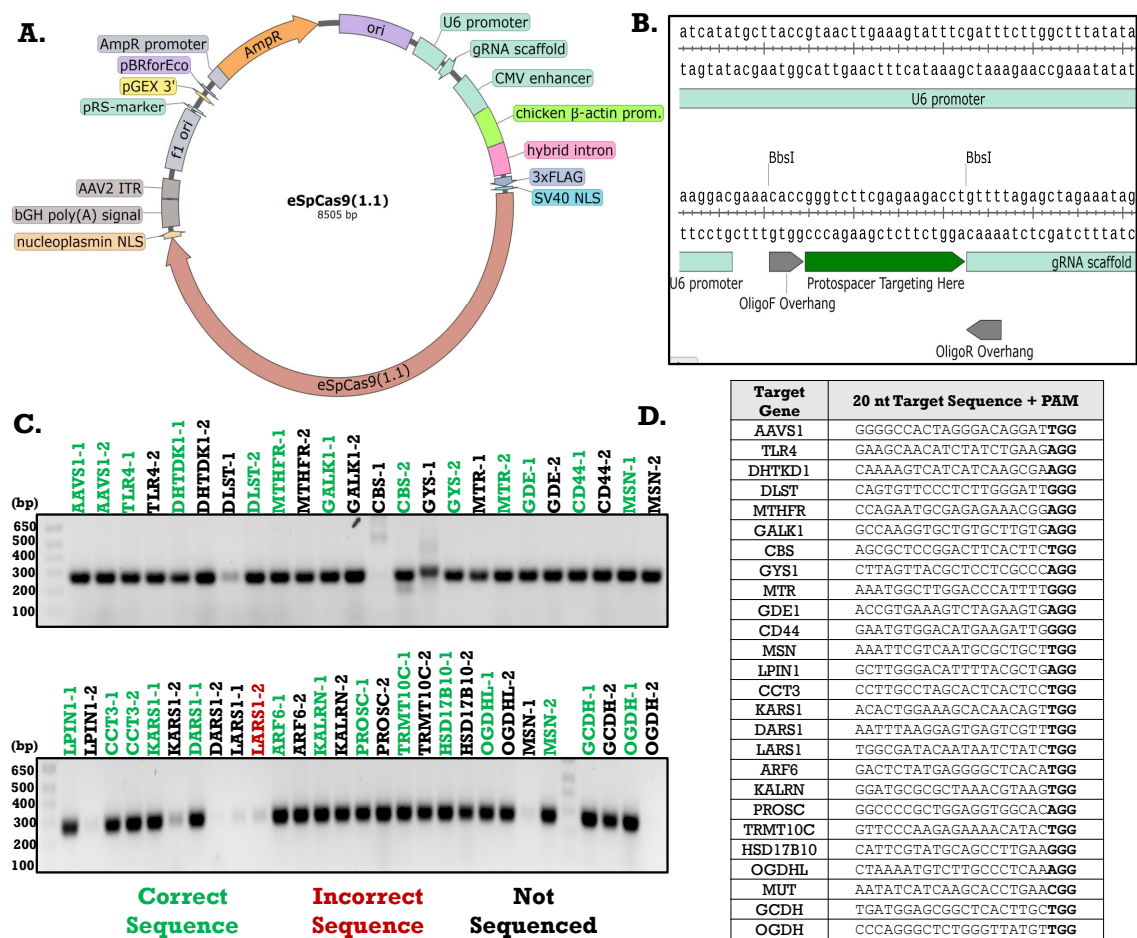


Figure 6.5 Cloning Target-Specific sgRNAs. **A.** The eSpCas9(1.1) plasmid. **B.** Zoomed-in view of the sgRNA cloning region, showing the U6-promoter, the targeting region, and the downstream sgRNA scaffold. Target-specific protospacers are inserted at the *BbsI* cut sites. **C.** Modified colony PCR, using hU6-seq-f1 and the target-specific reverse oligo, to check for correct insertion of the protospacer following restriction-ligation cloning. Clones verified with Sanger Sequencing, accounting for 96% of the clones sequenced, are displayed in green. **D.** Initial target gene list and the protospacer targeting sequence used as recommended by Schmid-Burgk *et al.* [511]. Note that the adjacent PAM is required for dsDNA cleavage by Cas9, but it is not included in the sgRNA itself.

format, I assembled a target list containing 26 different genes. A full list of the genes and their corresponding protospacer sequences is provided in Figure 6.5D. The targeting sequence for each selected gene was taken directly from the original pCRISPANT paper, wherein Schmid-Burgk *et al.* provided recommended protospacers for all human genes [511]. The recommended targeting sequences strictly minimize the number of residues which are truncated from the C-terminus of the target protein as a result of PAM restrictions. I will return shortly to my mistaken

choice to use the targeting regions closest to the stop codon rather than selecting the best predicted sgRNA within a reasonable distance in Section 6.7. The gene target list contains not only the standard *AAVS1*-targeting sgRNA, but also TLR4, metabolic enzymes, amino-acyl tRNA synthetases, and guanidine nucleotide binding proteins. The list was assembled with input from our group and other groups in the department, with an eye towards a specific knock-in application for each gene; the more generic motivation was to explore the parallelizability of the GlowPAINT method.

I next ordered and cloned the sgRNAs according to the standard method supplied by Dr. Hublitz except for one important improvement. For one-off cloning of a few sgRNAs at a time, Dr. Hublitz recommended miniprepping and sequencing three different clones for each sgRNA. However, for a higher-throughput system (here I handled 26 sgRNAs at once), this “prep and sequence” approach to finding a correct clone would have been tedious and expensive. Instead, I designed a colony PCR approach that picks out correct clones. I first designed and ordered a generic forward primer, called hU6-seq-f1, that sits inside the U6 promoter upstream of the protospacer insertion site. As the reverse primer for colony PCR, I used the target-specific reverse oligo used to clone the targeting region in the first place. Because I ordered the oligos in a 96-well plate, it was relatively trivial to add a unique reverse primer to each colony PCR reaction. If the picked colony contained the correct, target specific protospacer, a 300 bp amplicon would be generated. Figure 6.5C shows that this colony PCR approach was highly effective at identifying clones with the correct insertion. I picked two clones for each target, and in all but one case I observed a strong 300 bp amplicon for at least one clone. I submitted a subset of the clones for Sanger Sequencing and observed that *all* clones which generated a clear 300 bp amplicon contained the correctly inserted targeting region. The only incorrect sequencing run was generated from the LARS1-2 clone, which clearly gave a weaker amplicon (red text in Figure 6.5C). Therefore, my colony PCR approach was essentially 100% effective and represents a simple but important methodological improvement that facilitates high-throughput cloning of sgRNAs without the need to miniprep and sequence every resulting plasmid. In theory, this approach should work for 96 sgRNAs at once; in fact, I handled these 26 targets in 96-well plates throughout the procedure.

6.7 Checking sgRNA Cleavage Efficiency

With the sgRNAs cloned, I now had all three components to deploy GlowPAINT: a GFP¹⁻¹⁰ cell line, GFP¹¹-containing NHEJ donors, and target-specific sgRNAs. However, before piloting the entire system, I wanted to check the on-target cutting efficiency of the sgRNAs I had cloned. Selecting the right sgRNA is a challenge that many online tools, including the Broad Institute’s excellent CRISPick, attempt to solve by predicting on- and off-target cutting efficiency scores [521, 540, 541]. However, it is still best to experimentally validate that the selected sgRNA actually cleaves the target dsDNA in the cell line of interest—quite simply, if the sgRNA does not cut, there will be no insertion, no matter how elegant the design of the repair template. sgRNA efficiency is especially important to validate in the context of piloting a new system. In a few one-off pilot experiments (data not shown) with just the CCT3 sgRNA, I saw no insertion. However, without evidence that the sgRNA actually cleaved the target, I could not distinguish between no cutting (an easily corrected sgRNA problem) and no NHEJ-mediated knock-in (a more fundamental problem with the approach).

Therefore, I decided to test the cleavage efficiency of all of my cloned sgRNAs using a mismatch cleavage assay, often referred to as a Surveyor assay [542]. The assay is relatively simple and is diagrammed in Figure 6.6A. First, the relevant cells are transfected with just the eSpCas9(1.1)-sg*X* plasmid, which will introduce a dsDNA break. The more efficient the sgRNA, the more dsDNA breaks that will occur. The cells will then locally repair this break which, in the absence of a supplied repair template, will create indels at the cut site. After 48 hours, the gDNA is extracted from the mixed pools and the targeted region is amplified with PCR using target-specific primers. The primers sit 200–400 bp upstream and downstream of the predicted cut site, creating a moderately sized amplicon that contains both the wild type allele (from unedited cells) and the alleles containing indels (from cut cells). The amplicon is then denatured and re-annealed, which leads to the creation of heteroduplex DNA with a central mismatch at the site of the indel. T7 Endonuclease I, also marketed as the “Surveyor nuclease”, specifically recognizes this mismatched DNA and catalyzes cleavage of the heteroduplex DNA into two pieces. Cleavage of the amplicon can be semi-quantitatively tracked on a gel—the more efficient the original sgRNA, the higher fraction of edited cells and therefore a higher percentage of T7 Endonuclease I cleaved heteroduplex DNA. As a negative control, the target region from cells not transfected with the sgRNA is amplified, re-annealed, and

treated. The entire assay takes only five days to complete and provides a reliable way to rank sgRNAs and confirm that they are cleaving the gDNA as desired. The only drawback is that individual primer sets must be ordered for each cut.

To test the sgRNAs cloned in Section 6.6, I transfected 800,000 Expi293^{GFP10} cells with each eSpCas9(1.1)-sg X plasmid in 24 well plates. After 48 hours, I extracted gDNA using a commercially available reagent (DNAzolTM) and then completed the mismatch cleavage assay as described in Section 6.1.5. I used two sets of primers to amplify each target site and obtained amplicons of the expected size for 17 of the targets (65% of the tested sgRNAs) using touchdown PCR and the high fidelity Q5[®] Polymerase. As a negative control, I used the same primer sets to amplify the target region from wild-type gDNA. Re-annealed amplicons from the 17 successfully amplified targets and their respective negative controls were subjected to T7 Endonuclease I treatment and then analyzed on an agarose gel, shown in Figure 6.6B. Disappointingly, only three sgRNAs (marked in green) clearly cleaved their target: eSpCas9(1.1)-GDE1-sg1, eSpCas9(1.1)-LPIN-sg1, and eSpCas9(1.1)-CD44-sg1. For all the other sgRNAs, including the TLR4 sgRNA, I did not observe the appearance of two smaller cleaved bands relative to the negative control T7 Endonuclease I treatment. I used eSpCas9(1.1)-AAVS1-sg1 to generate the Expi293^{GFP10} cell line, so I also considered this a working sgRNA even though I could not obtain a specific gDNA amplicon during this experiment. This means that, of the 26 sgRNAs I originally cloned, I could only experimentally validate on-target cleavage for four of them (just over 15% efficiency). These results are summarized in a table for clarity in Figure 6.6C.

The low efficiency of the sgRNAs I cloned stems from my decision to use the target sequences recommended in the original pCRISPAIN^T paper. These guides were chosen to minimize the distance between the cut site and the stop codon in an effort to reduce the number of residues lost at the C-terminus in the fusion protein. However, especially given the low efficiency I observed for sgRNAs designed this way, this is a poor way to select a target region. Instead, focus should be placed on guides predicted to cut with high efficiency *within a biologically relevant truncation region*; for example, I would prefer to use a high efficiency guide that introduces an eight-residue truncation over a low efficiency guide that cuts directly at the C-terminus. If I were to repeat these experiments in the future, I would manually design sgRNAs for each target using the CRISPick tool and pick the best scoring target sequence within a tolerable window of the C-terminus, rather than simply

picking the most proximal target sequence. Furthermore, I would order multiple (2–4) sgRNAs per target gene to increase the probability of finding a strong cutter.

These results prove that sgRNA design and sgRNA validation are critical steps in the GlowPAINT workflow, which is one of the reasons why I describe the method as moderate-throughput. With the requirement for multiple sgRNAs manually designed per target and validated with a mismatch cleavage assay, the potential throughput drops significantly (in my opinion to probably 12 edits and 24–36 sgRNAs at a time). However, I pressed onward with validating the core features of the GlowPAINT method by working with just the three sgRNAs I validated in the mismatch cleavage assay.

6.8 Piloting the GlowPAINT Method with GDE1

Human glycerophosphodiester phosphodiesterase 1 (GDE1; UniProt Q9NZC3) is a 37.7 kDa, ER-localized transmembrane enzyme responsible for the hydrolysis of phosphodiester bonds in glycerophosphodiesters such as glycerophosphoinositol and glycerophosphoethanolamine [543]). GDE1 is involved in the regulation of GPCR signalling pathways, is ubiquitously expressed (including in HEK293 cells), and contains a C-terminal cytoplasmic tail that is separated from the ER-localized catalytic domain by the ER-membrane [544]. These facts, combined with my empirical observation that eSpCas9(1.1)-GDE1-sg1 efficiently cuts the *GDE1* target locus (see Figure 6.6), made GDE1 a useful target for piloting the full GlowPAINT method. Towards this end, with the assistance of Douglas Ferreira, another DPhil student in the group, I transfected Expi293^{GFP10} cells with the entire three-plasmid GlowPAINT system: eSpCas9(1.1)-GDE1-sg1 to cut, Frame Selector Plasmid to linearize the repair template, and pGlowPAINT-3xFLAG-PuroR to serve as the repair template for NHEJ-mediated knock-in of 3xFLAG-GFP¹¹-T2A-PuroR to the C-terminus of GDE1. Note that I tested all three Frame Selectors in this experiment; Frame Selector #0 is predicted to be most likely to generate an in-frame insertion of the tag at the GDE1 locus. However, because NHEJ can result in 1–3 bp deletions, it is not impossible for a different Frame Selector to generate an in-frame knock-in. Schmid-Burgk *et al.*'s data demonstrated that the Frame Selector predicted to be in-frame is always the best performing, but that the other Frame Selectors can still give lower efficiency in-frame knock-in [511]. A schematic of the *GDE1* locus is provided in Figure 6.7A, showing the PAM, the targeting region, and the predicted site of

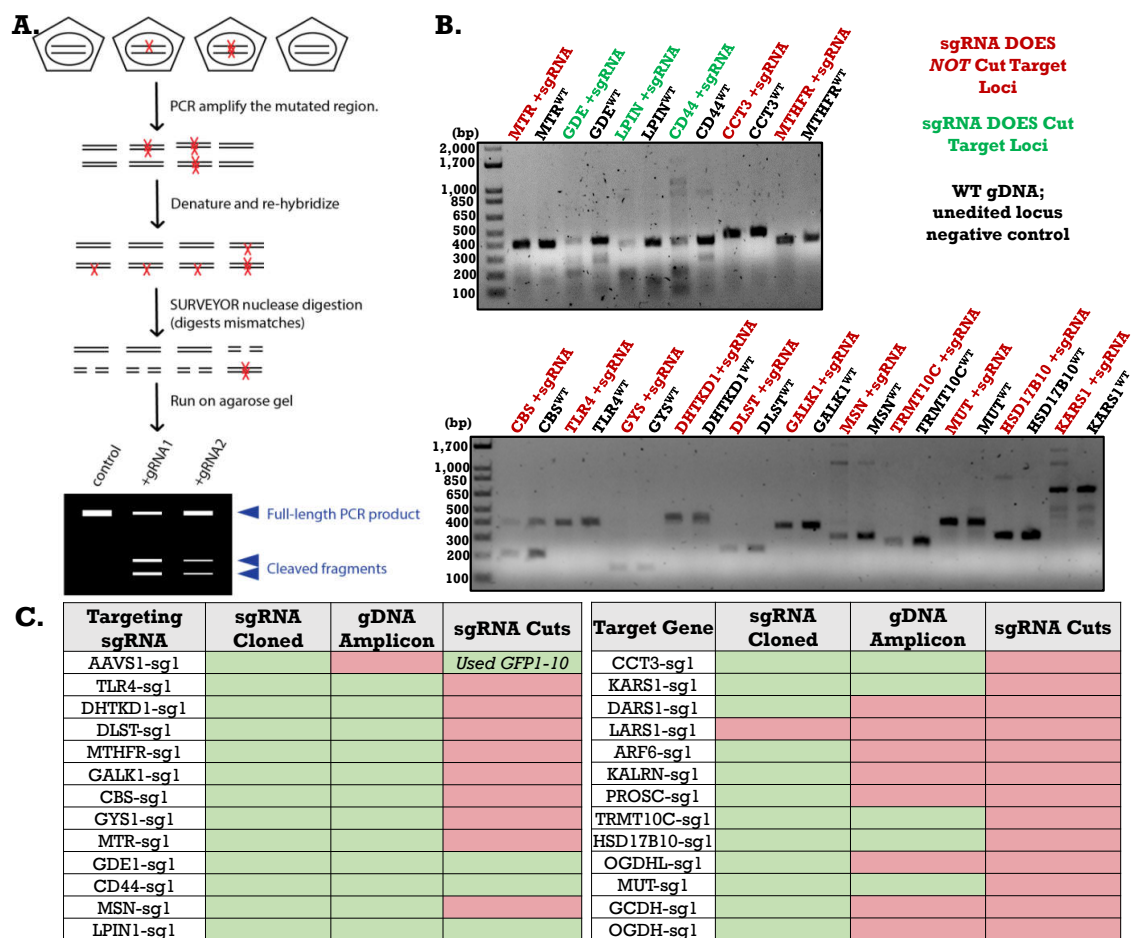


Figure 6.6 Checking sgRNA Efficiency. **A.** Schematic of the mismatch cleavage assay, reproduced from <https://blog.addgene.org/crispr-101-validating-your-genome-edit>. **B.** Results of a mismatch cleavage assay from 17 eSpCas9(1.1)-sgRNA vectors transfected into Expi293^{GFP10} cells. Guides which induced significant dsDNA breaks are highlighted in green. **C.** Overview of the performance of the 26 eSpCas9(1.1)-sgX vectors cloned and tested herein. Only four of the original 26 sgRNAs were experimentally confirmed to induce dsDNA breaks.

Cas9-mediated cleavage (which results in the truncation of a single residue, Phe331, from the C-terminus of the GDE1 protein). As described fully in Section 6.1.6, I also transfected individual plasmid components and empty transfection reagent to serve as negative controls, and after 48 hours I subjected the transfected bulk pool cells to selection with $2 \mu\text{g mL}^{-1}$ puromycin to generate polyclonal cell lines (“pc”). This selection step should enrich for cells containing the desired insertion.

I first examined the *GDE1* genomic locus using OUT-OUT PCR with primers that anneal 200 bp upstream and downstream of the predicted cut site. The results, shown in Figure 6.7B, suggested that knock-in occurred in the presence of Frame

Selector #0 but not with Frame Selector #1, #2, or the mock transfection. As discussed in Section 6.3, the CRISPAIN method results in the knock-in of the entire donor template plasmid, which is over 3 kb in length. The large, smeary band in the pcFS0 band is consistent with a large knock-in but an insufficiently long extension time in the PCR reaction; it could also be due to nuclease contamination causing degradation of the PCR amplicon. To validate this result, I performed an anti-FLAG western blot on the un-selected bulk pools, the results of which are shown in Figure 6.7C. While the band is faint (as expected given that these samples were taken from the un-selected bulk pools, where the edit will be a small minority of the overall cell population), there is a specific GDE1-3xFLAG-GFP11 band at the expected molecular weight when Frame Selector #0 was used. Note that the strong band at approximately 37 kDa is due to the presence of 3xFLAG-mCherry, which is included as a marker on the frame selector plasmids.

With suggestive evidence at the gDNA and protein level, I next subjected the pc cell lines (from Frame Selector #0 and #1) to limiting dilution to generate monoclonal (“mc”) cell lines. I generated five cell lines (mc1–mc5) which, after selection and scale-up, survived a second round of puromycin selection. I then subjected lysates from these lines to an anti-FLAG-tag western blot and observed a FLAG-tagged product in mc4 at approximately 60 kDa and 180 kDa. GDE1 has a molecular weight of approximately 40 kDa, but membrane proteins often run higher than their predicted MW, especially since GDE1 is known to be glycosylated [544]. The larger band could correspond to a trimer. In any case, I reproduced this result by scaling up mc4 (to 120×10^6 cells) and then performing an anti-FLAG pull-down followed by an anti-FLAG western blot. Yet again, I observed a strong FLAG-tagged product at approximately 60 kDa that was not present in the pull-down from an equal number of wild-type, unedited cells. Unfortunately, due to instrument access limitations during COVID-19, I could not analyze mc4 on a confocal microscope or FACS instrument to confirm Dox-inducible fluorescence. However, this pilot experiment does show that the pGlowPAINT vectors with a GFP¹¹ insertion behave as expected and allow targeted knock-in of a C-terminal affinity tag.

6.9 Conclusions and Future Directions

In this Chapter, I presented my efforts to establish a CRISPR-Cas9 system for the endogenous tagging of human genes for diverse applications. Motivated by the

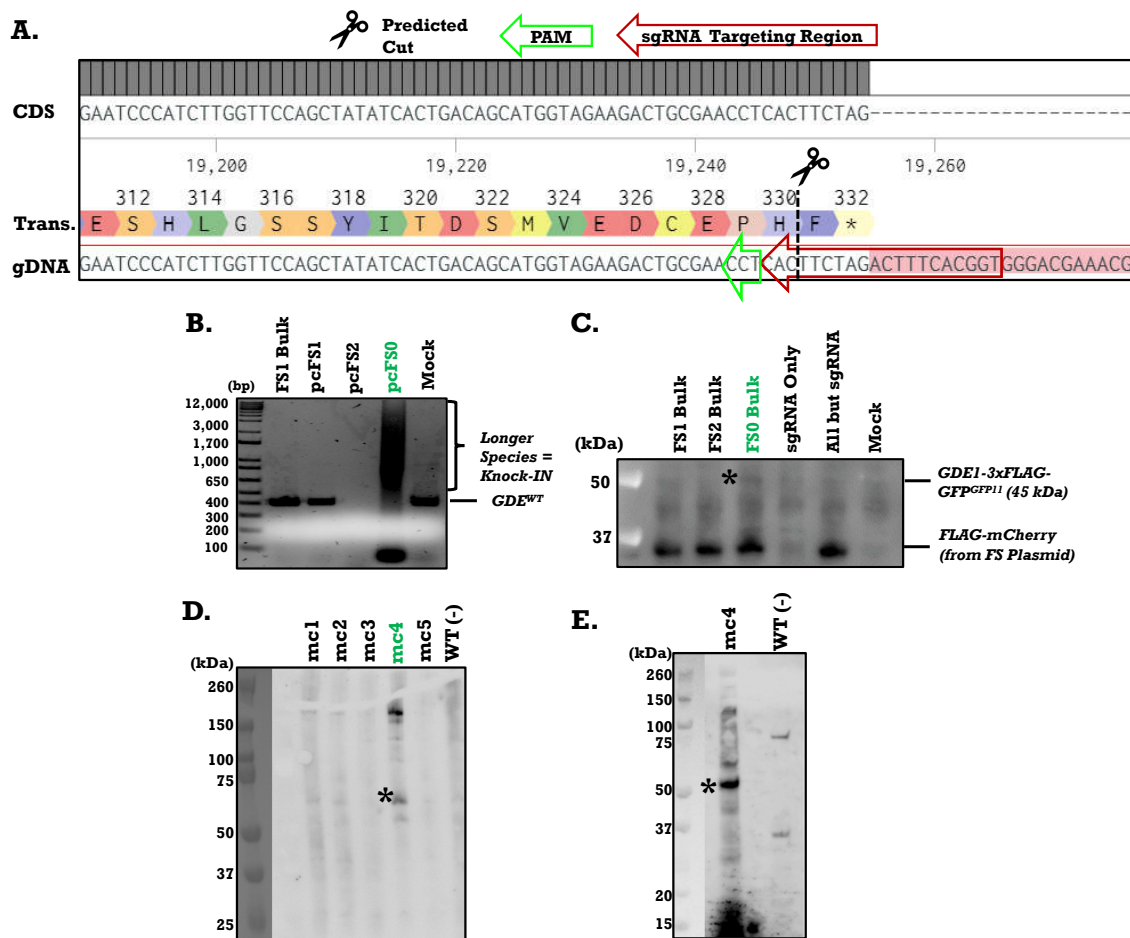


Figure 6.7 GlowPAINT Pilot with *GDE1*. **A.** Schematic of the *GDE1* genomic locus, focusing on the targeted C-terminus. The PAM, targeting sequence, and predicted Cas9-induced cut site are annotated. **B.** Expi293^{GFP10} cells subjected to OUT-OUT PCR after transfection with the *GDE*-targeting GlowPAINT components. All three frame selector plasmids (FS0, FS1, and FS2) were tested using pGlowPAINT-3xFLAG-PuroR as a repair template and eSpCas9(1.1)-*GDE1*-sg1 to generate a dsDNA break. “pc” denotes polyclonal lines following puromycin selection. **C.** Anti-FLAG western blot from the un-selected bulk pool cells as in B. The “sgRNA Only” negative control was transfected with only eSpCas9(1.1)-*GDE1*-sg1; the “All but sgRNA” negative control was transfected with all GlowPAINT components *except* for eSpCas9(1.1)-*GDE1*-sg1. **D.** Anti-FLAG-tag western blot from five Puro^R monoclonal cell lines. **E.** Eluates from an anti-FLAG pull-down using 120 x 10⁶ cells from mc4 or wild-type Expi293 cells analyzed with an anti-FLAG blot.

success of my colleague Dr. Kelly on human TRiC, and with an eye towards generating tagged TLR4 cell lines to address the unanswered questions from Chapter 5, I combined two previously reported methodologies (split-GFP and CRISPAIN) to generate a novel tagging platform which I coined GlowPAINT. In principle, Glow-

PAINT allows the C-terminal insertion of diverse functional epitopes to any human gene in a moderate-throughput manner without requiring a target-specific HDR template or tedious limiting dilution experiments. Using a homemade plasmid, I created a Dox-inducible GFP¹⁻¹⁰ cell line to serve as the foundation for editing. I next introduced the GFP¹¹ peptide to seven different pCRISPAIN universal donor plasmids to generate a useful set of pGlowPAINT plasmids. Finally, after cloning, testing, and triaging over two dozen gene-specific sgRNAs, I demonstrated knock-in of 3xFLAG-GFP¹¹ to the *GDE1* locus, providing proof of principle for the GlowPAINT method. Relative to the other results in this thesis, the work presented in Chapter 6 is the most incomplete due to the impact of COVID-19, and therefore there are obvious and important next steps that must be performed prior to routine deployment of GlowPAINT. First, the FACS-based selection of correctly edited, GFP-positive cells must be validated, and the feasibility of parallelizing the method with available instrumentation should be evaluated. Because of instrument limitations during the UK lockdown, I was unable to validate the efficiency of selecting for edited cells with FACS. While in principle this should be trivial, this validation is essential because the FACS-enrichment removes the requirement for single-cell dilution and screening, a rate-limiting step in the procedure. Next, the method should be tested across more diverse genes and more challenging cell lines; particularly, efforts to generate a THP1 monocyte cell line with tagged TLR4 (my ideal end-goal for GlowPAINT) would be valuable. Thirdly, the edited cell lines require downstream quality control to confirm that GlowPAINT does not radically change the expression level, sub-cellular localization, or biological activity of the tagged proteins. This validation is required in all CRISPR-workflows; however, it is especially important for GlowPAINT, which results in the incorporation of an uncommon bi-functional tag to the protein of interest. Finally, an important extension of the method would be to generate cell lines with GFP¹⁻¹⁰ expressed in specific sub-cellular compartments, including the mitochondria, ER, and nucleus. Using the vectors and cell-lines presented herein, only proteins with a cytosolic C-terminus can be selected by fluorescence (since the GFP¹⁻¹⁰ counterpart is expressed in the cytosol).

However, despite requiring further work-up, the overall design, molecular biology, and proof-of-principle experiments presented in Chapter 6 represent a meaningful step forward for CRISPR/Cas9-mediated endogenous tagging. Given the growing utility and accessibility of cryo-EM and proteomics, alongside the clear practical advantages of GlowPAINT versus other systems (especially in terms of simplicity,

throughout, and reliability of the fluorescence-based selection), I envision GlowPAINT becoming increasingly useful in the future. The aforementioned next steps are already being actively pursued by others in the laboratory, justifying my optimism that GlowPAINT will soon be fully optimized and ready for routine deployment as a standard workflow for CRISPR/Cas9-mediated generation of endogenously tagged human cell lines. In fact, in March 2021, another group independently developed the working principles of GlowPAINT, using ssODN's to insert mNeon¹¹ to over 1,000 endogenous loci for microscopic localization studies and miniaturized AP-MSMS experiments [545]. In a game-changing tour de force, this technology allowed Cho *et al.* to generate “OpenCell”, a proteome scale “cartography” of human protein localization and interactomes. However, Cho *et al.* do not report the use of the cell lines for endogenous purification or structural studies, leaving room for GlowPAINT to make an impact in the cryo-EM space. Importantly, this recent publication emphasizes the strength of the approach developed herein, and highlights the overall direction of the field towards high-throughput endogenous tagging to study biological systems.

Chapter 7

The COVID-19 Interlude: Antigen Production for Serological Assays

Chapter 7 concludes the results section of this thesis with my second method development vignette, focused on COVID-19-related protein production. For more than four full-time months during the height of the UK lockdown (March–August 2020), I stopped working on my thesis project and was instead seconded to the SGC’s Biotechnology Group in order to purify COVID-19 relevant proteins, including the SARS-CoV-2 spike protein (S) and receptor-binding domain (RBD), along with monoclonal antibodies. In Chapter 7, I will focus exclusively on my work on the S protein. Alongside the talented Dr. Alejandra Fernandez-Cid, the Senior Scientist in the Biotech Group, I co-managed a team of over twenty scientists who worked together to purify large quantities of S. This purified S antigen was required to support Oxford’s National Serology Testing Center, an automated platform that provides surveillance serological testing to support COVID-19 research and policy-making. I personally managed all in-lab scientists on a daily basis for this four-month period and drove key improvements in the purification of S, a protein that proved quite challenging. All of the protein-level data in this Chapter are my own, and I personally completed all of the key optimization experiments to ensure reproducibility. However, even though I directed and completed much of the work, throughout this section I will use first-person plural pronouns to emphasize the collaborative team effort, which was assembled under emergency conditions and on a short timeline. This is especially important for the cell culture work, the majority of which was handled by other scientists whom I will mention by name where appropriate.

To date, we have purified over 470 milligrams of S antigen which continues to be used in a serological assay that matches, and in most cases out-performs, other commer-

cially available diagnostics. We also paved the way for scaling up production of the S protein in bioreactors for other applications via the generation of over-expressing monoclonal cell lines. My work on secreted protein expression systems, discussed in Chapter 3, and diverse protein purification techniques, discussed primarily in Chapter 2, proved especially valuable for this effort and highlights how a commitment to robust methodology in one area can be leveraged to make an impact elsewhere. Furthermore, I include this Chapter under the University’s COVID-19 Student Policy as an explanation for the unfinished loose ends in Chapter 5 and 6 which, in a non-COVID world, I would have tied up before finishing my time at Oxford.

7.1 Materials and Methods

7.1.1 Transient Expression and Purification of SARS-CoV-2 S Protein

On Day 1, Expi293FTM cells (Life Technologies Limited A14528) were split to 1×10^6 cells per mL in roller bottles and grown at 170 RPM, 37 °C, and 8% CO₂ in 1 L cultures in FreeStyleTM 293 medium (Thermo Fisher 12338-018). On Day 2 the cells, now at approximately 1×10^6 cells per mL, were divided into 500 mL cultures in roller bottles prior to transfection. The transfection steps described below are for one flask containing 500 mL of cells. In a Falcon tube, 250 µg of DNA was added to 10 mL of pre-warmed Opti-MEM I Reduced-Serum Medium (Gibco 31985062). In a second tube, 1.5 mg of 25K Linear-PEI (PolySciences 23966-1) was added to 10 mL of pre-warmed Opti-MEM. These amounts correspond to 0.5 mg L^{-1} plasmid DNA and a 1:6 (w:w) ratio of DNA:PEI. The plasmid DNA mixture was added step-wise to the PEI mixture, mixed gently, and incubated for 30 min at room temperature. At this point, 10 mM sodium butyrate was added as an enhancer to the cells. Finally, the DNA/PEI mixture was added drop-wise to the cells, and the transfected cells were allowed to grow at 170 RPM, 30 °C, and 8% CO₂ for 68–72 hours. After three days, the cell viability was checked and recorded (routinely over 90% at harvest) and the medium was harvested via centrifugation at $900 \times g$ for 15 min at 18 °C. The clarified medium was transferred to a clean centrifuge tube and centrifuged again at $4,000 \times g$ for 15 min at 18 °C to further remove aggregates and cellular debris. At this point, the clarified supernatant was buffered by the addition of concentrated PBS to a final concentration of 1X PBS (pH = 7.2–7.4) prior to purification.

To purify the S protein from the clarified, buffered supernatant, 4 mL L⁻¹ loose, pre-equilibrated Ni²⁺ IMAC Sepharose (GE) was added. The samples were incubated for at least two hours at room temperature with gentle stirring to maximize capture of S. The bound resin was collected in large gravity flow columns and then washed with 50 CV of 1X PBS, 50 CV of 1X PBS + 25 mM imidazole, and 50 CV of 1X PBS + 40 mM imidazole prior to elution with 3 x 2 CV of 1X PBS + 300 mM imidazole. At this point, S-containing fractions were pooled, placed on ice, and maintained at 4 °C for the rest of the purification. S was desalted by serial dialysis back into 1X PBS, ensuring that the final [imidazole] was less than 0.1 mM. At this stage, depending on the downstream application, 0.1% Tween-20 was added to the sample to reduce protein loss during concentration. Using a Vivaspin Centrifugal Concentrator (GE), the sample was gently concentrated in increments of 3 min at 2,500 x *g*, with a mixing step between every spin. Once the concentration reached at least 1 mg mL⁻¹, the sample was removed from the concentrator and centrifuged for 30 min at 21,000 x *g* to remove aggregates. After removing a small sample for quality control, the protein was flash frozen in LN₂ and stored at -80 °C prior to use.

For quality control, 10 µg of the purified protein was analyzed via SDS-PAGE for purity and the putative S band was confirmed by in-gel tryptic digest MSMS. To confirm the presence of a monodisperse trimer, 10 µg of the purified protein was diluted in 110 µL of PBS and analyzed via analytical SEC on a Sepax 3000 column with a DionexTM Ultimate 3000 UHPLC System. The trimeric S eluted between 2.55–2.85 min, consistent with the expected trimer based on comparison to Gel Filtration Standards (Bio-Rad). 6 µg of S per reaction was also analyzed in triplicate via DSF using the method reported in Section 2.1.11.

7.1.2 Cloning S-c23

To clone S-c23, S-c16 was transferred into the Neo^R pOPINTTGneo vector backbone by PCR. pOPINTTGneo was linearized according to the manufacturer's (NEB) recommendations via double digest with *KpnI*-HF and *NotI*-HF followed by gel extraction of the backbone. S-c16, containing Met1–Gln1208 (UniProt: P0DTC2) of the SARS-CoV-2 S protein and a C-terminal 8His-Twin-Strep tag, was amplified by PCR using Q5[®] Polymerase with primers containing HiFi-compatible overhangs generated by NEBuilder (S-c23-f1 + S-c23-r1; sequences below). The 3.8 kb amplicon was checked on an agarose gel and then incubated overnight with 1 µL *DpnI*

at 37°C prior to spin purification. 50 ng of linearized pOPINTTGneo was HiFi-assembled according to the manufacturer’s recommendations (NEB) with 50 ng of the purified PCR amplicon, transformed into Mach 1 cells, and finally selected on Carb^R LB-agar plates. Five colonies were checked with Colony PCR using the vector-specific primers pOPINTTGneo-seq-f1 + pOPINTTGneo-seq-r1. One colony with an insert of the correct size (approximately 4 kb) was miniprepmed and the in-frame insertion of the S-c16 coding region was confirmed with Sanger Sequencing using pOPINTTGneo-seq-f1 + pOPINTTGneo-seq-r1. The rest of the coding sequence was confirmed using a set of overlapping internal sequencing primers (SPIKE-seq1 through SPIKE-seq7). The new construct was called S-c23 and used to generate stable cell lines as described in Section 7.1.3.

New primers used for PCR amplification of S-c16. HiFi overhangs in lowercase, S-c16 annealing regions in uppercase:

S-c23-f1 (5'-ggttgcgtagctgaaaccggtacc-
ATGTTTCGTGTTCTGCTGGTGCTGCTGCCTCTGG)

S-c23-r1 (5'-ggagccacccccaatttgagaaataatga-
GCTCGAGGCCGGCAAGGCCGGATCCC

Sequencing primers used for Colony PCR and Sanger Sequencing:

pOPINTTGneo-seq-f1 (5'-ATGGTTCGAGGTGAGGTGTG)

pOPINTTGneo-seq-r1 (5'-CTAGCTAGAGATCTCCGAGGG)

Internal sequencing primers to confirm the entirety of the S coding sequence:

SPIKE-seq1 (5'-ATGTTTCGTGTTCTGCTGGTGCT)

SPIKE-seq2 (5'-CGCGTATATTCGTCTGGCTAA)

SPIKE-seq3 (5'-GAAGGGCATCTACCAGACCA)

SPIKE-seq4 (5'-GGCTCATGGTGTAGGCGAT)

SPIKE-seq5 (5'-CTGCATGGCGAAAGGGATTT)

SPIKE-seq6 (5'-AGATCCAAGCAAGCCTAGCA)

SPIKE-seq7 (5'-TGTTCTGTCACGTGACCTAC)

7.1.3 Generation of S-Expressing Monoclonal Cell Lines

Stable cell lines for S-c17, S-c19, and S-c23 were generated via random genetic integration. Prior to transfection, S-c17 was linearized with *SacI*-HF, S-c19 was

linearized with *BsaI*, and S-c23 was linearized with *MluI* according to the manufacturer's recommendations (NEB). Both the circular and the linearized plasmid DNA were heat sterilized for 20 min at 80 °C prior to transfection. 10 mL of adherent Expi293F™ cells, grown in DMEM + 10% FBS + 1X Gentamicin, were seeded at 2×10^6 cells per mL and transfected with 5 µg plasmid and 15 µL FugeneHD® according to the manufacturer's recommendations (Promega). The cells were allowed to recover for 48 hours post-transfection before the addition of fresh medium containing 200 µg mL⁻¹ G418. After three additional days, the [G418] was raised to 800 µg mL⁻¹ and maintained for at least 8 additional days (until the cells in a negative control flask were visibly dead). At this point, the G418-resistant cell populations were trypsinized and a portion of the polyclonal bulk pools were frozen. The rest of the cells, from both the linearized and circular parental plasmid DNA, were subjected to limiting dilution and plated at a density of 0.5 cells per 200 µL into multiple 96-well plates. After 14 days, wells containing a single adherent colony were transferred into 24-well suspension blocks with 100% FreeStyle™ medium and grown for three days. At this point, samples of the supernatant were taken and 16 µL were analyzed via western blot as described in Section 6.1.4, except for the use of anti-Penta-His-HRP conjugate primary antibody (Qiagen 34460) to detect the His-tag on S. Monoclonal lines demonstrating the highest level of S expression were frozen in 10% DMSO and scaled up for large-scale purification. The best lines came from linearized S-c17 and S-c23 and circular S-c19 plasmid DNA.

7.2 Introduction to SARS-CoV-2 S and Serology Testing

As of March 2021, over 120 million people have been infected by the SARS-CoV-2 virus, and the resulting disease (COVID-19) has killed over 2.5 million people. SARS-CoV-2 belongs to the broad family of beta coronaviruses, which also includes the “near miss” pandemic viruses MERS-CoV (2002) and SARS-CoV (2012); SARS-CoV-2 shares 50% and 79% sequence identity, respectively, with these previous coronaviruses [546]. SARS-CoV-2 infects human cells via the high affinity interaction between the viral particle Spike (S) protein and the human transmembrane receptor angiotensin I converting enzyme 2 (ACE2) [547, 548]. S binding to ACE2 via the receptor-binding domain (RBD) within the S1 subunit triggers a large conformational rearrangement into the “post-fusion state”, which is facilitated by the proteolytic cleavage of S at the S1/S2 and S' sites by the human serine protease

TMPRSS2 [549, 550, 551]. This conformational change induces fusion of the viral and cellular membranes and intracellular release of the 29.9 kb SARS-CoV-2 mRNA genome, which contains the 16 non-structural proteins, four structural proteins, and the six or more accessory proteins required for viral replication and extracellular release of newly packaged virus particles [552].

S, being the largest surface-exposed viral protein and the viral molecule directly responsible for cellular entry, is the primary antigen for neutralizing antibodies generated by the adaptive immune system; as a result, it is the protein presented by all four of the approved vaccines (Moderna, Pfizer, Astra Zeneca, and Johnson & Johnson) [553, 554]. The sub-domain architecture (including the S1/S2 and S2' proteolytic processing sites) and overall structure of the S ectodomain, solved with remarkable speed by Wrapp *et al.*, is reproduced in Figure 7.1A [555]. S exists as a homotrimer with a combined molecular weight of 423 kDa, and contains a large, surface-exposed ectodomain (Gln14–Gln1208) and a transmembrane domain with a short internal C-terminal extension (Pro1213–Thr1273). The RBD (green) exists in a dynamic equilibrium between the “up” and “down” conformations at the top of S; it is in the up conformation that RBD binds to ACE2. Unlike other coronaviruses, the SARS-CoV-2 S protein has a unique polybasic furin cleavage site (“RRAR”) inserted at residue 682, which makes S more susceptible to processing by host proteases and contributes to the increased infectivity of SARS-CoV-2 [556]. S is also heavily glycosylated, which contributes to immune evasion by shielding more than 40% of the surface of S from circulating antibodies. This feature is emphasized in Figure 7.1B, which is reproduced from Grant *et al.*'s application of molecular dynamics to simulate the behavior of S glycoprotein [557]. Grant *et al.* used experimentally determined glycan structures, derived from glycoproteomics, at each of the 22 N-glycan sites on S protein as purified from HEK293 cells [558]. Therefore, the structure as presented in Figure 7.1B is most representative of the complete S glycoprotein that we purified from Expi293 cells in this section.

Because of the importance of S as the primary antigen for neutralizing antibodies, the vast majority of SARS-CoV-2 serological assays are “anti-S”, in that they detect the presence of S-binding antibodies from patient-derived sera [559, 560]. Serological assays are orthogonal to PCR-based diagnostics, which can only detect the presence of viral RNA during an ongoing infection. On the other hand, serological assays are often used for epidemiological purposes to survey the population and calculate the proportion of previously infected individuals in an area; these studies can also help

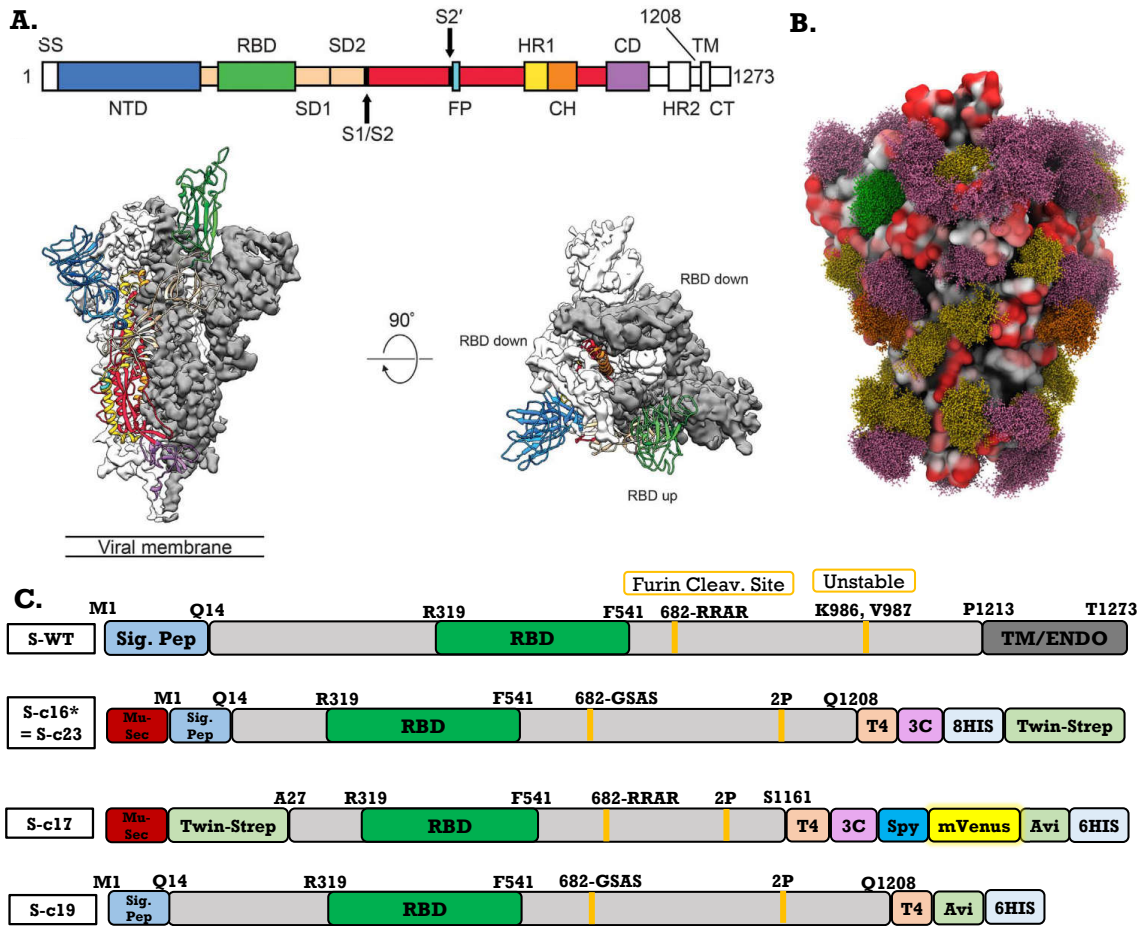


Figure 7.1 The SARS-CoV-2 S Protein. **A.** Sub-domain architecture and overall structure of pre-fusion S, reproduced from Wrapp *et al.* [555]. The ACE2-binding RBD in the “up” conformation is displayed in green. **B.** Extensive glycosylation envelope on S, reproduced from Grant *et al.* [557]. **C.** Comparison of WT S with the key constructs used herein. Stabilizing sequence changes at the polybasic furin cleavage site (682-RRAR to 682-GSAS) and the 2P mutation (K986P, V987P) are marked with a vertical yellow line.

to determine the asymptomatic infection rate, a key variable in computational modelling of the disease. Longitudinal serological assays can also be used to determine the strength and duration of anti-S antibody responses, which informs on the risk of re-infection and length of protection after an infection [560]. In the context of vaccine development, which was pursued in earnest by the Jenner Institute at Oxford while our work was ongoing, serological assays serve as a primary readout for the immune-activating efficacy of vaccine candidates, especially during pre-clinical and Phase 1 studies [561]. However, despite their vital importance to both research and policy making, serological assays are more challenging than PCR assays to develop, leading to conflicting results from different assay platforms and commercially

available kits [562, 563].

Therefore, in the early months of the pandemic (March 2020), the Nuffield Department of Medicine decided to establish a National Serology Testing Center in Oxford, lead by a large consortium of scientists that came to be known as The National SARS-CoV-2 Serology Assay Evaluation Group. Well over a hundred scientists worked together to develop a fully automated, in-house serology testing platform which outperformed other tests in an unbiased head-to-head comparison and performed with 99.1% sensitivity and 99% selectivity [564]. The platform was eventually scaled-up in partnership with Thermo Fisher, and it continues to be used to evaluate vaccine efficacy and to provide UK-wide data to the Office of National Statistics as part of the broader COVID-19 Infection Survey (<https://www.ox.ac.uk/news/2020-11-10-oxford-ramps-covid-19-testing-capability>). As part of the development consortium, the SGC Biotechnology Group was tasked with developing a method for the large-scale purification of S, which continues to be used as the sole antigen in the ELISA-based assay. For four months, I stopped working on my thesis project and instead served as the daily supervisor of the SGC's S purification efforts, working as the liaison between the in-lab scientists and the at-home management team (primarily Dr. Nicola-Burgess Brown and Dr. Alejandra Fernandez-Cid). The method we developed, built according to the methodological improvements described herein, has been used to purify over 470 mg of high quality S protein to date; furthermore, our S protein continues to be used in the Thermo Scientific Omnipath™ Combi SARS-CoV-2 IgG ELISA to test up to 50,000 patient samples per day.

7.3 Method for the Transient Mammalian Expression and Purification of S

In this Chapter, I will present data on four critical S ectodomain constructs, which are diagrammed in Figure 7.1C. The constructs include S-c16, the key serology testing construct used for transient S over-expression, and the identical S-c23, a selectable version of S-c16 used to generate stably expressing monoclonal cell lines (see Section 7.5). S-c16 and S-c23 contain a 3C-cleavable 8His-Twin-Strep tag. We also generated stable cell lines using S-c19, which contains the full-length S ectodomain (Met1–Gln1208) with a His-AviTag, and S-c17, which contains a multi-epitope fluorescent mVenus tag on a truncated S ectodomain (Ser27–Ser1161). Note that all tags were placed at the C-terminus to avoid interference with the N-terminal domain

(NTD; blue in Figure 7.1A). The constructs all contained a mutated furin-cleavage site, which prevents proteolytic degradation of the protein, as well as the “2P” mutation, which stabilizes S in the pre-fusion conformation and increases purified yields [565]. This mutation was inferred from previous work on the MERS-CoV S protein and was included in both the cryo-EM construct by Wrapp *et al.* and the vaccine candidates. Note that subsequent work led to an even more stable and higher yielding SARS-CoV-2 S construct containing six proline mutations and named “HexaPro” [566]. However, because at the start of the project only the 2P construct had been released, all work herein used the 2P mutations. Because the membrane-spanning portion of S was removed to solubilize all constructs, a C-terminal trimerization “T4” motif was used to maintain the biological oligomer [567, 568]. Finally, all constructs we used were secreted from mammalian Expi293 cells, driven by the native signal peptide and/or the orthogonal mu-sec signal sequence. Secretion of S is consistent with the biology of the virus, results in near-physiological glycosylation of S, and is the standard approach in the literature for preparation of S proteins.

At the start of our work to produce the S protein, the best available protocol was published by the Krammer Lab at Mount Sinai using a 2P stabilized construct (Met1–Pro1213) with the furin cleavage site mutated to a single alanine residue [569]. The Krammer Lab shared this construct, coined S-c11, with us; they also shared S-c12, an identical construct except for the lack of the 2P mutation present in S-c11. Both S-c11 and S-c12 contain a thrombin-cleavable T4-6His tag at the C-terminus. The Krammer Lab reported the use of transient transfection in the Expi293TM Expression System (Thermo Fisher), which includes expensive, proprietary transfection reagents, medium, and transfection enhancers. The Krammer Lab purified S by a one-step gravity flow purification on Ni-NTA resin to isolate S with a final reported yield of 2 mg mL⁻¹. However, in pilot experiments (see Section 7.4), we could not reproduce these yields in our in-house system using the less rich FreeStyleTM 293 medium and the cheaper polyethylenamine (PEI) as a transfection reagent. At best, we observed approximately 0.1 mg L⁻¹ with S-c11 and S-c12, and we struggled to pull-down all of the expressed protein with only a 6His tag (data not shown). This provided the impetus to clone the other constructs (S-c16, -c17, and -c19), which were generated by the Ray Owens Group in the Oxford Division of Structural Biology and passed to us. However, the Krammer method still provided the foundation of our optimization experiments, and we actively worked with the Krammer Lab at the beginning of the project.

Like many others, we struggled to express and purify S at the outset. The protein is heavily glycosylated, unstable above 1 mg mL^{-1} , poorly expressed, and prone to killing the Expi293 cells used as an expression host. Furthermore, transient transfection protocols require large amounts of plasmid DNA, up to 1 mg L^{-1} ; in fact, at the outset we were using more plasmid DNA in the transfection than we obtained in final purified protein (w:w). Furthermore, the purification of secreted proteins requires processing large culture volumes, capturing a dilute target protein, and carefully washing away abundant protein contaminants present in the medium. The purity of the final protein and the reproducibility of the method were of paramount importance, since the protein was to be used in serological assays where batch-to-batch variation would cripple the reliability of the diagnostic and compromise longitudinal comparison of assay results. These problems were amplified by the large amount of protein required for our serological assay (which at the peak consumed 10 mg of S protein per week), the short timelines required to establish the assay, and the prevailing lockdown restrictions in the UK. To address these challenges, we dedicated the entirety of the SGC's protein production capacity and experience to the problem, and we optimized every step of the method with a focus on scale, yield, reproducibility, and cost management.

Under my in-lab direction, we optimized the transfection and purification protocol into a reproducible method that yields 1 mg L^{-1} of pure, monodisperse, trimeric, and active S-c16. The final method, which supports the purification of up to 24 mg of S per week, is described in detail in Section 7.1.1 and outlined in Figure 7.2A. In short, Expi293 cells grown in roller bottles with FreeStyle™ 293 medium are transiently transfected with 0.5 mg L^{-1} gigaprepmed plasmid DNA using linear PEI as a transfection reagent at a 1:6 DNA:PEI ratio (w:w). The transfected cells are then treated with 10 mM sodium butyrate as an enhancer and are grown for 72 hours at 30°C with shaking at 150 RPM prior to harvest and buffering of the cell culture medium followed by a two hour incubation at 25°C with 4 mL L^{-1} Ni^{2+} IMAC Sepharose to capture S. A stringent and strictly reproduced wash series totalling over 150 CV is completed to ensure purity prior to elution with 300 mM imidazole, overnight serial dialysis into 1X PBS, and concentration to 1 mg mL^{-1} using centrifugal concentrator units in increments of three minutes per spin. The final sample is then centrifuged for 30 minutes at $21,000 \times g$ to remove aggregates prior to flash freezing for storage. Since the method calls for just a single step IMAC purification, a $10 \mu\text{g}$ sample of the final protein is finally subjected to analytical SEC on a Dionex™ HPLC to validate the presence of a monodisperse trimer.

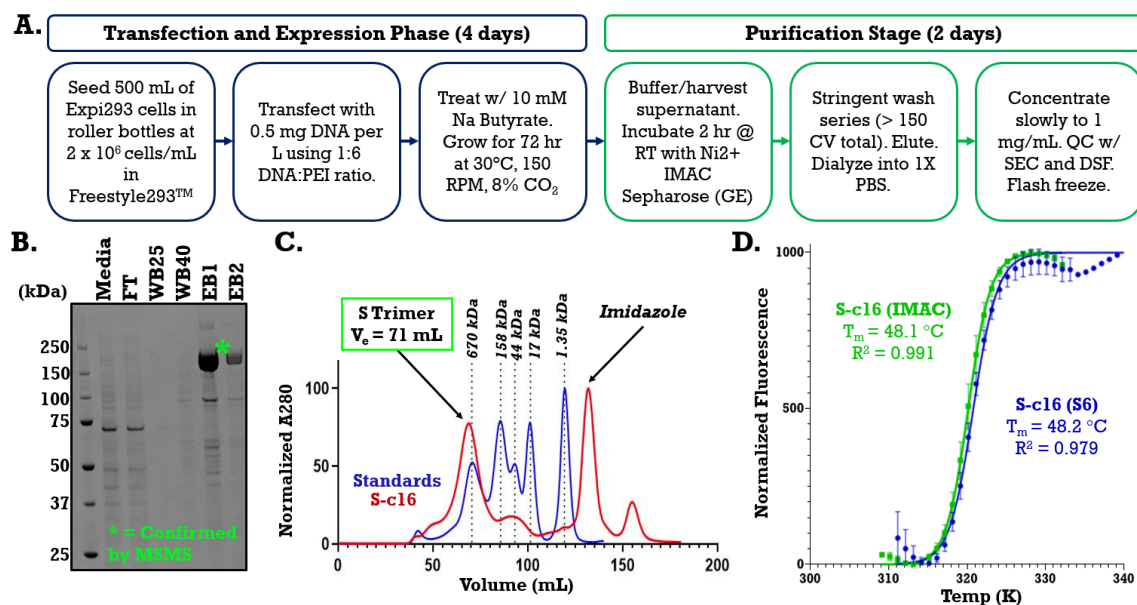


Figure 7.2 Final S-c16 Purification Method. **A.** Overview of the S purification pipeline developed in this chapter. **B.** Final purity of S-c16 after a single IMAC column and stringent washes. An extensive glycosylation envelope causes the 141 kDa S to run at approximately 200 kDa on the gel. **C.** Chromatogram from a Superdex S6 columns showing that the purified S-c16 is monodisperse and trimeric. **D.** DSF analysis of S-c16 shows that the protein is thermally stable ($T_m = 48^\circ\text{C}$) and that the protein melts the same before (green) and after SEC polishing (blue).

This method allows two scientists to work in parallel to purify up to 12 L of transfected cells over a two-day period, and the entire pipeline (from transfection through storage) can be completed twice per week, yielding up to 24 mg S protein per week at maximum production capacity. The final protein is pure (Figure 7.2B), monodisperse and trimeric (Figure 7.2C), and stable ($T_m = 48^\circ\text{C}$; Figure 7.2D), and critically it displays batch-to-batch reproducibility over millions of assays and dozens of protein preps. Furthermore, despite remaining expensive relative to other proteins, our S purification method relies on common medium, transfection reagents, and purification strategies, reducing the per-unit cost to a manageable level and allowing other labs to easily deploy the approach. In addition to its use internally for the serological assay, the final standard protocol (and, in some cases, the protein itself) has been shared with labs worldwide to support various applications, a sampling of which are provided in Section 7.6. In the next section, I walk through key experiments that we performed to optimize this S purification pipeline and settle on the final protocol.

7.4 Optimization of the Method

Because every milligram of S was precious, especially at the beginning of the project, we worked to optimize every conceivable variable in the purification process, totalling over three dozen unique optimization experiments. In this section, I present the critical results that pushed our final yields from less than 0.1 mg mL^{-1} to at least 1 mg mL^{-1} by the end of the project. The cell culture work in this section, including tedious maintenance of huge quantities of Expi293 cells, was supervised by Shubhashish Mukhopadhyay. Note that all expressions in this section were conducted at a minimum 50 mL scale in individual round-bottom flasks (i.e. no small scale 3 mL tests) to ensure that the results would seamlessly translate into larger culture vessels. Furthermore, all of the S protein shown on a gel in this section was quality controlled with SEC and/or DSF to ensure that the final product was consistent with the expected oligomeric state and melting temperature reported in Figure 7.2C and D.

Our first experiment focused on identifying the optimal S expression construct, transfection ratio, and expression temperature. We compared S-c11 and S-c12, obtained from the Krammer Lab, with the in-house generated S-c16. Linear PEI is a cheap and versatile transfection reagent that can facilitate high-efficiency transfection of large quantities of suspension cells; however, depending on the size of the plasmid cargo, the cell line, and the culture medium, different ratios of DNA to packaging PEI perform better than others [570]. Therefore, we compared our two standard ratios: 1:3 and 1:6 DNA:PEI (w:w). Finally, we compared yields after three days of expression at either 30 or 37 °C. Especially for unstable or toxic proteins, lower expression temperatures can increase yields [571], an observation that, in our experience, holds especially true for secreted proteins, which float unprotected in warm cell culture medium prior to harvest. The tested variables required 12 x 50 mL cultures in 250 mL round-bottom flasks, and the medium was pulled down with 500 μL of Ni^{2+} IMAC Sepharose prior to elution with 2 x 1 mL of high imidazole elution buffer. The resulting eluates, from equal volume of samples, were analyzed by SDS-PAGE and the results are presented in Figure 7.3A. Based on the intensities of the final S bands (all of which were confirmed by MSMS), we showed that S-c16 transfected at a 1:6 ratio and expressed at 30 °C was the optimal combination. S-c16 outperformed the two Krammer constructs, while S-c11 (containing the stabilizing 2P mutation) expressed better than S-c12 (without the 2P mutation) as expected. Our results also showed that expression at 30 °C represents a general solution for all

three constructs, and we fed this information back to the Krammer Lab to support their efforts.

Next, we moved forward with optimizing the type of expression vessel and the solid-phase resin used for purification. Since secreted proteins are subject to shear forces in the culture medium, we compared disposable plastic roller bottles (which shake at 170 RPM but are easier to handle and allow a larger number of bottles per incubator) to traditional round-bottom glass flasks (which shake slower at 110 RPM but have a larger footprint and require intensive cleaning and sterilization between runs). We also compared the use of standard Ni^{2+} IMAC Sepharose with Co^{2+} -charged TALON resin. For this experiment, the cells were transfected at a 250 mL scale with S-c16 and, after three days, the harvested medium was split into two halves prior to incubating for two hours with either 1 mL bed volume of Ni^{2+} IMAC Sepharose or TALON. After washing with 50 CV of wash buffer (40 mM imidazole for the Ni^{2+} , 10 mM imidazole for the TALON), S was eluted twice with high imidazole and an equal volume of the eluates were analyzed with SDS-PAGE. The results, displayed in Figure 7.3B, demonstrated that the round-bottom flask and roller bottles yield equivalent amounts of S-c16, allowing us to move forward with the smaller, disposable roller bottles. Secondly, we observed that more S-c16 was recovered using Ni^{2+} IMAC Sepharose for pull-down; while the TALON eluates were slightly purer, as expected, the trade-off in terms of final yield was severe, especially since the Ni^{2+} -eluates were themselves quite pure. Therefore, we selected roller bottles for expression and Ni^{2+} IMAC Sepharose for purification of S-c16.

Subsequently, we compared two different expression medium: FreeStyleTM 293 and Expi293TM. Both mediums are serum free and chemically defined, ensuring batch-to-batch reproducibility. FreeStyleTM medium is significantly cheaper and more widely available for bulk order, but Expi293TM medium is the recommended medium in Thermo's Expi293TM system because it is richer and facilitates longer growths at higher cell densities. In this experiment, we compared two different transfected cell densities (2 and 3 x 10⁶ cells per mL) in Expi293TM medium. We also checked to see if the cells transfected in FreeStyleTM medium could be harvested twice with a feeding step in-between, as a potential way to stretch a single transfection. So, we collected medium three days after transfection (Harvest #1), resuspended the cells in fresh FreeStyleTM, and then harvested again after three more days (Harvest #2). The experiment was conducted at a 500 mL scale for the FreeStyleTM samples and a 280 mL scale for the Expi293TM medium, all using S-c16. The resulting eluates from this experiment, following pull-down on Ni^{2+} IMAC Sepharose, are shown in Figure

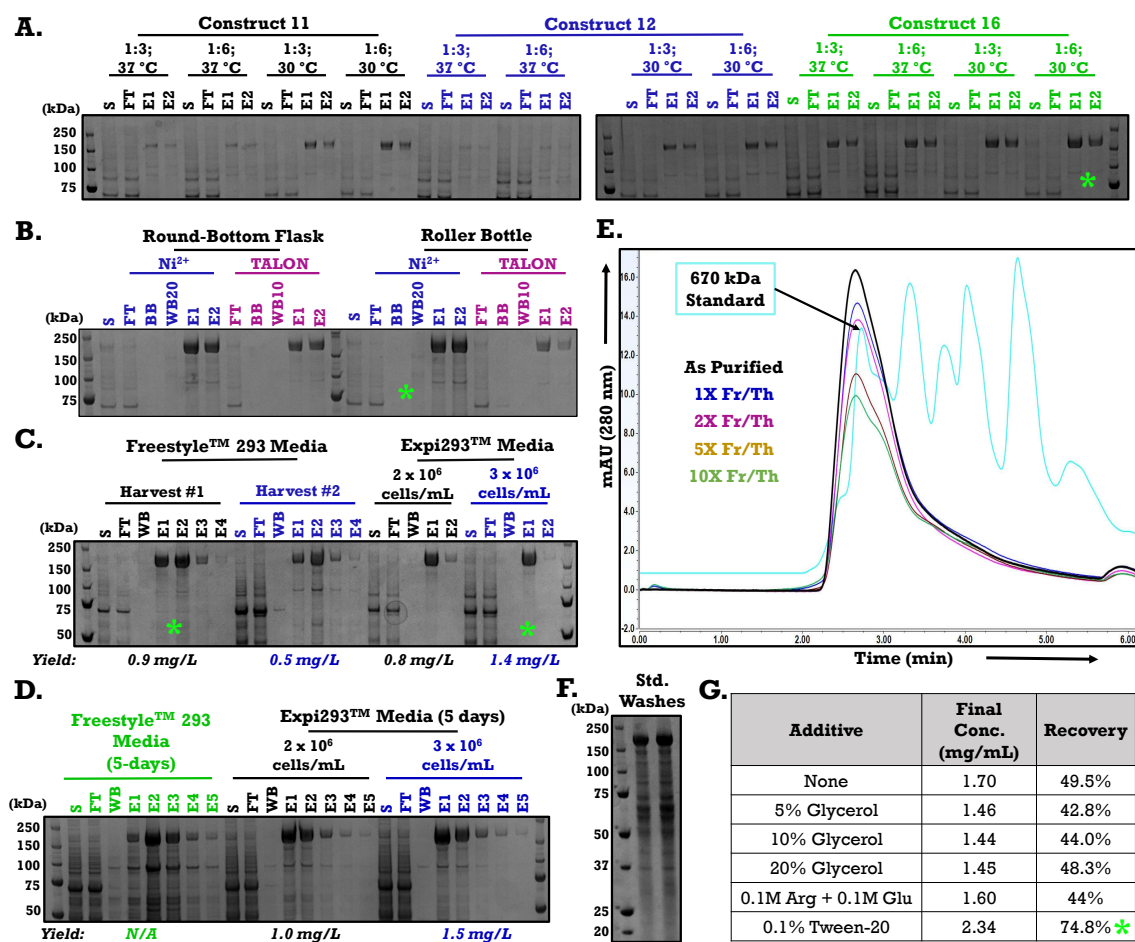


Figure 7.3 Optimization of S Expression and Purification. **A.** SDS-PAGE following pull-down from a 50 mL culture transfected with three different constructs at two different DNA:PEI ratios (w:w) and expressed at two different temperatures. In this and other figures, the best performing condition(s) are marked with green asterisks. **B.** SDS-PAGE of purified S-c16 following pull-down with either Ni²⁺ IMAC Sepharose or TALON from cells expressed at a 500 mL scale in either round-bottom flasks or roller bottles. **C.** SDS-PAGE of purified S-c16 comparing the use of FreeStyle™ 293 and Expi293™ medium. The FreeStyle™ samples were harvested after 3 days (Harvest #1) and then resuspended in fresh medium and harvested again on day six (Harvest #2). The Expi293 samples were harvested after three days following transfection at two different cell densities. Implied yields are reported beneath the gel. **D.** Same experiment as in C., except that the supernatants were harvested after five days. **E.** Analytical SEC traces, generated using a Dionex™ HPLC, following 0, 1, 2, 5, or 10 freeze-thaw cycles of S-c16 stored in 1X PBS. **F.** (Im)purity of S-c16 purified with IMAC using only standard washing procedures. **G.** 72 µg of purified S-c16 was concentrated in centrifugal concentrators with the indicated stabilizing additives. The final recovered amount was recorded. Only Tween-20 improved the final concentrated yield.

7.3C. In a nearly identical experiment, we compared a straight five-day expression in both mediums, without a feeding step at any stage. The final gel from this experiment is shown in Figure 7.3D. In both Figure 7.3C and D, the implied final yields are reported underneath the gel based on the protein concentration in the eluates and the original culture volume.

The first high-level conclusion from these two experiments was that longer expressions, for either five days without feeding or six days with a medium change in the middle, compromises the final protein purity. This seemed especially true in the less rich FreeStyleTM medium (Figure 7.3C, Harvest #2 and Figure 7.3D, green), and even in the Expi293TM medium the three day protein (Figure 7.3C) was cleaner than the five-day protein (Figure 7.3D). This observation was consistent with a notable decrease in cell viability at harvest after longer expression times, sometimes as low as 50%. We hypothesized that intracellular contaminants, released into the medium due to cell death, were responsible for this time-dependent decrease in S purity. Therefore, we strictly maintained a three-day expression window to ensure the S antigen was not contaminated with intracellular proteins. Secondly, the final implied yields reported in Figure 7.3C showed that, when transfected at a cell density of 3×10^6 cells per mL, S could be produced at marginally higher yields in Expi293TM medium. However, in subsequent large-scale experiments we realized that this small increase in yield was insufficient to justify abandoning the FreeStyleTM medium. Not only is FreeStyleTM 40% cheaper than Expi293TM medium, but Expi293TM medium contains a higher concentration of chelating reagents, which strip the Ni²⁺ IMAC Sepharose and reduce yields when larger volumes were prepped. We solved this problem by diluting the medium 1:2 in 1X PBS and then pulling down; this worked well for smaller volumes like this optimization experiment, but at scale this dilution transformed a manageable 12 L purification into a cumbersome 24 L purification. Therefore, despite many publications, including the original Krammer method, reporting the use of Expi293TM medium, we continued with FreeStyleTM medium for large-scale purification. When the medium cost and handling time were considered, the FreeStyleTM-prepped S protein was cheaper per mg, even with the slightly higher absolute yield in Expi293TM medium. In conclusion, these two experiments suggested that a three-day expression in Expi293TM medium at 3×10^6 cells per mL is optimal for small-scale purification of S, while a three-day expression in FreeStyleTM 293 medium at the standard 2×10^6 cells per mL is superior for a large-scale pipeline.

The next improvement in our protocol was the implementation of a strict washing protocol. Based on our combined experience after dozens of preps, we settled on the following wash series to clean S-c16 prior to elution: 50 CV of 1X PBS, 50 CV of 1X PBS + 25 mM imidazole, and 50 CV of 1X PBS + 40 mM imidazole. For a 12 L prep, requiring a 24 mL bed volume IMAC column, the 150 CV of washes require a total buffer volume of 3.6 L, which greatly exceeds the manufacturer's recommendations of 10 CV. However, the gains in purity were manifest. Figure 7.3F shows a preparation of S-c16 that was washed with only the standard, recommended volumes of wash buffer. This protein was highly contaminated, especially relative to the purity of our S-c16 presented repeatedly in Figures 7.2 and 7.3. Furthermore, MSMS analysis of the bands suggested that both intracellular proteins (including elongation factor 2 and TRiC subunits) and secreted proteins (including BSA and complement C3) were present in the final sample. As a result, we maintained the strict washes no matter the size of the purification, a choice which increased the manual handling required to complete the preps but greatly improved the final S purity. Because the protein was pure enough straight off of an IMAC column to use in the assay, these washes also allowed us to skip an SEC clean-up step and simply dialyze the protein into 1X PBS before storage. This increased our final yields (since protein is always lost of SEC columns) and reduced the required purification steps and handling time.

Finally, we wanted to ensure that the purified S was stable after freeze/thaw cycles so that the protein could be thawed as needed by the scientists handling the serological assay. To check this, we concentrated a freshly purified sample of S-c16 to the standard 1 mg mL^{-1} in 1X PBS and then subjected it to ten freeze/thaw cycles. Each cycle consisted of a dip into LN_2 and then a direct thaw in warm water, which is quite harsh. Equal volume samples of S after zero, one, two, five, and ten freeze/thaw cycles were then analyzed via analytical SEC, and the resulting chromatograms are shown in Figure 7.3E. As expected, the freshly purified S was monodisperse and trimeric (black). After one (blue) or two (pink) freeze/thaw cycles, we observed no change in the oligomeric state and only a small decrease in the height of the peak. However, we observed a noticeable decrease in the height of the peak and the appearance of a right shoulder after five (brown) or ten (green) freeze-thaw cycles. This is consistent with degradation of the S trimer after multiple freeze-thaw cycles. Therefore, as a general rule, we only subjected S to one freeze/thaw cycle prior to use in the assay (i.e. purify, freeze, thaw, use, discard), which is conservative given that the protein survived an additional freeze/thaw cycle in this experiment. Importantly, this result also relieved us of the pressure to generate S as needed,

and instead allowed us to stockpile the protein in the freezer to sustain the serology testing in the event of a shutdown of the SGC labs due to COVID-19.

At this point, we had optimized all of the standard transfection, expression, and purification parameters. However, one stage in the process continued to cause problems: centrifugal concentration. Like ourselves, all of the groups we were in contact with, including the Krammer and Owens groups, observed that S was unstable at concentrations above 1 mg mL^{-1} . In an effort to standardize the automated serological assay platform, we were asked to deliver S at exactly 1 mg mL^{-1} ; however, we routinely observed up to 50% sample loss as the S protein was concentrated up to 1 mg mL^{-1} following overnight dialysis. Likely due to the extensive glycan shell, S stuck to the concentration membrane and was prone to precipitating and aggregating as the concentration increased. We could partially reduce this protein loss by concentrating in cycles of just three minutes with intermittent mixing steps. However, in an effort to further improve this recovery, we screened three different concentration additives: glycerol, Arg/Glu, and Tween-20. These additives were added to $72 \mu\text{g}$ of S at a starting concentration of 0.2 mg mL^{-1} and then violently concentrated at $21,000 \times g$ in micro centrifugal concentrators. The final concentration and absolute recovery of S in the presence of these additives is reported in Figure 7.3G. The results suggested that incorporation of 0.1% of the non-ionic detergent Tween-20, a gentle and common additive in ELISA wash buffers, could improve the recovery of the S protein by 50% relative to the protein concentrated in empty 1X PBS.

7.5 Generation and Evaluation of S-Expressing Monoclonal Cell Lines

At the end of the optimization experiments presented in Section 7.4, we had developed a method that reproducibly yielded $0.8\text{--}1.0 \text{ mg mL}^{-1}$ S from transient transfection, sufficient to support the serological assay in perpetuity. However, for this and other applications moving forward, the use of transient transfection was especially cumbersome, since it required 0.5 mg L^{-1} of plasmid DNA; this DNA requirement was so extensive that we eventually relied on three full-time scientists completing up to six gigapreps of the S-c16 construct each week to support the pipeline. In an effort to bypass the DNA purification and transient transfection stages of the protocol, we moved forward with the generation of monoclonal over-expressing cell

lines. This became even more important when we were granted additional funds to purchase a 12 L Kuhner SB-10X orbital shaking bioreactor to further scale up our S production. Although large transient transfections are possible in bioreactors, stable cell lines fully capitalize on the increased scale by allowing continuous feeding of the cells and harvest of the spent medium for purification. In such a “perfusion” set-up, with continuous feeding and harvest, a stable cell line facilitates the purification of secreted proteins from the same bioreactor for four weeks or longer. This type of production dominates industrial purification of antibodies and therapeutic proteins and presented an opportunity for a step-change in our capacity to produce S [572, 573]. Therefore, with much of the cell biology spearheaded by Dr. Jong Fu Wong, we moved forward with the generation of monoclonal stable cell lines for our three key S constructs.

At the outset, only two of our S constructs were cloned in selectable vectors: S-c17 (in NeoR pNeoSec) and S-c19 (in NeoR pOPINTTGneo). We generated monoclonal Expi293 cell lines using random genetic integration of the parental S-expressing plasmid followed by selection with $800\text{ }\mu\text{g mL}^{-1}$ G418. Encouraged by reports that pre-linearization of the parental plasmid with restriction enzymes could increase insertion efficiency [574], we tested transfection with both circular and linear plasmid DNA. After selection, the cells were subjected to limiting dilution in 96-well plates, and wells with a single visible colony were scaled up and probed for S expression via anti-His-tag western blot of the culture medium. The highest S-expressing lines were further scaled up, suspension adapted, stored as frozen stocks, and subsequently used for large-scale purification runs. A detailed protocol for the generation of the monoclonal cell lines reported herein can be found in Section 7.1.3.

Figure 7.4A shows the selection of S-c17 expressing mc14D3, which contained a strong His-tagged band in the supernatant at the expected size, along with clear mVenus fluorescence under the microscope. The mc14D3 cell line was scaled up at a 500 mL scale and the medium was pulled down using a 25 mL bed volume IMAC Sepharose 6 Fast Flow column; the gel from this IMAC enrichment step is shown in Figure 7.4B. S-c17 is strongly over-expressed by mc14D3 as evidenced by the strength of the band in the un-purified medium (Lane 1). In fact, during the transient transfection of S-c16 (see Figure 7.2B and all of the gels in Figure 7.3), the target protein is not visible on a gel from the un-purified medium. Impressively, the 500 mL input volume actually overwhelmed the large IMAC column, and some of the S-c17 escaped in the flow-through (FT). However, even without accounting for this lost protein in the FT, the yield of S-c17 from mc14D3 was 120 mg L^{-1} ,

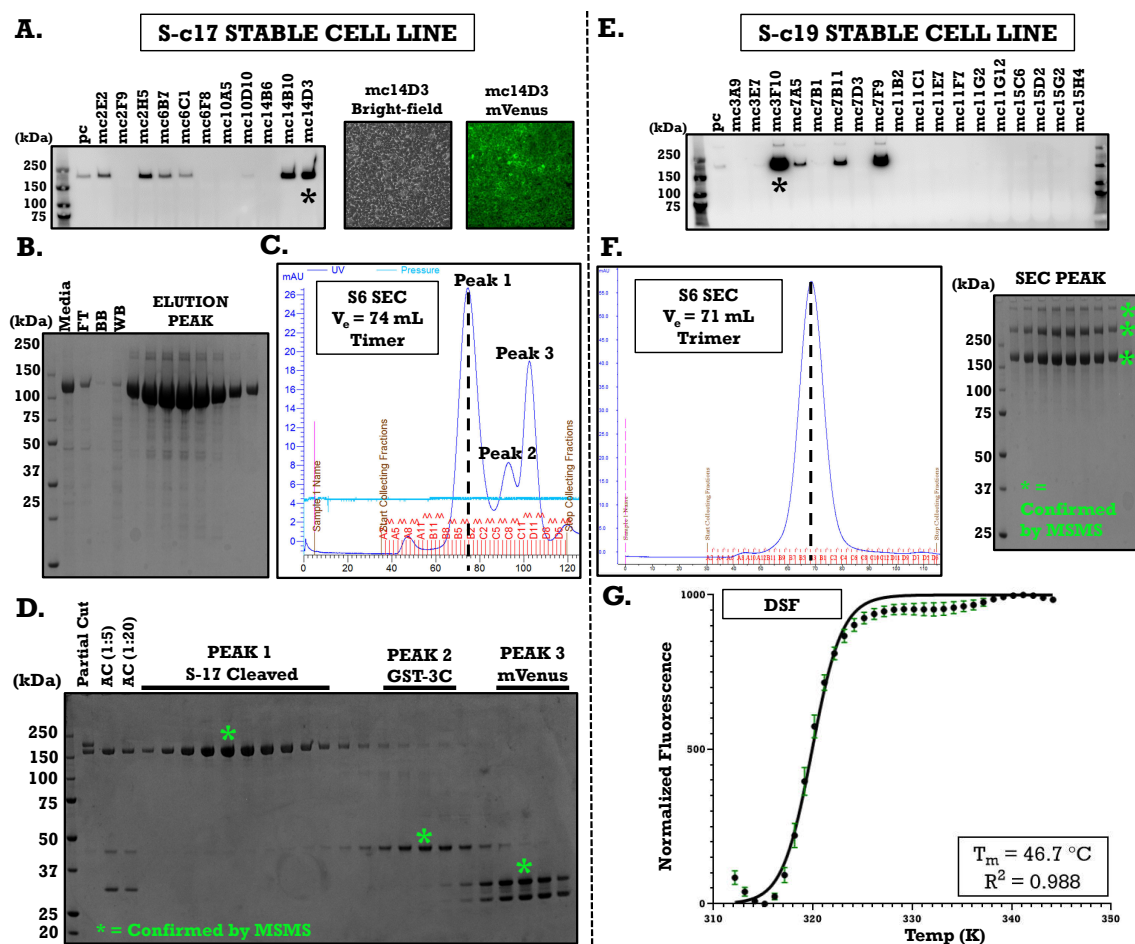


Figure 7.4 S-c17 and S-c19 Monoclonal Cell Lines. **A.** Anti-His-tag western blot of medium from S-c17 expressing monoclonal Expi293 cell lines following G418 selection and single-cell dilution. Fluorescent and bright-field microscope images from mc14D3 show a secreted mVenus-tagged S as expected. **B.** 500 mL culture of S-c17 mc14D3 purified with a 25 mL IMAC Sepharose 6 Fast Flow column. **C.** S6 chromatogram generated following 3C-cleavage of S-c17. **D.** SDS-PAGE of fractions from chromatogram in C. Peak 1 is cleaved S-c17, Peak 2 is GST-3C, and Peak 3 is the released mVenus. AC = after cut. **E.** Anti-His-tag western blot of medium from S-c19 expressing monoclonal Expi293 cell lines following G418 selection and single-cell dilution. **F.** S6 chromatogram and the corresponding SDS-PAGE of S-c19 following IMAC purification from 1 L of medium from mc3F10. **G.** DSF analysis of S-c19 purified from mc3F10.

essentially 100-fold better than S-c16 produced transiently. S-c17, containing a 13 residue deletion at the N-terminus and a 47 residue C-terminal deletion relative to S-c16 and S-c19, was reliably the highest expressing S construct we purified, suggesting these truncations stabilize the protein. The incorporation of a large, highly soluble mVenus tag might also improve the recovery of the fusion protein by protecting it from degradation and/or aggregation after secretion into the medium. As further

validation of the mc14D3 stable cell line, we subjected the IMAC-purified protein to treatment with C3-protease, which should cleave all of the C-terminal tags but retain the S trimer. After cleavage, we ran the protein over an S6 column and the chromatogram is shown in Figure 7.4C. As expected, we observed a monodisperse S-c17 trimer (Peak 1) at an elution volume of 74 mL, slightly smaller than the full-length S-c16 which eluted from this column at 71 mL (see Figure 7.2C). SDS-PAGE and MSMS analysis of the S6 fractions, shown in Figure 7.4D, demonstrated that Peak 2 was GST-3C and Peak 3 contained the C-terminal mVenus-containing tag. Therefore, mc14D3 could be used to generate large quantities of pure, monodisperse, un-tagged S-c17 appropriate for structural and functional studies. In our bioreactor, this cell line could theoretically yield up to 2.9 g of S-c17 in a single week.

Figure 7.4E shows the selection of S-c19 expressing mc3F10, which contained a strong His-tagged band in the supernatant at the expected molecular weight. Scale-up of mc3F10 at the 2 L scale and purification on a 25 mL IMAC Sepharose 6 Fast Flow column yielded 15 mg L^{-1} of full-length S-c19. Following IMAC, the protein was subjected to SEC on an S6 column; the chromatogram and corresponding gel is shown in Figure 7.4F. The S-c19 protein derived from mc3F10 is monodisperse and trimeric, with an observed elution volume of 68 mL. The peak is highly pure by SDS-PAGE, although S-c19 resolves as three distinct oligomeric bands on a gel (all confirmed to be S by MSMS). We hypothesized that the un-biotinylated AviTagTM on S-c19, which contains a mixture of positive and negatively charged residues, further stabilized the trimer relative to S-c16 and prevented complete dissociation of the oligomers during SDS-PAGE. S-c19 derived from mc3F10 behaved similarly to the transiently expressed S-c16 during DSF, with a slightly lower T_m of 46.7°C as calculated from the data in Figure 7.4G. When grown in our bioreactor, mc3F10 could yield up to 360 mg of full-length S-c19 each week. This S-c19 cell line is especially useful because the protein product contains the full-length, 2P S ectodomain that can be biotinylated with the AviTagTM for immobilization or detection.

Based on our success with monoclonal cell lines expressing S-c17 and S-c19, we wanted to move forward with the generation of a line expressing the serology testing construct S-c16. However, S-c16 was cloned into the backbone pHL-SEC, which lacked a selection marker and was therefore only suitable for transient transfection. Therefore, we first had to re-clone S-c16 into a selectable backbone. Because of our success with S-c19, we decided to move S-c16 into the pOPINTTGneo vector and re-name the construct S-c23. The pOPINTTGneo vector contains a tri-functional promoter that facilitates expression of the target protein in either bacterial, insect,

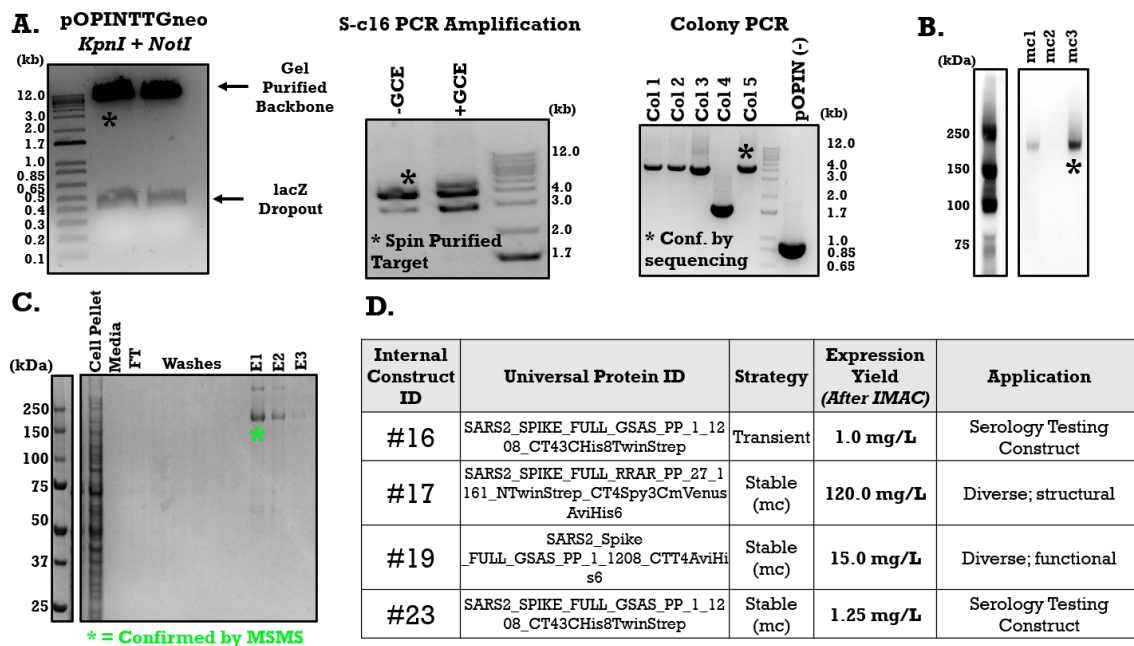


Figure 7.5 S-c23 Monoclonal Cell Line. **A.** Molecular biology to move S-c16 into a selectable backbone to generate S-c23. (left) *KpnI* and *NotI* double-digest of pOPINTTGneo. (center) PCR amplification of S-c16 with primers containing HiFi compatible overhangs. (right) Colony PCR from five colonies transformed with the HiFi assembled vector product. The pOPIN lane is a negative control. Colony 5, marked with an asterisk, contained the in-frame insertion of S-c16 by Sanger Sequencing and the plasmid was re-named S-c23. **B.** Anti-His-tag western blot of the medium from three monoclonal Expi293 lines transfected with S-c23 and selected with G418. mc3, marked with an asterisk, yielded the most protein. **C.** SDS-PAGE following IMAC purification from 2 L of mc3. **D.** Overview of the four workhorse constructs reported in this Chapter.

or mammalian cells [575], and it also contains a mu-sec signal sequence upstream of a *KpnI* cut site. A detailed protocol for cloning S-c23 is provided in Section 7.1.2, and key steps in the process are shown in Figure 7.5A. First, the pOPINTTGneo backbone was linearized by double digestion with *KpnI* and *NotI*, which released a lacZ dropout and opened the ORF directly downstream of the vector's signal peptide. This digestion is shown in the first panel in Figure 7.5A. S-c16 was then PCR amplified with Q5[®] Polymerase and primers containing HiFi-compatible 5' overhangs generated by NEBuilder. The PCR worked better in the absence of the Q5[®] GC-Enhancer, as demonstrated in the middle panel of 7.5A. The linearized vector and PCR amplicon were purified, assembled with a HiFi Assembly, and then transformed into Mach 1 cells prior to colony PCR analysis using vector-specific primers. As shown in the right hand panel of 7.5A, Colony 5 contained an insert of the expected size (approximately 4 kb) and Sanger Sequencing verified the in-frame

insertion of S-c16 with the backbone signal peptide. At the protein level, this new construct S-c23 is 100% identical to the serology testing construct S-c16.

After generating the new construct, we moved forward with generating S-c23-expressing Expi293 monoclonal cell lines exactly as before, and the results of the anti-His-tag western blot following single cell dilution are shown in Figure 7.5B. The mc3 line secreted significant amounts of a correctly sized His-tagged protein into the cell culture medium. The mc3 line was subsequently suspension adapted and scaled up at a 2L scale and purified in an identical manner to S-c16. SDS-PAGE following a one-step IMAC purification is shown in Figure 7.5C, and MSMS analysis confirmed that the expressed protein was S-c23. The final yield from mc3 was 1.25 mg L^{-1} , 25% higher than the transient transfection of S-c16. Interestingly, despite having the same S sequence, S-c19 expresses ten-fold better than S-c23, a result we reproduced many times. In other experiments (data not shown) we demonstrated that the Twin-Strep-tag in S-c16 is not accessible for pull-down, suggesting misfolding or aggregation of the tag region downstream of the T4-site in S-c16. Therefore, we hypothesized that the use of a His-AviTagTM is preferable to the use of the more common His-Twin-Strep-tag for the isolation of S, and we would recommend that other groups avoid the His-Twin-Strep-tag in S constructs moving forward. Nevertheless, the S-c23 expressing cell-line mc3 allowed us to isolate the serology testing construct at a slightly higher yield without requiring large quantities of plasmid DNA and tedious transient transfections. In our bioreactor, mc3 could yield as much as 30 mg of the serology testing construct each week.

As a final summary, Figure 7.5D provides an overview of the four workhorse constructs reported in Chapter 7. While we continued to use S-c16 for serology testing since the assay was optimized and working well with this construct, the three monoclonal cell lines we generated provide a superior route to purifying large quantities of S to support diverse applications in the future. As a sample of the impact from this work, I provide a few representative applications of our S protein in Section 7.6.

7.6 Conclusions and Future Directions

Most importantly, the S-c16 we generated was used to support the National Serology Testing Center and Thermo Scientific's OmnipathTM Combi SARS-CoV-2 IgG ELISA. As of March 2021, we delivered over 470 mg of S-c16 to support this effort. The serological assay developed with our S antigen continues to stand as one of the

most accurate tests available, due in no small part to the purity, reliability, and quantity of the S purified with the method developed herein [564]. Future work will focus on the transfer of S expression methods into our new bioreactor, facilitated by our work to generate monoclonal cell lines with orthogonal S constructs. Importantly, in addition to the serology testing platform, we were also able to deliver S to over a dozen different academic and industrial groups to support diverse SARS-CoV-2 research.

As one example, in collaboration with Agilent, our internal mass spectrometry supervisor Dr. Rod Chalk developed a cheap and reliable method for mapping the N-linked glycans on S without requiring ultra-sensitive instruments or specialized glycoproteomics approaches [576]. The Mass Retention Time Fingerprinting method reported by Chalk *et al.* will allow us and others to quality control prep-to-prep variation in the S glycosylation envelope and explore parameters which modify the S-glycans. As a second example, many reports suggested that the S protein can activate an innate immune response in human monocytes and macrophages, although there remains some controversy about the exact mechanism [577, 578, 579]. In an effort to validate this hypothesis, we provided LPS-free S-c19 to Anna Marzeda in the Kennedy Institute. Anna was unable to replicate the activation of primary human monocytes with the S protein, leading her to hypothesize that cell activation experiments with misfolded or LPS-contaminated S protein could be confusing the literature. These two examples highlight the continuing impact of our S protein purification methodologies, which will facilitate additional research; this is especially critical given the prohibitive cost (often above \$10,000 per milligram) of obtaining S protein from commercial vendors. As an example of the scale of our output, purchasing sufficient S to support the serology testing platform would have cost well over to \$4,500,000.

In conclusion, Chapter 7 presented my second method development vignette, focusing on the large-scale production of S, which concludes the results section of this thesis. While serving as the daily in-lab manager of the SGC's S purification pipeline, which required the teamwork of well over twenty scientists, I oversaw the development of robust methods to generate four different S constructs. The work was completed during the height of the UK lockdown in early- to mid-2020 and stalled my personal project for over four months. My in-lab efforts were extensive over this period; I personally completed all key optimization experiments and dozens of large-scale preps, totalling 38 unique electronic lab notebook records, and I spent over sixty in-lab hours working on the project each week. Therefore, I include this

Chapter under Oxford's COVID-19 Student Policy to highlight my activities during the pandemic and mitigate some of the unfinished experiments in Chapters 5 and 6. My work on secreted expression of TLR4, purification of diverse FReP constructs, and generation of stably expressing mammalian cell lines using CRISPR all facilitated my ability to help with this S purification project, which made a crucial impact on Oxford's COVID-19 response.

Chapter 8

Conclusion: A New (Old) Hypothesis

8.1 Receptorome from First Principles

In contrast to currently accepted dogma in the literature, I hypothesize that the FRePs, and perhaps even the entire super-family of DAMPs, *do not* directly engage TLR4 to activate pro-inflammatory signalling. In Chapter 4, using recombinant proteins and biophysical techniques, I could not detect a high-affinity interaction between TLR4 and any FReP, a result which seemed surprising at first but is actually consistent with the dearth of biophysical data suggesting a direct complex in the DAMP literature, as reviewed in Section 1.2. Furthermore, in Chapter 5, using primary human monocytes in an XL-AP-MSMS workflow, I did not detect a TLR4:FReP interaction but I did demonstrate FReP-mediated integrin activation and speculated on TNC binding to lipoprotein receptors. These data could explain the divergence between the cellular data showing a high-affinity, DAMP-mediated activation of TLR4 signalling and the lack of biophysical data demonstrating a direct interaction between either TLR4 or TLR4:MD-2 and DAMPs. I contest that TLR4 is necessary but insufficient for DAMP-activation of immune cells, and that TLR4 activation likely requires the assembly of a large, context-dependent accessory protein apparatus which I coined a “receptorome” herein. My data fundamentally do not support a direct TLR4:DAMP interaction, and I remain convinced that the immediate receptor ligation event is mediated by a protein(s) other than TLR4.

Furthermore, I want to argue that this new hypothesis can in fact be derived from first principles of the innate immune system. As discussed in Chapter 1, the canoni-

cal function of the TLR family is to detect pathogenic molecular signatures, initiate a defensive response, and alert the immune system to the presence of dangerous exogenous infections. This classic innate immune detection of PAMPs proceeds broadly within a simple “1 ligand, 1 receptor” paradigm. Direct PAMP binding, including LPS to TLR4:MD-2, acylated lipoproteins to the TLR2:TLR1 or TLR2:TLR6 heterodimers, dsRNA to TLR3, flagellin to TLR5, ssRNA to TLR8, and CpG DNA to TLR9 all immediately trigger TLR oligomerization and activate downstream pro-inflammatory signalling [52]. In the context of truly exogenous stimuli, this direct activation mechanism makes evolutionary sense. TLRs are remarkably ancient across all multi-cellular eukaryotes; even simple poriferans and cnidarians contain TLR orthologs which contribute to protection from infection [580]. For exogenous stimuli, then, the logic is simple across eukaryota: see something, say something.

However, for the detection of endogenous DAMPs in the context of higher eukaryotes, this mechanism poses problems. While the presence of a pathogen-derived PAMP implies a very high probability of infection and therefore a high likelihood of a beneficial immune response, immune cells can expect to randomly encounter endogenous DAMPs under normal conditions. If TLRs were to recognize self proteins directly, with very high affinity, the autoimmune toxicities would likely be disadvantageous, especially given the dozens of different DAMPs known to circulate. In this way, it makes evolutionary sense that TLR recognition of DAMPs should require additional thresholds of context specific regulation, and therefore the involvement of accessory proteins, that tune the response to endogenous ligands and prevent runaway inflammation—the known autoimmune toxicities of chronic DAMP up-regulation further support this claim. Therefore, my hypothesis regarding the requirement for a multiprotein regulatory apparatus to facilitate DAMP activation of TLR4 is fundamentally based on the contention that innate immune receptors *should not*, as a matter of principle, become activated by a single high affinity binding event to endogenous molecules. The broader DAMP field, in my view, has neglected the critical involvement of context-specific regulatory proteins and confined TLR-activation upon DAMP binding within the simplistic and inappropriate “see something, say something” paradigm suggested by PAMPs. For endogenous DAMPs, the innate immune system likely operates on a “maybe see something, maybe say something” basis.

8.2 History of a TLR4 Receptorome

In the above Section 8.1, I called this receptorome hypothesis “new”, and it does indeed diverge from the predominant model in the modern DAMP literature. However, many years ago, significant effort was made to characterize TLR4-associated megacomplexes, and these older findings strongly support the notion that TLR4 clusters into large, distinct lipidic microenvironments with many other proteins during activation. In some sense, then, my contention for a DAMP-induced, TLR4-associated receptorome is a callback to a previous hypothesis that I believe should be re-visited in earnest within the context of endogenous TLR4 activators.

As early as 2002, Martha and Kerry Triantafilou hypothesized that supramolecular complexes are required to support PAMP-activated TLR4 signalling, writing in a review:

“It is our belief that the mechanism of action of a couple of receptor molecules is an oversimplified model of LPS-induced activation. The existence of different supramolecular arrangements could explain how these relatively conserved individual molecules (TLR4 and CD14) have managed to cope with a rapidly changing pathogenic universe. Accumulating evidence is pointing towards the direction that responses to different pathogens vary depending on cell type, composition of supramolecular activation clusters and intracellular adaptor molecules...We propose that the innate recognition of bacteria involves the dynamic association of multiple receptors within one or more microdomains. The transient association of different receptors within the activation cluster could give rise to the recognition and/or discrimination of different ligands” [581].

Triantafilou and Triantafilou’s claim is based on years of biophysical experiments at the cell surface (including AP-MSMS, fluorescence recovery after photobleaching (FRAP), and FRET) which demonstrate the existence of LPS-activated, TLR4-associated receptor clusters. In 2001, using a combination of FRET, super-resolution microscopy, and sucrose density gradient flotation after detergent extraction to probe the LPS activation cluster in monocytes, Pfeiffer *et al.* reported a lipid-microdomain-associated cluster simultaneously involving TLR4, CD14, CD55 (*aka* complement decay-accelerating factor), CD16a (*aka* low affinity immunoglobulin gamma Fc region receptor III-A), CD11b/CD18 (*aka* integrin $\alpha M\beta 2$), CD32 (*aka* low affinity immunoglobulin gamma Fc region receptor II-b), CD64 (*aka* high affinity immunoglobulin gamma Fc receptor I), CD36 (*aka* scavenger receptor platelet

glycoprotein 4), FC γ RIIIa, and CD81 (CD81 antigen, known as a microdomain scaffold) [582]. Furthermore, by comparing the response to LPS, lipoteichoic acid (LTA), and ceramide, Pfeiffer *et al.* showed that the composition of the induced microdomain varies depending on the agonist used. Similarly, using AP-MSMS and FRET to study LPS-activated monocytes and epithelial cells in 2001, Triantafilou, Triantafilou, and Dedrick discovered a unique LPS-activation cluster consisting of HSP70, HSP90, CXCR4, and GDF5 (growth differentiation factor 5) [583]. Antibody blockade of any of these four proteins interferes with receptor clustering and inhibits TNF- α release upon treatment with even very high concentrations (1000 ng mL⁻¹) of LPS. In 2002, Triantafilou *et al.* reproduced these results by demonstrating that lipid rafts are responsible for co-localizing these receptor complexes (via direct extraction of the rafts and western blotting) and that lipid-raft-disrupting drugs inhibit LPS-induced TNF- α secretion in a monocytic cell line [56]. Triantafilou *et al.* nicely reproduced the earlier 2001 reports by directly capturing an LPS-activation complex within lipid rafts consisting of at least TLR4, CD14, HSP70, HSP90, CXCR4, and GDF5. These seminal original papers, and many subsequent ones, were summarized in two 2015 review articles which emphasized the surprising diversity and functional importance of TLR4 accessory proteins within lipid raft microdomains [584, 585]. Indeed, Pociennikowska *et al.* write: “...an increasing body of data suggests that other raft proteins, like CD44 and CD36, co-operate with TLR4 in the induction of the pro-inflammatory responses not only to LPS but also to endogenous ligands...” [585].

8.3 A Unifying Hypothesis

In summary, the first-principles receptorome hypothesis presented in Section 8.1 is broadly consistent with historical data presented in Section 8.2 showing that TLR4-associated supramolecular scaffolds within lipid rafts are critical for receptor activation. In the context of PAMP-activated TLR4 signalling, a direct, high-affinity complex between TLR4:MD-2:LPS is formed (see Section 1.1.3 for a discussion); yet, even in this context, TLR4 is known to form supramolecular clusters which modify its signalling. It is my hypothesis, then, that in the context of DAMP-activated TLR4 signalling, a similar supramolecular cluster forms but TLR4 is not the direct, high affinity receptor within the lipid raft. Indeed, I suggest that transient, context-specific protein complexes (involving integrins and lipoprotein receptors) serve as the true DAMP-binding scaffold, and that this information is indirectly communicated

to TLR4 via accessory proteins co-localized in the lipid rafts. This hypothesis could explain (1) the strict requirement of TLR4 for downstream NF- κ B activation by DAMPs (2) the phenotypic differences between PAMP- and DAMP-stimulated cells and (3) the lack of reliable biophysical data showing a direct interaction between DAMPs and TLR4. It is my belief that this hypothesis unifies conflicting data in the literature and may lead to the discovery of novel, and perhaps therapeutically tractable, molecular mechanisms behind DAMP activation of TLR4.

Appendix A

Supplementary Data and Figures

A.1 TNC, ANGPTL2, and ANGPTL6 Complete X-Ray Data Collection and Refinement Statistics

A.2 FAS2FURIOUS Manuscript

A.3 Raw XL-AP-MSMS Data

A.4 List of Abbreviations

References

- [1] Kenneth Murphy and Casey Weaver. *Janeway's Immunobiology*. Garland Science, Taylor and Francis, New York, 9th edition, 2017.
- [2] Luke A. J. O'Neill, Douglas Golenbock, and Andrew G Bowie. The history of Toll-like receptors — redefining innate immunity. *Nature Reviews Immunology*, 13(6):453–460, 2013.
- [3] Charles A Janeway. Approaching the asymptote? Evolution and revolution in immunology. *Cold Springs Harbor Symposia on Quantitative Biology*, 54:1–13, 1989.
- [4] Nicholas J. Gay and Fionna J. Keith. Drosophila Toll and IL-1 receptor. *Nature*, 351:355–356, 1991.
- [5] Ruth Steward. Dorsal , an Embryonic Polarity Gene in Drosophila , is Homologous to the Vertebrate Proto-oncogene c-rel. 238(4827):692–694, 1987.
- [6] Patrick A Baeuerle and Thomas Henkel. Function and Activation of NF- κ B in the Immune System. *Annual Review of Medicine. Rev. Immunol*, 12:141–179, 1994.
- [7] A. Heguy, C. T. Baldari, G. Macchia, J. L. Telford, and M. Melli. Amino acids conserved in interleukin-1 receptors (IL-1Rs) and the Drosophila Toll protein are essential for IL-1R signal transduction. *Journal of Biological Chemistry*, 267(4):2605–2609, 1992.
- [8] Steve Whitham, S.P. Dinesh-Kumar, Doil Choi, Reinhard Hehl, Catherine Corr, and Barbara Baker. The product of the tobacco mosaic virus resistance gene N: Similarity to toll and the interleukin-1 receptor. *Cell*, 78:1101–1115, 1994.

- [9] Y. Tony Ip, Michael Reach, Ylva Engstrom, Latha Kadalayil, Haini Cai, Sergio González-Crespo, Kazuaki Tatei, and Michael Levine. Dif, a dorsal-related gene that mediates an immune response in *Drosophila*. *Cell*, 75(4):753–763, 1993.
- [10] Bruno Lemaitre, Emmanuelle Nicolas, Lydia Michaut, Jean Marc Reichhart, and Jules A. Hoffmann. The dorsoventral regulatory gene cassette spatzle/Toll/Cactus controls the potent antifungal response in *Drosophila* adults. *Cell*, 86(6):973–983, 1996.
- [11] Bruno Lemaitre. The road to Toll. *Nature Reviews Immunology*, 4:1–8, 2004.
- [12] J.E. Sims and S.K. Dower. Interleukin-1 Receptors. *Eur. Cytokine Netw.*, 5:539–546, 1994.
- [13] Ruslan Medzhitov, Paula Preston-Hurlburt, and Charles A. Janeway. A human homologue of the *Drosophila* toll protein signals activation of adaptive immunity. *Nature*, 388(6640):394–397, 1997.
- [14] O. Westphal and O. Lüderitz. Chemische Erforschung von Lipopolysacchariden gramnegativer Bakterien. *Angewandte Chemie*, 66(13-14):407–417, 7 1954.
- [15] Ranjan Sen and David Baltimore. Inducibility of κ immunoglobulin enhancer-binding protein NF- κ B by a posttranslational mechanism. *Cell*, 47(6):921–928, 1986.
- [16] J. Han, J. D. Lee, P. S. Tobias, and R. J. Ulevitch. Endotoxin induces rapid protein tyrosine phosphorylation in 70Z/3 cells expressing CD14. *Journal of Biological Chemistry*, 268(33):25009–25014, 1993.
- [17] Peter S. Tobias, Katrin Soldau, and Richard J. Ulevitch. Isolation of a lipopolysaccharide-binding acute phase reactant from rabbit serum. *Journal of Experimental Medicine*, 164(3):777–793, 1986.
- [18] Samuel D Wright, Robert A Ramos, Peter S Tobias, Richard J Ulevitch, John C Mathison, Samuel D Wright, Robert A Ramos, Peter S Tobias, Richard J Ulevitch, and John C Mathison. CD14 , a Receptor for Complexes of Lipopolysaccharide LPS and LPS Binding Protein. *Science*, 249(4975):1431–1433, 1990.

- [19] R. Schumann, Leong, G. Flaggs, P. Gray, S. Wright, J. Mathison, P. Tobias, and R. Ulevitch. Structure and function of lipopolysaccharide binding protein. *Science*, 249(4975):1429–1431, 9 1990.
- [20] Ruey-bing Yang, Melanie R Mark, Alane Gray, Arthur Huang, Ming Hong Xie, Min Zhang, Audrey Goddard, William I Wood, Austin L Gurney, and Paul J Godowski. Toll-like receptor-2 mediates lipopolysaccharide-induced cellular signalling. *Nature*, 395(6699):284–288, 9 1998.
- [21] A Poltorak, X He, I Smirnova, M Y Liu, C Van Huffel, X Du, D Birdwell, E Alejos, M Silva, C Galanos, M Freudenberg, P Ricciardi-Castagnoli, B Layton, and B Beutler. Defective LPS signaling in C3H/HeJ and C57BL/10ScCr mice: mutations in Tlr4 gene. *Science (New York, N.Y.)*, 282(5396):2085–8, 12 1998.
- [22] Salman T. Qureshi, Line Larivière, Gary Leveque, Sophie Clermont, Karen J. Moore, Philippe Gros, and Danielle Malo. Endotoxin-tolerant Mice Have Mutations in Toll-like Receptor 4 (Tlr4). *Journal of Experimental Medicine*, 189(4):615–625, 2 1999.
- [23] Katsuaki Hoshino, Osamu Takeuchi, Taro Kawai, Hideki Sanjo, Tomohiko Ogawa, Yoshifumi Takeda, Kiyoshi Takeda, and Shizuo Akira. Toll-Like Receptor 4 (TLR4)-Deficient Mice Are Hyporesponsive to Lipopolysaccharide: Evidence for TLR4 as the Lps Gene Product. *Journal of Immunology*, 162:3749–3752, 1999.
- [24] R Shimazu, S Akashi, H Ogata, Y Nagai, K Fukudome, K Miyake, and M Kimoto. MD-2, a molecule that confers lipopolysaccharide responsiveness on Toll-like receptor 4. *Journal of Experimental Medicine*, 189(11):1777–82, 6 1999.
- [25] Andra B. Schromm, Egil Lien, Philipp Henneke, Jesse C. Chow, Atsutoshi Yoshimura, Holger Heine, Eicke Latz, Brian G. Monks, David A. Schwartz, Kensuke Miyake, and Douglas T. Golenbock. Molecular genetic analysis of an endotoxin nonresponder mutant cell line: A point mutation in a conserved region of MD-2 abolishes endotoxin-induced signaling. *Journal of Experimental Medicine*, 194(1):79–88, 2001.
- [26] Taro Kawai and Shizuo Akira. The roles of TLRs, RLRs and NLRs in pathogen recognition. *International immunology*, 21(4):317–37, 4 2009.

- [27] A Yoshimura, E Lien, R R Ingalls, E Tuomanen, R Dziarski, and D Golenbock. Cutting edge: recognition of Gram-positive bacterial cell wall components by the innate immune system occurs via Toll-like receptor 2. *Journal of immunology (Baltimore, Md. : 1950)*, 163(1):1–5, 1999.
- [28] Osamu Takeuchi, Taro Kawai, Peter F. Mührladt, Michael Morr, Justin D. Radolf, Arturo Zychlinsky, Kiyoshi Takeda, and Shizuo Akira. Discrimination of bacterial lipoproteins by Toll-like recepttor 6. *International Immunology*, 13(7):933–940, 2001.
- [29] Osamu Takeuchi, Shintaro Sato, Takao Horiuchi, Katsuaki Hoshino, Kiyoshi Takeda, Zhongyun Dong, Robert L. Modlin, and Shizuo Akira. Cutting Edge: Role of Toll-Like Receptor 1 in Mediating Immune Response to Microbial Lipoproteins. *The Journal of Immunology*, 169(1):10–14, 2002.
- [30] Tsung Hsien Chuang and Richard J. Ulevitch. Identification of hTLR10: A novel human Toll-like receptor preferentially expressed in immune cells. *Biochimica et Biophysica Acta - Gene Structure and Expression*, 1518(1-2):157–161, 2001.
- [31] Faith Fore, Cut Indriputri, Janet Mamutse, and Jusak Nugraha. TLR10 and its unique anti-inflammatory properties and potential use as a target in therapeutics. *Immune Network*, 20(3):1–10, 2020.
- [32] Marije Oosting, Shih Chin Cheng, Judith M. Bolscher, Rachel Vestering-Stenger, Theo S. Plantinga, Ineke C. Verschuere, Peer Arts, Anja Garritsen, Hans Van Eenennaam, Patrick Sturm, Bart Jan Kullberg, Alexander Hoischen, Gosse J. Adema, Jos W.M. Van Der Meer, Mihai G. Netea, and Leo A.B. Joosten. Human TLR10 is an anti-inflammatory pattern-recognition receptor. *Proceedings of the National Academy of Sciences of the United States of America*, 111(42):E4478–E4484, 2014.
- [33] Bethany M. Henrick, Xiao Dan Yao, Muhammad Atif Zahoor, Alash’le Abimiku, Sophia Osawe, and Kenneth L. Rosenthal. TLR10 senses HIV-1 proteins and significantly enhances HIV-1 infection. *Frontiers in Immunology*, 10(MAR):1–14, 2019.
- [34] Fumitaka Hayashi, Kelly D. Smith, Adrian Ozinsky, Thomas R. Hawn, Eugene C. Yi, David R. Goodlett, Jimmy K. Eng, Shizuo Akira, David M. Underhill, and Alan Aderem. The innate immune response to bacterial flagellin is mediated by Toll-like receptor 5. *Nature*, 410(6832):1099–1103, 2001.

- [35] Dekai Zhang, Guolong Zhang, Matthew S. Hayden, Matthew B. Greenblatt, Crystal Bussey, Richard A. Flavell, and Sankar Ghosh. A Toll-like Receptor That Prevent Infection by Uropathogenic Bacteria. *Science*, 303(5663):1522–1526, 2004.
- [36] Felix Yarovinsky, Dekai Zhang, John F. Andersen, Gerard L. Bannenberg, Charles N. Serhan, Matthew S. Hayden, Sara Hieny, Fayyaz S. Sutterwala, Richard A. Flavell, Sankar Ghosh, and Alan Sher. Immunology: TLR11 activation of dendritic cells by a protozoan profilin-like protein. *Science*, 308(5728):1626–1629, 2005.
- [37] A. Alicia Koblansky, Dragana Jankovic, Hyunju Oh, Sara Hieny, Waradon Sungnak, Ramkumar Mathur, Matthew S. Hayden, Shizuo Akira, Alan Sher, and Sankar Ghosh. Recognition of Profilin by Toll-like Receptor 12 Is Critical for Host Resistance to *Toxoplasma gondii*. *Immunity*, 38(1):119–130, 2013.
- [38] Warrison A. Andrade, Maria Do Carmo Souza, Espiridion Ramos-Martinez, Kamalpreet Nagpal, Miriam S. Dutra, Mariane B. Melo, Daniella C. Bartholomeu, Sankar Ghosh, Douglas T. Golenbock, and Ricardo T. Gazzinelli. Combined action of nucleic acid-sensing toll-like receptors and TLR11/TLR12 heterodimers imparts resistance to *toxoplasma gondii* in mice. *Cell Host and Microbe*, 13(1):42–53, 2013.
- [39] Asa Hidmark, Antonia von Saint Paul, and Alexander H. Dalpke. TLR13 Is a Receptor for Bacterial RNA. *The Journal of Immunology*, 189(6):2717–2721, 2012.
- [40] Hiroaki Hemmi, Osamu Takeuchi, Taro Kawai, Tsuneyasu Kaisho, Shintaro Sato, Hideki Sanjo, Makoto Matsumoto, Katsuaki Hoshino, Hermann Wagner, Kiyoshi Takeda, and Shizuo Akira. A Toll-like receptor recognizes bacterial DNA. *Nature*, 408(6813):740–745, 2000.
- [41] Anne Krug, Gary D. Luker, Winfried Barchet, David A. Leib, Shizuo Akira, and Marco Colonna. Herpes simplex virus type 1 activates murine natural interferon-producing cells through toll-like receptor 9. *Blood*, 103(4):1433–1437, 2004.
- [42] Jennifer Lund, Ayuko Sato, Shizuo Akira, Ruslan Medzhitov, and Akiko Iwasaki. Toll-like receptor 9-mediated recognition of Herpes simplex virus-2 by plasmacytoid dendritic cells. *Journal of Experimental Medicine*, 198(3):513–520, 2003.

- [43] Christian Bode, Gan Zhao, Folkert Steinhagen, Takeshi Kinjo, and Dennis M Klinman. CpG DNA as a vaccine adjuvant. *Expert Review of Vaccines*, 10(4):499–511, 4 2011.
- [44] Lena Alexopoulou, Agnieszka Czopik Holt, Ruslan Medzhitov, and Richard A. Flavell. Recognition of double-stranded RNA and activation of NF- κ B by Toll-like receptor 3. *Nature*, 413(6857):732–738, 10 2001.
- [45] Hiroaki Hemmi, Tsuneyasu Kaisho, Osamu Takeuchi, Shintaro Sato, Hideki Sanjo, Katsuaki Hoshino, Takao Horiuchi, Hideyuki Tomizawa, Kiyoshi Takeda, and Shizuo Akira. Small-antiviral compounds activate immune cells via the TLR7 MyD88-dependent signaling pathway. *Nature Immunology*, 3(2):196–200, 2002.
- [46] Florian Heil, Hiroaki Hemmi, Hubertus Hochrein, Franziska Ampenberger, Carsten Kirschning, Shizuo Akira, Grayson Lipford, Hermann Wagner, and Stefan Bauer. Species-Specific Recognition of Single-Stranded RNA via Toll-like Receptor 7 and 8. *Science*, 303(5663):1526–1529, 2004.
- [47] Sandra S. Diebold, Tsuneyasu Kaisho, Hiroaki Hemmi, Shizuo Akira, and Caetano Reis E Sousa. Innate Antiviral Responses by Means of TLR7-Mediated Recognition of Single-Stranded RNA. *Science*, 303(5663):1529–1531, 2004.
- [48] Jennifer M. Lund, Lena Alexopoulou, Ayuko Sato, Margaret Karow, Niels C. Adams, Nicholas W. Gale, Akiko Iwasaki, and Richard A. Flavell. Recognition of single-stranded RNA viruses by Toll-like receptor 7. *Proceedings of the National Academy of Sciences of the United States of America*, 101(15):5598–5603, 2004.
- [49] Jorge L. Cervantes, Bennett Weinerman, Chaitali Basole, and Juan C. Salazar. TLR8: The forgotten relative revindicated. *Cellular and Molecular Immunology*, 9(6):434–438, 2012.
- [50] Kenia Pedrosa Nunes, Amanda Almeida de Oliveira, Francesca Elisabeth Mowry, and Vinicia Campana Biancardi. Targeting TLR4 signaling pathways: can therapeutics pay the toll for hypertension? *British Journal of Pharmacology*, 1(334), 2018.
- [51] Marco Di Gioia and Ivan Zanoni. Toll-like receptor co-receptors as master regulators of the immune response. *Molecular Immunology*, 63(2):143–152, 2015.

- [52] Clare E. Bryant, Martyn Symmons, and Nicholas J. Gay. Toll-like receptor signalling through macromolecular protein complexes. *Molecular Immunology*, 63(2):162–165, 2015.
- [53] K Miyake. Roles for accessory molecules in microbial recognition by Toll-like receptors. *J Endotoxin Res*, 12(4):195–204, 2006.
- [54] Blake Bertani and Natividad Ruiz. Function and Biogenesis of Lipopolysaccharides. *EcoSal Plus*, 8(1):1–33, 2018.
- [55] S. D. Wright, P. S. Tobias, R. J. Ulevitch, and R. A. Ramos. Lipopolysaccharide (LPS) binding protein opsonizes LPS-bearing particles for recognition by a novel receptor on macrophages. *Journal of Experimental Medicine*, 170(4):1231–1241, 1989.
- [56] Martha Triantafilou, Kensuke Miyake, Douglas T. Golenbock, and Kathy Triantafilou. Mediators of innate immune recognition of bacteria concentrate in lipid rafts and facilitate lipopolysaccharide-induced cell activation. *Journal of Cell Science*, 115(12):2603–2611, 2002.
- [57] Ivan Zanoni, Renato Ostuni, Lorri R. Marek, Simona Barresi, Roman Barbalat, Gregory M. Barton, Francesca Granucci, and Jonathan C. Kagan. CD14 controls the LPS-induced endocytosis of toll-like receptor 4. *Cell*, 147(4):868–880, 2011.
- [58] Jonathan C Kagan, Tian Su, Tiffany Horng, Amy Chow, Shizuo Akira, and Ruslan Medzhitov. TRAM couples endocytosis of Toll-like receptor 4 to the induction of interferon- β . *Nature Immunology*, 9(4):361–368, 2008.
- [59] Yoshinori Nagai, Sachiko Akashi, Masakazu Nagafuku, Masato Ogata, Yoichiro Iwakura, Shizuo Akira, Toshio Kitamura, Atsushi Kosugi, Masao Kimoto, and Kensuke Miyake. Essential role of MD-2 in LPS responsiveness and TLR4 distribution. *Nature Immunology*, 3(7):667–672, 2002.
- [60] Sachiko Akashi, Shin Ichiroh Saitoh, Yasutaka Wakabayashi, Takane Kikuchi, Noriaki Takamura, Yoshinori Nagai, Yutaka Kusumoto, Koichi Fukase, Shoichi Kusumoto, Yoshiyuki Adachi, Atsushi Kosugi, and Kensuke Miyake. Lipopolysaccharide interaction with cell surface toll-like receptor 4-MD-2: Higher affinity than that with MD-2 or CD14. *Journal of Experimental Medicine*, 198(7):1035–1042, 2003.

- [61] A. Visintin, A. Mazzoni, J. A. Spitzer, and D. M. Segal. Secreted MD-2 is a large polymeric protein that efficiently confers lipopolysaccharide sensitivity to toll-like receptor 4. *Proceedings of the National Academy of Sciences of the United States of America*, 98(21):12156–12161, 2001.
- [62] Ricardo Núñez Miguel, Joyce Wong, Julian F. Westoll, Heather J. Brooks, Luke A.J. O’Neill, Nicholas J. Gay, Clare E. Bryant, and Tom P. Monie. A dimer of the toll-like receptor 4 cytoplasmic domain provides a specific scaffold for the recruitment of signalling adaptor proteins. *PLoS ONE*, 2(8), 2007.
- [63] Tae Ho Jang and Hyun Ho Park. Crystal structure of TIR domain of TLR6 reveals novel dimeric interface of TIR-TIR interaction for toll-like receptor signaling pathway. *Journal of Molecular Biology*, 426(19):3305–3313, 2014.
- [64] Beom Seok Park, Dong Hyun Song, Ho Min Kim, Byong-Seok Choi, Hayyoung Lee, and Jie-Oh Lee. The structural basis of lipopolysaccharide recognition by the TLR4-MD-2 complex. *Nature*, 458(7242):1191–5, 4 2009.
- [65] Luke A.J. O’Neill and Andrew G. Bowie. The family of five: TIR-domain-containing adaptors in Toll-like receptor signalling. *Nature Reviews Immunology*, 7(5):353–364, 2007.
- [66] Masahiro Yamamoto, Shintaro Sato, Hiroaki Hemmi, Hideki Sanjo, Satoshi Uematsu, Tsuneyasu Kaisho, Katsuaki Hoshino, Osamu Takeuchi, Masaya Kobayashi, Takashi Fujita, Kiyoshi Takeda, and Shizuo Akira. Essential role for TIRAP in activation of the signalling cascade shared by TLR2 and TLR4. *Nature*, 420(6913):324–9, 11 2002.
- [67] Tiffany Horng, Gregory M Barton, Richard A Flavell, and Ruslan Medzhitov. The adaptor molecule TIRAP provides signalling specificity for Toll-like receptors. *Nature*, 420(6913):329–33, 11 2002.
- [68] Kannan Badri Narayanan and Hyun Ho Park. Toll/interleukin-1 receptor (TIR) domain-mediated cellular signaling pathways. *Apoptosis*, 20(2):196–209, 2015.
- [69] Maria Loiarro, Grazia Gallo, Nicola Fantò, Rita De Santis, Paolo Carminati, Vito Ruggiero, and Claudio Sette. Identification of critical residues of the MyD88 death domain involved in the recruitment of downstream kinases. *Journal of Biological Chemistry*, 284(41):28093–28103, 2009.

- [70] Yong Chen Lu, Wen Chen Yeh, and Pamela S. Ohashi. LPS/TLR4 signal transduction pathway. *Cytokine*, 42(2):145–151, 2008.
- [71] Nicholas J. Gay, Monique Gangloff, and Luke A.J. O’Neill. What the Myddosome structure tells us about the initiation of innate immunity. *Trends in Immunology*, 32(3):104–109, 2011.
- [72] Li Deng, Chen Wang, Erika Spencer, Liyong Yang, Amy Braun, Jianxin You, Clive Slaughter, Cecile Pickart, and Zhijian J. Chen. Activation of the I κ b kinase complex by TRAF6 requires a dimeric ubiquitin-conjugating enzyme complex and a unique polyubiquitin chain. *Cell*, 103(2):351–361, 2000.
- [73] Atsuhiko Kanayama, Rashu B. Seth, Lijun Sun, Chee Kwee Ea, Mei Hong, Abdullah Shaito, Yu Hsin Chiu, Li Deng, and Zhijian J. Chen. TAB2 and TAB3 activate the NF- κ B pathway through binding to polyubiquitin chains. *Molecular Cell*, 15(4):535–548, 2004.
- [74] Jianhong Yao, Whan Kim Tae, Jinzhong Qin, Zhengfan Jiang, Youcun Qian, Hui Xiao, Yi Lu, Wen Qian, Muhammet Fatih Gulen, Nywana Sizemore, Joseph DiDonato, Shintaro Sato, Shizuo Akira, Bing Su, and Xiaoxia Li. Interleukin-1 (IL-1)-induced TAK1-dependent Versus MEKK3-dependent NF κ B activation pathways bifurcate at IL-1 receptor-associated kinase modification. *Journal of Biological Chemistry*, 282(9):6075–6089, 2007.
- [75] Ruaidhrí J. Carmody and Youhai H. Chen. Nuclear factor-kappaB: activation and regulation during toll-like receptor signaling. *Cellular & Molecular Immunology*, 4(1):31–41, 2007.
- [76] Eitan Shaulian and Michael Karin. AP-1 as a regulator of cell life and death. *Nature Cell Biology*, 4(5):E131–E136, 2002.
- [77] Luigi Franchi, Raul Muñoz-Planillo, and Gabriel Núñez. Sensing and reacting to microbes through the inflammasomes. *Nature Immunology*, 13(4):325–332, 2012.
- [78] Karen V. Swanson, Meng Deng, and Jenny P.Y. Ting. The NLRP3 inflammasome: molecular activation and regulation to therapeutics. *Nature Reviews Immunology*, 19(8):477–489, 2019.
- [79] M. Yamamoto, Shintaro Sato, Hiroaki Hemmi, Katsuaki Hoshino, Tsuneyasu Kaisho, Hideki Sanjo, Osamu Takeuchi, Masanaka Sugiyama, Masaru Okabe,

- Kiyoshi Takeda, and Shizuo Akira. Role of Adaptor TRIF in the MyD88-Independent Toll-Like Receptor Signaling Pathway. *Science*, 301(5633):640–643, 2003.
- [80] Masahiro Yamamoto, Shintaro Sato, Hiroaki Hemmi, Satoshi Uematsu, Katsuaki Hoshino, Tsuneyasu Kaisho, Osamu Takeuchi, Kiyoshi Takeda, and Shizuo Akira. TRAM is specifically involved in the Toll-like receptor 4-mediated MyD88-independent signaling pathway. *Nature Immunology*, 4(11):1144–1150, 2003.
- [81] Gagik Oganessian, Supriya K. Saha, Beichu Guo, Jeannie Q. He, Arash Shargian, Brian Zarnegar, Andrea Perry, and Genhong Cheng. Critical role of TRAF3 in the Toll-like receptor-dependent and -independent antiviral response. *Nature*, 439(7073):208–211, 2006.
- [82] Jae Young Kim, Amer A. Beg, and Eric B. Haura. Non-canonical IKKs, IKK ϵ and TBK1, as novel therapeutic targets in the treatment of non-small cell lung cancer. *Expert Opinion on Therapeutic Targets*, 17(10):1109–1112, 2013.
- [83] Kenya Honda and Tadatsugu Taniguchi. IRFs: master regulators of signalling by Toll-like receptors and cytosolic pattern-recognition receptors. *Nature Reviews Immunology*, 6(9):644–658, 2006.
- [84] Michael Carty, Rory Goodbody, Martina Schröder, Julianne Stack, Paul N. Moynagh, and Andrew G. Bowie. The human adaptor SARM negatively regulates adaptor protein TRIF-dependent Toll-like receptor signaling. *Nature Immunology*, 7(10):1074–1081, 2006.
- [85] J.Lee D.Song and Jie-Oh Lee. Sensing of microbial molecular patterns by Toll-like receptors. *Immunological Reviews*, 250:216–229, 2012.
- [86] Charles A. Dinarello. Immunological and inflammatory functions of the interleukin-1 family. *Annual Review of Immunology*, 27:519–550, 2009.
- [87] John E. Sims and Dirk E. Smith. The IL-1 family: Regulators of immunity. *Nature Reviews Immunology*, 10(2):89–102, 2010.
- [88] E. Gruys, M. J.M. Toussaint, T. A. Niewold, and S. J. Koopmans. Acute phase reaction and acute phase proteins. *Journal of Zhejiang University: Science*, 6 B(11):1045–1056, 2005.

- [89] Toshio Tanaka, Masashi Narazaki, and Tadamitsu Kishimoto. IL-6 in Inflammation, Immunity, and Disease. *Cold Spring Harbor Perspectives in Biology*, 6(10):a016295–a016295, 10 2014.
- [90] Richard M. Locksley, Nigel Killeen, and Michael J. Lenardo. The TNF and TNF receptor superfamilies: Integrating mammalian biology. *Cell*, 104(4):487–501, 2001.
- [91] George D. Kalliolias and Lionel B. Ivashkiv. TNF biology, pathogenic mechanisms and emerging therapeutic strategies. *Nature Reviews Rheumatology*, 12(1):49–62, 2016.
- [92] Bharat B. Aggarwal, Subash C. Gupta, and Ji Hye Kim. Historical perspectives on tumor necrosis factor and its superfamily: 25 years later, a golden journey. *Blood*, 119(3):651–665, 2012.
- [93] Nan Yan and Zhijian J. Chen. Intrinsic antiviral immunity. *Nature Immunology*, 13(3):214–222, 2012.
- [94] Finlay McNab, Katrin Mayer-Barber, Alan Sher, Andreas Wack, and Anne O’Garra. Type I interferons in infectious disease. *Nature Reviews Immunology*, 15(2):87–103, 2015.
- [95] Mi Sun Jin and Jie-Oh Lee. Structures of the toll-like receptor family and its ligand complexes. *Immunity*, 29(2):182–91, 8 2008.
- [96] J. Bella, K. L. Hindle, P. A. McEwan, and S. C. Lovell. The leucine-rich repeat structure. *Cellular and Molecular Life Sciences*, 65(15):2307–2333, 2008.
- [97] Bostjan Kobe and Andrey V. Kajava. The leucine-rich repeat as a protein recognition motif. *Current Opinion in Structural Biology*, 11(6):725–732, 2001.
- [98] Jin Young Kang and Jie Oh Lee. Structural biology of the toll-like receptor family. *Annual Review of Biochemistry*, 80:917–941, 2011.
- [99] Ho Min Kim, Beom Seok Park, Jung In Kim, Sung Eun Kim, Judong Lee, Se Cheol Oh, Purevjav Enkhbayar, Norio Matsushima, Hayyoung Lee, Ook Joon Yoo, and Jie Oh Lee. Crystal Structure of the TLR4-MD-2 Complex with Bound Endotoxin Antagonist Eritoran. *Cell*, 130(5):906–917, 2007.

- [100] Umeharu Ohto, Koichi Fukase, Kensuke Miyake, and Toshiyuki Shimizu. Structural basis of species-specific endotoxin sensing by innate immune receptor TLR4/MD-2. *Proceedings of the National Academy of Sciences of the United States of America*, 109(19):7421–6, 5 2012.
- [101] Umeharu Ohto, Natsuko Yamakawa, Sachiko Akashi-Takamura, Kensuke Miyake, and Toshiyuki Shimizu. Structural analyses of human toll-like receptor 4 polymorphisms D299G and T399I. *Journal of Biological Chemistry*, 287(48):40611–40617, 2012.
- [102] Shin Ichiroh Saitoh, Sachiko Akashi, Takenao Yamada, Natsuko Tanimura, Makiko Kobayashi, Kazunori Konno, Fumi Matsumoto, Koichi Fukase, Shoichi Kusumoto, Yoshinori Nagai, Yutaka Kusumoto, Atsushi Kosugi, and Kensuke Miyake. Lipid A antagonist, lipid IVa, is distinct from lipid A in interaction with Toll-like receptor 4 (TLR4)-MD-2 and ligand-induced TLR4 oligomerization. *International Immunology*, 16(7):961–969, 2004.
- [103] Lijing Su, Ying Wang, Junmei Wang, Yuto Mifune, Matthew D. Morin, Brian T. Jones, Eva Marie Y. Moresco, Dale L. Boger, Bruce Beutler, and Hong Zhang. Structural Basis of TLR2/TLR1 Activation by the Synthetic Agonist Diprovocim. *Journal of Medicinal Chemistry*, 62(6):2938–2949, 2019.
- [104] Ying Wang, Lijing Su, Matthew D. Morin, Brian T. Jones, Yuto Mifune, Hexin Shi, Kuan wen Wang, Xiaoming Zhan, Aijie Liu, Jianhui Wang, Xiaohong Li, Miao Tang, Sara Ludwig, Sara Hildebrand, Kejin Zhou, Daniel J. Siegwart, Eva Marie Y. Moresco, Hong Zhang, Dale L. Boger, and Bruce Beutler. Adjuvant effect of the novel TLR1/TLR2 agonist Diprovocim synergizes with anti-PD-L1 to eliminate melanoma in mice. *Proceedings of the National Academy of Sciences of the United States of America*, 115(37):E8698–E8706, 2018.
- [105] Jin Young Kang, Xuehua Nan, Mi Sun Jin, Suk Jun Youn, Young Hee Ryu, Shinjee Mah, Seung Hyun Han, Hayyoung Lee, Sang Gi Paik, and Jie Oh Lee. Recognition of Lipopeptide Patterns by Toll-like Receptor 2-Toll-like Receptor 6 Heterodimer. *Immunity*, 31(6):873–884, 2009.
- [106] Sung-il Yoon, Oleg Kurnasov, Venkatesh Natarajan, Minsun Hong, Andrei V Gudkov, Andrei L Osterman, and Ian A Wilson. Structural basis of TLR5-flagellin recognition and signaling. *Science*, 335(6070):859–64, 2012.

- [107] Y. Olguín, P. Villalobos, L. G. Carrascosa, M. Young, E. Valdez, L. Lechuga, and R. Galindo. Detection of flagellin by interaction with human recombinant TLR5 immobilized in liposomes. *Analytical and Bioanalytical Chemistry*, 405(4):1267–1281, 2013.
- [108] Lin Liu, Istvan Botos, Yan Wang, Joshua N Leonard, Joseph Shiloach, David M Segal, and David R Davies. Structural basis of toll-like receptor 3 signaling with double-stranded RNA. *Science*, 320(5874):379–81, 2008.
- [109] Joshua N. Leonard, Rodolfo Ghirlando, Janine Askins, Jessica K. Bell, David H. Margulies, David R. Davies, and David M. Segal. The TLR3 signaling complex forms by cooperative receptor dimerization. *Proceedings of the National Academy of Sciences of the United States of America*, 105(1):258–263, 2008.
- [110] Istvan Botos, Lin Liu, Yan Wang, David M. Segal, and David R. Davies. The Toll-like receptor 3:dsRNA signaling complex. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, 1789(9-10):667–674, 2009.
- [111] Hiromi Tanji, Umeharu Ohto, Takuma Shibata, Masato Taoka, Yoshio Yamauchi, Toshiaki Isobe, Kensuke Miyake, and Toshiyuki Shimizu. Toll-like receptor 8 senses degradation products of single-stranded RNA. *Nature Structural and Molecular Biology*, 22(2):109–116, 2015.
- [112] Umeharu Ohto, Takuma Shibata, Hiromi Tanji, Hanako Ishida, Elena Krayukhina, Susumu Uchiyama, Kensuke Miyake, and Toshiyuki Shimizu. Structural basis of CpG and inhibitory DNA recognition by Toll-like receptor 9. *Nature*, 520(7549):702–705, 2015.
- [113] H. von Bernuth, C. Picard, Z. Jin, R. Pankla, H. Xiao, C.-L. Ku, M. Chrabieh, I. B. Mustapha, P. Ghandil, Y. Camcioglu, J. Vasconcelos, N. Sirvent, M. Guedes, A. B. Vitor, M. J. Herrero-Mata, J. I. Arostegui, C. Rodrigo, L. Alsina, E. Ruiz-Ortiz, M. Juan, C. Fortuny, J. Yague, J. Anton, M. Pascal, H.-H. Chang, L. Janniere, Y. Rose, B.-Z. Garty, H. Chapel, A. Issekutz, L. Marodi, C. Rodriguez-Gallego, J. Banchereau, L. Abel, X. Li, D. Chaussabel, A. Puel, and J.-L. Casanova. Pyogenic Bacterial Infections in Humans with MyD88 Deficiency. *Science*, 321(5889):691–696, 8 2008.

- [114] David Walsh, Joanna McCarthy, Caitriona O’Driscoll, and Silvia Melgar. Pattern recognition receptors-Molecular orchestrators of inflammation in inflammatory bowel disease. *Cytokine and Growth Factor Reviews*, 24(2):91–104, 2013.
- [115] K. S. Midwood and A. M. Piccinini. DAMPening inflammation by modulating TLR signalling. *Mediators of Inflammation*, 2010(672395), 2010.
- [116] J. J. De Haan, M. B. Smeets, G. Pasterkamp, and F. Arslan. Danger signals in the initiation of the inflammatory response after myocardial infarction. *Mediators of Inflammation*, 2013, 2013.
- [117] Li Yu, Liantang Wang, and Shangwu Chen. Endogenous toll-like receptor ligands and their biological significance. *Journal of Cellular and Molecular Medicine*, 14(11):2592–2603, 2010.
- [118] Marco E. Bianchi. DAMPs, PAMPs and alarmins: all we need to know about danger. *Journal of Leukocyte Biology*, 81(1):1–5, 2007.
- [119] Clare E. Bryant, Nick J. Gay, Stephane Heymans, Sandra Sacre, Liliana Schaefer, and Kim S. Midwood. Advances in Toll-like receptor biology: Modes of activation by diverse stimuli. *Critical Reviews in Biochemistry and Molecular Biology*, 50(5):359–379, 2015.
- [120] Mansi Saxena and Garabet Yeretssian. NOD-like receptors: Master regulators of inflammation and cancer. *Frontiers in Immunology*, 5:1–16, 2014.
- [121] Young Keun Kim, Jeon Soo Shin, and Moon H. Nahm. NOD-like receptors in infection, immunity, and diseases. *Yonsei Medical Journal*, 57(1):5–14, 2016.
- [122] Jan Rehwinkel and Michaela U. Gack. RIG-I-like receptors: their regulation and roles in RNA sensing. *Nature Reviews Immunology*, 20(9):537–551, 2020.
- [123] Thomas Vogl, Klaus Tenbrock, Stephan Ludwig, Nadja Leukert, Christina Ehrhardt, Marieke A.D. Van Zoelen, Wolfgang Nacken, Dirk Foell, Tom Van Der Poll, Clemens Sorg, and Johannes Roth. Mrp8 and Mrp14 are endogenous activators of Toll-like receptor 4, promoting lethal, endotoxin-induced shock. *Nature Medicine*, 13(9):1042–1049, 2007.
- [124] Per Björk, Anders Björk, Thomas Vogl, Martin Stenström, David Liberg, Anders Olsson, Johannes Roth, Fredrik Ivars, and Tomas Leanderson. Identification of human S100A9 as a novel target for treatment of autoimmune disease via binding to quinoline-3-carboxamides. *PLoS Biology*, 7(4):0800–0812, 2009.

- [125] Thomas Vogl, Athanasios Stratis, Viktor Wixler, Tom Völler, Sumita Thuraiyagam, Selina K. Jorch, Stefanie Zenker, Alena Dreiling, Deblina Chakraborty, Mareike Fröhling, Peter Paruzel, Corinna Wehmeyer, Sven Hermann, Olympia Papantonopoulou, Christiane Geyer, Karin Loser, Michael Schäfers, Stephan Ludwig, Monika Stoll, Tomas Leanderson, Joachim L. Schultze, Simone König, Thomas Pap, and Johannes Roth. Autoinhibitory regulation of S100A8/S100A9 alarmin activity locally restricts sterile inflammation. *Journal of Clinical Investigation*, 128(5):1852–1866, 2018.
- [126] Zhifei He, Matteo Riva, Per Björk, Karl Swärd, Matthias Mörgelin, Tomas Leanderson, and Fredrik Ivars. CD14 is a co-receptor for TLR4 in the S100A9-induced pro-inflammatory response in monocytes. *PLoS ONE*, 11(5):1–15, 2016.
- [127] H. Christian Volz, Danai Laohachewin, Cathrin Seidel, Felix Lasitschka, Kirsten Keilbach, Alexandra R. Wienbrandt, Joachim Andrassy, Angelika Bierhaus, Ziya Kaya, Hugo A. Katus, and Martin Andrassy. S100A8/A9 aggravates post-ischemic heart failure through activation of RAGE-dependent NF- κ B signaling. *Basic Research in Cardiology*, 107(2), 2012.
- [128] Yoshinori Okamura, Michiko Watari, Elliot S. Jerud, Donna W. Young, Sally T. Ishizaka, Jeffrey Rose, Jesse C. Chow, and Jerome F. Strauss. The Extra Domain A of Fibronectin Activates Toll-like Receptor 4. *Journal of Biological Chemistry*, 276(13):10229–10233, 2001.
- [129] Ran You, Mingzhe Zheng, and Paula J. McKeown-Longo. The first type III repeat in fibronectin activates an inflammatory pathway in dermal fibroblasts. *Journal of Biological Chemistry*, 285(47):36255–36259, 2010.
- [130] Juan J. Lasarte, Noelia Casares, Marta Gorraiz, Sandra Hervás-Stubbs, Laura Arribillaga, Cristina Mansilla, Maika Durantez, Diana Llopiz, Pablo Sarobe, Francisco Borrás-Cuesta, Jesús Prieto, and Claude Leclerc. The Extra Domain A from Fibronectin Targets Antigens to TLR4-Expressing Cells and Induces Cytotoxic T Cell Responses In Vivo. *The Journal of Immunology*, 178(2):748–756, 2007.
- [131] Swati Bhattacharyya, Zenshiro Tamaki, Wenxia Wang, Monique Hinchcliff, Paul Hoover, Spiro Getsios, Eric S. White, and John Varga. FibronectinEDA promotes chronic cutaneous fibrosis through toll-like receptor signaling. *Science Translational Medicine*, 6(232), 2014.

- [132] Martin Kohan, Andres F. Muro, Eric S. White, and Neville Berkman. EDA-containing cellular fibronectin induces fibroblast differentiation through binding to $\alpha 4 \beta 7$ integrin receptor and MAPK/Erk 1/2-dependent signaling . *The FASEB Journal*, 24(11):4503–4512, 2010.
- [133] Arti V. Shinde, Christopher Bystroff, Chunyu Wang, Mariette G. Voegelzang, Peter A. Vincent, Richard O. Hynes, and Livingston Van De Water. Identification of the peptide sequences within the EIIIA (EDA) segment of fibronectin that mediate integrin $\alpha 9 \beta 1$ -dependent cellular activities. *Journal of Biological Chemistry*, 283(5):2858–2870, 2008.
- [134] Yung Feng Liao, Philip J. Gotwals, Victor E. Kotliansky, Dean Sheppard, and Livingston De Van Water. The EIIIA segment of fibronectin is a ligand for integrins $\alpha 9 \beta 1$ and $\alpha 9 \beta 1$ providing a novel mechanism for regulating cell adhesion by alternative splicing. *Journal of Biological Chemistry*, 277(17):14467–14474, 2002.
- [135] Rhiannon M. Kelsh-Lasher, Anthony Ambesi, Ceyda Bertram, and Paula J. McKeown-Longo. Integrin $\alpha 4 \beta 1$ and TLR4 Cooperate to Induce Fibrotic Gene Expression in Response to Fibronectin’s EDA Domain. *Journal of Investigative Dermatology*, 137(12):2505–2512, 2017.
- [136] Stephen T. Smiley, Jennifer A. King, and Wayne W. Hancock. Fibrinogen Stimulates Macrophage Chemokine Secretion Through Toll-Like Receptor 4. *The Journal of Immunology*, 167(5):2887–2894, 2001.
- [137] Torstein Jensen, Peter Kierulf, Per M. Sandset, Olav Klingenberg, Gunn B. Joø, Hans C. Godal, and Ole H. Skjøsberg. Fibrinogen and fibrin induce synthesis of proinflammatory cytokines from isolated peripheral blood mononuclear cells. *Thrombosis and Haemostasis*, 97(5):822–829, 2007.
- [138] Douglas B. Kuhns, Debra A. Long Priel, and John I. Gallin. Induction of human monocyte interleukin (IL)-8 by fibrinogen through the toll-like receptor pathway. *Inflammation*, 30(5):178–188, 2007.
- [139] Conrad P. Hodgkinson, Kunal Patel, and Shu Ye. Functional Toll-like receptor 4 mutations modulate the response to fibrinogen. *Thrombosis and Haemostasis*, 100(2):301–307, 2008.
- [140] Cameron T. Landers, Hui Ying Tung, J. Morgan Knight, Matthew C. Madison, Yifan Wu, Zhimin Zeng, Paul C. Porter, Antony Rodriguez, Matthew J. Flick,

- Farrah Kheradmand, and David B. Corry. Selective cleavage of fibrinogen by diverse proteinases initiates innate allergic and antifungal immunity through CD11b. *Journal of Biological Chemistry*, 294(22):8834–8847, 2019.
- [141] Jeremy Sokolove, Xiaoyan Zhao, Piyanka E. Chandra, and William H. Robinson. Immune complexes containing citrullinated fibrinogen costimulate macrophages via toll-like receptor 4 and Fc γ receptor. *Arthritis and Rheumatism*, 63(1):53–62, 2011.
- [142] Kim Midwood, Sandra Sacre, Anna M Piccinini, Julia Inglis, Annette Trebaul, Emma Chan, Stefan Drexler, Nidhi Sofat, Masahide Kashiwagi, Gertraud Orend, Fionula Brennan, and Brian Foxwell. Tenascin-C is an endogenous activator of Toll-like receptor 4 that is essential for maintaining inflammation in arthritic joint disease. *Nature Medicine*, 15(7):774–780, 2009.
- [143] L Patel, W Sun, S S Glasson, E A Morris, C R Flannery, and P S Chockalingam. Tenascin-C induces inflammatory mediators and matrix degradation in osteoarthritic cartilage. *BMC Musculoskeletal Disorders*, 12(164):1–14, 2011.
- [144] Lorena Zuliani-Alvarez, Anna Piccinini, and Kim Midwood. Screening for Novel Endogenous Inflammatory Stimuli Using the Secreted Embryonic Alkaline Phosphatase NF- κ B Reporter Assay. *Bio-Protocol*, 7(7):1–12, 2017.
- [145] Azhar Maqbool, Emma J Spary, Iain W Manfield, Michaela Ruhmann, Lorena Zuliani-Alvarez, Filomena O Gamboa-Esteves, Karen E Porter, Mark J Drinkhill, Kim S Midwood, and Neil A Turner. Tenascin C upregulates interleukin-6 expression in human cardiac myofibroblasts via toll-like receptor 4 . *World Journal of Cardiology*, 8(5):340, 2016.
- [146] Swati Bhattacharyya, Wenxia Wang, Luisa Morales-Nebreda, Gang Feng, Minghua Wu, Xiaodong Zhou, Robert Lafyatis, Jungwha Lee, Monique Hinchcliff, Carol Feghali-Bostwick, Katja Lakota, G. R. Scott Budinger, Kirtee Raparia, Zenshiro Tamaki, and John Varga. Tenascin-C drives persistence of organ fibrosis. *Nature Communications*, 7:11703, 2016.
- [147] Rui Liu, Yong He, Bo Li, Jun Liu, Yingang Ren, Wei Han, Xing Wang, and Lihua Zhang. Tenascin-C produced by oxidized LDL-stimulated macrophages increases foam cell formation through toll-like receptor-4. *Molecules and Cells*, 34(1):35–41, 2012.

- [148] Anna M. Piccinini, Lorena Zuliani-Alvarez, Jenny M.P. Lim, and Kim S. Midwood. Distinct microenvironmental cues stimulate divergent TLR4-mediated signaling pathways in macrophages. *Science Signaling*, 9(443), 2016.
- [149] Verena Haage, Nirmeen Elmadany, Lars Roll, Andreas Faissner, David H. Gutmann, Marcus Semtner, and Helmut Kettenmann. Tenascin C regulates multiple microglial functions involving TLR4 signaling and HDAC1. *Brain, Behavior, and Immunity*, 81(July):470–483, 2019.
- [150] Liliana Schaefer, Andrea Babelova, Eva Kiss, Heinz J. Hausser, Martina Baliova, Miroslava Krzyzankova, Gunther Marsche, Marian F. Young, Daniel Mihalik, Martin Götte, Ernst Malle, Roland M. Schaefer, and Hermann Josef Gröne. The matrix component biglycan is proinflammatory and signals through Toll-like receptors 4 and 2 in macrophages. *Journal of Clinical Investigation*, 115(8):2223–2233, 2005.
- [151] Rosetta Merline, Kristin Moreth, Janet Beckmann, Madalina V. Nastase, Jinyang Zeng-Brouwers, José Guilherme Tralhão, Patricia Lemarchand, Josef Pfeilschifter, Roland M. Schaefer, Renato V. Iozzo, and Liliana Schaefer. Signaling by the matrix proteoglycan decorin controls inflammation and cancer through PDCD4 and microRNA-21. *Science Signaling*, 4(199):1–15, 2011.
- [152] Kristin Moreth, Helena Frey, Mario Hubo, Jinyang Zeng-Brouwers, Madalina Viviana Nastase, Louise Tzung Harn Hsieh, Riad Haceni, Josef Pfeilschifter, Renato V. Iozzo, and Liliana Schaefer. Biglycan-triggered TLR-2- and TLR-4-signaling exacerbates the pathophysiology of ischemic acute kidney injury. *Matrix Biology*, 35:143–151, 2014.
- [153] Thomas Neill, Liliana Schaefer, and Renato V. Iozzo. Decoding the Matrix: Instructive Roles of Proteoglycan Receptors. *Biochemistry*, 54(30):4583–4598, 2015.
- [154] Kristin Moreth, Renato V. Iozzo, and Liliana Schaefer. Small leucine-rich proteoglycans orchestrate receptor crosstalk during inflammation. *Cell Cycle*, 11(11):2084–2091, 2012.
- [155] Andrea Babelova, Kristin Moreth, Wasiliki Tsalastra-Greul, Jinyang Zeng-Brouwers, Oliver Eickelberg, Marian F. Young, Peter Brucker, Josef Pfeilschifter, Roland M. Schaefer, Hermann Josef Gröne, and Liliana Schaefer.

Biglycan, a danger signal that activates the NLRP3 inflammasome via toll-like and P2X receptors. *Journal of Biological Chemistry*, 284(36):24035–24048, 2009.

- [156] Agnes D. Berendsen, Larry W. Fisher, Tina M. Kilts, Rick T. Owens, Pamela G. Robey, J. Silvio Gutkind, and Marian F. Younga. Modulation of canonical Wnt signaling by the extracellular matrix component biglycan. *Proceedings of the National Academy of Sciences of the United States of America*, 108(41):17022–17027, 2011.
- [157] Jeremy Kiripolsky, Rose Anne Romano, Eileen M. Kasperek, Guan Yu, and Jill M. Kramer. Activation of Myd88-Dependent TLRs Mediates Local and Systemic Inflammation in a Mouse Model of Primary Sjögren’s Syndrome. *Frontiers in Immunology*, 10(January), 2020.
- [158] Jinyang Zeng-Brouwers, Sony Pandey, Jonel Trebicka, Malgorzata Wygrecka, and Liliana Schaefer. Communications via the Small Leucine-rich Proteoglycans: Molecular Specificity in Inflammation and Autoimmune Diseases. *Journal of Histochemistry and Cytochemistry*, 68(12):887–906, 2020.
- [159] Heiko Roedig, Madalina V. Nastase, Helena Frey, Kristin Moreth, Jinyang Zeng-Brouwers, Chiara Poluzzi, Louise Tzung Harn Hsieh, Christian Brandts, Simone Fulda, Malgorzata Wygrecka, and Liliana Schaefer. Biglycan is a new high-affinity ligand for CD14 in macrophages. *Matrix Biology*, 77:4–22, 2019.
- [160] Chiara Poluzzi, Madalina Viviana Nastase, Jinyang Zeng-Brouwers, Heiko Roedig, Louise Tzung Harn Hsieh, Jonas B. Michaelis, Eva Miriam Buhl, Flavia Rezende, Yosif Manavski, André Bleich, Peter Boor, Ralf P. Brandes, Josef Pfeilschifter, Ernst H.K. Stelzer, Christian Münch, Ivan Dikic, Christian Brandts, Renato V. Iozzo, Malgorzata Wygrecka, and Liliana Schaefer. Biglycan evokes autophagy in macrophages via a novel CD44/Toll-like receptor 4 signaling axis in ischemia/reperfusion injury. *Kidney International*, 95(3):540–562, 2019.
- [161] Yury I. Miller, Suganya Viriyakosol, Christoph J. Binder, James R. Feramisco, Theo N. Kirkland, and Joseph L. Witztum. Minimally modified LDL binds to CD14, induces macrophage spreading via TLR4/MD-2, and inhibits phagocytosis of apoptotic cells. *Journal of Biological Chemistry*, 278(3):1561–1568, 2003.

- [162] Kimberly A. Walton, Xavier Hsieh, Nima Gharavi, Shirley Wang, Grace Wang, Michael Yeh, Amy L. Cole, and Judith A. Berliner. Receptors involved in the oxidized 1-palmitoyl-2-arachidonoyl-sn-glycero-3-phosphorylcholine-mediated synthesis of interleukin-8: A role for toll-like receptor 4 and a glycosylphosphatidylinositol-anchored protein. *Journal of Biological Chemistry*, 278(32):29661–29666, 2003.
- [163] Yury I. Miller, Mi Kyung Chang, Christoph J. Binder, Peter X. Shaw, and Joseph L. Witztum. Oxidized low density lipoprotein and innate immune receptors. *Current Opinion in Lipidology*, 14(5):437–445, 2003.
- [164] Cameron R. Stewart, Lynda M. Stuart, Kim Wilkinson, Janine M. Van Gils, Jiusheng Deng, Annett Halle, Katey J. Rayner, Laurent Boyer, Ruiqin Zhong, William A. Frazier, Adam Lacy-Hulbert, Joseph El Khoury, Douglas T. Golenbock, and Kathryn J. Moore. CD36 ligands promote sterile inflammation through assembly of a Toll-like receptor 4 and 6 heterodimer. *Nature Immunology*, 11(2):155–161, 2010.
- [165] Ke Yang, Xinhe Liu, Yan Liu, Xinqiong Wang, Lijuan Cao, Xiaojie Zhang, Chundi Xu, Weifeng Shen, and Tong Zhou. DC-SIGN and Toll-like receptor 4 mediate oxidized low-density lipoprotein-induced inflammatory responses in macrophages. *Scientific Reports*, 7(1):1–11, 2017.
- [166] Luis Chávez-Sánchez, Montserrat Guadalupe Garza-Reyes, José Esteban Espinosa-Luna, Karina Chávez-Rueda, María Victoria Legorreta-Haquet, and Francisco Blanco-Favela. The role of TLR2, TLR4 and CD36 in macrophage activation and foam cell formation in response to oxLDL in humans. *Human Immunology*, 75(4):322–329, 2014.
- [167] Silke Walter, Maryse Letiembre, Yang Liu, Holger Heine, Botond Penke, Wenlin Hao, Barbara Bode, Nicole Manietta, Jessica Walter, Walter Schulz-Schüffer, and Klaus Faßbender. Role of the toll-like receptor 4 in neuroinflammation in Alzheimer’s disease. *Cellular Physiology and Biochemistry*, 20(6):947–956, 2007.
- [168] Atsuko Deguchi, Takeshi Tomita, Tsutomu Omori, Akiko Komatsu, Umeharu Ohto, Satoshi Takahashi, Natsuko Tanimura, Sachiko Akashi-Takamura, Kensuke Miyake, and Yoshiro Maru. Serum Amyloid A3 Binds MD-2 To Activate p38 and NF- κ B Pathways in a MyD88-Dependent Manner. *The Journal of Immunology*, 191(4):1856–1864, 2013.

- [169] Shirong Liu, Yang Liu, Wenlin Hao, Lisa Wolf, Amanda J. Kiliaan, Botond Penke, Claudia E. Rübe, Jochen Walter, Michael T. Heneka, Tobias Hartmann, Michael D. Menger, and Klaus Fassbender. TLR2 Is a Primary Receptor for Alzheimer’s Amyloid β Peptide To Trigger Neuroinflammatory Activation. *The Journal of Immunology*, 188(3):1098–1107, 2012.
- [170] Maria L.D. Udan, Deepa Ajit, Nikkilina R. Crouse, and Michael R. Nichols. Toll-like receptors 2 and 4 mediate A β (1-42) activation of the innate immune response in a human monocytic cell line. *Journal of Neurochemistry*, 104(2):524–533, 2008.
- [171] Heledd H. Jarosz-Griffiths, Elizabeth Noble, Jo V. Rushworth, and Nigel M. Hooper. Amyloid- β receptors: The good, the bad, and the prion protein. *Journal of Biological Chemistry*, 291(7):3174–3183, 2016.
- [172] Jong Sung Park, Daiva Svetkauskaite, Qianbin He, Jae Yeol Kim, Derek Strassheim, Akitoshi Ishizaka, and Edward Abraham. Involvement of Toll-like Receptors 2 and 4 in Cellular Activation by High Mobility Group Box 1 Protein. *Journal of Biological Chemistry*, 279(9):7370–7377, 2004.
- [173] Man Yu, Haichao Wang, Aihao Ding, Douglas T. Golenbock, Eicke Latz, Christopher J. Czura, Matthew J. Fenton, Kevin J. Tracey, and Huan Yang. HMGB1 signals through toll-like receptor (TLR) 4 and TLR2. *Shock*, 26(2):174–179, 2006.
- [174] Jong Sung Park, Fabia Gamboni-Robertson, Qianbin He, Daiva Svetkauskaite, Jae-Yeol Kim, Derek Strassheim, Jang-Won Sohn, Shingo Yamada, Ikuro Maruyama, Anirban Banerjee, Akitoshi Ishizaka, and Edward Abraham. High mobility group box 1 protein interacts with multiple Toll-like receptors. *American Journal of Physiology-Cell Physiology*, 290(3):C917–C924, 3 2006.
- [175] Huan Yang, Hulda S. Hreggvidsdottir, Karin Palmblad, Haichao Wang, Mahendar Ochani, Jianhua Li, Ben Lu, Sangeeta Chavan, Mauricio Rosas-Ballina, Yousef Al-Abed, Shizuo Akira, Angelika Bierhaus, Helena Erlandsson-Harris, Ulf Andersson, and Kevin J. Tracey. A critical cysteine is required for HMGB1 binding to toll-like receptor 4 and activation of macrophage cytokine release. *Proceedings of the National Academy of Sciences of the United States of America*, 107(26):11942–11947, 2010.

- [176] Mingzhu He, Marco E. Bianchi, Tom R. Coleman, Kevin J. Tracey, and Yousef Al-Abed. Exploring the biological functional mechanism of the HMGB1/TLR4/MD-2 complex by surface plasmon resonance. *Molecular Medicine*, 24(1):1–9, 2018.
- [177] Christian Termeer, Frauke Benedix, Jonathon Sleeman, Christina Fieber, Ursula Voith, Thomas Ahrens, Kensuke Miyake, Marina Freudenberg, Christopher Galanos, and Jan Christoph Simon. Oligosaccharides of hyaluronan activate dendritic cells via Toll-like receptor 4. *Journal of Experimental Medicine*, 195(1):99–111, 2002.
- [178] Kristen R. Taylor, Janet M. Trowbridge, Jennifer A. Rudisill, Christian C. Termeer, Jan C. Simon, and Richard L. Gallo. Hyaluronan Fragments Stimulate Endothelial Recognition of Injury through TLR4. *Journal of Biological Chemistry*, 279(17):17079–17084, 2004.
- [179] Dianhua Jiang, Jiurong Liang, Juan Fan, Shuang Yu, Suping Chen, Yi Luo, Glenn D. Prestwich, Marcella M. Mascarenhas, Hari G. Garg, Deborah A. Quinn, Robert J. Homer, Daniel R. Goldstein, Richard Bucala, Patty J. Lee, Ruslan Medzhitov, and Paul W. Noble. Regulation of lung injury and repair by Toll-like receptors and hyaluronan. *Nature Medicine*, 11(11):1173–1179, 2005.
- [180] Dianhua Jiang, Jiurong Liang, and Paul W. Noble. Hyaluronan as an immune regulator in human diseases. *Physiological Reviews*, 91(1):221–264, 2011.
- [181] Sarbjeet Makkar, Terrence E. Riehl, Baosheng Chen, Yan Yan, David M. Alvarado, Matthew A. Ciorba, and William F. Stenson. Hyaluronic acid binding to TLR4 promotes proliferation and blocks apoptosis in colon cancer. *Molecular Cancer Therapeutics*, 18(12):2446–2456, 2019.
- [182] Kristen R. Taylor, Kenshi Yamasaki, Katherine A. Radek, Anna Di Nardo, Heidi Goodarzi, Douglas Golenbock, Bruce Beutler, and Richard L. Gallo. Recognition of hyaluronan released in sterile injury involves a unique receptor complex dependent on toll-like receptor 4, CD44, and MD-2. *Journal of Biological Chemistry*, 282(25):18265–18275, 2007.
- [183] Kara A. Scheibner, Michael A. Lutz, Sada Boodoo, Matthew J. Fenton, Jonathan D. Powell, and Maureen R. Horton. Hyaluronan Fragments Act as an Endogenous Danger Signal by Engaging TLR2. *The Journal of Immunology*, 177(2):1272–1281, 2006.

- [184] Aisha Zaman, Zheng Cui, Joseph P. Foley, Hengjiang Zhao, Paul C. Grimm, Horace M. DeLisser, and Rashmin C. Savani. Expression and role of the hyaluronan receptor RHAMM in inflammation after bleomycin injury. *American Journal of Respiratory Cell and Molecular Biology*, 33(5):447–454, 2005.
- [185] Rainer Ebid, Julia Lichtnekert, and Hans-Joachim Anders. Hyaluronan Is Not a Ligand but a Regulator of Toll-Like Receptor Signaling in Mesangial Cells: Role of Extracellular Matrix in Innate Immunity. *ISRN Nephrology*, 2014:1–11, 2014.
- [186] Ramanjaneyulu Allam, Christina Rebecca Scherbaum, Murthy Narayana Darisipudi, Shrikant R. Mulay, Holger Hägele, Julia Lichtnekert, Jan Henrik Hagemann, Khader Valli Rupanagudi, Mi Ryu, Claudia Schwarzenberger, Bernd Hohenstein, Christian Hugo, Bernd Uhl, Christoph A. Reichel, Fritz Krombach, Marc Monestier, Helen Liapis, Kristin Moreth, Liliana Schaefer, and Hans Joachim Anders. Histones from dying renal cells aggravate kidney injury via TLR2 and TLR4. *Journal of the American Society of Nephrology*, 23(8):1375–1388, 2012.
- [187] Jun Xu, Xiaomei Zhang, Rosana Pelayo, Marc Monestier, Concetta T. Ammollo, Fabrizio Semeraro, Fletcher B. Taylor, Naomi L. Esmon, Florea Lupu, and Charles T. Esmon. Extracellular histones are major mediators of death in sepsis. *Nature Medicine*, 15(11):1318–1321, 2009.
- [188] Yasunari Munemasa. Histone H2B induces retinal ganglion cell death through toll-like receptor 4 in the vitreous of acute primary angle closure patients. *Laboratory Investigation*, 100(8):1080–1089, 2020.
- [189] Theodora Dorita Tsourouktsoglou, Annika Warnatsch, Marianna Ioannou, Dennis Hoving, Qian Wang, and Venizelos Papayannopoulos. Histones, DNA, and Citrullination Promote Neutrophil Extracellular Trap Inflammation by Regulating the Localization and Activation of TLR4. *Cell Reports*, 31(5):107602, 2020.
- [190] Jun Xu, Xiaomei Zhang, Marc Monestier, Naomi L. Esmon, and Charles T. Esmon. Extracellular Histones Are Mediators of Death through TLR2 and TLR4 in Mouse Fatal Liver Injury. *The Journal of Immunology*, 187(5):2626–2631, 2011.

- [191] Hai Huang, John Evankovich, Wei Yan, Gary Nace, Lemeng Zhang, Mark Ross, Xinghua Liao, Timothy Billiar, Jun Xu, Charles T. Esmon, and Allan Tsung. Endogenous histones function as alarmins in sterile inflammatory liver injury through Toll-like receptor 9 in mice. *Hepatology*, 54(3):999–1008, 2011.
- [192] Fabrizio Semeraro, Concetta T. Ammollo, James H. Morrissey, George L. Dale, Paul Friese, Naomi L. Esmon, and Charles T. Esmon. Extracellular histones promote thrombin generation through platelet-dependent mechanisms: Involvement of platelet TLR2 and TLR4. *Blood*, 118(7):1952–1961, 2011.
- [193] Ji Eun Kim, Hyun Ju Yoo, Ja Yoon Gu, and Hyun Kyung Kim Kim. Histones induce the procoagulant phenotype of endothelial cells through tissue factor up-regulation and thrombomodulin down-regulation. *PLoS ONE*, 11(6):1–14, 2016.
- [194] Durba Pal, Suman Dasgupta, Rakesh Kundu, Sudipta Maitra, Gobardhan Das, Satinath Mukhopadhyay, Sukanta Ray, Subeer S. Majumdar, and Samir Bhattacharya. Fetuin-A acts as an endogenous ligand of TLR4 to promote lipid-induced insulin resistance. *Nature Medicine*, 18(8):1279–1285, 2012.
- [195] Ximei Shen, Liyong Yang, Sunjie Yan, Huanhuan Zheng, Liyu Liang, Xiuhui Cai, and Meng Liao. Fetuin A promotes lipotoxicity in β cells through the TLR4 signaling pathway and the role of pioglitazone in anti-lipotoxicity. *Molecular and Cellular Endocrinology*, 412:1–11, 2015.
- [196] Portia L. Thomas, Gladys Nangami, Tanu Rana, Adam Evans, Stephen D. Williams, Dylan Crowell, Anil Shanker, Amos M. Sakwe, and Josiah Ochieng. The rapid endocytic uptake of fetuin-A by adherent tumor cells is mediated by Toll-like receptor 4 (TLR4). *FEBS Open Bio*, 10(12):2722–2732, 2020.
- [197] Anton Scott Goustin and Abdul B. Abou-Samra. The 'thrifty' gene encoding Ahsg/Fetuin-A meets the insulin receptor: Insights into the mechanism of insulin resistance. *Cellular Signalling*, 23(6):980–990, 2011.
- [198] Michael Demetriou, Christoph Binkert, Balram Sukhu, Howard C. Tenenbaum, and James W. Dennis. Fetuin/ α 2-HS glycoprotein is a transforming growth factor- β type II receptor mimic and cytokine antagonist. *Journal of Biological Chemistry*, 271(22):12755–12761, 1996.

- [199] Dorothea I. Siegel-Axel, Susanne Ullrich, Norbert Stefan, Kilian Rittig, Felicia Gerst, Christian Klingler, Ulrike Schmidt, Birgit Schreiner, Elko Randrianarisoa, Hans Eberhard Schaller, Ulrich A. Stock, Cora Weigert, Alfred Königsrainer, and Hans Ulrich Häring. Fetuin-A influences vascular cell growth and production of proinflammatory and angiogenic proteins by human perivascular fat cells. *Diabetologia*, 57(5):1057–1066, 2014.
- [200] Satinath Mukhopadhyay, Samim A. Mondal, Manoj Kumar, and Deep Dutta. Proinflammatory and Antiinflammatory Attributes of Fetu I ν -A: A Novel Hepatokine Modulating Cardiovascular and Glycemic Outcomes in Metabolic Syndrome. *Endocrine Practice*, 20(12):1345–1351, 12 2014.
- [201] Alexander R. Moschen, Arthur Kaser, Barbara Enrich, Birgit Mosheimer, Milan Theurl, Harald Niederegger, and Herbert Tilg. Visfatin, an Adipocytokine with Proinflammatory and Immunomodulating Properties. *The Journal of Immunology*, 178(3):1748–1758, 2007.
- [202] Antonella Managò, Valentina Audrito, Francesca Mazzola, Leonardo Sorci, Federica Gaudino, Katuscia Gizzi, Nicoletta Vitale, Danny Incarnato, Gabriele Minazzato, Alice Ianniello, Antonio Varriale, Sabato D Auria, Giulio Mengozzi, Gianfranco Politano, Salvatore Oliviero, and Nadia Raffaelli. Extracellular nicotinate phosphoribosyltransferase binds Toll like receptor 4 and mediates inflammation. *Nature Communications*, pages 1–14, 2019.
- [203] Sara M. Camp, Ermelinda Ceco, Carrie L. Evenoski, Sergei M. Danilov, Tong Zhou, Eddie T. Chiang, Liliana Moreno-Vinasco, Brandon Mapes, Jieling Zhao, Gamze Gursoy, Mary E. Brown, Djanybek M. Adyshev, Shahid S. Siddiqui, Hector Quijada, Saad Sammani, Eleftheria Letsiou, Laleh Saadat, Mohammed Yousef, Ting Wang, Jie Liang, and Joe G.N. Garcia. Unique Toll-Like Receptor 4 Activation by NAMPT/PBEF Induces NF κ B Signaling and Inflammatory Lung Injury. *Scientific Reports*, 5(February):1–14, 2015.
- [204] Valentina Audrito, Vincenzo Gianluca Messina, and Silvia Deaglio. NAMPT and NAPRT: Two Metabolic Enzymes With Key Roles in Inflammation. *Frontiers in Oncology*, 10(March):1–17, 2020.
- [205] Trivikram R. Molugu, Radu C. Oita, Udeep Chawla, Sara M. Camp, Michael F. Brown, and Joe G.N. Garcia. Nicotinamide phosphoribosyltransferase purification using SUMO expression system. *Analytical Biochemistry*, 598(January):113597, 2020.

- [206] Antje Garten, Susanne Schuster, Melanie Penke, Theresa Gorski, Tommaso De Giorgis, and Wieland Kiess. Physiological and pathophysiological roles of NAMPT and NAD metabolism. *Nature Reviews Endocrinology*, 11(9):535–546, 2015.
- [207] Saurabh Yadav and Avadhesh Surolia. Lysozyme elicits pain during nerve injury by neuronal Toll-like receptor 4 activation and has therapeutic potential in neuropathic pain. *Science Translational Medicine*, 11(504):eaav4176, 2019.
- [208] Jennifer F.A. Swisher, Nicholas Burton, Silvia M. Bacot, Stefanie N. Vogel, and Gerald M. Feldman. Annexin A2 tetramer activates human and murine macrophages through TLR4. *Blood*, 115(3):549–558, 2010.
- [209] Shuang Zhang, Min Yu, Qiang Guo, Rongpeng Li, Guobo Li, Shirui Tan, Xuefeng Li, Yuquan Wei, and Min Wu. Annexin A2 binds to endosomes and negatively regulates TLR4-triggered inflammatory responses via the TRAM-TRIF pathway. *Scientific Reports*, 5(May):1–15, 2015.
- [210] Kristi L. Allen, Fabio V. Fonseca, Venkaiah Betapudi, Belinda Willard, Jainwei Zhang, and Keith R. McCrae. A novel pathway for human endothelial cell activation by antiphospholipid/anti- β 2 glycoprotein I antibodies. *Blood*, 119(3):884–893, 2012.
- [211] Madoka Ohya, Chiaki Nishitani, Hitomi Sano, Chieko Yamada, Hiroaki Mitsuzawa, Takeyuki Shimizu, Tsuyoshi Saito, Kelly Smith, Erika Crouch, and Yoshio Kuroki. Human pulmonary surfactant protein D binds the extracellular domains of Toll-like receptors 2 and 4 through the carbohydrate recognition domain by a mechanism different from its binding to phosphatidylinositol and lipopolysaccharide. *Biochemistry*, 45(28):8657–8664, 2006.
- [212] Loïc Guillot, Viviane Balloy, Francis X. McCormack, Douglas T. Golenbock, Michel Chignard, and Mustapha Si-Tahar. Cutting Edge: The Immunostimulatory Activity of the Lung Surfactant Protein-A Involves Toll-Like Receptor 4. *The Journal of Immunology*, 168(12):5989–5992, 2002.
- [213] Shanjana Awasthi, Kevin Brown, Catherine King, Vibhudutta Awasthi, and Rajkumar Bondugula. A toll-like receptor-4-interacting surfactant protein-a-derived peptide suppresses tumor necrosis factor- α release from mouse JAWS II dendritic cells. *Journal of Pharmacology and Experimental Therapeutics*, 336(3):672–681, 2011.

- [214] Chieko Yamada, Hitomi Sano, Takeyuki Shimizu, Hiroaki Mitsuzawa, Chiaki Nishitani, Tetsuo Himi, and Yoshio Kuroki. Surfactant protein A directly interacts with TLR4 and MD-2 and regulates inflammatory cellular response: Importance of supratrimeric oligomerization. *Journal of Biological Chemistry*, 281(31):21771–21780, 2006.
- [215] Morihito Sato, Hitomi Sano, Daisuke Iwaki, Kazumi Kudo, Masanori Konishi, Hiroki Takahashi, Toru Takahashi, Hitoshi Imaizumi, Yasufumi Asai, and Yoshio Kuroki. Direct Binding of Toll-Like Receptor 2 to Zymosan, and Zymosan-Induced NF- κ B Activation and TNF- α Secretion Are Down-Regulated by Lung Collectin Surfactant Protein A. *The Journal of Immunology*, 171(1):417–425, 2003.
- [216] Seiji Murakami, Daisuke Iwaki, Hiroaki Mitsuzawa, Hitomi Sano, Hiroki Takahashi, Dennis R. Voelker, Toyoaki Akino, and Yoshio Kuroki. Surfactant protein A inhibits peptidoglycan-induced tumor necrosis factor- α secretion in U937 cells and alveolar macrophages by direct interaction with toll-like receptor 2. *Journal of Biological Chemistry*, 277(9):6830–6837, 2002.
- [217] H Sano, H Sohma, T Muta, S Nomura, D R Voelker, and Y Kuroki. Pulmonary surfactant protein A modulates the cellular response to smooth and rough lipopolysaccharides by interaction with CD14. *Journal of immunology (Baltimore, Md. : 1950)*, 163(1):387–95, 1999.
- [218] Hitomi Sano, Hirofumi Chiba, Daisuke Iwaki, Hitoshi Sohma, Dennis R. Voelker, and Yoshio Kuroki. Surfactant proteins A and D bind CD14 by different mechanisms. *Journal of Biological Chemistry*, 275(29):22442–22451, 2000.
- [219] Vijay Ramani, Rakesh Madhusoodhanan, Stanley Kosanke, and Shanjana Awasthi. A TLR4-interacting SPA4 peptide inhibits LPS-induced lung inflammation. *Innate Immunity*, 19(6):596–610, 2013.
- [220] Lidan Liu, Chaim Z. Aron, Cullen M. Grable, Adrian Robles, Xiangli Liu, Yuying Liu, Nicole Y. Fatheree, J. Marc Rhoads, and Joseph L. Alcorn. Surfactant protein A reduces TLR4 and inflammatory cytokine mRNA levels in neonatal mouse ileum. *Scientific Reports*, 11(1):1–15, 2021.
- [221] Vicky Sender, Linda Lang, and Cordula Stamme. Surfactant Protein-A Modulates LPS-Induced TLR4 Localization and Signaling via β -Arrestin 2. *PLoS ONE*, 8(3), 2013.

- [222] James M. Devaney, Catherine M. Greene, Clifford C. Taggart, Tomás P. Carroll, Shane J. O'Neill, and Noel G. McElvaney. Neutrophil elastase up-regulates interleukin-8 via toll-like receptor 4. *FEBS Letters*, 544(1-3):129–132, 2003.
- [223] Rym Benabid, Julien Wartelle, Laurette Malleret, Nicolas Guyot, Sophie Gangloff, François Lebargy, and Azzaq Belaaouaj. Neutrophil elastase modulates cytokine expression: Contribution to host defense against pseudomonas aeruginosa-induced pneumonia. *Journal of Biological Chemistry*, 287(42):34883–34894, 2012.
- [224] Antti Hietaranta, Harri Mustonen, Pauli Puolakkainen, Reijo Haapiainen, and Esko Kempainen. Proinflammatory effects of pancreatic elastase are mediated through TLR4 and NF- κ B. *Biochemical and Biophysical Research Communications*, 323(1):192–196, 2004.
- [225] Flavia L. Ribeiro-Gomes, Maria Carolina A. Moniz-de Souza, Magna S. Alexandre-Moreira, Wagner B. Dias, Marcela F. Lopes, Marise P. Nunes, Giuseppe Lungarella, and George A. DosReis. Neutrophils Activate Macrophages for Intracellular Killing of Leishmania major through Recruitment of TLR4 by Neutrophil Elastase . *The Journal of Immunology*, 179(6):3988–3994, 2007.
- [226] Kang Yun Lee, Shu Chuan Ho, Horng Chyuan Lin, Shu Min Lin, Chien Ying Liu, Chien Da Huang, Chun Hua Wang, Kian Fan Chung, and Han Pin Kuo. Neutrophil-derived elastase induces TGF- β 1 secretion in human airway smooth muscle via NF- κ B pathway. *American Journal of Respiratory Cell and Molecular Biology*, 35(4):407–414, 2006.
- [227] Hisanori Domon, Kosuke Nagai, Tomoki Maekawa, Masataka Oda, Daisuke Yonezawa, Wataru Takeda, Takumi Hiyoshi, Hikaru Tamura, Masaya Yamaguchi, Shigetada Kawabata, and Yutaka Terao. Neutrophil elastase subverts the immune response by cleaving toll-like receptors and cytokines in pneumococcal pneumonia. *Frontiers in Immunology*, 9:1–11, 2018.
- [228] Tetsuya Hagio, Katsuya Kishikawa, Kazuhito Kawabata, Sadatomo Tasaka, Satoru Hashimoto, Naoki Hasegawa, and Akitoshi Ishizaka. Inhibition of neutrophil elastase reduces lung injury and bacterial count in hamsters. *Pulmonary Pharmacology and Therapeutics*, 21(6):884–891, 2008.

- [229] Marilia S. Faria, Flavia C. G. Reis, Ricardo L. Azevedo-Pereira, Lesley S. Morrison, Jeremy C. Mottram, and Ana Paula C. A. Lima. Leishmania Inhibitor of Serine Peptidase 2 Prevents TLR4 Activation by Neutrophil Elastase Promoting Parasite Survival in Murine Macrophages . *The Journal of Immunology*, 186(1):411–422, 2011.
- [230] Rui Yamaguchi, Takatoshi Yamamoto, Arisa Sakamoto, Shinji Narahara, Hiroyuki Sugiuchi, and Yasuo Yamaguchi. Neutrophil elastase enhances IL-12p40 production by lipopolysaccharide-stimulated macrophages via transactivation of the PAR-2/EGFR/TLR4 signaling pathway. *Blood Cells, Molecules, and Diseases*, 59:1–7, 2016.
- [231] David A. Bergin, Catherine M. Greene, Erwin E. Sterchi, Cliona Kenna, Patrick Geraghty, Abderazzaq Belaaouaj, Clifford C. Taggart, Shane J. O’Neill, and Noel G. McElvaney. Activation of the epidermal growth factor receptor (EGFR) by a novel metalloprotease pathway. *Journal of Biological Chemistry*, 283(46):31736–31744, 2008.
- [232] Ziad Julier, Mikaël M. Martino, Alexandre De Titta, Laura Jeanbart, and Jeffrey A. Hubbell. The TLR4 agonist fibronectin extra domain a is cryptic, Exposed by elastase-2; Use in a fibrin matrix cancer vaccine. *Scientific Reports*, 5:1–10, 2015.
- [233] Prasad Rallabhandi, Quan M. Nhu, Vladimir Y. Toshchakov, Wenji Piao, Andrei E. Medvedev, Morley D. Hollenberg, Alessio Fasano, and Stefanie N. Vogel. Analysis of proteinase-activated receptor 2 and TLR4 signal transduction: A novel paradigm for receptor cooperativity. *Journal of Biological Chemistry*, 283(36):24314–24325, 2008.
- [234] Sheikh Rayees, Ian Rochford, Jagdish Chandra Joshi, Bhagwati Joshi, Somnath Banerjee, and Dolly Mehta. Macrophage TLR4 and PAR2 Signaling: Role in Regulating Vascular Inflammatory Injury and Repair. *Frontiers in Immunology*, 11(September):1–13, 2020.
- [235] Claus W. Heizmann and Jos A. Cox. New perspectives on s100 proteins: A multi-functional Ca^{2+} -, Zn^{2+} - and Cu^{2+} -binding protein family. *BioMetals*, 11(4):383–397, 1998.
- [236] Susan E. Kelly, David B. Jones, and Stewart Fleming. Calgranulin expression in inflammatory dermatoses. *The Journal of Pathology*, 159(1):17–21, 1989.

- [237] Nancy Hogg, Catherine Allen, and Jonathan Edgeworth. Monoclonal antibody 5.5 reacts with p8,14, a myeloid molecule associated with some vascular endothelium. *European Journal of Immunology*, 19(6):1053–1061, 6 1989.
- [238] Johannes Roth, Thomas Vogl, Clemens Sorg, and Cord Sunderkötter. Phagocyte-specific S100 proteins: A novel group of proinflammatory molecules. *Trends in Immunology*, 24(4):155–158, 2003.
- [239] T. Vogl, J. Roth, C. Sorg, F. Hillenkamp, and K. Strupat. Calcium-induced noncovalently linked tetramers of MRP8 and MRP14 detected by ultraviolet matrix-assisted laser desorption/ionization mass spectrometry. *Journal of the American Society for Mass Spectrometry*, 10(11):1124–1130, 1999.
- [240] Michael J. Hunter and Walter J. Chazin. High level expression and dimer characterization of the S100 EF-hand proteins, migration inhibitory factor-related proteins 8 and 14. *Journal of Biological Chemistry*, 273(20):12427–12435, 1998.
- [241] Thomas Vogl, Nadja Leukert, Katarzyna Barczyk, Kerstin Strupat, and Johannes Roth. Biophysical characterization of S100A8 and S100A9 in the absence and presence of bivalent cations. *Biochimica et Biophysica Acta - Molecular Cell Research*, 1763(11):1298–1306, 2006.
- [242] J. Edgeworth, M. Gorman, R. Bennett, P. Freemont, and N. Hogg. Identification of p8,14 as a highly abundant heterodimeric calcium binding protein complex of myeloid cells. *Journal of Biological Chemistry*, 266(12):7706–7713, 1991.
- [243] Francesca Ometto, Lara Friso, Davide Astorri, Costantino Botsios, Bernd Raffener, Leonardo Punzi, and Andrea Doria. Calprotectin in rheumatic diseases. *Experimental Biology and Medicine*, 242(8):859–873, 2017.
- [244] C van den Bos, J Roth, H G Koch, M Hartmann, and C Sorg. Phosphorylation of MRP14, an S100 protein expressed during monocytic differentiation, modulates Ca(2+)-dependent translocation from cytoplasm to membranes and cytoskeleton. *Journal of Immunology*, 156(3):1247–54, 2 1996.
- [245] Anke Rammes, Johannes Roth, Matthias Goebeler, Martin Klempt, Michael Hartmann, and Clemens Sorg. Myeloid-related protein (MRP) 8 and MRP14, calcium-binding proteins of the S100 family, are secreted by activated monocytes via a novel, tubulin- dependent pathway. *Journal of Biological Chemistry*, 272(14):9496–9502, 1997.

- [246] G. Myklebust and J. T. Gran. A prospective study of 287 patients with polymyalgia rheumatica and temporal arteritis: Clinical and laboratory manifestations at onset of disease and at the time of diagnosis. *British Journal of Rheumatology*, 35(11):1161–1168, 1996.
- [247] Thomas Vogl, Michel Eisenblätter, Tom Völler, Stefanie Zenker, Sven Hermann, Peter Van Lent, Andreas Faust, Christiane Geyer, Beatrix Petersen, Kirsten Roebrock, Michael Schäfers, Christoph Bremer, and Johannes Roth. Alarmin S100A8/S100A9 as a biomarker for molecular imaging of local inflammatory activity. *Nature Communications*, 5(May):1–12, 2014.
- [248] Kwi Young Kang, Jung Won Woo, and Sung Hwan Park. S100A8/A9 as a biomarker for synovial inflammation and joint damage in patients with rheumatoid arthritis. *Korean Journal of Internal Medicine*, 29(1):12–19, 2014.
- [249] P. A. CLOHESSY and B. E. GOLDEN. Calprotectin-Mediated Zinc Chelation as a Biostatic Mechanism in Host Defence. *Scandinavian Journal of Immunology*, 42(5):551–556, 1995.
- [250] Emily M. Zygiel and Elizabeth M. Nolan. Transition Metal Sequestration by the Host-Defense Protein Calprotectin. *Annual Review of Biochemistry*, 87(120):621–643, 2018.
- [251] P. L.E.M. Van Lent, L. Grevers, A. B. Blom, A. Sloetjes, J. S. Mort, T. Vogl, W. Nacken, W. B. Van Den Berg, and J. Roth. Myeloid-related proteins S100A8/S100A9 regulate joint inflammation and cartilage destruction during antigen-induced arthritis. *Annals of the Rheumatic Diseases*, 67(12):1750–1758, 2008.
- [252] Dong Gun Lee, Jung Won Woo, Seung Ki Kwok, Mi La Cho, and Sung Hwan Park. MRP8 promotes Th17 differentiation via upregulation of IL-6 production by fibroblast-like synoviocytes in rheumatoid arthritis. *Experimental and Molecular Medicine*, 45(4):1–9, 2013.
- [253] Matteo Riva, Eva Källberg, Per Björk, Dora Hancz, Thomas Vogl, Johannes Roth, Fredrik Ivars, and Tomas Leanderson. Induction of nuclear factor- κ B responses by the S100A9 protein is Toll-like receptor-4-dependent. *Immunology*, 137(2):172–182, 2012.
- [254] Rebecca L. Rich and David G. Myszka. Grading the commercial optical biosensor literature - Class of 2008: 'The Mighty Binders'. *Journal of Molecular Recognition*, 23(1):1–64, 2010.

- [255] Katharina Brandl, Thomas Glück, Pia Hartmann, Bernd Salzberger, and Werner Falk. A designed TLR4/MD-2 complex to capture LPS. *Journal of Endotoxin Research*, 11(4):197–206, 2005.
- [256] Andrea N Noes, Ran Shi, and Michael J. Harms. Zinc-independent activation of Toll-like receptor 4 by S100A9. *bioArxiv*, 2019.
- [257] Wing S. To and Kim S. Midwood. Plasma and cellular fibronectin: Distinct and independent functions during tissue repair. *Fibrogenesis and Tissue Repair*, 4(1):1–17, 2011.
- [258] Purva Singh, Cara Carraher, and Jean E. Schwarzbauer. Assembly of fibronectin extracellular matrix. *Annual Review of Cell and Developmental Biology*, 26:397–419, 2010.
- [259] Josephine M.J. Stoffels, Chao Zhao, and Wia Baron. Fibronectin in tissue regeneration: Timely disassembly of the scaffold is necessary to complete the build. *Cellular and Molecular Life Sciences*, 70(22):4243–4253, 2013.
- [260] Julie S. Lefebvre, Tania Lévesque, Serge Picard, Guillaume Paré, Annie Gravel, Louis Flamand, and Pierre Borgeat. Extra domain a of fibronectin primes leukotriene biosynthesis and stimulates neutrophil migration through activation of Toll-like receptor 4. *Arthritis and Rheumatism*, 63(6):1527–1533, 2011.
- [261] Srie Prihianti Gondokaryono, Hiroko Ushio, François Niyonsaba, Mutsuko Hara, Hiroshi Takenaka, Sumanasiri T. M. Jayawardana, Shigaku Ikeda, Ko Okumura, and Hideoki Ogawa. The extra domain A of fibronectin stimulates murine mast cells via Toll-like receptor 4. *Journal of Leukocyte Biology*, 82(3):657–665, 2007.
- [262] Haruta Mogami, Patrick W. Keller, Haolin Shi, and R. Ann Word. Effect of thrombin on human amnion mesenchymal cells, mouse fetal membranes, and preterm birth. *Journal of Biological Chemistry*, 289(19):13295–13307, 2014.
- [263] Jascha N. Rybak, Christoph Roesli, Manuela Kaspar, Alessandra Villa, and Dario Neri. The extra-domain A of fibronectin is a vascular marker of solid tumors and metastases. *Cancer Research*, 67(22):10948–10957, 2007.

- [264] Mingzhe Zheng, David M. Jones, Carol Horzempa, Aparna Prasad, and Paula J. McKeown-Longo. The first type III domain of fibronectin is associated with the expression of cytokines within the lung tumor microenvironment. *Journal of Cancer*, 2(1):478–483, 2011.
- [265] Rhiannon Kelsh, Ran You, Carol Horzempa, Mingzhe Zheng, and Paula J. McKeown-Longo. Regulation of the innate immune response by fibronectin: Synergism between the III-1 and EDA domains. *PLoS ONE*, 9(7), 2014.
- [266] Nidhi Sofat, Saralili Dipa Robertson, and Robin Wait. Fibronectin III 13-14 domains induce joint damage via Toll-like receptor 4 activation and synergize with interleukin-1 and tumour necrosis factor. *Journal of Innate Immunity*, 4(1):69–79, 2011.
- [267] Alessandro Malara, Cristian Gruppi, Vittorio Abbonante, Daniele Cattaneo, Luigi De Marco, Margherita Massa, Alessandra Iurlo, Umberto Gianelli, Carlo L. Balduini, Maria E. Tira, Andr s F. Muro, Anil K. Chauhan, Vittorio Rosti, Giovanni Barosi, and Alessandra Balduini. EDA fibronectin–TLR4 axis sustains megakaryocyte expansion and inflammation in bone marrow fibrosis. *Journal of Experimental Medicine*, 216(3):587–604, 2019.
- [268] Zhengyu He, Yuan Gao, Yuxiao Deng, Wen Li, Yongming Chen, Shunpeng Xing, Xianyuan Zhao, Jia Ding, and Xiangrui Wang. Lipopolysaccharide induces lung fibroblast proliferation through toll-like receptor 4 signaling and the phosphoinositide3-kinase-Akt pathway. *PLoS ONE*, 7(4):1–7, 2012.
- [269] Sujeong Jang, Jong-Sung Park, Young-Ho Won, Sook-Jung Yun, and Seong-Jin Kim. The Expression of Toll-Like Receptors (TLRs) in Cultured Human Skin Fibroblast is Modulated by Histamine. *Chonnam Medical Journal*, 48(1):7, 2012.
- [270] Liliana Schaefer and Renato V. Iozzo. Biological functions of the small leucine-rich proteoglycans: From genetics to signal transduction. *Journal of Biological Chemistry*, 283(31):21305–21309, 2008.
- [271] Rosetta Merline, Roland M. Schaefer, and Liliana Schaefer. The matricellular functions of small leucine-rich proteoglycans (SLRPs). *Journal of Cell Communication and Signaling*, 3(3-4):323–335, 2009.
- [272] Shoujun Chen and David E. Birk. The regulatory roles of small leucine-rich proteoglycans in extracellular matrix assembly. *FEBS Journal*, 280(10):2120–2137, 2013.

- [273] Louise Tzung Harn Hsieh, Madalina Viviana Nastase, Jinyang Zeng-Brouwers, Renato V. Iozzo, and Liliana Schaefer. Soluble biglycan as a biomarker of inflammatory renal diseases. *International Journal of Biochemistry and Cell Biology*, 54:223–235, 2014.
- [274] Helena Frey, Nina Schroeder, Tina Manon-Jensen, Renato V. Iozzo, and Liliana Schaefer. Biological interplay between proteoglycans and their innate immune receptors in inflammation. *FEBS Journal*, 280(10):2165–2179, 2013.
- [275] Hans Joachim Anders and Liliana Schaefer. Beyond tissue injury - Damage-associated molecular patterns, toll-like receptors, and inflammasomes also drive regeneration and fibrosis. *Journal of the American Society of Nephrology*, 25(7):1387–1400, 2014.
- [276] Madalina V. Nastase, Marian F. Young, and Liliana Schaefer. Biglycan: A Multivalent Proteoglycan Providing Structure and Signals. *Journal of Histochemistry and Cytochemistry*, 60(12):963–975, 2012.
- [277] Renato V. Iozzo and Liliana Schaefer. Proteoglycans in health and disease: Novel regulatory signaling mechanisms evoked by the small leucine-rich proteoglycans. *FEBS Journal*, 277(19):3864–3875, 2010.
- [278] Liliana Schaefer. Complexity of danger: The diverse nature of damage-associated molecular patterns. *Journal of Biological Chemistry*, 289(51):35237–35245, 2014.
- [279] Kristin Moreth, Rebekka Brodbeck, Andrea Babelova, Norbert Gretz, Tilmann Spieker, Jinyang Zeng-Brouwers, Josef Pfeilschifter, Marian F. Young, Roland M. Schaefer, and Liliana Schaefer. The proteoglycan biglycan regulates expression of the B cell chemoattractant CXCL13 and aggravates murine lupus nephritis. *Journal of Clinical Investigation*, 120(12):4251–4272, 2010.
- [280] Jinyang Zeng-Brouwers, Janet Beckmann, Madalina Viviana Nastase, Renato V. Iozzo, and Liliana Schaefer. De novo expression of circulating biglycan evokes an innate inflammatory tissue response via MyD88/TRIF pathways. *Matrix Biology*, 35:132–142, 2014.
- [281] Tyler Mrozowich, Vanessa Meier-Stephenson, and Trushar R. Patel. Microscale thermophoresis: Warming up to a new biomolecular interaction technique. *Biochemist*, 41(2):8–12, 2019.

- [282] Moran Jerabek-Willemsen, Timon André, Randy Wanner, Heide Marie Roth, Stefan Duhr, Philipp Baaske, and Dennis Breitsprecher. MicroScale Thermophoresis: Interaction analysis and beyond. *Journal of Molecular Structure*, 1077:101–113, 2014.
- [283] Ruth Chiquet-Ehrismann, Eleanor J. Mackie, Carolyn A. Pearson, and Teruyo Sakakura. Tenascin: an extracellular matrix protein involved in tissue interactions during fetal development and oncogenesis. *Cell*, 47(1):131–139, 1986.
- [284] F. S. Jones, S. Hoffman, B. A. Cunningham, and G. M. Edelman. A detailed structural model of cytotactin: Protein homologies, alternative RNA splicing, and binding regions. *Proceedings of the National Academy of Sciences of the United States of America*, 86(6):1905–1909, 1989.
- [285] Sean P. Giblin and Kim S. Midwood. Tenascin-C: Form versus function. *Cell Adhesion and Migration*, 9(1-2):48–82, 2015.
- [286] M. Fluck, V. Tunc-Civelek, and M. Chiquet. Rapid and reciprocal regulation of tenascin-C and tenascin-Y expression by loading of skeletal muscle. *Journal of Cell Science*, 113(20):3583–3591, 2000.
- [287] Ruth Chiquet-Ehrismann, Gertraud Orend, Matthias Chiquet, Richard P. Tucker, and Kim S. Midwood. Tenascins in stem cell niches. *Matrix Biology*, 37:112–123, 2014.
- [288] E Aufderheide, R Chiquet-Ehrismann, and P Ekblom. Epithelial-mesenchymal interactions in the developing kidney lead to expression of tenascin in the mesenchyme. *Journal of Cell Biology*, 105(1):599–608, 7 1987.
- [289] Gertraud Orend and Ruth Chiquet-Ehrismann. Tenascin-C induced signaling in cancer. *Cancer Letters*, 244(2):143–163, 2006.
- [290] Richard P. Tucker and Ruth Chiquet-Ehrismann. The regulation of tenascin expression by tissue microenvironments. *Biochimica et Biophysica Acta - Molecular Cell Research*, 1793(5):888–892, 2009.
- [291] Kim S. Midwood and Gertraud Orend. The role of tenascin-C in tissue injury and tumorigenesis. *Journal of Cell Communication and Signaling*, 3(3-4):287–310, 2009.

- [292] Wentao Huang, Ruth Chiquet-Ehrismann, Gertraud Orend, José V. Moyano, and Angeles Garcia-Pardo. Interference of tenascin-C with syndecan-4 binding to fibronectin blocks cell adhesion and stimulates tumor cell proliferation. *Cancer Research*, 61(23):8586–8594, 2001.
- [293] Kim S. Midwood, Leyla V. Valenick, Henry C. Hsia, and Jean E. Schwarzbauer. Coregulation of Fibronectin Signaling and Matrix Contraction by Tenascin-C and Syndecan-4. *Molecular Biology of the Cell*, 15(12):5670–5677, 12 2004.
- [294] Ruth Chiquet-Ehrismann and Matthias Chiquet. Tenascins: Regulation and putative functions during pathological stress. *Journal of Pathology*, 200(4):488–499, 2003.
- [295] Fui G. Goh, Anna M. Piccinini, Thomas Krausgruber, Irina A. Udalova, and Kim S. Midwood. Transcriptional Regulation of the Endogenous Danger Signal Tenascin-C: A Novel Autocrine Loop in Inflammation. *The Journal of Immunology*, 184(5):2655–2662, 2010.
- [296] Kim S. Midwood, Matthias Chiquet, Richard P. Tucker, and Gertraud Orend. Tenascin-C at a glance. *Journal of Cell Science*, 129(23):4321–4327, 2016.
- [297] Irina A. Udalova, Michaela Ruhmann, Scott J.P. Thomson, and Kim S. Midwood. Expression and Immune Function of Tenascin-C. *Critical Reviews™ in Immunology*, 31(2):115–145, 2011.
- [298] Anna M. Marzeda and Kim S. Midwood. Internal Affairs: Tenascin-C as a Clinically Relevant, Endogenous Driver of Innate Immunity. *Journal of Histochemistry and Cytochemistry*, 66(4):289–304, 2018.
- [299] Kim S. Midwood, Thomas Hussenet, Benoit Langlois, and Gertraud Orend. Advances in tenascin-C biology. *Cellular and Molecular Life Sciences*, 68(19):3175–3199, 2011.
- [300] Vivien C. Yee, Kathleen P. Pratt, Hélène C.F. Côté, Isolde Le Trong, Dominic W. Chung, Earl W. Davie, Ronald E. Stenkamp, and David C. Teller. Crystal structure of a 30 kDa C-terminal fragment from the γ chain of human fibrinogen. *Structure*, 5(1):125–138, 1997.
- [301] V. A. LIGHTNER, C. A. SLEMP, and H. P. ERICKSON. Localization and Quantitation of Hexabrachion (Tenascin) in Skin, Embryonic Brain, Tumors,

and Plasma. *Annals of the New York Academy of Sciences*, 580(1):260–275, 1990.

- [302] Joseph B. Ventimiglia, Carol J. Wikstrand, Lawrence E. Ostrowski, Mario A. Bourdon, Virginia A. Lightner, and Darell D. Bigner. Tenascin expression in human glioma cell lines and normal tissues. *Journal of Neuroimmunology*, 36(1):41–55, 1992.
- [303] Lorena Zuliani-Alvarez, Anna M Marzeda, Claire Deligne, Anja Schwenzer, Fiona E. McCann, Brian D Marsden, Anna M Piccinini, and Kim S Midwood. Mapping tenascin-C interaction with toll-like receptor 4 reveals a new subset of endogenous inflammatory triggers. *Nature Communications*, 8(1):1595, 12 2017.
- [304] Swati Bhattacharyya, Wenxia Wang, Wenyi Qin, Kui Cheng, Sara Coulup, Sherry Chavez, Shuangshang Jiang, Kirtee Raparia, Lucia Maria V. De Almeida, Christian Stehlik, Zenshiro Tamaki, Hang Yin, and John Varga. TLR4-dependent fibroblast activation drives persistent organ fibrosis in skin and lung. *JCI Insight*, 3(13), 2018.
- [305] Susan R. Aungier, Alison J. Cartwright, Anja Schwenzer, Jennifer L. Marshall, Michael R. Dyson, Peter Slavny, Kothai Parthiban, Aneesh Karatt-Vellatt, Ilfita Sahbudin, Eric Culbert, Patrick Hextall, Felix I.L. Clanchy, Richard Williams, Brian D. Marsden, Karim Raza, Andrew Filer, Christopher Dominic Buckley, John McCafferty, and Kim S. Midwood. Targeting early changes in the synovial microenvironment: A new class of immunomodulatory therapy? *Annals of the Rheumatic Diseases*, 78(2):186–191, 2019.
- [306] Russell F. Doolittle, Kyle McNamara, and Kevin Lin. Correlating structure and function during the evolution of fibrinogen-related domains. *Protein Science*, 21(12):1808–1823, 2012.
- [307] Fredrik Ronquist, Maxim Teslenko, Paul Van Der Mark, Daniel L. Ayres, Aaron Darling, Sebastian Höhna, Bret Larget, Liang Liu, Marc A. Suchard, and John P. Huelsenbeck. Mrbayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, 61(3):539–542, 2012.
- [308] Ivica Letunic and Peer Bork. Interactive Tree of Life (iTOL) v4: Recent updates and new developments. *Nucleic Acids Research*, 47(W1):256–259, 2019.

- [309] Carmine Carbone, Geny Piro, Valeria Merz, Francesca Simionato, Raffaella Santoro, Camilla Zecchetto, Giampaolo Tortora, and Davide Melisi. Angiopoietin-like proteins in angiogenesis, inflammation and cancer. *International Journal of Molecular Sciences*, 19(2):1–22, 2018.
- [310] Matthias Blüher. Distinct roles of angiopoietin-like 4 in the regulation of central and peripheral lipid metabolism? *Molecular Metabolism*, 4(2):79–80, 2015.
- [311] Clémence Chomel, Aurelie Cazes, Clement Faye, Marine Bignon, Elisa Gomez, Corinne Ardidie-Robouant, Alain Barret, Sylvie Ricard-Blum, Laurent Muller, Stephane Germain, and Catherine Monnot. Interaction of the coiled-coil domain with glycosaminoglycans protects angiopoietin-like 4 from proteolysis and regulates its antiangiogenic activity. *The FASEB Journal*, 23(3):940–949, 2009.
- [312] Ying Hua Yang, Yu Wang, Karen S.L. Lam, Ming Hon Yau, Kenneth K.Y. Cheng, Jialiang Zhang, Weidong Zhu, Donghai Wu, and Aimin Xu. Suppression of the Raf/MEK/ERK Signaling cascade and inhibition of angiogenesis by the carboxyl terminus of angiopoietin-like protein 4. *Arteriosclerosis, Thrombosis, and Vascular Biology*, 28(5):835–840, 2008.
- [313] Mitsuru Ono, Tetsuya Shimizugawa, Mitsuru Shimamura, Kenichi Yoshida, Chisa Noji-Sakikawa, Yosuke Ando, Ryuta Koishi, and Hidehiko Furukawa. Protein Region Important for Regulation of Lipid Metabolism in Angiopoietin-like 3 (ANGPTL3): ANGPTL3 is cleaved and activated in vivo. *Journal of Biological Chemistry*, 278(43):41804–41809, 2003.
- [314] Gaetano Santulli. Angiopoietin-like proteins: A comprehensive look. *Frontiers in Endocrinology*, 5(JAN):5–10, 2014.
- [315] Nathalie Thorin-Trescases and Eric Thorin. Angiopoietin-like-2: A multifaceted protein with physiological and pathophysiological properties. *Studies in Second Language Acquisition*, 16(May):1–17, 2014.
- [316] Tsuyoshi Kadomatsu, Motoyoshi Endo, Keishi Miyata, and Yuichi Oike. Diverse roles of ANGPTL2 in physiology and pathophysiology. *Trends in Endocrinology and Metabolism*, 25(5):245–254, 2014.

- [317] Tatsuya Okada, Hiroto Tsukano, Motoyoshi Endo, Mitsuhsa Tabata, Keishi Miyata, Tsuyoshi Kadomatsu, Kazuya Miyashita, Kei Semba, Eiichi Nakamura, Michishi Tsukano, Hiroshi Mizuta, and Yuichi Oike. Synovial cell-derived angiopoietin-like protein 2 contributes to Synovial chronic inflammation in rheumatoid arthritis. *American Journal of Pathology*, 176(5):2309–2319, 2010.
- [318] David Padua, Xiang H.F. Zhang, Qiongqing Wang, Cristina Nadal, William L. Gerald, Roger R. Gomis, and Joan Massagué. TGF β Primes Breast Tumors for Lung Metastasis Seeding through Angiopoietin-like 4. *Cell*, 133(1):66–77, 2008.
- [319] Ming Jie Tan, Ziqiang Teo, Ming Keat Sng, Pengcheng Zhu, and Nguan Soon Tan. Emerging roles of angiopoietin-like 4 in human cancer. *Molecular Cancer Research*, 10(6):677–688, 2012.
- [320] Yueqing Zhang, Xiaobo Hu, Ruiyang Tian, Wangui Wei, Wei Hu, Xia Chen, Wei Han, Huayou Chen, and Yi Gong. Angiopoietin-related growth factor (AGF) supports adhesion, spreading, and migration of keratinocytes, fibroblasts, and endothelial cells through interaction with RGD-binding integrins. *Biochemical and Biophysical Research Communications*, 347(1):100–108, 2006.
- [321] Hiroki Tanigawa, Keishi Miyata, Zhe Tian, Jun Aoi, Tsuyoshi Kadomatsu, Satoshi Fukushima, Aki Ogata, Naoki Takeda, Jiabin Zhao, Shunshun Zhu, Kazutoyo Terada, Motoyoshi Endo, Jun Morinaga, Taichi Sugizaki, Michio Sato, Masaki Suimye Morioka, Ichiro Manabe, Youichi Mashimo, Akira Hata, Yoshitaka Taketomi, Kei Yamamoto, Makoto Murakami, Kimi Araki, Masatoshi Jinnin, Hironobu Ihn, and Yuichi Oike. Upregulation of ANGPTL6 in mouse keratinocytes enhances susceptibility to psoriasis. *Scientific Reports*, 6(October):1–12, 2016.
- [322] Junke Zheng, Masato Umikawa, Changhao Cui, Jiyuan Li, Xiaoli Chen, Chaozheng Zhang, Hoangdinh Hyunh, Xunlei Kang, Robert Silvan, Xuan Wan, Jingxiao Ye, Alberto Puig Cantó, Shu Hsia Chen, Huan You Wang, E. Sally Ward, and Cheng Cheng Zhang. Inhibitory receptors bind ANGPTLs and support blood stem cells and leukaemia development. *Nature*, 485(7400):656–660, 2012.
- [323] Tai Hato, Mitsuhsa Tabata, and Yuichi Oike. The Role of Angiopoietin-Like Proteins in Angiogenesis and Metabolism. *Trends in Cardiovascular Medicine*, 18(1):6–14, 2008.

- [324] Shozo Ide, Yuji Toiyama, Tadanobu Shimura, Mikio Kawamura, Hiromi Yasuda, Susumu Saigusa, Masaki Ohi, Koji Tanaka, Yasuhiko Mohri, and Masato Kusunoki. Angiopoietin-Like Protein 2 Acts as a Novel Biomarker for Diagnosis and Prognosis in Patients with Esophageal Cancer. *Annals of Surgical Oncology*, 22(8):2585–2592, 2015.
- [325] Carmine Carbone, Tania Moccia, Cihui Zhu, Genni Paradiso, Alfredo Budillon, Paul J. Chiao, James L. Abbruzzese, and Davide Melisi. Anti-VEGF treatment-resistant pancreatic cancers secrete proinflammatory factors that contribute to malignant progression by inducing an EMT cell phenotype. *Clinical Cancer Research*, 17(17):5822–5832, 2011.
- [326] Motoyoshi Endo, Masahiro Nakano, Tsuyoshi Kadomatsu, Shigetomo Fukuhara, Hiroaki Kuroda, Shuji Mikami, Tai Hato, Jun Aoi, Haruki Horiguchi, Keishi Miyata, Haruki Odagiri, Tetsuro Masuda, Masahiko Harada, Hirotoishi Horio, Tsunekazu Hishima, Hiroaki Nomori, Takaaki Ito, Yutaka Yamamoto, Takashi Minami, Seiji Okada, Takashi Takahashi, Naoki Mochizuki, Hirotaka Iwase, and Yuichi Oike. Tumor cell-derived angiopoietin-like protein ANGPTL2 is a critical driver of metastasis. *Cancer Research*, 72(7):1784–1794, 2012.
- [327] Lucy Stols, Minyi Gu, Lynda Dieckman, Rosemarie Raffen, Frank R. Collart, and Mark I. Donnelly. A new vector for high-throughput, ligation-independent cloning encoding a tobacco etch virus protease cleavage site. *Protein Expression and Purification*, 25(1):8–15, 2002.
- [328] Pavel Savitsky, James Bray, Christopher D.O. Cooper, Brian D. Marsden, Pravin Mahajan, Nicola A. Burgess-Brown, and Opher Gileadi. High-throughput production of human proteins for crystallization: The SGC experience. *Journal of Structural Biology*, 172(1):3–13, 2010.
- [329] Darren J. Korbie and John S. Mattick. Touchdown PCR for increased specificity and sensitivity in PCR amplification. *Nature Protocols*, 3(9):1452–1456, 2008.
- [330] Rebecca Page, Kin Moy, Eric C. Sims, Jeffrey Velasquez, Brendan McManus, Carina Grittini, Thomas L. Clayton, and Raymond C. Stevens. Scalable high-throughput micro-expression device for recombinant proteins. *BioTechniques*, 37(3):364–370, 2004.

- [331] Jia Tsing Ng, Carien Dekker, Markus Kroemer, Michael Osborne, and Frank Von Delft. Using textons to rank crystallization droplets by the likely presence of crystals. *Acta Crystallographica Section D: Biological Crystallography*, 70(10):2702–2718, 2014.
- [332] Fabrice Gorrec. The MORPHEUS protein crystallization screen. *Journal of Applied Crystallography*, 42(6):1035–1042, 2009.
- [333] Clemens Grimm, Ashwin Chari, Klaus Reuter, and Utz Fischer. A crystallization screen based on alternative polymeric precipitants. *Acta Crystallographica Section D: Biological Crystallography*, 66(6):685–697, 2010.
- [334] Janet Newman, David Egan, Thomas S. Walter, Ran Meged, Ian Berry, Marouane Ben Jelloul, Joel L. Sussman, David I. Stuart, and Anastassis Perrakis. Towards rationalization of crystallization screening for small- To medium-sized academic laboratories: The PACT/JCSG+ strategy. *Acta Crystallographica Section D: Biological Crystallography*, 61(10):1426–1431, 2005.
- [335] Fabrice Gorrec. Protein crystallization screens developed at the MRC Laboratory of Molecular Biology. *Drug Discovery Today*, 21(5):819–825, 2016.
- [336] Apirat Chaikuad, Stefan Knapp, and Frank Von Delft. Defined PEG smears as an alternative approach to enhance the search for crystallization conditions and crystal-quality improvement in reduced screens. *Acta Crystallographica Section D: Biological Crystallography*, 71:1627–1639, 2015.
- [337] G. Winter. Xia2: An expert system for macromolecular crystallography data reduction. *Journal of Applied Crystallography*, 43(1):186–190, 2010.
- [338] Paul D. Adams, Pavel V. Afonine, Gábor Bunkóczi, Vincent B. Chen, Ian W. Davis, Nathaniel Echols, Jeffrey J. Headd, Li Wei Hung, Gary J. Kapral, Ralf W. Grosse-Kunstleve, Airlie J. McCoy, Nigel W. Moriarty, Robert Oeffner, Randy J. Read, David C. Richardson, Jane S. Richardson, Thomas C. Terwilliger, and Peter H. Zwart. PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D: Biological Crystallography*, 66(2):213–221, 2010.
- [339] Airlie J. McCoy, Ralf W. Grosse-Kunstleve, Paul D. Adams, Martyn D. Winn, Laurent C. Storoni, and Randy J. Read. Phaser crystallographic software. *Journal of Applied Crystallography*, 40(4):658–674, 2007.

- [340] Pavel V. Afonine, Ralf W. Grosse-Kunstleve, Nathaniel Echols, Jeffrey J. Headd, Nigel W. Moriarty, Marat Mustyakimov, Thomas C. Terwilliger, Alexandre Urzhumtsev, Peter H. Zwart, and Paul D. Adams. Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallographica Section D: Biological Crystallography*, 68(4):352–367, 2012.
- [341] P. Emsley, B. Lohkamp, W. G. Scott, and K. Cowtan. Features and development of Coot. *Acta Crystallographica Section D: Biological Crystallography*, 66(4):486–501, 2010.
- [342] Vincent B. Chen, W. Bryan Arendall, Jeffrey J. Headd, Daniel A. Keedy, Robert M. Immormino, Gary J. Kapral, Laura W. Murray, Jane S. Richardson, and David C. Richardson. MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D: Biological Crystallography*, 66(1):12–21, 2010.
- [343] Swanand Gore, Eduardo Sanz García, Pieter M.S. Hendrickx, Aleksandras Gutmanas, John D. Westbrook, Huanwang Yang, Zukang Feng, Kumaran Baskaran, John M. Berrisford, Brian P. Hudson, Yasuyo Ikegawa, Naohiro Kobayashi, Catherine L. Lawson, Steve Mading, Lora Mak, Abhik Mukhopadhyay, Thomas J. Oldfield, Ardan Patwardhan, Ezra Peisach, Gaurav Sahni, Monica R. Sekharan, Sanchayita Sen, Chenghua Shao, Oliver S. Smart, Eldon L. Ulrich, Reiko Yamashita, Martha Quesada, Jasmine Y. Young, Haruki Nakamura, John L. Markley, Helen M. Berman, Stephen K. Burley, Sameer Velankar, and Gerard J. Kleywegt. Validation of Structures in the Protein Data Bank. *Structure*, 25(12):1916–1927, 2017.
- [344] Nathan David Wright, Patrick Collins, Romain Talon, Elliot Nelson, Lizbé Koekemoer, Mingda Ye, Radosław Nowak, Joseph Newman, Jia Tsing Ng, Nick Mitrovich, Helton Wiggers, and Frank von Delft. The Low-Cost, Semi-Automated Shifter Microscope Stage Transforms Speed and Robustness of Manual Protein Crystal Harvesting. *bioRxiv*, 2019.
- [345] Tobias Krojer, Romain Talon, Nicholas Pearce, Patrick Collins, Alice Douangamath, Jose Brandao-Neto, Alexandre Dias, Brian Marsden, and Frank Von Delft. The XChemExplorer graphical workflow tool for routine or large-scale protein-ligand structure determination. *Acta Crystallographica Section D: Structural Biology*, 73:267–278, 2017.

- [346] Marcin Wojdyr, Ronan Keegan, Graeme Winter, and Alun Ashton. DIMPLE - a pipeline for the rapid generation of difference maps from protein crystals with putatively bound ligands . *Acta Crystallographica Section A Foundations of Crystallography*, 69(a1):s299–s299, 2013.
- [347] Nicholas M. Pearce, Tobias Krojer, Anthony R. Bradley, Patrick Collins, Radosław P. Nowak, Romain Talon, Brian D. Marsden, Sebastian Kelm, Jiye Shi, Charlotte M. Deane, and Frank Von Delft. A multi-crystal method for extracting obscured crystallographic states from conventionally uninterpretable electron density. *Nature Communications*, 8:24–29, 2017.
- [348] Garib N. Murshudov, Pavol Skubák, Andrey A. Lebedev, Navraj S. Pannu, Roberto A. Steiner, Robert A. Nicholls, Martyn D. Winn, Fei Long, and Alexei A. Vagin. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D: Biological Crystallography*, 67(4):355–367, 2011.
- [349] Eric F. Pettersen, Thomas D. Goddard, Conrad C. Huang, Gregory S. Couch, Daniel M. Greenblatt, Elaine C. Meng, and Thomas E. Ferrin. UCSF Chimera - A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13):1605–1612, 2004.
- [350] Frank H. Niesen, Helena Berglund, and Masoud Vedadi. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nature Protocols*, 2(9):2212–2221, 2007.
- [351] Stephane Boivin, Sandra Kozak, and Rob Meijers. Optimization of protein purification and characterization using Thermofluor screens. *Protein Expression and Purification*, 91(2):192–206, 2013.
- [352] Ashwin Chari, David Haselbach, Jan Martin Kirves, Juergen Ohmer, Elham Paknia, Niels Fischer, Oleg Ganichkin, Vanessa Müller, Jeremiah J Frye, Georg Petzold, Marc Jarvis, Michael Tietzel, Clemens Grimm, Jan Michael Peters, Brenda A. Schulman, Kai Tittmann, Jürgen Markl, Utz Fischer, and Holger Stark. ProteoPlex: Stability optimization of macromolecular complexes by sparse-matrix screening of chemical space. *Nature Methods*, 12(9):859–865, 2015.

- [353] Rod Chalk, Georgina Berridge, Leela Shrestha, Claire Strain-Damerell, Pravin Mahajan, Wyatt Yue, Opher Gileadi, and Nicola Burgess-Brown. High-Throughput Mass Spectrometry Applied to Structural Genomics. *Chromatography*, 1(4):159–175, 2014.
- [354] Daniel N Perkins, Darryl J C Pappin, David M Creasy, and John S Cottrell. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis*, 20:3551–3567, 1999.
- [355] A. I. Derman and J. Beckwith. Escherichia coli alkaline phosphatase fails to acquire disulfide bonds when retained in the cytoplasm. *Journal of Bacteriology*, 173(23):7719–7722, 1991.
- [356] Anupam Singh, Vaibhav Upadhyay, Arun Kumar Upadhyay, Surinder Mohan Singh, and Amulya Kumar Panda. Protein recovery from inclusion bodies of Escherichia coli using mild solubilization process. *Microbial Cell Factories*, 14(1):1–10, 2015.
- [357] Sofia Costa, André Almeida, António Castro, and Lucília Domingues. Fusion tags for protein solubility, purification, and immunogenicity in Escherichia coli: The novel Fh8 system. *Frontiers in Microbiology*, 5(FEB):1–20, 2014.
- [358] F. J.M. Mergulhão, D. K. Summers, and G. A. Monteiro. Recombinant protein secretion in Escherichia coli. *Biotechnology Advances*, 23(3):177–202, 2005.
- [359] Katleen Denoncin and Jean François Collet. Disulfide bond formation in the bacterial periplasm: Major achievements and challenges ahead. *Antioxidants and Redox Signaling*, 19(1):63–71, 2013.
- [360] Hervé Nozach, Carole Fruchart-Gaillard, François Fenaille, Fabrice Beau, Oscar Henrique Pereira Ramos, Badreddine Douzi, Natalie J. Saez, Mireille Moutiez, Denis Servent, Muriel Gondry, Robert Thaï, Philippe Cuniasse, Renaud Vincentelli, and Vincent Dive. High throughput screening identifies disulfide isomerase DsbC as a very efficient partner for recombinant expression of small disulfide-rich proteins in E. coli. *Microbial Cell Factories*, 12(1):1–16, 2013.
- [361] Edward R. LaVallie, Elizabeth A. DiBlasio, Sharlotte Kovacic, Kathleen L. Grant, Paul F. Schendel, and John M. McCoy. A Thioredoxin Gene Fusion Expression System That Circumvents Inclusion Body Formation in the E. coli Cytoplasm. *Nature Biotechnology*, 11(2):187–193, 2 1993.

- [362] John M Walker. *SUMO Protocols*, volume 497 of *Methods in Molecular Biology*. Humana Press, Totowa, NJ, 2009.
- [363] Jeffrey G Marblestone, Suzanne C Edavettal, Yiting Lim, Peter Lim, Xun Zuo, and Tauseef R Butt. Comparison of SUMO fusion technology with traditional gene fusion systems: enhanced expression and solubility with SUMO. *Protein science : a publication of the Protein Society*, 15(1):182–9, 1 2006.
- [364] Rachel B Kapust and David S Waugh. Escherichia coli maltose-binding protein is effective in promoting solubility. *Protein Science*, (8):1668–1674, 1999.
- [365] Sreedevi Nallamsetty and David S. Waugh. Mutations that alter the equilibrium between open and closed conformations of Escherichia coli maltose-binding protein impede its ability to enhance the solubility of passenger proteins. *Biochemical and Biophysical Research Communications*, 364(3):639–644, 2007.
- [366] Pierre Douette, Rachel Navet, Pascal Gerkens, Moreno Galleni, Daniel Lévy, and Francis E. Sluse. Escherichia coli fusion carrier proteins act as solubilizing agents for recombinant uncoupling protein 1 through interactions with GroEL. *Biochemical and Biophysical Research Communications*, 333(3):686–693, 2005.
- [367] Paul H. Bessette, Fredrik Åslund, Jon Beckwith, and George Georgiou. Efficient folding of proteins with multiple disulfide bonds in the Escherichia coli cytoplasm. *Proceedings of the National Academy of Sciences of the United States of America*, 96(24):13703–13708, 1999.
- [368] D. Ritz, J. Lim, C. M. Reynolds, L. B. Poole, and J. Beckwith. Conversion of a peroxiredoxin into a disulfide reductase by a triplet repeat expansion. *Science*, 294(5540):158–160, 2001.
- [369] Yuji Yamamoto, Dani Ritz, Anne Gaëlle Planson, Thomas J. Jönsson, Melinda J. Faulkner, Dana Boyd, Jon Beckwith, and Leslie B. Poole. Mutant AhpC Peroxiredoxins Suppress Thiol-Disulfide Redox Deficiencies and Acquire Deglutathionylating Activity. *Molecular Cell*, 29(1):36–45, 2008.
- [370] Julie Lobstein, Charlie A. Emrich, Chris Jeans, Melinda Faulkner, Paul Riggs, and Mehmet Berkmen. SHuffle, a novel Escherichia coli protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm. *Microbial Cell Factories*, 11(1):1, 2012.

- [371] Raphael Levy, Robert Weiss, Gang Chen, Brent L. Iverson, and George Georgiou. Production of correctly folded fab antibody fragment in the cytoplasm of *Escherichia coli* *trxB gor* mutants via the coexpression of molecular chaperones. *Protein Expression and Purification*, 23(2):338–347, 2001.
- [372] Feras Hatahet, Van Dat Nguyen, Kirsi E.H. Salo, and Lloyd W. Ruddock. Disruption of reducing pathways is not essential for efficient disulfide bond formation in the cytoplasm of *E. coli*. *Microbial Cell Factories*, 9:67, 2010.
- [373] Anna Gaciarz, Narendar Kumar Khatri, M. Lourdes Velez-Suberbie, Mirva J. Saaranen, Yuko Uchida, Eli Keshavarz-Moore, and Lloyd W. Ruddock. Efficient soluble expression of disulfide bonded proteins in the cytoplasm of *Escherichia coli* in fed-batch fermentations on chemically defined minimal media. *Microbial Cell Factories*, 16(1):1–12, 2017.
- [374] Van D. Nguyen, Feras Hatahet, Kirsi E.H. Salo, Eveliina Enlund, Chi Zhang, and Lloyd W. Ruddock. Pre-expression of a sulfhydryl oxidase significantly increases the yields of eukaryotic disulfide bond containing proteins expressed in the cytoplasm of *E.coli*. *Microbial Cell Factories*, 10:1–13, 2011.
- [375] Iris H. Walker, Pei Chung Hsieh, and Paul D. Riggs. Mutations in maltose-binding protein that alter affinity and solubility properties. *Applied Microbiology and Biotechnology*, 88(1):187–197, 2010.
- [376] Victor Martin Bolanos-Garcia and Owen Richard Davies. Structural analysis and classification of native proteins from *E. coli* commonly co-purified by immobilised metal affinity chromatography. *Biochimica et Biophysica Acta - General Subjects*, 1760(9):1304–1313, 2006.
- [377] Aneika C. Leney and Albert J.R. Heck. Native Mass Spectrometry: What is in the Name? *Journal of the American Society for Mass Spectrometry*, 28(1):5–13, 2017.
- [378] Alexander McPherson and Jose A. Gavira. Introduction to protein crystallization. *Acta Crystallographica Section F: Structural Biology Communications*, 70(1):2–20, 2014.
- [379] L. Banci, I. Bertini, S. Cusack, R. N. De Jong, U. Heinemann, E. Y. Jones, K. Maskos, F. Kozielski, R. Owens, A. Perrakis, C. Siebold, I. Silman, A. Messerschmidt, A. Poterszman, G. Schneider, T. Sixma, J. L. Sussman, G. Stewart-Jones, J. C. Thierry, and Dino Moras. First steps towards effective

- methods in exploiting high-throughput technologies for the determination of human protein structures of high biomedical value. *Acta Crystallographica Section D: Biological Crystallography*, 62(10):1208–1217, 2006.
- [380] Wijnand T.M. Mooij, Eirini Mitsiki, and Anastassis Perrakis. ProteinCCD: Enabling the design of protein truncation constructs for expression and crystallization experiments. *Nucleic Acids Research*, 37:402–405, 2009.
 - [381] Krzysztof Ginalski. Comparative modeling for protein structure prediction. *Current Opinion in Structural Biology*, 16(2):172–177, 2006.
 - [382] Rune Linding, Lars Juhl Jensen, Francesca Diella, Peer Bork, Toby J. Gibson, and Robert B. Russell. Protein disorder prediction: Implications for structural proteomics. *Structure*, 11(11):1453–1459, 2003.
 - [383] Daniel W.A. Buchan and David T. Jones. The PSIPRED Protein Analysis Workbench: 20 years on. *Nucleic acids research*, 47(W1):W402–W407, 2019.
 - [384] Cédric Notredame, Desmond G. Higgins, and Jaap Heringa. T-coffee: A novel method for fast and accurate multiple sequence alignment. *Journal of Molecular Biology*, 302(1):205–217, 2000.
 - [385] Ruben Abagyan, Dmitry Kuznetsov, and Maxim Totrov. ICM - New Method for Protein Modeling and Design: Applications to Docking and Structure Prediction from. *Journal of Computational Chemistry*, 15(5):488–506, 1994.
 - [386] Thomas G.M. Schmidt, Lilia Batz, Lidia Bonet, Uwe Carl, Gerd Holzapfel, Klaus Kiem, Kamila Matulewicz, Dennis Niermeier, Isabel Schuchardt, and Kristian Stanar. Development of the Twin-Strep-tag® and its application for purification of recombinant proteins from cell culture supernatants. *Protein Expression and Purification*, 92(1):54–61, 2013.
 - [387] Xavier Robert and Patrice Gouet. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Research*, 42(W1):320–324, 2014.
 - [388] Ekaterina Biterova, Mariam Esmaeeli, Heli I Alanen, Mirva Saaranen, and Lloyd W Ruddock. Structures of Angptl3 and Angptl4, modulators of triglyceride levels and coronary artery disease. *Scientific Reports*, pages 1–12, 2018.

- [389] Virginie Garlatti, Nicolas Belloy, Lydie Martin, Monique Lacroix, Misao Matsushita, Yuichi Endo, Teizo Fujita, Juan Carlos Fontecilla-Camps, Gérard J Arlaud, Nicole M Thielens, and Christine Gaboriaud. Structural insights into the innate immune recognition specificities of L- and H-ficolins. *The EMBO journal*, 26(2):623–33, 2007.
- [390] Patrick M Collins, Alice Douangamath, Romain Talon, Alexandre Dias, Jose Brandao-Neto, Tobias Krojer, and Frank von Delft. Achieving a Good Crystal System for Crystallographic X-Ray Fragment Screening. *Methods in enzymology*, 610(September):251–264, 2018.
- [391] Disha Patel, Joseph D Bauman, and Eddy Arnold. Advantages of crystallographic fragment screening: Functional and mechanistic insights from a powerful platform for efficient drug discovery. *Progress in Biophysics and Molecular Biology*, 116(2-3):92–100, 11 2014.
- [392] Marc O’Reilly, Anne Cleasby, Thomas G. Davies, Richard J. Hall, R. Frederick Ludlow, Christopher W. Murray, Dominic Tisi, and Harren Jhoti. Crystallographic screening using ultra-low-molecular-weight ligands to guide drug design. *Drug Discovery Today*, 24(5):1081–1086, 5 2019.
- [393] Daniel Wood, José Daniel Lopez-Fernandez, Leanne Knight, Islam Al-Khawaldeh, Conghao Gai, Shengyin Lin, Mathew P. Martin, Duncan Miller, Celine Cano, Jane Endicott, Ian Hardcastle, Martin Noble, and Michael J. Waring. FragLites - minimal, halogenated fragments displaying pharmacophore doublets. An efficient approach to druggability assessment and hit generation. *Journal of Medicinal Chemistry*, 62:3741–3752, 2019.
- [394] Paul A Mcewan, Patrick Collins, Alex Dias, Michael Mazanetz, Jose Brandao-neto, Alice Douangamath, Jia Ng, Nathan Wright, Tobias Krojer, Romain Talon, John Barker, and Frank Von Delft. Industrial Use of Medium-Throughput Fragment Screening by Crystal Soaking through Highly Streamlined Sample Preparation at Diamond. In *Fragments 2015: Fifth RSC-BMCS Fragment-based Drug Discovery Meeting*, Cambridge, UK, 0.
- [395] John Liebeschuetz, Jana Hennemann, Tjelvar Olsson, and Colin R. Groom. The good, the bad and the twisted: A survey of ligand geometry in protein crystal structures. *Journal of Computer-Aided Molecular Design*, 26(2):169–183, 2012.

- [396] Masoud Vedadi, Frank H. Niesen, Abdellah Allali-Hassani, Oleg Y. Fedorov, Patrick J. Finerty, Gregory A. Wasney, Ron Yeung, Cheryl Arrowsmith, Linda J. Ball, Helena Berglund, Raymond Hui, Brian D. Marsden, Pär Nordlund, Michael Sundstrom, Johan Weigelt, and Aled M. Edwards. Chemical screening methods to identify ligands that promote protein stability, protein crystallization, and structure determination. *Proceedings of the National Academy of Sciences of the United States of America*, 103(43):15835–15840, 2006.
- [397] Eric F. Pettersen, Thomas D. Goddard, Conrad C. Huang, Elaine C. Meng, Gregory S. Couch, Tristan I. Croll, John H. Morris, and Thomas E. Ferrin. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Science*, 30(1):70–82, 2021.
- [398] Lorena Zuliani-Alvarez and Kim S. Midwood. Fibrinogen-Related Proteins in Tissue Repair: How a Unique Domain with a Common Structure Controls Diverse Aspects of Wound Healing. *Advances in Wound Care*, 4(5):273–285, 2015.
- [399] Kenji Yokoyama, Harold P. Erickson, Yasuo Ikeda, and Yoshikazu Takada. Identification of amino acid sequences in fibrinogen γ -chain and tenascin C C-terminal domains critical for binding to integrin $\alpha(v)\beta 3$. *Journal of Biological Chemistry*, 275(22):16891–16898, 2000.
- [400] William A. Barton, Dorothea Tzvetkova-Robev, Edward P. Miranda, Momchil V. Kolev, Kanagalaghatta R. Rajashankar, Juha P. Himanen, and Dimitar B. Nikolov. Crystal structures of the Tie2 receptor ectodomain and the angiopoietin-2-Tie2 complex. *Nature Structural and Molecular Biology*, 13(6):524–532, 2006.
- [401] Glen Spraggon, Stephen J. Everse, and Russell F. Doolittle. Crystal structures of fragment D from human fibrinogen and its crosslinked counterpart from fibrin. *Nature*, 389(6650):455–462, 1997.
- [402] Virginie Garlatti, Lydie Martin, Evelyne Gout, Jean Baptiste Reiser, Teizo Fujita, Gérard J. Arlaud, Nicole M. Thielens, and Christine Gaboriaud. Structural basis for innate immune sensing by M-ficolin and its control by a pH-dependent conformational switch. *Journal of Biological Chemistry*, 282(49):35814–35820, 2007.

- [403] Shmuel Hurwitz. Homeostatic control of plasma calcium concentration. *Critical Reviews in Biochemistry and Molecular Biology*, 31(1):41–100, 1996.
- [404] Gerda E. Breitwieser. Extracellular calcium as an integrator of tissue function. *International Journal of Biochemistry and Cell Biology*, 40(8):1467–1480, 2008.
- [405] G. N. Tzimas, M. Afshar, A. Emadali, E. Chevet, H. Vali, and P. P. Metrakos. Correlation of cell necrosis and tissue calcification with ischemia/reperfusion injury after liver transplantation. *Transplantation Proceedings*, 36(6):1766–1768, 2004.
- [406] Manuela Rossol, Matthias Pierer, Nora Raulien, Dagmar Quandt, Undine Meusch, Kathrin Rothe, Kristin Schubert, Torsten Schöneberg, Michael Schaefer, Ute Krügel, Sanela Smajilovic, Hans Bräuner-Osborne, Christoph Baerwald, and Ulf Wagner. Extracellular Ca^{2+} is a danger signal activating the NLRP3 inflammasome through G protein-coupled calcium sensing receptors. *Nature Communications*, 3, 2012.
- [407] V A Luckow, S C Lee, G F Barry, and P O Olins. Efficient generation of infectious recombinant baculoviruses by site-specific transposon-mediated insertion of foreign genes into a baculovirus genome propagated in *Escherichia coli*. *Journal of virology*, 67(8):4566–79, 1993.
- [408] Abhiram Dukkupati, Hyun Ho Park, Deepa Waghay, Suzanne Fischer, and K. Christopher Garcia. BacMam system for high-level expression of recombinant soluble and membrane glycoproteins for structural studies. *Protein Expression and Purification*, 62(2):160–170, 2008.
- [409] U. Ohto, K. Fukase, K. Miyake, and T. Shimizu. Structural basis of species-specific endotoxin sensing by innate immune receptor TLR4/MD-2. *Proceedings of the National Academy of Sciences*, 109(19):7421–7426, 2012.
- [410] Daniel J. Fitzgerald, Philipp Berger, Christiane Schaffitzel, Kazuhiro Yamada, Timothy J. Richmond, and Imre Berger. Protein complex expression by using multigene baculoviral vectors. *Nature Methods*, 3(12):1021–1032, 2006.
- [411] Régis P. Lemaitre, Aliona Bogdanova, Barbara Borgonovo, Jeffrey B. Woodruff, and David N. Drechsel. FlexiBAC: A versatile, open-source baculovirus vector system for protein expression, secretion, and proteolytic processing. *BMC Biotechnology*, 19(1):1–11, 2019.

- [412] Richard B. Hitchman, Robert D. Possee, Andrew T. Crombie, Adam Chambers, Kim Ho, Evangelia Siaterli, Olga Lissina, Heather Sternard, Robert Novy, Kathryn Loomis, Louise E. Bird, Raymond J. Owens, and Linda A. King. Genetic modification of a baculovirus vector for increased expression in insect cells. *Cell Biology and Toxicology*, 26(1):57–68, 2010.
- [413] Florian Weissmann, Georg Petzold, Ryan VanderLinden, Pim J. Huis In’t Veld, Nicholas G. Brown, Fabienne Lampert, Stefan Westermann, Holger Stark, Brenda A. Schulman, and Jan Michael Peters. BiGBac enables rapid gene assembly for the expression of large multisubunit protein complexes. *Proceedings of the National Academy of Sciences of the United States of America*, 113(19):E2564–E2569, 2016.
- [414] D.-F. Lee, C.-C. Chen, T.-A. Hsu, and J.-L. Juang. A Baculovirus Superinfection System: Efficient Vehicle for Gene Transfer into Drosophila S2 Cells. *Journal of Virology*, 74(24):11873–11880, 2000.
- [415] L. F. Carbonell, M. J. Klowden, and L. K. Miller. Baculovirus-mediated expression of bacterial genes in dipteran and mammalian cells. *Journal of Virology*, 56(1):153–160, 1985.
- [416] Ângela M. Moraes, Soraia A.C. Jorge, Renato M. Astray, Claudio A.T. Suazo, Camilo E. Calderón Riquelme, Elisabeth F.P. Augusto, Aldo Tonso, Marilena M. Pamboukian, Rosane A.M. Piccoli, Manuel F. Barral, and Carlos A. Pereira. Drosophila melanogaster S2 cells for expression of heterologous genes: From gene cloning to bioprocess development. *Biotechnology Advances*, 30(3):613–628, 2012.
- [417] I. Schneider. Cell lines derived from late embryonic stages of Drosophila melanogaster. *Journal of Embryology and Experimental Morphology*, 27(2):353–365, 1972.
- [418] Marilena Martins Pamboukian, Soraia Athie Calil Jorge, Mariza Gerdulo Santos, Adriana Yurie Yokomizo, Carlos Augusto Pereira, and Aldo Tonso. Insect cells respiratory activity in bioreactor. *Cytotechnology*, 57(1):37–44, 2008.
- [419] Willem Adriaan de Jongh, Sancha Salgueiro, and Charlotte Dyring. The use of Drosophila S2 cells in R&D and bioprocessing. *Pharmaceutical Bioprocessing*, 1(2):197–213, 6 2013.

- [420] Jan Zitzmann, Gundula Sprick, Tobias Weidner, Christine Schreiber, and Peter Czermak. Process Optimization for Recombinant Protein Expression in Insect Cells. In *New Insights into Cell Culture Technology*. InTech, 5 2017.
- [421] Gianni Vandenborre, Guy Smagghe, Bart Ghesquière, Gerben Menschaert, Rameshwaram Nagender Rao, Kris Gevaert, and Els J.M. van Damme. Diversity in protein glycosylation among insect species. *PLoS ONE*, 6(2), 2011.
- [422] Daniel G. Gibson, Lei Young, Ray Yuan Chuang, J. Craig Venter, Clyde A. Hutchison, and Hamilton O. Smith. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*, 6(5):343–345, 2009.
- [423] Brian A Rabe and Constance Cepko. A Simple Enhancement for Gibson Isothermal Assembly. *bioRxiv*, 2020.
- [424] Sean Munro and Hugh R.B. Pelham. An hsp70-like protein in the ER: Identity with the 78 kd glucose-regulated protein and immunoglobulin heavy chain binding protein. *Cell*, 46(2):291–300, 1986.
- [425] Yuki Ohmuro-Matsuyama and Hideki Yamaji. Modifications of a signal sequence for antibody secretion from insect cells. *Cytotechnology*, 70(3):891–898, 2018.
- [426] Hiroyuki Sonoda, Yoichi Kumada, Tomohisa Katsuda, and Hideki Yamaji. Production of single-chain Fv-Fc fusion protein in stably transformed insect cells. *Biochemical Engineering Journal*, 67:77–82, 2012.
- [427] Michelle E. Kimple, Allison L. Brill, and Renee L. Pasker. Overview of affinity tags for protein purification. *Current Protocols in Protein Science*, 9.9.1-9.9.:Supp. 73, 2013.
- [428] Soban Tufail, Moazur Rhaman, Osman Mirza, Raheem Ullah, Mazhar Iqbal, Majid Ali Shah, Fouzia Ismat, and Muhammad Imran. Activity of the Human Rhinovirus 3C Protease Studied in Various Buffers, Additives and Detergents Solutions for Recombinant Protein Production. *Plos One*, 11(4):e0153436, 2016.
- [429] David S. Waugh. An overview of enzymatic reagents for the removal of affinity tags. *Protein Expression and Purification*, 80(2):283–293, 2011.

- [430] Péter Ecsédi, Bence Kiss, Gergő Gógl, László Radnai, László Buday, Kitti Koprivanacz, Károly Liliom, Ibolya Leveles, Beáta Vértessy, Norbert Jeszenői, Csaba Hetényi, Gitta Schlosser, Gergely Katona, and László Nyitray. Regulation of the Equilibrium between Closed and Open Conformations of Annexin A2 by N-Terminal Phosphorylation and S100A4-Binding. *Structure*, 25(8):1195–1207, 2017.
- [431] Vishal Kamat and Ashique Rafique. Designing binding kinetic assay on the bio-layer interferometry (BLI) biosensor to characterize antibody-antigen interactions. *Analytical Biochemistry*, 536:16–31, 2017.
- [432] Rejane L. Petersen. Strategies using bio-layer interferometry biosensor technology for vaccine research and development. *Biosensors*, 7(4), 2017.
- [433] Yao Yu, Scott Mitchell, Heather Lynaugh, Michael Brown, R. Paul Nobrega, Xiaoyong Zhi, Tingwan Sun, Isabelle Caffry, Yuan Cao, Rong Yang, Irina Burnina, Yingda Xu, and Patricia Estep. Understanding fortebio’s sensors for high-throughput kinetic and epitope screening for purified antibodies and yeast culture supernatant. *Journal of Biomolecular Screening*, 21(1):88–95, 2016.
- [434] Suneale Banerji, Anthony J. Day, Jan D. Kahmann, and David G. Jackson. Characterization of a functional hyaluronan-binding domain from the human CD44 molecule expressed in Escherichia coil. *Protein Expression and Purification*, 14(3):371–381, 1998.
- [435] Volker Gerke and Stephen E. Moss. Annexins: From structure to function. *Physiological Reviews*, 82(2):331–371, 2002.
- [436] Siwen Wang, Rui Song, Ziyi Wang, Zhaocheng Jing, Shaoxiong Wang, and Jian Ma. S100A8/A9 in inflammation. *Frontiers in Immunology*, 9(JUN), 2018.
- [437] Jean Da Silva Correia and Richard J. Ulevitch. MD-2 and TLR4 N-linked glycosylations are important for a functional lipopolysaccharide receptor. *Journal of Biological Chemistry*, 277(3):1845–1854, 2002.
- [438] Dounia Djeghloul, Bhavik Patel, Holger Kramer, Andrew Dimond, Chad Whilding, Karen Brown, Anne Céline Kohler, Amelie Feytout, Nicolas Veland, James Elliott, Tanmay A.M. Bharat, Abul K. Tarafder, Jan Löwe, Bee L. Ng,

- Ya Guo, Jacky Guy, Miles K. Huseyin, Robert J. Klose, Matthias Merken-
schlager, and Amanda G. Fisher. Identifying proteins bound to native mitotic
ESC chromosomes reveals chromatin repressors are important for compaction.
Nature Communications, 11(1), 2020.
- [439] John J Kelly, Dale Tranter, Els Pardon, Gamma Chi, Holger Kramer, Kelly M
Knee, M Janz, Jan Steyaert, Christine Bulawa, Ville O Paavilainen, Juha T
Huiskonen, and Yue. Snapshots of actin and tubulin folding inside the TRiC
chaperonin. *bioArxiv*, 2021.
- [440] Jürgen Cox and Matthias Mann. MaxQuant enables high peptide identification
rates, individualized p.p.b.-range mass accuracies and proteome-wide protein
quantification. *Nature Biotechnology*, 26(12):1367–1372, 2008.
- [441] Dattatreya Mellacheruvu, Zachary Wright, Amber L. Couzens, Jean Philippe
Lambert, Nicole A. St-Denis, Tuo Li, Yana V. Miteva, Simon Hauri, Mi-
haela E. Sardi, Teck Yew Low, Vincentius A. Halim, Richard D. Bagshaw,
Nina C. Hubner, Abdallah Al-Hakim, Annie Bouchard, Denis Faubert,
Damian Fermin, Wade H. Dunham, Marilyn Goudreau, Zhen Yuan Lin,
Beatriz Gonzalez Badillo, Tony Pawson, Daniel Durocher, Benoit Coulombe,
Ruedi Aebersold, Giulio Superti-Furga, Jacques Colinge, Albert J.R. Heck,
Hyungwon Choi, Matthias Gstaiger, Shabaz Mohammed, Ileana M. Cristea,
Keiryn L. Bennett, Mike P. Washburn, Brian Raught, Rob M. Ewing,
Anne Claude Gingras, and Alexey I. Nesvizhskii. The CRAPome: A con-
taminant repository for affinity purification-mass spectrometry data. *Nature
Methods*, 10(8):730–736, 2013.
- [442] Hyungwon Choi, Brett Larsen, Zhen Yuan Lin, Ashton Breitkreutz, Dat-
tatreya Mellacheruvu, Damian Fermin, Zhaohui S. Qin, Mike Tyers,
Anne Claude Gingras, and Alexey I. Nesvizhskii. SAINT: Probabilistic scoring
of affinity purification-mass spectrometry data. *Nature Methods*, 8(1):70–73,
2011.
- [443] Hannes Hahne, Fiona Pachl, Benjamin Ruprecht, Stefan K. Maier, Susan
Klaeger, Dominic Helm, Guillaume Médard, Matthias Wilm, Simone Lemeer,
and Bernhard Kuster. DMSO enhances electrospray response, boosting sensi-
tivity of proteomic experiments. *Nature Methods*, 10(10):989–991, 2013.
- [444] Damaris Bausch-Fluck, Emanuela Sara Milani, and Bernd Wollscheid. Sur-
faceome nanoscale organization and extracellular interaction networks. *Cur-
rent Opinion in Chemical Biology*, 48:26–33, 2019.

- [445] Bernd Wollscheid, Damaris Bausch-Fluck, Christine Henderson, Robert O'Brien, Miriam Bibel, Ralph Schiess, Ruedi Aebersold, and Julian D. Watts. Mass-spectrometric identification and relative quantification of N-linked cell surface glycoproteins. *Nature Biotechnology*, 27(4):378–386, 2009.
- [446] Damaris Bausch-Fluck, Andreas Hofmann, Thomas Bock, Andreas P. Frei, Ferdinando Cerciello, Andrea Jacobs, Hansjoerg Moest, Ulrich Omasits, Rebekah L. Gundry, Charles Yoon, Ralph Schiess, Alexander Schmidt, Paulina Mirkowska, Anetta Härtlová, Jennifer E. Van Eyk, Jean Pierre Bourquin, Ruedi Aebersold, Kenneth R. Boheler, Peter Zandstra, and Bernd Wollscheid. A mass spectrometric-derived cell surface protein atlas. *PLoS ONE*, 10(4):1–22, 2015.
- [447] Andreas P. Frei, Ock Youm Jeon, Samuel Kilcher, Hansjoerg Moest, Lisa M. Henning, Christian Jost, Andreas Plückthun, Jason Mercer, Ruedi Aebersold, Erick M. Carreira, and Bernd Wollscheid. Direct identification of ligand-receptor interactions on living cells and tissues. *Nature Biotechnology*, 30(10):997–1001, 2012.
- [448] Andreas P. Frei, Hansjoerg Moest, Karel Novy, and Bernd Wollscheid. Ligand-based receptor identification on living cells and tissues using TRICEPS. *Nature Protocols*, 8(7):1321–1336, 2013.
- [449] Tammy Lynn Tremblay and Jennifer J. Hill. Biotin-Transfer from a trifunctional crosslinker for identification of cell surface receptors of soluble protein ligands. *Scientific Reports*, 7(April):1–9, 2017.
- [450] Julie K. Freed, Julia R. Smith, Peigang Li, and Andrew S. Greene. Isolation of signal transduction complexes using biotin and crosslinking methodologies. *Proteomics*, 7(14):2371–2374, 2007.
- [451] Adam Byron, Janet A. Askari, Jonathan D. Humphries, Guillaume Jacquemet, Ewa J. Koper, Stacey Warwood, Colin K. Choi, Matthew J. Stroud, Christopher S. Chen, David Knight, and Martin J. Humphries. A proteomic approach reveals integrin activation state-dependent control of microtubule cortical targeting. *Nature Communications*, 6, 2015.
- [452] Thomas Köcher and Giulio Superti-Furga. Mass spectrometry-based functional proteomics: From molecular machines to protein networks. *Nature Methods*, 4(10):807–815, 2007.

- [453] Michael Fairhead and Mark Howarth. Site-Specific Biotinylation of Purified Proteins Using BirA. In Arnaud Gautier and Marlon J. Hünnger, editors, *Site-Specific Protein Labeling: Methods and Protocols*, chapter 12, pages 171–184. Humana Press, New York, 2015.
- [454] Mathias Uhlén, Linn Fagerberg, Bjö M. Hallström, Cecilia Lindskog, Per Oksvold, Adil Mardinoglu, Åsa Sivertsson, Caroline Kampf, Evelina Sjöstedt, Anna Asplund, Ing Marie Olsson, Karolina Edlund, Emma Lundberg, Sanjay Navani, Cristina Al Khalili Szigyanto, Jacob Odeberg, Dijana Djureinovic, Jenny Ottosson Takanen, Sophia Hober, Tove Alm, Per Henrik Edqvist, Holger Berling, Hanna Tegel, Jan Mulder, Johan Rockberg, Peter Nilsson, Jochen M. Schwenk, Marica Hamsten, Kalle Von Feilitzen, Mattias Forsberg, Lukas Persson, Fredric Johansson, Martin Zwahlen, Gunnar Von Heijne, Jens Nielsen, and Fredrik Pontén. Tissue-based map of the human proteome. *Science*, 347(6220), 2015.
- [455] Kelvin F. Cho, Tess C. Branon, Namrata D. Udeshi, Samuel A. Myers, Steven A. Carr, and Alice Y. Ting. Proximity labeling in mammalian cells with TurboID and split-TurboID. *Nature Protocols*, pages 1–29, 2020.
- [456] Megan R. Chastney, Craig Lawless, and Martin J. Humphries. Multiplexed Proximity Biotinylation Coupled to Mass Spectrometry for Defining Integrin Adhesion Complexes. *Current Protocols in Cell Biology*, 88(1):e113, 2020.
- [457] Kumar Yugandhar, Shagun Gupta, and Haiyuan Yu. Inferring Protein-Protein Interaction Networks From Mass Spectrometry-Based Proteomic Approaches: A Mini-Review. *Computational and Structural Biotechnology Journal*, 17:805–811, 2019.
- [458] Rose Oughtred, Jennifer Rust, Christie Chang, Bobby Joe Breitkreutz, Chris Stark, Andrew Willems, Lorrie Boucher, Genie Leung, Nadine Kolas, Frederick Zhang, Sonam Dolma, Jasmin Coulombe-Huntington, Andrew Chatr-aryamontri, Kara Dolinski, and Mike Tyers. The BioGRID database: A comprehensive biomedical resource of curated protein, genetic, and chemical interactions. *Protein Science*, 30(1):187–200, 2021.
- [459] Ronen Zaidel-Bar, Shalev Itzkovitz, Avi Ma’ayan, Ravi Iyengar, and Benjamin Geiger. Functional atlas of the integrin adhesome. *Nature Cell Biology*, 9(8):858–867, 2007.

- [460] Sabina E. Winograd-Katz, Reinhard Fässler, Benjamin Geiger, and Kyle R. Legate. The integrin adhesome: From genes and proteins to human disease. *Nature Reviews Molecular Cell Biology*, 15(4):273–288, 2014.
- [461] Paul Shannon, Andrew Markiel, Owen Ozier, S. Nitin Baliga, Jonathan T. Want, Daniel Gamage, Nada Amin, Benno Schwikowski, and Trey Ideker. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research*, 13(11):2498–2504, 11 2003.
- [462] Paul D. Thomas, Michael J. Campbell, Anish Kejariwal, Huaiyu Mi, Brian Karlak, Robin Daverman, Karen Diemer, Anushya Muruganujan, and Apurva Narechania. PANTHER: A library of protein families and subfamilies indexed by function. *Genome Research*, 13(9):2129–2141, 2003.
- [463] Huaiyu Mi, Anushya Muruganujan, and Paul D. Thomas. PANTHER in 2013: Modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Research*, 41(D1):377–386, 2013.
- [464] Bijay Jassal, Lisa Matthews, Guilherme Viteri, Chuqiao Gong, Pascual Lorente, Antonio Fabregat, Konstantinos Sidiropoulos, Justin Cook, Marc Gillespie, Robin Haw, Fred Loney, Bruce May, Marija Milacic, Karen Rothfels, Cristoffer Sevilla, Veronica Shamovsky, Solomon Shorser, Thawfeek Varusai, Joel Weiser, Guanming Wu, Lincoln Stein, Henning Hermjakob, and Peter D’Eustachio. The reactome pathway knowledgebase. *Nucleic Acids Research*, 48(D1):D498–D503, 2020.
- [465] Chaofeng Han, Jing Jin, Sheng Xu, Haibo Liu, Nan Li, and Xuetao Cao. Integrin CD11b negatively regulates TLR-triggered inflammatory responses by activating Syk and promoting degradation of MyD88 and TRIF via Cbl-b. *Nature Immunology*, 11(8):734–742, 2010.
- [466] Yoshikazu Takada, Xiaojing Ye, and Scott Simon. The integrins. *Genome Biology*, 8(5), 2007.
- [467] Tobias G. Kapp, Florian Rechenmacher, Stefanie Neubauer, Oleg V. Maltsev, Elisabetta A. Cavalcanti-Adam, Revital Zarka, Ute Reuning, Johannes Notni, Hans Jürgen Wester, Carlos Mas-Moruno, Joachim Spatz, Benjamin Geiger, and Horst Kessler. A comprehensive evaluation of the activity and selectivity profile of ligands for RGD-binding integrins. *Scientific Reports*, 7:1–13, 2017.

- [468] Kenji Yokoyama, Xi Ping Zhang, Leonid Medved, and Yoshikazu Takada. Specific binding of integrin $\alpha v\beta 3$ to the fibrinogen γ and $\alpha(E)$ chain C-terminal domains. *Biochemistry*, 38(18):5872–5877, 1999.
- [469] C. C. Hsu, W. J. Chuang, C. H. Chang, Y. L. Tseng, H. C. Peng, and T. F. Huang. Improvements in endotoxemic syndromes using a disintegrin, rhodostomin, through integrin $\alpha v\beta 3$ -dependent pathway. *Journal of Thrombosis and Haemostasis*, 9(3):593–602, 2011.
- [470] Zhixia Chen, Shuang Wang, Yihui Chen, Zhenzhen Shao, Zhuang Yu, Shuya Mei, and Quan Li. Integrin $\beta 3$ Modulates TLR4-Mediated Inflammation by Regulation of CD14 Expression in Macrophages in Septic Condition. *Shock*, 53(3):335–343, 2020.
- [471] Alexander S. Antonov, Galina N. Antonova, David H. Munn, Nahid Mivechi, Rudolf Lucas, John D. Catravas, and Alexander D. Verin. $\alpha V\beta 3$ integrin regulates macrophage inflammatory responses via PI3 kinase/Akt-dependent NF- κ B activation. *Journal of Cellular Physiology*, 226(2):469–476, 2011.
- [472] Tatiana Gianni, Valerio Leoni, Liudmila S. Chesnokova, Lindsey M. Hutt-Fletcher, and Gabriella Campadelli-Fiume. $\alpha v\beta 3$ -Integrin Is a Major Sensor and Activator of Innate Immunity To Herpes Simplex Virus-1. *Proceedings of the National Academy of Sciences of the United States of America*, 109(48):19792–19797, 2012.
- [473] Gisa Gerold, Khalid Abu Ajaj, Michael Bienert, Hans Jürgen Laws, Arturo Zychlinsky, and Juana L. de Diego. A Toll-like receptor 2-integrin $\beta 3$ complex senses bacterial lipopeptides via vitronectin. *Nature Immunology*, 9(7):761–768, 2008.
- [474] Jenson Lim and Neil A. Hotchin. Signalling mechanisms of the leukocyte integrin $\alpha M\beta 2$: Current and future perspectives. *Biology of the Cell*, 104(11):631–640, 2012.
- [475] Tatiana P. Ugarova and Valentin P. Yakubenko. Recognition of fibrinogen by leukocyte integrins. *Annals of the New York Academy of Sciences*, 936:368–385, 2001.
- [476] Prerna C. Patel and Rene E. Harrison. Membrane Ruffles Capture C3bi-opsonized Particles in Activated Macrophages. *Molecular Biology of the Cell*, 19(11):4628–4639, 11 2008.

- [477] Guang Sheng Ling, Jason Bennett, Kevin J. Woollard, Marta Szajna, Liliane Fossati-Jimack, Philip R. Taylor, Diane Scott, Guido Franzoso, H. Terence Cook, and Marina Botto. Integrin CD11b positively regulates TLR4-induced signalling pathways in dendritic cells but not in macrophages. *Nature Communications*, 5:1–12, 2014.
- [478] Olachi J. Mezu-Ndubuisi and Akhil Maheshwari. The role of integrins in inflammation and angiogenesis. *Pediatric Research*, pages 1–8, 2020.
- [479] Dennis Wolf, Nathaly Anto-Michel, Hermann Blankenbach, Ansgar Wiedemann, Konrad Buscher, Jan David Hohmann, Bock Lim, Marina Bäuml, Alex Marki, Maximilian Mauler, Daniel Duerschmied, Zhichao Fan, Holger Winkels, Daniel Sidler, Philipp Diehl, Dirk M. Zajonc, Ingo Hilgendorf, Peter Stachon, Timoteo Marchini, Florian Willecke, Maximilian Schell, Björn Sommer, Constantin Von Zur Muhlen, Jochen Reinöhl, Teresa Gerhardt, Edward F. Plow, Valentin Yakubenko, Peter Libby, Christoph Bode, Klaus Ley, Karlheinz Peter, and Andreas Zirlik. A ligand-specific blockade of the integrin Mac-1 selectively targets pathologic inflammation while maintaining protective host-defense. *Nature Communications*, 9(1), 2018.
- [480] Daniel J. Loegering and Michelle R. Lennartz. Protein kinase C and toll-like receptor signaling. *Enzyme Research*, 2011(1), 2011.
- [481] Miho Kubo-Murai, Kaoru Hazeki, Naoe Sukenobu, Kyoko Yoshikawa, Kiyomi Nigorikawa, Kazumi Inoue, Toshiyoshi Yamamoto, Misako Matsumoto, Tsukasa Seya, Norimitsu Inoue, and Osamu Hazeki. Protein kinase C δ binds TIRAP/Mal to participate in TLR signaling. *Molecular Immunology*, 44(9):2257–2264, 2007.
- [482] Yubin Li, Mousa Komai-Koma, Derek S. Gilchrist, Daniel K. Hsu, Fu-Tong Liu, Tabitha Springall, and Damo Xu. Galectin-3 Is a Negative Regulator of Lipopolysaccharide-Mediated Inflammation. *The Journal of Immunology*, 181(4):2781–2789, 2008.
- [483] Uri Arad, Noa Madar-Balakirski, Avital Angel-Korman, Sharon Amir, Sharon Tzadok, Ortal Segal, Aharon Menachem, Aviram Gold, Ori Elkayam, and Dan Caspi. Galectin-3 is a sensor-regulator of toll-like receptor pathways in synovial fibroblasts. *Cytokine*, 73(1):30–35, 2015.

- [484] Miguel Angel Burguillos, Martina Svensson, Tim Schulte, Antonio Boza-Serrano, Albert Garcia-Quintanilla, Edel Kavanagh, Martiniano Santiago, Nikenza Viceconte, Maria Jose Oliva-Martin, Ahmed Mohamed Osman, Emma Salomonsson, Lahouari Amar, Annette Persson, Klas Blomgren, Adnane Achour, Elisabet Englund, Hakon Leffler, Jose Luis Venero, Bertrand Joseph, and Tomas Deierborg. Microglia-secreted galectin-3 acts as a toll-like receptor 4 ligand and contributes to microglial activation. *Cell Reports*, 10(9):1626–1638, 2015.
- [485] Jacques Neefjes, Marlieke L.M. Jongsma, Petra Paul, and Oddmund Bakke. Towards a systems understanding of MHC class I and MHC class II antigen presentation. *Nature Reviews Immunology*, 11(12):823–836, 2011.
- [486] Xingguang Liu, Zhenzhen Zhan, Dong Li, Li Xu, Feng Ma, Peng Zhang, Hangping Yao, and Xuetao Cao. Intracellular MHC class II molecules promote TLR-triggered innate immune responses by maintaining activation of the kinase Btk. *Nature Immunology*, 12(5):416–424, 2011.
- [487] Jeong Hoon Kim, Catherine K. Yang, Kyu Heo, Robert G. Roeder, Woojin An, and Michael R. Stallcup. CCAR1, a Key Regulator of Mediator Complex Recruitment to Nuclear Receptor Transcription Complexes. *Molecular Cell*, 31(4):510–519, 2008.
- [488] April R. Suriano, Amy N. Sanford, Nahmah Kim, Miae Oh, Sarah Kennedy, Mark J. Henderson, Kelly Dietzmann, and Kathleen E. Sullivan. GCF2/LRRFIP1 Represses Tumor Necrosis Factor Alpha Expression. *Molecular and Cellular Biology*, 25(20):9073–9081, 2005.
- [489] Andre L. Reed, Hitoshi Yamazaki, Joshua D. Kaufman, Yaffa Rubinstein, Barbara Murphy, and Alfred C. Johnson. Molecular cloning and characterization of a transcription regulator with homology to GC-binding factor. *Journal of Biological Chemistry*, 273(34):21594–21602, 1998.
- [490] Penggao Dai, Sun Yong Jeong, Yanbao Yu, Taohua Leng, Weidong Wu, Ling Xie, and Xian Chen. Modulation of TLR Signaling by Multiple MyD88-Interacting Partners Including Leucine-Rich Repeat Fli-I-Interacting Proteins. *The Journal of Immunology*, 182(6):3450–3460, 2009.

- [491] Pengyuan Yang, Huazhang An, Xingguang Liu, Mingyue Wen, Yuanyuan Zheng, Yaocheng Rui, and Xuetao Cao. The cytosolic nucleic acid sensor LR-RFIP1 mediates the production of type I interferon via a B-catenin-dependent pathway. *Nature Immunology*, 11(6):487–494, 2010.
- [492] Kenya Honda, Akinori Takaoka, and Tadatsugu Taniguchi. Type I Interferon Gene Induction by the Interferon Regulatory Factor Family of Transcription Factors. *Immunity*, 25(3):349–360, 2006.
- [493] Yasuaki Aratani. Myeloperoxidase: Its role for host defense, inflammation, and neutrophil function. *Archives of Biochemistry and Biophysics*, 640(January):47–52, 2018.
- [494] E. Malle, P. G. Furtmüller, W. Sattler, and C. Obinger. Myeloperoxidase: A target for new drug development? *British Journal of Pharmacology*, 152(6):838–854, 2007.
- [495] Valentina Loria, Ilaria Dato, Francesca Graziani, and Luigi M. Biasucci. Myeloperoxidase: A new biomarker of inflammation in ischemic heart disease and acute coronary syndromes. *Mediators of Inflammation*, 2008, 2008.
- [496] Dragana Odobasic, A. Richard Kitching, and Stephen R. Holdsworth. Neutrophil-mediated regulation of innate and adaptive immunity: The role of myeloperoxidase. *Journal of Immunology Research*, 2016, 2016.
- [497] Driss El Kebir, Levente József, Wanling Pan, and János G. Filep. Myeloperoxidase delays neutrophil apoptosis through CD11b/CD18 integrins and prolongs inflammation. *Circulation Research*, 103(4):352–359, 2008.
- [498] Denise Lau, Hanke Mollnau, Jason P. Eiserich, Bruce A. Freeman, Andreas Daiber, Ursula M. Gehling, Jens Brümmer, Volker Rudolph, Thomas Münzel, Thomas Heitzer, Thomas Meinertz, and Stephan Baldus. Myeloperoxidase mediates neutrophil activation by association with CD11b/CD18 integrins. *Proceedings of the National Academy of Sciences of the United States of America*, 102(2):431–436, 2005.
- [499] K. A. Bixby, M. H. Nanao, N. V. Shen, A. Kreusch, H. Bellamy, P. J. Pfaffinger, and S. Choe. Zn²⁺-binding and molecular determinants of tetramerization in voltage-gated K⁺ channels. *Nature Structural Biology*, 6(1):38–43, 1999.

- [500] Jochen Schwenk, Michaela Metz, Gerd Zolles, Rostislav Turecek, Thorsten Fritzius, Wolfgang Bildl, Etsuko Tarusawa, Akos Kulik, Andreas Unger, Klara Ivankova, Riad Seddik, Jim Y. Tiao, Mathieu Rajalu, Johana Trojanova, Volker Rohde, Martin Gassmann, Uwe Schulte, Bernd Fakler, and Bernhard Bettler. Native GABAB receptors are heteromultimers with a family of auxiliary subunits. *Nature*, 465(7295):231–235, 2010.
- [501] Y. Zhong, J. Yang, W. W. Xu, Y. Wang, C. C. Zheng, B. Li, and Q. Y. He. KCTD12 promotes tumorigenesis by facilitating CDC25B/CDK1/Aurora A-dependent G2/M transition. *Oncogene*, 36(44):6177–6189, 2017.
- [502] Leonard J. Banaszak and Wasantha K. Ranatunga. The assembly of apoB-containing lipoproteins: A structural biology point of view. *Annals of Medicine*, 40(4):253–267, 2008.
- [503] Matthew L. Brown, M. P. Ramprasad, Patrick K. Umeda, Akira Tanaka, Yasushi Kobayashi, Teruo Watanabe, Hiroaki Shimoyamada, Wen Lin Kuo, Ran Li, Ruiling Song, William A. Bradley, and Sandra H. Gianturco. A macrophage receptor for apolipoprotein B48: Cloning, expression, and atherosclerosis. *Proceedings of the National Academy of Sciences of the United States of America*, 97(13):7488–7493, 2000.
- [504] Akio Kawakami, Mariko Tani, Tsuyoshi Chiba, Katsumasa Yui, Shohei Shinozaki, Katsuyuki Nakajima, Akira Tanaka, Kentaro Shimokado, and Masayuki Yoshida. Pitavastatin inhibits remnant lipoprotein-induced macrophage foam cell formation through apoB48 receptor-dependent mechanism. *Arteriosclerosis, Thrombosis, and Vascular Biology*, 25(2):424–429, 2005.
- [505] Gwang Woong Go and Arya Mani. Low-density lipoprotein receptor (LDLR) family orchestrates cholesterol homeostasis. *Yale Journal of Biology and Medicine*, 85(1):19–28, 2012.
- [506] Yury I. Miller, Soo Ho Choi, Longhou Fang, and Richard Harkewicz. Toll-Like Receptor-4 and Lipoprotein Accumulation in Macrophages. *Trends in Cardiovascular Medicine*, 19(7):227–232, 2009.
- [507] J. Herz and D. K. Strickland. LRP: A multifunctional scavenger and signaling receptor. *Journal of Clinical Investigation*, 108(6):779–784, 2001.

- [508] Takahisa Kanekiyo and Guojun Bu. The low-density lipoprotein receptor-related protein 1 and amyloid- β clearance in Alzheimer's disease. *Frontiers in Aging Neuroscience*, 6:1–12, 2014.
- [509] Nicola Potere, Marco Giuseppe Del Buono, Adolfo Gabriele Mauro, Antonio Abbate, and Stefano Toldo. Low Density Lipoprotein Receptor-Related Protein-1 in Cardiac Inflammation and Infarct Healing. *Frontiers in Cardiovascular Medicine*, 6:1–18, 2019.
- [510] Jennifer N. Rauch, Gabriel Luna, Elmer Guzman, Morgane Audouard, Collin Challis, Youssef E. Sibih, Carolina Leshuk, Israel Hernandez, Susanne Wegmann, Bradley T. Hyman, Viviana Gradinaru, Martin Kampmann, and Kenneth S. Kosik. LRP1 is a master regulator of tau uptake and spread. *Nature*, 580(7803):381–385, 2020.
- [511] Jonathan L Schmid-Burgk, Klara Höning, Thomas S Ebert, and Veit Hornung. CRISPaint allows modular base-specific gene tagging using a ligase-4-dependent mechanism. *Nature Communications*, 7(12338):1–12, 2016.
- [512] John R. Gilbert and Jeffrey M. Vance. Isolation of Genomic DNA from Mammalian Cells. *Current Protocols in Human Genetics*, 19(1):1–6, 1998.
- [513] Andreas Untergasser, Ioana Cutcutache, Triinu Koressaar, Jian Ye, Brant C. Faircloth, Mairo Remm, and Steven G. Rozen. Primer3-new capabilities and interfaces. *Nucleic Acids Research*, 40(15):1–12, 2012.
- [514] Daichi Kamiyama, Sayaka Sekine, Benjamin Barsi-Rhyne, Jeffrey Hu, Baohui Chen, Luke A. Gilbert, Hiroaki Ishikawa, Manuel D. Leonetti, Wallace F. Marshall, Jonathan S. Weissman, and Bo Huang. Versatile protein tagging in cells with split fluorescent protein. *Nature Communications*, 7(11046):1–9, 2016.
- [515] Andrew V Anzalone, Luke W Koblan, and David R Liu. Genome editing with CRISPR-Cas nucleases, base editors, transposases and prime editors. *Nature Biotechnology*, 38(7):824–844, 7 2020.
- [516] F. Ann Ran, Patrick D. Hsu, Jason Wright, Vineeta Agarwala, David A. Scott, and Feng Zhang. Genome engineering using the CRISPR-Cas9 system. *Nature Protocols*, 8(11):2281–2308, 2013.

- [517] Adrian Pickar-Oliver and Charles A. Gersbach. The next generation of CRISPR–Cas technologies and applications. *Nature Reviews Molecular Cell Biology*, 2019.
- [518] Haifeng Wang, Marie La Russa, and Lei S. Qi. CRISPR/Cas9 in Genome Editing and beyond. *Annual Review of Biochemistry*, 85:227–264, 2016.
- [519] Mazhar Adli. The CRISPR tool kit for genome editing and beyond. *Nature Communications*, 9(1911):1–13, 2018.
- [520] Fuguo Jiang and Jennifer A Doudna. CRISPR–Cas9 Structures and Mechanisms. *Annual Review of Biophysics*, 46(1):505–529, 5 2017.
- [521] Vivien Marx. Guide RNAs: it’s good to be choosy. *Nature Methods*.
- [522] Shannon M Miller, Tina Wang, Peyton B Randolph, Mandana Arbab, Max W Shen, Tony P Huang, Zaneta Matuszek, Gregory A Newby, Holly A Rees, and David R Liu. Continuous evolution of SpCas9 variants compatible with non-G PAMs. *Nature Biotechnology*, 38(4):471–481, 4 2020.
- [523] Daniel Agudelo, Alexis Durringer, Lusiné Bozoyan, Caroline C. Huard, Sophie Carter, Jeremy Loehr, Dafni Synodinou, Mathieu Drouin, Jayme Salsman, Graham Dellaire, José Laganière, and Yannick Doyon. Marker-free coselection for CRISPR-driven genome editing in human cells. *Nature Methods*, 14(6):615–620, 2017.
- [524] Fabian Ocegüera-Yanez, Shin Il Kim, Tomoko Matsumoto, Ghee Wan Tan, Long Xiang, Takeshi Hatani, Takayuki Kondo, Makoto Ikeya, Yoshinori Yoshida, Haruhisa Inoue, and Knut Woltjen. Engineering the AAVS1 locus for consistent and scalable transgene expression in human iPSCs and their differentiated derivatives. *Methods*, 101:43–55, 2016.
- [525] Mathieu Dalvai, Jeremy Loehr, Karine Jacquet, Caroline C. Huard, Céline Roques, Pauline Herst, Jacques Côté, and Yannick Doyon. A Scalable Genome-Editing-Based Approach for Mapping Multiprotein Complexes in Human Cells. *Cell Reports*, 13(3):621–633, 2015.
- [526] Y. Doyon and J. Côté. Preparation and Analysis of Native Chromatin-Modifying Complexes. *Methods in Enzymology*, 573:303–318, 2016.

- [527] Alexandre Paix, Andrew Folkmann, Daniel H. Goldman, Heather Kulaga, Michael J. Grzelak, Dominique Rasoloson, Supriya Paidemarry, Rachel Green, Randall R. Reed, and Geraldine Seydoux. Precision genome editing using synthesis-dependent repair of Cas9-induced DNA breaks. *Proceedings of the National Academy of Sciences of the United States of America*, 114(50):E10745–E10754, 2017.
- [528] Oguz Kanca, Jonathan Zirin, Jorge Garcia-Marques, Shannon Marie Knight, Donghui Yang-Zhou, Gabriel Amador, Hyunglok Chung, Zhongyuan Zuo, Liwen Ma, Yuchun He, Wen-Wen Lin, Ying Fang, Ming Ge, Shinya Yamamoto, Karen L. Schulze, Yanhui Hu, Allan C. Spradling, Stephanie E. Mohr, Norbert Perrimon, and Hugo J. Bellen. An efficient CRISPR-based strategy to insert small and large fragments of DNA using short homology arms. *eLife*, 8:1–22, 11 2019.
- [529] Paul T. Manna, Luther J. Davis, and Margaret S. Robinson. Fast and cloning-free CRISPR/Cas9-mediated genomic editing in mammalian cells. *Traffic*, 20(12):974–982, 2019.
- [530] St  phanie Cabantous, Thomas C. Terwilliger, and Geoffrey S. Waldo. Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein. *Nature Biotechnology*, 23(1):102–107, 2005.
- [531] Chiara Foglieni, St  phanie Papin, Agnese Salvad  , Tariq Afroz, Sandra Pinton, Giona Pedrioli, Giorgio Ulrich, Magdalini Polymenidou, and Paolo Paganetti. Split GFP technologies to structurally characterize and quantify functional biomolecular interactions of FTD-related proteins. *Scientific Reports*, 7(1):1–15, 2017.
- [532] Amanda M. Lanza, Do Soon Kim, and Hal S. Alper. Evaluating the influence of selection markers on obtaining selected pools and stable cell lines in human cells. *Biotechnology Journal*, 8(7):811–821, 2013.
- [533] Ian M. Slaymaker, Linyi Gao, Bernd Zetsche, David A. Scott, Winston X. Yan, and Feng Zhang. Rationally engineered Cas9 nucleases with improved specificity. *Science*, 351(6268):84–88, 2016.
- [534] Georgyi V Los, Lance P Encell, Mark G. McDougall, Danette D Hartzell, Natasha Karassina, Chad Zimprich, Monika G. Wood, Randy Learish, Rachel Friedman Ohana, Marjeta Urh, Dan Simpson, Jacqui Mendez, Kris Zimmerman, Paul Otto, Gediminas Vidugiris, Ji Zhu, Aldis Darzins, Dieter H.

- Klaubert, Robert F. Bulleit, and Keith V. Wood. HaloTag: A Novel Protein Labeling Technology for Cell Imaging and Protein Analysis. *ACS Chemical Biology*, 3(6):373–382, 6 2008.
- [535] Christopher G. England, Emily B. Ehlerding, and Weibo Cai. NanoLuc: A Small Luciferase Is Brightening Up the Field of Bioluminescence. *Bioconjugate Chemistry*, 27(5):1175–1187, 2016.
- [536] Michael Haugwitz, Tatiana Garachtchenko, Omar Nourzaie, Suvarna Gandlur, and Hiroaki Sagawa. Rapid, on-demand protein stabilization and destabilization using the ProteoTuner™ systems. *Nature Methods*, 5(10):iii–iv, 2008.
- [537] Bijan Zakeri, Jacob O. Fierer, Emrah Celik, Emily C. Chittock, Ulrich Schwarz-Linek, Vincent T. Moy, and Mark Howarth. Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin. *Proceedings of the National Academy of Sciences of the United States of America*, 109(12), 2012.
- [538] Irsyad N.A. Khairil Anuar, Anusuya Banerjee, Anthony H. Keeble, Alberto Carella, Georgi I. Nikov, and Mark Howarth. Spy&Go purification of SpyTag-proteins using pseudo-SpyCatcher to access an oligomerization toolbox. *Nature Communications*, 10(1):1–13, 2019.
- [539] Natasha C. Dale, Elizabeth K.M. Johnstone, Carl W. White, and Kevin D.G. Pfleger. NanoBRET: The bright future of proximity-based assays. *Frontiers in Bioengineering and Biotechnology*, 7:1–13, 2019.
- [540] John G. Doench, Ella Hartenian, Daniel B. Graham, Zuzana Tothova, Mudra Hegde, Ian Smith, Meagan Sullender, Benjamin L. Ebert, Ramnik J. Xavier, and David E. Root. Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nature Biotechnology*, 32(12):1262–1267, 2014.
- [541] Kendall R. Sanson, Ruth E. Hanna, Mudra Hegde, Katherine F. Donovan, Christine Strand, Meagan E. Sullender, Emma W. Vaimberg, Amy Goodale, David E. Root, Federica Piccioni, and John G. Doench. Optimized libraries for CRISPR-Cas9 genetic screens with multiple modalities. *Nature Communications*, 9(1):1–15, 2018.
- [542] Léna Vouillot, Aurore Thélie, and Nicolas Pollet. Comparison of T7E1 and surveyor mismatch cleavage assays to detect mutations triggered by engineered nucleases. *G3: Genes, Genomes, Genetics*, 5(3):407–415, 2015.

- [543] Noriyuki Yanaka. Mammalian glycerophosphodiester phosphodiesterases. *Bio-science, Biotechnology and Biochemistry*, 71(8):1811–1818, 2007.
- [544] Bin Zheng, Christopher P. Berrie, Daniela Corda, and Marilyn G. Farquhar. GDE1/MIR16 is a glycerophosphoinositol phosphodiesterase regulated by stimulation of G protein-coupled receptors. *Proceedings of the National Academy of Sciences of the United States of America*, 100(4):1745–1750, 2003.
- [545] Nathan H Cho, Keith C Cheveralls, Andreas-david Brunner, Kibeom Kim, André C Michaelis, Preethi Raghavan, Hirofumi Kobayashi, Laura Savy, Jason Y Li, Hera Canaj, James Y S Kim, M Edna, Christian Gnann, Frank Mccarthy, Joana P Cabrera, Rachel M Brunetti, Bryant B Chhun, Marco Y Hein, Bo Huang, Shalin B Mehta, Jonathan S Weissman, and Rafael Gómez-sjöberg. OpenCell : proteome-scale endogenous tagging enables the cartography of human cellular organization. *bioArxiv*, 2021.
- [546] Roujian Lu, Xiang Zhao, Juan Li, Peihua Niu, Bo Yang, Honglong Wu, Wenling Wang, Hao Song, Baoying Huang, Na Zhu, Yuhai Bi, Xuejun Ma, Faxian Zhan, Liang Wang, Tao Hu, Hong Zhou, Zhenhong Hu, Weimin Zhou, Li Zhao, Jing Chen, Yao Meng, Ji Wang, Yang Lin, Jianying Yuan, Zhihao Xie, Jinmin Ma, William J. Liu, Dayan Wang, Wenbo Xu, Edward C. Holmes, George F. Gao, Guizhen Wu, Weijun Chen, Weifeng Shi, and Wenjie Tan. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *The Lancet*, 395(10224):565–574, 2020.
- [547] Wanbo Tai, Lei He, Xiujuan Zhang, Jing Pu, Denis Voronin, Shibo Jiang, Yusen Zhou, and Lanying Du. Characterization of the receptor-binding domain (RBD) of 2019 novel coronavirus: implication for development of RBD protein as a viral attachment inhibitor and vaccine. *Cellular & Molecular Immunology*, 17(6):613–620, 6 2020.
- [548] Alexandra C. Walls, Young Jun Park, M. Alejandra Tortorici, Abigail Wall, Andrew T. McGuire, and David Veasley. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell*, 181(2):281–292, 2020.
- [549] Markus Hoffmann, Hannah Kleine-Weber, Simon Schroeder, Nadine Krüger, Tanja Herrler, Sandra Erichsen, Tobias S. Schiergens, Georg Herrler, Nai-Huei Wu, Andreas Nitsche, Marcel A. Müller, Christian Drosten, and Stefan Pöhlmann. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell*, pages 1–10, 2020.

- [550] Alexandra C. Walls, M. Alejandra Tortorici, Joost Snijder, Xiaoli Xiong, Berend Jan Bosch, Felix A. Rey, and David Veeler. Tectonic conformational changes of a coronavirus spike glycoprotein promote membrane fusion. *Proceedings of the National Academy of Sciences of the United States of America*, 114(42):11157–11162, 2017.
- [551] Xiaoyi Fan, Duanfang Cao, Lingfei Kong, and Xinzhen Zhang. Cryo-EM analysis of the post-fusion structure of the SARS-CoV spike glycoprotein. *Nature Communications*, 11(1):1–10, 2020.
- [552] Yaara Finkel, Orel Mizrahi, Aharon Nachshon, Shira Weingarten-Gabbay, David Morgenstern, Yfat Yahalom-Ronen, Hadas Tamir, Hagit Achdout, Dana Stein, Ofir Israeli, Adi Beth-Din, Sharon Melamed, Shay Weiss, Tomer Israely, Nir Paran, Michal Schwartz, and Noam Stern-Ginossar. The coding capacity of SARS-CoV-2. *Nature*, 589(7840):125–130, 2021.
- [553] Yuan Huang, Chan Yang, Xin feng Xu, Wei Xu, and Shu wen Liu. Structural and functional properties of SARS-CoV-2 spike protein: potential antiviral drug development for COVID-19. *Acta Pharmacologica Sinica*, 41(9):1141–1149, 2020.
- [554] Simran Preet Kaur and Vandana Gupta. COVID-19 Vaccine: A comprehensive status report. *Virus Research*, 288(July):198114, 2020.
- [555] Daniel Wrapp, Nianshuang Wang, Kizzmekia S. Corbett, Jory A. Goldsmith, Ching Lin Hsieh, Olubukola Abiona, Barney S. Graham, and Jason S. McLellan. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science*, 367(6483):1260–1263, 2020.
- [556] Anwarul Hasan, Bilal Ahamad Paray, Arif Hussain, Fikry Ali Qadir, Farnoosh Attar, Falah Mohammad Aziz, Majid Sharifi, Hossein Derakhshankhah, Behnam Rasti, Masoumeh Mehrabi, Koorosh Shahpasand, Ali Akbar Saboury, and Mojtaba Falahati. A review on the cleavage priming of the spike protein on coronavirus by angiotensin-converting enzyme-2 and furin. *Journal of Biomolecular Structure and Dynamics*, 0(0):1–9, 2020.
- [557] Emma. K. Grant, David J. Fallon, Michael M. Hann, Ken G. M. Fantom, Chad Quinn, Francesca Zappacosta, Roland S. Annan, Chun-wa Chung, Paul Bamborough, David P. Dixon, Peter Stacey, David House, Vipul Kumar K. Patel, Nicholas C. O. Tomkinson, and Jacob Bush. PhotoAffinity

Bits: A photoaffinity-based fragment screening platform for efficient identification of protein ligands. *Angewandte Chemie International Edition*, page anie.202008361, 8 2020.

- [558] Yasunori Watanabe, Joel D. Allen, Daniel Wrapp, Jason S. McLellan, and Max Crispin. Site-specific glycan analysis of the SARS-CoV-2 spike. *Science*, 369(6501):330–333, 2020.
- [559] Fatima Amanat, Daniel Stadlbauer, Shirin Strohmeier, Thi H.O. Nguyen, Veronika Chromikova, Meagan McMahon, Kaijun Jiang, Guha Asthagiri Arunkumar, Denise Jurczyszak, Jose Polanco, Maria Bermudez-Gonzalez, Giulio Kleiner, Teresa Aydillo, Lisa Miorin, Daniel S. Fierer, Luz Amarilis Lugo, Erna Milunka Kojic, Jonathan Stoeber, Sean T.H. Liu, Charlotte Cunningham-Rundles, Philip L. Felgner, Thomas Moran, Adolfo García-Sastre, Daniel Caplivski, Allen C. Cheng, Katherine Kedzierska, Olli Vapalahti, Jussi M. Hepojoki, Viviana Simon, and Florian Krammer. A serological assay to detect SARS-CoV-2 seroconversion in humans. *Nature Medicine*, 26(7):1033–1036, 2020.
- [560] Yannick Galipeau, Matthew Greig, George Liu, Matt Driedger, and Marc André Langlois. Humoral Responses and Serological Assays in SARS-CoV-2 Infections. *Frontiers in Immunology*, 11(December):1–19, 2020.
- [561] Pedro M. Folegatti, Katie J. Ewer, Parvinder K. Aley, Brian Angus, Stephan Becker, Sandra Belij-Rammerstorfer, Duncan Bellamy, Sagida Bibi, Mustapha Bittaye, Elizabeth A. Clutterbuck, Christina Dold, Saul N. Faust, Adam Finn, Amy L. Flaxman, Bassam Hallis, Paul Heath, Daniel Jenkin, Rajeka Lazarus, Rebecca Makinson, Angela M. Minassian, Katrina M. Pollock, Maheshi Ramasamy, Hannah Robinson, Matthew Snape, Richard Tarrant, Merryn Voysey, Catherine Green, Alexander D. Douglas, Adrian V.S. Hill, Teresa Lambe, Sarah C. Gilbert, Andrew J. Pollard, Jeremy Aboagye, Kelly Adams, Aabidah Ali, Elizabeth Allen, Jennifer L. Allison, Rachel Anslow, Edward H. Arbe-Barnes, Gavin Babbage, Kenneth Baillie, Megan Baker, Natalie Baker, Philip Baker, Ioana Baleanu, Juliana Ballaminut, Eleanor Barnes, Jordan Barrett, Louise Bates, Alexander Batten, Kirsten Beadon, Rebecca Beckley, Eleanor Berrie, Lisa Berry, Amy Beveridge, Kevin R. Bewley, Else Margreet Bijker, Tracey Bingham, Luke Blackwell, Caitlin L. Blundell, Emma Bolam, Elena Boland, Nicola Borthwick, Thomas Bower, Amy Boyd, Tanja Brenner, Philip D. Bright, Charlie Brown-O’Sullivan, Emily

Brunt, Jamie Burbage, Sharon Burge, Karen R. Buttigieg, Nicholas Byard, Ingrid Cabera Puig, Anna Calvert, Susana Camara, Michelangelo Cao, Federica Cappuccini, Melanie Carr, Miles W. Carroll, Victoria Carter, Katrina Cathie, Ruth J. Challis, Sue Charlton, Irina Chelysheva, Jee Sun Cho, Paola Cicconi, Liliana Cifuentes, Helen Clark, Elizabeth Clark, Tom Cole, Rachel Colin-Jones, Christopher P. Conlon, Aislinn Cook, Naomi S. Coombes, Rachel Cooper, Catherine A. Cosgrove, Karen Coy, Wendy E.M. Crocker, Christina J. Cunningham, Brad E. Damratoski, Lynne Dando, Mehreen S. Dato, Hannah Davies, Hans De Graaf, Tesfaye Demissie, Claudio Di Maso, Isabelle Dietrich, Tao Dong, Francesca R. Donnellan, Naomi Douglas, Charlotte Downing, Jonathan Drake, Rachael Drake-Brockman, Ruth Elizabeth Drury, Susanna Jane Dunachie, Nick J. Edwards, Frances D.L. Edwards, Chris J. Edwards, Sean C. Elias, Michael J. Elmore, Katherine R.W. Emary, Marcus Rex English, Susanne Fagerbrink, Sally Felle, Shuo Feng, Samantha Field, Carine Fixmer, Clare Fletcher, Karen J. Ford, Jamie Fowler, Polly Fox, Emma Francis, John Frater, Julie Furze, Michelle Fuskova, Eva Galiza, Diane Gbesemete, Ciaran Gilbride, Kerry Godwin, Giacomo Gorini, Lyndsey Goulston, Caroline Grabau, Lara Gracie, Zoe Gray, Lucy Belle Guthrie, Mark Hackett, Sandro Halwe, Elizabeth Hamilton, Joseph Hamlyn, Brama Hanumunthadu, Irasha Harding, Stephanie A. Harris, Andrew Harris, Daisy Harrison, Clare Harrison, Thomas C. Hart, Louise Haskell, Sophia Hawkins, Ian Head, John Aaron Henry, Jennifer Hill, Susanne H.C. Hodgson, Mimi M. Hou, Elizabeth Howe, Nicola Howell, Cecilia Hutlin, Sabina Ikram, Catherine Isitt, Poppy Iveson, Susan Jackson, Frederic Jackson, Sir William James, Megan Jenkins, Elizabeth Jones, Kathryn Jones, Christine E. Jones, Bryony Jones, Reshma Kailath, Konstantinos Karampatsas, Jade Keen, Sarah Kelly, Dearbhla Kelly, David Kerr, Simon Kerridge, Liaquat Khan, Uzma Khan, Annabel Killen, Jasmin Kinch, Thomas B. King, Lloyd King, Jade King, Lucy Kingham-Page, Paul Klenerman, Francesca Knapper, Julian C. Knight, Daniel Knott, Stanislava Koleva, Alexandra Kupke, Colin W. Larkworthy, Jessica P.J. Larwood, Anna Laskey, Alison M. Lawrie, Arlene Lee, Kim Yee Ngan Lee, Emily A. Lees, Helen Legge, Alice Lelliott, Nana Marie Lemm, Amelia M. Lias, Aline Linder, Samuel Lipworth, Xinxue Liu, Shuchang Liu, Raquel Lopez Ramon, May Lwin, Francesca Mabesa, Meera Madhavan, Garry Mallett, Kushal Mansatta, Ines Marcal, Spyridoula Marinou, Emma Marlow, Julia L. Marshall, Jane Martin, Joanne McEwan, Lorna McInroy, Gretchen Meddaugh, Alexander J. Mentzer, Neginsadat Mirtorabi, Maria Moore, Edward Moran, Ella Morey,

Victoria Morgan, Susan Jane Morris, Hazel Morrison, Gertraud Morshead, Richard Morter, Yama F. Mujadidi, Jilly Muller, Tatiana Munera-Huertas, Claire Munro, Alasdair Munro, Sarah Murphy, Vincent J. Munster, Philomena Mweu, Andrés Noé, Fay L. Nugent, Elizabeth Nuthall, Katie O'Brien, Daniel O'Connor, Blanché Oguti, Jennifer L. Oliver, Catarina Oliveira, Peter John O'Reilly, Mairead Osborn, Piper Osborne, Cathy Owen, Daniel Owens, Nelly Owino, Mihaela Pacurar, Kaye Parker, Helena Parracho, Maia Patrick-Smith, Victoria Payne, Jennifer Pearce, Yanchun Peng, Marco Polo Peralta Alvarez, James Perring, Katja Pfafferott, Dimitra Pipini, Emma Plested, Helen Pluess-Hall, Katrina Pollock, Ian Poulton, Laura Presland, Samuel Provstgaard-Morys, David Pulido, Kajal Radia, Fernando Ramos Lopez, Jade Rand, Helen Ratcliffe, Thomas Rawlinson, Sarah Rhead, Amy Riddell, Adam John Ritchie, Hannah Roberts, Joanna Robson, Sophie Roche, Cornelius Rohde, Christine S. Rollier, Rossana Romani, Indra Rudiansyah, Stephen Saich, Sara Sajjad, Stephannie Salvador, Lidia Sanchez Riera, Helen Sanders, Katherine Sanders, Shari Sapaun, Chloe Sayce, Ella Schofield, Gavin Screaton, Beatrice Selby, Calum Semple, Hannah R. Sharpe, Imam Shaik, Adam Shea, Holly Shelton, Sarah Silk, Laura Silva-Reyes, Donal T. Skelly, Heather Smee, Catherine C. Smith, David J. Smith, Rinn Song, Alexandra J. Spencer, Elizabeth Stafford, Amy Steele, Elena Stefanova, Lisa Stockdale, Anna Szigeti, Abdessamad Tahiri-Alaoui, Moira Tait, Helen Talbot, Rachel Tanner, Iona Jennifer Taylor, Victoria Taylor, Rebecca Te Water Naude, Nazia Thakur, Yrene Themistocleous, Andreas Themistocleous, Merin Thomas, Tonia M. Thomas, Amber Thompson, Samantha Thomson-Hill, Jennifer Tomlins, Susan Tonks, James Towner, Nguyen Tran, Julia A. Tree, Adam Truby, Kate Turkentine, Cheryl Turner, Nicola Turner, Sally Turner, Toby Tuthill, Marta Ulaszewska, Rachel Varughese, Neeltje Van Doremalen, Kristin Veighey, Marije K. Verheul, Iason Vichos, Elia Vitale, Laura Walker, Marion E.E. Watson, Benjamin Welham, Julie Wheat, Caroline White, Rachel White, Andrew T. Worth, Danny Wright, Suzie Wright, Xin Li Yao, and Yasmine Yau. Safety and immunogenicity of the ChAdOx1 nCoV-19 vaccine against SARS-CoV-2: a preliminary report of a phase 1/2, single-blind, randomised controlled trial. *The Lancet*, 396(10249):467–478, 2020.

- [562] Andrea Tešija Kuna, Milena Hanžek, Ines Vukasović, Nora Nikolac Gabaj, Valentina Vidranski, Ivana Čelap, Marijana Miler, Nevenka Stančin, Brankica Šimac, Marcela Živković, Marko Žarak, Marta Kmet, Marijana Jovanović, Sanja Tadinac, Sandra Šupraha Goreta, Josipa Periša, Ivan Šamija, and Mario

Štefanović. Comparison of diagnostic accuracy for eight SARS-CoV-2 serological assays. *Biochemia Medica*, 31(1):1–13, 2021.

- [563] Jeffrey D. Whitman, Joseph Hiatt, Cody T. Mowery, Brian R. Shy, Ruby Yu, Tori N. Yamamoto, Ujjwal Rathore, Gregory M. Goldgof, Caroline Whitty, Jonathan M. Woo, Antonia E. Gallman, Tyler E. Miller, Andrew G. Levine, David N. Nguyen, Sagar P. Bapat, Joanna Balcerek, Sophia A. Bylsma, Ana M. Lyons, Stacy Li, Allison Wai yi Wong, Eva Mae Gillis-Buck, Zachary B. Steinhart, Youjin Lee, Ryan Apathy, Mitchell J. Lipke, Jennifer Anne Smith, Tina Zheng, Ian C. Boothby, Erin Isaza, Jackie Chan, Dante D. Acenas, Jinwoo Lee, Trisha A. Macrae, Than S. Kyaw, David Wu, Dianna L. Ng, Wei Gu, Vanessa A. York, Haig Alexander Eskandarian, Perri C. Callaway, Lakshmi Warriar, Mary E. Moreno, Justine Levan, Leonel Torres, Lila A. Farrington, Rita P. Loudermilk, Kanishka Koshal, Kelsey C. Zorn, Wilfredo F. Garcia-Beltran, Diane Yang, Michael G. Astudillo, Bradley E. Bernstein, Jeffrey A. Gelfand, Edward T. Ryan, Richelle C. Charles, A. John Iafrate, Jochen K. Lennerz, Steve Miller, Charles Y. Chiu, Susan L. Stramer, Michael R. Wilson, Aashish Manglik, Chun Jimmie Ye, Nevan J. Krogan, Mark S. Anderson, Jason G. Cyster, Joel D. Ernst, Alan H.B. Wu, Kara L. Lynch, Caryn Bern, Patrick D. Hsu, and Alexander Marson. Evaluation of SARS-CoV-2 serology assays reveals a range of test performance. *Nature Biotechnology*, 38(10):1174–1183, 2020.
- [564] Mark Ainsworth, Monique Andersson, Kathryn Auckland, J Kenneth Bailie, Eleanor Barnes, Sally Beer, Amy Beveridge, Sagida Bibi, Luke Blackwell, Martyna Borak, Abbie Bown, Tim Brooks, Nicola A Burgess-Brown, Susana Camara, Matthew Catton, Kevin K. Chau, Thomas Christott, Elizabeth Clutterbuck, Jesse Coker, Richard J Cornall, Stuart Cox, David Crawford-Jones, Derrick W Crook, Silvia D’Arcangelo, Wanwisa Dejnirattasai, Julie M M Dequaire, Stavros Dimitriadis, Kate E Dingle, George Doherty, Christina Dold, Tao Dong, Susanna J Dunachie, Daniel Ebner, Marc Emmenegger, Alexis Espinosa, David W Eyre, Rory Fairhead, Shayan Fassih, Conor Feehily, Sally Felle, Alejandra Fernandez-Cid, Maria Fernandez Mendoza, Thomas H Ford, Thomas Fordwoh, Deborah Fox McKee, John Frater, Veronica Gallardo Sanchez, Nick Gent, Dominique Georgiou, Christopher J Groves, Bassam Hallis, Peter M Hammond, Stephanie B. Hatch, Heli J Harvala, Jennifer Hill, Sarah J Hoosdally, Bryn Horsington, Alison Howarth, Tim James, Katie Jeffery, Elizabeth Jones, Anita Justice, Fredrik Karpe, James Kavanagh, David S

Kim, Richard Kirton, Paul Klenerman, Julian C Knight, Leonidas Koukoulis, Andrew Kwok, Ullrich Leuschner, Robert Levin, Aline Linder, Teresa Lockett, Sheila F Lumley, Spyridoula Marinou, Brian D Marsden, Jose Martinez, Lucas Martins Ferreira, Lara Mason, Philippa C Matthews, Alexander J Mentzer, Alexander Mobbs, Juthathip Mongkolsapaya, Jordan Morrow, Shubhashish M M Mukhopadhyay, Matthew J Neville, Sarah Oakley, Marta Oliveira, Ashley Otter, Kevin Paddon, Jordan Pascoe, Yanchun Peng, Elena Perez, Prem K Perumal, Timothy E A Peto, Hayleah Pickford, Rutger J Ploeg, Andrew J Pollard, Anastasia Richardson, Thomas G Ritter, David J Roberts, Gillian Rodger, Christine S Rollier, Cathy Rowe, Justine K Rudkin, Gavin Sreaton, Malcolm G Semple, Alex Sienkiewicz, Laura Silva-Reyes, Donal T Skelly, Alberto Sobrino Diaz, Lizzie Stafford, Lisa Stockdale, Nicole Stoesser, Teresa Street, David I Stuart, Angela Sweed, Adan Taylor, Hannah Thraves, Hoi P Tsang, Marije K Verheul, Richard Vipond, Timothy M Walker, Susan Wareing, Yolanda Warren, Charlie Wells, Clare Wilson, Kate Withycombe, and Rebecca K Young. Performance characteristics of five immunoassays for SARS-CoV-2: a head-to-head benchmark comparison. *The Lancet Infectious Diseases*, 20(12):1390–1400, 12 2020.

- [565] Jesper Pallesen, Nianshuang Wang, Kizzmekia S. Corbett, Daniel Wrapp, Robert N. Kirchdoerfer, Hannah L. Turner, Christopher A. Cottrell, Michelle M. Becker, Lingshu Wang, Wei Shi, Wing Pui Kong, Erica L. Andres, Arminja N. Kettenbach, Mark R. Denison, James D. Chappell, Barney S. Graham, Andrew B. Ward, and Jason S. McLellan. Immunogenicity and structures of a rationally designed prefusion MERS-CoV spike antigen. *Proceedings of the National Academy of Sciences of the United States of America*, 114(35):E7348–E7357, 2017.
- [566] Ching-lin Hsieh, Jory A Goldsmith, Jeffrey M Schaub, Andrea M. DiVenere, Hung-che Kuo, Kamyab Javanmardi, Kevin C Le, Daniel Wrapp, Alison G Lee, Yutong Liu, Chia-wei Chou, Patrick O Byrne, Christy K Hjorth, Nicole V Johnson, John Ludes-Meyers, Annalee W Nguyen, Juyeon Park, Nianshuang Wang, Dzifa Amengor, Jason J Lavinder, Gregory C Ippolito, Jennifer A Maynard, Ilya J Finkelstein, and Jason S. McLellan. Structure-based design of prefusion-stabilized SARS-CoV-2 spikes. *Science*, 369(6510):1501–1505, 9 2020.
- [567] Katerina Papanikolopoulou, Vincent Forge, Pierrette Goeltz, and Anna Mittraki. Formation of Highly Stable Chimeric Trimers by Fusion of an Adenovirus

Fiber Shaft Fragment with the Foldon Domain of Bacteriophage T4 Fibrin. *Journal of Biological Chemistry*, 279(10):8991–8998, 2004.

- [568] Florian Krammer, Irina Margine, Gene S. Tan, Natalie Pica, Jens C. Krause, and Peter Palese. A carboxy-terminal trimerization domain stabilizes conformational epitopes on the stalk domain of soluble recombinant Hemagglutinin substrates. *PLoS ONE*, 7(8), 2012.
- [569] Daniel Stadlbauer, Fatima Amanat, Veronika Chromikova, Kaijun Jiang, Shirin Strohmeier, Guha Asthagiri Arunkumar, Jessica Tan, Disha Bhavsar, Christina Capuano, Ericka Kirkpatrick, Philip Meade, Ruhi Nichalle Brito, Catherine Teo, Meagan McMahon, Viviana Simon, and Florian Krammer. SARS-CoV-2 Seroconversion in Humans: A Detailed Protocol for a Serological Assay, Antigen Production, and Test Setup. *Current Protocols in Microbiology*, 57(1):1–15, 2020.
- [570] Otmane Boussif, Frank LezoualC’H, Maria Antonietta Zanta, Mojgan Djavaheri Mergny, Daniel Scherman, Barbara Demeneix, and Jean Paul Behr. A versatile vector for gene and oligonucleotide transfer into cells in culture and in vivo: Polyethylenimine. *Proceedings of the National Academy of Sciences of the United States of America*, 92(16):7297–7301, 1995.
- [571] Chi Yen Lin, Zhen Huang, Wei Wen, Andrew Wu, Congzhou Wang, and Li Niu. Enhancing protein expression in HEK-293 cells by lowering culture temperature. *PLoS ONE*, 10(4):1–19, 2015.
- [572] Annamarie C. Dalton and William A. Barton. Over-expression of secreted proteins from mammalian cell lines. *Protein Science*, 23(5):517–525, 2014.
- [573] Jean Marc Bielser, Moritz Wolf, Jonathan Souquet, Hervé Broly, and Massimo Morbidelli. Perfusion mammalian cell culture for recombinant protein manufacturing – A critical review. *Biotechnology Advances*, 36(4):1328–1340, 2018.
- [574] Grant Stuchbury and Gerald Münch. Optimizing the generation of stable neuronal cell lines via pre-transfection restriction enzyme digestion of plasmid DNA. *Cytotechnology*, 62(3):189–194, 2010.
- [575] Nick S. Berrow, David Alderton, Sarah Sainsbury, Joanne Nettleship, Rene Asensberg, Nahid Rahman, David I. Stuart, and Raymond J. Owens. A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. *Nucleic Acids Research*, 35(6), 2007.

- [576] R. Chalk, W. Greenland, T. Moreira, J. Coker, S. M.M. Mukhopadhyay, E. Williams, C. Manning, T. Bohstedt, R. McCrorie, A. Fernandez-Cid, and N. A. Burgess-Brown. Identification, Mapping and Relative Quantitation of SARS-CoV-2 Spike Glycopeptides by Mass-Retention Time Fingerprinting. *bioRxiv*, 2020.
- [577] Shahanshah Khan, Mahnoush S Shafiei, Christopher Longoria, John Schoggins, and C Rashmin. SARS-CoV-2 spike protein induces inflammation via TLR2-dependent activation of the NF- κ B pathway. *bioRxiv*, 2021.
- [578] Sebastian Theobald, Alexander Simonis, Christoph Kreer, Matthias Zehner, Julia Fischer, Marie-Christine Albert, Jakob Malin, Jessica Gräb, Sandra Winter, Ute Sandaradura de Silva, Boris Böll, Philipp Köhler, Henning Gruell, Isabelle Suárez, Michael Hallek, Gerd Fätkenheuer, Norma Jung, Oliver Cornely, Clara Lehmann, Hamid Kashkar, Florian Klein, and Jan Rybníček. The SARS-CoV-2 spike protein primes inflammasome-mediated interleukin-1- β secretion in COVID-19 patient-derived macrophages. *Research Square*, pages 1–14, 2020.
- [579] Chao Gao, Junwei Zeng, Nan Jia, Kathrin Stavenhagen, Yasuyuki Matsumoto, Hua Zhang, Jiang Li, Adam J. Hume, Elke Mühlberger, Irma van Die, Julian Kwan, Kelan Tantisira, Andrew Emili, and Richard D. Cummings. SARS-CoV-2 spike protein interacts with multiple innate immune receptors. *bioRxiv*, 2020.
- [580] Joseph J. Brennan and Thomas D. Gilmore. Evolutionary origins of toll-like receptor signaling. *Molecular Biology and Evolution*, 35(7):1576–1587, 2018.
- [581] Martha Triantafyllou and Kathy Triantafyllou. Lipopolysaccharide recognition : CD14 , TLRs and the LPS-activation cluster. *Trends in Immunology*, 23(6):301–304, 2002.
- [582] Alexandra Pfeiffer, Alfred Böttcher, Evelyn Ors, Michael Kapinsky, Peter Nagy, Andrea Bodnér, Ingo Spreitzer, Gerhard Liebisch, Wolfgang Drobnik, Klaus Gempel, Markus Horn, Stefan Holmer, Thomas Hartung, Gabriele Multhoff, Gerhard Scharf, Hansgeorg Schindler, Artur J. Ulmer, Holger Heine, Felix Stelter, Christine Schütt, Gregor Rothe, János Szili, Sándor Damjanovich, and Gerd Schmitz. Lipopolysaccharide and ceramide docking to CD14 provokes ligand-specific receptor clustering in rafts. *European Journal of Immunology*, 31(11):3153–3164, 2001.

- [583] Kathy Triantafilou, Martha Triantafilou, and Russell L Dedrick. A CD14-independent LPS receptor cluster. *Nature Immunology*, 2(4):338–345, 4 2001.
- [584] Jean Marie Ruyschaert and Caroline Loney. Role of lipid microdomains in TLR-mediated signalling. *Biochimica et Biophysica Acta - Biomembranes*, 1848(9):1860–1867, 2015.
- [585] Agnieszka Płóciennikowska, Aneta Hromada-Judycka, Kinga Borzęcka, and Katarzyna Kwiatkowska. Co-operation of TLR4 and raft proteins in LPS-induced pro-inflammatory signaling. *Cellular and Molecular Life Sciences*, 72(3):557–581, 2015.